TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Inst. f. Statistik u.Wahrscheinlichkeitstheorie

Statistik

und

Wahrscheinlichkeitstheorie

Unterlagen

zur

Übung

LVA–Nr.: 107.369 [2h]

WS 2010|11

W. Gurker

A–1040 Wien
Wiedner Hauptstrasse 8-10|107

Ass.Prof. Dipl.-Ing. Dr.techn. Werner GURKER
Inst. f. Statistik u. Wahrscheinlichkeitstheorie
Technische Universität Wien
Wiedner Hauptstr. 8-10 | 107
A-1040 Wien

Tel.: 58801 - 10724
E-Mail: W.Gurker@tuwien.ac.at
Spr.: Di u. Do von 11-12

Vorwort

Die folgende Aufgabensammlung bildet die Grundlage für die Übung *Statistik und Wahrscheinlichkeitstheorie* [107.369] im WS 2010|11. Sie orientiert sich in Aufbau und Inhalt an der gleichnamigen Vorlesung [107.254]. Der überwiegende Teil der Aufgaben illustriert die in der Vorlesung präsentierten Konzepte. Gelegentlich gibt es aber auch Aufgaben, die sich mit Konzepten beschäftigen, die in der Vorlesung nicht oder nur am Rande behandelt werden. Diese ergänzenden Aufgaben werden nach Maßgabe der verfügbaren Zeit behandelt.

Für die statistische/graphische Aufbereitung und Auswertung von Datensätzen und für sonstige Berechnungen wird ein Statistikpaket benötigt. In dieser Übung wird das unter der GNU–Lizenz frei verfügbare R verwendet (`http://www.r-project.org`). Erfahrungsgemäß bereitet dieses – im universitären Bereich weit verbreitete – Paket am Anfang einige Schwierigkeiten. Neben einer wachsenden Zahl von Lehrbüchern*) findet der/die Neueinsteiger/in im Internet zahlreiche Hilfestellungen. Überdies sind denjenigen Aufgaben, die mit Hilfe von R zu bearbeiten sind, entsprechende Skripts oder Hinweise beigefügt. Dazu empfiehlt sich die Installation eines auf die Verwendung mit R abgestimmten Editors (z.B. Tinn–R). (*Bem*: Die Datensätze und Skripts werden auf TISS zur Verfügung gestellt.)

In *Anhängen* (zu den Kapiteln und am Schluß) finden sich Ergänzungen, ein paar Dinge aus der Mathematik und übersichtsartige Darstellungen von diskreten und stetigen Verteilungen, von Konfidenzintervallen und Tests, sowie Wahrscheinlichkeitsnetze und oft benötigte Tabellen.

Wien, September 2010 W. G.

*) Drei Empfehlungen:

Dalgaard, P. (2008), *Introductory Statistics with R*, 2nd Ed., Springer.
Groß, J. (2010), *Grundlegende Statistik mit R*, Vieweg/Teubner.
Verzani, J. (2005), *Using R for Introductory Statistics*, Chapman & Hall/CRC.

Inhaltsverzeichnis

1 Einleitung

1.1. [*Kreisdiagramm*] Von der ACEA (*European Automobile Manufacturers'Association*; `www.acea.be`) werden u.a. Daten über Neuzulassungen von Kraftfahrzeugen gesammelt. Für das Jahr 2009 ergab sich für die PKW–Neuzulassungen das folgende Bild, aufgeschlüsselt nach Herstellergruppen (Zahlen für Westeuropa; Datenfile: `pkw-neuzul09.dat`):

GROUP	TOTAL
ASTON MARTIN	2247
BMW	694712
CHINA	138
CRYSLER	52088
DAIMLER	671341
FIAT	1200252
FORD	1425086
GM	1213169
JAGUAR LAND ROVER	83914
JAPAN	1156928
KOREA	527267
PORSCHE	32560
PSA	1817517
RENAULT	1237527
TOYOTA	676739
VOLKSWAGEN	2854573
OTHER	19616

(*Bem*: Die Herstellergruppe JAPAN umfaßt die Marken DAIHATSU, HONDA, MAZDA, MITSUBISHI, NISSAN, SUBARU, SUZUKI, und andere.)

Fassen Sie die Herstellergruppen mit einem Anteil von weniger als 3% mit der Gruppe OTHER zusammen und erstellen Sie ein Kreisdiagramm („Tortendiagramm"). Für eine bessere Lesbarkeit des Diagramms empfiehlt sich eine Darstellung nach der Größe der geordneten Anteile.

R: Das Kreisdiagramm wird mittels `pie` gezeichnet. Für die Zusammenfassung bzw. Ordnung der Daten sind einige Vorbereitungen nötig:

```
pkw <- read.table("pkw-neuzul09.dat", header=TRUE, sep=";")
TO <- sum(pkw$TOTAL)
pkw2 <- subset(pkw, subset=100*(pkw$TOTAL/TO) > 3)
TO2 <- sum(pkw2$TOTAL)
pkw2.other <- data.frame(GROUP="OTHER", TOTAL=TO-TO2)
(pkw3 <- rbind(pkw2, pkw2.other))
# Der Größe nach ordnen:
ra <- sort(pkw3$TOTAL, index.return=TRUE)
pie(pkw3$TOTAL[ra$ix], labels=pkw3$GROUP[ra$ix],
  col=gray(seq(0.5,1.0,length=dim(pkw3)[1])),
  main="PKW Neuzulassungen 2009 (Western Europe)")
```

1.2. [*Balkendiagramm*] Die beliebtesten Gastländer österreichischer Erasmus–Studierender im Studienjahr 2008/09 waren wie folgt (Datenquelle: Beilage zum Standard vom März 2010; Datenfile: `erasmus0809.dat`):

	Studium	Praktikum
Deutschland	291	502
Spanien	677	71
Frankreich	493	32
Italien	395	20
Schweden	352	24
Großbritannien	308	63
Finnland	265	7
Niederlande	220	30
Irland	132	12
Norwegen	131	7
Dänemark	124	4
Portugal	111	4
Belgien	87	22
Tschechien	72	19
Polen	57	14
Türkei	68	2
Ungarn	40	21
Griechenland	41	5
Estland	27	3
Island	25	1
Slowenien	24	1
Litauen	19	1
Rumänien	15	5
Slowakei	15	4
Lettland	14	2
Malta	14	1
Bulgarien	4	3
Liechtenstein	1	4
Luxemburg	0	2
Zypern	0	0

Erstellen Sie Balkendiagramme für (1) Studium, für (2) Praktikum und für die (3) Gessamtzahlen (Studium + Praktikum). Ordnen Sie die Länder jeweils nach Beliebtheit.

