

Stochastik in den Ingenieurwissenschaften

Eine Einführung mit R

(Wahrscheinlichkeitsrechnung und Statistik in den
Ingenieurwissenschaften)

Prof. Dr. Christine Müller
Technische Universität Dortmund

Wintersemester 2011/2012

Inhaltsverzeichnis

1	Fragestellungen	7
I	Beschreibende Statistik	15
2	Erste Schritte mit R	17
2.1	Herunterladen der freien Software R	17
2.2	Grafische Darstellungen mit R	17
3	Merkmale und Datentypen	21
3.1	Merkmale	21
3.2	Datentypen	22
4	Tabellarische und grafische Darstellungen von univariaten Daten	25
4.1	Qualitative Daten	25
4.2	Quantitativ diskrete Daten	29
4.3	Quantitative stetige Daten	32
5	Statistische Kennzahlen für die Lage	37
5.1	Nominale Daten	37
5.2	Ordinale Daten	38
5.3	Quantitative Daten	40
6	Statistische Kennzahlen für die Streuung	43
7	Bivariate Daten	49

7.1	Tabellarische und grafische Darstellungen	49
7.2	Zusammenhangsmaße (Korrelationsmaße)	52
7.3	Zusammenhangsmaße für nominale Daten	53
7.4	Zusammenhangsmaße für ordinale Daten	55
7.5	Zusammenhangsmaße für quantitative Daten	57
7.6	Lineare Regression	59
7.7	+ Nichtlineare Regression +	62
II	Wahrscheinlichkeitstheorie	63
8	Simulationen mit R	65
8.1	Zufallszahlen und deren grafische Darstellung	65
9	Statistische Kennzahlen für Zufallszahlen	67
9.1	Beschreibung des Zufalls	70
10	Mengentheoretische Grundlagen	75
11	Wahrscheinlichkeitsmaße, Wahrscheinlichkeitsräume	79
11.1	Diskrete Wahrscheinlichkeitsräume	80
11.2	Laplace-Räume	82
12	Zufallsvariablen und deren Verteilungen	85
12.1	Verteilung eindimensionaler Zufallsvariablen	86
12.2	Verteilung mehrdimensionaler Zufallsvariablen	90
13	Wichtige Wahrscheinlichkeitsverteilungen	93
13.1	Diskrete Wahrscheinlichkeitsverteilungen	93
13.2	Stetige Wahrscheinlichkeitsverteilungen	96
13.3	Erzeugung von Zufallszahlen mittels <code>runif</code>	102
14	Zwei- und mehrdimensionale Wahrscheinlichkeitsverteilungen	103
14.1	Zweidimensionale diskrete Wahrscheinlichkeitsverteilungen	103

14.2	Zweidimensionale Normalverteilung	103
14.3	Mehrdimensionale diskrete Verteilungen	108
14.4	Mehrdimensionale stetige Verteilungen	108
15	Bedingte Wahrscheinlichkeiten und stochastische Unabhängigkeit	109
15.1	Bedingte Wahrscheinlichkeiten	109
15.2	Stochastische Unabhängigkeit von Ereignissen	113
15.3	Unabhängigkeit von Zufallsvariablen	115
15.4	Verteilung von Funktionen von Zufallsvariablen	117
16	Erwartungswert und Varianz	119
16.1	Diskrete Wahrscheinlichkeitsverteilungen	119
16.2	Stetige Wahrscheinlichkeitsverteilungen	121
16.3	Eigenschaften des Erwartungswertes und der Varianz	122
16.4	Gesetz der großen Zahlen	125
16.5	Stochastische Simulation von Erwartungswert und Varianz	125
17	Weitere wahrscheinlichkeitstheoretische Kennzahlen	127
17.1	p -Quantile	127
17.2	Abhängigkeitsmaße	129
18	Markovketten	133
III	Schließende Statistik	143
19	Fragestellungen der Schließenden Statistik	145
20	Allgemeine Form eines statistischen Testes	147
21	Tests für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ bei Normalverteilung	151
21.1	Herleitung des Testes	151
21.2	Fehlerwahrscheinlichkeiten 2. Art	153
21.3	$+$ Wahl des Stichprobenumfanges $+$	155

22 Weitere Tests bei Normalverteilung	159
22.1 Tests von Hypothesen über den Erwartungswert	159
22.2 Tests von Hypothesen über die Varianz	160
22.3 Zweistichproben-Tests für gepaarte Stichproben	161
22.4 Zweistichproben-Tests für ungepaarte Stichproben	162
22.5 ⁺ Relevanz- und Äquivalenzttests ⁺	165
22.6 Test auf linearen Zusammenhang zwischen gepaarten Stichproben	169
22.7 Tests auf Normalverteilung	170
23 Tests bei nichtnormalverteilten Daten	173
23.1 Einstichproben- und Zweistichproben-Tests	173
23.2 Tests auf Zusammenhang	174
24 Intervallschätzungen/Konfidenzbereiche *	177
24.1 Allgemeine Eigenschaften	177
24.2 Konfidenzintervalle für den Erwartungswert einer Normalverteilung	178
24.3 Konfidenzintervalle für die Varianz einer Normalverteilung	179
24.4 Konfidenzintervalle für die Standardabweichung bei Normalverteilung	179
24.5 Konfidenzintervall für die Differenz $d = \mu_1 - \mu_2$ bei unbekannter (gleicher) Varianz .	179
25 Abkürzungen	181

Wichtige Hinweise zur Vorlesung und zur Klausur

⁺ Kapitel oder Abschnitte markiert mit einem Pluszeichen ⁺ werden in der Vorlesung nicht behandelt und werden auch nicht in der Klausur abgefragt. Sie dienen nur der zusätzlichen Information.

* Kapitel oder Abschnitte markiert mit einem Stern * werden in der Vorlesung behandelt. Sie werden aber nicht in der Klausur abgefragt.

Kapitel 1

Fragestellungen

Die Stochastik ist die Lehre des Zufalls. Lange ist man davon ausgegangen, dass diese Lehre für die Ingenieurwissenschaft nicht von Bedeutung ist, da alles auf Mechanik und anderen deterministischen Prozessen basiert. Aber das ist nicht der Fall. Maschinen produzieren nur mit einer bestimmten Genauigkeit und Teile der Maschinen fallen zufällig aus. Baumaterialien sind nicht immer gleich sondern schwanken in ihrer Beschaffenheit. Es stellt sich auch die Frage, wie Maschinen und Bauwerke konstruiert werden sollen, so dass bestimmte Anforderungen möglichst gut erfüllt werden. Im Zeitalter der Globalisierung mit ihrem harten Konkurrenzdruck ist es besonders wichtig, dass die Anforderungen so gut wie möglich erfüllt werden.

1.0.1 Beispiel (Kugellager)

In einer großen industriellen Kooperation war das Ziel, Stahlkugellager (steel ball bearings) von einem Durchmesser von 1 Mikrometer zu produzieren. Am Ende eines Tages wurden 10 Kugellager zufällig von einer Produktionslinie ausgewählt und deren Durchmesser ermittelt. In einem anderen Experiment wurden 10 Kugellager von einer anderen Produktionslinie ausgewählt. Es ergaben sich folgende in der Datei `STEEL.DAT` enthaltenen Durchmesser

Erste Linie	1.18	1.42	0.69	0.88	1.62	1.09	1.53	1.02	1.19	1.32
Zweite Linie	1.72	1.62	1.69	0.79	1.79	0.77	1.44	1.29	1.96	0.99

Siehe (Romano 1977), entnommen aus (Hand et al. 1994).

Verschiedene Fragen können hier gestellt werden:

1. Wird die Zielgröße von 1 Mikrometer im Mittel von allen Kugellagern der beiden Produktionslinien eingehalten?
2. Wird eine Abweichung von 0.3 Mikrometer von der Zielgröße toleriert, so stellt sich die Frage ob diese Toleranzbreite von allen Kugellagern eingehalten wird bzw. welcher Anteil aller Kugellagern diese Toleranzbreite einhält.
3. Hält eine Produktionslinie die Zielvorgabe besser ein? Unterscheiden sich die Ergebnisse der beiden Produktionslinien im mittleren Wert bzw. in der Streuung?

1.0.2 Beispiel (Lebenszeiten von Glühbirnen)

Die folgende Tabelle, die in der Datei LAMPS.DAT enthalten ist, stellt die Häufigkeitstabelle der Lebenszeiten von 300 Glühbirnen dar.

Lebenszeit (in Stunden)	Absolute Häufigkeit
950-1000	2
1000-1050	2
1050-1100	3
1100-1150	6
1150-1200	7
1200-1250	12
1250-1300	16
1300-1350	20
1350-1400	24
1400-1450	27
1450-1500	29
1500-1550	29
1550-1600	28
1600-1650	25
1650-1700	21
1700-1750	16
1750-1800	12
1800-1850	8
1850-1900	6
1900-1950	3
1950-2000	2
2000-2050	1
2050-2100	1

Siehe (Gupta52), entnommen aus (Hand et al. 1994), S. 108.

Die sogenannte Normalverteilung beschreibt diese Lebenszeiten sehr gut, weshalb die Frage aufkam, ob das echte Daten sind. Normalerweise werden Lebenszeiten mit anderen Verteilungen beschrieben. Es stellt sich die Frage, ob eine typische Lebenszeitverteilung die Daten genauso gut beschreibt. Eine andere Fragestellung, die insbesondere Gupta interessierte hatte, ist, wie man die mittlere Lebenszeit schätzen würde, wenn man die die Glühbirnen nur maximal z.B. 1700 Stunden beobachtet hätte, d.h. wenn man nicht gewartet hätte, bis alle Glühbirnen kaputt gegangen sind. Solche Daten nennt man auch zensierte Daten.

1.0.3 Beispiel (Wirksamkeit von Rostschutzmitteln)

Zur Überprüfung der Wirksamkeit zweier Rostschutzmittel R_1 und R_2 wurden Proben genommen und die Ergebnisse den Kategorien *geringe*, *mittlere* und *starke* Wirksamkeit zugeordnet. Insgesamt ergab sich folgende Kontingenztafel:

	gering	mittel	stark
R_1	65	103	106
R_2	74	85	47

Gibt es einen Zusammenhang zwischen Rostschutzmittel und Rostanfälligkeit?

1.0.4 Beispiel (Hitzeentwicklung beim Zementansetzen)

Es sollte untersucht werden, wie die Zusammensetzung des Zements die Hitzeentwicklung beeinflusst. Neben der Hitzeentwicklung in Kalorien pro Gramm des Zements wurden die prozentuellen Anteile von

a = tricalcium aluminate

b = tricalcium silicate

c = tetracalcium aluminoferrite

d = dicalcium silicate

im Zement ermittelt und in der Datei `SETTING.DAT` aufgeführt.

a	b	c	d	Hitze
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

Siehe (Woods et al. 1932), entnommen aus (Hand et al. 1994).

Es stellt hier insbesondere die Frage, ob anhand einer beliebigen vorgegebenen Zusammensetzung des Zements die Hitzeentwicklung vorausgesagt werden kann.

1.0.5 Beispiel (Aufwickeln von Kammgarn)

Die folgenden Daten zeigen die Anzahl der Zyklen, bis Kammgarn falsch aufgewickelt wurde. Diese Anzahlen wurden unter folgenden experimentellen Bedingungen eines $3 \times 3 \times 3$ -faktoriellen Versuchsplanes ermittelt:

x_1 : Länge des Kammgarnes(250, 300, 350 mm)

x_2 : Amplitude der Zyklen (8, 9, 10 mm)

x_3 : Gewicht (40, 45, 50 g)

Dabei werden in der Datei WOOL.DAT die jeweils 3 Stufen der Faktoren Länge, Amplitude und Gewicht mit -1, 0 und +1 bezeichnet.

Faktorstufen			Anzahl der Zyklen
x_1	x_2	x_3	
-1	-1	-1	674
-1	-1	0	370
-1	-1	+1	292
-1	0	-1	338
-1	0	0	266
-1	0	+1	210
-1	+1	-1	170
-1	+1	0	118
-1	+1	+1	90
0	-1	-1	1414
0	-1	0	1198
0	-1	+1	634
0	0	-1	1022
0	0	0	620
0	0	+1	438
0	+1	-1	442
0	+1	0	332
0	+1	+1	220
+1	-1	-1	3636
+1	-1	0	3184
+1	-1	+1	2000
+1	0	-1	1568
+1	0	0	1070
+1	0	+1	566
+1	+1	-1	1140
+1	+1	0	884
+1	+1	+1	360

Siehe (1), entnommen aus (Hand et al. 1994).

Interessant ist hier bei welchen Einstellungen es am längsten dauert, bis das Garn schief aufgewickelt wird, und zwar nicht für den konkreten Datensatz sondern für zukünftiges Aufwickeln. Es stellt sich

auch die Frage, ob einer der Faktoren Länge, Amplitude und Gewicht überhaupt einen Einfluss auf die Anzahl der Zyklen hat, d.h. ob es bei einem der Faktoren egal ist, auf welcher Stufe er eingestellt ist.

1.0.6 Beispiel (Penicillin-Herstellung)

Vier Methoden A, B, C und D zur Herstellung von Penicillin wurden in einem sogenannten randomisierten Block-Experiment miteinander verglichen. Die Blöcke waren Mischungen, die genügend Material für vier Durchläufe enthielten. Die folgende Tabelle enthält vier Doppelspalten. Von den Doppelspalten enthält die erste Spalte jeweils den Ertrag und die zweite Spalte jeweils die Position im Block. Dabei gibt hier die Position im Block an, an welchem Durchlauf die Methode verwendet wurde. So bedeutet zum Beispiel die Position 3 bei Block 4 und Methode B, dass die Methode B im 3. Durchlauf beim Verwenden der 4. Mischung angewendet wurde. Die Datei `PNICILL.DAT` enthält die folgenden Ergebnisse.

Block	Methode							
	A		B		C		D	
1	89	1	88	3	97	2	94	4
2	84	4	77	2	92	3	79	1
3	81	2	87	1	87	4	85	3
4	87	1	92	3	89	2	84	4
5	79	3	81	4	80	1	88	2

Siehe (Box et al. 1978), entnommen aus (Hand et al. 1994), S. 347. Hier ist von Interesse, ob sich die 4 Methoden unterscheiden, und wenn ja, welche die beste ist. Auch hier bezieht sich die Frage nicht auf den konkreten Datensatz sondern generell auf die Herstellung.

1.0.7 Beispiel (Druckfestigkeit von Beton)

In Kassel wurde eine neue Herstellungsweise von besonders hartem Beton entwickelt. Um festzustellen, ob andere Institutionen auch in der Lage sind, diesen Beton herzustellen, wurde ein sogenannter Ringversuch gestartet, an dem sich mehrere Universitäten beteiligten. Da man annahm, dass die Schleifmethode und die Methode zum Prüfen der Druckfestigkeit auch von den Institutionen abhängen könnte, wurden das Schleifen und das Prüfen auch auf die beteiligten Universitäten verteilt. Ein Ausschnitt aus den Prüfergebnissen enthält die Datei `Druckfestigkeit.csv`, die folgende Spalten enthält:

	Druck	Festbetonrohrdichte	H	S	P
1	168.500	2.495	Aachen	Aachen	Kassel
2	167.100	2.516	Aachen	Aachen	Kassel
3	158.700	2.485	Aachen	Aachen	Kassel
4	150.500	2.485	Aachen	Kassel	Kassel
⋮	⋮	⋮	⋮	⋮	⋮
122	147.700	2.525	Kassel	Kassel	Kassel
123	145.300	2.495	Kassel	Kassel	Kassel
124	144.700	2.525	Kassel	Kassel	Kassel
125	142.100	2.502	Kassel	Kassel	Kassel

Dabei steht **Druck** für die Druckfestigkeit, **H** für den Herstellungsort, **S** für den Schleifort und **P** für den Prüfort. Die Datei enthält 125 Untersuchungen.

1.0.8 Beispiel

Ein Fahrzeugpark enthält 6 Fahrzeuge. Jedes Fahrzeug hat die Wahrscheinlichkeit $\frac{1}{5}$, dass es an einem Tag kaputt geht. Pro Tag kann aber nur ein Fahrzeug repariert werden. Wie groß ist die Wahrscheinlichkeit, dass alle Fahrzeuge kaputt sind? Oder mit wie vielen Tagen im Jahr muss gerechnet werden, an denen alle Fahrzeuge nicht benutzbar sind?

In dieser Vorlesung wird nur ein sehr kleiner Teil dieser Fragestellungen behandelt. Diese Vorlesung dient in erster Linie dazu, die stochastischen Grundlagen zur Behandlung dieser Fragestellungen zu legen.

Die Vorlesung gliedert sich in drei Teile: Beschreibende Statistik, Beschreibung des Zufalls und Schließende Statistik. Bei der Beschreibenden Statistik werden Daten nur beschrieben, wobei man in der Schließenden Statistik Schlüsse auf die zugrunde liegenden Zufallsgesetzmäßigkeiten ziehen möchte. Dazu wird ein vertieftes Verständnis von Zufallsgesetzmäßigkeiten gebraucht.

Es gibt dabei prinzipiell drei Möglichkeiten das erforderliche Verständnis für Zufallsgesetzmäßigkeiten zu erlangen.

1. Man führt entsprechend viele Zufallsexperimente durch, um das entsprechende Verständnis über das Verhalten des Zufalls zu erhalten. So wurden in vorangegangenen Jahrhunderten unzählige Experimente mit dem Würfel durchgeführt. Auch jeder, der viele Würfelspiele gespielt hat, hat eine gute Kenntnis davon, wie sich der Zufall bei Würfeln auswirkt. Aber neben den Würfelexperimenten gibt es noch viel mehr andere Zufallsexperimente von ganz anderer Natur. Da müsste jemand sehr, sehr viele Experimente durchführen, um ein halbwegs gutes Verständnis zu bekommen. Das ist unmöglich.
2. Man beschäftigt sich mit der mathematischen Modellierung, in die die jahrhundertlange Erfahrung mit der Beschäftigung des Zufalls eingegangen ist. Aber das ist nur ein Zugang, der vor allem für Mathematiker geeignet ist.
3. Man simuliert die Zufallsexperimente am Computer und sammelt dadurch in recht kurzer Zeit die Erfahrungen, die ansonsten nur durch langwierige reale Experimente möglich sind.

In der Vorlesung wird vor allem der letzte Weg gegangen, auch wenn die mathematischen Hintergründe aufgezeigt werden. Das hat auch den Vorteil, dass damit neue Entwicklungen erfasst werden können, bei denen die heutige Mathematik an ihre Grenzen stößt. Bei komplizierten Fragestellungen kommt die Mathematik heute nicht mehr ohne Simulationen aus. Das heißt noch lange nicht, dass die Mathematik überflüssig geworden ist. Denn, um Simulationen sinnvoll anwenden zu können, muss dann doch vorher Mathematik betrieben werden.

Teil I

Beschreibende Statistik

Kapitel 2

Erste Schritte mit R

Simulationen können einfach mit der freien Software *R*, siehe (R 2010), durchgeführt werden. Natürlich kann auch jede andere Software, die Zufallszahlen zur Verfügung stellt, benutzt werden. Die Software *R* hat nicht nur den Vorteil, dass sie frei ist, sondern dass sie immer verbreiteter wird, so dass immer mehr Routinen im Internet dazu zu finden sind. Hier wird daher nur diese Software benutzt.

2.1 Herunterladen der freien Software R

Unter <http://cran.r-project.org/> finden Sie die neuesten R-Versionen für Windows, Mac und Linux. Windows-Nutzer klicken nach der Auswahl des Betriebssystems auf **base**. Hier finden Sie auch Installationsanleitungen.

Nun kann R gestartet werden. Dazu ist es sinnvoll R mit einem eigenen Arbeitsverzeichnis zu verknüpfen, weil ansonsten von R erstellte Dateien im Programm-Ordner abgelegt werden. Unter dem Menü-Punkt Hilfe, findet sich unter anderem eine ausführliche Einführung in R auf Englisch. Kennen Sie den Namen der R-Funktion, können Sie eine Hilfe zu der entsprechenden Funktion direkt über die Eingabe *?Name der Funktion* aufrufen. Hier finden Sie auch die Liste der zu übergebenen Argumente und die Ausgaben, sowie Beispiele für die Anwendung der Funktion.

2.2 Grafische Darstellungen mit R

Eindimensionale Grafiken können einfach mit der Funktion **plot** erzeugt werden. Dies wird hier anhand der Gaußschen Normalverteilung, die sicherlich jedem noch von dem 10 DM-Schein bekannt ist, demonstriert. Möchte man zum Beispiel die Dichte der sogenannten Standardnormalverteilung im Intervall $[-4, 4]$ darstellen, so erzeugt man zuerst einen Vektor, der x genannt werden kann, der in aufsteigender Reihenfolge Werte aus dem Intervall $[-4, 4]$ enthält und damit das Intervall $[-4, 4]$ rastert. Je feiner die Rasterung, desto glatter wird die Funktion. Anschließend wendet man die Funktion, die die Dichte der Standardnormalverteilung liefert, auf x an:

```
> x<-seq(-4,4,0.1)
```

```
> plot(x,dnorm(x,mean=0,sd=1),type="l",ylab="f(x)",
+ main="Dichte der Standardnormalverteilung f(x)")
```

Statt `dnorm(x,mean=0,sd=1)` kann auch `dnorm(x)` geschrieben werden, da `mean=0` und `sd=1` die Voreinstellungen sind. Lässt man `type="l"` weg, so werden nur Punkte an den Komponenten von x gezeichnet. Das Argument `ylab="f(x)"` ergibt eine Beschriftung für die y-Achse und mit `main` wird die Überschrift bestimmt.

Will man mehrere Dichten von Normalverteilungen in einer Grafik darstellen, können die weiteren Dichten mit `lines` hinzugefügt werden:

```
> x<-seq(-4,4,0.1)
> plot(x,dnorm(x,mean=0,sd=1),type="l",ylab="f(x)",
+ main="Dichten der Normalverteilung f(x)")
> lines(x,dnorm(x,mean=0,sd=2),lty=2)
> legend(-3.5,0.35,legend=c("sd=1","sd=2"),lty=c(1,2))
```

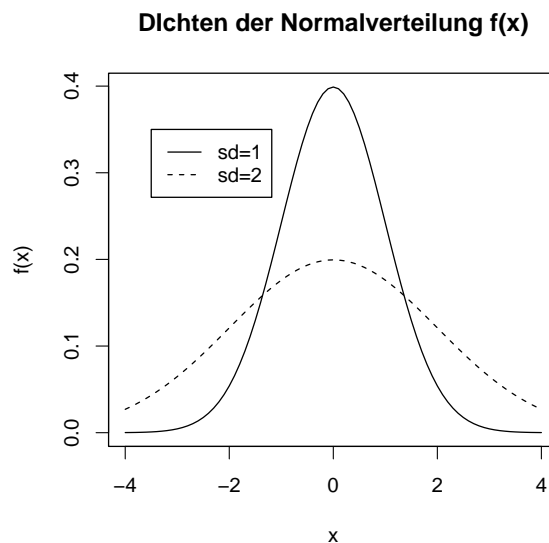


Abbildung 2.2.1: Zwei Dichten der Normalverteilung

Dabei wird mit dem Argument `lty` der Linientyp bestimmt, z.B. `lty=1` für eine durchgezogene Linie, `lty=2` für eine gestrichelte Linie. Eine Legende kann mit der R-Funktion `legend` hinzugefügt werden. Grundsätzlich können, wie schon oben erwähnt, die Argumente und die Ausgaben einer R-Funktion mit dem Befehl `?Funktionsname` abgefragt werden. Mit

```
> ?legend
```

werden zum Beispiel die Argumente der R-Funktion `legend` angezeigt.

Die Dichte der eindimensionalen Normalverteilung mit Parameter `mean=` μ und `sd=` σ ist eine Funktion von \mathbb{R} nach \mathbb{R} und ist gegeben durch

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

2.2.1 Übungsaufgabe

Stellen Sie die folgenden Dichten der Normalverteilung jeweils in einer Grafik dar. Wählen Sie dabei geeignete Bereiche für die x-Achse:

1. Die Dichte der Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=1$ und die Dichte der Normalverteilung mit $\mu=\text{mean}=0$, $\sigma=\text{sd}=1$ in einer Grafik.
2. Die Dichte der Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=1$ und die Dichte der Normalverteilung mit $\mu=\text{mean}=2$, $\sigma=\text{sd}=1$ in einer Grafik.
3. Die Dichte der Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=1$ und die Dichte der Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=2$ in einer Grafik.

Welche Schlüsse ziehen Sie daraus?

Kapitel 3

Merkmale und Datentypen

3.1 Merkmale

3.1.1 Definition (Merkmal, Merkmalsausprägung, Wertebereich)

Eine **statistische Gesamtheit (Kollektiv, Population)** ist die Menge der Einzelobjekte, die in bestimmten, vom Untersuchungsziel vorgegebenen Identifikationskriterien übereinstimmen. Eine untersuchungsrelevante Eigenschaft der Einzelobjekte heißt (statistisches) **Merkmal**. Die Werte, die ein Merkmal haben kann, heißen **Merkmalsausprägungen**. Die Menge der Merkmalsausprägungen heißt **Wertebereich**.

3.1.2 Tabelle (Merkmale und Ausprägungen)

Merkmal	mögliche Ausprägungen
Produktionsort	Kassel, Aachen, Karlsruhe, Leipzig
Alter	$t \geq 0, t \in \mathbb{R}$
Alter	0-10, 11-20, 21-30, \dots , 71-80, 81+
Alter	Kind, Jugendlicher, Erwachsener
Farbe	violett, grün
Farbe	$\lambda, 0.39 < \lambda < 0.69$ (μm , Wellenlänge)
Unfallhäufigkeit	$n \in \mathbb{N}_0$

3.2 Datentypen

3.2.1 Definition (Skalentypen, s. Bortz (1999))

Nominalskala: Eine Nominalskala ordnet den Objekten eines empirischen relationalen Systems Zahlen zu, die so geartet sind, dass Objekte mit gleicher Merkmalsausprägung gleiche Zahlen und Objekte mit verschiedener Merkmalsausprägung verschiedene Zahlen erhalten.

Ordinalskala: Eine Ordinalskala ordnet den Objekten eines empirischen relationalen Systems Zahlen zu, die so geartet sind, dass von jeweils zwei Objekten das Objekt mit der größeren Merkmalsausprägung die größere Zahl erhält.

Intervallskala: Eine Intervallskala ordnet den Objekten eines empirischen relationalen Systems Zahlen zu, die so geartet sind, dass die Rangordnung der Zahlendifferenzen zwischen je zwei Objekten der Rangordnung der Merkmalsunterschiede zwischen je zwei Objekten entspricht.

Verhältnisskala: Eine Verhältnisskala ordnet den Objekten eines empirischen relationalen Systems Zahlen zu, die so geartet sind, dass das Verhältnis zwischen je zwei Zahlen dem Verhältnis der Merkmalsausprägungen der jeweiligen Objekte entspricht.

3.2.2 Tabelle (Beispiele für Skalentypen)

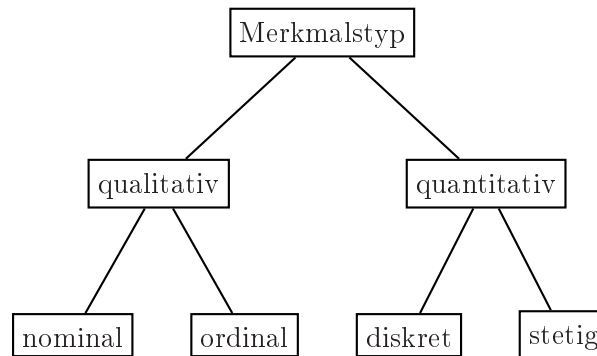
Skalentyp	mögliche Aussagen	Beispiele
Nominalskala	Gleichheit	Telefonnummern, Geschlecht
	Verschiedenheit	Krankheitsklassifikation
Ordinalskala	größer-kleiner Relationen	Ränge, Schulnoten
		Windstärken, Jahreszeiten
Intervallskala	Gleichheit von Differenzen	Temperatur (in ° Celsius)
		Intelligenzquotient
Verhältnisskala	Gleichheit von Verhältnissen	Längen- und Gewichtsmessung

3.2.3 Bezeichnung

Merkmale, die mit Nominal- bzw. Ordinalskala gemessen werden, heißen **qualitativ**, solche, die intervall- oder verhältnisskaliert sind, **quantitativ** oder **metrisch**.

Im folgenden werden die quantitativen Merkmalstypen **diskret** und **stetig** unterschieden.

Merkmalstyp	Anzahl der Ausprägungen	Beispiele
diskret	endlich oder abzählbar unendlich viele	Anzahl richtiger Antworten in einem Test, Anzahl erkrankter Personen in einer Population
stetig	überabzählbar viele	Reaktionszeit, Größe, Gewicht

3.2.4 Bemerkung (Merkmalstypen)

Außerdem wird zwischen **univariaten**, **bivariaten** und **multivariaten Daten** unterschieden. Bei univariaten Daten wird nur ein Merkmal beobachtet, bei bivariaten Daten werden zwei Merkmale beobachtet und bei multivariaten Merkmalen werden mehr als zwei Merkmale erfasst. Bivariate und multivariate Daten werden in Form einer Tabelle dargestellt.

3.2.5 Übungsaufgabe

Für die statistische Auswertung von Datensätzen ist es wichtig zunächst einmal Merkmalsträger, die Merkmale und den Datentyp der Merkmale zu bestimmen, da die anzuwendende statistische Methode abhängig vom Datentyp ist. Geben Sie daher bei den einzelnen Datensätzen aus Kapitel 1 die Merkmalsträger, die Merkmale und den Datentyp der Merkmale an.

Kapitel 4

Tabellarische und grafische Darstellungen von univariaten Daten

4.0.6 Bezeichnung

Merkmale werden mit großen lateinischen Buchstaben (z.B. X, Y, Z) bezeichnet, die zugehörigen Ausprägungen mit dem entsprechenden kleinen lateinischen Buchstaben (z.B. x, y, z).

4.1 Qualitative Daten

Voraussetzungen und Bezeichnungen.

$M_N = \{e_1, \dots, e_N\}$	Population bestehend aus Objekten e_1, \dots, e_N ,
X	nominales bzw. ordinales Merkmal,
$x(j), j \in J$	Merkmalsausprägungen von X ,
$W_X = \{x(j) \mid j \in J\}$	
$= \{x(1), \dots, x(J)\}$	Wertebereich von X mit den Merkmalsausprägungen $x(j), j \in J$,
$D_N = \{x_n \mid n = 1, \dots, N\}$	Datensatz (Urliste) aus der Messung von X in der Population M_N ,
$= \{x_1, \dots, x_N\}$	d.h. $x_n = X(e_n), n = 1, \dots, N$.

Zusätzliche Voraussetzung bei Ordinalskala: $x(1) < x(2) < \dots < x(J)$. J bezeichnet die Anzahl der Ausprägungen von X .

Die (deskriptive) Auswertung eines Datensatzes x_1, \dots, x_N erfolgt in den Schritten:

- Auszählung der Ausprägungen \longrightarrow **absolute Häufigkeit** von Ausprägung $x(j)$: $N_j = N(x(j))$, $j \in J$. D.h. N_j ist die Anzahl der Daten x_n mit $x_n = x(j)$. Dies lässt sich durch eine **Strichliste** leicht umsetzen.
- Bildung der **relativen Häufigkeit** von Ausprägung $x(j)$, $j \in J$: $f_j = \frac{N_j}{N}$, wobei $N = \sum_{j \in J} N_j$

der Stichprobenumfang (Anzahl Beobachtungen) ist. f_1, \dots, f_J heißt **Häufigkeitsverteilung (HV)** von X .

Darstellung einer Häufigkeitsverteilung in einer Tabelle

Ausprägung $x(j)$	absolute Häufigkeit n_j	relative Häufigkeit $f_j = N_j/N$
$x(1)$	N_1	$f_1 = N_1/N$
\vdots	\vdots	\vdots
$x(J)$	N_J	$f_J = N_J/N$
$\sum_{j \in J} N_j = N$		$\sum_{j \in J} f_j = 1$

4.1.1 Beispiel (Festigkeit von Beton)

Liegt die Datei **Druckfestigkeit.csv** im Arbeitsordner von R, so können die Daten aus dieser Datei mit folgendem Befehl in R eingelesen werden:

```
> beton<-read.csv("Druckfestigkeit.csv",header=T,dec=".",sep=";")
```

Mit der R-Funktion **table** können die absoluten Häufigkeiten der Herstellungsorte ermittelt werden:

```
> table(beton$H)
```

Aachen	Darmstadt	Dresden	Karlsruhe	Kassel	Leipzig
12	12	12	11	66	12

Die relativen Häufigkeiten erhält man, in dem durch die Länge des Vektors **beton\$H** (die Länge der Spalte **H**) geteilt wird:

```
> table(beton$H)/length(beton$H)
```

Aachen	Darmstadt	Dresden	Karlsruhe	Kassel	Leipzig
0.096	0.096	0.096	0.088	0.528	0.096

Zur besseren Veranschaulichung des Datensatzes werden graphische Methoden verwendet. Im folgenden werden zwei Methoden am obigen Beispiel vorgeführt.

Balkendiagramm: Beim Balkendiagramm ist die Höhe der einzelnen Balken durch die absoluten bzw. relativen Häufigkeiten der Merkmalsausprägungen gegeben.

Kreis- oder Tortendiagramm: Beim Kreis- oder Tortendiagramm sind die Flächen bzw. die Winkel der Tortenstücke proportional zu den Häufigkeiten der Merkmalsausprägungen.

4.1.2 Beispiel (Fortsetzung: Druckfestigkeit von Beton)

Die Balkendiagramme in Abbildung 4.1.1 wurden mit folgenden Befehlen erzeugt:

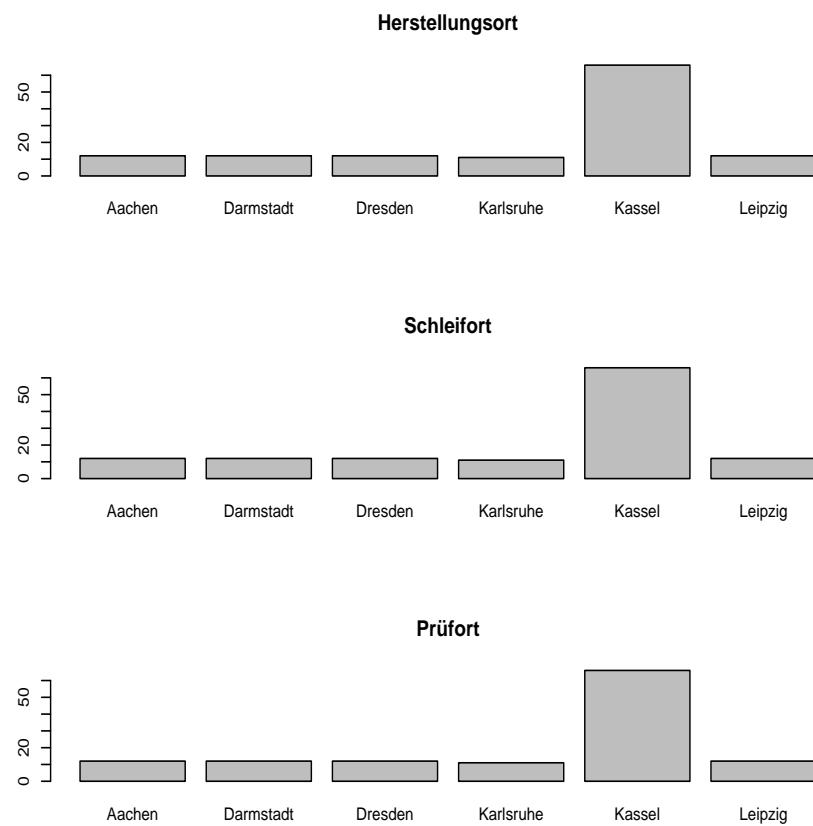


Abbildung 4.1.1: Balkendiagramme für die Herstellungs-, Schleif- und Prüforte

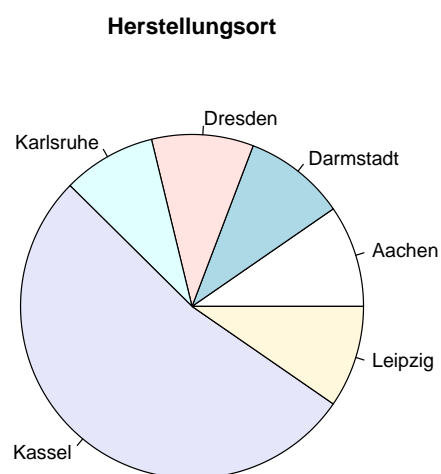


Abbildung 4.1.2: Kreisdiagramm für die Herstellungsorte

```
> par(mfrow=c(3,1))
> barplot(table(beton$H),main="Herstellungsort")
> barplot(table(beton$S),main="Schleifort")
> barplot(table(beton$P),main="Prüfort")
```

Das Kreisdiagramm in Abbildung 4.1.2 wurde mit folgenden Befehlen erzeugt:

```
> par(mfrow=c(1,1))
> pie(table(beton$H),main="Herstellungsort")
```

4.1.3 Übungsaufgabe

Erstellen Sie die Häufigkeitstabelle und die Kreisdiagramme für die Schleiforte aus dem Datensatz Druckfestigkeit.csv. Was lässt sich ablesen?

4.2 Quantitativ diskrete Daten

Voraussetzungen und Bezeichnungen.

$M_N = \{e_1, \dots, e_N\}$	Population bestehend aus Objekten e_1, \dots, e_N ,
X	quantitatives diskretes Merkmal,
$x, x \in W_X$	Merkmalsausprägungen von X ,
$W_X = \{x(j) \mid j \in J\}$ $= \{x(1), \dots, x(J)\}$	Wertebereich von X ,
$D_N = \{x_n \mid n = 1, \dots, N\}$ $= \{x_1, \dots, x_N\}$	Datensatz (Urliste) aus der Messung von X in der Population M_N , d.h. $x_n = X(e_n)$, $n = 1, \dots, N$.

In Analogie zu qualitativen Merkmalen wird die Häufigkeitsverteilung f_1, \dots, f_J gebildet. Zudem berechnet man die **relativen Summenhäufigkeiten** s_1, \dots, s_J , aus denen die **empirische Verteilungsfunktion** gebildet wird. Die relative Summenhäufigkeit s_j ist als

$$s_j = \sum_{k=1}^j f_k = \frac{\text{Anzahl der } x_n \text{ mit } x_n \leq x(j)}{n}$$

definiert.

Darstellung der Tabelle mit Häufigkeiten und Summenhäufigkeiten

Ausprägung $x(j)$	absolute Häufigkeit N_j	relative Häufigkeit f_j	relative Summenhäufigkeit s_j
$x(1)$	N_1	$f_1 = \frac{N_1}{N}$	$s_1 = f_1$
$x(2)$	N_2	$f_2 = \frac{N_2}{N}$	$s_2 = f_1 + f_2$
$x(3)$	N_3	$f_3 = \frac{N_3}{N}$	$s_3 = f_1 + f_2 + f_3$
\vdots	\vdots	\vdots	\vdots
$x(J-1)$	N_{J-1}	$f_{J-1} = \frac{N_{J-1}}{N}$	$s_{J-1} = \sum_{k=1}^{J-1} f_k$
$x(J)$	N_J	$f_J = \frac{N_J}{N}$	$s_J = \sum_{k=1}^J f_k = 1$
$\sum_{j \in J} N_j = N$		$\sum_{j \in J} f_j = 1$	

4.2.1 Beispiel (Anzahl von Unfällen an einer Kreuzung)

Die Datei `Unfall1.txt` enthält die Unfallanzahlen an einer Kreuzung pro Jahr für 20 Jahre. Diese Anzahlen sind:

2 2 6 1 1 1 2 1 0 1 2 0 7 3 1 0 3 2 0 1

Die absoluten, die relativen Häufigkeiten und die Summenhäufigkeiten werden in R wie folgt ermittelt:

```

> unfall<-scan("Unfall.txt")
Read 20 items
> table(unfall)
unfall
0 1 2 3 6 7
4 7 5 2 1 1
> table(unfall)/length(unfall)
unfall
  0    1    2    3    6    7
0.20 0.35 0.25 0.10 0.05 0.05
> cumsum(table(unfall))/length(unfall)
  0    1    2    3    6    7
0.20 0.55 0.80 0.90 0.95 1.00

```

Geeignete grafische Darstellungen sind das Stabdiagramm und die empirische Verteilungsfunktion.

Stabdiagramm: Die Höhe der Stäbe ist durch die absoluten oder relativen Häufigkeiten gegeben, wobei die Stäbe genau auf den Punkten der reellen Achse aufgetragen werden, wo sich die Merkmalsausprägungen befinden.

Empirische Verteilungsfunktion: Bei der empirischen Verteilungsfunktion werden die Summenhäufigkeiten genau oberhalb der Punkte der reellen Achse als Punkte aufgetragen werden, wo sich die Merkmalsausprägungen befinden. Damit eine Funktion entsteht, werden diese Punkte in horizontaler Richtung so lange fortgesetzt, bis die nächste Merkmalsausprägung kommt. Mathematisch ist die Verteilungsfunktion wie folgt definiert:

4.2.2 Definition

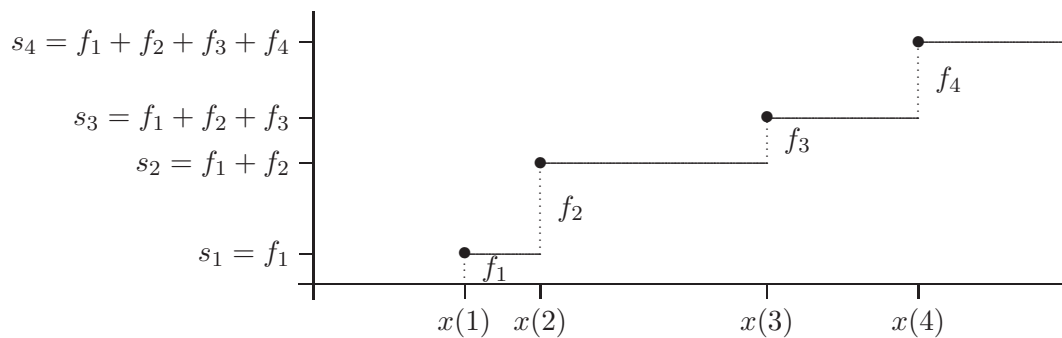
Sei X ein quantitatives diskretes Merkmal mit Merkmalsausprägungen $x(1) < x(2) < \dots < x(J)$ und zugehörigen relativen Häufigkeiten f_1, f_2, \dots, f_J . Dann heißt die Funktion $F_N : \mathbb{R} \rightarrow [0, 1]$ definiert durch

$$F_N(x) = \begin{cases} 0, & \text{für } x < x(1), \\ s_j = \sum_{k=1}^j f_k, & \text{für } x(j) \leq x < x(j+1), j = 1, \dots, J-1 \\ 1, & \text{für } x \geq x(J) \end{cases}$$

empirische Verteilungsfunktion des Merkmals X .

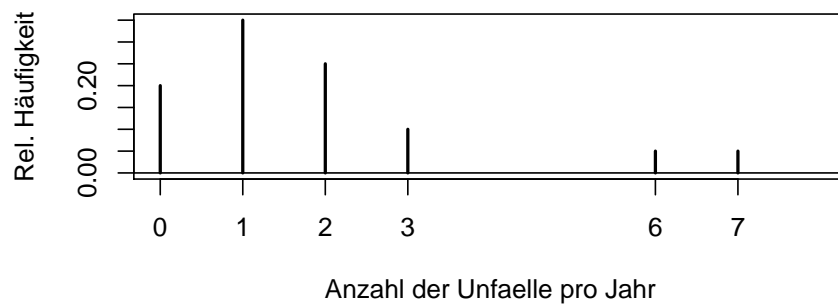
Die empirische Verteilungsfunktion ist eine sogenannte **Treppenfunktion**. Die Sprunghöhe an der Stelle $x(j)$ ist die relative Häufigkeit f_j . Der Graph einer empirischen Verteilungsfunktion sieht etwa

folgendermaßen aus:



Der Punkt \bullet markiert den Funktionswert der empirischen Verteilungsfunktion an der Stelle $x(j)$ und ist durch die Summenhäufigkeit gegeben.

Stabdiagramm



Verteilungsfunktion

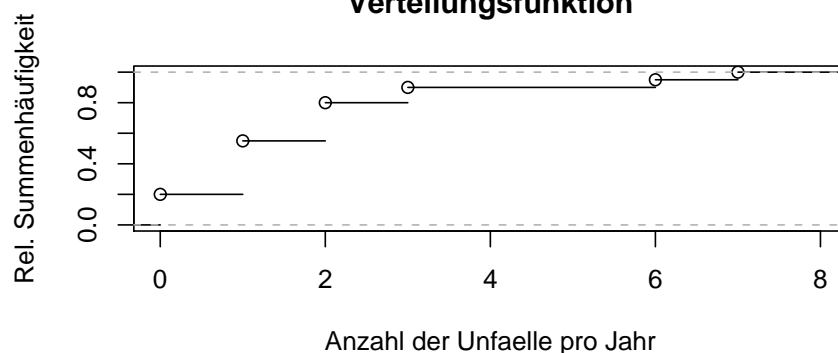


Abbildung 4.2.1: Stabdiagramm und empirische Verteilungsfunktion für die Unfallanzahlen

4.2.3 Beispiel (Fortsetzung von Beispiel 4.2.1)

Das Stabdiagramm und die empirische Verteilungsfunktion in Abbildung 4.2.1 wurden mit folgenden Befehlen erzeugt:

```
> par(mfrow=c(2,1))
```

```
> plot(table(unfall)/length(unfall),type="h",xlab="Anzahl der Unfaelle pro Jahr",
  ylab="Rel. Häufigkeit",xlim=c(0,8),main="Stabdiagramm")
> abline(0,0)
> plot(ecdf(unfall),xlab="Anzahl der Unfaelle pro Jahr",ylab="Rel. Summenhäufigkeit",
  xlim=c(0,8),main="Verteilungsfunktion")
```

4.3 Quantitative stetige Daten

Voraussetzungen und Bezeichnungen.

$M_N = \{e_1, \dots, e_N\}$	Population bestehend aus Objekten e_1, \dots, e_N ,
X	quantitatives stetiges Merkmal
$x, x \in W_X$	Merkmalsausprägungen von X ,
$W_X = \bigcup_{j=1}^J K_j$	Klassierter (kategorisierter) Wertebereich,
$K_j = (v_{j-1}, v_j]$	Merkmalsklasse mit Klassengrenzen $v_0 < v_1 < \dots < v_J$,
$D_N = \{x_n \mid n = 1, \dots, N\}$ $= \{x_1, \dots, x_N\}$	Datensatz (Urliste) aus der Messung von X in der Population M_N , d.h. $x_n = X(e_n)$, $n = 1, \dots, N$.

Quantitative stetige Daten zeichnen sich insbesondere dadurch aus, dass gleiche Merkmalsausprägungen nur sehr selten auftreten. Somit ist eine reine Häufigkeitstabelle nicht sinnvoll und nicht übersichtlich. Um eine übersichtlichere Darstellung zu erhalten, geht man in diesen Fällen oft zur **Kategorisierung oder Klassierung der Meßwerte** über. Der Wertebereich W_X wird in J Klassen K_1, \dots, K_J eingeteilt, wobei $K_j = (v_{j-1}, v_j]$ ein Intervall mit den Klassengrenzen v_{j-1} und v_j ist. Die Beobachtungen werden nun den Klassen zugeordnet. Für die Klassen erhält man dadurch eine Häufigkeitsverteilung:

Klasse K_j	absolute Häufigkeit $N(K_j)$	relative Häufigkeit $f(K_j)$	relative Summenhäufigkeit $s(K_j)$
$K_1 = (v_0, v_1]$	$N(K_1)$	$f(K_1) = \frac{N(K_1)}{N}$	$s(K_1) = f(K_1)$
$K_2 = (v_1, v_2]$	$N(K_2)$	$f(K_2) = \frac{N(K_2)}{N}$	$s(K_2) = f(K_1) + f(K_2)$
\vdots	\vdots	\vdots	\vdots
$K_{J-1} = (v_{J-2}, v_{J-1}]$	$N(K_{J-1})$	$f(K_{J-1}) = \frac{N(K_{J-1})}{N}$	$s(K_{J-1}) = \sum_{k=1}^{J-1} f(K_k)$
$K_J = (v_{J-1}, v_J]$	$N(K_J)$	$f(K_J) = \frac{N(K_J)}{N}$	$s(K_J) = \sum_{k=1}^J f(K_k) = 1$
$\sum_{j=1}^J N(K_j) = N$		$\sum_{j=1}^J f(K_j) = 1$	

Die absoluten Häufigkeiten $N(K_j)$, relativen Häufigkeiten $f(K_j)$ und relativen Summenhäufigkeiten $s(K_j)$ berechnen sich daher gemäß

$$N(K_j) = \text{Anzahl der } x_n \text{ mit } x_n \in K_j, \quad f(K_j) = \frac{N(K_j)}{N}, \quad s(K_j) = \sum_{k=1}^j f(K_k).$$

$f(K_1), \dots, f(K_J)$ heißt **Häufigkeitsverteilung** des Merkmals X zur Klasseneinteilung K_1, \dots, K_J oder **klassierte Häufigkeitsverteilung**. Die Punkte $b_j = v_j - v_{j-1}$ heißen **Klassenbreiten**, $j = 1, \dots, J$.

Die graphische Darstellung dieser Häufigkeitsverteilung erfolgt über ein sogenanntes **Histogramm**. Dies ist ein Flächendiagramm, in dem Rechtecke das Häufigkeitsverhältnis der einzelnen Klassen wiedergeben. Es gilt der Grundsatz:

Die Fläche des Rechtecks über dem Intervall K_j ist proportional zur relativen Häufigkeit $f(K_j)$, d.h.

$$f(K_j) = b_j \cdot h_j = (v_j - v_{j-1})h_j, \quad j = 1, \dots, J,$$

wobei h_j die Höhe des Rechtecks bezeichne. h_j berechnet sich daher gemäß:

$$h_j = \frac{f(K_j)}{b_j}, \quad j = 1, \dots, J.$$

Neben dem Histogramm kann auch die **empirische Verteilungsfunktion**, so wie sie in Abschnitt 4.2 eingeführt wurde, benutzt werden.