R: Balkendiagramme werden mittels `barplot` gezeichnet. Für 'Studium' lauten entsprechende Commands z.B. wie folgt:

```
dat <- read.table("erasmus0809.dat", header=TRUE, skip=1)
Stud0809 <- dat$Stud
names(Stud0809) <- dat$Länder
ind.stud <- sort(Stud0809, decreasing=TRUE, index.return=TRUE)
barplot(Stud0809[ind.stud$ix], las=2, cex.names=0.7, axis.lty=1,
  main="Beliebteste Gastländer österreichischer Studierender\
  Studium 2008/09")
```

1.3. [*Paretodiagramm*] Eine Variante des Balkendiagramms ist das nach dem italienischen Ökonomen Vilfredo Federico Pareto (1848 – 1932) benannte Diagramm. Es ist eine graphische Umsetzung des in vielen Bereichen anwendbaren *Paretoprinzips*. (Recherchieren Sie, was darunter zu verstehen ist!) Erstellen Sie Paretodiagramme für die PKW–Neuzulassungen von Aufgabe 1.1 und für die Erasmus–Studierenden von Aufgabe 1.2. Interpretieren Sie die Diagramme.

R: Das Paretodiagramm wird mittels `pareto.chart` erstellt. Dazu muß zuerst das Package `qcc` installiert und geladen werden:

```
install.packages("qcc")
library(qcc)
```

1.4. [*Histogramm, Stem-and-Leaf–Plot*] Der Datensatz `alt.dat` enthält Messungen (Einheit: Stunden) der Lebensdauer von 40 elektronischen Komponenten, die einem beschleunigten Lebensdauertest (d.i. bei höherer Temperatur als beim üblichen Gebrauch) unterworfen wurden.

(a) Konstruieren Sie ein flächentreues Histogramm (d.i. ein Histogramm, dessen Fläche gleich 1 beträgt).

R: Ein unstrukturierter Datensatz wird mittels `scan` eingelesen. Histogramme werden mittels `hist` gezeichnet. Die Flächentreue der Darstellung erreicht man mit der Option `freq=FALSE` (oder `prob=TRUE`).

```
x <- scan("alt.dat")
hist.alt <- hist(x, freq=FALSE, xlab="Stunden", col="lightgrey")
```

(Wie lautet der Default für die Klasseneinteilung?)

(b) Erstellen Sie einen *Stem-and-Leaf–Plot*. Dies ist eine Art Strichliste und wird nur bei kleineren Datensätzen, deren Elemente nur wenige Dezimalstellen aufweisen, angewendet. Interpretieren Sie den Plot.

R: Die Funktion lautet `stem`.

1.5. [*Summenpolygon, Empirische Verteilungsfunktion*] Fortsetzung von Aufgabe 1.4:

(a) Zeichnen Sie das zu der von `hist` gewählten Klasseneinteilung passende Summenpolygon der relativen Häufigkeiten.

R: In `hist.alt` (ein Objekt vom Typ `list`) steht die benötigte Information:

```
n <- length(x)
plot(hist.alt$breaks, c(0,cumsum(hist.alt$counts))/n, type="o",
  pch=19, lwd=2, xlab="Stunden", ylab="Kumul. rel. Häufigkeiten")
```

(b) Zeichnen Sie die empirische Verteilungsfunktion.

R: Die emprische Verteilungsfunktion wird mittels `ecdf` bestimmt und mit `plot` gezeichnet (dabei gibt es noch weitere Darstellungsoptionen):

```
plot(ecdf(x), verticals=TRUE, do.points=FALSE, xlab="Stunden",
  main="Empirische Verteilungsfunktion")
```

1.6. [*Histogramm*] Die Datensätze `euroweight4.dat` und `euroweight6.dat` enthalten Messungen der Gewichte von jeweils 250 neuen 1€–Münzen. (*Bem.:* Teil einer größeren Studie an insgesamt 2000 belgischen 1€–Münzen.) Aus Gründen, die später in der Vorlesung erläutert werden, besteht die Vermutung, daß die Gewichte näherungsweise einer *Normalverteilung* („Glockenkurve“) folgen. Dies soll an den beiden Datensätzen näher untersucht werden.

(a) Betrachten Sie zunächst Batch 4 und erstellen Sie ein flächentreues Histogramm (d.h. ein Histogramm, dessen Fläche gleich 1 beträgt). Nehmen Sie dazu die folgende Klasseneinteilung:

$$[7.400, 7.410], (7.410, 7.420], \dots, (7.640, 7.650]$$

R: Die folgenden Commands zeichnen das Histogramm:

```
euro4 <- read.table("euroweight4.dat", header=TRUE, skip=1)[,2]
brk <- seq(7.400, 7.650, by=0.010)
hist(euro4, breaks=brk, freq=FALSE, main=paste("Batch",4),
  xlab="Gewicht [g]", col="lightgrey")
```

(b) Betrachten Sie nun Batch 6 und erstellen Sie ebenfalls ein flächentreues Histogramm. Nehmen Sie dazu dieselbe Klasseneinteilung wie für Batch 4. Plazieren Sie beide Histogramme zum einfacheren Vergleich in eine Abbildung. Kommentieren Sie das Ergebnis.

R: Um die beiden Histogramme in ein 2×1–Array zu zeichnen, kann man die folgenden Commands verwenden:

```
par(mfrow=c(2,1))
[Histogramm für Batch 4]
[Histogramm für Batch 6]
par(mfrow=c(1,1))
```

1.7. [*Kerndichteschätzung*] Die Klassen eines Histogramms bilden gewissermaßen Fenster, durch die die Daten betrachtet werden. Diese Vorstellung läßt sich dahingehend verallgemeinern, daß man nicht feste sondern gleitende Fenster betrachtet. Ist $x_1, \dots, x_n$ der Datensatz und $K(z)$ eine stetige, symmetrische *Kernfunktion* mit den Eigenschaften:

$$K(z) \geq 0, \qquad \int_{-\infty}^{\infty} K(z)\,dz = 1$$

dann nennt man die Funktion:

$$\widehat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R}$$

einen *Kerndichteschätzer* mit der *Bandbreite h*. Letztere bestimmt die „Glattheit“ der Schätzung. Bebräuchliche Kerne sind der *Normalkern* (vgl. Kapitel 3), der *Rechteckskern* ($K(z) = 1/2$ für $|z| \leq 1$, gleich 0 sonst) und der *Epanechnikoff–Kern*:

$$K(z) = \frac{3}{4}\,(1 - z^2), \quad |z| \leq 1 \text{ (gleich 0 sonst)}$$

(a) Zeigen Sie, daß die Epanechnikoff–Funktion eine Kernfunktion ist.

(b) Zeichnen Sie über die Histogramme von Aufgabe 1.6 die Kerndichteschätzungen. Nehmen Sie dazu den Normalkern und experimentieren Sie mit verschiedenen Bandbreiten.

R: Die Funktion `density` bestimmt die Kerndichteschätzung (Default ist der Normalkern; mittels der Option `bw` läßt sich die Bandbreite einstellen), einem bereits bestehenden Plot (Histogramm) hinzugefügt wird sie mittels `lines`:

```
[Commands für das Histogramm]
lines(density(euro4), lty=1, lwd=2)
```

1.8. [*Lageparameter*] Die folgenden Daten (`ozon.dat`) sind (der Größe nach geordnete) Meßwerte der Ozonbelastung an 38 Meßstellen an zwei aufeinanderfolgenden Tagen. (Datenquelle: Bundesumweltamt, `www.umweltbundesamt.at`; Meßwerte vom 4. und 5. Juli 2010, NOÖ)

	μg/m^3									
Tag 1	122	123	124	124	126	126	127	127	127	127
	128	128	129	130	131	132	132	133	133	134
	134	136	137	139	140	140	140	140	141	141
	146	148	149	149	152	154	157	161		
Tag 2	102	109	109	110	112	113	113	113	114	115
	116	116	117	117	118	118	118	118	118	119
	120	122	122	122	123	125	125	127	127	128
	128	128	129	129	130	130	131	135		

Berechnen Sie für beide Datensätze: Mittelwert, Median, die Quartile und die Hinges (*Bem.*: Die *Hinges* sind die Mediane der unteren/oberen Hälfte der geordneten Daten. Bei einer ungeraden Anzahl von Werten gehört der mittlere Wert zu beiden Hälften. Die Hinges entsprechen in etwa dem 1. und 3. Quartil.) Rechnen Sie „mit der Hand" und mit R.

R: Die Funktionen lauten `mean`, `median`, `quantile`. Verwenden Sie auch die Funktionen `summary`, `fivenum` sowie die (eigene) Funktion `kennz`.

Bem: Abgesehen vom Median sind empirische Quantile nicht einheitlich definiert. Praktisch sind die Unterschiede aber meist nicht relevant. In R werden 9 Typen unterschieden. Grundsätzlich kann man die Definitionen danach unterteilen, ob für Quantile nur Datenwerte oder auch (interpolierte) Werte dazwischen in Frage kommen. Zu letzterer Gruppe gehört auch die auf dem Summenpolygon basierende Definition der VO. (Vgl. für eine R–Implementierung die eigene Funktion `quant.vo`).

1.9. [*Streuungsparameter*] Berechnen Sie für die Ozondaten von Aufgabe 1.8: Spannweite, Quartilabstand/Hingeabstand, MAD, Varianz, Streuung und Variationskoeffizient. Rechnen Sie „mit der Hand" und mit R.

R: Die Funktionen lauten `IQR`, `var`, `sd`. Für andere Größen kann man ganz einfach eigene Funktionen schreiben (vgl. auch die eigene Funktion `kennz`). Verwenden Sie auch `summary` und `fivenum`.

1.10. [*Boxplot, Dotplot*] Fortsetzung von Aufgabe 1.8:

(a) Zeichnen Sie für beide Datensätze einen *Boxplot*. Letzterer (auch *Box-Whisker–Plot* genannt) ist quasi eine graphische Darstellung von `fivenum`. (Vgl. für weitere Details/Varianten z.B. WIKIPEDIA.)

R: Die Funktion lautet `boxplot`. Mittels des folgenden Commands werden die Boxplots für beide Datensätze nebeneinander gezeichnet:

```
boxplot(Ozon ~ Tag, data=dat, notch=TRUE, col="lightgrey")
```

(b) Zeichnen Sie für beide Datensätze einen *Dotplot*. Letzterer stellt die Werte einer quantitativen Variablen als Punkte auf einer Linie dar.

R: Die Funktion lautet `stripchart`. Mittels des folgenden Commands werden die Dotplots für beide Datensätze übereinander gezeichnet:

```
stripchart(Ozon ~ Tag, data=dat, method="stack")
```

1.11. [*Empirische Varianz*] Zeigen Sie die folgende alternative Berechnungsmöglichkeit für die empirische Varianz S_n^2:

$$S_n^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2\right]$$

(*Bem*: Die obige Darstellung von S_n^2 heißt auch *(empirischer) Verschiebungssatz*.)