4.3.1 Beispiel (Fortsetzung von Beispiel 1.0.7)

Die Druckfestigkeit in Beispiel 1.0.7 ist ein quantitative stetiges Merkmal. Um die absoluten und relativen Häufigkeiten dafür zu bekommen, wenn der Beton in Kassel hergestellt wurde, gibt man folgendes ein:

```
> hist(beton[beton$H=="Kassel","Druck"],plot=F)
$breaks
[1] 110 120 130 140 150 160 170 180
$counts
[1] 4 2 9 19 23 5 4
$intensities
[1] 0.006060605 0.003030303 0.013636364 0.028787879 0.034848485 0.007575758
[7] 0.006060606
$density
[1] 0.006060605 0.003030303 0.013636364 0.028787879 0.034848485 0.007575758
[7] 0.006060606
$mids
[1] 115 125 135 145 155 165 175
$xname
[1] "beton[beton$H == \"Kassel\", \"Druck\"]"
$equidist
[1] TRUE
attr(,"class")
[1] "histogram"
```

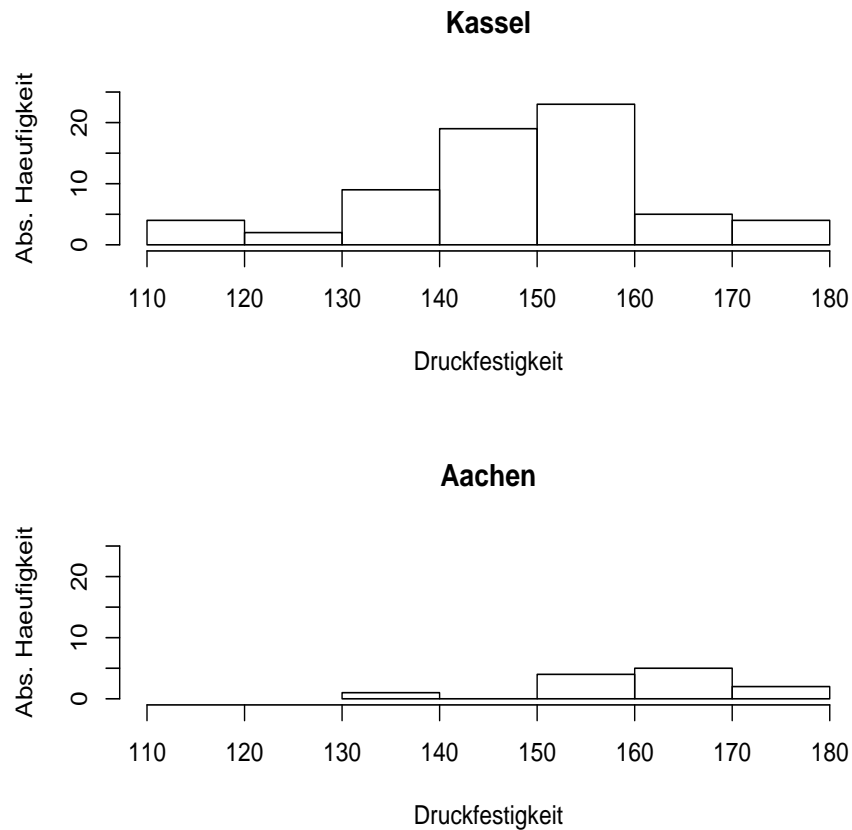


Abbildung 4.3.1: Histogramme der Druckfestigkeitswerte bei Herstellung in Kassel und Aachen

Dabei gibt `counts` die relativen Häufigkeiten an und `intensities` sowie `densities` geben die relativen Häufigkeiten an. `breaks` gibt die benutzten Klassengrenzen an. Diese kann jeder beim Aufruf von `hist` auch selber festlegen, in dem man z.B. `breaks=c(110,140,150,180)` oder `breaks=3` angibt, wenn man nur drei Klassen haben möchte. Im zweiten Fall wählt R geeignete Klassengrenzen, während man im ersten Fall die Klassengrenzen auch festlegt. Die Histogramme in Abbildung 4.3.1 wurden mit folgenden Befehlen erzeugt:

```
> par(mfrow=c(2,1))
> hist(beton[beton$H=="Kassel","Druck"],xlab="Druckfestigkeit",ylab="Abs.
  Häufigkeit",main="Kassel",xlim=c(110,180),ylim=c(0,25))
> hist(beton[beton$H=="Aachen","Druck"],xlab="Druckfestigkeit",ylab="Abs.
  Häufigkeit",main="Aachen",xlim=c(110,180),ylim=c(0,25))
```

Die empirischen Verteilungsfunktionen in Abbildung 4.3.2 erhält man wie folgt:

```
> par(mfrow=c(2,1))
> par(cex=0.7,cex.lab=1.5)
> plot(ecdf(beton[beton$H=="Kassel","Druck"]),xlim=c(110,180),xlab="Druckfestigkeit",
  main="Kassel",pch=16)
```

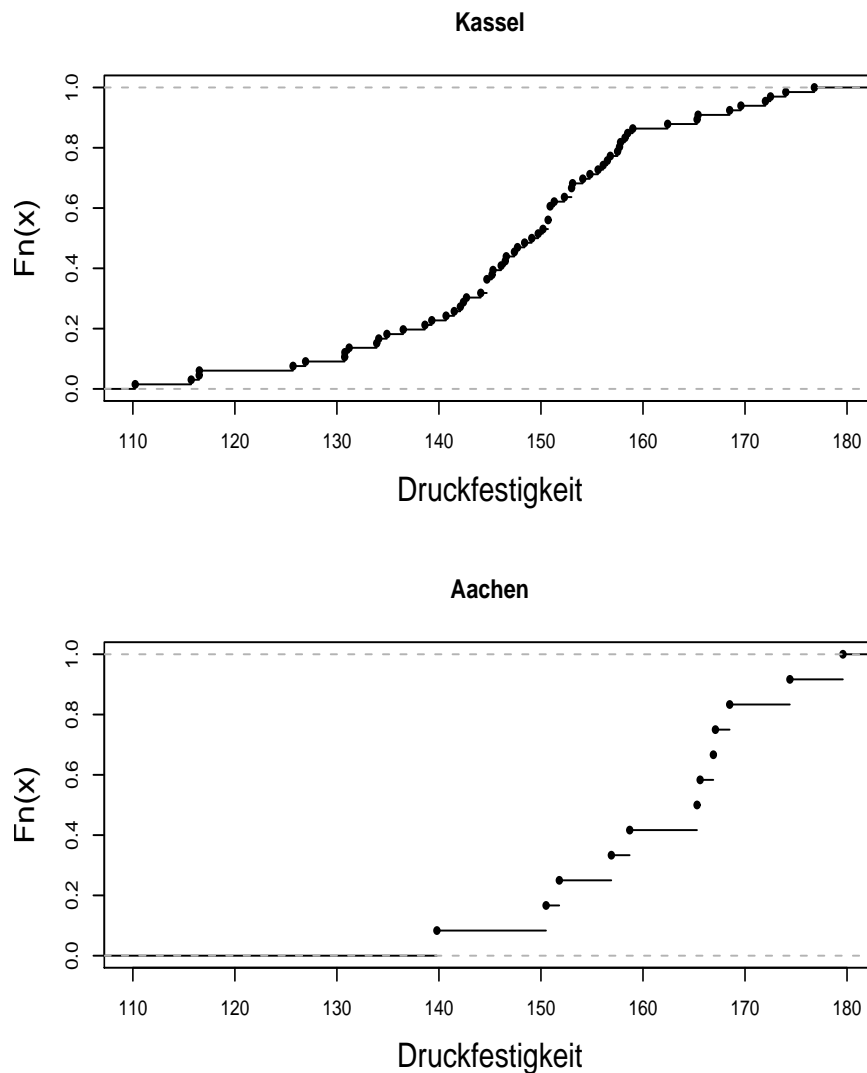


Abbildung 4.3.2: Empirische Verteilungsfunktionen der Druckfestigkeitswerte bei Herstellung in Kassel und Aachen

```
> plot(ecdf(beton[beton$H=="Aachen", "Druck"]), xlim=c(110,180), xlab="Druckfestigkeit",
      main="Aachen", pch=16)
```

Mit `cex` und `cex.lab` werden die Größe der Symbole und der Beschriftungen beeinflusst. Das Argument `pch=16` bewirkt, dass statt Kreisen ausgefüllte Punkte gezeichnet werden.

4.3.2 Übungsaufgabe

Die Durchmesser der Kugellager aus den beiden Produktionslinien von Beispiel 1.0.1 sollen verglichen werden. Stellen Sie dazu die Daten im Datensatz `STEEL.DAT` mittels Histogrammen und Verteilungsfunktionen gegenüber. Geben Sie auch die Tabelle mit den absoluten und relativen Häufigkeiten sowie den Summenhäufigkeiten an. Lesen Sie den Datensatz `STEEL.DAT` mit folgenden

Befehlen ein:

```
> steel<-scan("STEEL.DAT")
Read 20 items
> steel1<-steel[1:10]
> steel2<-steel[11:20]
```

Welche Unterschiede gibt es?

Kapitel 5

Statistische Kennzahlen für die Lage

In den vorherigen Abschnitten wurden Methoden zur Darstellung von Daten verwendet, die nur ein geringes Maß an Reduktion vornahmen (z.B. Häufigkeitsverteilung, Histogramm, empirische Verteilungsfunktion, Bildung von Klassen, empirische Verteilungsfunktion etc.). Oft ist man aber daran interessiert, Datensätze mit wenigen Kenngrößen zu beschreiben (etwa um zwei Datensätze miteinander zu vergleichen).

Lageparameter dienen zur Beschreibung des „Zentrums/Schwerpunktes“ eines Datensatzes. Sie geben Anhaltspunkte für die Lage eines untersuchten Merkmals (\rightarrow „Durchschnittswert“). Im folgenden werden für die unterschiedlichen Merkmalstypen diverse Lageparameter vorgestellt.

5.1 Nominale Daten

Wie in den vorhergehenden Abschnitten werden folgende Bezeichnungen verwendet: $x(j)$, $j = 1, \dots, J$: Merkmalsausprägungen von X ; W_X : Wertebereich von X ; x_1, \dots, x_N : Datensatz aus einer Messung von X .

5.1.1 Definition (Modus, Modalwert)

Der **Modus** x_{mod} des Datensatzes x_1, \dots, x_N ist definiert durch diejenige Ausprägung $x(j^*) \in W_X$ mit

$$N(x(j^*)) = \max\{N(x(1)), \dots, N(x(J))\}, \quad x_{mod} = x(j^*).$$

x_{mod} bezeichnet also die Ausprägung des Merkmals X , die im Datensatz x_1, \dots, x_N die größte absolute (bzw. relative) Häufigkeit besitzt. Wird der Modus für einen speziellen Datensatz ausgewertet, so heißt der resultierende Wert **Modalwert** des Datensatzes.

5.1.2 Beispiel (Druckfestigkeit von Beton)

Aus Beispiel 4.1.1 ist ersichtlich, dass Kassel als Herstellungsort am häufigsten vorkommt. Also ist $x_{mod} = \text{Kassel}$ bzw. **Kassel** ist der Modalwert.

5.2 Ordinale Daten

Wie im Fall eines nominalen Merkmals kann der Modus zur Beschreibung der Lage verwendet werden. Ein weiterer sinnvoller Lageparameter ist der **Median**. Dazu wird der **Rangwert** einer Beobachtung im vorliegenden Datensatz benötigt.

5.2.1 Definition (Rangwert, Rang, Maximum, Minimum)

Seien x_1, \dots, x_N Beobachtungen eines ordinalen Merkmals. Im (aufsteigend) geordneten Datensatz

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

bezeichnet $x_{(n)}$, $n = 1, \dots, N$, den **n -ten Rangwert**. Der 1-te Rangwert $x_{(1)}$ heißt auch **Minimum**, der N -te Rangwert $x_{(N)}$ heißt auch **Maximum** des Datensatzes x_1, \dots, x_N .

Mit dem **Rang** $R(x_n)$ der **Beobachtung** x_n wird die Position der n -ten Beobachtung x_n in der geordneten Folge $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ bezeichnet. Tritt ein Wert x_n s -mal im Datensatz auf, d.h.,

$$x_{(r-1)} < \underbrace{x_{(r)} = \dots = x_{(r+s-1)}}_{\text{alle } s \text{ Werte sind gleich } x_n} < x_{(r+s)},$$

so wird der Rang von x_n definiert als $\frac{1}{s} \sum_{l=r}^{r+s-1} l$.

5.2.2 Bemerkung

Die Bildung einer Anordnung der Beobachtungen setzt zwingend ein ordinal skaliertes Merkmal voraus. Für rein nominale Datensätze machen Rangwerte und Ränge daher keinen Sinn.

5.2.3 Beispiel (Elastizität von Kunststoff)

Die Elastizität eines Kunststoffes wurde in 0, 1, 2, 3, 4 klassifiziert, wobei 4 für die höchste Elastizität steht. Bei 10 Kunststoffteilen wurden folgende Elastizitätswerte ermittelt:

2 3 3 1 3 2 0 1 4 1

Diese werde wie folgt in R eingegeben:

```
> plastik<-c(2,3,3,1,3,2,0,1,4,1)
```

Die sortierten Daten und die Ränge (des unsortierten Datensatzes) erhält man dann wie folgt:

```
> sort(plastik)
[1] 0 1 1 1 2 2 3 3 3 4
> rank(plastik)
[1] 5.5 8.0 8.0 3.0 8.0 5.5 1.0 3.0 10.0 3.0
```

Mittels der Rangwerte wird nun der Median definiert.

5.2.4 Definition (Median für ordinale Merkmale)

Sei $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ der geordnete Datensatz. Dann wird der **Median** $\tilde{x}_{0.5}$ definiert durch

$$\tilde{x}_{0.5} = \begin{cases} x_{(k)}, & k = \frac{N+1}{2}, N \text{ ungerade,} \\ \{x_{(k)}, x_{(k+1)}\}, & k = \frac{N}{2}, N \text{ gerade.} \end{cases}$$

Eine alternative Definition des Medians ist die folgende: $\tilde{x}_{0.5}$ ist ein Wert, für den gilt:

- a) Die Anzahl der x_n mit $x_n \leq \tilde{x}_{0.5}$ ist $\geq \frac{N}{2}$ bzw. mindestens 50% der x_n erfüllen $x_n \leq \tilde{x}_{0.5}$.
- b) Die Anzahl der x_n mit $x_n \geq \tilde{x}_{0.5}$ ist $\geq \frac{N}{2}$ bzw. mindestens 50% der x_n erfüllen $x_n \geq \tilde{x}_{0.5}$.

5.2.5 Bemerkung

- Zur Berechnung des Medians wird der geordnete Datensatz verwendet. Es wird lediglich die Reihenfolge der Beobachtungen vernachlässigt.
- Ist der Stichprobenumfang N ungerade, so ergibt sich die in der Mitte des geordneten Datensatzes liegende Beobachtung als Median. Ist N gerade, so existiert ein in der Mitte liegender Wert nicht. In diesem Fall ist der Median eine Menge aus i.a. zwei Werten!

Eine Verallgemeinerung des Medians sind die **p -Quantile**. Der Datensatz x_1, \dots, x_N wird in zwei Anteile aufgeteilt, wobei der erste Teil ca. $100p\%$ der gesamten Beobachtungen umfaßt und der zweite Teil die restlichen $100(1-p)\%$ der Beobachtungen enthält ($p \in (0, 1)$):

$$\underbrace{x_{(1)} \leq \dots \leq x_{(k)}}_{\approx 100p\%} \leq \underbrace{x_{(k+1)} \leq \dots \leq x_{(N)}}_{\approx 100(1-p)\%}.$$

5.2.6 Definition (p -Quantil für ordinale Merkmale)

Für $p \in (0, 1)$ wird das **p -Quantil** \tilde{x}_p des Datensatzes x_1, \dots, x_N definiert durch

$$\tilde{x}_p = \begin{cases} x_{(k)}, & Np < k < Np + 1, Np \notin \mathbb{N}, \\ \{x_{(k)}, x_{(k+1)}\}, & k = Np \in \mathbb{N}. \end{cases}$$

5.2.7 Bemerkung

- Für $p = 0.5$ ergibt sich der Median, d.h., das 0.5-Quantil entspricht dem Median.
- Ist $Np \in \mathbb{N}$, so ist das p -Quantil i.a. eine aus zwei Werten bestehende Menge (vgl. Median: $N \cdot \frac{1}{2} \in \mathbb{N}$ ist gleichbedeutend damit, dass N gerade ist!).
- Für gewisse Werte von p werden für das zugehörige p -Quantil eigene Bezeichnungen verwendet:

p	Bezeichnung
0.5	Median
0.25	unteres Quartil
0.75	oberes Quartil
$\frac{k}{10}$	k -tes Dezantil ($k = 1, \dots, 9$)
$\frac{k}{100}$	k -tes Perzentil ($k = 1, \dots, 99$)

- Das p -Quantil \tilde{x}_p ist ein Wert im geordneten Datensatz $x_{(1)} \leq \dots \leq x_{(N)}$, so dass mindestens $100p\%$ der beobachteten Werte nicht größer sind und mindestens $100(1-p)\%$ der beobachteten Werte nicht kleiner sind. Für $p = 0.05$ gibt das 0.05-Quantil also einen Beobachtungswert an, für den mindestens 5% der Beobachtungen nicht größer und mindestens 95% der Beobachtungen nicht kleiner sind.

5.2.8 Beispiel (Fortsetzung von Beispiel 5.2.3)

Den Median, das Minimum, das 0.25-Quantil, das 0.75-Quantil, das Maximum und das 0.1-Quantil der Elastizitätsklassen aus Beispiel 5.2.3 erhält man mit folgenden Befehlen:

```
> median(plastik)
[1] 2
> quantile(plastik,type=1)
 0%  25%  50%  75% 100%
  0    1    2    3    4
> quantile(plastik,0.10,type=1)
10%
 0
```

5.3 Quantitative Daten

In Analogie zu qualitativen Daten können die vorgestellten Lageparameter auch zur Beschreibung des Datensatzes x_1, \dots, x_N bei vorliegendem quantitativem Merkmal verwendet werden. Der Modus ist wie bei qualitativen Merkmalen definiert als die Ausprägung mit der größten absoluten/relativen Häufigkeit. Es ist zu beachten, dass der Modus bei stetigen quantitativen Merkmalen meist keinen Sinn ergibt, da i.a. jeder Beobachtungswert nur einmal auftreten wird. Bei diskreten Merkmalen ist er aber ein adäquates Beschreibungselement. Der Median bzw. die p -Quantile werden abweichend definiert, da mit den Ausprägungen quantitativer Merkmale „gerechnet“ werden kann.

5.3.1 Definition (Median, p -Quantil für quantitative Merkmale)

Für $p \in (0, 1)$ wird das **p -Quantil** \tilde{x}_p des Datensatzes x_1, \dots, x_N definiert durch

$$\tilde{x}_p = \begin{cases} x_{(k)}, & Np < k < Np + 1, \quad Np \notin \mathbb{N}, \\ \frac{1}{2}(x_{(k)} + x_{(k+1)}), & k = Np \in \mathbb{N}. \end{cases}$$

Das 0.5-Quantil wird auch als **Median** bezeichnet.

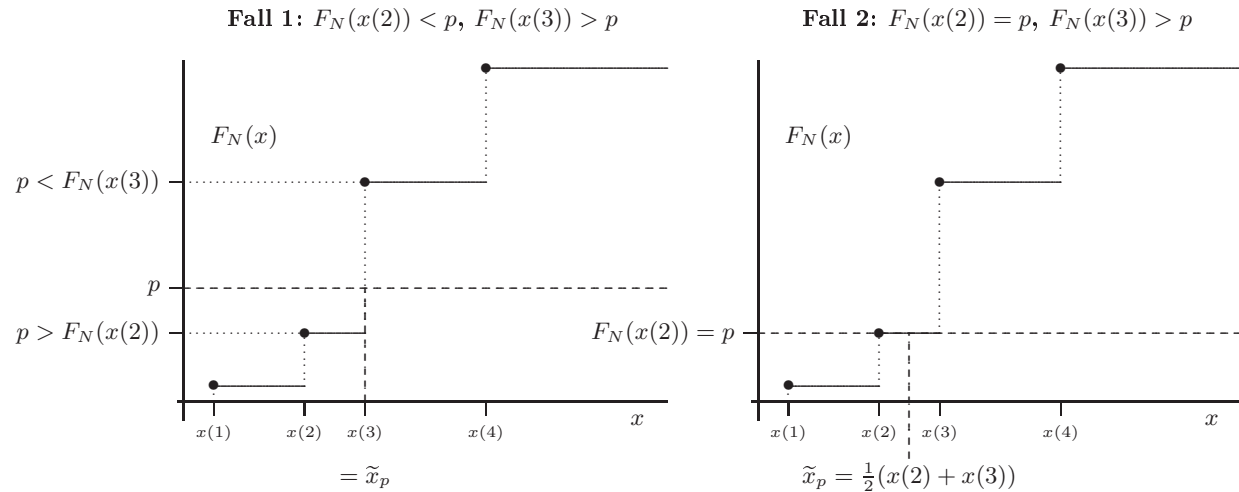
5.3.2 Bemerkung (Berechnung der p -Quantile mittels empirischer Verteilungsfunktion)

Sei $p \in (0, 1)$. Bei **diskretem** Merkmal X mit Ausprägungen $x(1) < \dots < x(J)$ lässt sich das p -Quantil \tilde{x}_p eines Datensatzes x_1, \dots, x_N auch aus der empirischen Verteilungsfunktion F_N berechnen. Für die als p -Quantil in Frage kommende Ausprägung $x(j)$ muss nämlich entweder

$$(F_N(x(j-1)) < p \quad \text{und} \quad F_N(x(j)) > p) \quad \text{oder} \quad F_N(x(j)) = p$$

erfüllt sein. Im ersten Fall gilt $\tilde{x}_p = x(j)$. Im zweiten Fall gilt $\tilde{x}_p = \frac{1}{2}(x(j) + x(j+1))$, falls $F_N(x(j+k)) = p$ für $k = 1, \dots, l-1$ und $F_N(x(j+l)) > p$ erfüllt sind. In den meisten Fäl-

len wird im zweiten Fall $l = 1$, d.h. $F_N(x(j+1)) > p$, gelten. Die graphische Darstellung der beiden Situationen ist in den folgenden Abbildungen illustriert.



Bei quantitativen Daten x_1, \dots, x_N wird oft das arithmetische Mittel \bar{x} als Lageparameter verwendet. Umgangssprachlich wird es mit „Mittelwert, Durchschnittswert“ o.ä. bezeichnet.

5.3.3 Definition (arithmetisches Mittel)

Seien x_1, \dots, x_N Beobachtungen eines quantitativen Merkmals X . Dann heißt

$$\bar{x}_N = \frac{1}{N} \sum_{n=1}^N x_n$$

arithmetisches Mittel von x_1, \dots, x_N . Sind Mißverständnisse ausgeschlossen, so wird der Index n weggelassen, und man schreibt einfach \bar{x} .

Ist der Datensatz durch eine Häufigkeitsverteilung f_1, \dots, f_J beschrieben, d.h. f_j ist die relative Häufigkeit von Ausprägung $x(j)$, so wird \bar{x} definiert durch:

$$\bar{x} = \sum_{j=1}^J f_j \cdot x(j).$$

In diesem Fall heißt \bar{x} auch **gewogenes arithmetisches Mittel**.

5.3.4 Bemerkung

Das arithmetische Mittel kann durch Ausreißer sehr verfälscht werden. Dies ist nicht der Fall beim Median. Unterscheiden sich Mittelwert und Median stark, so muss man davon ausgehen, dass Ausreißer vorhanden sind.

5.3.5 Beispiel (Festigkeit von Beton, Fortsetzung von Beispiel 1.0.7)

Die Mediane und die arithmetischen Mittel der Druckfestigkeiten bei Herstellung in Kassel und Aachen erhält man wie folgt:

```
> median(beton[beton$H=="Kassel", "Druck"])
```

```
[1] 149.415
> median(beton[beton$H=="Aachen","Druck"])
[1] 165.45
> mean(beton[beton$H=="Kassel","Druck"])
[1] 147.8405
> mean(beton[beton$H=="Aachen","Druck"])
[1] 162.0917
> quantile(beton[beton$H=="Kassel","Druck"])
      0%      25%      50%      75%     100%
110.2300 141.6500 149.4150 156.4075 176.8000
> quantile(beton[beton$H=="Aachen","Druck"])
      0%      25%      50%      75%     100%
139.800 155.625 165.450 167.450 179.600
```

Alle Lagemaßzahlen sind bei Aachen höher als bei Kassel. Die Mediane und die arithmetischen Mittel unterscheiden sich innerhalb eines Herstellungsortes wenig, so dass man davon ausgehen kann, dass keine Ausreißer stark die Lage verfälschen.

5.3.6 Übungsaufgabe (Fortsetzung von Übungsaufgabe 4.3.2)

Berechnen Sie zum Vergleich der Produktionslinien alle sinnvollen Lagekennzahlen für die Daten im Datensatz `STEEL.DAT`.

5.3.7 Übungsaufgabe

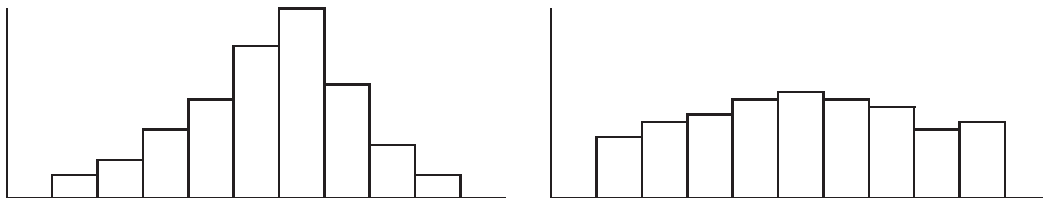
Kleine Stahlproben wurden Tausenden von Lastwechseln ausgesetzt. Nach 0, 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 12.000, 14.000, 16.000, 18.000 Lastwechseln wurden mit einem Mikroskop Aufnahmen der Probenoberfläche aufgenommen. Dabei wurden pro Zeitpunkt 54 Bildausschnitte von der Probenoberfläche aufgenommen. Mit einem Risserkennungsprogramm wurden Mikrorisse identifiziert. Aus diesen identifizierten Rissen wurden dann zu jedem Bildausschnitt und Zeitpunkt die Anzahl der Risse bestimmt. Die Datei `CrackCounts.dat` enthält diese Rissanzahlen. Dabei bezeichnen die Spalte `No` die Nummer des Bildausschnittes, die Spalten `T0`, `T1`, `T2`, `T3`, `T4`, `T5`, `T6`, `T7`, `T8`, `T9`, `T10`, `T12`, `T14`, `T16`, `T18` die Zeitpunkte 1 bis 18 (in tausend Lastwechseln). Erstellen Sie Histogramme von den Rissanzahlen zu den Zeitpunkten 0, 5, 10 und 18 und berechnen Sie die Mediane und die arithmetischen Mittel. Lesen Sie den Datensatz `CrackCounts.dat` mit folgendem Befehl ein:

```
> CrackCounts<-read.table("CrackCounts.dat",header=T)
```

Kapitel 6

Statistische Kennzahlen für die Streuung

Während im vorhergehenden Abschnitt ein Datensatz durch verschiedene Lageparameter beschrieben wurde, sollen im folgenden Größen angegeben werden, die die Streuung eines Merkmals darstellen. Die Streuung in den Daten resultiert daraus, dass bei Messungen eines Merkmals i.a. verschiedene Werte beobachtet werden (z.B. Körpergrößen in einer Gruppe von Personen oder erreichte Noten in einem Examen). Zwar ermöglichen Lageparameter die Beschreibung eines Datensatzes durch Angabe eines mittleren Wertes, jedoch können zwei Datensätze mit gleichem oder zumindest nahezu gleichem Lageparameter sehr unterschiedliche Streuung um diesen Lageparameter aufweisen. Der Lageparameter liefert somit eine nur unzureichende Information über die Struktur des Datensatzes. Dieses Verhalten läßt sich auch aus den Häufigkeitsverteilungen bzw. den zugehörigen Histogrammen ablesen:



Streuungsparameter ergänzen die in den Lageparametern enthaltene Information und dienen dazu, das Abweichungsverhalten (des Merkmals) in einer Population zu quantifizieren. Sie werden unterschieden in diejenigen, die auf der

- Differenz zwischen zwei Lageparametern beruhen (etwa Differenz zwischen Maximum und Minimum = Spannweite) und solchen, die die
- Abweichung zwischen den beobachteten Werten und einem Lageparameter (etwa quadratische Abweichung zwischen Beobachtungen und arithmetischem Mittel = Varianz) zur Berechnung nutzen.

Zudem unterscheidet man Streuungsparameter, die die gleiche Dimension wie die Daten haben (z.B. Spannweite, Quartilsabstand, Standardabweichung), und solche, die dimensionslos sind (z.B. Quartilskoeffizient und Variationskoeffizient).

Zur Interpretation von Streuungsparametern läßt sich festhalten: Je größer der Wert eines Streuungsparameters ist, desto größer streuen die Beobachtungen. Ist der Wert klein, so sind die Beobachtungen um einen Wert konzentriert.

Grundsätzlich setzen alle Streuungsparameter einen Abstandsbegriff voraus, so dass immer ein **quantitatives Merkmal** angenommen werden muß. Dabei ist unwesentlich, ob das Merkmal diskret oder stetig ist.

6.0.8 Definition (Streuungsparameter)

Seien x_1, \dots, x_N Beobachtungswerte eines Merkmals X . Dann heißen:

1. die Differenz zwischen Maximum und Minimum **Spannweite**: $R = x_{(N)} - x_{(1)}$.
2. die Differenz zwischen oberem und unterem Quartil **Quartilsabstand**: $Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}$.
3. **Median der absoluten Abweichungen**: $d_{MAD} = \text{Median von } |x_1 - \tilde{x}_{0.5}|, \dots, |x_N - \tilde{x}_{0.5}|$.
4. **Standardabweichung**: $s = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2} = \sqrt{\frac{1}{N-1} \left(\sum_{n=1}^N x_n^2 - N \cdot \bar{x}^2 \right)}$.

Von Streuungsparametern abgeleitete Größen:

1. **Varianz** (mittlere quadratische Abweichung): $s^2 = \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2$.
2. **Quartilskoeffizient**: $Q_{\text{koeff}} = \frac{2Q}{\tilde{x}_{0.25} + \tilde{x}_{0.75}} = \frac{2(\tilde{x}_{0.75} - \tilde{x}_{0.25})}{\tilde{x}_{0.25} + \tilde{x}_{0.75}}$, wobei X nur nicht-negative Ausprägungen besitzt..
3. **Variationskoeffizient**: $V = \frac{s}{\bar{x}}$, wobei X nur nicht-negative Ausprägungen besitzt.

6.0.9 Bemerkung

Werden die Daten durch eine Häufigkeitsverteilung f_1, \dots, f_m mit zugehörige Ausprägungen $x(1) < \dots < x(m)$ beschrieben, so können folgende alternative Formeln verwendet werden:

1. **Spannweite**: $R = \max\{x(j) \mid f_j > 0\} - \min\{x(j) \mid f_j > 0\}$. Gilt $f_1 > 0$ und $f_J > 0$, so folgt: $R = x(J) - x(1)$.
2. **Varianz**: $s^2 = \frac{N}{N-1} \sum_{j=1}^J f_j (x(j) - \bar{x})^2$.
3. **Standardabweichung**: $s = \sqrt{\frac{N}{N-1} \sum_{j=1}^J f_j (x(j) - \bar{x})^2}$.

6.0.10 Bemerkung

1. Werden mehrere Datensätze wie x_1, \dots, x_N und y_1, \dots, y_N betrachtet, so wird zur Identifikation der Varianz und der Standardabweichung ein x bzw. ein y angehängt, d.h. statt s^2 und s werden s_x^2 , s_y^2 und s_x , s_y benutzt.
2. Wird zu einem Datensatz x_1, \dots, x_N der linear transformierte Datensatz y_1, \dots, y_N mit $y_n = ax_n + b$, $n = 1, \dots, N$, betrachtet, so gilt für die Varianz und die Standardabweichung:

$$s_y^2 = a^2 s_x^2 \quad \text{bzw.} \quad s_y = |a| s_x.$$

Die Eigenschaft für die Standardabweichung gilt auch für die Spannweite, den Quartilsabstand, die mittlere absolute Abweichung und den Median der absoluten Abweichungen.

3. Die Spannweite und die Standardabweichung (bzw. Varianz) sind sehr Ausreißer-empfindlich. Dagegen sind der Quartilsabstand und der Median der absoluten Abweichungen sehr unempfindlich (robust) gegenüber Ausreißern.
4. Die Streuungsparameter Quartilskoeffizient und Variationskoeffizient sind dimensionslose Parameter. Der Quartilskoeffizient setzt den Quartilsabstand ins Verhältnis zum Lageparameter *Quartilsmitte*

$$Q_{\text{mitte}} = \frac{1}{2}(\tilde{x}_{0.25} + \tilde{x}_{0.75}),$$

d.h. $Q_{\text{koeff}} = \frac{Q}{Q_{\text{mitte}}}$, während der Variationskoeffizient Standardabweichung und Mittelwert nutzt. Beide Streuungsparameter sind hilfreich beim Vergleich zweier Datensätze. Ist z.B. der Mittelwert in zwei Datensätzen recht unterschiedlich, so hilft der Variationskoeffizient die Streuung der Datensätze miteinander zu vergleichen.

6.0.11 Beispiel (Druckfestigkeit von Beton, Fortsetzung von Beispiel 1.0.7)

Für die beiden Herstellungsorte Kassel und Aachen erhält man für Druckfestigkeit folgende Streuungsparameter.

Spannweite:

```
> max(beton[beton$H=="Kassel", "Druck"])-min(beton[beton$H=="Kassel", "Druck"])
[1] 66.57
> max(beton[beton$H=="Aachen", "Druck"])-min(beton[beton$H=="Aachen", "Druck"])
[1] 39.8
```

Quartilsabstand:

```
> quantile(beton[beton$H=="Kassel", "Druck"], 0.75)-quantile(beton[beton$H=="Kassel", "Druck"], 0.25)
14.7575
> quantile(beton[beton$H=="Aachen", "Druck"], 0.75)-quantile(beton[beton$H=="Aachen", "Druck"], 0.25)
11.825
```

Median der absoluten Abweichungen (MAD):

```
> mad(beton[beton$H=="Kassel", "Druck"])
[1] 10.89711
> mad(beton[beton$H=="Aachen", "Druck"])
[1] 11.34189
```

Standardabweichung:

```
> sd(beton[beton$H=="Kassel", "Druck"])
[1] 14.08459
> sd(beton[beton$H=="Aachen", "Druck"])
[1] 11.06004
```

Bis auf den MAD sind die Streuungsparameter bei Kassel höher als bei Aachen. Dass der MAD sich anders verhält, kann darauf hindeuten, dass Ausreißer die höhere Streuung von Kassel bewirken.

Mittels der p -Quantile läßt sich eine einfache graphische Methode angeben, die zur Darstellung von Lage und Streuung eines Datensatzes verwendet werden kann, der sogenannte **Box-Plot** (Kasten-Diagramm). Diese Diagramme sind besonders zum Vergleich von mehreren Datensätzen geeignet und visualisieren sowohl Lageparameter als auch Streuungsparameter. Die verwendeten Lage- und Streuungsparameter können unterschiedlich gewählt werden, üblich sind das untere Quartil als untere, bzw. linke Begrenzung der Box, und das obere Quartil als obere, bzw. rechte Begrenzung der Box. Außerdem wird in der Box noch die Lage des Medians mithilfe eines Querbalkens deutlich gemacht. Innerhalb der Grenzen der Box liegen dann mindestens 50% der Daten. Bei den **Box-Whisker-Plots** werden zusätzlich noch sogenannte Whisker eingetragen. Dabei ist

- der **obere Whisker** w_o die größte Beobachtung, die kleiner als $\tilde{x}_{0.75} + 1.5(\tilde{x}_{0.75} - \tilde{x}_{0.25})$ ist,
- der **untere Whisker** w_u die kleinste Beobachtung, die größer als $\tilde{x}_{0.25} - 1.5(\tilde{x}_{0.75} - \tilde{x}_{0.25})$ ist.

Alle Beobachtungen, die nicht im Intervall $[w_u, w_o]$ liegen, werden als Ausreißer aufgefasst.

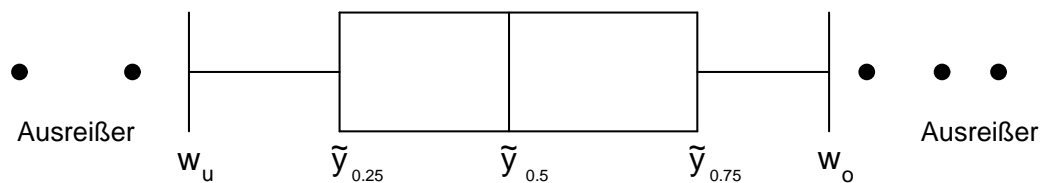


Abbildung 6.0.1: Box-Whisker-Plot

6.0.12 Beispiel (Druckfestigkeit von Beton, Fortsetzung von Beispiel 1.0.7)

Die in Abbildung 6.0.2 gegebenen Box-Whisker-Plots wurden wie folgt gewonnen:

```
> boxplot(Druck~H,data=beton)
```

Aus diesen sieht man deutlich, dass bei Kassel einige Ausreißer auftreten. Dies ist der Grund dafür, dass sich der MAD anders als die anderen Streuungsparameter beim Vergleich von Kassel und Aachen verhielt.

6.0.13 Übungsaufgabe

Bestimmen Sie für die vier prozentualen Anteile der Zementkomponenten a, b, c und d im Datensatz `SETTING.DAT` den Median und die 25%- und 75%- Quantile. Welche statistischen Kennzahlen könnten noch berechnet werden? Lesen Sie den Datensatz `SETTING.DAT` mit folgendem Befehl ein:

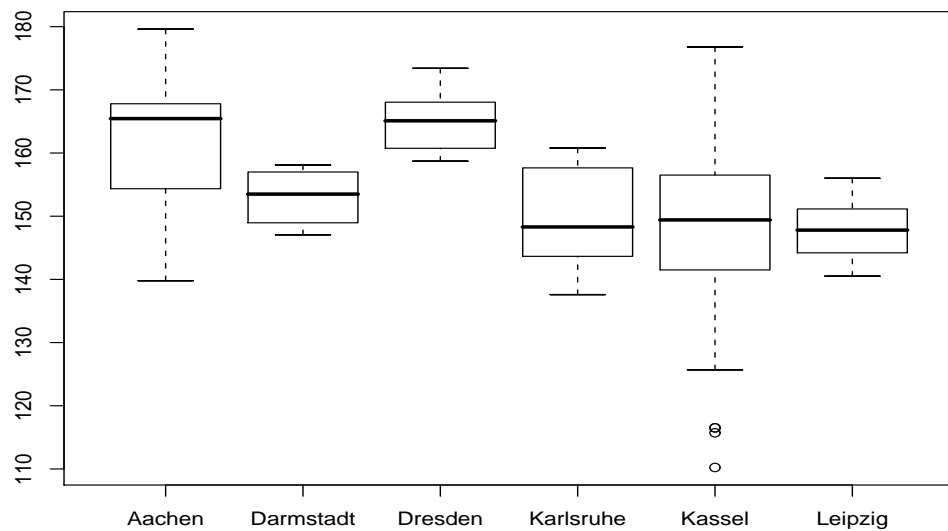


Abbildung 6.0.2: Box-Whisker-Plots für die Herstellungsorte

```
> read.table("SETTING.DAT",header=F,col.names=c("a","b","c","d","Hitze"))
```

6.0.14 Übungsaufgabe (Fortsetzung von Übungsaufgabe 4.3.2)

Stellen Sie die Daten der beiden Produktionslinien im Datensatz `STEEL.DAT` mittels Box-Whisker-Plots gegenüber und berechnen Sie jeweils alle möglichen Streuungsparameter.

6.0.15 Übungsaufgabe

Betrachten Sie den Datensatz `CrackCounts.dat` von Übungsaufgabe 5.3.7. Erstellen Sie die Boxplots der Rissanzahlen für alle 15 Zeitpunkte und interpretieren Sie die Ergebnisse.

Kapitel 7

Bivariate Daten

7.1 Tabellarische und grafische Darstellungen

In vielen Anwendungen wird nicht ein Merkmal eines Objektes gemessen, sondern mehrere (z.B. Körpergröße, Körpergewicht, Blutdruck von Personen oder Geschlecht und Rauchverhalten einer Person). Im folgenden werden ausschließlich Paare von Merkmalen X und Y eines Objektes betrachtet (Notation: (X, Y) oder (X_1, X_2) etc.). Die Ergebnisse lassen sich jedoch auch auf die Kombination von drei oder mehr Merkmalen übertragen. Das Paar (X, Y) heißt auch **bivariates Merkmal** mit Komponenten X und Y .

Bezeichnungen.

$x(1), \dots, x(J)$	Merkmalsausprägungen von Merkmal X , falls X diskret
$y(1), \dots, y(K)$	Merkmalsausprägungen von Merkmal Y , falls Y diskret
$(x(j), y(k)), j = 1, \dots, J, k = 1, \dots, K$	Merkmalsausprägungen von Merkmal (X, Y) , falls X und Y diskret
W_X, W_Y	Wertebereiche

Entsprechend zum Fall eines Merkmals können klassierte Daten betrachtet werden. Zudem sind beliebige Kombination von Merkmalstypen möglich, z.B., X nominal und Y stetig quantitativ oder X ordinal und Y diskret quantitativ.

Wie im Fall eines Merkmals werden zunächst Methoden zur Darstellung solcher Daten vorgestellt. Grundlage der Untersuchungen ist der bivariate Datensatz

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N).$$

Hierbei ist zu beachten, dass die Meßwerte (x_n, y_n) immer als Paare betrachtet werden müssen. Ausnahmen sind dann gegeben, wenn das Ziel lediglich die Analyse einer Komponente des bivariaten Merkmals (X, Y) ist.

In Analogie zur Situation eines Merkmals wird die Häufigkeitsverteilung des (diskreten) bivariaten Merkmals (X, Y) definiert, d.h. es werden die Merkmalskombinationen $(x(j), y(k))$ des Merkmals

(X, Y) ausgezählt und absolute Häufigkeiten gebildet:

$$N_{jk} = n((x(j), y(k))) = \text{Anzahl der Paare mit Ausprägung } (x(j), y(k)), \\ j = 1, \dots, J; k = 1, \dots, K.$$

Die Werte $f_{jk} = \frac{N_{jk}}{N}$ heißen Häufigkeitsverteilung. Sie werden in einer Tabelle dargestellt, der sogenannten **Kontingenztafel**:

absolute Häufigkeiten						relative Häufigkeiten					

```
> table(beton$H,beton$S)
```

	Aachen	Darmstadt	Dresden	Karlsruhe	Kassel	Leipzig
Aachen	9	0	0	0	3	0
Darmstadt	0	9	0	0	3	0
Dresden	0	0	9	0	3	0
Karlsruhe	0	0	0	8	3	0
Kassel	3	3	3	3	51	3
Leipzig	0	0	0	0	3	9

```
> table(beton$H,beton$S)/length(beton$H)
```

	Aachen	Darmstadt	Dresden	Karlsruhe	Kassel	Leipzig
Aachen	0.072	0.000	0.000	0.000	0.024	0.000
Darmstadt	0.000	0.072	0.000	0.000	0.024	0.000
Dresden	0.000	0.000	0.072	0.000	0.024	0.000
Karlsruhe	0.000	0.000	0.000	0.064	0.024	0.000
Kassel	0.024	0.024	0.024	0.024	0.408	0.024
Leipzig	0.000	0.000	0.000	0.000	0.024	0.072

7.1.3 Bemerkung

Wird statt eines diskreten Merkmals ein klassiertes stetiges Merkmal betrachtet, so werden die Ausprägungen in den Kontingenztafeln durch Klassen ersetzt. Die Häufigkeitsverteilungen werden völlig analog angegeben.

Sind beide Merkmale X und Y **stetig quantitativ**, so werden die Daten oft in einem **Streudiagramm (Scatterplot)** graphisch dargestellt. Die Beobachtungspaare (x_n, y_n) werden in einem x - y -Koordinatensystem als Punkte markiert.

7.1.4 Beispiel (Druckfestigkeit und Festbetonrohddichte, Fortsetzung von Beispiel 1.0.7)

Das Streudiagramm in Abbildung 7.1.1 wurde wie folgt erzeugt:

```
> plot(beton$Druck,beton$Festbetonrohddichte,xlab="Druckfestigkeit",
      ylab="Festbetonrohddichte")
```

7.1.5 Übungsaufgabe

Stellen Sie die Rissanzahlen aus dem Datensatz `CrackCounts.dat` aus der Übungsaufgabe 5.3.7 für die Zeitpunkte 5 und 10 sowie 10 und 18 in Streudiagrammen dar.

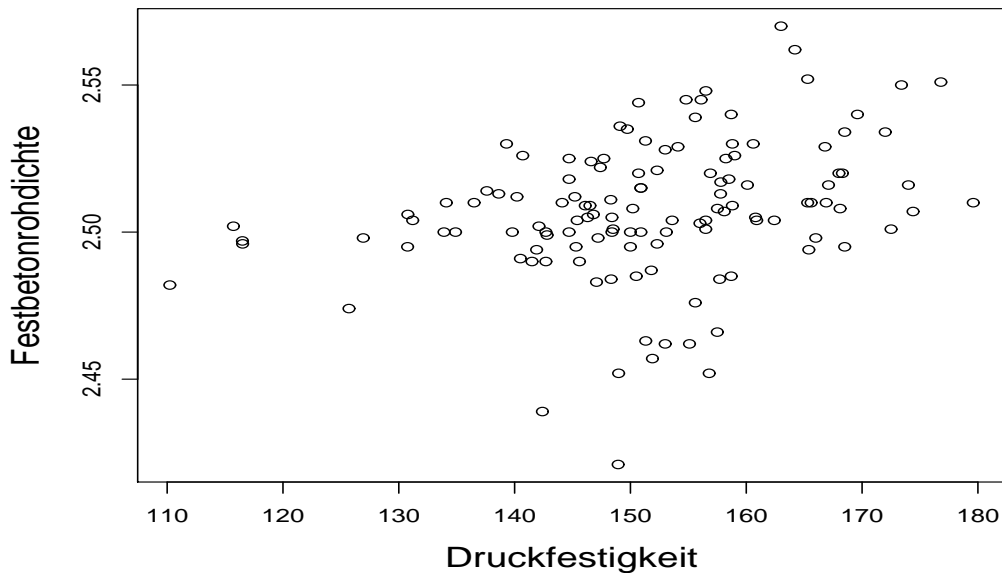


Abbildung 7.1.1: Streudiagramm für die Druckfestigkeit und die Festbetonrohndichte

7.2 Zusammenhangsmaße (Korrelationsmaße)

Im folgenden werden bivariate Merkmale (X, Y) mit beobachteten Werten

$$(x_1, y_1), \dots, (x_N, y_N)$$

betrachtet. Wurden in den vorhergehenden Abschnitten Lage- und Streuungsparameter für die Merkmalskomponenten X und Y vorgeschlagen, um die empirischen Eigenschaften dieser Merkmale mit wenigen Kennziffern zu beschreiben, so sollen nun Maße angegeben werden, die einen Zusammenhang zwischen den Komponenten X und Y quantifizieren. Dabei wird keinerlei Aussage getroffen, welcher Art dieser Zusammenhang ist (z.B. kausal oder wechselseitig).

Die Anwendbarkeit eines Zusammenhangsmaßes ist — wie bei Lage- und Streuungsparametern — abhängig von Skalen- und Merkmalstypen. Aus diesem Grund wird angenommen, dass X und Y jeweils mit dem gleichen Skalentyp gemessen werden. Dabei darf natürlich X metrisch und Y nominal skaliert sein. Es kann dann aber lediglich eine Methode für nominale Datentypen verwendet werden. Das Merkmal mit dem geringsten Skalenniveau bestimmt daher die verwendete Methode. Folgende Fälle werden unterschieden:

nominal, ordinal, quantitativ.

7.3 Zusammenhangsmaße für nominale Daten

In Analogie zu Lage- und Streuungsparametern können zur quantitativen Beschreibung nur die (absoluten/relativen) Häufigkeiten benutzt werden. Das folgende Maß beruht auf der Kontingenztafel der absoluten Häufigkeiten. Es kann auch mit den zugehörigen relativen Häufigkeiten geschrieben werden. Die Methode kann ebenfalls zur Analyse von bivariaten Merkmalen mittels klassierter Daten verwendet werden.

7.3.1 Definition (χ^2 -Größe, Kontingenzkoeffizient von Pearson)

Seien N_{jk} die absolute Häufigkeit der Ausprägung $(x(j), y(k))$ und $N_{j\bullet}$, $N_{\bullet k}$ die entsprechenden absoluten Randhäufigkeiten. Dann heißt

$$\chi^2 = \sum_{j=1}^J \sum_{k=1}^K \frac{(N_{jk} - \nu_{jk})^2}{\nu_{jk}}, \quad \text{wobei } \nu_{jk} = \frac{N_{j\bullet} \cdot N_{\bullet k}}{N},$$

χ^2 -Größe. Der (korrigierte) **Kontingenzkoeffizient von Pearson** C ist bestimmt durch

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N} \cdot \frac{\min\{J, K\}}{\min\{J, K\} - 1}}.$$

7.3.2 Bemerkung

1. Für die χ^2 -Größe gilt: $\chi^2 = N \left(\sum_{j=1}^J \sum_{k=1}^K \frac{N_{jk}^2}{N_{j\bullet} \cdot N_{\bullet k}} - 1 \right)$. Gilt ferner $J = K = 2$, so vereinfacht sich die Darstellung:

$$\chi^2 = N \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{N_{1\bullet}N_{2\bullet}N_{\bullet 1}N_{\bullet 2}}.$$

2. Es gilt: $0 \leq C \leq 1$. Dabei wird der Wert 0 als geringste Abhängigkeit (oder **Unabhängigkeit**) der Merkmale X und Y interpretiert, während der Wert 1 als Kennziffer höchster Abhängigkeit gesehen wird. Der Fall $C = 0$ ist gleichbedeutend damit, dass

$$N_{jk} = \frac{N_{j\bullet} \cdot N_{\bullet k}}{N} \quad \text{bzw.} \quad f_{jk} = f_{j\bullet} f_{\bullet k}, \quad j = 1, \dots, J; k = 1, \dots, K.$$

Die absoluten/relativen Häufigkeiten N_{jk} , f_{jk} sind in dieser Situation durch die Randhäufigkeiten vollständig bestimmt. Diese Eigenschaft wird auch als (**empirische**) **Unabhängigkeit** bezeichnet.

7.3.3 Beispiel (Zusammenhang zwischen Herstellungs- und Schleifort, Fortsetzung von Beispiel 1.0.7)

Die χ^2 -Größe für den Zusammenhang zwischen Herstellungs- und Schleifort berechnet sich zu

```
> chisq.test(table(beton$H, beton$S))$statistic
X-squared
311.4669
```

Warning message:

```
Chi-squared approximation may be incorrect in: chisq.test(table(beton$H, beton$S))
```

Daraus wird der Kontingenzkoeffizient von Pearson berechnet:

```
> chi<-chisq.test(table(beton$H,beton$S))$statistic
Warning message:
Chi-squared approximation may be incorrect in: chisq.test(table(beton$H, beton$S))
> sqrt(chi/(chi+length(beton$H))*(6/5))
X-squared
0.9253817
```

Mit 0.9253817 liegt der Kontingenzkoeffizient von Pearson nahe bei 1, so dass ein starker Zusammenhang zwischen Herstellungsort und Schleifort besteht. Das ist auch nicht verwunderlich, da die meisten Proben in Kassel hergestellt und/oder geschliffen wurden.

7.3.4 Übungsaufgabe

Bestimmen Sie die χ^2 -Größe und den Kontingenzkoeffizienten von Pearson für das Beispiel 1.0.3 (Wirksamkeit von Rostschutzmitteln) im Skript. Lesen Sie dazu die Häufigkeitstabelle mit folgenden Befehlen ein und interpretieren Sie das Ergebnis.

```
Tabelle<-matrix(c(65,103,106,74,85,47),ncol=3,byrow=T)
```

7.4 Zusammenhangsmaße für ordinale Daten

Sind beide Komponenten des bivariaten Merkmals (X, Y) ordinal skaliert, so werden anstelle der Häufigkeiten die Ordnungen der Beobachtungswerte zur Berechnung eines Zusammenhangsmaßes genutzt. Bei der Definition dieser Kenngröße wird auf die **Ränge** $R(x_n)$ bzw. $R(y_n)$ der Beobachtungen zurückgegriffen. Diese werden getrennt nach Datenreihen berechnet, d.h. $R(x_n)$ ist der Rang der Beobachtung x_n im Datensatz x_1, \dots, x_N , $R(y_n)$ ist der Rang der Beobachtung y_n im Datensatz y_1, \dots, y_N (vgl. Definition 3.4). Auf diese Weise erhält man die Rangpaare:

$$(R(x_1), R(y_1)), \dots, (R(x_N), R(y_N)).$$

Diese werden zur Definition des **Spearman'schen Rangkorrelationskoeffizienten** benutzt:

7.4.1 Definition

Seien $(x_1, y_1), \dots, (x_N, y_N)$ Beobachtungen eines bivariaten, ordinal skalierten Merkmals (X, Y) mit zugehörigen Rangpaaren $(R(x_1), R(y_1)), \dots, (R(x_N), R(y_N))$. Dann heißt

$$r_{xy}^{Sp} = \frac{\sum_{n=1}^N (R(x_n) - \bar{R}(x))(R(y_n) - \bar{R}(y))}{\sqrt{\sum_{n=1}^N (R(x_n) - \bar{R}(x))^2 \sum_{n=1}^N (R(y_n) - \bar{R}(y))^2}}$$

Rangkorrelationskoeffizient von Spearman, wobei $\bar{R}(x) = \frac{1}{N} \sum_{n=1}^N R(x_n)$ und $\bar{R}(y) = \frac{1}{N} \sum_{n=1}^N R(y_n)$ die arithmetischen Mittel der Ränge sind.

7.4.2 Bemerkung

1. Der **Spearman'sche Rangkorrelationskoeffizient** benutzt lediglich die Reihenfolge der Beobachtungswerte. Die Werte selbst sind für den Koeffizienten irrelevant.
2. Sind jeweils alle Beobachtungswerte der Datenreihe x_1, \dots, x_N und der Datenreihe y_1, \dots, y_N verschieden, so läßt sich r_{xy}^{Sp} einfacher gemäß der Formel

$$r_{xy}^{Sp} = 1 - \frac{6}{n(n^2 - 1)} \sum_{n=1}^N (R(x_n) - R(y_n))^2$$

berechnen.

3. Es gilt $-1 \leq r_{xy}^{Sp} \leq 1$, wobei $r_{xy}^{Sp} = 1$ genau dann gilt, wenn aus $x_n \prec x_m$ für die zugehörigen y -Werte folgt $y_n \prec y_m$. Dies bedeutet, dass das Ordnungsverhalten gleichsinnig in x und y ist: Wächst x , so wächst auch y . $r_{xy}^{Sp} = -1$ gilt hingegen, wenn aus $x_n \prec x_m$ für die zugehörigen y -Werte folgt $y_n \succ y_m$. In diesem Fall ist das Ordnungsverhalten gegensinnig, d.h. steigt der Wert von x , so fällt der Wert von y .

7.4.3 Beispiel (Preis und Lebensdauer eines Bauteils)

Bei einem elektronischen Bauteiltyp wurde der Preis in 10 Preisklassen erfasst, wobei 10 die höchste Preisklasse war. Ebenso wurde die Lebensdauer in 30 Klassen erfasst, wobei 30 die Klasse mit den höchsten Lebenszeiten war. Es wurden folgende Daten erhalten:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_n	8	3	9	7	10	2	6	2	1	9	10	9	5	4	7	7	3	4	4	10
y_n	28	11	30	26	29	7	25	18	7	27	24	28	23	15	26	26	5	16	23	26

Daraus können folgende Ränge ermittelt werden:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$R(x_n)$	14.0	4.5	16.0	12.0	19.0	2.5	10.0	2.5	1.0	16.0	19.0	16.0	9.0	7.0	12.0
$R(y_n)$	17.5	4.0	20.0	13.5	19.0	2.5	11.0	7.0	2.5	16.0	10.0	17.5	8.5	5.0	13.5
n	16	17	18	19	20										
$R(x_n)$	12.0	4.5	7.0	7.0	19.0										
$R(y_n)$	13.5	1.0	6.0	8.5	13.5										

Es gibt zwei Möglichkeiten den Spearmanschen Rangkorrelationskoeffizienten zu bestimmen:

```
> cor(rank(Bauteil$preis),rank(Bauteil$lebensdauer))
[1] 0.8539384
```

oder

```
> cor.test(Bauteil$preis,Bauteil$lebensdauer,method="spearman")$estimate
      rho
0.8539384
Warning message:
Cannot compute exact p-values with ties in: cor.test.default(Bauteil$preis,
Bauteil$lebensdauer, method = "spearman")
```

Mit 0.8539384 ist der Spearmansche Rangkorrelationskoeffizient nahe bei 1, so dass es einen monoton wachsenden Zusammenhang zwischen Preis und Lebensdauer gibt, d.h. ein höherer Preis bedeutet auch eine längere Lebensdauer.

7.5 Zusammenhangsmaße für quantitative Daten

Diese arithmetisches Mittel und Standardabweichung geben zwar Auskunft über Lage und Streuung der Komponenten X und Y . Sie liefern jedoch keinen Hinweis über die Zusammenhänge der Merkmale. Um den Zusammenhang zu quantifizieren, definiert man die **(empirische) Kovarianz der Merkmale X und Y** . Im Gegensatz zu den bisherigen Zusammenhangsmaßen wird nun direkt auf die gemessenen Werte zurückgegriffen.

7.5.1 Definition (Kovarianz)

Seien $(x_1, y_1), \dots, (x_N, y_N)$ Meßwerte eines bivariaten, quantitativen Merkmals (X, Y) . Dann heißt

$$s_{xy} = \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})$$

(empirische) Kovarianz der Merkmale X und Y .

7.5.2 Bemerkung

1. Die Kovarianz s_{xy} läßt sich auch berechnen gemäß

$$s_{xy} = \frac{1}{N-1} \left(\sum_{n=1}^N x_n y_n - N \cdot \bar{x} \cdot \bar{y} \right).$$

2. Es gilt: $s_{xx} = s_x^2$, $s_{yy} = s_y^2$. Zum Nachweis betrachtet man die bivariaten Datensätze $(x_1, x_1), \dots, (x_N, x_N)$ bzw. $(y_1, y_1), \dots, (y_N, y_N)$.
3. Der Wert der Kovarianz kann eine beliebige reelle Zahl sein.

Mittels der Kovarianz wird nun der **Korrelationskoeffizient von Bravais-Pearson** definiert:

7.5.3 Definition

Seien s_{xy} die (empirische) Kovarianz von X und Y , sowie s_x und s_y die Standardabweichungen von X bzw. Y . Dann heißt

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Korrelationskoeffizient von Bravais-Pearson.

7.5.4 Bemerkung

1. Es gilt: $-1 \leq r_{xy} \leq 1$, wobei $r_{xy} = -1$ bedeutet: es gibt eine **negative** Zahl $a < 0$ und eine reelle Zahl $b \in \mathbb{R}$ mit $y_n = ax_n + b$, $i = 1, \dots, n$. Es gibt also einen linearen Zusammenhang zwischen dem Merkmal X und Y derart, dass wenn X eine Einheit steigt, dann fällt Y um a

Einheiten. Insbesondere ist das Ordnungsverhalten gegensinnig. Ist hingegen $r_{xy} = 1$, so gibt es eine **positive** Zahl $a > 0$ und eine reelle Zahl $b \in \mathbb{R}$ mit $y_n = ax_n + b$, $i = 1, \dots, n$. Es gibt also einen linearen Zusammenhang zwischen dem Merkmal X und Y derart, dass wenn X eine Einheit steigt, dann steigt Y ebenfalls, und zwar um a Einheiten. Insbesondere ist das Ordnungsverhalten gleichsinnig.

2. r_{xy} ist symmetrisch in X und Y , d.h. $r_{xy} = r_{yx}$.
3. Gilt $r_{xy} = 0$, so spricht man von **(empirischer) Unkorreliertheit** der Merkmale X und Y .
4. Setzt man in der Definition von r_{xy} anstelle der Beobachtungswerte x_n bzw. y_n die entsprechenden Ränge ein, so erhält man den Spearmanschen Rangkorrelationskoeffizienten r_{xy}^{Sp} .

7.5.5 Beispiel (Druckfestigkeit und Festbetonrohddichte, Fortsetzung von Beispiel 1.0.7)

Der Korrelationskoeffizient von Bravais-Pearson zwischen Druckfestigkeit und Festbetonrohddichte berechnet sich zu:

```
> cor(beton$Druck,beton$Festbetonrohddichte)
[1] 0.3071468
```

Wegen 0.3071468 besteht eine positive Korrelation, d.h. eine höhere Druckfestigkeit tritt oft auch mit einer höheren Festbetonrohddichte auf. Allerdings ist dieser Zusammenhang nicht besonders ausgeprägt, da 0.3071468 weit von 1 entfernt ist.

7.5.6 Übungsaufgabe

Stellen Sie die prozentualen Anteile der Zementkomponenten b und d im Datensatz `SETTING.DAT` in einem Streudiagramm dar und berechnen Sie alle möglichen Zusammenhangsmaße. Interpretieren Sie die Ergebnisse.

7.6 Lineare Regression

Im vorhergehenden Abschnitt wurden Maße diskutiert, die zur Quantifizierung einer Abhängigkeit zwischen zwei Merkmalen X und Y benutzt werden. Mit diesen Methoden wurde lediglich eine gewisse Abhängigkeit der Merkmale festgestellt. Offen blieb etwa, ob es einen funktionalen Zusammenhang zwischen den Merkmalen X und Y gibt. In diesem Abschnitt wird ein Verfahren für quantitativen Merkmale vorgestellt, mit dem die funktionale Gestalt eines solchen Zusammenhangs beschrieben werden kann. Eine derartige Vorgehensweise ist etwa begründet, wenn sich aus theoretischen Überlegungen eines Fachwissenschaftlers ergibt, dass das Merkmal Y als Funktion des Merkmals X geschrieben werden kann, d.h. $Y = f(X)$ mit einer Funktion f . In der Regel wird f zumindest teilweise unbekannt sein oder von unbekannten Parametern abhängen. Ziel ist es daher, mittels eines Datensatzes $(x_1, y_1), \dots, (x_N, y_N)$ Aussagen über die Funktion f zu erhalten. Die Funktion f dient daher einerseits zur Beschreibung der Abhängigkeitsstruktur der Merkmale und andererseits zur Analyse eines Trendverhaltens. Im folgenden wird ein Zusammenhang der Form

$$Y = f(X)$$

unterstellt. Das Merkmal X heißt **Regressor** oder **erklärende Variable**, Y wird als **Regressand** bzw. als **abhängige Variable** bezeichnet. f heißt **Regressionsfunktion**. Die Werte $\hat{y}_n = f(x_n)$, $n = 1, \dots, n$, heißen **Regressionswerte**.

Im allgemeinen wird für die Regressionswerte \hat{y}_n gelten: $\hat{y}_n \neq y_n$, d.h. die Funktionswerte von f weichen an den Stellen x_n von den tatsächlich gemessenen Werten y_n ab. Dieses Phänomen kann durch (natürliche) Schwankungen in den Eigenschaften der Objekte oder durch Meßfehler und Meßungenauigkeiten hervorgerufen sein. Es wird daher versucht, die Regressionsfunktion f in einer Klasse von Funktionen *möglichst gut* anzupassen.

Im folgenden wird zunächst die allgemeine Vorgehensweise an der Klasse der linearen Funktionen illustriert. Es wird also davon ausgegangen, dass die Regressionsfunktion f linear ist, d.h.

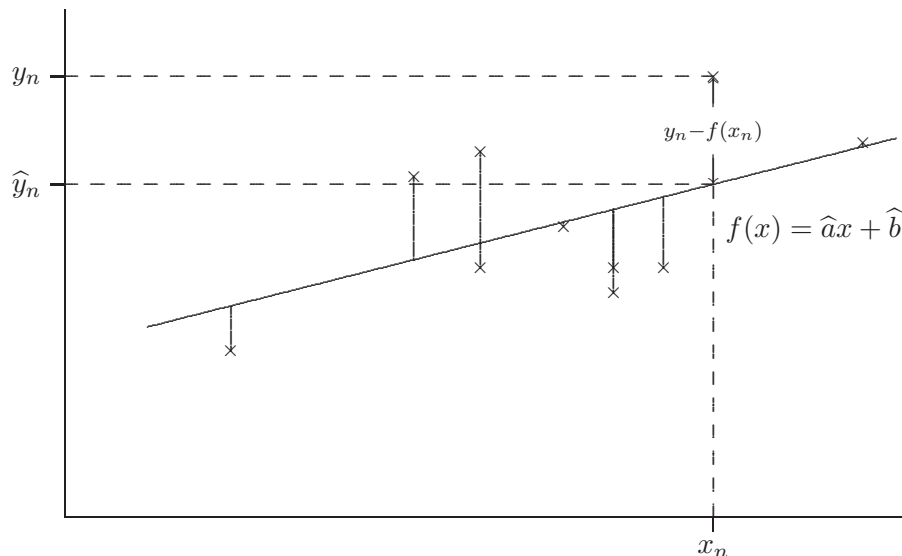
$$f(x) = ax + b, \quad x \in \mathbb{R},$$

wobei die Parameter a und b unbekannt sind. Diese müssen nun aus den Beobachtungen $(x_1, y_1), \dots, (x_N, y_N)$ möglichst gut geschätzt werden. In dieser Situation heißt die Regressionsfunktion auch **Regressionsgerade**. Als Verfahren bedient man sich der **Methode der kleinsten Quadrate**:

Die Parameter a und b der Regressionsfunktion $f(x) = ax + b$ werden berechnet als Lösung des Minimierungsproblems:

$$\min_{a,b} Q(a,b), \quad \text{wobei } Q(a,b) = \sum_{n=1}^N (y_n - ax_n - b)^2.$$

Graphisch läßt sich das Verfahren folgendermaßen interpretieren:



Mittels der Methode der kleinsten Quadrate wird daher die Summe der quadratischen Abstände $(y_n - f(x_n))^2$ bzgl. der Parameter a und b minimiert. Als Lösung erhält man die Schätzungen

$$\hat{a} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})}{\sum_{n=1}^N (x_n - \bar{x})^2} = \frac{\sum_{n=1}^N x_n y_n - n\bar{x} \cdot \bar{y}}{\sum_{n=1}^N x_n^2 - n\bar{x}^2} \quad \text{und} \quad \hat{b} = \bar{y} - \hat{a} \cdot \bar{x}.$$

Die Regressionsgerade hat daher die Gleichung: $f(x) = \hat{a}x + \hat{b}$.

7.6.1 Beispiel (Druckfestigkeit und Festbetonrohddichte, Fortsetzung von Beispiel 1.0.7)

Auch wenn der Korrelationskoeffizient von Bravais-Pearson zwischen Druckfestigkeit und Festbetonrohddichte nicht nahe bei 1 oder -1 ist, kann eine Gerade durch die Daten gelegt werden. Die Schätzungen \hat{a} für die Steigung und \hat{b} für den Intercept erhält man wie folgt:

```
> lsfit(beton$Druck, beton$Festbetonrohddichte)$coef
      Intercept              X
2.4207578924 0.0005723389
```

Mit

```
> plot(beton$Druck, beton$Festbetonrohddichte, xlab="Druckfestigkeit",
      ylab="Festbetonrohddichte")
> abline(lsfit(beton$Druck, beton$Festbetonrohddichte)$coef)
```

wird die Abbildung 7.6.1 erzeugt, bei der die geschätzte Gerade in das Streudiagramm eingezeichnet ist.

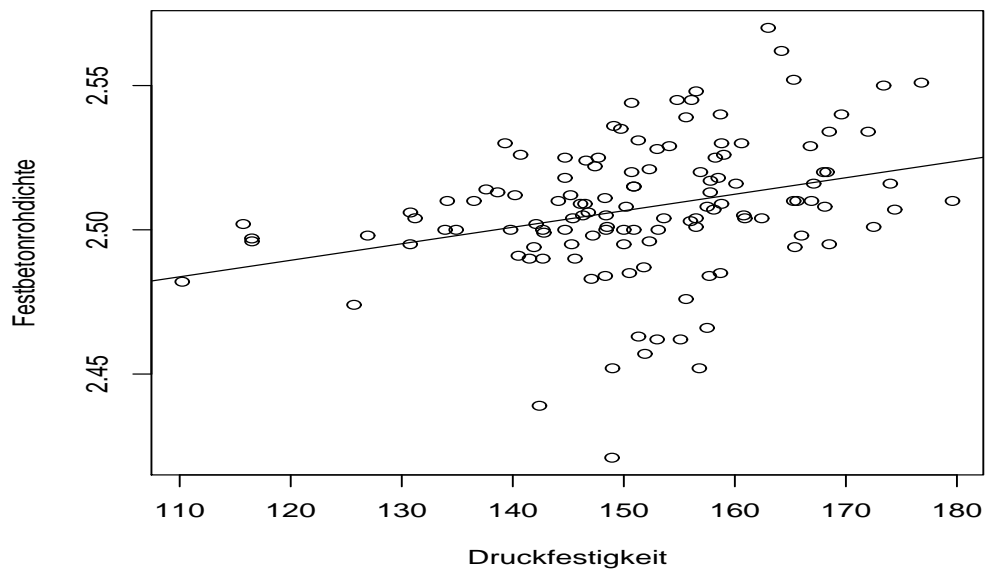


Abbildung 7.6.1: Gerade durch das Streudiagramm für die Druckfestigkeit und die Festbetonrohndichte

7.6.2 Übungsaufgabe

Berechnen Sie alle möglichen Zusammenhangsmaße der Rissanzahlen aus dem Datensatz `CrackCounts.dat` für die Zeitpunkte 5 und 10 sowie 10 und 18 und tragen Sie die Regressionsgerade in die Streudiagramme ein. Interpretieren Sie die Ergebnisse.

7.7 + Nichtlineare Regression +

Die Annahme, dass die Regressionsfunktion linear ist, wird in der Praxis häufig getroffen. Zum einen lässt sich die die Regressionsfunktion leicht berechnen und zum anderen ist ein linearer Zusammenhang auch aus praktischen Erfahrungen oft plausibel. Mithilfe der Methode der kleinsten Quadrate lässt sich aber auch für kompliziertere Regressionsfunktionen der Zusammenhang zwischen den Variablen schätzen. Ebenfalls kann die Methode der kleinsten Quadrate genutzt, wenn es mehr als eine erklärende Variable gibt.

Ist die Regressionsfunktion von der Form $f(x) = ag(x) + b$ für eine bekannte Funktion g , z.B. $g(x) = x$, $g(x) = x^2$, $g(x) = \log(x)$, ändert sich nichts. Werden die Daten entsprechend transformiert, kann die Schätzung der Parameter a und b direkt mit den Formeln der linearen Regression berechnet werden. Es muss nur mit $x^* = g(x)$ anstelle von x und anstelle der Beobachtungen x_n mit den transformierten Beobachtungen $x_n^* = g(x_n)$ gearbeitet werden.

Häufig können die Formeln der linearen Regression aber auch in anderen Fällen genutzt werden. Besteht zum Beispiel der funktionelle Zusammenhang $f(x) = b \cdot a^x$, dann ist f nicht von der Form $f(x) = ag(x) + b$, aber logarithmiert man beide Seiten, so erhält man $f^*(x) = \ln f(x) = \ln b + \ln ax = b^* + a^*x$, also wieder einen linearen Zusammenhang.

Allgemein gilt: Kann man die nichtlineare Regressionsfunktion $f(x) = \varphi(x, a, b)$ (also eine nichtlineare Funktion in x , mit unbekannten Parametern a und b) mit Transformationen g, g_0, g_1 und h so umformen, dass gilt

$$f^*(x) = h(f(x)) = g_0(b) + g_1(a) \cdot g(x),$$

dann können die Parameter $b^* = g_0(b)$ und $a^* = g_1(a)$ wieder mithilfe der Formeln der linearen Regression geschätzt werden, indem statt der Daten $(x_1, y_1), \dots, (x_N, y_N)$ die Daten $(x_1^*, y_1^*) = (g(x_1), h(y_1)), \dots, (x_N^*, y_N^*) = (g(x_N), h(y_N))$ benutzt werden.

Lassen sich aber keine solche Transformationen finden, so kann die Kleinste-Quadrat-Summen-Methode dennoch genutzt werden, um die beiden unbekannten Parameter a, b der nichtlinearen Funktion zu schätzen. Sie wird dann wie folgt definiert.

7.7.1 Definition

Gilt für die Regressionsfunktion $f(x) = \varphi(x, a, b)$, dann heißt $(\hat{a}, \hat{b})^T$ Kleinste-Quadrat-Summen-Schätzung, falls gilt

$$(\hat{a}, \hat{b})^T = \arg \min_{(a,b)^T \in \mathbb{R}^2} \sum_{n=1}^N (y_n - \varphi(x_n, a, b))^2. \quad (7.1)$$

Genauso kann die Kleinste-Quadrat-Summen-Schätzung auch für Regressionfunktionen mit mehr unbekannten Parametern oder mehr erklärende Variablen x_1, x_2, \dots definiert werden.

Die Lösung von (7.1) kann im Allgemeinen nicht explizit angegeben werden. Sie kann aber mithilfe numerischer Verfahren bestimmt werden, in R kann dazu zum Beispiel die Funktion `nls` (nonlinear least squares) benutzt werden, hierbei müssen immer auch sinnvolle Startwerte für die Schätzer übergeben werden.

Teil II

Wahrscheinlichkeitstheorie

Kapitel 8

Simulationen mit R

8.1 Zufallszahlen und deren grafische Darstellung

Für stochastische Simulationen werden Zufallszahlen gebraucht. In R können Zufallszahlen mittels verschiedener Wahrscheinlichkeitsverteilungen erzeugt werden. Während **dVerteilung** die Dichte der Verteilung ergibt, erhält man mit **rVerteilung** Zufallszahlen, die gemäß der Verteilung verteilt sind. Mit

```
> rnorm(100, mean=0, sd=1)
```

bzw.

```
> rnorm(100)
```

erzeugt man 100 Zufallszahlen, die gemäß der Standardnormalverteilung verteilt sind. Man kann diesen Zufallszahlen ein Histogramm (Säulendiagramm) erstellen, wobei R die Einteilung der Klassen automatisch vornimmt, wenn man sie nicht über das Argument **break** eingibt. Um einen Vergleich mit Abbildung 2.2.1 zu ermöglichen, sollte der Bereich der x-Achse genauso als $[-4, 4]$ mit **xlim** gewählt werden.

```
> hist(rnorm(100, mean=0, sd=1), freq=F, xlim=c(-4, 4),  
+ main="Histogramm von 100 Zufallszahlen", xlab="x")
```

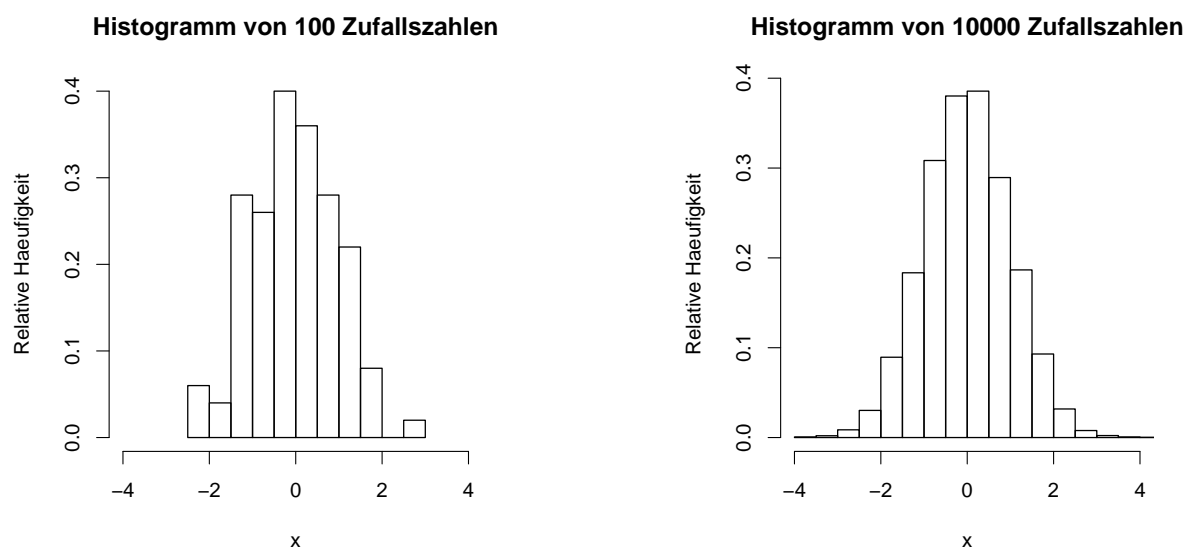


Abbildung 8.1.1: Histogramme der Zufallszahlen der Standardnormalverteilung

Kapitel 9

Statistische Kennzahlen für Zufallszahlen

Wie für Datensätze können statistische Maßzahlen für Zufallszahlen berechnet werden. Ist $x = (x_1, \dots, x_N)$ der Vektor von N Zufallszahlen so erhält man das Arithmetische Mittel

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$$

mittels der R-Funktion `mean`:

```
> x<-rnorm(100,mean=0,sd=1)
> mean(x)
[1] 0.09353081
> x<-rnorm(10000,mean=0,sd=1)
> mean(x)
[1] 0.005399328
```

Die empirische Varianz

$$s(x)^2 = \frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2$$

wird in R mit `var` berechnet und die Standardabweichung

$$s(x) = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2}$$

mit `sd`:

```
> x<-rnorm(100,mean=0,sd=1)
> var(x)
[1] 0.7260218
> sd(x)
[1] 0.8520691
```

```
> x<-rnorm(10000,mean=0,sd=1)
> var(x)
[1] 1.023311
> sd(x)
[1] 1.011588
```

Mit der logischen Verknüpfung & können auch relative Häufigkeiten in Intervallen schnell bestimmt werden. Mit folgendem Befehl wird so die relative Häufigkeit der erzeugten normalverteilten Zufallszahlen im Intervall $(0, 1]$ bestimmt.

```
> x<-rnorm(100)
> sum(x<=1 & x>0)/length(x)
[1] 0.37
```

Die R-Funktion `length` ergibt die Länge des Vektors, hier als N . Mit `x<=1 & x>0` wird ein Vektor mit logischen Werten `TRUE` und `FALSE` erzeugt, wobei `TRUE` bei der n 'ten Komponente gesetzt wird, wenn $x_n \in (0, 1]$ erfüllt ist. Werden Rechenoperationen mit logischen Werten durchgeführt (hier die Summation), wird jedes `TRUE` in 1 und jedes `FALSE` in 0 umgewandelt.

```
> x<-rnorm(10)
> x
[1] -0.34824036  0.41497294 -2.25316848  0.75629867  1.35189920  0.03752934
[7] -2.20324783 -1.40893928 -0.41448444  1.36826166
> x<=1 & x>0
[1] FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
```

9.0.1 Übungsaufgabe

Erzeugen Sie jeweils 100 und 10000 Zufallszahlen von folgenden Normalverteilungen

- (a) Normalverteilung mit $\mu=\text{mean}=0$, $\sigma=\text{sd}=1$,
- (b) Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=1$,
- (c) Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=2$.

Erstellen Sie in allen Fällen ein Histogramm und berechnen Sie in allen Fällen das arithmetische Mittel, den Median, die empirische Varianz und die Standardabweichung. Welche Schlüsse ziehen Sie daraus? Vergleichen Sie die Histogramme auch mit den grafischen Darstellungen von Übungsaufgabe 2.2.1.

9.0.2 Übungsaufgabe

Erzeugen Sie 10000 Zufallszahlen x_1, \dots, x_{10000} einer Normalverteilung mit $\mu=\text{mean}=1$, $\sigma=\text{sd}=1$. Bestimmen Sie die relativen Häufigkeiten von folgenden Fällen:

- (a) für die Zufallszahl x_n gilt $x_n \leq -2$,
- (b) für die Zufallszahl x_n gilt $x_n > -5$,
- (c) für die Zufallszahl x_n gilt $-5 < x_n \leq -2$,

(d) für die Zufallszahl x_n gilt $-2 < x_n \leq 4$,

(e) für die Zufallszahl x_n gilt $x_n \leq -2$ oder $x_n > 4$.

Welche der obigen relativen Häufigkeiten geben aufsummiert eine andere der obigen relativen Häufigkeiten?

9.1 Beschreibung des Zufalls

Aus der Abbildung 8.1.1 ist ersichtlich, dass bei 10000 standardnormalverteilten Zufallszahlen das Histogramm der Dichte der Standardnormalverteilung in Abbildung 2.2.1 viel mehr ähnelt als bei 100 Zufallszahlen. Auch ist aus Abschnitt 9 ersichtlich, dass bei 10000 Zufallszahlen das arithmetische Mittel dichter beim Parameter $\mu = \text{mean} = 0$ der Normalverteilung liegt als bei 100 Zufallszahlen. Die gleiche Beziehung gilt zwischen Standardabweichung und dem Parameter $\sigma = \text{sd} = 1$. Das ist kein Zufall. Erzeugt man immer wieder die Zufallszahlen, so wird man immer wieder diese Beobachtungen machen. Es ist ein Gesetz, das **Gesetz der großen Zahlen**. Da man früher keine Zufallszahlen schnell erzeugen konnte, hat es ziemlich lange gedauert dieses Gesetz zu finden. Früher konnte man die meiste Erfahrung am besten mit dem Würfel sammeln. Auch den Würfelwurf können wir jetzt am Computer simulieren.

9.1.1 Beispiel (Würfelwurf)

Analysiert man die physikalischen Eigenschaften eines symmetrischen Würfels, so liegt die Vermutung nahe, dass keine Seite (Ziffer) beim Wurf bevorzugt wird. Diese Eigenschaft wird beispielsweise in einer Vielzahl von Spielen genutzt. Alle Seiten (Ziffern) sind somit als *gleichwahrscheinlich* zu betrachten, und es wird daher etwa erwartet, dass in sechs Würfeln eine Sechs fällt. Diese Annahme wird später als mathematisches Modell formuliert. Zufallszahlen, die dieser Annahme folgen, können mittels Zufallszahlen erzeugt werden, die der **gleichförmigen Verteilung** auf $[0, 1]$ folgen. Bei dieser Verteilung hat jede Zahl zwischen $[0, 1]$ die gleiche Wahrscheinlichkeit. Teilt man nun dieses Intervall in 6 gleich große Intervalle auf, so hat jedes Teilintervall eine Wahrscheinlichkeit von $1/6$. Dieses Teilintervalle werden dann den Zahlen 1,2,3,4,5,6 zugeordnet, so dass jede dieser Zahlen mit der Wahrscheinlichkeit $1/6$ auftritt, so wie es beim fairen Würfel der Fall ist. Die Datei `wuerfel.asc` enthält die Funktion zur Simulation des Würfelwurfes:

```
"wuerfel" <-
function (N)
{
# Simuliert N Wuerfelwuerfe und gibt die relative Häufigkeit der Sechs aus
u<-runif(N)
u[u<=1/6]<-6
u[1/6<u & u<=2/6]<-5
u[2/6<u & u<=3/6]<-4
u[3/6<u & u<=4/6]<-3
u[4/6<u & u<=5/6]<-2
u[5/6<u & u<=1]<-1
list(Wuerfelergebnisse=u,Anteil=sum(u==6)/N)
}
```

Laden dieser Datei und der Aufruf der Funktion `wuerfel` ergibt z.B. für $N = 10$:

```
> source("wuerfel.asc")
> wuerfel(N=10)
$Wuerfelergebnisse
[1] 2 1 6 6 3 3 4 6 4 3
```

```
$Anteil
[1] 0.3
```

Der Vorteil der Funktion `wuerfel` besteht darin, dass sie schnell so umgeschrieben werden kann, dass auch ein „gefälschter“ Würfel simuliert werden kann, also ein Würfel, bei dem nicht alle Zahlen von 1 bis 6 gleichwahrscheinlich sind. Ansonsten kann der Würfelwurf auch mit der Funktion `sample` simuliert werden, hierbei muss aber dann der Anteil der Sechsen extra berechnet werden, das Argument `size` gibt dabei die Anzahl der Würfe an und `replace=TRUE` bedeutet, dass bei jedem Wurf wieder aus den Zahlen von 1 bis 6 gezogen wird:

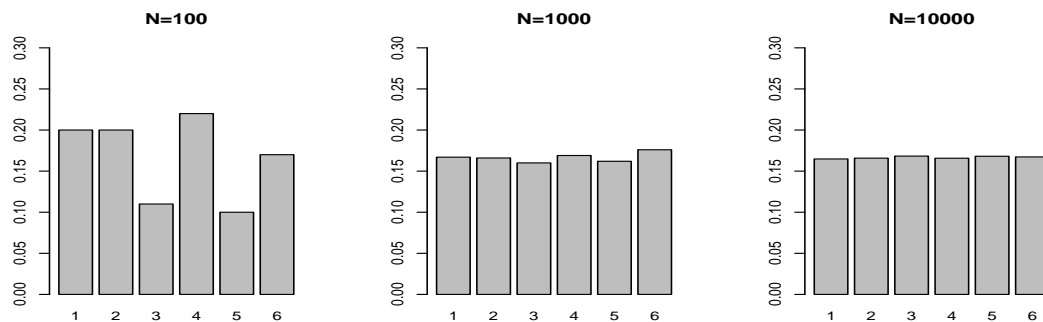
```
> wuerfelergebnisse<-sample(1:6,size=10,replace=TRUE)
> wuerfelergebnisse
[1] 1 4 4 3 4 6 3 3 6 5
> Anteil<-sum(wuerfelergebnisse==6)/length(wuerfelergebnisse)
> Anteil
[1] 0.2
```

Mit den R-Funktionen `table` und `barplot` können die absoluten Häufigkeiten der Zahlen 1,2,3,4,5,6 bei N Würfelwürfen ermittelt und anschließend grafisch in einem Balkendiagramm dargestellt werden. Um drei Grafiken in einer darzustellen, wird hier außerdem die R-Funktion `par` mit dem Argument `mfrow=c(1,3)` benutzt, wobei `c(1,3)` bedeutet, dass die Grafiken in Form einer 1×3 -Matrix angeordnet werden, d.h. eine Zeile mit 3 Spalten.

```
> par(mfrow=c(1,3))
> w<-wuerfel(100)
> t<-table(w$Wuerfelergebniss)
> barplot(t/100,main="N=100",ylim=c(0,0.3))
> w<-wuerfel(1000)
> t<-table(w$Wuerfelergebniss)
> barplot(t/1000,main="N=1000",ylim=c(0,0.3))
> w<-wuerfel(10000)
> t<-table(w$Wuerfelergebniss)
> barplot(t/10000,main="N=10000",ylim=c(0,0.3))
```

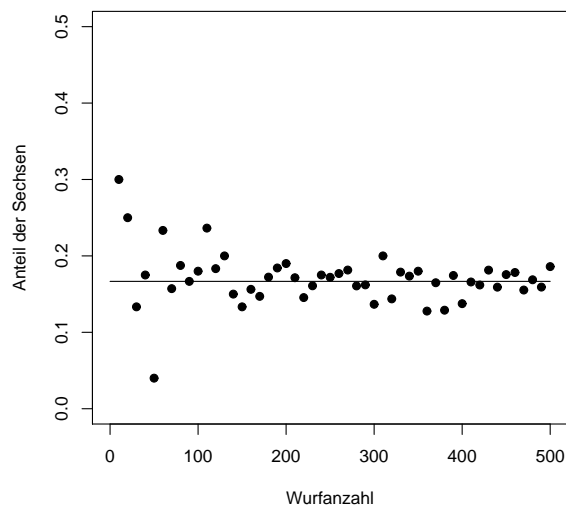
Insbesondere können wir untersuchen, wie sich der Anteil der Sechsen entwickelt, wenn die Anzahl N der Würfelwürfe immer größer wird. Die Datei `Anteil_Sechsen_Plot.asc` enthält die Funktion `Anteil.der.Sechsen.Plot`, nämlich:

```
"Anteil.der.Sechsen.Plot" <-
function ()
{
# Plottet den Anteil der Sechsen für Wurffanzahlen von 10 bis 500
plot(seq(0,500,50),seq(0,500,50)*0+1/6,xlab="Wurffanzahl",
     ylab="Anteil der Sechsen", ylim=c(0,1/2),type="l")
for(N in seq(10,500,10)){
points(N,wuerfel(N)$Anteil,pch=16)
}
}
```

Abbildung 9.1.1: Relative Häufigkeiten bei N Würfelwürfen

Diese Funktion zeichnet mit `plot` als erstes eine waagerechte Linie bei $1/6$. Dann werden in einer `for`-Schleife sukzessive mit `points` die Anteile der Sechsen als Punkte hinzugefügt. Sie wird mit folgendem Befehl aufgerufen:

```
> Anteil.der.Sechsen.Plot()
```

Abbildung 9.1.2: Relative Häufigkeiten der Sechsen mit wachsendem N

Die Abbildung 9.1.2 zeigt, dass die relative Häufigkeit der Sechsen um den Wert $1/6$ streut und dass diese Streuung um so geringer wird, desto höher die Anzahl N der Würfelwürfe wird. Es sieht so aus, dass die relative Häufigkeit der Sechsen gegen $1/6$ konvergiert. Natürlich haben wir dieses Wissen schon in die Konstruktion der Zufallszahlen, die die Würfelergebnisse simulieren, hineingesteckt. Aber mit einem richtigen Würfel würde man das gleiche Verhalten bekommen.

Das Auftreten der Sechs im Würfelwurf mit der Wahrscheinlichkeit $1/6$ zu verbinden, war historisch kein einfacher Schritt. Dieser Wert, der als Grenzwert auftritt, wird ja nie im endlichen Fall beobachtet. Im endlichen Fall treten mit den relativen Häufigkeiten, die man nur beobachten kann, alle möglichen Werte auf und, wenn N nicht durch 6 teilbar ist, tritt $1/6$ als relative Häufigkeit gar nicht auf. Noch schwieriger war es, ein mathematisches Konzept zu entwickeln, mit dem das **Gesetz der großen Zahlen** auch mathematisch bewiesen werden kann. Dieses Konzept wurde erst vollständig - in axiomatischer Weise - in den zwanziger Jahren des letzten Jahrhundert entwickelt.

In den folgenden Abschnitten wird dieser axiomatischer Ansatz zur Beschreibung des Zufalls vorgestellt und dann wird gezeigt, wie man damit arbeiten kann. Für dieses Konzept sind einige mengentheoretische Begriffe notwendig.

Kapitel 10

Mengentheoretische Grundlagen

Bei der Modellierung zufallsabhängiger Vorgänge werden zunächst die möglichen Ergebnisse angegeben, die als Konsequenz dieser Vorgänge möglich sind. Zwei Konsequenzen müssen einander ausschließen.

10.0.2 Definition

Die Menge aller möglichen **Ergebnisse** ω eines Zufallsvorgangs (Zufallsexperiments) wird **Grundmenge** (Grundraum, Ergebnisraum) genannt und mit dem griechischen Buchstaben Ω bezeichnet:

$$\Omega = \{\omega \mid \omega \text{ ist mögliches Ergebnis eines zufallsabhängigen Vorgangs}\}.$$

Eine **Menge von Ergebnissen** heißt **Ereignis**. Ereignisse werden mit großen lateinischen Buchstaben A, B, C, \dots bezeichnet. Ein Ereignis, das genau ein Ergebnis als Element besitzt, heißt **Elementarereignis**.

10.0.3 Beispiel (Würfelwurf)

Wir betrachten das Zufallsexperiment eines einfachen Würfelwurfs. Die möglichen Ergebnisse sind die Ziffern $1, \dots, 6$, d.h. die Grundmenge ist $\Omega = \{1, 2, 3, 4, 5, 6\}$. Elementarereignisse sind $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$ und $\{6\}$. Man beachte, dass Elementarereignisse Mengen sind, während Ergebnisse Elemente sind. Andere Ereignisse sind etwa:

- gerade Ziffer: $A = \{2, 4, 6\}$,
- ungerade Ziffer: $B = \{1, 3, 5\}$,
- Ziffer kleiner als 5: $C = \{1, 2, 3, 4\}$,
- Ziffer ist Summe zweier verschiedener Ziffern: $D = \{3, 4, 5, 6\}$,
- Ziffer ist Primzahl: $E = \{2, 3, 5\}$,
- Ziffer ist kleiner gleich 6: $F = \{1, 2, 3, 4, 5, 6\} = \Omega$.

10.0.4 Beispiel (Ergebnisraum, Grundraum)

Experiment	Grundraum Ω	Element ω
a) Würfelwurf	$\{1, \dots, 6\}$	gewürfelte Augenzahl
b) Münzwurf	$\{ \text{Kopf (K)}, \text{Zahl (Z)} \}$	obere Seite der geworfenen Münze
c) Roulette	$\{0, 1, \dots, 36\}$	Ergebnis des Drehrades
d) Warten auf erste „6“ beim Würfeln	$\mathbb{N} \cup \{\infty\}$	$\omega \in \mathbb{N}$: erste „6“ tritt im ω -ten Wurf auf; $\omega = \infty$: es tritt keine „6“ auf
e) Lotto „6 aus 49“	$\{(\omega_1, \dots, \omega_6) : 1 \leq \omega_1 < \dots < \omega_6 \leq 49\}$	ω_i ist Nr. der i -t größten gezogenen Kugel
f) Defektstelle einer Leitung (Gas, Telefon) der Länge 1km	$[0, 1]$	Defekt an der Stelle ω
g) Windrichtung an einer Meßstelle	$[0, 360)$	ω bezeichnet den Winkel der Windrichtung (z.B. Nord: $\omega = 0^\circ$; Ost: $\omega = 90^\circ$, etc.)
h) Wartezeit bis zum Ausfall einer Maschine	$[0, T)$	Zeit bis zum Ausfall (Höchstgrenze T)
i) Temperaturverlauf am Tage x an einer Wetterstation	$\{\omega : [0, 24) \rightarrow [-273, \infty), \omega \text{ stetige Funktion} \}$	Temperaturverlauf von 0 Uhr bis 24 Uhr in $^\circ$ Celsius, $\omega(t)$ ist Temperatur zur Zeit t .

Ereignisse sind Teilmengen des Grundraums Ω . Damit können die für Mengen definierten Verknüpfungen auf Ereignisse angewendet werden. Als Abkürzung für die Aussage *Das Ergebnis ω ist im Ereignis A enthalten* wird kurz $\omega \in A$ geschrieben. Seien A, B, C und $A_i, i \in I$, Ereignisse, wobei I eine Indexmenge ist.

- **Schnittereignis** zweier Mengen: $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ und } \omega \in B\}$
Gilt für die Ereignisse A und B : $A \cap B = \emptyset$, so heißen A und B **disjunkte** Ereignisse.
- **Schnittereignis** beliebig vieler Mengen: $\bigcap_{i \in I} A_i = \{\omega \in \Omega \mid \omega \in A_i \text{ für jedes } i \in I\}$
Die Ereignisse $A_i, i \in I$, heißen **paarweise disjunkt**, falls für jeweils zwei verschiedene Indizes $i, j \in I$ gilt: $A_i \cap A_j = \emptyset$.
- **Vereinigungseignis** zweier Mengen: $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ oder } \omega \in B\}$
- **Vereinigungseignis** beliebig vieler Mengen: $\bigcup_{i \in I} A_i = \{\omega \in \Omega \mid \text{es gibt ein } i \in I \text{ mit } \omega \in A_i\}$
- **Teilereignis**: $A \subset B$: Für jedes $\omega \in A$ gilt $\omega \in B$
- **Komplementäreignis**: $A^c = \Omega \setminus A = \{\omega \in \Omega \mid \omega \notin A\} = \{\omega \in \Omega \mid \omega \text{ nicht Element von } A\}$;
 A^c heißt **Komplement** von A (in Ω)

- **Differenzereignis:** $B \setminus A = \{\omega \in \Omega \mid \omega \in B \text{ und } \omega \notin A\} = B \cap A^c = B \setminus (A \cap B)$;
 $B \setminus A$ heißt **Komplement** von A in B

- **Distributivgesetz:**

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

- **Regeln von de Morgan:**

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

Die Sprechweisen sind in folgender Tabelle zusammengestellt.

Mathematisches Objekt	Interpretation
Ω	Grundraum, Ergebnisraum
$\omega \in \Omega$	(mögliches) Ergebnis
A	Ereignis
\mathfrak{A}	Menge aller Ereignisse
Ω	sicheres Ereignis
\emptyset	unmögliches Ereignis
$\omega \in A$	Ereignis A tritt ein
$\omega \in A^c$	Ereignis A tritt nicht ein
$\omega \in A \cup B$	Ereignis A oder Ereignis B tritt ein
$\omega \in A \cap B$	Ereignis A und Ereignis B treten ein
$A \subset B$	Eintreten von Ereignis A impliziert das Eintreten von Ereignis B
$A \cap B = \emptyset$	Ereignisse A und B schließen einander aus
$\omega \in \bigcup_{i \in I} A_i$	mindestens ein Ereignis A_i , $i \in I$, tritt ein
$\omega \in \bigcap_{i \in I} A_i$	alle Ereignisse A_i , $i \in I$, treten ein

10.0.5 Beispiel

Gegeben sei das Zufallsexperiment eines Würfelwurfs mit zwei symmetrischen Würfeln. Dann ist der Grundraum Ω gegeben durch

$$\Omega = \{(i, j) \mid i, j = 1, \dots, 6\} = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (5, 6), (6, 1), \dots, (6, 6)\}.$$

Wir betrachten nun die Ereignisse A , B , C und D mit $A \hat{=}$ *erste Ziffer gerade*, $B \hat{=}$ *Summe beider Ziffern ist gerade*, $C \hat{=}$ *erste Ziffer ist eine Sechs* und $D \hat{=}$ *beide Ziffern sind gleich*. Dann gilt:

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

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

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

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

Dann gilt etwa:

$$A \cap B = \{(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}$$

$$A \cup B = \Omega \setminus \{(1, 2), (1, 4), (1, 6), (3, 2), (3, 4), (3, 6), (5, 2), (5, 4), (5, 6)\}$$

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

$$C \setminus A = \emptyset$$

$$D \cap B = D$$

$$C \cap D = \{(6, 6)\}$$

Kapitel 11

Wahrscheinlichkeitsmaße, Wahrscheinlichkeitsräume

11.0.6 Definition (Kolmogorov-Axiome, Wahrscheinlichkeitsmaß)

Seien Ω ein Grundraum und \mathfrak{A} die Menge aller Ereignisse über Ω . Dann heißt die Abbildung

$$P : \mathfrak{A} \rightarrow [0, 1], \quad A \mapsto P(A),$$

ein **Wahrscheinlichkeitsmaß**, falls sie folgende Eigenschaften besitzt:

1. $0 \leq P(A) \leq 1$ für jedes Ereignis $A \in \mathfrak{A}$
2. $P(\Omega) = 1$
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ für alle paarweise disjunkten Ereignisse A_1, A_2, \dots

Der Wert $P(A)$ für ein Ereignis A heißt **Wahrscheinlichkeit** von A . Das Tripel $(\Omega, \mathfrak{A}, P)$ heißt **Wahrscheinlichkeitsraum**.

Die in obiger Definition angegebenen drei Eigenschaften heißen **Kolmogorov-Axiome**. Aus ihnen lassen sich einige Folgerungen für Wahrscheinlichkeitsmaße ableiten.

11.0.7 Bemerkung (Eigenschaften von Wahrscheinlichkeitsmaßen)

1. $P(\emptyset) = 0$, d.h. die Wahrscheinlichkeit des unmöglichen Ereignisses ist null.
2. Gilt $A \subset B$, so folgt $P(B \setminus A) = P(B) - P(A)$.
3. $P(A^c) = 1 - P(A)$.
4. Gilt $A \subset B$, so folgt $P(A) \leq P(B)$.
5. $P(A \cup B) = P(A) + P(B)$, falls A und B disjunkt sind.

6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ für **beliebige** Ereignisse A und B .
7. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$.
8. $P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D) - P(A \cap B) - P(B \cap C) - P(A \cap C) - P(A \cap D) - P(B \cap D) - P(C \cap D) + P(A \cap B \cap C) + P(A \cap B \cap D) + P(A \cap D \cap C) + P(D \cap B \cap C) - P(A \cap B \cap C \cap D)$.
9. Poincaré-Sylvester-Formel:

$$P\left(\bigcup_{n=1}^N A_n\right) = \sum_{m=1}^N (-1)^{m+1} \sum_{1 \leq n_1 < \dots < n_m \leq N} P(A_{n_1} \cap \dots \cap A_{n_m}).$$

11.0.8 Beispiel

Eine Maschine kann durch genau drei Ursachen ausfallen. Die Ereignisse A_1, A_2, A_3 bezeichnen die Ereignisse, dass die Maschine wegen Ursache 1, 2 bzw. 3 ausfällt. Aus Erfahrung kann ein Maschinist sagen, mit welcher Wahrscheinlichkeit eine bestimmte Ursache bzw. eine Kombination von Ursachen auftritt. Der Maschinist gibt folgende Wahrscheinlichkeiten an:

$$\begin{aligned} P(A_1) &= 0.08, & P(A_2) &= 0.12, & P(A_3) &= 0.05 \\ P(A_1 \cap A_2) &= P(A_1 \cap A_3) = P(A_2 \cap A_3) &= 0.02, & & P(A_1 \cap A_2 \cap A_3) &= 0.01 \end{aligned}$$

Daraus werden folgende Wahrscheinlichkeiten berechnet:

- die Wahrscheinlichkeit, dass die Maschine ausfällt, ist: $P(A_1 \cup A_2 \cup A_3) = 0.2$.
- die Wahrscheinlichkeit, dass die Maschine nicht ausfällt, ist:
 $P((A_1 \cup A_2 \cup A_3)^c) = 0.8$.
- die Wahrscheinlichkeit, dass die Maschine nur wegen Ursache 2 ausfällt, ist: $P(A_2 \cap (A_1 \cup A_3)^c) = 0.09$.

11.1 Diskrete Wahrscheinlichkeitsräume

Im vorhergehenden Abschnitt wurden Wahrscheinlichkeitsmaße über Grundräumen definiert. Dabei muß gemäß Definition die Wahrscheinlichkeit für jedes Ereignis A definiert werden, wobei gewisse Nebenbedingungen erfüllt sein müssen. Das aufwendige Verfahren, jedem Ereignis eine Wahrscheinlichkeit zuweisen zu müssen, läßt sich in einigen Fällen deutlich vereinfachen. Ein erstes Beispiel sind diskrete Wahrscheinlichkeitsräume.

11.1.1 Definition (diskreter Wahrscheinlichkeitsraum)

Seien $\Omega = \{\omega_1, \omega_2, \dots\}$ ein **endlicher** oder **abzählbar unendlicher** Grundraum und P ein Wahrscheinlichkeitsmaß auf Ω . Dann heißt (Ω, P) **diskreter Wahrscheinlichkeitsraum**.

11.1.2 Beispiel

Seien $\Omega = \{0, 1\}$ und $P(\{0\}) = P(\{1\}) = \frac{1}{2}$, $P(\{0, 1\}) = 1$. Dann ist (Ω, P) ein diskreter Wahrscheinlichkeitsraum.

Im Fall eines diskreten Wahrscheinlichkeitsraumes reicht es die Wahrscheinlichkeiten für die Elementarereignisse $\{\omega_i\}$ festzulegen. Diese ermöglichen dann mittels der Rechenregeln für Wahrscheinlichkeiten die Berechnung der Wahrscheinlichkeit eines beliebigen Ereignisses. Für jedes Ergebnis $\omega_i \in \Omega$ genügt es daher, eine Zahl $p_i \in [0, 1]$ anzugeben und die Wahrscheinlichkeit von $\{\omega_i\}$ durch $P(\{\omega_i\}) = p_i$ festzulegen. Dabei muß die Bedingung $\sum_{i=1}^{\infty} p_i = 1$ erfüllt sein. Für ein beliebiges Ereignis A berechnet sich die Wahrscheinlichkeit $P(A)$ dann gemäß der Vorschrift:

$$P(A) = \sum_{i: \omega_i \in A} p_i.$$

11.1.3 Beispiel (Würfelwurf)

Bei einem symmetrischen Würfel wird angenommen, dass die Wahrscheinlichkeit eine bestimmte Ziffer zu werfen gleich ist. Das bedeutet für die Wahrscheinlichkeiten der Elementarereignisse: $p_i = P(\{i\}) = \frac{1}{6}$, $i = 1, \dots, 6$. Mittels dieser Definition lassen sich die Wahrscheinlichkeiten aller anderen Ereignisse bestimmen, z.B.:

- gerade Zahl: $P(\{2, 4, 6\}) = p_2 + p_4 + p_6 = \frac{3}{6} = \frac{1}{2}$.
- Primzahl: $P(\{2, 3, 5\}) = p_2 + p_3 + p_5 = \frac{3}{6} = \frac{1}{2}$.
- Quadratzahl: $P(\{1, 4\}) = p_1 + p_4 = \frac{2}{6} = \frac{1}{3}$.
- Summe zwei verschiedener Zahlen: $P(\{3, 4, 5, 6\}) = p_3 + p_4 + p_5 + p_6 = \frac{4}{6} = \frac{2}{3}$.

11.1.4 Beispiel

Bei einem Glücksrad mit vier Feldern ist der Anteil der Felder wie folgt gegeben:

$$\omega_1 \hat{=} \text{Feld 1: } \frac{1}{5} \quad \omega_2 \hat{=} \text{Feld 2: } \frac{1}{4} \quad \omega_3 \hat{=} \text{Feld 3: } \frac{1}{2}.$$

Aus diesen Vorgaben ergibt sich wegen $P(\Omega) = 1$ für die Wahrscheinlichkeit des vierten Feldes ($\hat{=} \omega_4$):

$$\begin{aligned} p_4 &= P(\{\omega_4\}) = 1 - P(\{\omega_4\}^c) \\ &= 1 - P(\{\omega_1, \omega_2, \omega_3\}) = 1 - (P(\{\omega_1\}) + P(\{\omega_2\}) + P(\{\omega_3\})) \\ &= 1 - p_1 - p_2 - p_3 = 1 - \frac{1}{5} - \frac{1}{4} - \frac{1}{2} = \frac{1}{20}. \end{aligned}$$

Erhält man einen Gewinn, wenn ein Feld mit gerader Nummer als Ergebnis auftritt, so beträgt die Wahrscheinlichkeit zu gewinnen:

$$P(\{2, 4\}) = p_2 + p_4 = \frac{1}{4} + \frac{1}{20} = \frac{3}{10} = 0.3.$$

Umgekehrt verliert man seinen Einsatz mit Wahrscheinlichkeit 0.7.

11.2 Laplace-Räume

Eine weitere Möglichkeit Wahrscheinlichkeitsmaße zu definieren ist durch folgende Vorgehensweise gegeben. Sei $\Omega = \{\omega_1, \dots, \omega_n\}$ eine **endliche** Menge mit n Elementen. Dann wird durch die Vorschrift

$$P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n}, \quad A \text{ Ereignis, } |A| \text{ bezeichnet die Anzahl der Elemente von } A,$$

ein Wahrscheinlichkeitsmaß definiert. Das auf diese Weise definierte Wahrscheinlichkeitsmaß P wird als **Laplace-Verteilung** oder **diskrete Gleichverteilung** auf Ω bezeichnet. (Ω, P) heißt **Laplace-Raum** über Ω . Insbesondere ist ein Laplace-Raum ein diskreter Wahrscheinlichkeitsraum.

In Laplace-Räumen ist die Berechnung von Wahrscheinlichkeiten aber besonders einfach. Zunächst gilt für jedes Elementarereignis:

$$P(\{\omega_i\}) = \frac{1}{|\Omega|} = \frac{1}{n},$$

d.h. die Wahrscheinlichkeit eines jeden Elementarereignisses ist gleich $\frac{1}{n}$. Die Wahrscheinlichkeit eines beliebigen Ereignisses berechnet sich aus der Anzahl von Elementen des Ereignisses. Bezeichnet man diese Ergebnisse als *günstige Fälle*, so erhält man die Merkregel:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl möglicher Fälle}}.$$

11.2.1 Beispiel

Der einfache Würfelwurf wird modelliert durch die Grundmenge $\Omega = \{1, \dots, 6\}$ und die Laplace-Verteilung auf Ω . Für ein beliebiges Ereignis A erhält man daher die Wahrscheinlichkeit durch $P(A) = \frac{|A|}{6}$.