1.12. [*Normalanpassung*] Wie in der VO angesprochen, versucht man in der schließenden Statistik u.a. den empirisch gegebenen Verteilungen (Histogrammen) theoretische Verteilungen (Dichten) anzupassen. Versuchen Sie dies für die EURO–Gewichte von Batch 4 und Batch 6 (Aufgabe 1.6) mit der Anpassung einer *Normaldichte*. Nehmen Sie für die beiden *Parameter* dieser Verteilung (μ, σ^2) die entsprechenden empirischen Größen ($\overline{x}$, s^2).

R: Die folgenden Commands leisten das Gewünschte (für Batch 4):

```
x <- read.table("euroweight4.dat", header=TRUE, skip=1)$weight
brk <- seq(7.400, 7.650, by=0.010)
hist.x <- hist(x, breaks=brk, plot=FALSE)
m <- mean(x); s <- sd(x)
z <- seq(m-4*s, m+4*s, length=100)
yR <- range(0, hist.x$density, dnorm(0, 0, s))
hist(x, breaks=brk, prob=TRUE, col="lightgrey", ylim=yR,
  xlab="Gewicht [g]", main="Normalanpassung: Batch 4")
lines(z, dnorm(z, m, s), lty=1, lwd=2)
```

1.13. [*Quantil-Quantil–Plot*] Der *QQ–Plot* ist eine einfache graphische Methode, um herauszufinden, ob zwei Datensätze aus derselben Verteilung stammen. Dazu zeichnet man die der Größe nach geordneten Werte der kleineren Stichprobe (Größe n) gegen die $(i-1)/(n-1)$–Quantile ($i = 1, \ldots, n$) der anderen Stichprobe. Liegen diese Punkte annähernd auf einer Geraden, so ist dies ein Indiz dafür, daß dieselbe Verteilung (abgesehen von möglichen Lage– und Skalierungsunterschieden) zugrunde liegt. Zeichnen Sie den QQ–Plot für die beiden Ozonmeßreihen von Aufgabe 1.8.

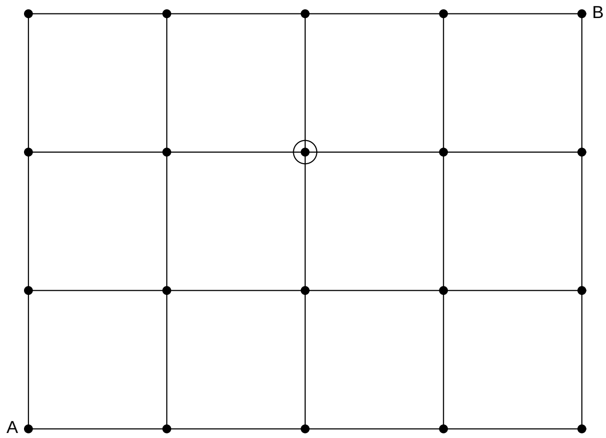
R: Die Funktion lautet `qqplot`.

1.14. [*Kombinatorik*] Eine Studentin möchte 10 Fachücher in ein Regal einordnen. Von den 10 Büchern beschäftigen sich 4 mit Mathematik, 3 mit Chemie, 2 mit Geschichte und eines mit Englisch. Die Studentin möchte die Bücher so anordnen, daß Bücher zur selben Thematik nebeneinander stehen. Wieviele verschiedene Anordnungen gibt es?

1.15. [*Kombinatorik*] Betrachten Sie das unten angegebene Gitter von Punkten. Sie starten im Punkt A und möchten nach Punkt B. Dabei können Sie nur jeweils einen Schritt nach oben oder nach rechts machen.

(a) Wieviele verschiedene Wege von A nach B gibt es?

(b) Wieviele verschiedene Wege von A nach B verlaufen durch den eingeringelten Punkt?



1.16. [*Kombinatorik*] Zeigen Sie, daß es $\binom{n-1}{r-1}$ verschiedene *positive* ganzzahlige Lösungsvektoren $(x_1, x_2, \ldots, x_r)$ der folgenden Gleichung gibt:

$$x_1 + x_2 + \cdots + x_r = n, \quad x_i > 0,\ i = 1, \ldots, r$$

(*Hinweis*: Betrachten Sie n nichtunterscheidbare Symbole, die Sie in r nichtleere Gruppen aufteilen möchten. Z.B lautet für $n = 8$ und $r = 3$ eine mögliche Aufteilung 000|000|00.) Wieviele verschiedene *nichtnegative* ganzzahlige Lösungsvektoren $(x_1, x_2, \ldots, x_r)$ hat die obige Gleichung? Illustrieren Sie die Ergebnisse an einem einfachen Beispiel, etwa an den Lösungen von $x_1 + x_2 = 3$.

1.17. [*Kombinatorik*] Betrachten Sie ein Turnier an dem n Personen teilnehmen. Das Ergebnis des Turniers besteht in einer Gruppierung der Teilnehmer derart, daß die erste Gruppe aus den Personen besteht, die sich den ersten Platz teilen, die nächste Gruppe aus den Personen, die sich den nächstbesten Platz teilen, usw. $N(n)$ sei die Zahl der verschiedenen möglichen Ergebnisse des Turniers. (Beispielsweise gilt $N(2) = 3$: Bei 2 Teilnehmern kann Person 1 oder Person 2 den alleinigen ersten Platz einnehmen, oder beide teilen sich den ersten Platz.)

(a) Wie lauten alle möglichen Ergebnisse für $n = 3$?

(b) Zeigen Sie *ohne* Rechnung, daß (mit $N(0) := 1$):

$$N(n) = \sum_{i=1}^{n} \binom{n}{i} N(n-i)$$

(c) Zeigen Sie, daß $N(n)$ auch wie folgt geschrieben werden kann:

$$N(n) = \sum_{i=0}^{n-1} \binom{n}{i} N(i)$$

(d) Verwenden Sie die Rekursion zur Berechnung von $N(3)$ und $N(4)$.