11.2.2 Beispiel (Ein Problem des Chevalier de Méry, 1607-1685)

Analog zu Beispiel 6.4 wird der dreifache Würfelwurf durch den Grundraum

$$\Omega = \{(i, j, k) \mid i, j, k = 1, \dots, 6\}$$

modelliert. P bezeichne die Laplace-Verteilung auf Ω , d.h., es wird angenommen, jedes Ergebnis (i, j, k) besitzt die gleiche Wahrscheinlichkeit. Wir betrachten die Ereignisse $A = \text{Augensumme } 11$ und $B = \text{Augensumme } 12$ und fragen, welches Ereignis höhere Wahrscheinlichkeit hat. Zunächst betrachten wir die Möglichkeiten, die Zahlen 11 und 12 als Summe dreier Zahlen aus der Menge $\{1, \dots, 6\}$ zu schreiben:

$$\begin{aligned} 11 : & \quad 1 + 4 + 6; 1 + 5 + 5; 2 + 3 + 6; 2 + 4 + 5; 3 + 3 + 5; 3 + 4 + 4 \\ 12 : & \quad 1 + 5 + 6; 2 + 4 + 6; 2 + 5 + 5; 3 + 3 + 6; 3 + 4 + 5; 4 + 4 + 4. \end{aligned}$$

Insgesamt hat man daher jeweils 6 Möglichkeiten, die Zahlen 11 und 12 zu kombinieren, was gleiche Wahrscheinlichkeit beider Ereignisse nahelegt. Zur Berechnung der Wahrscheinlichkeit schreiben wir

die Ereignisse als Teilmengen von Ω und erhalten:

$$\begin{aligned}
 A &= \{(i, j, k) \in \Omega \mid i + j + k = 11\} \\
 &= \{(1,4,6), (1,6,4), (4,1,6), (4,6,1), (6,1,4), (6,4,1), (1,5,5), (5,1,5), (5,5,1), (2,3,6), (2,6,3), (3,2,6), (3,6,2), (6,2,3), (6,3,2), \\
 &\quad (2,4,5), (2,5,4), (4,2,5), (4,5,2), (5,2,4), (5,4,2), (3,3,5), (3,5,3), (5,3,3), (3,4,4), (4,3,4), (4,4,3)\}, \\
 B &= \{(i, j, k) \in \Omega \mid i + j + k = 12\} \\
 &= \{(1,5,6), (1,6,5), (5,1,6), (5,6,1), (6,1,5), (6,5,1), (2,4,6), (2,6,4), (4,2,6), (4,6,2), (6,2,4), (6,4,2), (2,5,5), (5,2,5), (5,5,2), \\
 &\quad (3,3,6), (3,6,3), (6,3,3), (3,4,5), (3,5,4), (4,3,5), (4,5,3), (5,3,4), (5,4,3), (4,4,4)\}.
 \end{aligned}$$

Daher gilt $|A| = 27$ und $|B| = 25$, so dass wegen $|\Omega| = 6^3 = 216$: $P(A) = \frac{27}{216} = \frac{1}{8} = 0.125$ und $P(B) = \frac{25}{216} = 0.116$. Das Ereignis A hat somit eine höhere Eintrittswahrscheinlichkeit. De Meré hatte diesen Unterschied bemerkt, konnte ihn aber nicht nachweisen.

Liegt ein Laplace-Raum vor, so reduziert sich die Berechnung von Wahrscheinlichkeiten also auf das Abzählen von Elementen eines Ereignisses. Mittels einer aufzählenden Darstellung eines Ereignisses läßt sich die Berechnung auf simple, wenn auch oft mühsame Weise, durchführen. Das **systematische Abzählen** von Mengen ist Gegenstand der Kombinatorik, auf die hier aber nicht weiter eingegangen wird.

11.2.3 Übungsaufgabe

Ändern Sie die Funktion `wuerfel` so ab, dass die Wahrscheinlichkeit eine 6 zu würfeln $1/4$ ist und die Wahrscheinlichkeit eine 1, 2, 3, 4 oder 5 zu würfeln jeweils gleich ist. Simulieren Sie diesen verfälschten Würfel 1000 mal und vergleichen Sie den Anteil der Sechsen mit denen des unverfälschten Würfels.

Kapitel 12

Zufallsvariablen und deren Verteilungen

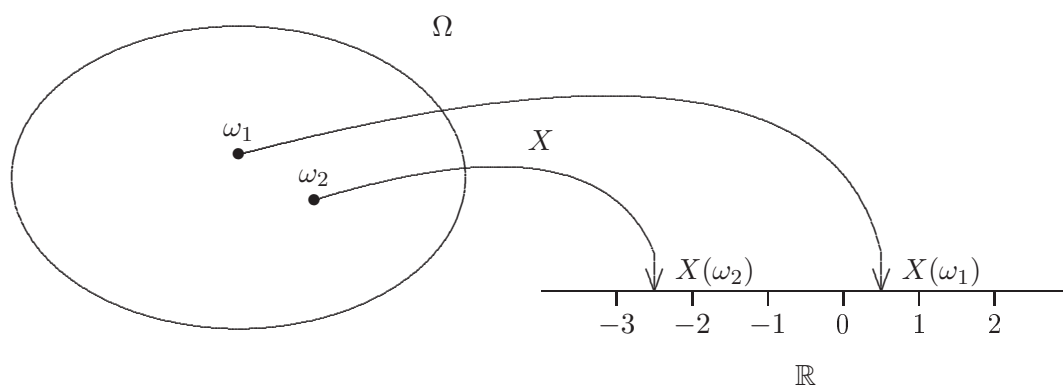
In vielen Situationen sind die Elemente ω des Grundraumes Ω eines Zufallsexperiments selbst nicht von Interesse, sondern lediglich eine bestimmte Eigenschaft der Elemente ω , wie etwa die Brenndauer einer Glühbirne, die Summe der Augenzahlen zweier Würfel, die Anzahl von Sechsen bei 1000 Würfelwürfen.

12.0.4 Beispiel

Beim zweifachen Würfelwurf ist der Grundraum gegeben durch $\Omega = \{(i, j) \mid i, j \in \{1, \dots, 6\}\}$. Ist die Augensumme beider Würfel von Interesse, so läßt sich dies durch eine **Abbildung** X beschreiben:

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ (i, j) &\mapsto i + j. \end{aligned}$$

Eine Abbildung $X : \Omega \longrightarrow \mathbb{R}$ wird im folgenden als **Zufallsvariable** bezeichnet. Jedem Element $\omega \in \Omega$ wird eine reelle Zahl als Funktionswert zugeordnet. Dieser Wert $X(\omega)$ heißt **Realisation** von X .



Beispiele für Zufallsvariablen sind etwa:

- Münzwurf: $\Omega = \{\text{Kopf}, \text{Zahl}\}$, $X : \Omega \rightarrow \mathbb{R}$ definiert durch $X(\text{Kopf}) = 0$, $X(\text{Zahl}) = 1$.
- zweifacher Würfelwurf: $X = \text{Augensumme}$.

- Y = Anzahl Sechsen bei 1000 Würfelwürfen
- Z = Anzahl Kunden eines Geschäftes an einem Tag.
- W = Reifendruck.
- V = Laufzeit einer Maschine.
- U = Abweichung einer Füllmenge von einem Sollwert.

12.0.5 Bemerkung

In den obigen Beispielen wurde meist der zugrunde liegende Grundraum Ω nicht angegeben, sondern lediglich die Abbildung spezifiziert. Es wird also nur angegeben, welche Größe untersucht werden soll (z.B. Reifendruck, Laufzeit, Abweichung, etc.) Diese Vorgehensweise ist typisch, da im allgemeinen die **Realisationen** einer Zufallsgröße X von Interesse sind. Der Grundraum Ω wird i.a. vernachlässigt. Die Verteilungsannahmen werden bei diesem Vorgehen unmittelbar an die Zufallsvariable X gestellt.

Oft sind auch **mehrdimensionale Zufallsvariablen** von Interesse (z.B. Laufzeiten zweier Maschinen). Mehrdimensionale Zufallsvariablen werden als Vektoren geschrieben, z.B.:

$$(X, Y), \quad (U, V, W), \quad (X_1, \dots, X_n).$$

(X_1, \dots, X_n) hat dann Werte $(x_1, \dots, x_n) \in \mathbb{R}^n$ und heißt **n -dimensionale Zufallsvariable**. Von besonderer Bedeutung ist im folgenden der Fall $n = 2$. Für diesen wird i.a. die Schreibweise (X, Y) verwendet.

12.1 Verteilung eindimensionaler Zufallsvariablen

Die **Wahrscheinlichkeitsverteilung** oder kurz **Verteilung** einer Zufallsvariablen X ist definiert durch

$$P^X(B) = P(X \in B) = P(\{\omega \in \Omega \mid X(\omega) \in B\}), \quad B \subset \mathbb{R}.$$

Diese Festlegung der Verteilung P^X von X über die Ereignisse B ist i.a. zu aufwendig. Zur Beschreibung der Verteilung P^X reicht es aber die Wahrscheinlichkeit von Intervallen des Typs $(-\infty, x]$, $x \in \mathbb{R}$, anzugeben. Diese legen die Verteilung P^X eindeutig fest und haben eine interessante Interpretation. $P^X((-\infty, x]) = P(X \leq x)$ ist die Wahrscheinlichkeit dafür, dass die Zufallsvariable X den Wert x nicht übersteigt. Mittels dieser Intervalle wird die **Verteilungsfunktion** der Zufallsvariablen X definiert.

12.1.1 Definition (Verteilungsfunktion)

Sei X eine Zufallsvariable über einem Grundraum Ω . Dann heißt die Funktion $F = F^X : \mathbb{R} \rightarrow [0, 1]$ definiert durch

$$F(x) = P^X((-\infty, x]) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}), \quad x \in \mathbb{R}.$$

Verteilungsfunktion von X .

12.1.2 Bemerkung

Die Verteilungsfunktion ist das wahrscheinlichkeitstheoretische Pendant zur **empirischen Verteilungsfunktion**. Es besitzt daher auch ähnliche Eigenschaften:

- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$
- F ist monoton wachsend, d.h. für $x < y$ gilt: $F(x) \leq F(y)$
- F ist rechtsseitig stetig, d.h. $\lim_{x \downarrow z+} F(x) = F(z)$
- $P(a < X \leq b) = F(b) - F(a)$ für $a < b$
- $P(X > a) = 1 - F(a)$

Wie bei Merkmalen unterscheidet man zwischen diskreten und stetigen Zufallsvariablen. In der ersten Situation ist die zugehörige Verteilungsfunktion F eine **Treppenfunktion** mit höchstens abzählbar vielen Sprungstellen, in der zweiten Situation ist die Verteilungsfunktion der Zufallsvariablen X stetig. In Analogie bezeichnet man die zugehörigen Verteilungen als diskret bzw. stetig.

12.1.3 Spezialfall (Diskrete Zufallsvariablen)

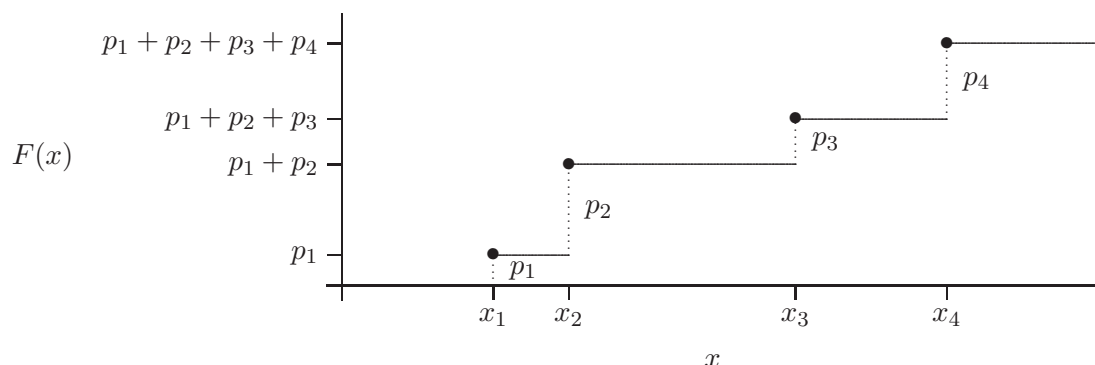
F ist eine Treppenfunktion mit höchstens abzählbar vielen Sprungstellen, so dass F zwischen zwei Sprungstellen insbesondere konstant ist. Im folgenden sei angenommen, dass es eine kleinste Sprungstelle gebe. Betrachten wir nun die Menge der Sprungstellen $\{x_1, x_2, x_3, \dots\}$, wobei $x_1 < x_2 < x_3 < \dots$, und definieren

$$p_i = P(X = x_i) = P(\{\omega \in \Omega \mid X(\omega) = x_i\}),$$

so erhält man folgende Darstellung der Verteilungsfunktion:

$$F(x) = P(X \leq x) = \sum_{i: x_i \leq x} p_i.$$

Insbesondere erhält man die (auch für empirische Verteilungsfunktionen) charakteristische Darstellung:



Daraus folgen für eine diskreten Zufallsvariable die Eigenschaften:

1. $P(X = x_i) = F(x_i) - F(x_{i-1}) = p_i, i \geq 2.$
2. $P(X = x) = 0, x \notin \{x_1, x_2, \dots\}.$

Eine diskrete Zufallsvariable hat also höchstens abzählbar viele Werte, die positive Wahrscheinlichkeit besitzen. Die mittels der Werte p_1, p_2, \dots definierte Funktion

$$p : \mathbb{R} \rightarrow [0, 1], \quad p(x) = \begin{cases} p_i, & x = x_i, i \in \mathbb{N}, \\ 0, & x \neq x_i. \end{cases}$$

heißt **Zähldichte** von X . Durch Angabe dieser Zähldichte ist die Verteilung einer diskreten Zufallsvariable eindeutig festgelegt. Im nächsten Abschnitt werden einige wichtige diskrete Verteilungen vorgestellt.

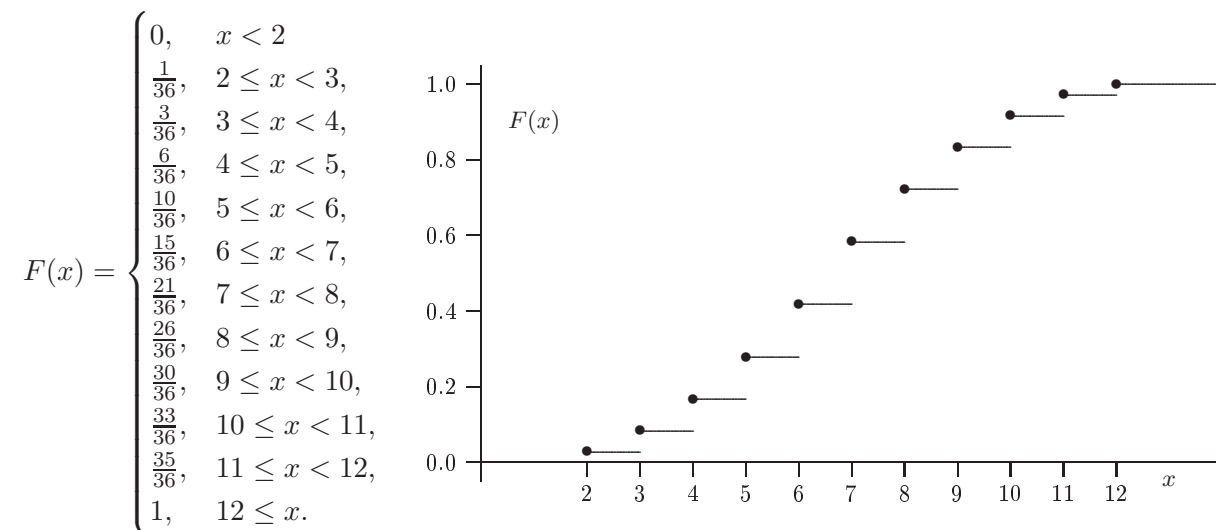
12.1.4 Beispiel

Im Experiment des zweifachen Würfelwurfs wird die Zufallsvariable $X = \text{Augensumme beider Würfel}$ betrachtet: $X((i, j)) = i + j$. Aus dem Würfelexperiment ist klar, dass als Augensumme nur die Zahlen $2, \dots, 12$ (mit positiver Wahrscheinlichkeit) auftreten. Wir berechnen daher die Wahrscheinlichkeiten $P(X = k)$, $k = 2, \dots, 12$.

Für $k = 2$ gilt: $P(X = 2) = P(\{(1, 1)\}) = \frac{1}{36}$. Für die Augensumme 4 erhält man $P(X = 4) = P(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36}$. Insgesamt erhält man:

k	2	3	4	5	6	7	8	9	10	11	12
$P(X = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Dies liefert die Verteilungsfunktion:



12.1.5 Beispiel

Ein Münzwurf werde beliebig oft unabhängig durchgeführt, wobei die Wahrscheinlichkeit für Kopf durch eine Zahl $p \in (0, 1)$ gegeben sei. Die Zufallsvariable X beschreibe die Wartezeit bis zum ersten Mal *Kopf* auftritt. Da das erste Mal *Kopf* in einem beliebigen Wurf $1, 2, 3, \dots$ auftreten kann, sind für die Wartezeit bis zum Eintritt dieses Ereignisses die Werte $0, 1, 2, \dots$ möglich. Bezeichne A_j , $j \in \mathbb{N}$, das Ereignis, dass im j -ten Wurf *Zahl* auftritt. Dann gilt nach Voraussetzung A_1, A_2, \dots sind (gemeinsam) stochastisch unabhängig und $P(A_j) = 1 - p$, $j \in \mathbb{N}$. Sei k die Wartezeit, dann gilt für

- $k = 0$: $P(X = 0) = P(A_1^c) = p$

- $k = 1$: $P(X = 1) = P(A_1 \cap A_2^c) = P(A_1) \cdot P(A_2^c) = (1 - p)p$
- $k \in \mathbb{N}$: $P(X = k) = P(A_1 \cap \dots \cap A_k \cap A_{k+1}^c) = \left(\prod_{j=1}^k P(A_j) \right) P(A_{k+1}^c) = (1 - p)^k p$

Insgesamt gilt also $P(X = k) = p(1 - p)^k$, $k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Diese Verteilung heißt **geometrische Verteilung** mit Parameter p .

12.1.6 Spezialfall (Stetige Zufallsvariablen)

Eine Zufallsvariable X heißt stetig, wenn ihre Verteilungsfunktion F eine stetige Funktion ist. Entsprechen diskrete Zufallsvariablen den quantitativ-diskreten Merkmalen im Bereich der deskriptiven Statistik, so bilden stetige Zufallsvariablen das Pendant zu quantitativ-stetigen Merkmalen.

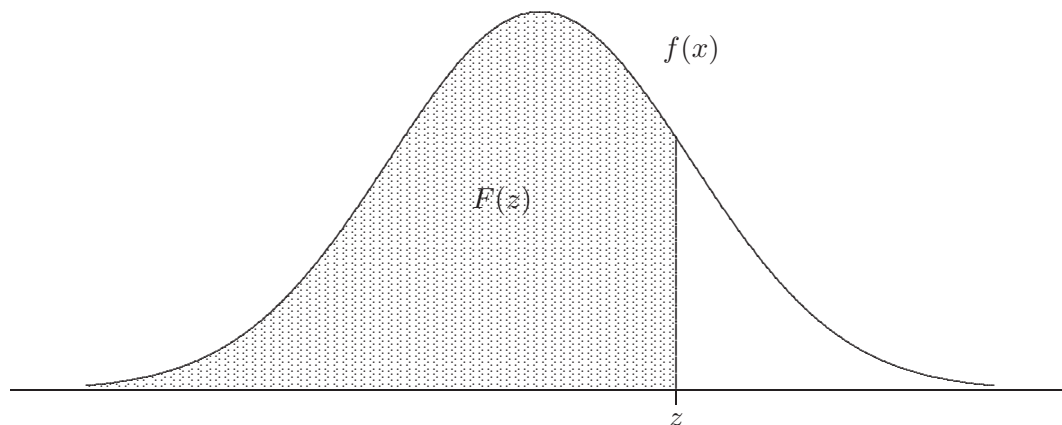
Im folgenden werden zur Vereinfachung nur solche Verteilungsfunktionen betrachtet, die sich mittels eines **Riemann-Integrals** darstellen lassen:

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R},$$

wobei f eine reellwertige, Riemann-integrierbare Funktion ist mit

$$f(x) \geq 0, \quad x \in \mathbb{R}, \quad \text{und} \quad \int_{-\infty}^{\infty} f(t) dt = 1.$$

f wird als **Dichtefunktion** von X bzw. kurz als **Dichte** von X bezeichnet. Der Wert der Verteilungsfunktion F an der Stelle z ist somit der Wert der Fläche, die durch die x -Achse und den Graphen der Funktion f eingeschlossen wird und nach rechts durch die senkrechte Gerade durch den Punkt $(z, 0)$ begrenzt ist (schattierte Fläche):



Für die hier betrachteten stetigen Verteilungsfunktionen gelten folgende Eigenschaften:

1. $P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = F(b) - F(a)$, $a \leq b$, d.h. es spielt keine Rolle, ob die Grenzen dazu genommen werden oder nicht!
2. $P(X = x) = 0$ für alle $x \in \mathbb{R}$, d.h. jeder Wert $x \in \mathbb{R}$ tritt mit Wahrscheinlichkeit 0 auf!
3. $F'(x) = f(x)$, d.h. die Ableitung der Verteilungsfunktion ist die Dichtefunktion.

12.1.7 Beispiel

Von der Wartezeit X [in min] eines Kunden an einer Supermarktkasse werde angenommen, dass die Verteilung P^X die Verteilungsfunktion

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\frac{x}{2}}, & x \geq 0, \end{cases}$$

habe. Dann gilt für die Wahrscheinlichkeit, dass der Kunde höchstens 5 min warten muß: $P(X \leq 5) = F(5) = 1 - e^{-2.5} = 0.918$, d.h. der Kunde muß mit einer Wahrscheinlichkeit von höchstens 8.2% länger als 5 min warten. Die Wahrscheinlichkeit, dass er zwischen 3 und 6 Minuten anstehen muß, beträgt:

$$P(3 \leq X \leq 6) = F(6) - F(3) = e^{-1.5} - e^{-3} = 0.173.$$

Für die Dichtefunktion f gilt: $f(x) = 0$, falls $x < 0$, und $f(x) = \frac{1}{2}e^{-\frac{x}{2}}$, $x \geq 0$.

Weitere Verteilungsbeispiele sind im nächsten Abschnitt angegeben.

12.1.8 Übungsaufgabe

Simulieren Sie einen einfachen Würfelwurf mit $N = 1000$ und erstellen Sie dafür mithilfe von `plot(ecdf(wuerfel(1000)$Wuerfelergebnisse))` die empirische Verteilungsfunktion der Zufallszahlen, die die theoretische Verteilungsfunktion approximiert. Ändern Sie die Funktion `wuerfel` so ab, dass diese Zufallszahlen gemäß der Zufallsvariable X eines verfälschten Würfels ergibt mit

$$P(X = 6) = \frac{1}{5}, \quad P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5).$$

Simulieren Sie diesen verfälschten Würfel auch 1000 mal und stellen Sie die empirische Verteilungsfunktion dieser Zufallszahlen wie für den unverfälschten Würfel dar.

12.1.9 Übungsaufgabe

Die Verteilungsfunktion F der Zufallsvariablen X sei folgendermaßen gegeben:

$$F(x) = \begin{cases} 0, & \text{für } x < 2, \\ 1 - \frac{1}{k} & \text{für } k \leq x < k+1, \quad k \in \mathbb{N} \setminus \{0, 1\}. \end{cases}$$

Skizzieren Sie diese Verteilungsfunktion und bestimmen Sie aus der Verteilungsfunktion folgende Wahrscheinlichkeiten:

- (a) $P(X < 2)$, (b) $P(X \leq 2)$, (c) $P(X \leq 3)$,
- (d) $P(X > 3)$, (e) $P(2 < X \leq 3)$, (f) $P(X \notin (2, 3])$,
- (g) $P(2 < X < 3)$, (h) $P(X \in [2, 3])$.

12.2 Verteilung mehrdimensionaler Zufallsvariablen

Wir betrachten im folgenden den Fall zweier Zufallsvariablen X und Y . Der Zufallsvektor (X, Y) nehme Werte im \mathbb{R}^2 an, d.h. die Realisationen von (X, Y) sind Paare (x, y) mit $x, y \in \mathbb{R}$. Die

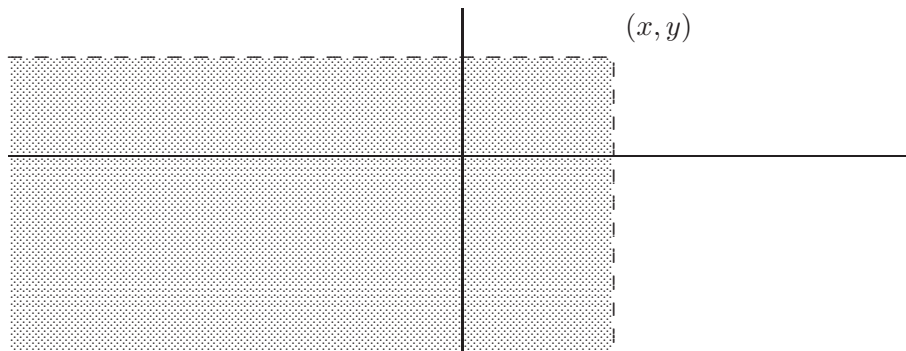
Verteilung $P^{(X,Y)}$ wird analog zum eindimensionalen Fall definiert:

$$P^{(X,Y)}(B) = P((X, Y) \in B) = P(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in B\}), \quad B \subset \mathbb{R}^2.$$

Wie im eindimensionalen Fall kann eine Verteilungsfunktion definiert werden:

$$F(x, y) = P(X \leq x, Y \leq y) = P(\{\omega \in \Omega \mid X(\omega) \leq x, Y(\omega) \leq y\}), \quad x, y \in \mathbb{R}.$$

Der Wert der Verteilungsfunktion F an der Stelle (x, y) ist die Wahrscheinlichkeit, dass die Werte des Zufallsvektors (X, Y) in den folgenden schattierten Bereich fallen:



Diese hat folgende Eigenschaften:

1. $\lim_{x, y \rightarrow -\infty} F(x, y) = 0$, $\lim_{x, y \rightarrow \infty} F(x, y) = 1$
2. $\lim_{y \rightarrow \infty} F(x, y) = F^X(x)$, $\lim_{x \rightarrow \infty} F(x, y) = F^Y(y)$
3. F ist monoton wachsend in jeder Komponente, d.h. $F(x_1, y) \leq F(x_2, y)$ für $x_1 < x_2$ und $F(x, y_1) \leq F(x, y_2)$ für $y_1 < y_2$.

Die Verteilungen F^X und F^Y heißen **Randverteilungen** von $P^{(X,Y)}$ bzw. $F = F^{(X,Y)}$. Wie im eindimensionalen Fall werden diskrete und stetige Zufallsvariablen unterschieden. Es werden nur solche Fälle betrachtet, in denen X und Y beide diskret bzw. beide stetig sind. Entsprechend zum eindimensionalen Fall können die Verteilungen durch (Zähl-) Dichten beschrieben werden:

- **Diskreter Fall:** $p_{ij} = P(X = x_i, Y = y_j)$, wobei $x_1 < x_2 < \dots$ und $y_1 < y_2 < \dots$ die Sprungstellen der Verteilungsfunktionen F^X bzw. F^Y sind, so dass

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{i, j: x_i \leq x, y_j \leq y} p_{ij}.$$

Die Zahlen p_{ij} heißen Zähldichten von $P^{(X,Y)}$ bzw. von (X, Y) . Die (Rand-) Zähldichten von X bzw. Y können berechnet werden durch:

$$p_i^X = \sum_j p_{ij}, \quad p_j^Y = \sum_i p_{ij}.$$

• **Stetiger Fall:**

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds,$$

wobei f eine Riemann-integrierbare Funktion ist mit

$$f(s, t) \geq 0, \quad \text{für alle } s, t \in \mathbb{R} \quad \text{und} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s, t) dt ds = 1.$$

Für die (Rand-) Dichten von X bzw. Y gelten:

$$f^X(x) = \int_{-\infty}^{\infty} f(x, t) dt, \quad f^Y(y) = \int_{-\infty}^{\infty} f(s, y) ds.$$

12.2.1 Beispiel

Der Zufallsvektor (X, Y) habe eine diskrete Verteilung mit

$$p_{ij} = \binom{n}{j} \frac{1}{2^{n+i+1}}, \quad i \in \mathbb{N}_0, j \in \{0, \dots, n\}.$$

Dann gilt für die Randdichten von X bzw. Y :

$$p_i^X = \sum_{j=0}^n \binom{n}{j} \frac{1}{2^{n+i+1}} = \frac{1}{2^{i+1}} \sum_{j=0}^n \binom{n}{j} \frac{1}{2^n} = \frac{1}{2^{i+1}}, \quad i \in \mathbb{N}_0,$$

$$p_j^Y = \sum_{i=0}^{\infty} \binom{n}{j} \frac{1}{2^{n+i+1}} = \binom{n}{j} \frac{1}{2^{n+1}} \sum_{i=0}^{\infty} \frac{1}{2^i} = \binom{n}{j} \frac{1}{2^n}, \quad j = 0, \dots, n.$$

Die Verteilung von X ist eine sogenannte **geometrische Verteilung** mit Parameter $\frac{1}{2}$, während die Verteilung von Y die **Binomialverteilung** mit Parametern n und $\frac{1}{2}$ ist.

12.2.2 Beispiel

Die Zufallsvariable (X, Y) habe die Verteilungsfunktion:

$$F(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-(y-x)}), & 0 \leq x \leq y, \\ 1 - e^{-y}, & 0 \leq y < x, \\ 0, & \text{sonst.} \end{cases}$$

Für die Randverteilungen gilt:

$$F^X(x) = \lim_{y \rightarrow \infty} F(x, y) = 1 - e^{-x}, \quad x \geq 0, \quad F^X(x) = 0, \quad x < 0,$$

$$F^Y(y) = \lim_{x \rightarrow \infty} F(x, y) = 1 - e^{-y}, \quad y \geq 0, \quad F^Y(y) = 0, \quad y < 0.$$

X und Y haben also die gleiche Verteilung. Diese heißt **Exponentialverteilung** mit Parameter 1 oder auch Standardexponentialverteilung.

Kapitel 13

Wichtige Wahrscheinlichkeitsverteilungen

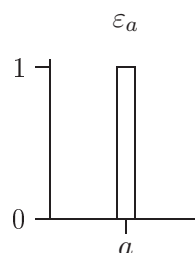
In diesem Abschnitt werden einige wichtige Wahrscheinlichkeitsverteilungen vorgestellt. Dabei wird für die Aussage *X ist verteilt gemäß einer Verteilung P* die Schreibweise $X \sim P$ verwendet. Die Verteilungen werden dargestellt mittels ihrer (Zähl-) Dichte. Zum Teil werden auch die Verteilungsfunktionen dargestellt. In *R* erhält man (Zähl-) Dichten mittels `ddistname`, Verteilungsfunktionen mittels `pdistname` und Zufallszahlen der Verteilung mittels `rdistname`, wobei *distname* für den Verteilungsnamen steht. In den aufgeführten Verteilungen wird nur dieser Verteilungsname in *R* aufgeführt.

13.1 Diskrete Wahrscheinlichkeitsverteilungen

Der **Träger** T_X einer **diskreten Verteilung** P^X ist die Menge der Werte, die positive Wahrscheinlichkeit besitzen.

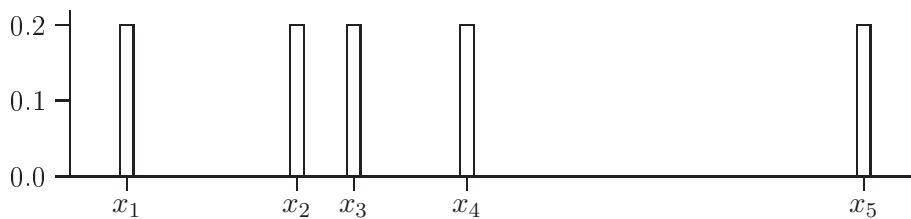
13.1.1 Beispiel (Einpunktverteilung ε_a)

- **Zähldichte:** $P(X = a) = 1$ für ein $a \in \mathbb{R}$
- **Träger:** $T_X = \{a\}$
- **Diagramm der Zähldichte:**

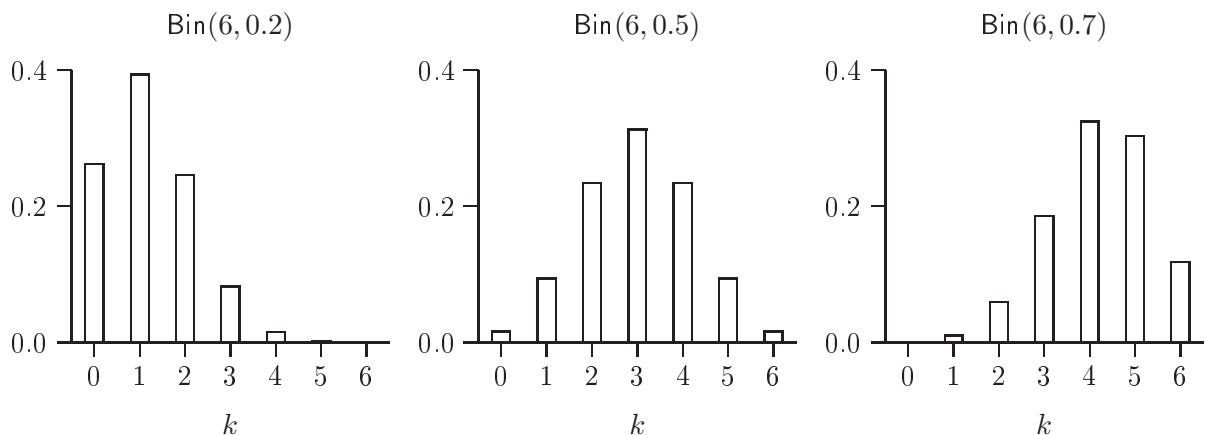


13.1.2 Beispiel (Diskrete Gleichverteilung $G(x_1, \dots, x_n)$)

- **Zähldichte:** $P(X = x_k) = \frac{1}{n}, x_1 < \dots < x_n, n \in \mathbb{N}$
- **Träger:** $T_X = \{x_1, \dots, x_n\}$
- **Diagramm der Zähldichte:** $G(x_1, \dots, x_5)$

**13.1.3 Beispiel (Binomialverteilung $\text{Bin}(n, p)$)**

- **Verteilungsname in R:** `binom` mit Argumenten `size` für n und `prob` für p
- **Zähldichte:** $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; n \in \mathbb{N}, p \in (0, 1)$
- **Träger:** $T_X = \{0, 1, \dots, n\}$
- **Diagramme von Zähldichten:**

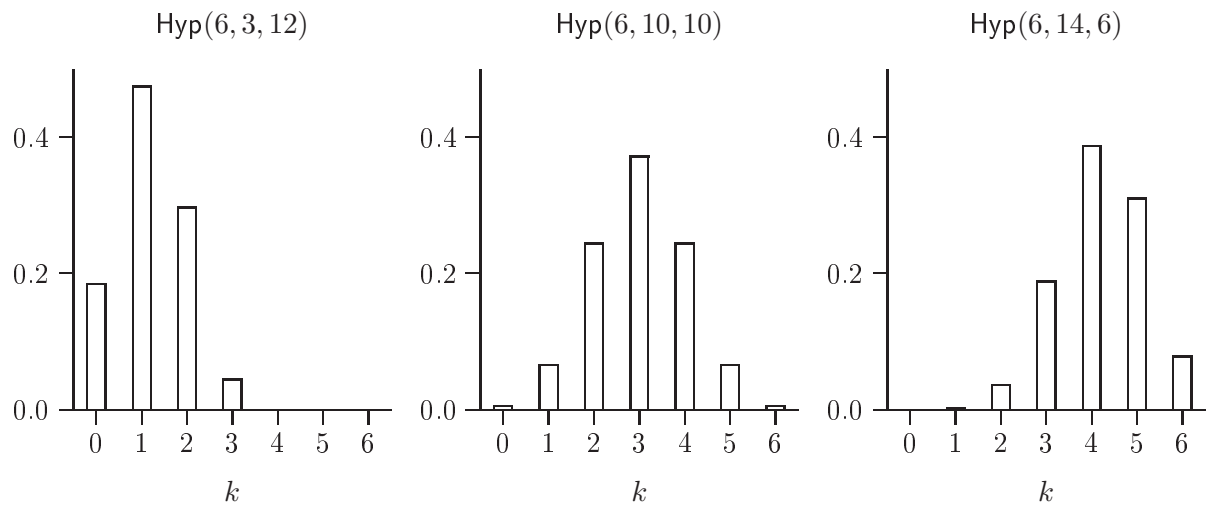


- **Bemerkung:** Für $n = 1$ heißt die Binomialverteilung auch **Bernoulli-Verteilung** oder **Zweipunktverteilung** mit Parameter p .
- **Anwendung:** Häufigkeit eines Zustandes, wenn nur zwei Zustände auftreten können, z.B. Gut-Schlecht-Prüfung (Ziehen mit Zurücklegen)

13.1.4 Beispiel (Hypergeometrische Verteilung $\text{Hyp}(n, r, s)$)

- **Verteilungsname in R:** `hyper` mit Argumenten `k` für n , `m` für r und `n` für s
- **Zähldichte:** $P(X = k) = \frac{\binom{r}{k} \binom{s}{n-k}}{\binom{r+s}{n}}, n, r, s \in \mathbb{N}, n \leq r + s$, wobei $P(X = k) = 0$, falls $k > r$ oder $n - k > s$
- **Träger:** $T_X = \{\max\{0, n - s\}, \dots, \min\{r, n\}\}$

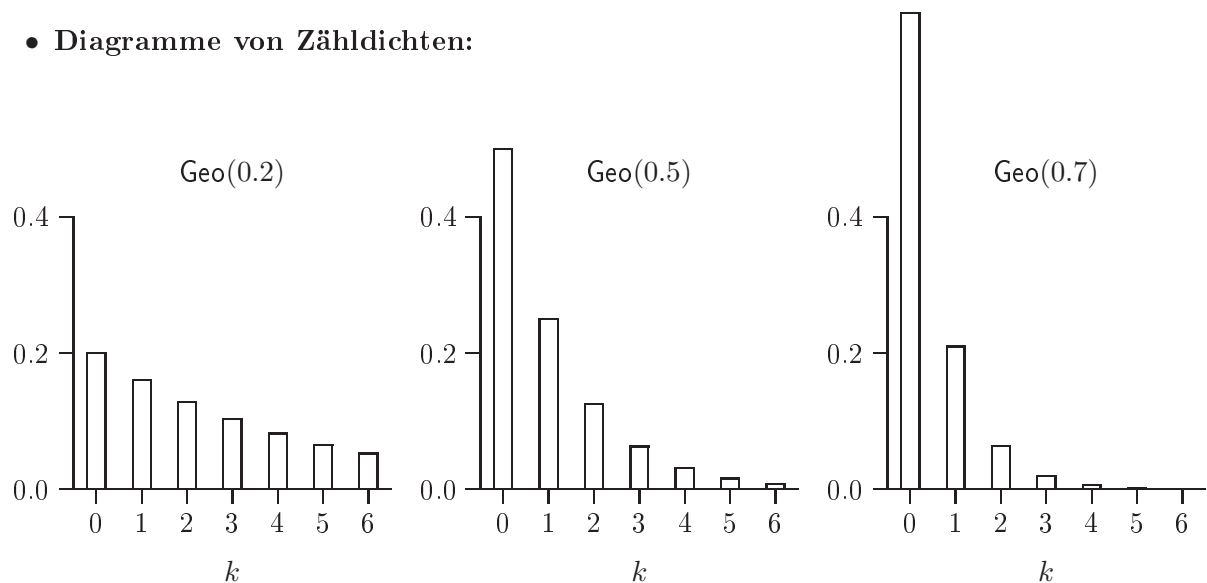
- **Diagramme von Zähldichten:**



- **Anwendung:** Häufigkeit eines Zustandes, wenn nur zwei Zustände auftreten können, z.B. Gut-Schlecht-Prüfung (n -maliges Ziehen aus einer Urne mit r roten und s schwarzen Kugeln ohne Zurücklegen)

13.1.5 Beispiel (Geometrische Verteilung $\text{Geo}(p)$)

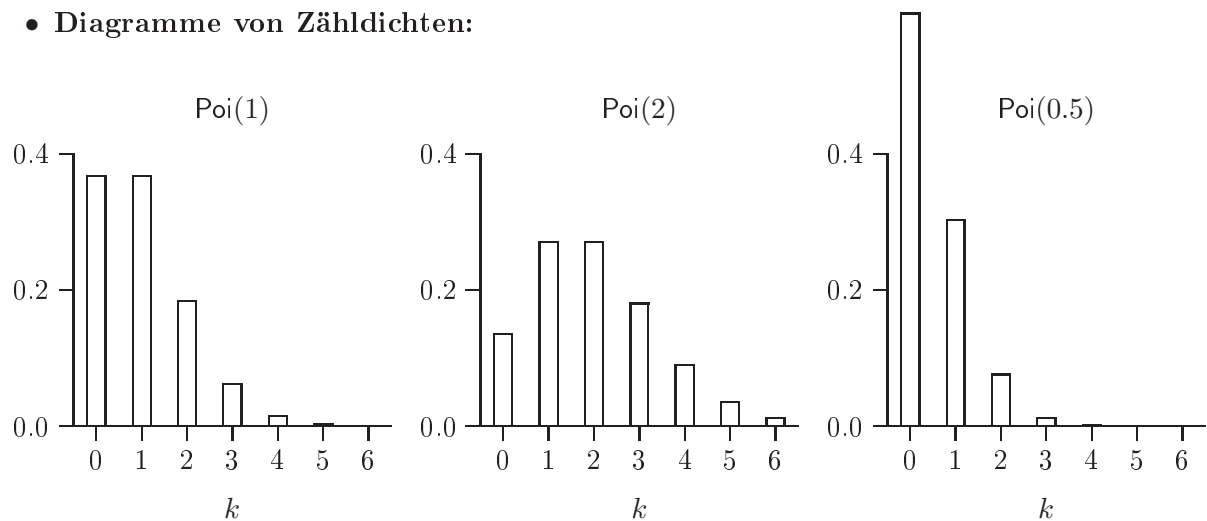
- **Verteilungsname in R:** `geom` mit Argument `prob` für p
- **Zähldichte:** $P(X = k) = p(1 - p)^k$, $p \in (0, 1)$
- **Träger:** $T_X = \mathbb{N}_0$
- **Diagramme von Zähldichten:**



- **Anwendung:** Wartezeit, bis einer von zwei möglichen Zuständen eingetreten ist.

13.1.6 Beispiel (Poisson-Verteilung $\text{Poi}(\lambda)$)

- **Verteilungsname in R:** `pois` mit Argument `lambda` für λ
- **Zähldichte:** $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $\lambda > 0$
- **Träger:** $T_X = \mathbb{N}_0$
- **Diagramme von Zähldichten:**



- **Anwendung:** Seltene Ereignisse

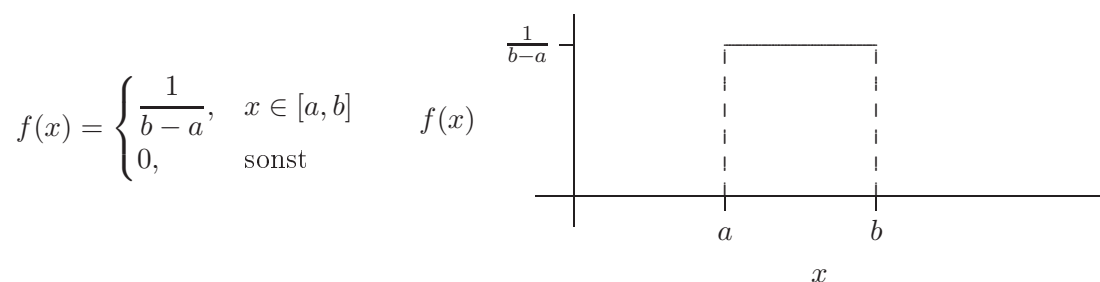
13.2 Stetige Wahrscheinlichkeitsverteilungen

Der **Träger** T_X einer **stetigen Verteilung** ist die Menge der Punkte, für die die Dichtefunktion f^X positiv ist.

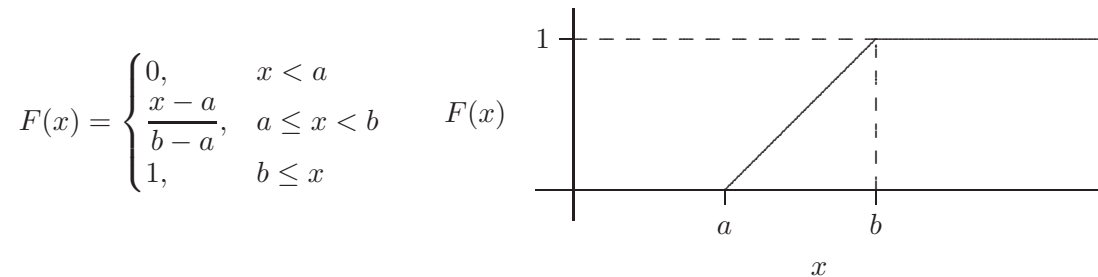
13.2.1 Beispiel (Rechteckverteilung (stetige Gleichverteilung) $R[a, b]$ mit Parametern a und b)

Träger: $T_X = [a, b]$ mit $a, b \in \mathbb{R}$, $a < b$

- **Dichte:**

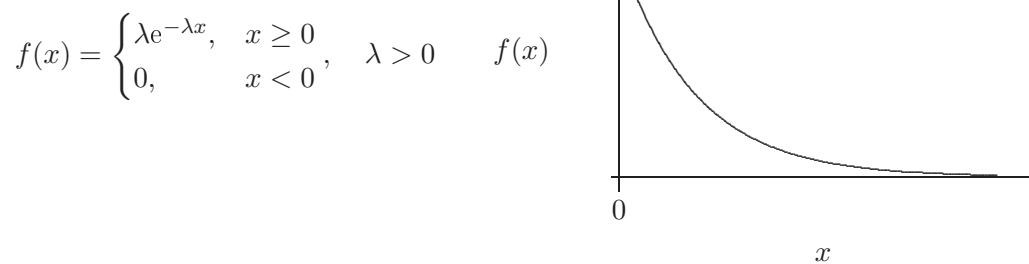


- **Verteilungsfunktion:**

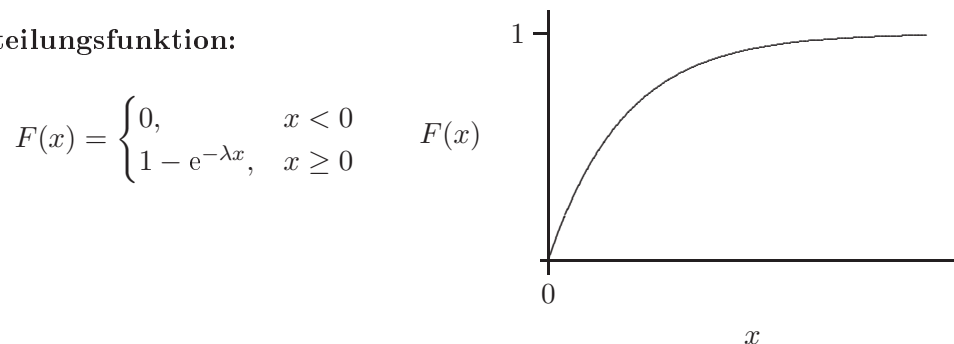


13.2.2 Beispiel (Exponentialverteilung $\text{Exp}(\lambda)$ mit Parameter λ)

- **Verteilungsname in R:** `exp` mit Argument `rate` für λ
- **Träger:** $T_X = [0, \infty)$
- **Dichte:**



- **Verteilungsfunktion:**



- **Anwendung:** Lebenszeit von Glühlampen, elektrischen Teilen etc.

13.2.3 Beispiel (Weibull-Verteilung $W(\alpha, \beta)$ mit Parametern α und β)

- **Verteilungsname in R:** `weibull` mit Argumenten `shape` für α und `scale` für β
- **Träger:** $T_X = [0, \infty)$
- **Dichte:**

$$f(x) = \begin{cases} 0, & x < 0 \\ \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}, & x \geq 0 \end{cases}, \quad \alpha, \beta > 0$$

- **Verteilungsfunktion:**

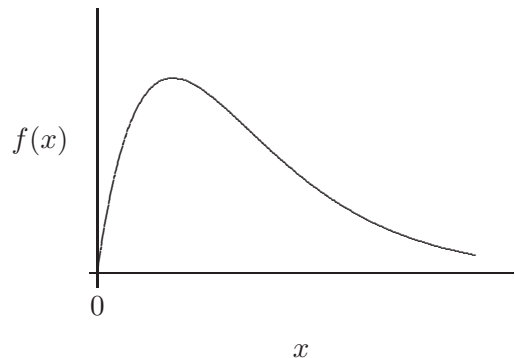
$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp^{-(x/\beta)^\alpha}, & x \geq 0 \end{cases}$$

- **Bemerkung:** Für $\alpha = 1$, $\beta = 1/\lambda$ ergibt sich die Exponentialverteilung.

13.2.4 Beispiel (Gamma-Verteilung $\Gamma(\alpha, \beta)$ mit Parametern α und β)

- **Verteilungsname in R:** `gamma` mit Argumenten `shape` für β und `rate` für α bzw. `scale` für $\frac{1}{\alpha}$
- **Träger:** $T_X = [0, \infty)$
- **Dichte:** $(\Gamma(\beta) = \int_0^\infty t^{\beta-1} e^{-t} dt)$

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



- **Verteilungsfunktion:** Geschlossene Darstellung nur für $\beta \in \mathbb{N}$.
- **Bemerkung:** Für $\alpha = \lambda$, $\beta = 1$ ergibt sich die Exponentialverteilung.

13.2.5 Beispiel (χ^2 -Verteilung χ_f^2 mit f Freiheitsgraden)

- **Verteilungsname in R:** `chisq` mit Argument `df` für f (`df`=degree of freedom)
- **Träger:** $T_X = [0, \infty)$
- **Dichte:**

$$f(x) = \begin{cases} \frac{1}{2^{f/2} \Gamma(f/2)} x^{f/2-1} e^{-x/2}, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad f \in \mathbb{N}$$

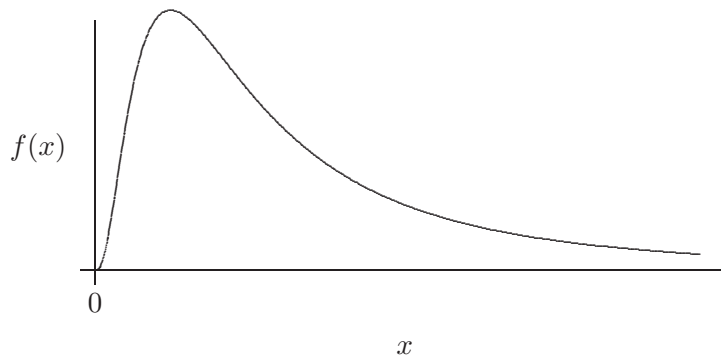
- **Verteilungsfunktion:** Geschlossene Darstellung nur für gerade $f \in \mathbb{N}$.
- **Bemerkung:** Die χ_f^2 -Verteilung ist eine Gamma-Verteilung mit Parametern $\frac{1}{2}$ und $\frac{f}{2}$.

13.2.6 Beispiel (F-Verteilung F_{f_1, f_2} mit f_1 und f_2 Freiheitsgraden)

- **Verteilungsname in R:** `f` mit Argumenten `df1` für f_1 und `df2` für f_2
- **Träger:** $T_X = [0, \infty)$

- **Dichte:**

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{f_1+f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\Gamma\left(\frac{f_2}{2}\right)} \left(\frac{f_1}{f_2}\right)^{f_1/2} \frac{x^{f_1/2-1}}{(1+f_1x/f_2)^{(f_1+f_2)/2}}, & x \geq 0, \\ 0, & x < 0 \end{cases}, \quad f_1, f_2 \in \mathbb{N}$$

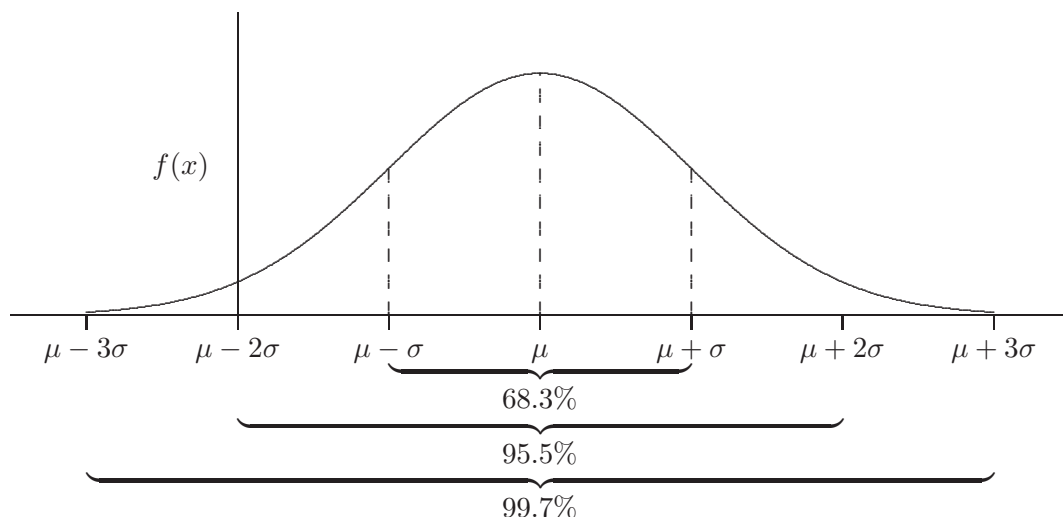


- **Verteilungsfunktion:** Nicht geschlossen darstellbar.

13.2.7 Beispiel (Normalverteilung $N(\mu, \sigma^2)$ mit Parametern μ und σ^2)

- **Verteilungsname in R:** norm mit Argumenten **mean** für μ und **sd** für σ
- **Träger:** $T_X = \mathbb{R}$
- **Dichte:** ('Gaußsche Glockenkurve')

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma^2 > 0.$$



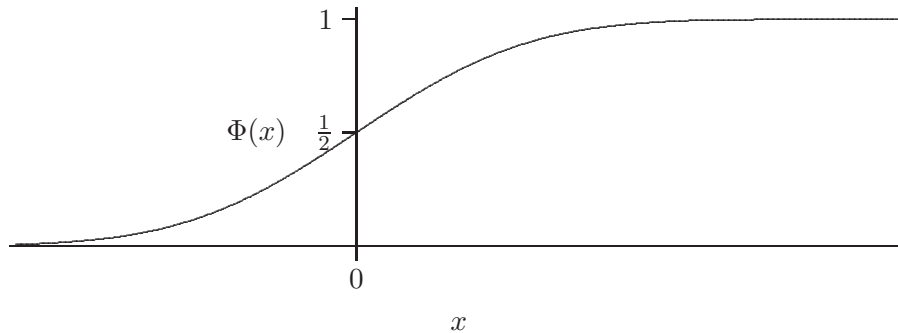
Die Prozentzahlen geben den Wert der Fläche über dem jeweiligen Intervall an. Die Fläche über dem Intervall $[\mu - \sigma, \mu + \sigma]$ beträgt etwa 0.683.

- **Verteilungsfunktion:** Eine geschlossene Darstellung existiert nicht. Die Verteilungsfunktion der Standardnormalverteilung (i.e. $\mu = 0, \sigma^2 = 1$) wird i.a. mit Φ bezeichnet. Die Verteilungsfunktion Φ_{μ, σ^2} einer $N(\mu, \sigma^2)$ -Verteilung berechnet sich aus Φ gemäß:

$$\Phi_{\mu, \sigma^2}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}.$$

Da die Dichte der Standardnormalverteilung symmetrisch um 0 ist, gilt für deren Verteilungsfunktion

$$\Phi(x) = 1 - \Phi(-x), \quad x \in \mathbb{R}.$$



- **Six Sigma Qualität:** In der Ingenieursstatistik spielt vor allem die sogenannte Six Sigma Qualität eine große Rolle. In der Regel variieren angefertigte Teile in ihrer Größe und in der Qualitätssicherung muss sichergestellt werden, dass die Variation in bestimmten zulässigen Grenzen passiert. Sind die Grenzen durch USL (=upper specification limits) und LSL (=lower specification limits) gegeben und folgt die Verteilung der Größe der Teile einer Normalverteilung mit μ , so besitzt die Produktion die Six Sigma Qualität, wenn für den Parameter σ der Normalverteilung folgendes gilt:

$$\sigma \leq \frac{1}{6} \min\{USL - \mu, \mu - LSL\}.$$

Damit gilt insbesondere

$$\mu - LSL \geq 6\sigma \implies LSL \leq \mu - 6\sigma \quad (13.1)$$

sowie

$$USL - \mu \geq 6\sigma \implies USL \geq \mu + 6\sigma \quad (13.2)$$

Aus (13.1) erhalten wir somit für eine Zufallsvariable X , die die Größe des Teiles beschreibt, folgendes

$$\begin{aligned} P(X < LSL) &\leq P(X < \mu - 6\sigma) \\ &= P\left(\frac{X - \mu}{\sigma} < -6\right) = \Phi(-6) = 0.000\,000\,000\,986\,5876, \end{aligned}$$

d.h. die Wahrscheinlichkeit, dass ein Teil produziert wird, dessen Größe die untere zulässige Grenze LSL unterschreitet, ist kleiner als 1 zu einer Milliarde (1 part per billion=PPB). Das gleiche gilt für das Überschreiten der oberen Grenze ULS mittels (13.2):

$$\begin{aligned} P(X > USL) &\leq P(X > \mu + 6\sigma) \\ &= P\left(\frac{X - \mu}{\sigma} > 6\right) = 1 - \Phi(6) = 0.000\,000\,000\,986\,5876. \end{aligned}$$

Da das vorgegebene μ in der Regel nicht genau eingehalten werden kann, schreibt man vor, dass die Differenz zwischen wahren μ_T und vorgegebenen μ nicht größer als 1.5σ sein darf. Damit wird (13.1) zu

$$LSL \leq \mu - 6\sigma = \mu_T + \mu - \mu_T - 6\sigma \leq \mu_T + 1.5\sigma - 6\sigma = \mu_T - 4.5\sigma$$

und (13.2) zu

$$USL \geq \mu + 6\sigma = \mu_T + \mu - \mu_T + 6\sigma \geq \mu_T - 1.5\sigma + 6\sigma = \mu_T + 4.5\sigma,$$

und es gilt

$$P(X < LSL) \leq P\left(\frac{X - \mu_T}{\sigma} < -4.5\right) = \Phi(-4.5) = 0.000\,003\,397\,673 < 0.000\,003\,4,$$

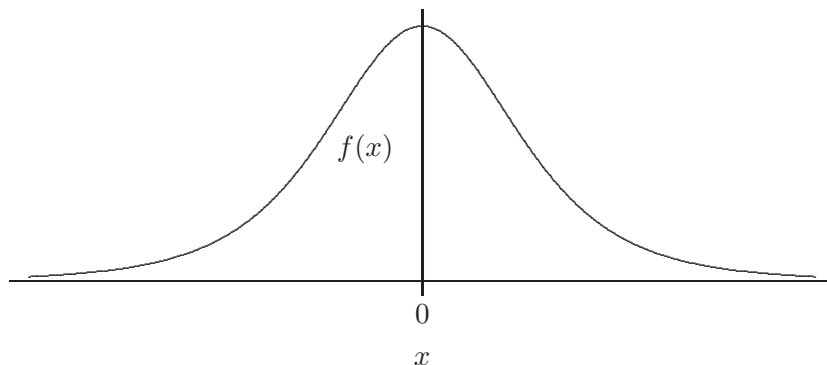
d.h. die Wahrscheinlichkeit, dass ein Teil produziert wird, dessen Größe die untere zulässige Grenze LSL unterschreitet, ist dann kleiner als 3.4 zu einer Million (3.4 part per million=PPM). Das gleiche gilt für das Überschreiten der oberen Grenze.

Beispiel: Es sei z.B. vorgegeben, dass der innere Durchmesser von produzierten Schrauben zwischen den Grenzen 20.5 und 22.0 Millimeter liegen soll. Folgt der Durchmesser der Schraube einer Normalverteilung mit $\mu = 21.3$, so besagt, die Six Sigma Qualität, dass dann der Parameter σ dieser Normalverteilung kleiner als das Minimum von $\frac{1}{6}(22.0 - 21.3) = 0.1166667$ und $\frac{1}{6}(21.3 - 20.5) = 0.1333333$, d.h. kleiner als 0.1166667 sein muss. Ob das der Fall ist und ob das wahre μ_T weniger als $1.5 \cdot 0.1166667$ von 21.3 abweicht, müssen mit Methoden der *Schließenden Statistik* geklärt werden.

13.2.8 Beispiel (t-Verteilung t_f mit f Freiheitsgraden)

- **Verteilungsname in R:** `t` mit Argument `df` für f
- **Träger:** $T_X = \mathbb{R}$
- **Dichte:**

$$f(x) = \frac{\Gamma\left(\frac{f+1}{2}\right)}{\sqrt{f\pi}\Gamma\left(\frac{f}{2}\right)} \left(1 + \frac{x^2}{f}\right)^{-(f+1)/2}, \quad x \in \mathbb{R}, f \in \mathbb{N}$$



- **Verteilungsfunktion:** Nicht geschlossen darstellbar.

13.3 Erzeugung von Zufallszahlen mittels `runif`

Jede beliebige Verteilung kann aus der Rechteck-Verteilung $R[0, 1]$ auf $[0, 1]$ erzeugt werden, die in R durch `runif` gegeben ist. Für diskrete Verteilungen kann eine Einteilung des Intervalls $[0, 1]$ vorgenommen werden, die den einzelnen Wahrscheinlichkeiten entspricht. Sind diese Wahrscheinlichkeiten durch $p_1 = P(\{n_1\})$, $p_2 = P(\{n_2\})$, $p_3 = P(\{n_3\})$, ... gegeben, so wird $[0, 1]$ in die Intervalle $[0, p_1]$, $[p_1, p_1 + p_2]$, $[p_1 + p_2, p_1 + p_2 + p_3]$, ... eingeteilt. Ergibt `runif` einen Wert in $[\sum_{i=1}^{j-1} p_i, \sum_{i=1}^j p_i]$, so wird als Zufallszahl n_j benutzt. Das gleiche Ergebnis erhält man, wenn man die Verteilungsfunktion F der Verteilung benutzt. In diesem Fall wird die Zufallszahl als $F^{-1}(u) := \inf\{z \in \mathbb{R}; F(z) \geq u\}$ gesetzt, wenn `runif` den Wert u ergibt. Für stetige Verteilungen kann nur die zweite Methode benutzt werden.

13.3.1 Übungsaufgabe

Erzeugen Sie 1000 Zufallszahlen von folgenden Verteilungen:

1. Poissonverteilung mit Parameter $\lambda = 1$, $\lambda = 2$, $\lambda = 4$,
2. Exponentialverteilung mit Parameter $\lambda = \frac{1}{2}$, $\lambda = 1$, $\lambda = 2$,
3. Weibullverteilung mit Parameter $(\alpha, \beta) = (1, 1)$, $(\alpha, \beta) = (1, 2)$, $(\alpha, \beta) = (2, 1)$, $(\alpha, \beta) = (2, 2)$,
4. Gammaverteilung mit Parameter $(\alpha, \beta) = (1, 1)$, $(\alpha, \beta) = (1, 2)$, $(\alpha, \beta) = (2, 1)$, $(\alpha, \beta) = (2, 2)$,
5. χ^2 -Verteilung mit 2, 3 und 10 Freiheitsgraden,
6. t -Verteilung mit 2, 3 und 10 Freiheitsgraden.

Stellen Sie die Zufallszahlen grafisch dar und stellen Sie diese Grafiken mit den grafischen Darstellungen der entsprechenden Dichten gegenüber.

Hinweis: Bei Zufallszahlen von einer stetigen Verteilung ist das Histogramm die richtige Darstellung, während bei Zufallszahlen von einer diskreten Verteilung ein Stabdiagramm gegeben durch `plot(x, type="h")` die angemessene Darstellung ist.

Kapitel 14

Zwei- und mehrdimensionale Wahrscheinlichkeitsverteilungen

14.1 Zweidimensionale diskrete Wahrscheinlichkeitsverteilungen

Zweidimensionale diskrete Wahrscheinlichkeitsverteilungen werden angegeben durch ihre Zähldichte $(p_{ij})_{i,j}$. Ein Beispiel ist die Verteilung aus Beispiel 12.2.1, die als Randverteilungen eine geometrische Verteilung und eine Binomialverteilung besitzt.

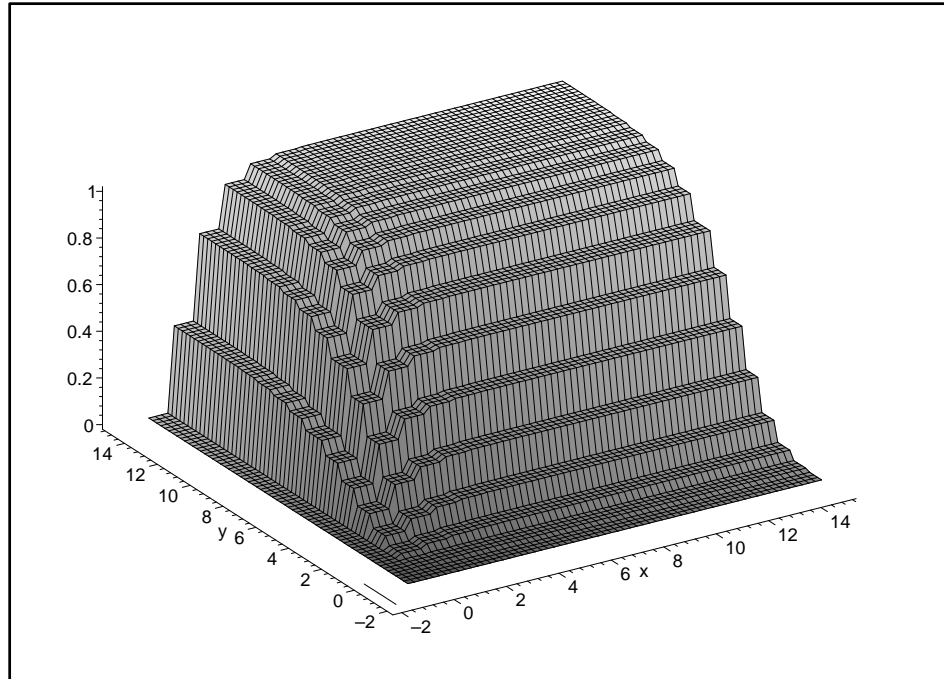
14.2 Zweidimensionale Normalverteilung

Eine zweidimensionale stetige Wahrscheinlichkeitsverteilung von besonderer Bedeutung ist die **zweidimensionale Normalverteilung**. Sie besitzt in ihrer allgemeinsten Form fünf Parameter $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$ und $\rho \in (-1, 1)$. Entsprechend wird sie bezeichnet mit $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Ihre Dichtefunktion ist gegeben durch $(x, y \in \mathbb{R})$:

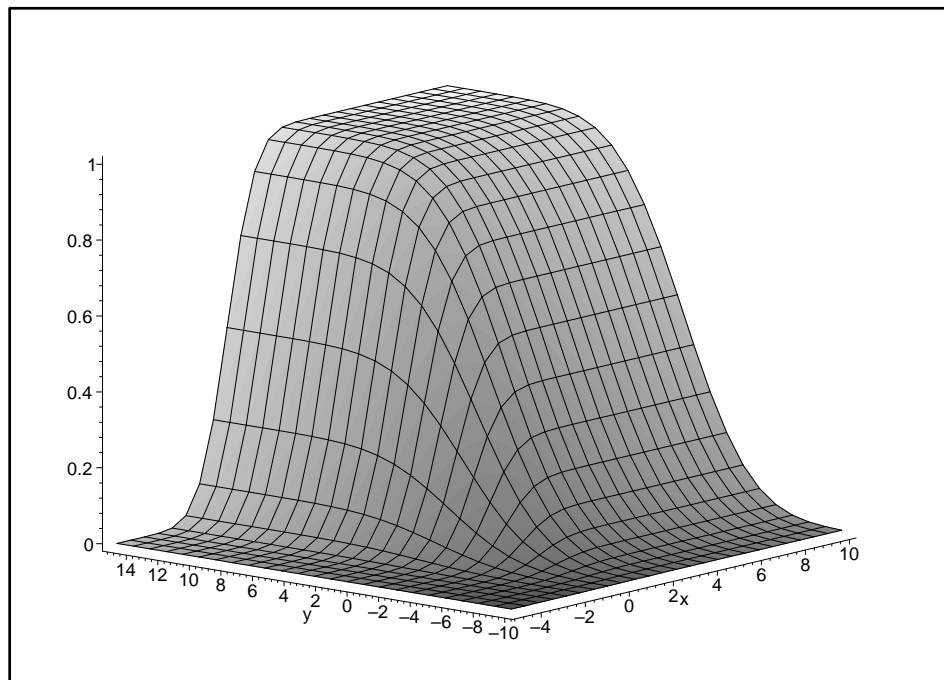
$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}.$$

Gilt $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ und $\rho = 0$, so heißt die entsprechende Verteilung **zweidimensionale Standardnormalverteilung**.

Zweidimensionale diskrete Verteilungsfunktion



Zweidimensionale stetige Verteilungsfunktion



Die zweidimensionale Normalverteilung ist im Standardpaket von R nicht enthalten. Sie kann aber schnell selber geschrieben werden. Mit der Eingabe

```
> source("d2norm.asc")
```

kann aber diese Funktion als vorhandenes Programm eingelesen werden. Dazu müssen Sie nur die Datei `d2norm.asc` in den Ordner kopieren, auf das R zugreift. Mit der Eingabe

```
> fix(d2norm)
```

wird ein Fenster geöffnet, in der die selbstgeschriebene Funktion `d2norm` steht. Sie können in diesem Fenster auch die Funktion abändern, wobei die Änderung erhalten bleibt, wenn Sie speichern. Grundsätzlich können mit dem Befehl `fix(Funktionsname)` neue Funktionen geschrieben und vorhandene selbstgeschriebene Funktionen abgeändert werden. Der Aufruf von `fix(d2norm)` ergibt folgendes:

```
function (x,y,m1=0,m2=0,sd1=1,sd2=1,rho=0)
{
# Dichte der zweidimensionalen Normalverteilung
f<-(1/(2*pi*sd1*sd2*sqrt(1-rho^2)))*exp((-1/(2*(1-rho^2)))*((x-m1)^2/sd1^2
-2*rho*(x-m1)*(y-m2)/(sd1*sd2)+(y-m2)^2/sd2^2))
f
}
```

Diese Funktion hat neben den Argumenten x und y die Parameter $\mu_1 = m1$, $\mu_2 = m2$, $\sigma_1 = s1$, $\sigma_2 = s2$ und $\rho = rho$, wobei die Parameter mit Voreinstellungen versehen sind, die die zweidimensionale Standardnormalverteilung ergeben. Die Dichte der zweidimensionalen Normalverteilung kann man sich als einen Berg vorstellen. Zur grafischen Darstellungen eines „Berges“ gibt es mehrere Möglichkeiten in R. Mit der R-Funktion `persp` erhält man eine perspektivische Darstellung, mit `contour` eine Darstellung mit Höhenlinien und mit `image` ein Bild, in dem helle Werte hohe Werte anzeigen. Für die Anwendung dieser Funktionen muss ein Intervall von \mathbb{R}^2 gerastert werden. Mit zwei Vektoren x und y kann diese Rasterung erzeugt werden. Dann muss eine Matrix mit den Funktionswerten an den Rasterpunkten erstellt werden. Hier wird diese Matrix `f` genannt, auch wenn sie in den R-Funktionen `z` genannt wird. Die Programm-Datei `dnorm_plot.asc` enthält die folgende Funktion, die die Dichtefunktion der Normalverteilung über dem Intervall $[-3, 3] \times [-3, 3]$ darstellt.

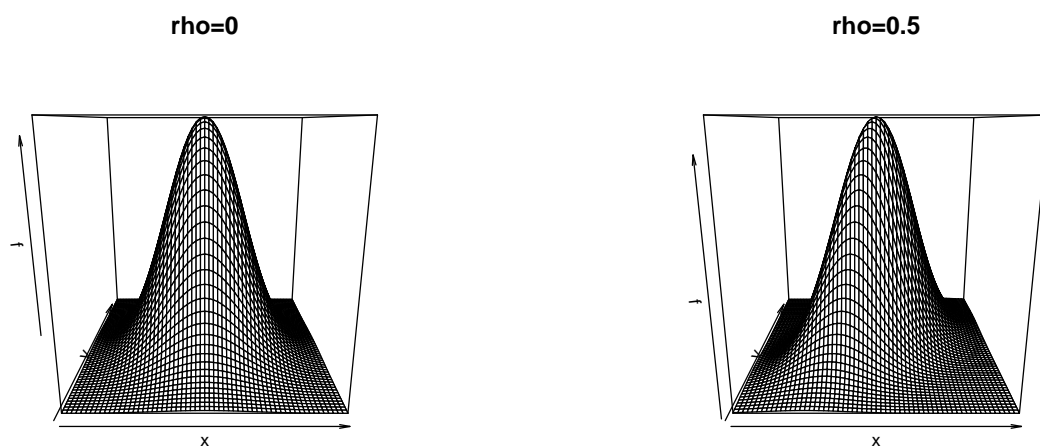
```
"d2norm.plot" <-
function (m1=0,m2=0,sd1=1,sd2=1,rho=0.5)
{
# Plottet die zweidimensionale Normalverteilung
par(pty="s")
x<-seq(-3,3,0.1)
y<-seq(-3,3,0.1)
I<-length(x)
J<-length(y)
f<-matrix(rep(0,I*J),ncol=J)
for(i in 1:I){
```

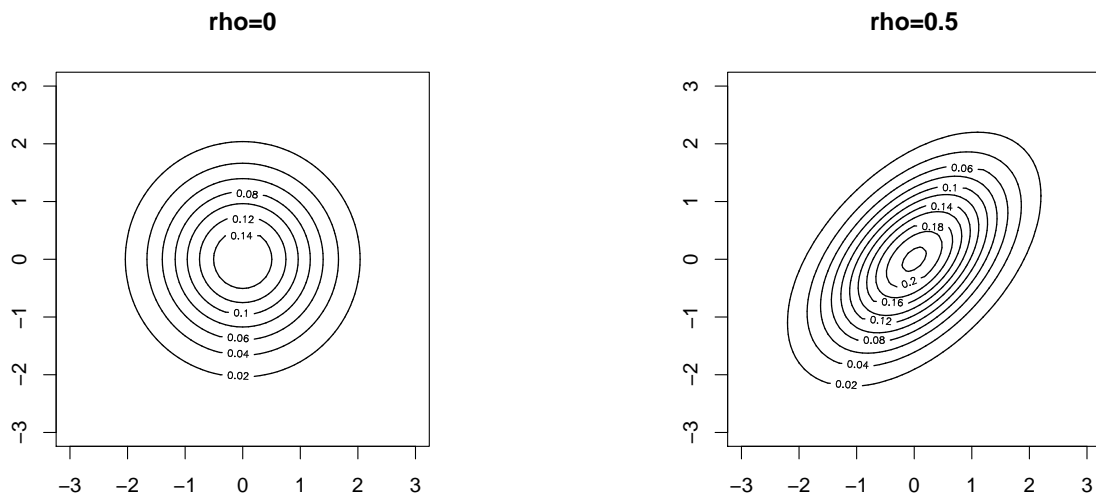
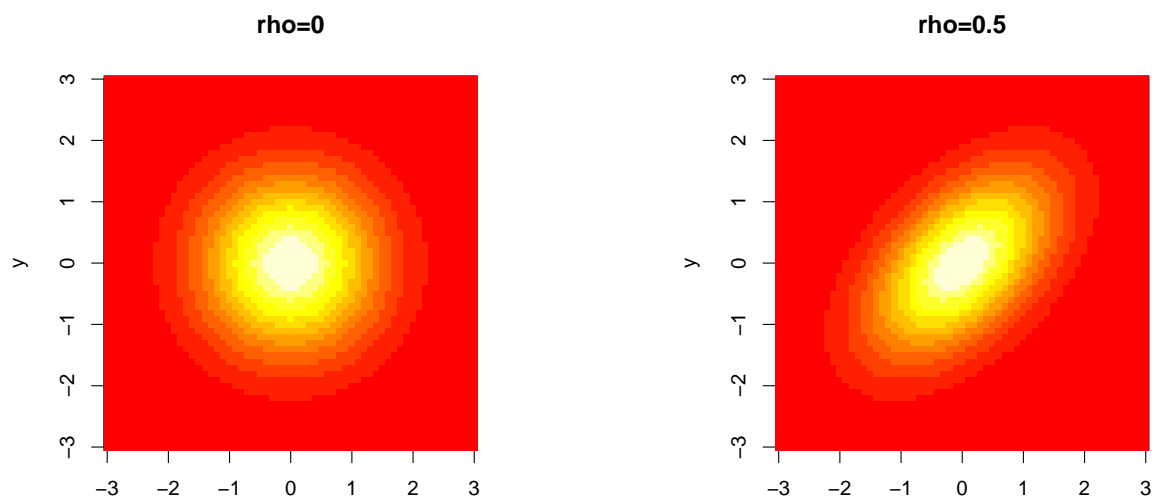
```

for(j in 1:J){
  f[i,j]<-d2norm(x[i],y[j],m1=m1,m2=m2,sd1=sd1,sd2=sd2,rho=rho)
}
}
persp(x,y,f)
# contour(x,y,f)
# image(x,y,f)
title("rho=0.5")
}

```

`par(pty="s")` erzeugt einen quadratischen Darstellungsbereich für die Grafik. Das ist insbesondere für `contour` und `image` wichtig, weil ansonsten das Bild verzerrt ist. Alles, was hinter `#` steht, wird ignoriert, so dass dahinter Kommentare geschrieben werden können. Hier werden insbesondere die Alternativen zu `persp` angezeigt. Sollen `contour` oder `image` benutzt werden, muss das Kommentarzeichen `#` entsprechend umgeändert werden. Die Funktion muss auch geändert werden, wenn andere Normalverteilungen über anderen Intervallen dargestellt werden sollen.

Abbildung 14.2.1: Darstellung mit `persp`

Abbildung 14.2.2: Darstellung mit `contour`Abbildung 14.2.3: Darstellung mit `image`

Die zweidimensionale Normalverteilung hat einige interessante Eigenschaften:

Sei $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Beide Randverteilungen sind eindimensionale Normalverteilungen, d.h. es gilt:

$$f^X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}, \quad x \in \mathbb{R}, \quad f^Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}, \quad y \in \mathbb{R}.$$

Die Zufallsvariablen X und Y sind stochastisch unabhängig genau dann, wenn $\rho = 0$ erfüllt ist. Dies lässt sich leicht über die Produktforderung an die Dichten nachweisen.

14.2.1 Übungsaufgabe

Stellen Sie die folgenden Dichten von zweidimensionalen Normalverteilungen dar:

1. Mit `persp`, `contour`, `image` die Dichte der zweidimensionalen Normalverteilung mit $\mu_1=\mathbf{m1}=0$, $\mu_2=\mathbf{m2}=1$, $\sigma_1=\mathbf{sd1}=1$, $\sigma_2=\mathbf{sd2}=1$ und $\rho=\mathbf{rho}=0$.
2. Mit `contour` die Dichte der zweidimensionalen Normalverteilung mit $\mu_1=\mathbf{m1}=0$, $\mu_2=\mathbf{m2}=1$, $\sigma_1=\mathbf{sd1}=1$, $\sigma_2=\mathbf{sd2}=2$ und $\rho=\mathbf{rho}=0$.
3. Mit `contour` die Dichte der zweidimensionalen Normalverteilung mit $\mu_1=\mathbf{m1}=0$, $\mu_2=\mathbf{m2}=1$, $\sigma_1=\mathbf{sd1}=1$, $\sigma_2=\mathbf{sd2}=1$ und $\rho=\mathbf{rho}=0.5$.
4. Mit `contour` die Dichte der zweidimensionalen Normalverteilung mit $\mu_1=\mathbf{m1}=2$, $\mu_2=\mathbf{m2}=1$, $\sigma_1=\mathbf{sd1}=1$, $\sigma_2=\mathbf{sd2}=1$ und $\rho=\mathbf{rho}=0$.
5. Mit `contour` die Dichte der zweidimensionalen Normalverteilung mit $\mu_1=\mathbf{m1}=0$, $\mu_2=\mathbf{m2}=1$, $\sigma_1=\mathbf{sd1}=1$, $\sigma_2=\mathbf{sd2}=1$ und $\rho=\mathbf{rho}=-0.8$.

Was bewirken die Änderungen der verschiedenen Parameter?

14.3 Mehrdimensionale diskrete Verteilungen

Eine wichtige mehrdimensionale Verteilung ist die **Multinomialverteilung**, die die Binomialverteilung verallgemeinert.

- **Zähldichte:** $\text{Mult}(n_1, \dots, n_r; n; p_1, \dots, p_r) = \frac{n!}{n_1!n_2!\dots n_r!} \prod_{i=1}^r p_i^{n_i}$ mit $n_1 + n_2 + \dots + n_r = n$, $p_1 + p_2 + \dots + p_r = 1$.
- **Träger:** $T_X = \{1, \dots, n\}^r$
- **Anwendung:** Häufigkeiten von Zuständen, wenn mehr als zwei Zustände auftreten.

14.4 Mehrdimensionale stetige Verteilungen

Die wichtigste mehrdimensionale Verteilung ist die mehrdimensionale Normalverteilung. Diese hat im Fall von s Dimensionen als Parameter einen s -dimensionalen Vektor $\mu \in \mathbb{R}^s$ und eine symmetrische, positiv definite Matrix $\Sigma \in \mathbb{R}^{s \times s}$ ($A \in \mathbb{R}^{s \times s}$ ist positiv definit $\Leftrightarrow c^\top A c > 0$ für alle $c \in \mathbb{R}^s \setminus \{0\}$). Ist $x \in \mathbb{R}^s$, so ist die Dichte bei $x = (x_1, \dots, x_s)^\top$ gegeben durch

$$f(x) = \frac{1}{\sqrt{(2\pi)^s \det(\Sigma)}} \exp \left\{ -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right\},$$

dabei bezeichnet \det die Determinante einer Matrix. Die zweidimensionale Normalverteilung ist ein Spezialfall der mehrdimensionalen Normalverteilung mit

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Kapitel 15

Bedingte Wahrscheinlichkeiten und stochastische Unabhängigkeit

15.1 Bedingte Wahrscheinlichkeiten

In vielen Situationen sind gewisse Vorinformationen bezüglich des Ereignisses A verfügbar. Beispielsweise ist bekannt, dass ein Ereignis B bereits eingetreten ist. Diese Zusatzinformation beeinflusst die Einschätzung der Wahrscheinlichkeiten.

15.1.1 Beispiel

Beim einfachen Würfelwurf beträgt die Wahrscheinlichkeit eine vorgegebene Ziffer zu würfeln jeweils $\frac{1}{6}$. Ist bekannt, dass beim Wurf eine gerade Ziffer herausgekommen ist, so würde man mit Wahrscheinlichkeit $\frac{1}{3}$ erwarten, dass das Ergebnis eine 2 ist. Weiterhin würde niemand nach dieser Information auf eine ungerade Zahl setzen, d.h. die Wahrscheinlichkeit für eine ungerade Ziffer wäre Null.

Die im obigen Beispiel dargestellte Situation wird durch den Begriff der **bedingten Wahrscheinlichkeit** formalisiert.

15.1.2 Definition

Seien (Ω, P) ein Wahrscheinlichkeitsraum und B ein Ereignis mit $P(B) > 0$. Dann wird die **bedingte Wahrscheinlichkeit eines Ereignisses A gegeben B** definiert durch

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

Gelegentlich wird die Schreibweise $P_B(A) = P(A|B)$ benutzt. P_B heißt **bedingte Verteilung unter B** .

15.1.3 Bemerkung

- P_B ist ein Wahrscheinlichkeitsmaß auf dem Grundraum B , d.h. P_B erfüllt die Kolmogorov-Axiome mit Grundraum B .

- In der Definition der bedingten Wahrscheinlichkeit wird für das Ereignis B positive Wahrscheinlichkeit vorausgesetzt. Im folgenden sei dies grundsätzlich erfüllt, wenn bedingte Wahrscheinlichkeiten betrachtet werden.

15.1.4 Beispiel

Im Beispiel 15.1.1 bezeichne B das Ereignis *gerade Zahl*, und A sei das Ereignis *es wird eine 2 gewürfelt*. Dann gilt $A \cap B = \{2\}$ und somit

$$P(A|B) = \frac{P(\{2\})}{P(\{2, 4, 6\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Bezeichnet ferner C das Ereignis *ungerade Zahl*, so gilt offensichtlich $B \cap C = \emptyset$. Insgesamt gilt daher $P(C|B) = 0$.

Die Definition der bedingten Wahrscheinlichkeit $P(A|B)$ läßt sich auch in anderer Weise lesen:

$$P(A \cap B) = P(A|B) \cdot P(B).$$

Die Wahrscheinlichkeit des Schnittereignisses $A \cap B$ läßt sich also berechnen, wenn die bedingte Wahrscheinlichkeit von A gegeben B und die Wahrscheinlichkeit von B bekannt sind. Dies führt sogleich zu der Formel:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Dies ist ein Spezialfall des **Satzes von der totalen Wahrscheinlichkeit**.

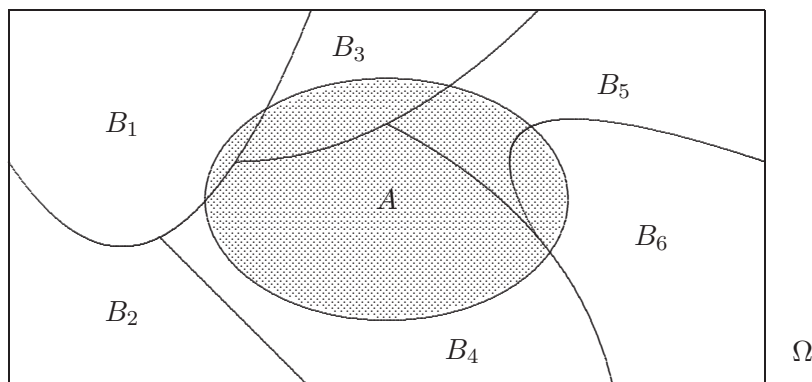
15.1.5 Satz (Satz von der totalen Wahrscheinlichkeit)

Seien B_1, \dots, B_n eine Zerlegung des Grundraumes Ω , d.h. B_1, \dots, B_n sind paarweise disjunkt und $\bigcup_{i=1}^n B_i = \Omega$. Ferner gelte $P(B_i) > 0$, $i = 1, \dots, n$.

Dann gilt für die Wahrscheinlichkeit eines beliebigen Ereignisses A :

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

Folgende Graphik illustriert den Begriff *Zerlegung von Ω* für $n = 6$.



15.1.6 Beispiel

Zwei Firmen V und W stellen ein Maschinenteil her. Sei B das Ereignis, dass das Maschinenteil von Firma V hergestellt wurde, und B^c das Ereignis, dass das Maschinenteil von Firma W hergestellt wurde. A sei das Ereignis, dass das Maschinenteil kaputt ist. Folgende Wahrscheinlichkeiten sind bekannt:

Wahrscheinlichkeit, dass das Maschinenteil von Firma V hergestellt wurde: $P(B) = 0.04$

Wahrscheinlichkeit, dass das Maschinenteil kaputt ist,

wenn es von Firma V hergestellt wurde: $P(A|B) = 0.7$

Wahrscheinlichkeit, dass das Maschinenteil kaputt ist,

wenn es von Firma W hergestellt wurde: $P(A|B^c) = 0.01$

Zunächst folgt aus den Angaben $P(B^c) = 1 - P(B) = 0.96$, so dass

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = 0.7 \cdot 0.04 + 0.01 \cdot 0.96 = 0.0376.$$

Daher beträgt die Wahrscheinlichkeit, dass das Maschinenteil kaputt ist, 3.76%.

Eine weitere wichtige Regel für bedingte Wahrscheinlichkeiten ist die **Formel von Bayes**:

15.1.7 Satz (Formel von Bayes)

Seien B_1, \dots, B_n eine Zerlegung des Grundraumes Ω , d.h. B_1, \dots, B_n sind paarweise disjunkt und $\bigcup_{i=1}^n B_i = \Omega$. Ferner gelte $P(B_i) > 0$, $i = 1, \dots, n$, und $P(A) > 0$.

Dann gilt:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}.$$

15.1.8 Bemerkung

- Bei der Bayesschen Formel wird von der **Wirkung** A auf die **Ursache** B_i zurückgeschlossen. A kann als Symptom interpretiert werden, das von verschiedenen Krankheiten B_1, \dots, B_n verursacht wird. Gesucht ist die Wahrscheinlichkeit, dass der Patient bei Beobachtung des Symptoms A an Krankheit B_i leidet. In dieser Situation ist man oft daran interessiert die Krankheit B_i zu finden, die die größte Wahrscheinlichkeit $P(B_i|A)$ besitzt.
- $P(B_1), \dots, P(B_n)$ heißen **a-priori Wahrscheinlichkeiten**; $P(B_1|A), \dots, P(B_n|A)$ heißen **a-posteriori Wahrscheinlichkeiten**. Die a-priori Wahrscheinlichkeiten sind als subjektive Vorbewertung zu interpretieren, die durch zusätzliche Information (Eintreten des Ereignisses A) in die a-posteriori Wahrscheinlichkeiten überführt werden. Dieser Zusammenhang wird durch die Bayessche Formel hergestellt.

15.1.9 Beispiel

Die Formel von Bayes kann zur Analyse von medizinischen Labortests verwendet werden. Dabei sind prinzipiell zwei Situationen von Falschergebnissen zu unterscheiden: der Test erkennt eine Krankheit nicht, obwohl die Person erkrankt ist (*falsch negativ*), bzw. der Test ergibt eine Erkrankung, obwohl die Person tatsächlich gesund ist (*falsch positiv*). Als Beispiel sei der ELISA-Test zur Erkennung einer HIV-Infektion gewählt (siehe N. Henze (1997) *Stochastik für Einsteiger*. Vieweg. S. 108-109).

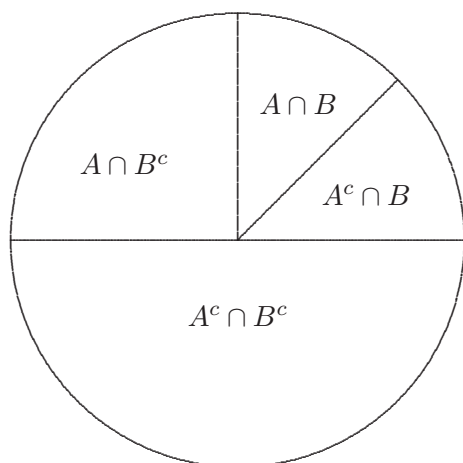
Der ELISA-Test erkennt mit Wahrscheinlichkeit 0.997, dass eine erkrankte Person auch tatsächlich erkrankt ist. Die Wahrscheinlichkeit, dass eine gesunde Person als krank erkannt wird, beträgt 0.001.

Nimmt man nun an, dass die Wahrscheinlichkeit krank zu sein $q \in [0, 1]$ ist, so beträgt die Wahrscheinlichkeiten tatsächlich krank (Ereignis B) zu sein, wenn ein positiver Befund vorliegt (Ereignis A):

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)} = \frac{0.997 \cdot q}{0.997 \cdot q + 0.001(1 - q)}.$$

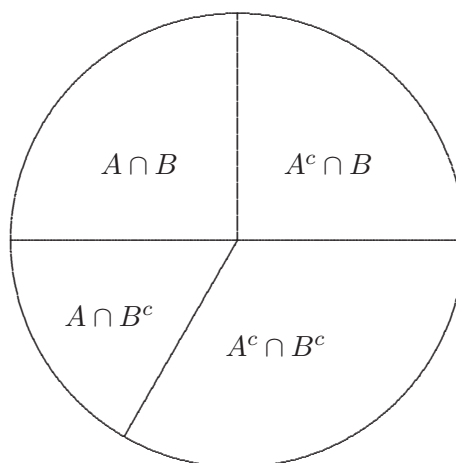
Für $q = 0.02$ (d.h. das Krankheitsrisiko liegt bei 2%) ergibt sich $P(B|A) = 0.953$. Die Wahrscheinlichkeit bei positivem Befund auch tatsächlich erkrankt zu sein, liegt daher bei etwa 95%. Bei einem Risikowert von 0.2% liegt die Wahrscheinlichkeit bei ca. 67%. Insbesondere ist ersichtlich, dass die Wahrscheinlichkeit bei positivem Testergebnis erkrankt zu sein, stark von der subjektiven Einschätzung des Risikos q abhängt.

Mit etwas unrealistischen bedingten Wahrscheinlichkeiten und höherer Krankheitswahrscheinlichkeit, kann dieser Effekt auch graphisch veranschaulicht werden. Dazu sei $P(A|B) = 1/2$ die Wahrscheinlichkeit, dass der Test eine kranke Person auch als krank erkennt, und $P(A|B^c) = 1/3$ die Wahrscheinlichkeit, dass der Test eine gesunde Person als krank erkennt. Für $P(B) = 1/4$ und $P(B) = 1/2$ können die einzelnen Wahrscheinlichkeiten als Anteile eines Kreises dargestellt werden:



$$P(B) = 1/4:$$

$$P(B|A) = \frac{1/2 \cdot 1/4}{1/2 \cdot 1/4 + 1/3 \cdot 3/4} = 1/3 = 0.333.$$



$$P(B) = 1/2:$$

$$P(B|A) = \frac{1/2 \cdot 1/2}{1/2 \cdot 1/2 + 1/3 \cdot 1/2} = 3/5 = 0.6.$$

15.1.10 Übungsaufgabe

Eine Firma bezieht ein Bauteil von vier Lieferanten A, B, C, D in folgenden Anteilen:

A : 30%, B : 20%, C : 40%, D : 10%.

Der Anteil der fehlerhaften gelieferten Bauteile beträgt bei

A : $\frac{10}{100}$, B : $\frac{5}{100}$, C : $\frac{5}{100}$, D : $\frac{90}{100}$.

Die Firma baut zufällig eins von den gelieferten Bauteilen ein.

- (a) Wie groß ist dann die Wahrscheinlichkeit, dass das eingebaute Bauteil fehlerhaft ist?
- (b) Wie groß ist die Wahrscheinlichkeit, dass ein fehlerhaftes eingebautes Bauteil von der Firma A stammt?

15.1.11 Übungsaufgabe

Sei X eine Zufallsvariable mit Exponentialverteilung mit Parameter $\lambda = 1$. Bestimmen Sie folgende Wahrscheinlichkeiten:

$$P(X \geq 3), \quad P\left(X \geq \frac{1}{2}\right), \quad P(X \geq 1), \quad P\left(X \geq 1 | X \geq \frac{1}{2}\right), \\ P(X \geq 2 | X \geq 1), \quad P(X \geq 4 | X \geq 1).$$

Welche Vermutung leiten Sie davon ab?

15.2 Stochastische Unabhängigkeit von Ereignissen

Im vorhergehenden Abschnitt wurde der Begriff der bedingten Wahrscheinlichkeit eingeführt. Dabei wurde der Einfluß eines Ereignisses B auf ein Ereignis A durch die Wahrscheinlichkeit $P(A|B)$ modelliert. Intuitiv würde man die Ereignisse A und B als unabhängig betrachten, wenn die Wahrscheinlichkeit von A nicht davon abhängt ob B eingetreten ist oder nicht, d.h.

$$P(A|B) = P(A). \quad (\text{unab})$$

15.2.1 Beispiel

Aus einer Urne mit zwei roten und drei schwarzen Kugeln werden mit Zurücklegen zwei Kugeln entnommen. A bezeichne das Ereignis 2. Kugel schwarz, B sei das Ereignis 1. Kugel rot. Dann gilt:

$$P(A) = \frac{3}{5}, \quad P(B) = \frac{2}{5}, \quad P(A \cap B) = \frac{6}{25}, \quad P(A|B) = \frac{3}{5} = P(A).$$

Dieser Sachverhalt spiegelt die Situation wider, dass die Ziehungen sich durch das Zurücklegen der Kugeln nicht beeinflussen. Wird das Experiment mittels Ziehen ohne Zurücklegen durchgeführt, so gilt $P(A|B) = 0.75 \neq 0.6 = P(A)$. Dies zeigt, dass die Wahrscheinlichkeit ansteigt im zweiten Zug eine schwarze Kugel zu ziehen, wenn bekannt ist, dass im ersten eine rote entnommen wurde.

In der Bedingung (unab) wird vorausgesetzt, dass $P(B) > 0$ ist, da sonst die bedingte Wahrscheinlichkeit nicht definiert ist. Um diese Schwierigkeit zu umgehen wird die **stochastische Unabhängigkeit** zweier Ereignisse A und B definiert durch:

15.2.2 Definition

Seien (Ω, P) ein Wahrscheinlichkeitsraum und A, B Ereignisse.

Dann heißen A und B **stochastisch unabhängig**, wenn gilt:

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

15.2.3 Bemerkung

- Sind A und B stochastisch unabhängig, so auch A, B^c ; A^c, B und A^c, B^c .
- Ereignisse A_1, \dots, A_n heißen **paarweise stochastisch unabhängig**, falls $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ für alle $i \neq j$.
- Ereignisse A_1, \dots, A_n heißen **gemeinsam stochastisch unabhängig**, falls für alle Auswahlen von Indizes $\{i_1, \dots, i_s\} \subset \{1, \dots, n\}$ gilt:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_s}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_s}).$$

Für $n = 3$ bedeutet dies: die Ereignisse A_1, A_2, A_3 sind gemeinsam stochastisch unabhängig, falls:

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1)P(A_2), & P(A_1 \cap A_3) &= P(A_1)P(A_3), & P(A_2 \cap A_3) &= P(A_2)P(A_3), \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3). \end{aligned}$$

Zur paarweisen stochastischen Unabhängigkeit kommt daher noch eine zusätzliche Forderung dazu. Dies bedeutet insbesondere, dass die gemeinsame stochastische Unabhängigkeit von Ereignissen deren paarweise Unabhängigkeit impliziert. Die Umkehrung ist aber i.a. nicht richtig (siehe Beispiel 15.2.4). Wenn im folgenden von stochastischer Unabhängigkeit mehrerer Ereignisse gesprochen wird, so meint dies immer gemeinsame stochastische Unabhängigkeit.

15.2.4 Beispiel

Seien $\Omega = \{1, 2, 3, 4\}$ und P die Laplace-Verteilung auf Ω . Die Ereignisse $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$ sind dann paarweise stochastisch unabhängig, da

$$P(A) = P(B) = P(C) = \frac{1}{2}, \quad P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}.$$

Wegen $A \cap B \cap C = \emptyset$ gilt jedoch:

$$P(A \cap B \cap C) = 0 \neq \frac{1}{8} = P(A)P(B)P(C),$$

d.h. A, B, C sind nicht gemeinsam stochastisch unabhängig.

15.2.5 Beispiel

In Beispiel 10.0.5 wurden beim zweifachen Würfelwurf die Ereignisse $A = \text{erste Ziffer gerade}$ und $B = \text{Summe beider Ziffern ist gerade}$ betrachtet. Als Wahrscheinlichkeit ergeben sich

$$P(A) = \frac{18}{36} = \frac{1}{2}, \quad P(B) = \frac{18}{36} = \frac{1}{2} \quad \text{und} \quad P(A \cap B) = \frac{9}{36} = \frac{1}{4}.$$

Daher gilt $P(A \cap B) = \frac{1}{4} = P(A)P(B)$, so dass die Ereignisse A und B stochastisch unabhängig sind. Die Ereignisse $D = \text{Beide Ziffern sind gleich}$ und B sind hingegen nicht stochastisch unabhängig.

15.3 Unabhängigkeit von Zufallsvariablen

In Abschnitt 15.2 wurde die Unabhängigkeit zweier Ereignisse A und B definiert durch die Eigenschaft:

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

Die Forderung dieser Produkteigenschaft wird nun auf Zufallsvariablen übertragen.

15.3.1 Definition (Stochastische Unabhängigkeit von Zufallsvariablen)

Seien X und Y Zufallsvariablen mit Verteilungsfunktionen F^X , F^Y und $F^{(X,Y)}$. Dann heißen X und Y stochastisch unabhängig, falls

$$F^{(X,Y)}(x,y) = F^X(x) \cdot F^Y(y) \quad \text{für alle } x, y \in \mathbb{R}.$$

Zufallsvariablen X_1, \dots, X_n heißen stochastisch unabhängig, falls

$$F^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n F^{X_i}(x_i) \quad \text{für alle } x_1, \dots, x_n \in \mathbb{R}.$$

15.3.2 Bemerkung

Die Unabhängigkeit von Zufallsvariablen lässt sich auch mittels der (Zähl-) Dichten überprüfen.

- **Diskreter Fall:** Äquivalent zur Definition 15.3.1 ist die Bedingung an die Zähldichten der Zufallsvariablen:

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \quad \text{für alle } i \text{ und } j.$$

Diese Bedingung erinnert an die ursprüngliche Definition der Unabhängigkeit von Ereignissen. Für Zufallsvariablen X_1, \dots, X_n lautet die Bedingung:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) \quad \text{für alle } x_i \in T_{X_i}, i = 1, \dots, n.$$

- **Stetiger Fall:** Äquivalent zur Definition 15.3.1 ist die Bedingung an die Dichten der Zufallsvariablen:

$$f^{(X,Y)}(x,y) = f^X(x) \cdot f^Y(y) \quad \text{für alle } x \text{ und } y.$$

Für Zufallsvariablen X_1, \dots, X_n lautet die Bedingung:

$$f^{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f^{X_i}(x_i) \quad \text{für alle } x_1, \dots, x_n \in \mathbb{R}.$$

Je nach Fragestellung kann es einfacher sein die Definition zu benutzen oder auf diese äquivalenten Formulierungen zurückzugreifen.

15.3.3 Beispiel

In Beispiel 12.2.1 ergibt sich für das Produkt der Zähldichten p_i^X , $i \in \mathbb{N}_0$, und p_j^Y , $j = 0, \dots, n$:

$$p_i^X \cdot p_j^Y = \frac{1}{2^{i+1}} \cdot \binom{n}{j} \frac{1}{2^n} = \binom{n}{j} \frac{1}{2^{n+i+1}} = p_{ij}.$$

Daher ist die Produkteigenschaft für die Zähldichten erfüllt, und X und Y sind stochastisch unabhängig.

In Beispiel 12.2.2 kann die Produkteigenschaft der Verteilungsfunktionen zur Untersuchung der Unabhängigkeit verwendet werden. Wären die Zufallsvariablen X und Y stochastisch unabhängig, so müßte $F^{(X,Y)} = F^X(x) \cdot F^Y(y)$ für alle $x, y \in \mathbb{R}$ gelten. Wegen

$$F^X(x) \cdot F^Y(y) = (1 - e^{-x})(1 - e^{-y}), \quad x, y \geq 0$$

und $F^{(X,Y)}(x, y) = 1 - e^{-y}$, $0 \leq y < x$, gilt offensichtlich

$$F^X(x) \cdot F^Y(y) = (1 - e^{-x})(1 - e^{-y}) \neq 1 - e^{-y} = F^{(X,Y)}(x, y), \quad 0 \leq y < x,$$

so dass X und Y nicht stochastisch unabhängig sind.

15.3.4 Beispiel

Besitzt (X, Y) eine zweidimensionale Normalverteilung (siehe Abschnitt 14.2), so sind X und Y genau dann stochastisch unabhängig, wenn $\rho = 0$ gilt. Denn es gilt $\rho = 0$ genau dann, wenn

$$\begin{aligned} f^{(X,Y)}(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\} \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{1}{2\sigma_1^2}(x-\mu_1)^2 \right\} \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2\sigma_2^2}(y-\mu_2)^2 \right\} = f^X(x) f^Y(y). \end{aligned}$$

15.3.5 Übungsaufgabe

Die Zufallsvariablen X und Y beschreiben die Bewertungen eines Bauteiles in zwei Tests. Die Verteilung von (X, Y) ist durch folgende unvollständige Tabelle gegeben:

$P(X = i, Y = j)$		Y			$P(X = i)$
		1	2	3	
X	1	$\frac{1}{20}$	0		0.05
	2	$\frac{3}{20}$		$\frac{1}{20}$	0.25
	3	$\frac{3}{20}$		$\frac{1}{20}$	0.3
	4	$\frac{1}{20}$	$\frac{1}{5}$		

a) Vervollständigen Sie die Tabelle.

b) Sind X und Y stochastisch unabhängig oder gibt es einen Zusammenhang zwischen den Tests?

15.4 Verteilung von Funktionen von Zufallsvariablen

In diesem Abschnitt wird die Wahrscheinlichkeitsverteilung von Funktionen diskreter und stetiger Zufallsvariablen diskutiert.

Seien X und Y diskrete Zufallsvariablen mit Zähldichte $P(X = x_i, Y = y_j)$ und $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine Funktion. Dann ist die Verteilung der Zufallsvariablen $h(X, Y)$ bestimmt durch

$$P(h(X, Y) = k) = P^{(X, Y)}(\{(x_i, y_j) \mid h(x_i, y_j) = k\}) = \sum_{(i, j): h(x_i, y_j) = k} P(X = x_i, Y = y_j).$$

15.4.1 Beispiel

Seien $h(x, y) = x \cdot y$, $X \sim \text{Bin}(2, p)$, $p \in (0, 1)$, und Y eine Zufallsvariable mit Zähldichte

$$P(Y = 1) = 1 - t, \quad P(Y = 2) = t, \quad t \in (0, 1).$$

Ferner seien X und Y stochastisch unabhängig. Dann besitzt $h(i, j)$ für $i \in \{0, 1, 2\}$, $j \in \{1, 2\}$ offensichtlich die möglichen Werte $\{0, 1, 2, 4\}$. Damit gilt für $P(XY = k)$, $k = 0, 1, 2, 4$:

$$\begin{aligned} P(XY = 0) &= P(X = 0 \text{ oder } Y = 0) = P(X = 0) = (1 - p)^2 \\ P(XY = 1) &= P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = 2p(1 - p)(1 - t) \\ P(XY = 2) &= P(X = 1, Y = 2 \text{ oder } X = 2, Y = 1) = P(X = 1, Y = 2) + P(X = 2, Y = 1) \\ &= P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1) = 2p(1 - p)t + p^2(1 - t) \\ P(XY = 4) &= P(X = 2, Y = 2) = P(X = 2)P(Y = 2) = p^2t \end{aligned}$$

15.4.2 Beispiel (Summe von $\text{Bin}(1, p)$ -verteilten Zufallsvariablen)

Seien $h(x, y) := x + y$ und X, Y stochastisch unabhängige Zufallsvariable mit $X, Y \sim \text{Bin}(1, p)$, $p \in (0, 1)$. Dann gilt für die Verteilung von $X + Y$:

$$\begin{aligned} P(X + Y = 0) &= P(X = 0, Y = 0) = P(X = 0)P(Y = 0) = (1 - p)^2 \\ P(X + Y = 1) &= P(X = 0, Y = 1 \text{ oder } X = 1, Y = 0) = P(X = 0, Y = 1) + P(X = 1, Y = 0) \\ &= P(X = 0)P(Y = 1) + P(X = 1)P(Y = 0) = 2p(1 - p) \\ P(X + Y = 2) &= P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = p^2. \end{aligned}$$

Damit gilt $X + Y \sim \text{Bin}(2, p)$. Die Summe zweier stochastisch unabhängiger $\text{Bin}(1, p)$ -verteilter Zufallsvariablen ist also wiederum binomialverteilt, jedoch mit den Parametern 2 und p .