1.18. [*Klassische Wahrscheinlichkeit*] Bestimmen Sie die Wahrscheinlichkeiten der einzelnen Gewinnränge beim *Joker*. (*Bem*: Details zum Spiel finden Sie auf der Homepage der Österreichischen Lotterien, `www.win2day.at`.)

1.19. [*Klassische Wahrscheinlichkeit*] Einem üblichen Kartenpaket aus 52 Karten (4 Farben: Kreuz, Herz, Pik, Karo) werden 10 Karten zufällig entnommen. Jede der gezogenen Karten wird abhängig von der Farbe auf einen Stapel gelegt. Mit welcher Wahrscheinlichkeit umfaßt der höchste Stapel 4 Karten, der nächst höhere 3 Karten, der nächst höhere 2 Karten und der niedrigste 1 Karte?

1.20. [*Klassische Wahrscheinlichkeit*] Eine Lade enthält 10 verschiedene Sockenpaare. Wenn 8 Socken zufällig entnommen werden, wie groß ist die Wahrscheinlichkeit, daß (a) kein Paar, (b) genau ein Paar darunter ist?

1.21. [*Klassische Wahrscheinlichkeit*] Ein Array der Länge N wird auf zufällige Weise mit Elementen belegt.

(a) Mit welcher Wahrscheinlichkeit gibt es unter den ersten n Belegungen *keine* Kollision?

(b) Mit welcher Wahrscheinlichkeit kommt es bei der n–ten Belegung zur *ersten* Kollision?

(c) Betrachten Sie (a) und (b) konkret für den Fall $N = 365$. (*Bem*: Dieser Fall heißt auch „Geburtstagsproblem".) Wie groß muß n mindestens sein, sodaß die Wahrscheinlichkeit von (a) kleiner als $1/2$ ist?

1.22. [*Klassische Wahrscheinlichkeit*] Ein Behälter enthalte 20 rote und 10 blaue Kugeln. Die Kugeln werden zufällig nacheinander ohne Zurücklegen entnommen. Mit welcher Wahrscheinlichkeit werden auf diese Weise alle rote Kugeln vor allen blauen Kugeln entnommen? (*Hinweis*: Dies ist genau dann der Fall, wenn die letzte entnommene Kugel blau ist.)

1.23. [*Runs*] Betrachten Sie eine Binärfolge bestehend aus N Elementen, m der einen und n der anderen Art ($m + n = N$). Ein *Run* ist eine Teilfolge aus identischen Elementen, begrenzt auf beiden Seiten von einem Element der anderen Art (oder vom Anfang/Ende der Folge). Beispielsweise gibt es in der Folge der Länge $N = 20$ ($m = 11$, $n = 9$):

$$\underline{0}\ \underline{1\ 1}\ \underline{0\ 0}\ \underline{1\ 1}\ \underline{0}\ \underline{1\ 1\ 1}\ \underline{0\ 0\ 0\ 0\ 0}\ \underline{1\ 1}\ \underline{0\ 0}$$

insgesamt neun Runs, fünf 0–Runs und vier 1–Runs.

(a) Ist R die Gesamtzahl der Runs in einer zufällig angeordneten Binärfolge der Länge N, mit m Elementen der einen und n Elementen der anderen Art, so bestimme man die Wahrscheinlichkeit, daß (1) $R = 2k$ (d.h., eine gerade Zahl) und daß (2) $R = 2k + 1$ (d.h., eine ungerade Zahl) ist. (*Hinweis*: Lösung u.a. auf Wikipedia, Stichwort 'Run–Test'; vgl. zum Verständnis der Lösung Aufgabe 1.16.)

(b) Bestimmen Sie speziell die Verteilung der Zahl der Runs für $m = 2$ und $n = 5$. Welche Anzahl ist am wahrscheinlichsten?

1.24. [*Geometrische Wahrscheinlichkeit*] Angenommen, zwei Signale erreichen (unabhängig voneinander) einen Empfänger zu einem beliebigen Zeitpunkt im Intervall $[0, T]$. Der Empfänger blockiert, wenn die Zeitdifferenz zwischen den Signalen kleiner als τ ist. Mit welcher Wahrscheinlichkeit blockiert der Empfänger?

1.25. [*Geometrische Wahrscheinlichkeit*] Ein dünner Stab der Länge $L = 200$ mm wird zunächst an zwei willkürlich gewählten Stellen x und y $(0 < x, y < L)$ markiert und dann an diesen Stellen durchgesägt, wodurch drei Stücke entstehen. Mit welcher Wahrscheinlichkeit ist zumindest eines dieser Stücke nicht länger als 10 mm ?

1.26. [*Geometrische Wahrscheinlichkeit*] Ein Satellit, dessen Orbit zwischen 60° nördlicher und 60° südlicher Breite liegt, droht abzustürzen. Wenn jeder Punkt auf dieser Erdkugelzone mit gleicher Wahrscheinlichkeit als Absturzstelle in Frage kommt, mit welcher Wahrscheinlichkeit wird der Satellit oberhalb 30° nördlicher Breite abstürzen? (*Hinweis*: Die Fläche einer Kugelzone ist $A = 2\pi rh$, h = Höhe der Zone.)

1.27. [*Geometrische Wahrscheinlichkeit*] Jemand verläßt zufällig zwischen 16 und 17 Uhr den Arbeitsplatz und begibt sich zur U–Bahn. Die Mutter lebt in der Nähe der einen Endstation, die/der Freund/in in der Nähe der anderen. Er/Sie will fair sein und nimmt jeweils diejenige U–Bahn, welche als erste eintrifft. Nach einiger Zeit beklagt sich die Mutter darüber, daß er/sie nur ganz selten zum Abendessen kommt, an den letzten 20 Arbeitstagen nur zweimal. Kommt dieses Ungleichgewicht zufällig zustande oder gibt es eine andere Erklärung dafür? (*Hinweis*: Nehmen Sie an, daß die U–Bahn ganz regelmäßig fährt.)