Allgemein gilt: Sind $X \sim \text{Bin}(n, p)$ und $Y \sim \text{Bin}(m, p)$ stochastisch unabhängig, so gilt $X + Y \sim \text{Bin}(n + m, p)$. Die ersten Parameter n und m werden also addiert, während der zweite Parameter unverändert bleibt.

Eine weitere Verallgemeinerung ist: Sind $X_1, \dots, X_N \sim \text{Bin}(1, p)$ stochastisch unabhängig, so gilt:

$$\sum_{n=1}^N X_n \sim \text{Bin}(N, p).$$

Die Summe von N stochastisch unabhängigen Bernoulli-verteilten Zufallsvariablen ist daher binomialverteilt mit Parametern N und p .

15.4.3 Satz

X_1, \dots, X_n seien stochastisch unabhängig.

1. Gilt $X_n \sim \text{Bin}(1, p)$ für $n = 1, \dots, N$, so gilt

$$\sum_{n=1}^N X_n \sim \text{Bin}(N, p),$$

d.h. die Summe unabhängiger Bernoulli-verteilter Zufallsvariablen ist binomialverteilt.

2. Gilt $X_n \sim \text{Poi}(\lambda)$ für $n = 1, \dots, N$, so gilt

$$\sum_{n=1}^N X_n \sim \text{Poi}(N\lambda),$$

d.h. die Summe unabhängiger Poission-verteilter Zufallsvariablen ist wieder Poisson-verteilt.

3. Gilt $X_n \sim N(\mu, \sigma^2)$ für $n = 1, \dots, N$, so gilt

$$\sum_{n=1}^N X_n \sim N(N\mu, N\sigma^2),$$

d.h. die Summe unabhängiger normalverteilter Zufallsvariablen ist wieder normalverteilt.

15.4.4 Übungsaufgabe

Während eines Fluges versage jedes Triebwerk eines Flugzeuges unabhängig von den anderen mit Wahrscheinlichkeit $p = 0.5$. Das Flugzeug bleibe flugfähig, wenn mindestens die Hälfte der Triebwerke funktioniert.

Vergleichen Sie die Zuverlässigkeit von Flugzeugen mit zwei bzw. vier Triebwerken, d.h. berechnen Sie die Wahrscheinlichkeit, dass das jeweilige Flugzeug funktionsfähig ist.

Kapitel 16

Erwartungswert und Varianz

Wie Zufallszahlen und damit auch Daten durch einige wenige Parameter wie arithmetisches Mittel, Standardabweichung, p -Quantil beschrieben werden können, können Wahrscheinlichkeitsverteilungen durch Verteilungsparameter beschrieben werden.

Die Analoga zu arithmetischem Mittel und empirischer Varianz sind Erwartungswert und Varianz.

16.1 Diskrete Wahrscheinlichkeitsverteilungen

Sei X eine diskrete Zufallsvariable mit Träger $T_X = \{x_1, x_2, \dots\}$ und Zähldichte $p_i = P(X = x_i)$. Alle im folgenden auftretenden Grenzwerte werden als existent vorausgesetzt.

16.1.1 Definition

Der **Erwartungswert** der Zufallsvariablen X ist definiert durch

$$E(X) = \sum_{x_i \in T_X} x_i P(X = x_i) = \begin{cases} \sum_{i=1}^n x_i p_i, & T_X = \{x_1, \dots, x_n\}, \\ \sum_{i=1}^{\infty} x_i p_i, & T_X = \{x_1, x_2, \dots\}. \end{cases}$$

Die **Varianz** der Zufallsvariablen X ist definiert durch

$$\text{var}(X) = E((X - EX)^2) = \sum_{x_i \in T_X} (x_i - EX)^2 P(X = x_i) = \begin{cases} \sum_{i=1}^n (x_i - EX)^2 p_i, & T_X = \{x_1, \dots, x_n\}, \\ \sum_{i=1}^{\infty} (x_i - EX)^2 p_i, & T_X = \{x_1, x_2, \dots\}. \end{cases}$$

Die **Standardabweichung** der Zufallsvariablen X ist definiert durch $\sqrt{\text{var}X}$.

Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion, so gilt für den Erwartungswert der transformierten Zufallsgröße $h(X)$:

$$\mathbb{E}(h(X)) = \sum_{x_i \in T_X} h(x_i) \mathbb{P}(X = x_i) = \begin{cases} \sum_{i=1}^n h(x_i) p_i, & T_X = \{x_1, \dots, x_n\}, \\ \sum_{i=1}^{\infty} h(x_i) p_i, & T_X = \{x_1, x_2, \dots\}. \end{cases}$$

Speziell für $h(x) = x^k$ erhält man das k -te **Moment** von X : $m_k = E(X^k)$, $k \in \mathbb{N}$.

Im folgenden werden für einige Verteilungen Erwartungswerte und Varianzen berechnet.

16.1.2 Beispiel

1. Sei X eine diskrete Zufallsvariable mit Verteilung

k	2	4	10	15	20
$\mathbb{P}(X = k)$	0.1	0.5	0.3	0.08	0.02

Dann gilt für den Erwartungswert $\mathbb{E}(X) = 2 \cdot 0.1 + 4 \cdot 0.5 + 10 \cdot 0.3 + 15 \cdot 0.08 + 20 \cdot 0.02 = 0.2 + 2 + 3 + 1.2 + 0.4 = 6.8$. Für die Varianz erhält man

$$\text{var}(X) = (2-6.8)^2 \cdot 0.1 + (4-6.8)^2 \cdot 0.5 + (10-6.8)^2 \cdot 0.3 + (15-6.8)^2 \cdot 0.08 + (20-6.8)^2 \cdot 0.02 = 23.536.$$

2. **Einpunktverteilung** $X \sim \varepsilon_a$: Da $T_X = \{a\}$ gilt, folgt sofort aus Definition 13.1:

$$\mathbb{E}(X) = \sum_{i=1}^1 a \mathbb{P}(X = a) = a, \quad \text{var}(X) = \sum_{i=1}^1 (a - \mathbb{E}X)^2 \mathbb{P}(X = a) = 0.$$

3. **Diskrete Gleichverteilung** $X \sim \mathbf{G}(x_1, \dots, x_n)$: Da $\mathbb{P}(X = x_i) = \frac{1}{n}$ gilt, folgt:

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \quad \text{var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s_x^2.$$

4. **Poissonverteilung** $X \sim \text{Poi}(\lambda)$: Für den Erwartungswert der Poissonverteilung gilt:

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

16.1.3 Bemerkung

1. Ist der Träger T_X der Zufallsvariablen X endlich, so entspricht der Erwartungswert $\mathbb{E}(X)$ in der deskriptiven Statistik dem gewogenen arithmetischen Mittel der Trägerpunkte x_1, \dots, x_n bzgl. der Häufigkeitsverteilung $f_j = \mathbb{P}(X = x_j)$. Eine analoge Aussage gilt für die Varianz.
2. Der Erwartungswert $\mathbb{E}(X)$ kann bei einem Glücksspiel als durchschnittliche Auszahlung pro Spiel interpretiert werden, wenn das Spiel sehr oft durchgeführt wird.

Vereinbart man beim einfachen Würfelwurf eine Auszahlung in EUR in Höhe der gewürfelten Augenzahl, so ergibt sich ein Erwartungswert von $E(X) = 3.5$ [EUR]. Daher ist dieses Spiel fair, wenn der Einsatz 3.5 EUR beträgt. Liegt der Einsatz niedriger, etwa bei 3 EUR, so liegt der **erwartete Gewinn** eines Spielers bei $3.5 - 3 = 0.5$ EUR. Ist der Einsatz höher, etwa 4 EUR, so ist der erwartete Gewinn $3.5 - 4 = -0.5$ EUR. Der Spieler erzielt also einen **erwarteten Verlust** von 0.5 EUR.

16.1.4 Übungsaufgabe

Sei die Verteilung von (X, Y) durch die Tabelle in Übungsaufgabe 15.3.5 gegeben. Bestimmen Sie $E(X)$, $E(Y)$, $\text{var}(X)$ und $\text{var}(Y)$.

16.2 Stetige Wahrscheinlichkeitsverteilungen

Sei X eine stetige Zufallsvariable mit Dichtefunktion f . Alle im folgenden auftretenden Integrale werden als existent vorausgesetzt.

16.2.1 Definition

Der **Erwartungswert** der Zufallsvariablen X ist definiert durch

$$E(X) = \int_{-\infty}^{\infty} t \cdot f(t) dt.$$

Die **Varianz** der Zufallsvariablen X ist definiert durch

$$\text{var}(X) = E(X - EX)^2 = \int_{-\infty}^{\infty} (t - EX)^2 \cdot f(t) dt.$$

Die **Standardabweichung** der Zufallsvariablen X ist definiert durch $\sqrt{\text{var}(X)}$.

Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion, so gilt für den Erwartungswert der transformierten Zufallsvariablen $h(X)$:

$$E(h(X)) = \int_{-\infty}^{\infty} h(t) \cdot f(t) dt.$$

Speziell für $h(x) = x^k$ erhält man das **k -te Moment** von X : $m_k = E(X^k)$, $k \in \mathbb{N}$.

Im folgenden werden für einige Verteilungsbeispiele Erwartungswerte und Varianzen berechnet.

16.2.2 Beispiel

1. **Rechteckverteilung** $X \sim \mathbf{R}[a, b]$:

$$\begin{aligned} E(X) &= \int_a^b t \cdot \frac{1}{b-a} dt = \frac{1}{b-a} \frac{1}{2} t^2 \Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}, \\ \text{var}(X) &= \int_a^b \left(t - \frac{b+a}{2}\right)^2 \cdot \frac{1}{b-a} dt = \frac{1}{b-a} \frac{1}{3} \left(t - \frac{b+a}{2}\right)^3 \Big|_a^b \\ &= \frac{1}{b-a} \left(\frac{1}{3} \left(\frac{b-a}{2}\right)^3 - \frac{1}{3} \left(\frac{a-b}{2}\right)^3 \right) = \frac{1}{12} (b-a)^2. \end{aligned}$$

2. **Exponentialverteilung** $X \sim \mathbf{Exp}(\lambda)$:

$$\mathbb{E}(X) = \int_0^\infty t \cdot \lambda e^{-\lambda t} dt = -te^{-\lambda t} \Big|_0^\infty + \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda}.$$

3. Für die **Normalverteilung** $\mathbf{N}(\mu, \sigma^2)$ gilt:

$$\mathbb{E}(X) = \mu, \quad \text{var}(X) = \sigma^2.$$

16.3 Eigenschaften des Erwartungswertes und der Varianz

In diesem Abschnitt werden Eigenschaften von Erwartungswerten und Varianzen betrachtet. Hierbei ist eine Unterscheidung in diskrete und stetige Verteilungen nicht notwendig. **Die Eigenschaften gelten unabhängig vom Verteilungstyp!**

Eigenschaften von Erwartungswerten.

Seien X, Y, X_1, \dots, X_n Zufallsvariablen.

1. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
2. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ für $a, b \in \mathbb{R}$
3. $\mathbb{E}\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n a_i \mathbb{E}(X_i) + b$ für $a_1, \dots, a_n, b \in \mathbb{R}$
4. Sind X und Y stochastisch unabhängig, so gilt: $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.
5. Sind alle Trägerpunkte $x \in T_X$ nicht-negativ, d.h. $x \geq 0$ für $x \in T_X$ (kurz: $X \geq 0$), so gilt: $\mathbb{E}(X) \geq 0$.

Eigenschaften von Varianzen.

Seien X, Y, X_1, \dots, X_n Zufallsvariablen.

1. $\text{var}(X) \geq 0$
2. $\text{var}(X) = 0$ genau dann, wenn $X \sim \varepsilon_a$ für ein $a \in \mathbb{R}$
3. $\text{var}(aX + b) = a^2 \text{var}(X)$ für alle $a, b \in \mathbb{R}$.
4. Sind X und Y stochastisch unabhängig, so gilt: $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.
5. Sind X_1, \dots, X_n stochastisch unabhängig, so gilt

$$\text{var}\left(\sum_{i=1}^n a_i X_i + b\right) = \sum_{i=1}^n a_i^2 \text{var}(X_i) \quad \text{für } a_1, \dots, a_n, b \in \mathbb{R}.$$

6. **Steinersche Regel:** $\mathbb{E}(X - a)^2 = \text{var}(X) + (\mathbb{E}(X) - a)^2$ für beliebige $a \in \mathbb{R}$.
7. $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

8. Tschebyscheff-Ungleichung:

$$P(|X - EX| \geq \varepsilon) \leq \frac{\text{var}(X)}{\varepsilon^2}.$$

Mittels der Varianz kann daher eine einfache Abschätzung für die Wahrscheinlichkeit angegeben werden, dass die Zufallsvariable X um mehr als ε von ihrem Erwartungswert EX abweicht.

16.3.1 Beispiel

Für eine $\text{Bin}(1, p)$ -verteilte Zufallsvariable X_1 erhalten wir

$$E(X_1) = 0 \cdot (1 - p) + 1 \cdot p = p, \quad \text{var}(X_1) = p(1 - p).$$

Nach Beispiel 11.2 gilt für stochastisch unabhängige Zufallsvariablen X_1, \dots, X_n mit $X_i \sim \text{Bin}(1, p)$, $1 \leq i \leq n$: $\sum_{i=1}^n X_i \sim \text{Bin}(n, p)$, d.h. wir erhalten eine binomialverteilte Zufallsgröße als Summe von stochastisch unabhängigen Bernoulli-verteilten Zufallsvariablen. Unter Ausnutzung der Linearität des Erwartungswertes und der Varianz (bei Unabhängigkeit!) folgt:

$$E(X) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = np,$$
$$\text{var}(X) = \text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) = np(1 - p).$$

Mittels der Rechenregeln lassen sich die Verteilungsparameter daher leicht berechnen.

Erwartungswert und Varianz einiger Verteilungen

124

Christine Müller, Stochastik in den Ingenieurwissenschaften, WS 2011/2012

16.3 Eigenschaften des Erwartungswertes und der Varianz

Diskrete Verteilungen				
Verteilung	Bezeichnung	Träger	Erwartungswert	Varianz
Einpunktverteilung	ε_a	$\{a\}$	a	0
Diskrete Gleichverteilung	$G(x_1, \dots, x_n)$	$\{x_1, \dots, x_n\}$	\bar{x}	s_x^2
Binomialverteilung	$\text{Bin}(n, p)$	$\{0, \dots, n\}$	np	$np(1-p)$
Bernoulli-Verteilung	$\text{Bin}(1, p)$	$\{0, \dots, 1\}$	p	$p(1-p)$
Hypergeometrische Verteilung	$\text{Hyp}(n, r, s)$	$\{\max\{0, n-s\}, \dots, \min\{r, n\}\}$	$n \frac{r}{r+s}$	$n \frac{r}{r+s} \frac{s}{r+s} \frac{r+s-n}{r+s-1}$
Geometrische Verteilung	$\text{Geo}(p)$	\mathbb{N}_0	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson-Verteilung	$\text{Poi}(\lambda)$	\mathbb{N}_0	λ	λ

Stetige Verteilungen				
Rechteckverteilung	$R[a, b]$	$[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential-Verteilung	$\text{Exp}(\lambda)$	$[0, \infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Weibull-Verteilung	$W(\alpha, \beta)$	$[0, \infty)$	$\beta \frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right)$	$\beta^2 \left(\frac{2}{\alpha} \Gamma\left(\frac{2}{\alpha}\right) - \left(\frac{1}{\alpha} \Gamma\left(\frac{1}{\alpha}\right) \right)^2 \right)$
Gamma-Verteilung	$\Gamma(\alpha, \beta)$	$[0, \infty)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
χ^2 -Verteilung	χ_f^2	$[0, \infty)$	f	$2f$
F-Verteilung	F_{f_1, f_2}	$[0, \infty)$	$\frac{f_2}{f_2-2}$ $(f_2 > 2)$	$\frac{2f_2^2(f_1+f_2-2)}{f_1(f_2-2)^2(f_2-4)}$ $(f_2 > 4)$
Normalverteilung	$N(\mu, \sigma^2)$	\mathbb{R}	μ	σ^2
t-Verteilung	t_f	\mathbb{R}	0	$\frac{f}{f-2}$ $(f > 2)$

16.3.2 Übungsaufgabe

Die Dreiecks-Verteilung ist eine stetige Wahrscheinlichkeitsverteilung, deren Dichte auf dem Träger eine Dreiecksgestalt hat. Bestimmen Sie die Dichte der Dreiecksverteilung für den Träger $[0, 4]$. Sei X eine Zufallsvariable, die diese Dreiecksverteilung besitzt. Bestimmen Sie auch $E(X)$ und $\text{var}(X)$.

16.4 Gesetz der großen Zahlen**16.4.1 Satz** (Gesetz der großen Zahl)

Sind X_1, X_2, \dots stochastisch unabhängige Zufallsvariablen, die alle die gleiche Verteilung mit endlicher Varianz besitzen, dann konvergiert das arithmetische Mittel $\frac{1}{N} \sum_{n=1}^N X_n$ mit wachsendem N gegen den Erwartungswert $E(X_1)$. Die Konvergenz gilt dabei in einem der folgenden Sinne

$$\lim_{N \rightarrow \infty} P \left(\left| \frac{1}{N} \sum_{n=1}^N X_n - E(X_1) \right| > \epsilon \right) = 0 \text{ für alle } \epsilon > 0 \text{ (schwaches Gesetz der großen Zahlen)}$$

oder

$$P \left(\left\{ \omega; \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N X_n(\omega) = E(X_1) \right\} \right) = 1 \text{ (starkes Gesetz der großen Zahlen).}$$

Die Zufallsvariablen X_1, X_2, \dots können auch Indikatorfunktionen von folgender Form sein

$$X_n(\omega) = 1_A(Y_n(\omega)) = \begin{cases} 1, & \text{falls } Y_n(\omega) \in A, \\ 0, & \text{falls } Y_n(\omega) \notin A. \end{cases}$$

Dann ist das arithmetische Mittel $\frac{1}{N} \sum_{n=1}^N X_n(\omega)$ die relative Häufigkeit der Ereignisse, dass $Y_n(\omega)$ in A fällt. Sind Y_1, Y_2, \dots stochastisch unabhängig und besitzen sie die gleiche Verteilung, so gilt nach Satz 16.4.1, dass diese relative Häufigkeit von Y_n in A gegen $E(Y_n) = P(Y_n \in A) = P^{Y_n}(A)$ konvergiert. Da Zufallszahlen gerade die Realisierungen der Zufallsvariablen Y_1, Y_2, \dots bei einem speziellen ω darstellen, erhalten wir die am Anfang in Teil II gegebene Motivation von Wahrscheinlichkeit.

16.5 Stochastische Simulation von Erwartungswert und Varianz

Können Erwartungswert oder Varianz nicht explizit berechnet werden, können sie per Simulation bestimmt werden. Dazu erzeugt man von der Verteilung genügend Zufallszahlen x_1, x_2, \dots, x_N . Der Erwartungswert wird dann nach dem Gesetz der großen Zahlen (Satz 16.4.1) durch das arithmetische Mittel \bar{x} der Zufallszahlen approximiert.

Da die Varianz ein spezieller Erwartungswert ist, kann auch die Varianz durch

$$\frac{1}{N} \sum_{n=1}^N (x_n - E(X))^2$$

approximiert werden, wenn der Erwartungswert $E(X)$ bekannt ist. Ist dieser aber unbekannt, kann dieser durch das arithmetische Mittel der Zufallszahlen ersetzt werden, so dass folgende Approximation benutzt werden kann

$$\frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2.$$

Das ist bis auf den Faktor $\frac{1}{N}$ die empirische Varianz der Zufallszahlen, bei der der Faktor $\frac{1}{N-1}$ benutzt wird. Bei sehr großem N , was bei Simulationen der Fall sein sollte, ist es egal ob $\frac{1}{N}$ oder $\frac{1}{N-1}$ benutzt wird. Somit kann die Varianz durch die empirische Varianz der Zufallszahlen approximiert werden. Für kleine Datensätzen ist der Faktor $\frac{1}{N-1}$ und damit die empirische Varianz sogar besser.

Kapitel 17

Weitere wahrscheinlichkeitstheoretische Kennzahlen

17.1 p -Quantile

In Abschnitt 9 wurden p -Quantile für Zufallszahlen vorgestellt. Ein Analogon kann auch für Wahrscheinlichkeitsverteilungen definiert werden. Diese p -Quantile spielen innerhalb der induktiven Statistik eine grundlegende Rolle.

17.1.1 Definition (p -Quantil)

Sei X eine Zufallsvariable mit Verteilungsfunktion F .

Für eine Zahl $p \in (0, 1)$ wird das **p -Quantil** $Q_p = Q_p(X)$ der durch die Verteilungsfunktion $F = F^X$ festgelegten Verteilung P^X definiert durch den kleinsten Wert $x \in \mathbb{R}$, für den gilt:

$$F(x) \geq p.$$

Das 0.5-Quantil heißt **Median**, das 0.25-Quantil **unteres Quartil** und das 0.75-Quantil **oberes Quartil**.

17.1.2 Bemerkung

1. Die obige Definition stimmt nicht ganz mit der in Abschnitt 5.3 überein. Sie könnte aber genauso gegeben werden. Der Vorteil der obigen Definition ist aber, dass dadurch das p -Quantil immer eindeutig ist und wie folgt über die Verteilungsfunktion F bestimmt werden kann.
2. Ist F eine Treppenfunktion, d.h. die zugrundeliegende Verteilung ist diskret (etwa mit Träger $T_X = \{x_1, x_2, \dots\}$), so ist ein Trägerpunkt gleich dem p -Quantil. Es werden zwei Situationen unterschieden, wobei $x_1 < x_2 < \dots$ gilt:
 - Es gibt **einen** Trägerpunkt x_j mit $F(x_j) = p$. Dann ist x_j das p -Quantil: $Q_p = x_j$.
 - Es gibt **keinen** Trägerpunkt x_j mit $F(x_j) = p$. Dann gibt es zwei Fälle:

1. $F(x_1) > p$: Dann ist $Q_p = x_1$;
2. $F(x_1) < p$: Dann gibt es Trägerpunkte x_{j-1} und x_j mit $F(x_{j-1}) < p$ und $F(x_j) > p$, und es gilt: $Q_p = x_j$.

Sei nun $X \sim \text{Bin}(3, 0.5)$. Dann gilt:

x_i	0	1	2	3
$P(X = x_i)$	0.125	0.375	0.375	0.125
$F(x_i)$	0.125	0.5	0.875	1

Ist $p = 0.5$, so wird der Wert von der Verteilungsfunktion angenommen an der Stelle $x = 1$. Daher ist $x = 1$ der Median oder das 0.5-Quantil $Q_{0.5}$. Ist $p = 0.1$, so ist $F(x_1) = F(0) = 0.125 > 0.1$ und $Q_{0.1} = 0$. Ist $p = 0.9$, so gilt $F(x_3) = F(2) = 0.875 < 0.9$ und $F(x_4) = F(3) = 1 > 0.9$, so dass $Q_{0.9} = 3$.

3. Bei stetigen Verteilungen ist das p -Quantil der kleinste x -Wert mit $F(x) = p$. Ist insbesondere F eine streng monotone Funktion, so ist Q_p die **eindeutig bestimmte Lösung der Gleichung** $F(x) = p$.

Sei $X \sim \text{Exp}(\lambda)$, d.h. X hat die Verteilungsfunktion

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

Die Gleichung $F(x) = p$ besitzt für $p \in (0, 1)$ eine eindeutige Lösung:

$$Q_p = -\frac{1}{\lambda} \log(1 - p).$$

4. Die Verteilungsfunktion der Standardnormalverteilung ist zwar streng monoton steigend, jedoch nicht in expliziter Form verfügbar. Daher müssen in diesem Fall die p -Quantile numerisch bestimmt werden.

Soll etwa $Q_{0.1}$ bestimmt werden, so kann benutzt werden, dass die Dichtefunktion der Standardnormalverteilung symmetrisch um den Nullpunkt ist. Für die Verteilungsfunktion gilt daher: $\Phi(x) = 1 - \Phi(-x)$. Für ein p -Quantil mit $p < 0.5$ nutzt man daher die Beziehung

$$Q_p = -Q_{1-p}.$$

Für das 0.1-Quantil erhält man daher $Q_{0.1} = -Q_{0.9} = -1.29$. Der Median ist $Q_{0.5} = 0$.

5. p -Quantile von Wahrscheinlichkeitsverteilungen erhält man in R aber ganz einfach über `qdistname`, wobei *distname* für den Verteilungsnamen steht. So liefert z.B. `qnorm(0.1)` das 0.1-Quantil der Standardnormalverteilung.

17.1.3 Übungsaufgabe

Die Zufallsvariable X besitze die diskrete Gleichverteilung auf $1, 2, \dots, 6$. Bestimmen Sie das 0.25-Quantil $Q_{0.25}(X)$ und das 0.75-Quantil $Q_{0.75}(X)$.

17.2 Abhängigkeitsmaße

Wenn zwei Zufallsvariablen X und Y nicht stochastisch unabhängig sind, können bedingte Wahrscheinlichkeiten berechnet werden. Man kann aber die Abhängigkeit auch durch eine Kennzahl erfassen. Dazu gibt es mehrere Möglichkeiten. Hier wird nur die Kovarianz und die Korrelation behandelt.

Sei $P^{(X,Y)}$ die gemeinsame Verteilung von X und Y . Sind X und Y diskret, so bezeichne $p_{ij} = P(X = x_i, Y = y_j)$ die Zähldichte von $P^{(X,Y)}$. Sind X und Y stetig, so bezeichne $f^{(X,Y)}$ die zugehörige Dichtefunktion. Für den Erwartungswert des Produktes $X \cdot Y$ gilt dann:

$$E(XY) = \begin{cases} \sum_{i,j} x_i \cdot y_j \cdot P(X = x_i, Y = y_j), & X, Y \text{ diskret,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f^{(X,Y)}(x, y) dy dx, & X, Y \text{ stetig.} \end{cases}$$

In analoger Weise wird der Erwartungswert von $h(X, Y)$ definiert, wobei $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine Funktion ist:

$$E(h(X, Y)) = \begin{cases} \sum_{i,j} h(x_i, y_j) \cdot P(X = x_i, Y = y_j), & X, Y \text{ diskret,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f^{(X,Y)}(x, y) dy dx, & X, Y \text{ stetig.} \end{cases}$$

Basierend auf dieser Definition werden nun Kovarianz und Korrelation der Zufallsvariablen X und Y definiert.

17.2.1 Definition (Kovarianz, Korrelation)

Seien X und Y Zufallsvariablen. Dann heißt

$$\text{kov}(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))]$$

Kovarianz von X und Y . Die Größe

$$\text{korr}(X, Y) = \frac{\text{kov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

heißt **Korrelation** von X und Y .

17.2.2 Bemerkung

1. Kovarianz und Korrelation sind symmetrisch, d.h. es gilt: $\text{kov}(X, Y) = \text{kov}(Y, X)$ und $\text{korr}(X, Y) = \text{korr}(Y, X)$.
2. $\text{kov}(X, Y) = E(XY) - E(X) \cdot E(Y)$.
3. $-1 \leq \text{korr}(X, Y) \leq 1$, wobei $\text{korr}(X, Y) = 1$ genau dann, wenn es Zahlen $a > 0$ und $b \in \mathbb{R}$ gibt mit $P(Y = aX + b) = 1$. $\text{korr}(X, Y) = -1$ ist genau dann erfüllt, wenn es Zahlen $a < 0$ und $b \in \mathbb{R}$ gibt mit $P(Y = aX + b) = 1$. Y ist in beiden Fällen mit Wahrscheinlichkeit 1 eine lineare Funktion von X .

4. Gilt $\text{korr}(X, Y) < 0$, so heißen die Zufallsvariablen X und Y **negativ korreliert**. Gilt $\text{korr}(X, Y) > 0$, so heißen die Zufallsvariablen X und Y **positiv korreliert**.
5. Gilt $\text{korr}(X, Y) = 0$, so heißen die Zufallsvariablen X und Y **unkorreliert**. Es gilt: Sind X und Y stochastisch unabhängig, so sind sie auch unkorreliert. **Die Umkehrung ist i.a. aber nicht richtig!** Eine Ausnahme ist die zweidimensionale Normalverteilung.
6. $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{kov}(X, Y)$, d.h. die Formel $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ gilt auch für unkorrelierte Zufallsvariablen.
7. $\text{kov}(aX + b, cY + d) = ac \cdot \text{kov}(X, Y)$ für alle $a, b, c, d \in \mathbb{R}$.
8. $\text{kov}(X, X) = \text{var}(X)$, $\text{kov}(Y, Y) = \text{var}(Y)$

17.2.3 Beispiel

Seien X und Y Zufallsvariablen mit folgender Zähldichte:

		Y				
		P(X = x _i , Y = y _j)	0	1	2	P(X = x _i)
X	-1		$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{2}{5}$
	1		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{3}{5}$
		P(Y = y _j)	$\frac{7}{20}$	$\frac{9}{20}$	$\frac{1}{5}$	1

Dann gilt:

$$E(X) = (-1) \cdot \frac{2}{5} + 1 \cdot \frac{3}{5} = \frac{1}{5} = 0.2$$

$$E(X^2) = (-1)^2 \cdot \frac{2}{5} + 1^2 \cdot \frac{3}{5} = 1$$

$$\text{var}(X) = E(X^2) - (E(X))^2 = 1 - \frac{1}{25} = \frac{24}{25} = 0.96$$

$$E(Y) = 0 \cdot \frac{7}{20} + 1 \cdot \frac{9}{20} + 2 \cdot \frac{1}{5} = \frac{17}{20} = 0.85$$

$$E(Y^2) = 0^2 \cdot \frac{7}{20} + 1^2 \cdot \frac{9}{20} + 2^2 \cdot \frac{1}{5} = \frac{5}{4} = 1.25$$

$$\text{var}(Y) = E(Y^2) - (E(Y))^2 = \frac{5}{4} - \frac{17^2}{20^2} = \frac{211}{400} = 0.5275$$

$$E(XY) = (-1) \cdot 0 \cdot \frac{1}{10} + (-1) \cdot 1 \cdot \frac{1}{5} + (-1) \cdot 2 \cdot \frac{1}{10} + 1 \cdot 0 \cdot \frac{1}{4} + 1 \cdot 1 \cdot \frac{1}{4} + 1 \cdot 2 \cdot \frac{1}{10} = \frac{1}{20} = 0.05$$

$$\text{kov}(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{20} - \frac{1}{5} \cdot \frac{17}{20} = -\frac{3}{25} = -0.12$$

$$\text{korr}(X, Y) = \frac{\text{kov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}} = \frac{-\frac{3}{25}}{\sqrt{\frac{24}{25} \cdot \frac{211}{400}}} = -\sqrt{\frac{6}{211}} = -0.169.$$

17.2.4 Beispiel

Ist $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, so gilt

$$\text{korr}(X, Y) = \rho,$$

und X und Y sind stochastisch unabhängig genau dann, wenn $\text{korr}(X, Y) = \rho = 0$.

17.2.5 Übungsaufgabe

Sei die Verteilung von (X, Y) durch die Tabelle in Übungsaufgabe 15.3.5 gegeben. Bestimmen Sie $\text{kov}(X, Y)$ und $\text{korr}(X, Y)$.

Kapitel 18

Markovketten

Ein wichtiges Thema der Wahrscheinlichkeitstheorie ist das Studium von stochastischen Prozessen, d.h. von Familien von Zufallsvariablen, die meist die zeitliche, gelegentlich die räumliche, Entwicklung eines Zufallsgeschehens beschreiben. Neben den Folgen von unabhängigen Zufallsvariablen, die bisher betrachtet wurden, ist eine Klasse von Prozessen besonders wichtig, die man Markovsche Ketten oder Markov-Ketten nennt. Sie sind durch eine spezielle übersichtliche Form der Abhängigkeit der Variablen charakterisiert. Markov-Ketten werden zum Beispiel zur Simulation von Warteschlangen oder im Bereich der Qualitätssicherung in der Fertigungskontrolle benutzt.

18.0.6 Definition (Stochastischer Prozess)

Eine Familie $\{X_t; t \in T\}$ von Zufallsvariablen mit Werten in I heißt **stochastischer Prozess mit Parameterbereich T und Zustandsraum I** .

18.0.7 Definition (Markov-Kette)

Eine **Markov-Kette** ist ein stochastischer Prozess $\{X_n; n \in \mathbb{N} \cup \{0\}\}$ mit abzählbarem Zustandsraum I , der die folgende **Markovsche Eigenschaft** besitzt: Für alle $n \in \mathbb{N} \cup \{0\}$ und alle $i_0, i_1, \dots, i_{n+1} \in I$ mit $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) > 0$ gilt

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n).$$

Die Zufallsvariable X_n wird als Zustand eines Systems zur Zeit n interpretiert. Der Prozess hat die Markovsche Eigenschaft, wenn die Wahrscheinlichkeit, zur Zeit $n+1$ in einem beliebigen Zustand zu gelangen, nur vom Zustand zur Zeit n und von n abhängt, aber nicht davon, in welchen Zuständen das System früher war. Hängt die Wahrscheinlichkeit, dass das System zur Zeit $n+1$ in einem beliebigen Zustand gelangt, nur vom Zustand zur Zeit n und nicht von n ab, so heißt die Markov-Kette **homogen** oder Kette mit **stationären Übergangswahrscheinlichkeiten**. Für homogene Markov-Ketten gilt für alle $i, j \in I$ und alle $n \in \mathbb{N} \cup \{0\}$

$$P(X_{n+1} = i | X_n = j) = p_{ij}.$$

Ist der Zustandsraum I endlich, so kann mit den Übergangswahrscheinlichkeiten $P(X_{n+1} = i | X_n = j)$ eine Matrix $\mathcal{P}_{n,n+1} = (P(X_{n+1} = i | X_n = j))_{i,j \in I}$ gebildet werden. Diese Matrix wird **Über-**

gangsmatrix genannt und ist eine **stochastische Matrix**, d.h. es gilt

$$P(X_{n+1} = i | X_n = j) \geq 0 \text{ für alle } i, j \in I \text{ und } \sum_{i \in I} P(X_{n+1} = i | X_n = j) = 1 \text{ für alle } j \in I.$$

Kennt man die Wahrscheinlichkeiten der Zustände $j \in I$ zum Zeitpunkt n , so können bei einer Markov-Kette mit dem Satz von der totalen Wahrscheinlichkeit (Satz 15.1.5) die Wahrscheinlichkeiten der Zustände zum Zeitpunkt $n + 1$ berechnet werden:

$$P(X_{n+1} = i) = \sum_{j \in I} P(X_{n+1} = i | X_n = j) P(X_n = j).$$

Diese Beziehung kann auch mit der Übergangsmatrix $\mathcal{P}_{n,n+1}$ ausgedrückt werden, wenn generell p_n der Vektor mit den Wahrscheinlichkeiten $P(X_n = i)$ ist, d.h. $p_n = (P(X_n = i))_{i \in I}$:

$$p_{n+1} = \mathcal{P}_{n,n+1} p_n. \quad (18.1)$$

Da bei einer homogenen Markovkette die Übergangswahrscheinlichkeiten nicht von der Zeit n abhängen, hängt auch die Übergangsmatrix nicht von n ab und es kann $\mathcal{P} = \mathcal{P}_{n,n+1} = (p_{ij})_{i,j \in I}$ gesetzt werden. Damit vereinfacht sich (18.1) zu

$$p_{n+1} = \mathcal{P} p_n.$$

Wiederholtes Anwenden ergibt

$$p_{n+m} = \mathcal{P}^m p_n \text{ für } m \in \mathbb{N}.$$

Ist p_n ein Eigenvektor von \mathcal{P} zum Eigenwert 1 (d.h. es gilt $\mathcal{P} p_n = p_n$), so ändert sich die Wahrscheinlichkeitsverteilungen bei $n + m$ nicht mehr, denn dann gilt

$$p_{n+m} = \mathcal{P}^m p_n = \mathcal{P}^{m-1} p_n = \dots = \mathcal{P} p_n = p_n.$$

18.0.8 Definition

Sei die homogene Markovkette gegeben durch die Übergangsmatrix \mathcal{P} . Eine Wahrscheinlichkeitsverteilung gegeben durch $\pi = (\pi_i)_{i \in I}$ heißt **invariante Wahrscheinlichkeitsverteilung** der homogenen Markovkette, falls gilt

$$\pi = \mathcal{P} \pi.$$

18.0.9 Satz

Ist die homogene Markovkette gegeben durch die Übergangsmatrix \mathcal{P} und gibt es $l \in \mathbb{N}$, so dass die Matrix \mathcal{P}^l nur echt positive Einträge hat, so gilt für alle p_n :

$$\mathcal{P}^M p_n \text{ konvergiert mit } M \rightarrow \infty \text{ gegen die invariante Wahrscheinlichkeitsverteilung } \pi.$$

18.0.10 Bemerkung

- Die Bedingung, dass es $l \in \mathbb{N}$ gibt, so dass die Matrix \mathcal{P}^l nur echt positive Einträge hat, impliziert, dass mit positiver Wahrscheinlichkeit von jedem Zustand j nach endlich vielen Schritten zu jedem Zustand i gelangt werden kann. Man sagt auch, dass die Zustände i und j kommunizieren. Allerdings gilt nicht allgemein, dass es ein l gibt, so dass \mathcal{P}^l nur echt positive Einträge hat, wenn alle Zustände miteinander kommunizieren. Ein Gegenbeispiel erhält man mit zwei Zuständen 0 und 1 und $p_{01} = 1, p_{10} = 1$.
- Satz 18.0.9 gilt auch, wenn p_n durch ein Einpunkt-Maß gegeben ist, d.h. wenn p_n ein Vektor ist, der eine Eins und ansonsten lauter Nullen enthält. Das kann so interpretiert werden, dass zum Zeitpunkt n in einem bestimmten Zustand, nämlich der Zustand, der die Wahrscheinlichkeit 1 hat, gestartet wird.
- Generell wird eine Markovkette wie folgt realisiert: Man startet zur Zeit $n = 0$ in einem Zustand $i_0 \in I$. Mit der Wahrscheinlichkeitsverteilung $P(X_1 = i | X_0 = i_0), i \in I$, wird ein Zustand i_1 zur Zeit $n = 1$ realisiert. Ausgehend von diesem Zustand i_1 wird mit der Wahrscheinlichkeitsverteilung $P(X_2 = i | X_1 = i_1), i \in I$, ein Zustand i_2 zur Zeit $n = 2$ realisiert und so weiter. Im ganzen erhält man eine Folge von Zuständen $i_0, i_1, i_2, i_3, \dots$. Man kann nun auch noch zeigen, dass unter den Voraussetzungen des Satzes 18.0.9 die relativen Häufigkeiten der Zustände $i \in I$ innerhalb der Folge $i_0, i_1, i_2, i_3, \dots, i_N$ mit $N \rightarrow \infty$ gegen die Wahrscheinlichkeiten $\pi_i, i \in I$, der invarianten Wahrscheinlichkeitsverteilung konvergieren. Dies soll anhand der Fragestellung von Beispiel 1.0.8 demonstriert werden.

18.0.11 Beispiel (Anzahl kaputter Fahrzeuge, siehe Beispiel 1.0.8)

Wenn von 6 Fahrzeugen pro Tag nur ein Fahrzeug repariert werden kann und die Fahrzeuge unabhängig von einander kaputt gehen, so können am Ende eines Tages 0,1,2,3,4,5,6 Fahrzeuge kaputt sein. Beobachtet man die Anzahl X_n der kaputten Fahrzeuge über mehrere Tage n , so haben wir einen stochastischen Prozess mit einem Zustandsraum $I = \{0, 1, 2, 3, 4, 5, 6\}$ und einem Parameterraum, der durch die Tage n gegeben ist. Dieser stochastische Prozess ist eine Markov-Kette, da die Anzahl der kaputten Fahrzeuge an einem Tag nur davon abhängt, wie viele Fahrzeuge am Tag davor kaputt waren. Es ist auch eine homogene Markov-Kette, da die Übergangswahrscheinlichkeiten von einem Tag zum nächsten Tag nicht vom speziellen Tag abhängen (Sonntage und Feiertage werden hier einfach nicht berücksichtigt). Ist p die Wahrscheinlichkeit, dass ein Fahrzeug an einem Tag kaputt geht, so werden die Übergangswahrscheinlichkeiten wie folgt bestimmt (dabei wird angenommen, dass jedes Fahrzeug mit gleicher Wahrscheinlichkeit p kaputt geht und dass die Reparatur eines Fahrzeuges nur am Morgen eines Tages vor Inbetriebnahme der Fahrzeuge erfolgt):

1. Ist am Abend des Vortrages **kein** Fahrzeug kaputt, so muss am Morgen des Folgetages kein Fahrzeug repariert werden und im Laufe des Tages können 0,1,2,3,4,5,6 Fahrzeuge kaputt gehen. Die Wahrscheinlichkeiten dieser 7 Zustände sind durch die Binomialverteilung $\text{Bin}(6, p)$ gegeben:

$$P(X_{n+1} = i | X_n = 0) = \binom{6}{i} p^i (1-p)^{6-i} \text{ für } i = 0, 1, 2, 3, 4, 5, 6.$$

2. Ist am Abend des Vortrages **ein** Fahrzeug kaputt, so wird dieses am Morgen des folgenden Tages repariert und im Laufe des Tages können 0,1,2,3,4,5,6 Fahrzeuge kaputt gehen. In diesem

Fall liegt wieder eine $\text{Bin}(6, p)$ -Verteilung vor:

$$P(X_{n+1} = i | X_n = 1) = \binom{6}{i} p^i (1-p)^{6-i} \text{ für } i = 0, 1, 2, 3, 4, 5, 6.$$

3. Sind am Abend des Vortrages j Fahrzeuge kaputt ($j = 2, 3, 4, 5$), so sind nach der Reparatur eines Fahrzeuges am Morgen des folgenden Tages $j - 1$ Fahrzeuge kaputt und im Laufe des Tages können noch $0, 1, \dots, 6 - j + 1$ weitere Fahrzeuge kaputt gehen, was gemäß einer $\text{Bin}(6 - j + 1, p)$ -Verteilung eintritt. Am Ende des Tages können dann $j - 1, j, \dots, 6$ Fahrzeuge kaputt sein. Es gelten also folgende Übergangswahrscheinlichkeiten:

$$P(X_{n+1} = i | X_n = j) = \binom{6-j+1}{i-j+1} p^{i-j+1} (1-p)^{6-i} \text{ für } i = j-1, j, \dots, 6,$$

$$P(X_{n+1} = i | X_n = j) = 0 \text{ für } i = 0, 1, \dots, j-2.$$

4. Sind am Abend des Vortrages **6** Fahrzeuge kaputt, so sind nach der Reparatur eines Fahrzeuges am Morgen des folgenden Tages 5 Fahrzeuge weiterhin kaputt und im Laufe des Tages können noch 0 oder 1 Fahrzeug kaputt gehen, was gemäß einer $\text{Bin}(1, p)$ -Verteilung eintritt. Am Ende des Tages können dann 5 oder 6 Fahrzeuge kaputt sein. Diese 2 Zustände treten mit folgenden Wahrscheinlichkeiten auf:

$$P(X_{n+1} = 5 | X_n = 6) = 1 - p, \quad P(X_{n+1} = 6 | X_n = 6) = p.$$

Ansonsten gilt $P(X_{n+1} = i | X_n = 6) = 0$ für $i = 0, 1, 2, 3, 4$.

Es ist klar, dass X_1, X_2, \dots eine homogene Markovkette bildet, die kommunizierende Zustände besitzt, da von jeder Anzahl von kaputten Fahrzeugen man in endliche vielen Tagen mit positiver Wahrscheinlichkeit zu einer anderen Anzahl kaputter Fahrzeuge gelangen kann.

Die Datei `UeMatFahrzeuge.asc` enthält die Funktion, die die Übergangsmatrix erzeugt:

```
"UeMatFahrzeuge" <-
function (p=0.4,k=6)
{
# Berechnet Uebergangsmatrix mit den Uebergangswahrscheinlichkeiten
# für das Fahrzeuge Beispiel
uemat<-dbinom(0:k,size=k,prob=p)
uemat<-c(uemat,uemat)
for(j in 2:k){
  uemat<-c(uemat, rep(0,j-1), dbinom(0:(6-j+1),size=6-j+1,prob=p))
}
uemat<-matrix(uemat,ncol=(k+1), byrow=F)
uemat
}
```

Folgender Aufruf der Funktion `UeMatFahrzeuge` liefert z.B. die Übergangsmatrix für $p = 0.4$:


```
> UeMatFahrzeuge(p=0.4)
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,] 0.046656 0.046656 0.000000 0.000000 0.000000 0.000000 0.000000
[2,] 0.186624 0.186624 0.077776 0.000000 0.000000 0.000000 0.000000
[3,] 0.311040 0.311040 0.259200 0.129600 0.000000 0.000000 0.000000
[4,] 0.276480 0.276480 0.345600 0.345600 0.216000 0.000000 0.000000
[5,] 0.138240 0.138240 0.230400 0.345600 0.432000 0.360000 0.000000
[6,] 0.036864 0.036864 0.076800 0.153600 0.288000 0.480000 0.600000
[7,] 0.004096 0.004096 0.010240 0.025600 0.064000 0.160000 0.400000
```

Die Übergangsmatrix ist also

$$P = \begin{pmatrix} 0.0467 & 0.0467 & 0.0000 & 0.0000 & 0.000 & 0.00 & 0.0 \\ 0.1866 & 0.1866 & 0.0778 & 0.0000 & 0.000 & 0.00 & 0.0 \\ 0.3110 & 0.3110 & 0.2592 & 0.1296 & 0.000 & 0.00 & 0.0 \\ 0.2765 & 0.2765 & 0.3456 & 0.3456 & 0.216 & 0.00 & 0.0 \\ 0.1382 & 0.1382 & 0.2304 & 0.3456 & 0.432 & 0.36 & 0.0 \\ 0.0369 & 0.0369 & 0.0768 & 0.1536 & 0.288 & 0.48 & 0.6 \\ 0.0041 & 0.0041 & 0.0102 & 0.0256 & 0.064 & 0.16 & 0.4 \end{pmatrix}.$$

Mit der Funktion `MarkovKette` in der Datei `MarkovKette.asc` kann nun eine homogene Markovkette zu einer beliebigen Übergangsmatrix erzeugt werden:

```
"MarkovKette" <-
function (uemat, N=4, start=1)
{
# Simuliert eine Markovkette mit der Übergangsmatrix uemat
# für die Zustände 0,1,...,ncol(uemat)-1
# Startwert:
i<- start
mk<-i-1
for(n in 1:N){
  u<-runif(1)
  for(l in 1:ncol(uemat)){
    if(i == l){
      psum<-cumsum(uemat[,l])
      for(k in 1:nrow(uemat)){
        if(k==1){
          if(u<=psum[1])
            {j<-1}
        }
        else{
          if(psum[k-1]<u & u<=psum[k])
            {j<-k}
        }
      }
    }
  }
}
```

```

    }
  }
  i<-j
  mk<-c(mk,i-1)
}
list(Markovkette=mk)
}

```

Wendet man diese Funktion z.B. auf die Übergangsmatrix für das Fahrzeug-Beispiel an, so erhält man z.B. folgende Markovketten:

```

> MarkovKette(UeMatFahrzeuge(p=0.4),N=30)
$Markovkette
[1] 0 3 4 3 2 5 6 5 4 4 4 4 4 4 4 6 6 6 5 4 5 4 3 3 2 3 3 4 5 5
> MarkovKette(UeMatFahrzeuge(p=0.2),N=30)
$Markovkette
[1] 0 1 2 2 2 3 3 2 1 0 3 2 4 4 3 3 4 3 3 2 2 3 4 3 3 2 3 3 4 4 5

```

Die invariante Wahrscheinlichkeitsverteilung π kann über drei Methoden bestimmt werden:

1. Durch Bestimmung eines Eigenvektors von \mathcal{P} zum Eigenwert 1, der nur positive Einträge hat und normiert ist.
2. Durch Approximation über $\mathcal{P}^M p_n$ mit großem M , wobei für $p_n = (\frac{1}{k}, \dots, \frac{1}{k})$ benutzt werden kann, wenn $p_n \in \mathbb{R}^k$ gilt.
3. Durch Ermittlung der relativen Häufigkeiten der Zustände in einer langen simulierten Markovkette.

Diese drei Möglichkeiten werden in der Funktion `InvariantesP`, die in der Datei `InvariantesP.asc` gegeben ist, durchgeführt und die Ergebnisse für die bestimmten invarianten Wahrscheinlichkeitsverteilungen werden mittels Balkendiagramme gegenübergestellt.

```

"InvariantesP" <-
function (UeMat=UeMatFahrzeuge(0.4),epsilon=0.00001,M=100,N=100)
{
  # Bestimmt die invariante Verteilung der
  # Markovkette gegen durch die Uebergangsmatrix UeMat
  # 1. durch Approximation mit maximal M Schritten und
  #   Genauigkeit epsilon
  # 2. exakt über Bestimmung des Eigenvektors zum Eigenwert 1
  # 3. durch Simulation der Markovkette über N Perioden
  k<-ncol(UeMat)
  par(mfrow=c(1,3))
  # Approximation der invarianten Verteilung
  p<-rep(1,k)/k
  p1<-UeMat%*%p
  m<-1
  while(sum((p-p1)^2)>epsilon & m<M){

```

```

    p<-p1
    p1<-UeMat%*%p1
    m<-m+1
  }
  cat("Anzahl der Durchlaufe bis Abbruch: ",m,"\n")
  p1<-as.vector(p1)
  names(p1)<-as.character(((1:k)-1))
  barplot(p1,main="Approx. invariante Verteilung")
  # Berechnung der exakten invarianten Verteilung
  Qeigen<-eigen(UeMat)
  q<- eigen(UeMat)$vectors[,Qeigen$values<=1+epsilon & Qeigen$values>=1-epsilon]
  q<-q/sum(q)
  names(q)<-as.character(((1:k)-1))
  barplot(q,main="Invariante Verteilung")
  # Simulation der invarianten Verteilung
  r<-table(MarkovKette(UeMat,N))/N
  barplot(r,main="Simulierte invariante Verteilung")
  list(Approximiert=p1,Exakt=q,Simuliert=r)
}

```

Die Abbildung 18.0.1, die die Ergebnisse der drei Methoden zur Bestimmung der invarianten Wahrscheinlichkeitsverteilung zeigt, wurde mit folgendem Aufruf erzeugt:

```

> InvariantesP(UeMat=UeMatFahrzeuge(0.4),epsilon=0.00001,M=100,N=1000)
Anzahl der Durchlaufe bis Abbruch: 10
$Approximiert
      0      1      2      3      4      5
0.0001174179 0.0022831675 0.0227945837 0.1221133099 0.3273884327 0.3833037689
      6
0.1419993192

$Exakt
      0      1      2      3      4      5
0.0001041711 0.0021285766 0.0220150754 0.1204806640 0.3269297519 0.3852140366
      6
0.1431277245

$Simuliert
      0      1      2      3      4      5      6
0.002 0.003 0.027 0.111 0.326 0.402 0.130

```

Somit ist die invariante Verteilung durch

$$\pi = (\pi_0, \pi_1, \dots, \pi_6)^\top \approx (0.0001, 0.0021, 0.0220, 0.1205, 0.3269, 0.3852, 0.1431)^\top$$

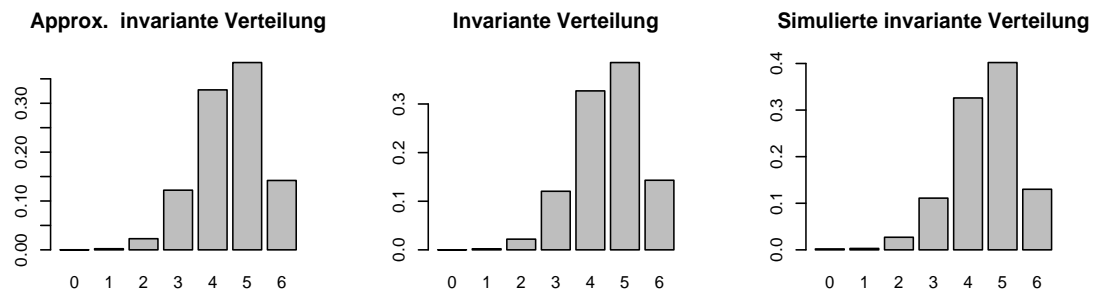


Abbildung 18.0.1: Verschiedene Methoden zur Bestimmung der invarianten Wahrscheinlichkeitsverteilung

gegeben. Nachrechnen der Definition der invarianten Verteilung ergibt:

$$P\pi = \begin{pmatrix} 0.0467 & 0.0467 & 0.0000 & 0.0000 & 0.000 & 0.00 & 0.0 \\ 0.1866 & 0.1866 & 0.0778 & 0.0000 & 0.000 & 0.00 & 0.0 \\ 0.3110 & 0.3110 & 0.2592 & 0.1296 & 0.000 & 0.00 & 0.0 \\ 0.2765 & 0.2765 & 0.3456 & 0.3456 & 0.216 & 0.00 & 0.0 \\ 0.1382 & 0.1382 & 0.2304 & 0.3456 & 0.432 & 0.36 & 0.0 \\ 0.0369 & 0.0369 & 0.0768 & 0.1536 & 0.288 & 0.48 & 0.6 \\ 0.0041 & 0.00401 & 0.0102 & 0.0256 & 0.064 & 0.16 & 0.4 \end{pmatrix} \cdot \begin{pmatrix} 0.0001 \\ 0.0021 \\ 0.0220 \\ 0.1205 \\ 0.3269 \\ 0.3852 \\ 0.1431 \end{pmatrix} \approx \begin{pmatrix} 0.0001 \\ 0.0021 \\ 0.0220 \\ 0.1205 \\ 0.3269 \\ 0.3852 \\ 0.1431 \end{pmatrix} = \pi$$

Das Ergebnis der invarianten Verteilung kann dazu benutzt werden, zu ermitteln, wie groß die Wahrscheinlichkeit ist, dass eine bestimmte Anzahl von Fahrzeugen pro Tag kaputt ist. So ist die Wahrscheinlichkeit für einen Tag ohne kaputte Fahrzeuge mit $\pi_0 = 0.0001$ sehr klein. Die Wahrscheinlichkeit für einen Tag, bei dem alle sechs Fahrzeuge kaputt sind, ist dagegen $\pi_6 = 0.1431$. Man kann auch bestimmen, mit wie viel kaputten Fahrzeugen man pro Tag im Mittel rechnen muss. Dazu sei X eine Zufallsvariable, die die invariante Verteilung als Verteilung hat, d.h. $P(X = i) = \pi_i$ für $i = 0, 1, \dots, 6$. Dann ist deren Erwartungswert die Anzahl der kaputten Fahrzeugen, mit denen man pro Tag im Mittel rechnen muss. Wegen

$$E(X) = 0 \cdot 0.0001 + 1 \cdot 0.0021 + 2 \cdot 0.0220 + 3 \cdot 0.1205 + 4 \cdot 0.3269 + 5 \cdot 0.3852 + 6 \cdot 0.1431 = 4.4998$$

muss man also mit ungefähr 4.5 kaputten Fahrzeugen pro Tag rechnen. Die Variabilität dieser Anzahl wird mit der Standardabweichung angegeben. Mit

$$E(X^2) = 0^2 \cdot 0.0001 + 1^2 \cdot 0.0021 + 2^2 \cdot 0.0220 + 3^2 \cdot 0.1205 + 4^2 \cdot 0.3269 + 5^2 \cdot 0.3852 + 6^2 \cdot 0.1431 = 21.1866$$

ergibt sich eine Varianz von

$$\text{var}(X) = E(X^2) - E(X)^2 = 21.1866 - 4.4998^2 = 0.9384$$

und somit eine Standardabweichung von $\sqrt{\text{var}(X)} = 0.9687105$. Also man muss mit ungefähr 4.5 kaputten Fahrzeugen plus/minus einem kaputten Fahrzeug pro Tag rechnen, bzw. 4.5 ± 1 ist die erwartete Anzahl von kaputten Fahrzeugen pro Tag.

18.0.12 Bemerkung

In Beispiel 18.0.11 wurde gezeigt, dass alle drei Methoden zur Bestimmung der invarianten Wahrscheinlichkeitsverteilung sehr ähnliche Ergebnisse liefern. Dabei ist die Methode, die die Markovkette simuliert, die aufwendigste Methode. Allerdings ist diese Methode eine sinnvolle Methode, wenn die Übergangsmatrix sehr groß ist. Sie ist sogar die einzige mögliche Methode, wenn die Übergangsmatrix nicht explizit angegeben werden kann, da man gar nicht alle möglichen Zustände kennt. Das ist zum Beispiel bei den sogenannten **Markovketten Monte Carlo Verfahren** der Fall.

18.0.13 Übungsaufgabe

Simulieren Sie die Markovkette für das Fahrzeugbeispiel aus Beispiel 1.0.8 für folgende Wahrscheinlichkeiten p :

$$p = 0.1 \text{ und } p = 0.3.$$

Bestimmen Sie für beide Fälle die invariante Verteilung und zwar über Approximation mit $M = 100$, exakt und über Simulation einer Markovkette mit $N = 1000$. Stellen Sie die Ergebnisse grafisch dar.

- (a) *Wie groß ist die Wahrscheinlichkeit bei den gewonnenen invarianten Verteilungen, dass an einem Tag alle Fahrzeuge kaputt sind?*
- (b) *Wie groß ist die erwartete Anzahl kaputter Fahrzeuge pro Tag? Wie stark streut diese erwartete Anzahl?*

Teil III

Schließende Statistik

Kapitel 19

Fragestellungen der Schließenden Statistik

Wir haben gesehen, dass die Zufallszahlen die Eigenschaften der Verteilung widerspiegeln, von der sie erzeugt werden. Wenn man nun in der Statistik Daten erhebt, so folgen sie auch einer Verteilung. Der Unterschied ist nur, dass diese Verteilung unbekannt ist und dass man in der Regel nur wenige Daten hat. D.h. man hat Daten x_1, \dots, x_N oder Datenpaare $(x_1, y_1), \dots, (x_N, y_N)$, bei denen N recht klein ist. Auf jeden Fall ist N nicht so groß, wie das bei den Zufallszahlen der Fall ist, wo N ohne weiteres als 1.000 oder sogar 10.000 gewählt werden kann. Wie man nun trotzdem mit relativ kleinen Datensätzen auf die zu Grunde liegende Wahrscheinlichkeitsverteilung geschlossen werden kann, ist Aufgabe der „Schließenden Statistik“. Ein anderer Name für „Schließende Statistik“ ist „Inferenz-Statistik“.

Wir werden hier vor allem den Fall betrachten, dass die Daten x_1, \dots, x_N eindimensional sind. Dabei werden wir immer annehmen, dass die Daten x_1, \dots, x_N Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen X_1, \dots, X_N , die alle die gleiche Verteilung X_0 besitzen. D.h. x_1, \dots, x_N könnten auch Zufallszahlen von der Verteilung X_0 sein. Die Annahme, dass die Zufallsvariablen X_1, \dots, X_N stochastisch unabhängig und identisch verteilt sind, ist so häufig, dass diese Eigenschaft mit **u.i.v** (unabhängig und identisch verteilt) oder noch kürzer mit **iid** (independent and identically distributed) abgekürzt wird. $X = (X_1, \dots, X_N)$ ist dann ein Zufallsvektor, der als Realisierung den Stichprobenvektor $x = (x_1, \dots, x_N)$ besitzt.

In der Regel ist nicht das Ziel, die ganze unbekannte zu Grunde liegende Verteilung P^{X_0} zu bestimmen, sondern man ist nur an bestimmten Kennzahlen der Wahrscheinlichkeitsverteilung interessiert.

19.0.14 Beispiel

Von den Daten x_1, \dots, x_N weiß man, dass sie im Prinzip jeden Wert in \mathbb{R} annehmen können, d.h. die zu Grunde liegende Wahrscheinlichkeitsverteilung ist eine stetige Verteilung. Z.B. können x_1, \dots, x_N Schraubendurchmesser sein. Von diesen Schraubendurchmessern aber interessiert eventuell nicht die ganze Verteilung, sondern nur der Erwartungswert und die Varianz, d.h. die Frage, ob die Schraubendurchmesser eine vorgegebene Zielgröße im Mittel einhalten und nicht zu sehr um diese Zielgröße streuen. Sind x_1, \dots, x_N Realisierungen von unabhängigen Zufallsvariablen X_1, \dots, X_N , die alle die gleiche Verteilung besitzen, so interessiert also $\mu = \mathbf{E}(X_1) = \dots = \mathbf{E}(X_N)$ sowie $\sigma^2 = \text{var}(X_1) = \dots = \text{var}(X_N)$. Diese sind unbekannt und anhand der Daten x_1, \dots, x_N soll eine Aussage

über μ und σ^2 gemacht werden.

In vieler Hinsicht vereinfacht sich das Problem, wenn man aus irgendwelchem Grunde weiß, dass die Schraubendurchmesser eine Normalverteilung $N(\mu, \sigma^2)$ besitzen. Dann sind μ und σ^2 die Parameter der Normalverteilung und die Verteilung wird durch diese beiden Parameter vollständig bestimmt. Allerdings bleibt das Problem, dass μ und σ^2 unbekannt sind und anhand von x_1, \dots, x_N bestimmt werden müssen.

In der Beschreibenden Statistik lassen sich drei wichtige Grundtypen von Verfahren angeben, die für unterschiedliche Arten von Aussagen verwendet werden können:

- **Punktschätzungen:** Hierbei soll ein spezieller Wert, der für die zu Grunde liegende Wahrscheinlichkeitsverteilung charakteristisch ist, geschätzt werden (etwa mittlerer Schraubendurchmesser μ oder Streuung σ). Die in Teil I behandelten statistischen Kennzahlen für die Lage, die Streuung und den Zusammenhang sind solche Punktschätzungen für die wahrscheinlichkeitstheoretischen Analoga. Z.B. ist das arithmetische Mittel von x_1, \dots, x_N eine Punktschätzung für den Erwartungswert μ .
- **Intervallschätzungen:** Da Punktschätzungen i.a. nur sehr ungenaue Prognosen liefern, werden oft **Konfidenzintervalle** angegeben. Diese Bereiche werden so konstruiert, daß die Wahrscheinlichkeit, daß der untersuchte (unbekannte) Parameter in dem angegebenen Bereich liegt, einer vorgegebenen (hohen) Wahrscheinlichkeit entspricht. In obigem Beispiel bedeutet dies etwa, ein Intervall $[\hat{\mu}_1, \hat{\mu}_2]$ anzugeben mit $P(\mu \in [\hat{\mu}_1, \hat{\mu}_2]) \geq 0.95$.
- **Hypothesentests:** In vielen Fällen sollen konkrete Hypothesen bzgl. des untersuchten Parameters untersucht werden. Kennzeichnend für ihr Konstruktionsprinzip ist, daß richtige Hypothesen nur mit einer kleinen Wahrscheinlichkeit abgelehnt werden sollen. Im obigen Beispiel kann etwa die Hypothese, dass die Streuung σ der zu Grunde liegenden Verteilung nicht größer als ein vorgegebener Wert σ_0 ist, untersucht werden.

Hypothesentests und Intervallschätzungen werden im folgenden für verschiedene Modellannahmen untersucht. Dabei wird immer wieder auf Resultate aus der Wahrscheinlichkeitstheorie zurückgegriffen werden.

Kapitel 20

Allgemeine Form eines statistischen Testes

Bis auf entartete Fälle wird das arithmetische Mittel zum Beispiel von Schraubendurchmessern nicht mit einer vorgegebenen Zielgröße μ_0 übereinstimmen. Es stellt sich die Frage, wie weit muss das arithmetische Mittel von μ_0 abweichen, damit man schließen kann, dass der Produktionsprozess der Schrauben die vorgegebene Zielgröße nicht erfüllt, d.h. dass man den Produktionsprozess stoppen muss.

Auch wenn die Zufallsvariablen X_0 und Y_0 stochastisch unabhängig sind, so wird die empirische Kovarianz von Zufallszahlenpaaren $(x_1, y_1), \dots, (x_N, y_N)$ die gemäß der Verteilung von (X_0, Y_0) erzeugt wurden, ungleich 0 sein, obwohl $\text{kov}(X_0, Y_0) = 0$ gilt. Die Abweichung von 0 wird umso größer sein, je kleiner die Anzahl N ist. Hier stellt sich die Frage, wie sehr muss die empirische Kovarianz von 0 abweichen, damit man schließen kann, dass X_0 und Y_0 nicht stochastisch unabhängig sind.

Solche Fragen beantworten die **statistischen Tests**.

Generell hat man eine

Nullhypothese oder Hypothese H_0 ,

und eine

Alternative-Hypothese oder Alternative H_1 ,

wie zum Beispiel:

$H_0 : E(X_0) = \mu_0$, d.h. die Produktion hält die Zielgröße ein.

$H_1 : E(X_0) \neq \mu_0$, d.h. die Zielgröße μ_0 wird nicht eingehalten.

oder

$H_0 : X_0$ und Y_0 sind stochastisch unabhängig.

$H_1 : \text{Es gibt einen Zusammenhang zwischen } X_0 \text{ und } Y_0.$

Um solche Fragen zu beantworten, wird aus dem Datensatz x eine **Teststatistik** $T(x)$ berechnet, die in der Regel durch die Schätzung für die entsprechende Fragestellung gegeben ist. Da der Daten-

satz ein zufälliges Ergebnis ist, ist auch $T(x)$ ein zufälliges Ergebnis. $T(x)$ ist damit die Realisierung eines Zufallprozesses $T(X)$, der mathematisch auch Zufallsvariable genannt wird. Man betrachtet nun die Wahrscheinlichkeit, dass die Teststatistik einen höheren oder gleichen Wert als den Wert $T(x)$ annimmt, der beim Datensatz x beobachtet wurde, falls die Nullhypothese gilt. Diese Wahrscheinlichkeit ist der **P-Wert**, der von Statistik-Programmen ausgegeben wird.

20.0.15 Definition (P-Wert)

Der P-Wert ist die maximale Wahrscheinlichkeit unter der Nullhypothese, dass die Zufallsvariable $T(X)$ einen Wert größer gleich $T(x)$ annimmt, der beim Datensatz x beobachtet wurde, d.h. der P-Wert ist also

$$P_{H_0}(T(X) \geq T(x)).$$

Der P-Wert, d.h. die maximale Wahrscheinlichkeit unter der Nullhypothese, dass die Zufallsvariable $T(X)$ einen höheren Wert als $T(x)$ annimmt, ist klein, wenn $T(x)$ einen sehr hohen (extremen) Wert angenommen hat. Ist diese Wahrscheinlichkeit zu klein, wird die Nullhypothese abgelehnt. Dabei muss von vornherein festgelegt, was als zu klein betrachtet wird. Dazu legt man $\alpha \in (0, 1)$ fest und spricht dann von einem **Test zum Niveau α** , wenn gilt

$$P_{H_0}(T(X) \geq T(x)) \leq \alpha \implies H_0 \text{ wird abgelehnt bzw. die Daten sprechen signifikant gegen } H_0, \text{ bzw. Entscheidung für } H_1.$$

$$P_{H_0}(T(X) \geq T(x)) > \alpha \implies \text{Die Daten sprechen nicht gegen } H_0, \text{ eventuell sind es zu wenige Daten.}$$

In der Regel wird α wie folgt festgelegt:

$$\alpha = 0.05, \quad \text{falls ein Test am gleichen Datensatz gemacht wird.}$$

$$\alpha = \frac{0.05}{\text{Anzahl der Tests}}, \quad \text{falls mehrere Tests am gleichen Datensatz gemacht werden. (20.1)}$$

Die zweite Festlegung von α , auch **α -Adjustierung** genannt, wird sehr oft verletzt. Dabei muss diese α -Adjustierung auch gemacht werden, wenn mehrere Variablen des Datensatzes unabhängig voneinander getestet werden, denn diese Variablen gehören zum gleichen Datensatz. Diese α -Adjustierung muss auch gemacht werden, wenn nur Teile der Einzelbeobachtungen in mehrere Tests eingehen.

Ist $P_{H_0}(T(X) \geq T(x)) > \alpha$, kann nur die schwache Aussage gemacht werden, dass die Daten nicht gegen H_0 sprechen. Damit spricht aber **nichts** für H_0 . Es ist ein weit verbreiteter Fehlschluss, dass aus $P_{H_0}(T(X) \geq T(x)) > \alpha$ geschlossen wird, dass H_0 dann gelten muss. $P_{H_0}(T(X) \geq T(y)) > \alpha$ erhält man nämlich insbesondere, wenn man zu wenige Daten beobachtet hat. Umgekehrt kann man aber von $P_{H_0}(T(X) \geq T(x)) \leq \alpha$ auf H_1 schließen. Das liegt an folgenden möglichen Fehlentscheidungen:

	Entscheidung für H_0	Entscheidung für H_1
H_0 gilt	keine Fehlentscheidung	Fehlentscheidung 1. Art (α -Fehler)
H_1 gilt	Fehlentscheidung 2. Art (β -Fehler)	keine Fehlentscheidung

Entscheidet man sich für H_1 , so gilt wegen $P_{H_0}(T(X) \geq T(x)) \leq \alpha$ die Beziehung $T(x) \geq t$, wobei t der kleinste Wert mit $P_{H_0}(T(X) \geq t) \leq \alpha$ ist. Dann kann die Entscheidung richtig sein oder es liegt eine Fehlentscheidung 1. Art vor. Die Wahrscheinlichkeit für die Fehlentscheidung 1. Art ist aber

$$P_{H_0}(\text{Fehlentscheidung 1. Art}) \leq P_{H_0}(T(X) \geq t) \leq \alpha,$$

also sehr klein, wenn α sehr klein ist. Das bedeutet, dass, wenn viele einen Test zum Niveau $\alpha = 0.05$ anwenden, höchstens 5% der Anwender diese Fehlentscheidung machen. Das heißt, höchstens 5% der Anwender entscheiden sich fälschlicherweise für H_1 , d.h. höchstens 5% der Anwender entscheiden sich für H_1 , obwohl H_0 richtig ist.

Würde man aber umgekehrt bei $P_{H_0}(T(X) \geq T(x)) > \alpha$ eine Entscheidung für H_0 zulassen, dann muss man mit der Fehlentscheidung 2. Art rechnen. Nun kann aber $P_{H_1}(\text{Fehlentscheidung 2. Art})$ oft gar nicht berechnet werden und, wenn es doch berechnet werden kann, gilt in der Regel

$$P_{H_1}(\text{Fehlentscheidung 2. Art}) \geq 1 - \alpha,$$

womit die Wahrscheinlichkeit für diese Fehlentscheidung 2. Art extrem groß sein würde. Das bedeutet, dass, wenn viele einen Test zum Niveau $\alpha = 0.05$ anwenden, bis zu 95% der Anwender die Fehlentscheidung 2. Art treffen würden. Das heißt, bis zu 95% der Anwender würden sich dann für H_0 entscheiden, obwohl H_1 gilt. Das ist natürlich nicht akzeptabel. Daher interpretiert man eine Entscheidung für H_0 nur dahin, dass die Daten nicht gegen H_0 sprechen. Das ist aber kein Beleg für H_0 .

Genauso wird im Gericht vorgegangen, wo gilt „Im Zweifelsfall für den Angeklagten“: Wenn nicht genug Indizien vorhanden sind, um einen Angeklagten zu verurteilen, muss er freigesprochen werden, auch wenn er vielleicht der Täter war. Nur wenn genug Indizien (Daten) für seine Verurteilung vorhanden sind, kann der Angeklagte verurteilt werden (kann H_0 verworfen werden). D.h. die Nullhypothese H_0 entspricht der These der Verteidigung, dass der Angeklagte nicht der Täter ist.

	Entscheidung für H_0 $\hat{=}$ Freispruch	Entscheidung für H_1 $\hat{=}$ Verurteilung
H_0 gilt $\hat{=}$ Angeklagter ist nicht Täter	keine Fehlentscheidung	Fehlentscheidung 1. Art (α -Fehler)
H_1 gilt $\hat{=}$ Angeklagter ist Täter	Fehlentscheidung 2. Art (β -Fehler)	keine Fehlentscheidung

Die unterschiedlichen Wahrscheinlichkeiten für die Fehlentscheidungen 1. und 2. Art können auch durch stochastische Simulation ermittelt werden. Im folgenden werden die wichtigsten Tests vorgestellt und von einigen werden die Wahrscheinlichkeiten für die Fehlentscheidungen 1. und 2. Art dargestellt.

Kapitel 21

Tests für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ bei Normalverteilung

Kann man annehmen, dass die Daten x_1, \dots, x_N von einer Normalverteilung stammen, d.h. dass sie Realisierungen von unabhängigen, normalverteilten Zufallsvariablen X_1, \dots, X_N sind, vereinfacht sich vieles. So ist die Verteilung dann bis auf die beiden Parameter μ und σ^2 der Normalverteilung bekannt.

Wenn man eine Normalverteilung annimmt, so sind Hypothesen über den Erwartungswert Hypothesen über den Parameter μ . Hier wird das Testproblem $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ ausführlicher betrachtet.

21.1 Herleitung des Testes

Die Daten x_1, \dots, x_N sprechen gegen die Nullhypothese $H_0 : \mu = \mu_0$, wenn das arithmetische Mittel \bar{x} zu sehr von μ_0 abweicht. Da aber Daten verschieden stark streuen, ist es sinnvoll die Differenz $\bar{x} - \mu_0$ mit einer Streuungsschätzung zu normieren. Die geeignete Streuungsschätzung ist die Standardabweichung $s(x)$, d.h. die Wurzel aus der empirischen Varianz. Somit ist eine sinnvolle Teststatistik

$$T(x) = \sqrt{N} \frac{|\bar{x} - \mu_0|}{s(x)}.$$

Diese wird noch mit \sqrt{N} multipliziert, weil nämlich gilt:

$$X_1, \dots, X_N \sim N(\mu_0, \sigma^2) \implies \sqrt{N} \frac{\bar{X} - \mu_0}{s(X)} \sim t_{N-1},$$

d.h. $\sqrt{N} \frac{\bar{X} - \mu_0}{s(X)}$ besitzt eine t -Verteilung mit $N - 1$ -Freiheitsgraden, wenn die Nullhypothese richtig ist. Da die Nullhypothese nur aus dem Wert μ_0 besteht, ist der P-Wert

$$\begin{aligned} P_{H_0}(T(X) > T(x)) &= P_{\mu_0} \left(\sqrt{N} \frac{|\bar{X} - \mu_0|}{s(X)} > T(x) \right) \\ &= P_{\mu_0} \left(\sqrt{N} \frac{\bar{X} - \mu_0}{s(X)} > T(x) \right) + P_{\mu_0} \left(\sqrt{N} \frac{\bar{X} - \mu_0}{s(X)} < -T(x) \right) \\ &= 1 - F_{t_{N-1}}(T(x)) + F_{t_{N-1}}(-T(x)) = 2 F_{t_{N-1}}(-T(x)), \end{aligned}$$

wobei $F_{t_{N-1}}$ die Verteilungsfunktion der t -Verteilung mit $N - 1$ -Freiheitsgraden ist. Da jede t -Verteilung wie die Standardnormalverteilung symmetrisch um 0 ist, gilt die Gleichheit $1 - F_{t_{N-1}}(T(x)) + F_{t_{N-1}}(-T(x)) = 2 F_{t_{N-1}}(-T(x))$. Die Verteilungsfunktion der t -Verteilung ist in R durch `pt` gegeben. Da so ein Test auf der t -Verteilung basiert, wird dieser Test auch **t-Test** genannt.

21.1.1 Beispiel (Fortsetzung von Beispiel 1.0.1)

Bei Beispiel 1.0.1 stellte sich die Frage, ob die Zielgröße von 1 Mikrometer bei den Kugellagerdurchmessern von beiden Produktionslinien eingehalten wird. Nehmen wir eine Normalverteilung für die Durchmesser an, so können wir

$$H_0 : \mu = 1 \text{ gegen } H_1 : \mu \neq 1$$

testen. Für die erste Produktionslinie bekommen wir folgende Werte für das arithmetische Mittel, die Standardabweichung, die Teststatistik $T(x)$ und den P-Wert:

```
> k1<-c(1.18,1.42,0.69,0.88,1.62,1.09,1.53,1.02,1.19,1.32)
> mean(k1)
[1] 1.194
> sd(k1)
[1] 0.2896818
> sqrt(10)*abs(mean(k1)-1)/sd(k1)
[1] 2.117778
> tx<-sqrt(10)*abs(mean(k1)-1)/sd(k1)
> 2*pt(-tx,9)
[1] 0.06326551
```

Da der P-Wert größer als 0.05 ist, wird die Nullhypothese $H_0 : \mu = 1$ nicht abgelehnt.

Das Ergebnis können wir auch einfacher mit der R-Funktion `t.test` bekommen:

```
> t.test(k1,alternative="two.sided",mu=1)
```

One Sample t-test

```
data: k1
t = 2.1178, df = 9, p-value = 0.06327
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 0.9867741 1.4012259
```



```
sample estimates:
mean of x
1.194
```

Für die zweite Produktionslinie gilt $\bar{x} = 1.406$, $s(x) = 0.4283093$ und somit $T(x) = 2.9976$, was zu einem P-Wert von 0.01502 führt. Bei dieser Produktionslinie muss also die Nullhypothese $H_0 : \mu = 1$ abgelehnt werden.

21.2 Fehlerwahrscheinlichkeiten 2. Art

Auch wenn Statistik-Pakete P-Werte ausgeben, ist es für die Untersuchung der Wahrscheinlichkeit für die Fehlentscheidung 2. Art sinnvoll die Entscheidungsregel expliziter anzugeben. Für das Testproblem $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ gilt folgendes:

$$P_{H_0}(T(X) > T(x)) \leq \alpha \iff T(x) > t_{N-1;1-\alpha/2},$$

wobei $t_{N-1;1-\alpha/2}$ das $1 - \alpha/2$ -Quantil der t -Verteilung mit $N - 1$ Freiheitsgraden ist. Quantile der t -Verteilung bekommt man in *R* mittels `qt`.

21.2.1 Beispiel (Fortsetzung von Beispiel 21.1.1)

```
> qt(0.975,df=9)
[1] 2.262157
```

Da in der ersten Produktionslinie $T(x) = 2.1178$ und in der zweiten Produktionslinie $T(x) = 2.9976$ gilt, ist klar, dass in der ersten Produktionslinie die Nullhypothese nicht abgelehnt wird, während in der zweiten Produktionslinie die Ablehnung erfolgt.

Um die Wahrscheinlichkeiten für die Fehlentscheidung 2. Art zu bekommen, muss nun

$$P_{H_1}(\text{Entscheidung für } H_0) = P_\mu(T(X) < t_{N-1;1-\alpha/2})$$

für alle $\mu \neq \mu_0$ bestimmt werden. Das kann mittels stochastischer Simulation geschehen: Man erzeugt M Stichproben $x^m = (x_1^m, \dots, x_N^m)$, $m = 1, \dots, M$, mittels einer Normalverteilung $N(\mu, \sigma^2)$. Dabei kann $\sigma^2 = 1$ gesetzt werden. Für jede Stichprobe x^m wird $T(x^m)$ berechnet. Die relative Häufigkeit der Fälle $T(x^m) < t_{N-1;1-\alpha/2}$ ist dann eine Approximation von $P_\mu(T(X) < t_{N-1;1-\alpha/2})$, die nach dem Gesetz der großen Zahlen umso besser ist, desto größer die Anzahl M der Simulationen ist.

Man kann in diesem Fall $P_\mu(T(X) < t_{N-1;1-\alpha/2})$ aber auch exakt berechnen. Es gilt nämlich:

$$X_1, \dots, X_N \sim N(\mu, \sigma^2) \implies \sqrt{N} \frac{\bar{X} - \mu_0}{s(X)} \sim t_{N-1}(\sqrt{N}(\mu - \mu_0)/\sigma),$$

wobei $t_{N-1}(\delta)$ die nichtzentrale t -Verteilung mit $N - 1$ -Freiheitsgraden und Nichtzentralitätspara-

meter δ ist. Also gilt für die Wahrscheinlichkeit der Fehlentscheidung 2. Art

$$\begin{aligned} P_\mu(T(X) \leq t_{N-1;1-\alpha/2}) &= P_\mu\left(\sqrt{N} \frac{|\bar{X} - \mu_0|}{s(X)} \leq t_{N-1;1-\alpha/2}\right) \\ &= P_\mu\left(-t_{N-1;1-\alpha/2} \leq \sqrt{N} \frac{\bar{X} - \mu_0}{s(X)} \leq t_{N-1;1-\alpha/2}\right) \\ &= F_{t_{N-1}(\sqrt{N}(\mu - \mu_0)/\sigma)}(t_{N-1;1-\alpha/2}) - F_{t_{N-1}(\sqrt{N}(\mu - \mu_0)/\sigma)}(-t_{N-1;1-\alpha/2}), \end{aligned}$$

wobei $F_{t_{N-1}(\sqrt{N}(\mu - \mu_0)/\sigma)}$ die Verteilungsfunktion der nichtzentralen t -Verteilung mit $N-1$ -Freiheitsgraden und Nichtzentralitätsparameter $\sqrt{N}(\mu - \mu_0)/\sigma$ ist. In *R* erhält man die Verteilungsfunktion der nichtzentralen t -Verteilung mittels des Arguments **ncp**. Per Voreinstellung gilt **ncp=0**, was die zentrale t -Verteilung ist.

Statt $P_\mu(T(X) < t_{N-1;1-\alpha/2})$ kann auch $P_\mu(T(X) \geq t_{N-1;1-\alpha/2})$ betrachtet werden. Dies als Funktion $\gamma(\mu)$ von μ aufgefasst wird auch **Gütefunktion** genannt:

$$\begin{aligned} \gamma(\mu) &= P_\mu(T(X) > t_{N-1;1-\alpha/2}) \\ &= 1 - F_{t_{N-1}(\sqrt{N}(\mu - \mu_0)/\sigma)}(t_{N-1;1-\alpha/2}) + F_{t_{N-1}(\sqrt{N}(\mu - \mu_0)/\sigma)}(-t_{N-1;1-\alpha/2}). \end{aligned}$$

Dann gilt

$$\gamma(\mu_0) = \text{Wahrscheinlichkeit für Fehlentscheidung 1. Art}$$

und

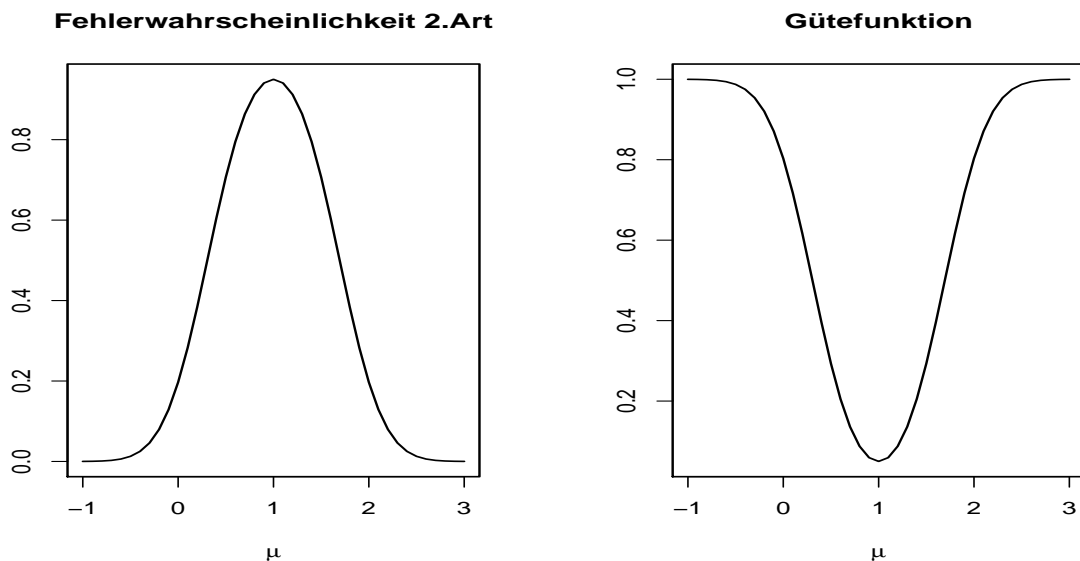
$$1 - \gamma(\mu) = \text{Wahrscheinlichkeit für Fehlentscheidung 2. Art, falls } \mu \neq \mu_0.$$

21.2.2 Beispiel (Fortsetzung von Beispiel 21.1.1)

Da σ nicht bekannt ist, nehmen wir hier einfach $\sigma = 1$ an. Mit folgenden *R*-Befehlen können die Wahrscheinlichkeiten der Fehlentscheidungen 2. Art (die Fehlerwahrscheinlichkeiten 2. Art) und die Gütefunktion grafisch dargestellt werden.

```
> mu<-seq(-4,6,0.1)
> plot(mu,pt(qt(0.975,df=9),df=9,ncp=sqrt(10)*(mu-1))-pt(-qt(0.975,df=9),
+ df=9,ncp=sqrt(10)*(1-mu)),type="l",ylab=" ",xlab=expression(mu))
> title("Fehlerwahrscheinlichkeit 2.Art")
> plot(mu,1-pt(qt(0.975,df=9),df=9,ncp=sqrt(10)*(mu-1))+pt(-qt(0.975,df=9),
+ df=9,ncp=sqrt(10)*(1-mu)),type="l",ylab=" ",xlab=expression(mu))
> title("Gütefunktion")
```

Daraus sieht man deutlich, dass bei Werten μ nahe bei $\mu_0 = 1$ die Wahrscheinlichkeiten der Fehlentscheidungen 2. Art sehr groß sind. Da bei $\mu = 1 = \mu_0$ die Gütefunktion den Wert $\alpha = 0.05$ annimmt, werden die Fehlerwahrscheinlichkeiten 2. Art ganz nahe bei 1 bis zu $0.95 = 1 - \alpha$ groß. Das ist eine allgemeine Eigenschaft von Tests.

Abbildung 21.2.1: Fehlerwahrscheinlichkeiten 2. Art und Gütefunktion bei $N=10$ und $\mu_0 = 1$

21.2.3 Bemerkung

- Durch Erhöhung des Stichprobenumfanges können die Fehlerwahrscheinlichkeiten 2. Art verringert werden und somit die Chancen eines Testes erhöht werden, dass er ein signifikantes Ergebnis liefert, wenn tatsächlich die Nullhypothese nicht erfüllt ist.
- Liegen die Fehlerwahrscheinlichkeiten 2. Art für bestimmte Alternativen ebenfalls unter 0.05, können diese Alternativen auch signifikant abgelehnt werden, so dass ein signifikantes Ergebnis für die Nullhypothese vorliegt.

21.3 + Wahl des Stichprobenumfanges +

Generell gilt also, dass die Fehlerwahrscheinlichkeiten 2. Art nahe bei μ_0 bis zu $1 - \alpha$ groß werden können. Möchte man diese aber in einer gewissen Entfernung von μ_0 kleiner als $\beta < 1 - \alpha$ bekommen, kann das mit einer genügend großen Stichprobe erreicht werden. Soll die Fehlerwahrscheinlichkeiten 2. Art bei allen Parametern μ mit $|\mu - \mu_0| > \delta \cdot \sigma$ kleiner als β sein, so bestimmt man den Stichprobenumfang N als ein Wert N mit

$$F_{t_{N-1}(\sqrt{N}\delta)}(t_{N-1;1-\alpha/2}) - F_{t_{N-1}(\sqrt{N}\delta)}(-t_{N-1;1-\alpha/2}) \leq \beta.$$

Wird unter diesen Umständen bei einem $\beta = 0.05$ die Nullhypothese $\mu = \mu_0$ nicht abgelehnt, so kann zumindest geschlossen werden, dass das unbekannte μ im Intervall $[\mu_0 - \delta \cdot \sigma, \mu_0 + \delta \cdot \sigma]$ liegt. Es kann aber weiterhin nicht geschlossen werden, dass $\mu = \mu_0$ gilt. So ein Schluss ist nie möglich!

21.3.1 Beispiel (Six Sigma Qualität)

Bei der Six-Sigma-Qualitätsanforderung (siehe Abschnitt 13.2.7) wird verlangt, dass die Abweichung

von einer Zielgröße μ_0 nicht mehr als 1.5σ betragen darf. Mit dem obigen Test kann $|\mu - \mu_0| \leq 1.5\sigma$ nur geschlossen werden, wenn die Nullhypothese $H_0 : \mu = \mu_0$ nicht abgelehnt wird und der Stichprobenumfang N so groß gewählt ist, dass die Fehlerwahrscheinlichkeit 2. Art für $|\mu - \mu_0| \geq 1.5\sigma$ kleiner gleich $\beta = 0.05$ ist. Damit ist $\delta = 1.5$. Durch Ausprobieren erhält man, dass der Stichprobenumfang mindestens $N = 8$ betragen muss:

```
> N<-7
> pt(qt(0.975,df=N-1),df=N-1,ncp=sqrt(N)*1.5)-pt(-qt(0.975,df=N-1),
+ df=N-1,ncp=sqrt(N)*1.5)
[1] 0.09233692
> N<-8
> pt(qt(0.975,df=N-1),df=N-1,ncp=sqrt(N)*1.5)-pt(-qt(0.975,df=N-1),
+ df=N-1,ncp=sqrt(N)*1.5)
[1] 0.04904818
```

21.3.2 Beispiel (Fortsetzung von Beispiel 21.1.1)

Mit $N = 10$ wird die erforderliche Stichprobengröße für die Six Sigma Qualität für beide Produktionslinien erreicht. Da in der ersten Produktionslinie die Nullhypothese $H_0 : \mu = 1$ nicht abgelehnt wurde, kann geschlossen werden, dass $|\mu - 1| \leq 1.5\sigma$ gilt. Bei der zweiten Produktionslinie wurde $H_0 : \mu = 1$ abgelehnt. Hier kann geschlossen werden, dass $\mu \neq 1$ ist, aber nicht mehr. Es könnte immer noch der Fall sein, dass $\mu \neq 1$ und $|\mu - 1| \leq 1.5\sigma$ gilt.

Man kann aber auch eine Verschärfung der Six Sigma Qualität betrachten: Soll z.B. mit dem t-Test auf $\mu \in [1 - 0.1\sigma, 1 + 0.1\sigma]$ geschlossen werden können, so ist $\delta = 0.1$ und die Stichprobengröße kann für $\beta = 0.05$ durch Ausprobieren wie folgt bestimmt werden:

```
> N<-1301
> pt(qt(0.975,df=N-1),df=N-1,ncp=sqrt(N)*0.1)-pt(-qt(0.975,df=N-1),
+ df=N-1,ncp=sqrt(N)*0.1)
[1] 0.0500563
> N<-1302
> pt(qt(0.975,df=N-1),df=N-1,ncp=sqrt(N)*0.1)-pt(-qt(0.975,df=N-1),
+ df=N-1,ncp=sqrt(N)*0.1)
[1] 0.04991328
```

Also muss die Stichprobengröße als $N = 1302$ gewählt werden, was in beiden Produktionslinien nicht erfüllt ist.

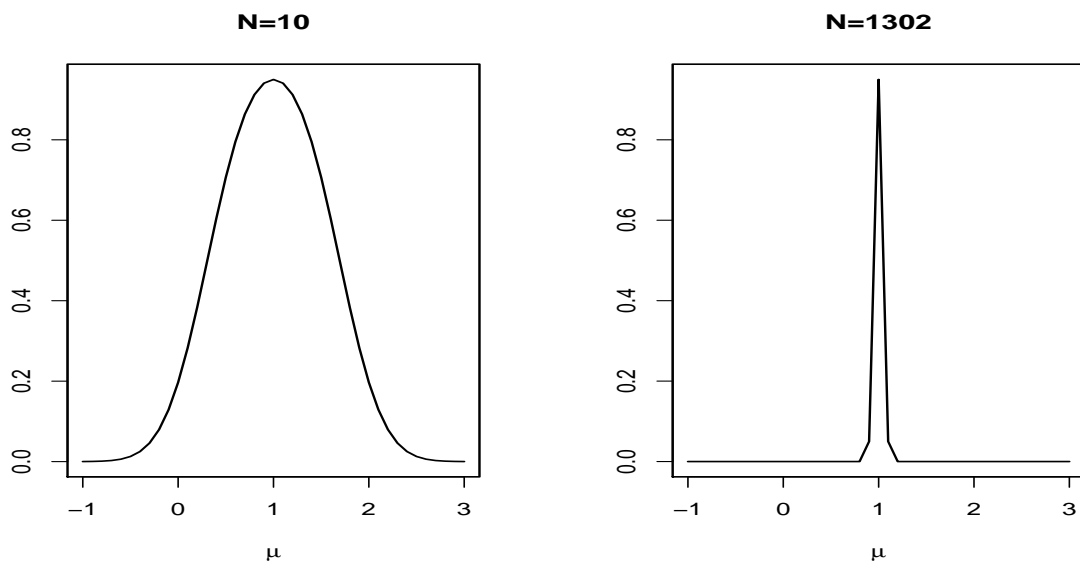


Abbildung 21.3.1: Fehlerwahrscheinlichkeiten 2. Art für $H_0 : \mu_0 = 1$ bei $\sigma = 1$ bei verschiedenen Stichprobengrößen

Kapitel 22

Weitere Tests bei Normalverteilung

In Abschnitt 21 wurde ausführlich das Testproblem $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ behandelt. Hier werden weitere Tests für die Normalverteilung vorgestellt. Wie beim Testproblem $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ sind bei einigen Tests auch Aussagen über die Fehlerwahrscheinlichkeiten 2. Art möglich. Kann aber keine Normalverteilung angenommen werden, müssen allerdings alternative Tests benutzt werden (siehe Abschnitt 23).

22.1 Tests von Hypothesen über den Erwartungswert

Wenn man eine Normalverteilung annimmt, so sind Hypothesen über den Erwartungswert Hypothesen über den Parameter μ . Neben der Fragestellung $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ können noch folgende Hypothesen betrachtet werden, wobei μ_0 und δ bekannte vorgegebene Werte sind:

1. $H_0 : \mu = \mu_0$ gegen $H_1 : \mu > \mu_0$
2. $H_0 : \mu = \mu_0$ gegen $H_1 : \mu < \mu_0$
3. $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$
4. $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$

In Abschnitt 21 wurde folgender Test hergeleitet:

Zweiseitiger t-Test für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$:

$$\text{Lehne } H_0 \text{ ab, falls } \sqrt{N} \frac{|\bar{x} - \mu_0|}{s(x)} > t_{N-1; 1-\alpha/2}.$$

Die Fehlerwahrscheinlichkeit 2. Art für $|\mu - \mu_0| = \delta \sigma > 0$ ist

$$F_{t_{N-1}(\sqrt{N}\delta)}(t_{N-1; 1-\alpha/2}) - F_{t_{N-1}(\sqrt{N}\delta)}(-t_{N-1; 1-\alpha/2}).$$

Aufruf in R: `t.test(x, alternative="two.sided", mu=...)`

Analog erhält man die anderen Tests. Dabei muss bei Hypothesen der Form $H_0 : \mu \leq \mu_0$ und $H_0 : \mu \geq \mu_0$ der P-Wert nur für den Wert μ am Rand der Hypothesenmenge bestimmt werden. Somit werden folgende t-Tests erhalten.

Einseitiger t-Test für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu > \mu_0$ oder $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$:

Lehne H_0 ab, falls $\sqrt{N} \frac{\bar{x} - \mu_0}{s(x)} > t_{N-1;1-\alpha}$.

Die Fehlerwahrscheinlichkeit 2. Art für $\mu - \mu_0 = \delta \sigma > 0$ ist

$$F_{t_{N-1}(\sqrt{N}\delta)}(t_{N-1;1-\alpha}).$$

Aufruf in R: `t.test(x, alternative="greater", mu=...)`

Einseitiger t-Test für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu < \mu_0$ oder $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$:

Lehne H_0 ab, falls $\sqrt{N} \frac{\mu_0 - \bar{x}}{s(x)} > t_{N-1;1-\alpha}$.

Die Fehlerwahrscheinlichkeit 2. Art für $\mu - \mu_0 = \delta \sigma < 0$ ist

$$F_{t_{N-1}(\sqrt{N}\delta)}(t_{N-1;1-\alpha}).$$

Aufruf in R: `t.test(x, alternative="less", mu=...)`

22.2 Tests von Hypothesen über die Varianz

Generell gilt:

$$X_1, \dots, X_N \sim N(\mu, \sigma_0^2) \implies (N-1) \frac{s(X)}{\sigma_0} \sim \chi_{N-1}^2,$$

wobei χ_{N-1}^2 die χ^2 -Verteilung mit $N-1$ Freiheitsgraden ist. Das α -Quantil der χ^2 -Verteilung mit $N-1$ Freiheitsgraden wird mit $\chi_{N-1;\alpha}^2$ bezeichnet.

Zweiseitiger Varianz-Test für $H_0 : \sigma = \sigma_0$ gegen $H_1 : \sigma \neq \sigma_0$:

Lehne H_0 ab, falls $(N-1) \frac{s(x)}{\sigma_0} > \chi_{N-1;1-\alpha/2}^2$ oder $(N-1) \frac{s(x)}{\sigma_0} < \chi_{N-1;\alpha/2}^2$.

Einseitiger Varianz-Test für $H_0 : \sigma \leq \sigma_0$ gegen $H_1 : \sigma > \sigma_0$:

Lehne H_0 ab, falls $(N-1) \frac{s(x)}{\sigma_0} > \chi_{N-1;1-\alpha}^2$.

Einseitiger Varianz-Test für $H_0 : \sigma \geq \sigma_0$ gegen $H_1 : \sigma < \sigma_0$:

Lehne H_0 ab, falls $(N-1) \frac{s(x)}{\sigma_0} < \chi_{N-1;\alpha}^2$.

22.2.1 Beispiel (Six Sigma Qualität der Kugellager, Fortsetzung von Beispiel 21.1.1)

Die wichtigste Anforderung der Six Sigma Qualität ist

$$\sigma \leq \frac{1}{6} \min\{USL - \mu_0, \mu_0 - LSL\}, \quad (22.1)$$

wobei LSL und USL die unteren und oberen Grenzen sind, innerhalb der eine Größe (z.B. Kugellagerdurchmesser) variieren darf (siehe Abschnitt 13.2.7). Um die Aussage (22.1) schließen zu können, muss die Nullhypothese des einseitigen Testes $H_0 : \sigma \geq \sigma_0$ gegen $H_1 : \sigma < \sigma_0$ abgelehnt werden, wobei $\sigma_0 = \frac{1}{6} \min\{USL - \mu_0, \mu_0 - LSL\}$ ist.

Wird in Beispiel 1.0.1 $\mu_0 = 1$, $LSL = 0.9$ und $USL = 1.2$ gesetzt, ist also $\sigma_0 = 0.1/6 = 0.01666667$.

```
> 9*sd(k1)/(0.1/6)
[1] 156.4282
> 9*sd(k2)/(0.1/6)
[1] 231.2870
> qchisq(0.05,df=9)
[1] 3.325113
```

Also gilt für die erste Produktionslinie

$$(N-1) \frac{s(x)}{\sigma_0} = 156.4282 > 3.325113 = \chi_{N-1;\alpha}^2$$

und für die zweite Produktionslinie

$$(N-1) \frac{s(x)}{\sigma_0} = 231.2870 > 3.325113 = \chi_{N-1;\alpha}^2,$$

weshalb in beiden Fällen $H_0 : \sigma \geq 0.01666667$ nicht abgelehnt wird. Also kann nicht auf Six-Sigma-Qualität geschlossen werden.

Im Gegenteil kann sogar geschlossen werden, dass dieser Qualitätsstandard nicht eingehalten wird. Dazu wird $H_0 : \sigma \leq \sigma_0$ gegen $H_1 : \sigma > \sigma_0$ getestet. In diesem Fall ist der kritische Wert $\chi_{N-1;1-\alpha}^2 = 16.91898$, der bei beiden Produktionslinien von der Teststatistik überschritten wird. Also kann $\sigma > 0.01666667$ geschlossen werden, weshalb beide Produktionslinien die Six Sigma Qualität nicht erfüllen.

22.3 Zweistichproben-Tests für gepaarte Stichproben

Hat man zwei sogenannte **gepaarte Stichproben** x_1, \dots, x_N und y_1, \dots, y_N , die eigentlich in der Form $(x_1, y_1), \dots, (x_N, y_N)$ vorliegen, so kann man immer zu $x_1 - y_1, \dots, x_N - y_N$ übergehen und hat dann einen eindimensionalen Datensatz. Gilt, dass x_1, \dots, x_N Realisierungen von iid $X_1, \dots, X_N \sim N(\mu_1, \sigma_1^2)$ und y_1, \dots, y_N Realisierungen von iid $Y_1, \dots, Y_N \sim N(\mu_2, \sigma_2^2)$ sind, so gilt $X_n - Y_n \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 - 2\text{kov}(X_n, Y_n))$. Damit kann z.B. $H_0 : \mu_1 = \mu_2$ gegen $H_1 : \mu_1 \neq \mu_2$ mit dem zweiseitigen t-Test für $H_0 : \mu_1 - \mu_2 = 0$ gegen $H_1 : \mu_1 - \mu_2 \neq 0$ basierend auf der einen Stichprobe $x_1 - y_1, \dots, x_N - y_N$ getestet werden. In R kann man aber die beiden Stichproben getrennt in den t-Test `t.test` eingeben, muss aber das Argument `paired = TRUE` dazu eingeben, also `t.test(x,y,alternative = "two.sided",paired=TRUE)`.

22.3.1 Beispiel (Fehlerhäufigkeiten einer Sortiermaschine)

Ein Hersteller produzierte Maschinen zum Sortieren von Hausmüll. Diese hatten aber eine hohe Fehlerquote beim Sortieren. Daher wurden sie umgebaut. Es ergaben sich vor und nach dem Umbau folgende Fehlerhäufigkeiten in Prozent:

Maschine	1	2	3	4	5
Fehlerhäufigkeit vor dem Umbau	30	61	59	44	62
Fehlerhäufigkeit nach dem Umbau	25	62	46	45	60

Der Hersteller möchte wissen, ob der Umbau etwas gebracht hat und ob es sinnvoll ist, auch andere Maschinen umzubauen. Stellt μ_x die erwartete Fehlerhäufigkeit vor dem Umbau und μ_y die erwartete Fehlerhäufigkeit nach dem Umbau dar, so sollte $H_0 : \mu_x \leq \mu_y$ gegen $H_0 : \mu_x > \mu_y$ getestet werden.

```
> x<-c(30,61,59,44,62)
> y<-c(25,62,46,45,60)
> t.test(x,y,alternative="greater",paired=TRUE)
      Paired t-test
data:  x and y
t = 1.3846, df = 4, p-value = 0.1192
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.942802      Inf
sample estimates:
mean of the differences
      3.6
```

Da der P-Wert nicht kleiner als 0.05 ist, spricht nichts gegen die Nullhypothese, d.h. es gibt kein Indiz dafür, dass der Umbau etwas gebracht hat. Das kann aber auch daran liegen, dass zu wenige Maschinen untersucht wurden.

22.4 Zweistichproben-Tests für ungepaarte Stichproben

Etwas ganz anderes gilt, wenn zwei **ungepaarte Stichproben** vorliegen. Dann sollten die beiden Stichproben stochastisch unabhängig sein und in der Regel sind auch die Stichprobenumfänge verschieden groß. D.h. man hat, dass x_1, \dots, x_{N_1} Realisierungen von iid $X_1, \dots, X_{N_1} \sim N(\mu_1, \sigma_1^2)$ und y_1, \dots, y_{N_2} Realisierungen von iid $Y_1, \dots, Y_{N_2} \sim N(\mu_2, \sigma_2^2)$ sind.

Tests zum Vergleich der beiden Varianzen σ_1^2 und σ_2^2 basieren auf der F-Verteilung. Das α -Quantil der F-Verteilung mit N und M Freiheitsgraden wird mit $F_{N,M;\alpha}$ bezeichnet.

Zweiseitiger Varianz-Test für zwei Stichproben für $H_0 : \sigma_1 = \sigma_2$ gegen $H_1 : \sigma_1 \neq \sigma_2$:

Lehne H_0 ab, falls $\frac{s(x)^2}{s(y)^2} > F_{N_1-1, N_2-1; 1-\alpha/2}$ oder $\frac{s(x)^2}{s(y)^2} < F_{N_1-1, N_2-1; \alpha/2}$.

Aufruf in R: `var.test(x,y,alternative = "two.sided")`.

Einseitiger Varianz-Test für zwei Stichproben für $H_0 : \sigma_1 \leq \sigma_2$ gegen $H_1 : \sigma_1 > \sigma_2$:

Lehne H_0 ab, falls $\frac{s(x)^2}{s(y)^2} > F_{N_1-1, N_2-1; 1-\alpha}$.

Aufruf in R: `var.test(x,y,alternative = "greater")`.

Einseitiger Varianz-Test für zwei Stichproben für $H_0 : \sigma_1 \geq \sigma_2$ gegen $H_1 : \sigma_1 < \sigma_2$:

Lehne H_0 ab, falls $\frac{s(x)^2}{s(y)^2} < F_{N_1-1, N_2-1; \alpha}$.

Aufruf in R: `var.test(x,y,alternative = "less")`.

Tests für den Vergleich der Erwartungswerte μ_1 und μ_2 basierenden wieder auf der t-Verteilung und werden auch t-Tests genannt. In der einfachen Form wird dabei angenommen, dass $\sigma_1^2 = \sigma_2^2 = \sigma^2$ gilt, d.h. dass die zugrunde liegenden Varianzen in beiden Stichproben gleich sind. Dann kann diese Varianz σ^2 mit

$$\begin{aligned} s_{12}(x, y)^2 &= \frac{1}{N_1 + N_2 - 2} ((N_1 - 1)s(x) + (N_2 - 1)s(y)) \\ &= \frac{1}{N_1 + N_2 - 2} \left(\sum_{n=1}^{N_1} (x_n - \bar{x})^2 + \sum_{n=1}^{N_2} (y_n - \bar{y})^2 \right) \end{aligned}$$

geschätzt werden. Außerdem wird die Abkürzung

$$K = \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

benutzt.

Zweiseitiger t-Test für zwei Stichproben für $H_0 : \mu_1 = \mu_2$ gegen $H_1 : \mu_1 \neq \mu_2$:

Lehne H_0 ab, falls $K \frac{|\bar{x} - \bar{y}|}{s_{12}(x, y)} > t_{N_1+N_2-2; 1-\alpha/2}$.