1.28. [*Empirisches GGZ*] Ein Beispiel aus den Anfängen der modernen Wahrscheinlichkeitsrechnung: Der französische Offizier und Schriftsteller Chevalier de Méré (1607 – 1684) wandte sich im Jahre 1654 mit der folgenden Frage an Blaise Pascal (1623 – 1662): Was ist vorteilhafter, beim Spiel mit einem Würfel auf das Eintreten mindestes eines Sechsers in vier Würfen oder beim Spiel mit zwei Würfeln auf das Eintreten eines Doppelsechsers in 24 Würfen zu setzen? De Méré wußte aus Erfahrung, daß die erste Wette für ihn vorteilhaft ist. Bei der zweiten Wette, von der er annahm, daß sie nur eine Variante der ersten sei, gestalteten sich die Einnahmen aber nicht ganz nach seinen Vorstellungen.

Bearbeiten Sie das Problem zunächst empirisch unter Verwendung der Funktion `demere`. Bestimmen Sie anschließend die exakten Wahrscheinlichkeiten.

Anhang 1

A.1 <u>Binomialkoeffizienten:</u>

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{(n)_r}{r!}$$

$$\binom{n}{0} = \binom{n}{n} = 1, \quad i < 0 \text{ oder } i > n : \binom{n}{i} = 0$$

$$1 \le r \le n: \quad \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \quad \text{(Pascal'sches Dreieck)}$$

A.2 Binomischer Lehrsatz:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

A.3 Multinomialkoeffizienten:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

$$\text{Spezialfall}: \binom{n}{r, n-r} = \binom{n}{r}$$

$$\binom{n}{n_1, n_2, \ldots, n_r} = \binom{n-1}{n_1 - 1, n_2, \ldots, n_r} + \binom{n-1}{n_1, n_2 - 1, \ldots, n_r} + \cdots$$

$$\cdots + \binom{n-1}{n_1, n_2, \ldots, n_r - 1}$$

A.4 Multinomialsatz:

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{(n_1,\ldots,n_r): \\ n_1 + \cdots + n_r = n}} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

A.5 Kombinatorik: Anzahlen möglicher Anordnungen oder Auswahlen von unterscheidbaren oder nicht unterscheidbaren Objekten mit oder ohne Beachtung der Reihenfolge.

1. Allgemeines Zählprinzip: Wenn eine Aufgabe durch eine Abfolge von k Schritten beschrieben werden kann, und wenn Schritt 1 auf n_1 verschiedene Arten erledigt werden kann, und wenn Schritt 2 – für jede Art der ersten Stufe – auf n_2 verschiedene Arten erledigt werden kann, usf., dann ist die Zahl der verschiedenen Möglichkeiten, die Aufgabe zu erledigen, gegeben durch:

$$n_1 n_2 \cdots n_k$$

2. Permutationen: Anordnungen von n Objekten, wobei alle Objekte vorkommen, mit Beachtung der Reihenfolge.

   (a) Unterscheidbare Objekte: Die Zahl der Permutationen von n verschiedenen Objekten beträgt:

$$n! = n(n-1)(n-2)\cdots(2)(1)$$

(b) Objekte mehrerer Klassen: Die Zahl der Permutationen von n Objekten, die in k Klassen zu je $n_1, n_2, \ldots, n_k$ ($\sum_{i=1}^{k} n_k = n$) gleichen Objekten vorliegen, beträgt:

$$\frac{n!}{n_1!\, n_2! \cdots n_k!} = \binom{n}{n_1, n_2, \ldots, n_k}$$

Bsp: Wieviele verschiedene Barcodes aus vier dicken, drei mittleren und zwei dünnen Linien gibt es?

$$\frac{9!}{4!\, 3!\, 2!} = 1260$$

3. Variationen: Auswählen von Objekten mit Beachtung der Reihenfolge.

(a) Ohne Zurücklegen: Die Zahl der Möglichkeiten aus n verschiedenen Objekten k ($\leq n$) Objekte ohne Zurücklegen und unter Beachtung der Reihenfolge auszuwählen, beträgt:

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

Bsp: Auf einer Platine gibt es acht verschiedene Stellen, an denen eine Komponente plaziert werden kann. Wenn vier verschiedene Komponenten plaziert werden sollen, wieviele verschiedene Designs gibt es?

$$(8)_4 = (8)(7)(6)(5) = \frac{8!}{4!} = 1680$$

(b) Mit Zurücklegen: Für die Auswahl von k Objekten aus n verschiedenen Objekten mit Zurücklegen und unter Beachtung der Reihenfolge gibt es n^k Möglichkeiten.

4. Kombinationen: Auswählen von Objekten ohne Beachtung der Reihenfolge.

(a) Ohne Zurücklegen: Die Zahl der Möglichkeiten aus n verschiedenen Objekten k ($\leq n$) Objekte ohne Zurücklegen und ohne Beachtung der Reihenfolge auszuwählen, beträgt:

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

Bsp: Wieviele Möglichkeiten gibt es, aus den Zahlen von 1 bis 45 sechs Zahlen ohne Zurücklegen und ohne Beachtung der Reihenfolge auszuwählen?

$$\binom{45}{6} = 8145060$$

(b) Mit Zurücklegen: Die Zahl der Möglichkeiten aus n verschiedenen Objekten k Objekte mit Zurücklegen und ohne Beachtung der Reihenfolge auszuwählen, beträgt:

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

Bsp: Ein gefüllter Getränkeautomat bietet 15 verschiedene Softdrinks an. Wenn Sie drei Flaschen entnehmen möchten, wobei die Marke egal ist, wieviele Möglichkeiten haben Sie?