Die Fehlerwahrscheinlichkeit 2. Art für $|\mu_1 - \mu_2| = \delta \sigma > 0$ ist

$$F_{t_{N_1+N_2-2}(K\delta)}(t_{N_1+N_2-2; 1-\alpha/2}) - F_{t_{N_1+N_2-2}(K\delta)}(-t_{N_1+N_2-2; 1-\alpha/2}).$$

Aufruf in R: `t.test(x,y,alternative="two.sided", var.equal = TRUE)`

Einseitiger t-Test für $H_0 : \mu_1 \leq \mu_2$ gegen $H_1 : \mu_1 > \mu_2$:

Lehne H_0 ab, falls $K \frac{|\bar{x} - \bar{y}|}{s_{12}(x, y)} > t_{N_1+N_2-2; 1-\alpha}$.

Die Fehlerwahrscheinlichkeit 2. Art für $\mu_1 - \mu_2 = \delta \sigma > 0$ ist

$$F_{t_{N_1+N_2-2}(K\delta)}(t_{N_1+N_2-2; 1-\alpha}).$$

Aufruf in R: `t.test(x,y,alternative="greater", var.equal = TRUE)`

Einseitiger t-Test für $H_0 : \mu_1 \geq \mu_2$ gegen $H_1 : \mu_1 < \mu_2$:

Lehne H_0 ab, falls $K \frac{|\bar{x} - \bar{y}|}{s_{12}(x, y)} < t_{N_1+N_2-2; \alpha/2}$.

Die Fehlerwahrscheinlichkeit 2. Art für $\mu_1 - \mu_2 = \delta \sigma < 0$ ist

$$F_{t_{N_1+N_2-2}(K\delta)}(t_{N_1+N_2-2; 1-\alpha}).$$

Aufruf in R: `t.test(x,y,alternative="less",var.equal = TRUE)`

Lehnt der Varianz-Test die Gleichheit der Varianzen σ_1^2 und σ_2^2 ab, so kann eine Modifikation des Zweistichproben-t-Testes benutzt werden, der sogenannte Welch-Zweistichproben-t-Test. In R erhält man den Welch-Zweistichproben-t-Test mit der Voreinstellung `var.equal = FALSE` in `t.test`.

22.4.1 Beispiel (Fortsetzung von Beispiel 21.1.1)

Bei Beispiel 1.0.1 kann man sich auch fragen, ob in beiden Produktionslinien die gleiche Streuung auftritt. Dazu wird der Varianz-Test für $H_0 : \sigma_1 = \sigma_2$ gegen $H_1 : \sigma_1 \neq \sigma_2$ durchgeführt:

```
> var.test(k1,k2,alternative="two.sided")
```

F test to compare two variances

data: k1 and k2

F = 0.4574, num df = 9, denom df = 9, p-value = 0.2595

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.1136199 1.8416221

sample estimates:

ratio of variances

0.4574329

Da der P-Wert mit 0.2595 größer als $\alpha = 0.05$ kann die Nullhypothese $H_0 : \sigma_1 = \sigma_2$ nicht abgelehnt werden. D.h. nichts spricht gegen die Gleichheit der Varianzen. Das bedeutet nicht, dass die Varianzen wirklich gleich sein müssen. Es können auch nur zu wenige Daten vorliegen, um die Nullhypothese widerlegen zu können. Da aber die vorliegenden Daten nicht gegen die Gleichheit der Varianzen sprechen, kann der einfache t-Test zum Vergleich der Erwartungswerte μ_1 und μ_2 benutzt werden:

```
> t.test(k1,k2,alternative="two.sided",var.equal=TRUE)
```

Two Sample t-test

data: k1 and k2

t = -1.2965, df = 18, p-value = 0.2112

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.5555277 0.1315277

sample estimates:

mean of x mean of y

1.194 1.406

Da der P-Wert mit 0.2112 größer als $\alpha = 0.05$ kann die Nullhypothese $H_0 : \mu_1 = \mu_2$ nicht abgelehnt werden. D.h. nichts spricht gegen die Gleichheit der Erwartungswerte. Und dieses Ergebnis erhalten wir, obwohl der Einstichproben-t-Test für $H_0 : \mu = 1$ gegen $H_1 : \mu \neq 1$ in der ersten Produktionslinie nicht abgelehnt wurde und in der zweiten Produktionslinie aber abgelehnt wurde. Das zeigt aber nur, dass das **Nichtablehnen der Nullhypothese keine Aussagekraft** hat.

Obwohl nichts gegen die Gleichheit der Varianzen spricht, kann trotzdem auch der Welch-Zweistichproben-t-Test für verschiedene Varianzen durchgeführt werden:

```
> t.test(k1,k2,alternative="two.sided",var.equal=FALSE)
```

Welch Two Sample t-test

```
data: k1 and k2
t = -1.2965, df = 15.809, p-value = 0.2134
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5589723  0.1349723
sample estimates:
mean of x mean of y
  1.194      1.406
```

Der P-Wert des Welch-t-Testes ist mit 0.2134 ein wenig größer als beim einfachen t-Test, wo der P-Wert 0.2112 ist. Da das Ziel ein P-Wert kleiner als $\alpha = 0.05$ ist, damit die Nullhypothese signifikant abgelehnt werden kann und man auf die Alternativhypothese schließen kann, sollte der Welch-t-Test nur benutzt werden, wenn die Varianzen wirklich verschieden sind. Ansonsten fällt der P-Wert ein wenig zu hoch aus. Meistens wird das nichts ausmachen, aber manchmal kann es gerade bewirken, dass mit dem Welch-t-Test keine Ablehnung der Nullhypothese möglich wird.

22.5 + Relevanz- und Äquivalenztests +

Wird der Stichprobenumfang sehr groß, wird eine Nullhypothese der Form $H_0 : \mu = \mu_0$ oder $H_0 : \mu_1 = \mu_2$ mit immer höherer Wahrscheinlichkeit abgelehnt, da in der Regel die Gleichheit nicht exakt vorliegt und mit wachsendem Stichprobenumfang auch sehr kleine Abweichungen von der Gleichheit erkannt werden. Aber oft will man gar nicht sehr kleine Abweichungen erkennen, sondern man möchte nur relevante Abweichungen aufspüren. Ein relevanter Unterschied wird am besten als Vielfaches der Standardabweichung σ ausgedrückt, d.h. durch $\delta\sigma$ mit $\delta > 0$. Im Einstichproben-Fall lauten dann die Null- und Alternativhypothesen wie folgt:

$$H_0 : |\mu - \mu_0| \leq \delta\sigma \text{ gegen } H_1 : |\mu - \mu_0| > \delta\sigma.$$

Der zugehörige Tests ist der sogenannte **Relevanz-t-Test** für $H_0 : |\mu - \mu_0| \leq \delta\sigma$ gegen $H_1 : |\mu - \mu_0| > \delta\sigma$:

$$\text{Lehne } H_0 \text{ ab, falls } \sqrt{N} \frac{|\mu_0 - \bar{x}|}{s(x)} > c_\alpha,$$

wobei c_α der Wert ist mit

$$\alpha = 1 - F_{t_{N-1}(\sqrt{N}\delta)}(c_\alpha) + F_{t_{N-1}(\sqrt{N}\delta)}(-c_\alpha).$$

Der Relevanz-t-Test ist dem zweiseitigen t-Test vorzuziehen, wenn der Stichprobenumfang N sehr groß ist. Bei Ablehnung der Nullhypothese des Relevanz-t-Testes kann man dann schließen, dass eine relevante Abweichung von μ_0 vorliegen muss, nämlich $|\mu - \mu_0| > \delta\sigma$.

Bei der Qualitätssicherung möchte man aber gar nicht einen relevanten Unterschied zeigen sondern nur „beweisen“, dass z.B. die These $\mu = \mu_0$ annähernd gilt. Das kann aber weder mit dem einfachen

zweiseitigen t-Test für $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ noch mit dem Relevanz-t-Test gezeigt werden, da das Annehmen der Nullhypothese kein signifikantes Ergebnis ist und unter Umständen nur durch zu wenige Daten erreicht wurde. Man kann die Forderung $\mu \approx \mu_0$ nur belegen, wenn diese Forderung als Alternativhypothese formuliert wird. Allerdings kann $H_0 : \mu \neq \mu_0$ gegen $H_1 : \mu = \mu_0$ nicht erfolgreich getestet werden. Denn, wenn die Gütefunktion (s. Abschnitt 21.2) für so einen Test für alle $\mu \neq \mu_0$ einen Wert kleiner gleich α ergibt, muss sie wegen der Stetigkeit auch für $\mu = \mu_0$ einen Wert kleiner gleich α ergeben. Dann ist aber die Wahrscheinlichkeit für ein signifikantes Ergebnis sehr klein, auch wenn $\mu = \mu_0$ gilt. Daher kann nur

$$H_0 : |\mu - \mu_0| \geq \delta\sigma \text{ gegen } H_1 : |\mu - \mu_0| < \delta\sigma$$

sinnvoll getestet werden. Dies wird mit dem **Äquivalenz-t-Test für $H_0 : |\mu - \mu_0| \geq \delta\sigma$ gegen $H_1 : |\mu - \mu_0| < \delta\sigma$** getestet:

$$\text{Lehne } H_0 \text{ ab, falls } \sqrt{N} \frac{|\mu_0 - \bar{x}|}{s(x)} < \tilde{c}_\alpha,$$

wobei \tilde{c}_α der Wert ist mit

$$\alpha = F_{t_{N-1}(\sqrt{N}\delta)}(\tilde{c}_\alpha) - F_{t_{N-1}(\sqrt{N}\delta)}(-\tilde{c}_\alpha).$$

22.5.1 Beispiel (Six Sigma Qualität)

Nach der Six Sigma Qualität (siehe Abschnitt 13.2.7) darf der wahre mittlere Wert μ nicht mehr als um 1.5σ von der Vorgabe μ_0 abweichen. Man kann auf $|\mu - \mu_0| \leq 1.5\sigma$ schließen, wenn der Äquivalenz-t-Test für $H_0 : |\mu - \mu_0| \geq \delta\sigma$ gegen $H_1 : |\mu - \mu_0| < \delta\sigma$ die Nullhypothese ablehnt. Zur Bestimmung des kritischen Wertes \tilde{c}_α des Äquivalent-t-Testes für $H_0 : |\mu - 1| \geq 1.5\sigma$ gegen $H_1 : |\mu - 1| < 1.5\sigma$ wird einfach die Funktion

$$g(c) = F_{t_{N-1}(\sqrt{N}1.5)}(c) - F_{t_{N-1}(\sqrt{N}1.5)}(-c)$$

grafisch dargestellt und von der Grafik der Wert c abgelesen mit $g(c) = \alpha$.

```
> N<-10
> c<-seq(1,3,0.01)
> plot(c,pt(c,df=N-1,ncp=sqrt(N)*1.5)-pt(-c,df=N-1,ncp=sqrt(N)*1.5),
+   type="l",xlab="c",ylab="g(c)")
> abline(0.05,0)
```

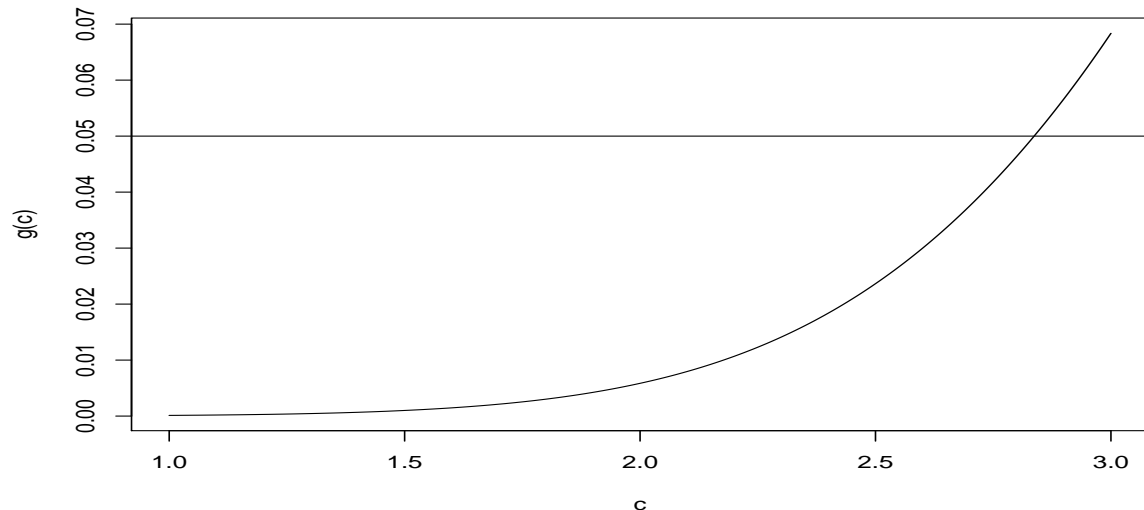
Von der Grafik wird für \tilde{c}_α ein Wert nahe bei 2.8 abgelesen. Durch Ausprobieren erhält man

```
> pt(2.84,df=N-1,ncp=sqrt(N)*1.5)-pt(-2.84,df=N-1,ncp=sqrt(N)*1.5)
[1] 0.05031328
> pt(2.83,df=N-1,ncp=sqrt(N)*1.5)-pt(-2.83,df=N-1,ncp=sqrt(N)*1.5)
[1] 0.04930968
```

Also ist $\tilde{c}_\alpha = 2.83$ für $\alpha = 0.05$ und $N = 10$.

22.5.2 Beispiel (Fortsetzung von Beispiel 21.1.1)

Da in der ersten Produktionslinie $T(x) = 2.1178$ und in der zweiten Produktionslinie $T(x) = 2.9976$

Abbildung 22.5.1: Die Funktion $g(c)$ zur Bestimmung von \tilde{c}_α bei $N = 10$ und $\delta = 1.5$

gilt, wird in der ersten Produktionslinie der kritische Wert des Äquivalenz-t-Testes für $H_0 : |\mu - 1| \geq 1.5\sigma$ gegen $H_1 : |\mu - 1| < 1.5\sigma$ unterschritten, also $H_0 : |\mu - 1| \geq 1.5\sigma$ abgelehnt. Also kann für die erste Produktionslinie $|\mu - 1| < 1.5\sigma$ geschlossen werden. Für die zweite Produktionslinie wird $H_0 : |\mu - 1| \geq 1.5\sigma$ nicht abgelehnt, so dass ein weitergehender Schluss als $\mu \neq 1$, der mit dem zweiseitigen t-Test gewonnen wurde, mit diesen Tests nicht möglich ist. Man könnte aber für die zweite Produktionslinie einen Relevanz-t-Test daraufhin durchführen, ob eine relevante Abweichung von der Vorgabe $\mu_0 = 1$ vorliegt.

Da aber in Beispiel 22.2.1 gezeigt wurde, dass $H_0 : \sigma \leq \sigma_0 = \frac{1}{6} \min\{USL - \mu_0, \mu_0 - LSL\} = 0.1/6$ in beiden Produktionslinien abgelehnt wird, ist die Six Sigma Qualität sowieso nicht in beiden Produktionslinien erfüllt. Man sieht also, dass die Begingung $|\mu - \mu_0| \leq 1.5\sigma$ nicht so schwer zu erfüllen ist, aber dass die einschneidende Bedingung die Bedingung an die Streuung ist.

Man betrachte nun noch die verschärfte Fragestellung $H_0 : |\mu - 1| \geq 0.1\sigma$ gegen $H_1 : |\mu - 1| < 0.1\sigma$. Zur Bestimmung des kritischen Wertes \tilde{c}_α des Äquivalenz-t-Testes für diese verschärfte Fragestellung wird wieder die Funktion

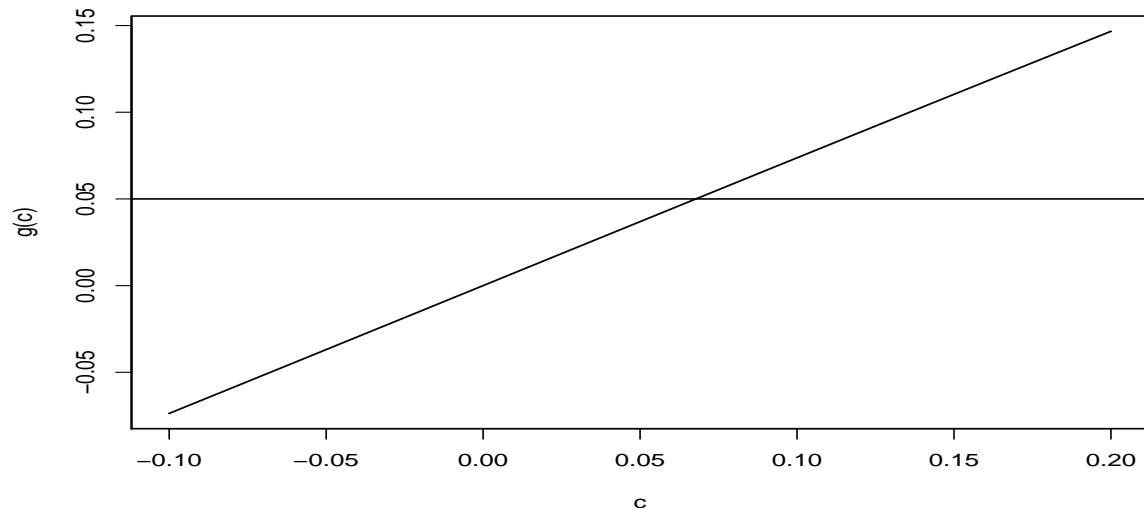
$$g(c) = F_{t_{N-1}(\sqrt{N}0.1)}(c) - F_{t_{N-1}(\sqrt{N}0.1)}(-c)$$

grafisch dargestellt und von der Grafik der Wert c abgelesen mit $g(c) = \alpha$.

```
> c<-seq(-0.1,0.2,0.01)
> plot(c,pt(c,df=N-1,ncp=sqrt(N)*0.1)-pt(-c,df=N-1,ncp=sqrt(N)*0.1),
+ type="l",xlab="c",ylab="g(c)")
> abline(0.05,0)
```

Von der Grafik wird für \tilde{c}_α ein Wert nahe bei 0.07 abgelesen. Durch Ausprobieren erhält man

```
> pt(0.068,df=N-1,ncp=sqrt(N)*0.1)-pt(-0.068,df=N-1,ncp=sqrt(N)*0.1)
```

Abbildung 22.5.2: Die Funktion $g(c)$ zur Bestimmung von \tilde{c}_α bei $N = 10$ und $\delta = 0.1$

```
[1] 0.05016034
> pt(0.067,df=N-1,ncp=sqrt(N)*0.1)-pt(-0.067,df=N-1,ncp=sqrt(N)*0.1)
[1] 0.04942379
```

Also kann $\tilde{c}_\alpha = 0.067$ gesetzt werden. Da in der ersten Produktionslinie $T(x) = 2.1178$ und in der zweiten Produktionslinie $T(x) = 2.9976$ gilt, wird in beiden Fällen die Nullhypothese $H_0 : |\mu - 1| \geq 0.1\sigma$ nicht abgelehnt. Also kann nicht $|\mu - 1| < 0.1\sigma$ geschlossen werden. Das aber kann daran liegen, dass mit $N = 10$ der Stichprobenumfang extrem klein ist, siehe Abschnitt 21.3.

Wie für eine Stichprobe gibt es auch für **zwei Stichproben** Relevanz- und Äquivalenz-Tests:

Relevanz-t-Test für $H_0 : |\mu_1 - \mu_2| \leq \delta\sigma$ gegen $H_1 : |\mu_1 - \mu_2| > \delta\sigma$:

$$\text{Lehne } H_0 \text{ ab, falls } K \frac{|\bar{x} - \bar{y}|}{s_{12}(x, y)} > c_\alpha,$$

wobei c_α der Wert ist mit

$$\alpha = 1 - F_{t_{N_1+N_2-2}(K\delta)}(c_\alpha) + F_{t_{N_1+N_2-2}(K\delta)}(-c_\alpha).$$

Äquivalenz-t-Test für $H_0 : |\mu_1 - \mu_2| \geq \delta\sigma$ gegen $H_1 : |\mu_1 - \mu_2| < \delta\sigma$:

$$\text{Lehne } H_0 \text{ ab, falls } K \frac{|\bar{x} - \bar{y}|}{s_{12}(x, y)} < \tilde{c}_\alpha,$$

wobei \tilde{c}_α der Wert ist mit

$$\alpha = F_{t_{N_1+N_2-2}(K\delta)}(\tilde{c}_\alpha) - F_{t_{N_1+N_2-2}(K\delta)}(-\tilde{c}_\alpha).$$

Dabei werden die gleichen Bezeichnungen wie für den t-Test für zwei Stichproben in Abschnitt 22.3 benutzt. Der Äquivalent-Test zeigt dabei, dass zwei Stichproben bezüglich des Erwartungswertes äquivalent sind.

22.6 Test auf linearen Zusammenhang zwischen gepaarten Stichproben

Sind die Daten der Form $(x_1, y_1), \dots, (x_N, y_N)$ Realisierungen von iid Zufallsvariablen $(X_1, Y_1), \dots, (X_N, Y_N)$, so drückt die Korrelation $Korr(X_n, Y_n)$ den linearen Zusammenhang zwischen X_n und Y_n aus. Besitzen überdies $(X_1, Y_1), \dots, (X_N, Y_N)$ eine zweidimensionale Normalverteilung mit Parametern $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$, so gilt $Korr(X_n, Y_n) = \rho$ und $Korr(X_n, Y_n) = 0$ bedeutet, dass X_n und Y_n stochastisch unabhängig sind. Tests von Hypothesen über ρ basieren auf dem **Pearsonschen Korrelationskoeffizienten** $r(x, y)$:

$$r(x, y) = \frac{\sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^N (x_n - \bar{x})^2 \sum_{n=1}^N (y_n - \bar{y})^2}} = \frac{s(x, y)}{s(x) s(y)}.$$

Zweiseitiger Korrelations-Test für $H_0 : \rho = 0$ gegen $H_1 : \rho \neq 0$:

$$\text{Lehne } H_0 \text{ ab, falls } \sqrt{N-2} \frac{|r(x, y)|}{\sqrt{1-r(x, y)^2}} > t_{N-2; 1-\alpha/2}.$$

Aufruf in R: `cor.test(x,y,alternative = "two.sided", method = "pearson")`.

Einseitiger Korrelations-Test für $H_0 : \rho \leq 0$ gegen $H_1 : \rho > 0$:

$$\text{Lehne } H_0 \text{ ab, falls } \sqrt{N-2} \frac{r(x, y)}{\sqrt{1-r(x, y)^2}} > t_{N-2; 1-\alpha}.$$

Aufruf in R: `cor.test(x,y,alternative = "greater", method = "pearson")`.

Einseitiger Korrelations-Test für $H_0 : \rho \geq 0$ gegen $H_1 : \rho < 0$:

$$\text{Lehne } H_0 \text{ ab, falls } \sqrt{N-2} \frac{r(x, y)}{\sqrt{1-r(x, y)^2}} < t_{N-2; \alpha}.$$

Aufruf in R: `cor.test(x,y,alternative = "less", method = "pearson")`.

22.6.1 Beispiel (Fortsetzung von Beispiel 1.0.4)

Bei der Hitzeentwicklung beim Zementansetzen besteht die Frage, ob es einen linearen Zusammenhang zwischen der Höhe eines der vier untersuchten Anteile und der Hitzeentwicklung gibt. Die Daten des Datensatzes `SETTING.DAT` können dabei einfach mit der R-Funktion `read.table` eingelesen werden.

```
> SETTING<-read.table("SETTING.DAT")
> colnames(SETTING)<-c("a", "b", "c", "d", "Hitze")
> cor.test(SETTING$a, SETTING$Hitze, alternative="two.sided", method="pearson")
```

Pearson's product-moment correlation

```
data:  SETTING$a and SETTING$Hitze
t = 3.55, df = 11, p-value = 0.004552
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3008647 0.9137954
sample estimates:
      cor
0.7307175
```

Analog erhält man als P-Werte: 0.0006648 für „b“, 0.05976 für „c“ und 0.0005762 für „d“.

Da vier Tests durchgeführt werden, kann nach (20.1) zur Festlegung des Niveaus α die Nullhypothese $H_0 : \rho = 0$ nur abgelehnt werden, wenn der P-Wert kleiner als $\alpha = 0.05/4 = 0.0125$ ist. Dies ist bei den Zementanteilen „a“, „b“ und „d“ der Fall, d.h. bei diesen drei Zementanteilen kann man davon ausgehen, dass die Höhe des Anteils und die Hitzeentwicklung nicht stochastisch unabhängig sind. Bei dem Zementanteil „c“ ist keine Aussage möglich, da die Nullhypothese nicht zum Niveau $\alpha = 0.0125$ abgelehnt wird.

22.7 Tests auf Normalverteilung

Außer bei sehr kleinen Stichprobenumfängen wie $N \leq 10$ muss in der Regel überprüft werden, ob die Daten gemäß der Normalverteilung verteilt sind, bevor Tests angewendet werden können, die die Normalverteilung voraussetzen. Diese Überprüfung kann nur wegfallen, wenn in vorangegangenen Studien die Normalverteilung belegt wurde oder inhaltliche Gründe für eine Normalverteilung sprechen. Da aber eine Normalverteilung nie richtig belegt werden kann, wird in den meisten Fällen ein Test auf Normalverteilung nötig sein.

Dazu gibt es mehrere Möglichkeiten wie den Shapiro-Wilk-Test gegeben durch `shapiro.test`. Um einen grafischen Eindruck zu bekommen, ob eine Normalverteilung vorliegen kann, kann man auch die Quantil-Quantil-Darstellung gegeben durch `qqnorm` benutzen. Liegen die Daten bei der Quantil-Quantil-Darstellung mehr oder weniger auf einer Geraden, kann man von einer Normalverteilung ausgehen. Spricht nichts gegen die Normalverteilung, so können die Tests, die eine Normalverteilung voraussetzen, benutzt werden.

22.7.1 Beispiel (Fortsetzung von Beispiel 21.1.1)

Obwohl die Stichprobenumfänge der Kugellagerdurchmesser nicht größer als 10 sind, wird für beide Produktionslinien der Shapiro-Wilk-Test durchgeführt:

```
> shapiro.test(k1)
      Shapiro-Wilk normality test
data:  k1
W = 0.9833, p-value = 0.9804

> shapiro.test(k2)
      Shapiro-Wilk normality test
```

```
data: k2
W = 0.9082, p-value = 0.2685
```

In beiden Fällen kann wie erwartet bei den kleinen Stichprobenumfängen die Normalverteilung nicht abgelehnt werden. Allerdings ist der P-Wert bei der zweiten Produktionslinie deutlich kleiner. Dass dort eventuell die Normalverteilung doch nicht vorliegen könnte, zeigt auch die Quantil-Quantil-Darstellung:

```
> qqnorm(k1,pch=16)
> qqnorm(k2,pch=16)
```

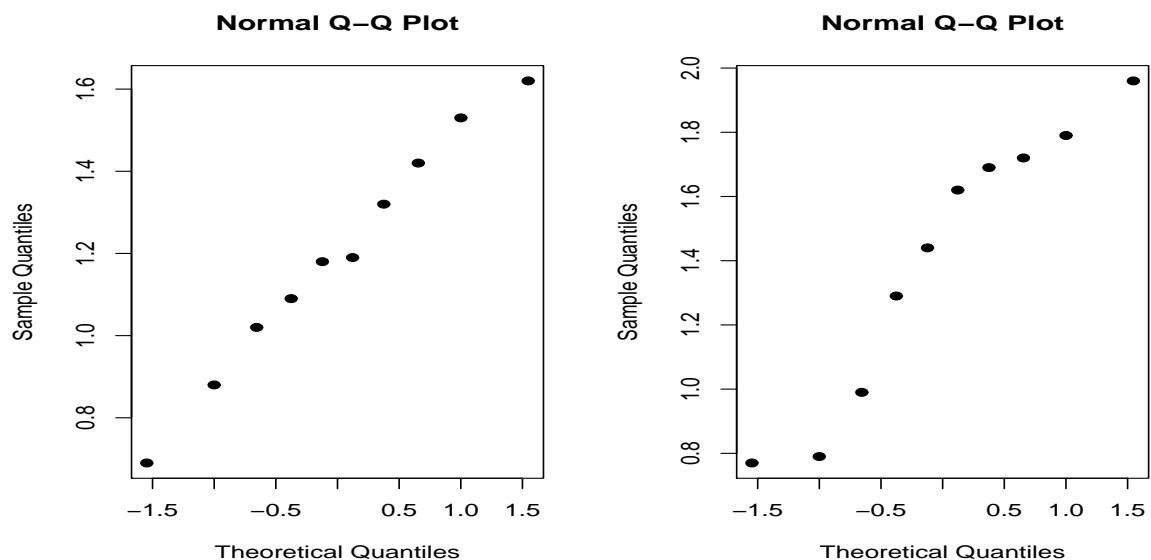


Abbildung 22.7.1: Quantil-Quantil-Darstellung: Links für die erste Produktionslinie, rechts für die zweite Produktionslinie

22.7.2 Bemerkung (Vortests bei mehreren Tests am gleichen Datensatz)

Vortests wie Tests auf Normalverteilung werden bei der α -Adjustierung (20.1) nicht mitgezählt, da diese Tests nur dazu dienen bestimmte Voraussetzungen für die eigentlichen Tests zu überprüfen. Also wird beim Vorhandensein von Vortests

$$\alpha = \frac{0.05}{\text{Anzahl der Tests am Datensatz} - \text{Anzahl der Vortests}}$$

gesetzt.

Kapitel 23

Tests bei nichtnormalverteilten Daten

23.1 Einstichproben-und Zweistichproben-Tests

Mit dem Wilcoxon-Rangsummen-Test kann getestet werden, ob eine Stichprobe im Mittel den Wert μ_0 annimmt. Außerdem kann damit getestet, ob zwei Stichproben im Mittel übereinstimmen. Dabei gehen nur die Ränge der Beobachtungen ein, so dass dieser Test auch bei ordinalen Daten angewendet werden kann. Die Funktion `wilcox.test` liefert den Wilcoxon-Rangsummen-Test.

23.1.1 Beispiel (Fortsetzung von 21.1.1)

Testet man in beiden Produktionslinien $H_0 : \mu_0 = 1$ gegen $H_0 : \mu_0 \neq 1$ mit dem Wilcoxon-Rangsummen-Test, so erhält man:

```
> wilcox.test(k1, alternative="two.sided", mu=1)
```

```
Wilcoxon signed rank test
```

```
data: k1
```

```
V = 46, p-value = 0.06445
```

```
alternative hypothesis: true mu is not equal to 1
```

```
> wilcox.test(k2, alternative="two.sided", mu=1)
```

```
Wilcoxon signed rank test
```

```
data: k2
```

```
V = 49, p-value = 0.02734
```

```
alternative hypothesis: true mu is not equal to 1
```

Man erhält das gleiche Ergebnis wie mit dem t-Test. Allerdings ist bei der ersten Produktionslinie der P-Wert mit 0.06445 etwas größer als beim t-Test, wo er 0.06327 beträgt. Noch größer ist der Unterschied bei der zweiten Produktionslinie: Dort ist der P-Wert des Wilcoxon-Rangsummen-Testes 0.02734 und des t-Testes 0.01502. Das zeigt, dass man den t-Test benutzen sollte, wenn nichts gegen die Normalverteilung spricht.

Das gleiche gilt eigentlich auch für die Zweistichproben-Tests für den Vergleich von zwei Erwartungswerten, d.h. für das Testen von $H_0 : \mu_1 = \mu_2$ gegen $H_0 : \mu_1 \neq \mu_2$. Allerdings tritt das Phänomen in diesem Beispiel nicht auf. Bei dem einfachen Zweistichproben-t-Test für gleiche Varianzen wurde ein P-Wert von 0.2112 und beim Zweistichproben-Welch-t-Test ein P-Wert von 0.2134 erzielt.

```
> wilcox.test(k1,k2,alternative="two.sided")
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: k1 and k2
```

```
W = 32.5, p-value = 0.1986
```

```
alternative hypothesis: true mu is not equal to 0
```

```
Warning message:
```

```
cannot compute exact p-value with ties in: wilcox.test.default(k1, k2,  
alternative = "two.sided")
```

23.2 Tests auf Zusammenhang

Es gibt auch einen Test auf einen Zusammenhang zwischen zwei nichtnormalverteilten Zufallsvariablen X_n und Y_n . Dabei werden im Pearsonschen Korrelationskoeffizienten $r(x, y)$ die Beobachtungen durch deren Ränge ersetzt. Damit erhält man den **Spearman'schen Rangkorrelationskoeffizienten**. Ein Tests für $H_0 : \rho = 0$ gegen $H_1 : \rho \neq 0$ basierend auf dem Spearman'schen Rangkorrelationskoeffizienten ist durch die Funktion `cor.test` mit dem Argument `method=spearman` gegeben. Man beachte, dass $\rho = \text{korr}(X_n, Y_N) = 0$ für nichtnormalverteilte Zufallsvariablen X_n und Y_n nicht stochastische Unabhängigkeit bedeuten muss. Mann spricht daher nur von Unkorreliertheit. Siehe Bemerkung 17.2.2.

23.2.1 Beispiel (Fortsetzung von Beispiel 22.6.1)

```
> cor.test(SETTING$a,SETTING$Hitze,alternative="two.sided",method="spearman")
```

```
Spearman's rank correlation rho
```

```
data: SETTING$a and SETTING$Hitze
```

```
S = 76.0036, p-value = 0.001275
```

```
alternative hypothesis: true rho is not equal to 0
```

```
sample estimates:
```

```
rho
```

```
0.791199
```

```
Warning message:
```

```
Cannot compute exact p-values with ties in: cor.test.default(SETTING$a,  
SETTING$Hitze, alternative = "two.sided",
```

Analog erhält man als P-Werte: 0.00403 für „b“, 0.124 für „c“ und 0.003023 für „d“. Bis auf den

Test für Anteil „a“ sind die P-Werte wieder alle größer als bei den Tests, die Normalverteilung voraussetzen.

Nehmen die zweidimensionalen Beobachtungen $(x_1, y_1), \dots, (x_N, y_N)$ nur wenige verschiedene Werte an, dann werden sie auch oft durch eine Tabelle mit deren Häufigkeiten gegeben. In diesem Fall kann auf Zusammenhang von X_n und Y_n mit dem χ^2 -Test auf Unabhängigkeit getestet werden (in R `chisq.test`). Dabei betrifft die Nullhypothese nicht nur Unkorreliertheit ausgedrückt durch $\rho = 0$ sondern richtige stochastische Unabhängigkeit. D.h. es wird

$H_0 : X_n \text{ und } Y_n \text{ sind stochastisch unabhängig}$

gegen

$H_1 : X_n \text{ und } Y_n \text{ sind nicht stochastisch unabhängig}$

getestet.

23.2.2 Beispiel (Fortsetzung von Beispiel 1.0.3)

Bei Beispiel 1.0.3 stellt sich die Frage, ob es einen Zusammenhang zwischen Rostschutzmittel und Wirksamkeit gibt.

```
> rost<-matrix(c(65 , 103 , 106 , 74 , 85 , 47),ncol=3,byrow=T)
> rost
      [,1] [,2] [,3]
[1,]   65  103  106
[2,]   74   85   47
> chisq.test(rost)
```

Pearson's Chi-squared test

```
data:  rost
X-squared = 15.7403, df = 2, p-value = 0.000382
```

Da der P-Wert kleiner als $\alpha = 0.05$ ist, kann geschlossen werden, dass es einen Zusammenhang zwischen Rostschutzmittel und Wirksamkeit gibt.

Kapitel 24

Intervallschätzungen/Konfidenzbereiche *

24.1 Allgemeine Eigenschaften

Während eine Punktschätzung einen einzigen Wert für einen (unbekannten) Parameter ϑ liefert, wird durch eine Intervallschätzung oder einen Konfidenzbereich¹ eine Menge von Werten angegeben. Diese Vorgehensweise beruht auf der Idee, daß das Ergebnis einer Punktschätzung ohnehin nicht den wahren Wert des Parameters liefert, sondern aufgrund des vorliegenden Zufalls immer vom tatsächlichen Wert abweicht. Eine Intervallschätzung liefert einen Bereich, in dem der unbekannte Parameter mit einer vorgegebenen Wahrscheinlichkeit liegt. Dazu sei wieder X_1, \dots, X_N eine unabhängige und identisch verteilte Stichprobe von X_0 und $X = (X_1, \dots, X_N)$ der Stichprobenvektor mit Realisierung $x = (x_1, \dots, x_N)$.

24.1.1 Definition (Intervallschätzung, Konfidenzintervall)

Seien $\alpha \in (0, 1)$ eine feste, vorgegebene Wahrscheinlichkeit und ϑ ein unbekannter Parameter.

Eine (zufällige) Menge $[\hat{\vartheta}_u(X), \hat{\vartheta}_o(X)]$ heißt **Intervallschätzung** oder **Konfidenzintervall**² zum Niveau $1 - \alpha$ für den Parameter ϑ , falls gilt:

$$P_{\vartheta}(\vartheta \in [\hat{\vartheta}_u(X), \hat{\vartheta}_o(X)]) \geq 1 - \alpha.$$

Hierbei sind $\hat{\vartheta}_u(X) = \hat{\vartheta}_u(X_1, \dots, X_N)$ und $\hat{\vartheta}_o(X) = \hat{\vartheta}_o(X_1, \dots, X_N)$ Statistiken, die von der Stichprobe X_1, \dots, X_N abhängen und für jede Realisation x_1, \dots, x_N die Ungleichung $\hat{\vartheta}_u(x_1, \dots, x_N) \leq \hat{\vartheta}_o(x_1, \dots, x_N)$ erfüllen. Der Parameter ϑ an der Wahrscheinlichkeit P deutet an, dass die Wahrscheinlichkeitsverteilung mit diesem Parameter benutzt wird. Wenn klar ist, welche Wahrscheinlichkeitsverteilung gemeint ist, wird oft auch dieser Parameter weggelassen.

Es ist zu beachten, daß die Wahrscheinlichkeit für das Konfidenzintervall größer gleich $1 - \alpha$ sein soll. Umgekehrt bedeutet dies, daß die Wahrscheinlichkeit für das Ereignis, daß der Parameter ϑ außerhalb des Konfidenzbereiches liegt, kleiner als α ist. Deshalb wird α üblicherweise klein gewählt. Typische Werte sind $\alpha = 0.1, 0.05, 0.01$. Oft wird statt $[\hat{\vartheta}_u(x), \hat{\vartheta}_o(x)]$ auch kurz $[\hat{\vartheta}_u, \hat{\vartheta}_o]$ geschrieben.

¹auch: Vertrauensbereich.

²Alternativ wird ein Konfidenzintervall zum Niveau $1 - \alpha$ auch kurz als $(1 - \alpha)$ -Konfidenzintervall bezeichnet.

Wie bei der Anwendung der statistischen Tests hängt die Wahl eines Konfidenzbereiches von den vorausgesetzten Verteilungsannahmen ab. Es werden daher für verschiedene Modelle Konfidenzbereiche angegeben, wobei die vorgestellten Konfidenzintervalle die gebräuchlichsten sind. Diese Wahl ist jedoch keineswegs eindeutig, und es lassen sich auch andere Bereiche angeben. Grundlegend zur Bestimmung von Konfidenzbereichen ist die Kenntnis der Wahrscheinlichkeitsverteilung der verwendeten Statistiken.

Dabei kann folgender Zusammenhang zu statistischen Tests ausgenutzt werden:

1. Hat man für alle möglichen Parameter ϑ Tests zum Niveau α für $H_0 : \vartheta = \vartheta_0$ gegen $H_0 : \vartheta \neq \vartheta_0$ der Form:

Lehne H_0 ab, falls $T_{\theta_0}(x) > c(\theta_0)$,

so ist

$$\{\theta_0; T_{\theta_0}(x) \leq c(\theta_0)\}$$

ein Konfidenzbereich zum Niveau $1 - \alpha$.

2. Hat man umgekehrt Konfidenzintervalle $[\hat{\vartheta}_u, \hat{\vartheta}_o]$ für θ zum Niveau $1 - \alpha$, so ist ein Test für $H_0 : \vartheta = \vartheta_0$ gegen $H_0 : \vartheta \neq \vartheta_0$ zum Niveau α gegeben durch:

Lehne H_0 ab, falls $\vartheta_0 \notin [\hat{\vartheta}_u, \hat{\vartheta}_o]$.

Im folgenden werden basierend auf einer Stichprobe X_1, \dots, X_N eines $N(\mu, \sigma^2)$ -verteilten Merkmals $(1-\alpha)$ -Konfidenzintervalle für die Parameter μ und σ^2 angegeben. Als Punktschätzer werden hierbei das arithmetische Mittel $\bar{X} = \frac{1}{N} \sum_{n=1}^N X_n$ und die empirische Varianz $s(x)^2 = \frac{1}{N-1} \sum_{n=1}^N (X_n - \bar{X})^2$ verwendet.

24.2 Konfidenzintervalle für den Erwartungswert einer Normalverteilung

In Analogie zur vorhergehenden Situation lassen sich Konfidenzintervalle angeben. Die Quantile der Standardnormalverteilung werden jedoch durch die entsprechenden Quantile der t_{N-1} -Verteilung ersetzt, die mit $t_{N-1;\beta}$ bezeichnet werden. Die Aussagen beruhen wieder auf der Tatsache, dass die Statistik

$$T = \sqrt{N} \frac{\bar{X} - \mu}{s(x)}$$

eine t -Verteilung mit $N - 1$ Freiheitsgraden besitzt. Eine Überprüfung, dass es sich tatsächlich um Konfidenzintervalle handelt, benutzt diese Eigenschaft.

- **Zweiseitiges** Konfidenzintervall: $[\hat{\mu}_u, \hat{\mu}_o] = \left[\bar{X} - t_{N-1;1-\alpha/2} \frac{s(x)}{\sqrt{N}}, \bar{X} + t_{N-1;1-\alpha/2} \frac{s(x)}{\sqrt{N}} \right]$
- **Einseitiges**, unteres Konfidenzintervall: $(-\infty, \hat{\mu}_o] = \left(-\infty, \bar{X} + t_{N-1;1-\alpha} \frac{s(x)}{\sqrt{N}} \right]$
- **Einseitiges**, oberes Konfidenzintervall: $[\hat{\mu}_u, \infty) = \left[\bar{X} - t_{N-1;1-\alpha} \frac{s(x)}{\sqrt{N}}, \infty \right)$.

24.3 Konfidenzintervalle für die Varianz einer Normalverteilung

Für den Parameter σ^2 lassen sich ebenfalls ein- bzw. zweiseitige Konfidenzintervalle bestimmen. In völliger Analogie zur obigen Vorgehensweise erhält man folgende Intervallschätzungen. Als Quantil werden wieder die Quantile der χ^2 -Verteilung mit $N - 1$ Freiheitsgraden verwendet.

- **Zweiseitiges** Konfidenzintervall: $[\hat{\sigma}_u^2, \hat{\sigma}_o^2] = \left[\frac{N-1}{\chi_{N-1;1-\alpha/2}^2} s(x)^2, \frac{N-1}{\chi_{N-1;\alpha/2}^2} s(x)^2 \right]$
- **Einseitiges**, unteres Konfidenzintervall $[0, \hat{\sigma}_o^2] = \left[0, \frac{N-1}{\chi_{N-1;\alpha}^2} s(x)^2 \right]$
- **Einseitiges**, oberes Konfidenzintervall $[\hat{\sigma}_u^2, \infty) = \left[\frac{N-1}{\chi_{N-1;1-\alpha}^2} s(x)^2, \infty \right)$.

Zum Beispiel erhält man das zweiseitige Konfidenzintervall wie folgt: Wegen $\frac{(N-1)s(x)^2}{\sigma^2} \sim \chi_{N-1}^2$ haben wir

$$\begin{aligned} 1 - \alpha &= P_\sigma \left(\chi_{N-1;\alpha/2}^2 \leq \frac{(N-1)s(x)^2}{\sigma^2} \leq \chi_{N-1;1-\alpha/2}^2 \right) \\ &= P_\sigma \left(\frac{1}{\chi_{N-1;1-\alpha/2}^2} \leq \frac{\sigma^2}{(N-1)s(x)^2} \leq \frac{1}{\chi_{N-1;\alpha/2}^2} \right) \\ &= P_\sigma \left(\sigma^2 \in \left[\frac{N-1}{\chi_{N-1;1-\alpha/2}^2} s(x)^2, \frac{N-1}{\chi_{N-1;\alpha/2}^2} s(x)^2 \right] \right). \end{aligned}$$

24.4 Konfidenzintervalle für die Standardabweichung bei Normalverteilung

Für σ gewinnt man geeignete Konfidenzintervalle durch Ziehen der Quadratwurzel aus den entsprechenden Intervallgrenzen $\hat{\sigma}_u^2$ bzw. $\hat{\sigma}_o^2$.

24.5 Konfidenzintervall für die Differenz $d = \mu_1 - \mu_2$ bei unbekannter (gleicher) Varianz

Seien X_1, \dots, X_{N_1} eine iid Stichprobe eines $N(\mu_1, \sigma^2)$ -verteilten Merkmals und Y_1, \dots, Y_{N_2} eine iid Stichprobe eines $N(\mu_2, \sigma^2)$ -verteilten Merkmals, wobei σ^2 unbekannt und $N_1, N_2 \geq 2$ sind. Ferner seien alle $N_1 + N_2$ Zufallsvariablen stochastisch unabhängig. Setze wieder

$$K = \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

Ein **zweiseitiges** Konfidenzintervall $[\hat{d}_u, \hat{d}_o]$ für die Differenz der Erwartungswerte $d = \mu_1 - \mu_2$ ist gegeben durch:

$$\left[\hat{d} - t_{N_1+N_2-2;1-\alpha/2} s_{12}(x, y) \cdot \frac{1}{K}, \hat{d} + t_{N_1+N_2-2;1-\alpha/2} s_{12}(x, y) \cdot \frac{1}{K} \right],$$

wobei $\hat{d} = \bar{X} - \bar{Y}$ eine Punktschätzung für die Differenz der Erwartungswerte d ,

$$s_{12}(x, y)^2 := \frac{1}{N_1 + N_2 - 2} \left(\sum_{n=1}^{N_1} (x_n - \bar{x})^2 + \sum_{n=1}^{N_2} (y_n - \bar{y})^2 \right) = \frac{N_1 - 1}{N_1 + N_2 - 2} s(x)^2 + \frac{N_2 - 1}{N_1 + N_2 - 2} s(y)^2$$

wieder die kombinierte Varianzschätzung und $t_{N_1+N_2-2;\beta}$ das β -Quantil der $t_{N_1+N_2-2}$ -Verteilung sind. Die Aussage beruht wieder auf der Eigenschaft:

$$\sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \frac{\hat{d} - (\mu_1 - \mu_2)}{s_{12}(x, y)} \sim t_{N_1+N_2-2}.$$

Einseitige Konfidenzintervalle lassen sich analog konstruieren. Zu beachten ist, daß das vorgestellte Konfidenzintervall auf der Annahme beruht, daß die Varianz der Zufallsvariablen X_n und Y_N gleich sind.

24.5.1 Beispiel (Fortsetzung von Beispiel 21.1.1)

Die R-Funktion `t.test` liefert auch die Konfidenzintervalle. Für die erste Produktionslinie erhalten wir:

```
> t.test(k1, alternative="two.sided", var.equal=TRUE)$conf.int
[1] 0.9867741 1.4012259
attr(,"conf.level")
[1] 0.95
```

Damit ist das Konfidenzintervall zum Niveau 0.95 für den Erwartungswert μ der ersten Produktionslinie $[0.9867741, 1.4012259]$. Analog erhält man $[1.099606, 1.712394]$ für die zweite Produktionslinie. Da $\mu_0 = 1$ im Konfidenzintervall für die erste Produktionslinie enthalten ist, wurde $H_0 : \mu = 1$ nicht abgelehnt. Dagegen ist $\mu_0 = 1$ nicht im Konfidenzintervall für die zweite Produktionslinie enthalten, weshalb $H_0 : \mu = 1$ abgelehnt wurde.

Für die Differenz der beiden Erwartungswerte $\mu_1 - \mu_2$ wird $[-0.5555277, 0.1315277]$ als Konfidenzintervall zum Niveau 0.95 erhalten. Da dieses Intervall die 0 enthält, wurde die Nullhypothese $H_0 : \mu_1 = \mu_2$ nicht abgelehnt.

```
> t.test(k1, k2, alternative="two.sided", var.equal=TRUE)$conf.int
[1] -0.5555277 0.1315277
attr(,"conf.level")
[1] 0.95
```

Kapitel 25

Abkürzungen

x	$= (x_1, \dots, x_N)^\top$ Datenvektor
y	$y = (y_1, \dots, y_N)^\top$ Datenvektor
\bar{x}	arithmetisches Mittel von x
$s = s_x = s(x)$	Standardabweichung von x
$s(x)^2$	empirische Varianz von x
$s_{xy} = s(x, y)$	empirische Kovarianz von x und y
$r_{xy} = r(x, y)$	empirische Korrelation von x und y
Ω	Ergebnismenge
ω	Ergebnis
A, B, C etc.	Mengen, interessierende Ereignisse
\mathfrak{A}	Menge aller Ereignisse
P	Wahrscheinlichkeit (Probability)
X, Y, Z, X_1, \dots, X_n etc.	Zufallsvariable
x, y, z, x_1, \dots, x_n etc.	Realisierungen der Zufallsvariablen
$P^X(B)$	$= P(X \in B) = P(\{\omega \in \Omega \mid X(\omega) \in B\})$
f	Dichtefunktion
$F(x)$	$= P^X((-\infty, x]) = P(X \leq x)$ Verteilungsfunktion an der Stelle x
Φ	Verteilungsfunktion der Standard-Normalverteilung
LSL	lower specification limit
USL	upper specification limit
$P(A B)$	bedingte Wahrscheinlichkeit von A gegeben B
$E(X)$	Erwartungswert der Zufallsvariablen X
$\text{var}(X)$	Varianz der Zufallsvariablen X
$\text{kov}(X, Y)$	Kovarianz der Zufallsvariablen X und Y
$\text{kor}(X, Y)$	Korrelation der Zufallsvariablen X und Y
iid	stochastisch unabhängig und identisch verteilt
ϑ	unbekannter Parameter einer Verteilung

$\text{Bin}(n, p)$	Binomialverteilung
$\text{Bin}(1, p)$	Bernoulliverteilung
$\text{Poi}(\lambda)$	Poissonverteilung
$\text{Exp}(\lambda)$	Exponentialverteilung
$\text{N}(\mu, \sigma^2)$	Normalverteilung
t_N	zentrale t -Verteilung mit N Freiheitsgraden
$t_N(\delta)$	nichtzentrale t -Verteilung mit Nichtzentralitätsparameter δ
$F_{t_N(\delta)}$	Verteilungsfunktion der nichtzentrale t -Verteilung
$t_{N;\alpha}$	α -Quantil der t -Verteilung
χ_N^2	χ^2 -Verteilung mit N Freiheitsgraden
$\chi_{N;\alpha}^2$	α -Quantil der χ^2 -Verteilung
$F_{N,M}$	F -Verteilung mit Freiheitsgraden N und M
$F_{N,M;\alpha}$	α -Quantil der F -Verteilung

Literaturverzeichnis

- [Allen 2006] Allen, T.T. (2006). *Introduction to Engineering Statistics and Six Sigma. Statistical Quality Control and Design of Experiments and Systems*. Springer, London.
- [Bortz 1999] Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Springer, Berlin.
- [Bosch 1999] Bosch, K. (1999). *Brückenkurs Mathematik*. Oldenbourg, München.
- [1] Box, G.E.P. and Cox, D.R. (1964). An analysis of transforms (with discussion). *Journal of the Royal Statistical Society, Series A*, **143**, 383-430.
- [Box et al. 1978] Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for experiments*. Wiley, New York, Table 7.1.
- [Dalgard 2002] Dalgaard, P. (2002). *Introductory Statistics with R*. Springer, Berlin.
- [DIALEKT 2002] DIALEKT-Projekt (2002). *Statistik interaktiv. Deskriptive Statistik*. Springer, Berlin.
- [Fahrmeier et al. 1997] Fahrmeier, L., Künstler, R., Pigeot, I., und Tutz, G. (1997). *Statistik*. Springer, Berlin.
- [Gupta52] Gupta, A.K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika* **39**, 260-273.
- [Hand et al. 1994] Hand et al. (1994). *A Handbook of Small Data Sets*, Chapman & Hall, London, S. 131.
- [Hartung et al. 1998] Hartung, J., Elpelt, B., und Klösener, H.P. (1998). *Statistik*. Oldenbourg, München.
- [Krengel 1998] Krengel, U. (1998). *Einführung in die Wahrscheinlichkeitsrechnung und Statistik*. Vieweg, Braunschweig.
- [Moeschlin 2003] Moeschlin, O. (2003). *Experimental Stochastics*. Springer, Berlin.
- [R 2010] R Development Core Team 2010, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>
- [Rinne 1997] Rinne, H. (1997). *Taschenbuch der Statistik*. Harry Deutsch, Thun.
- [Romano 1977] Romano, A. (1977). *Applied Statistics for Science and Industry*. Allyn and Bacon, Boston.
- [Schlittgen 2005] Schlittgen, R. (2005). *Das Statistiklabor. Einführung und Benutzerhandbuch*. Springer, Berlin.
- [Verzani 2004] Verzani, J. (2004). *Using R for Introductory Statistics*. Chapman & Hall /CRC, London.
- [Woods et al. 1932] Woods, H., Steinour, H.H. und Starke, H.R. (1932). Effects of composition of Portland Cement on heat evolved during hardening. *Industrial and Engineering Chemistry* **24**, 1207-1214.