$$\binom{15+3-1}{3} = \binom{17}{3} = 680$$

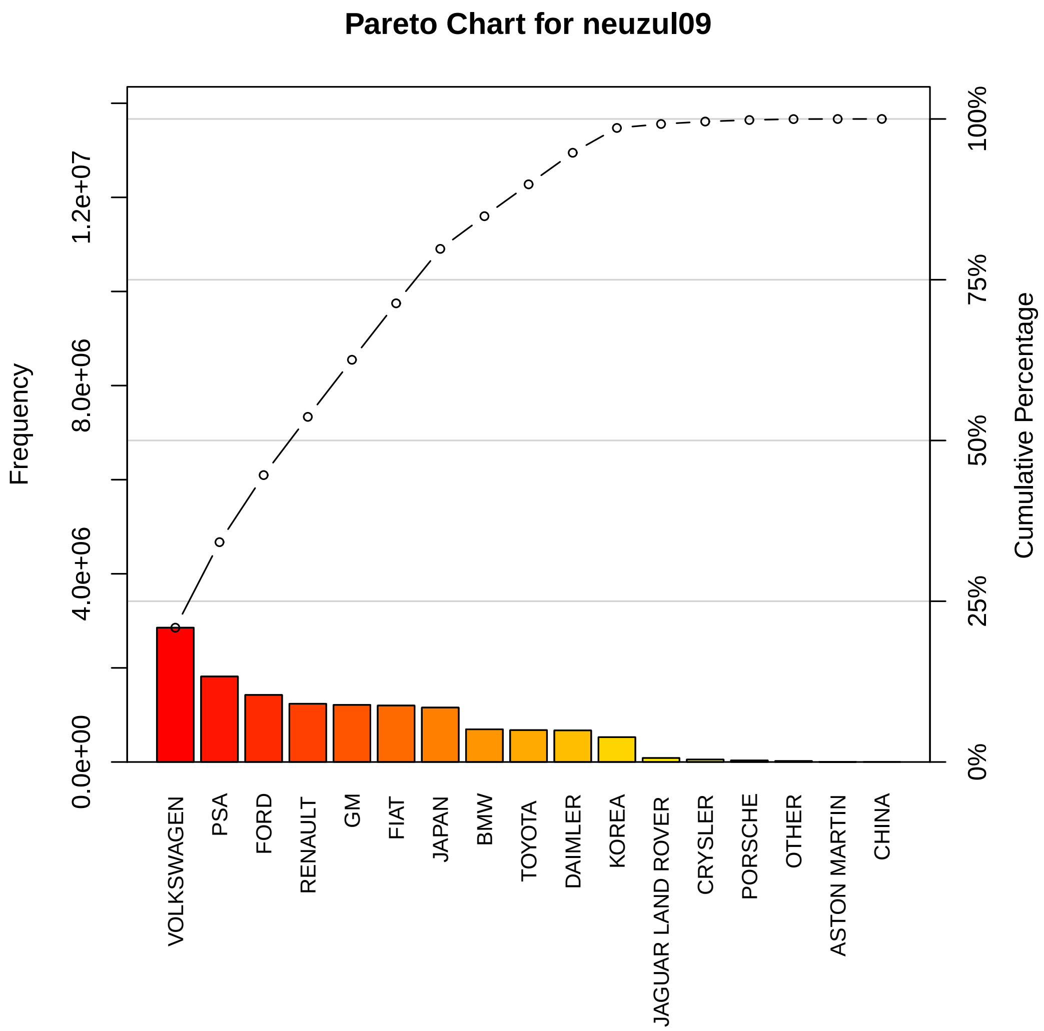
Lösungen/Bemerkungen zu ausgewählten Aufgaben

1.3. Das Paretodiagramm für die PKW–Neuzulassungen 2009 (Originaldaten) bekommt man wie folgt:

```
pkw <- read.table("pkw-neuzul09.dat", header=TRUE, sep=";")
install.packages("qcc")
library(qcc)
neuzul09 <- pkw$TOTAL
names(neuzul09) <- pkw$GROUP
pareto.chart(neuzul09, cex.names=0.8)
```

```
Pareto chart analysis for neuzul09
                    Frequency Cum.Freq.   Percentage Cum.Percent.
  VOLKSWAGEN          2854573   2854573 20.888636740     20.88864
  PSA                 1817517   4672090 13.299870903     34.18851
  FORD                1425086   6097176 10.428215981     44.61672
  RENAULT             1237527   7334703  9.055733365     53.67246
  GM                  1213169   8547872  8.877491150     62.54995
  FIAT                1200252   9748124  8.782969651     71.33292
  JAPAN               1156928  10905052  8.465941746     79.79886
  BMW                  694712  11599764  5.083627782     84.88249
  TOYOTA               676739  12276503  4.952108473     89.83460
  DAIMLER              671341  12947844  4.912608043     94.74720
  KOREA                527267  13475111  3.858331466     98.60554
  JAGUAR LAND ROVER     83914  13559025  0.614049479     99.21958
  CRYSLER               52088  13611113  0.381159393     99.60074
  PORSCHE               32560  13643673  0.238261208     99.83901
  OTHER                 19616  13663289  0.143542133     99.98255
  ASTON MARTIN           2247  13665536  0.016442658     99.99899
  CHINA                   138  13665674  0.001009829    100.00000
```

Dem Output von `pareto.chart` und dem Diagramm kann man z.B. entnehmen, daß ca. 80% der Neuzulassungen auf das Konto der Herstellergruppen VOLKSWAGEN, PSA, FORD, RENAULT, GM, FIAT und JAPAN gehen. Die anderen spielen bei den Neuzulassungen nur eine untergeordnete Rolle.

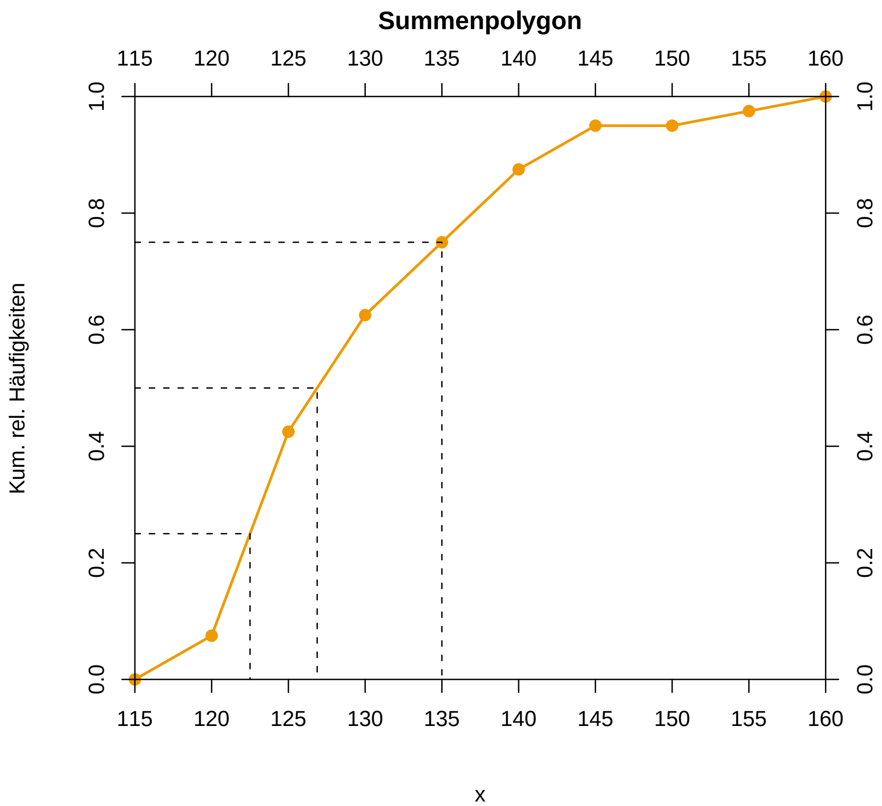


1.8. Bei mehreren Teildatensätzen kann man die Größen auch mit einem Command berechnen, beispielsweise:

```
dat <- read.table("ozon.dat", header=TRUE)
attach(dat)

by(Ozon, Tag, summary)

Tag: 1
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  122.0   127.2   133.5   136.0   140.8   161.0
---------------------------------------------------------
Tag: 2
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  102.0   115.2   118.5   120.2   127.0   135.0

detach(dat)
```

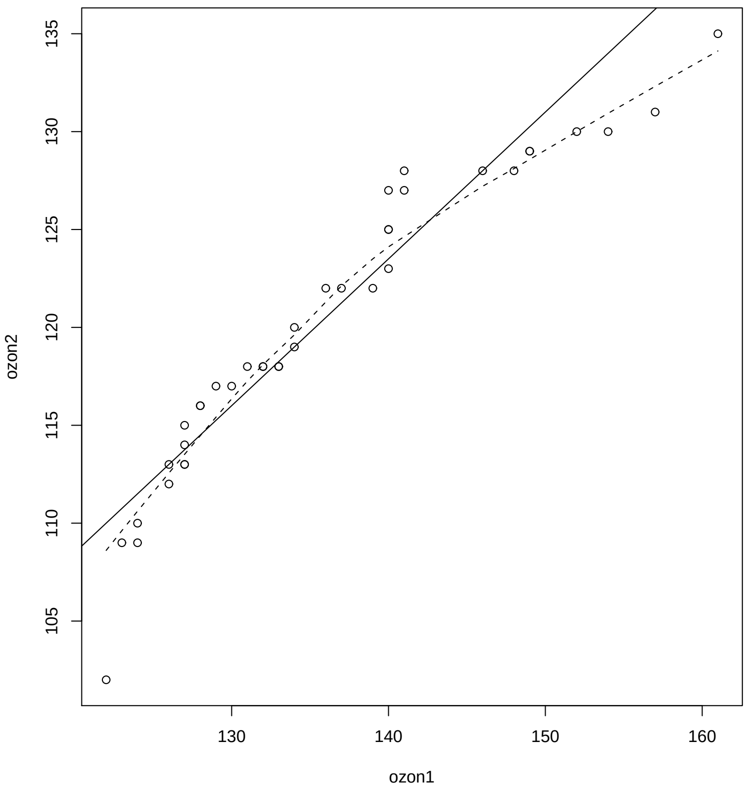
Beispiel (`alt.dat`) für die Verwendung von `quant.vo`:

```
x <- scan("alt.dat")
n <- length(x)
par(mfrow=c(2,1))
hist.alt <- hist(x, plot=FALSE)
quant.vo(x, hist.alt$breaks, c(0.25,0.5,0.75))
[1] 122.500 126.875 135.000
```



1.13. Der mittels `qqplot` gezeichnete QQ–Plot wird noch durch eine „robuste" Ausgleichsgerade (verläuft durch das 1. und 3. Quartil) und einen „Scatterplotsmoother" ergänzt. (*Bem*: Die Funktion `qqline` funktioniert nur zusammen mit `qqnorm`; vgl. Kapitel 6.)

```
dat <- read.table("ozon.dat", header=TRUE)
attach(dat)
ozon1 <- Ozon[Tag == 1]
ozon2 <- Ozon[Tag == 2]
ans.qq <- qqplot(ozon1, ozon2)
abline(line(ans.qq$x, ans.qq$y))          # robuste Gerade
lines(lowess(ans.qq$x, ans.qq$y), lty=2)  # Smoother
detach(dat)
```



Die deutliche Krümmung im Diagramm zeigt, daß die beiden Meßreihen aus verschiedenen Verteilungen stammen.

Bem: Bei Datensätzen gleichen Umfangs werden beim QQ–Plot einfach die der Größe nach geordneten Daten gegeneinander gezeichnet.

1.17. Die Darstellung (b) sieht man wie folgt: Es gibt $\binom{n}{i}$ Möglichkeiten für die Auswahl von i Personen, die sich den *letzten* Platz teilen. Für die restlichen $n-i$ Personen gibt es jeweils $N(n-i)$ mögliche Turnierergebnisse.

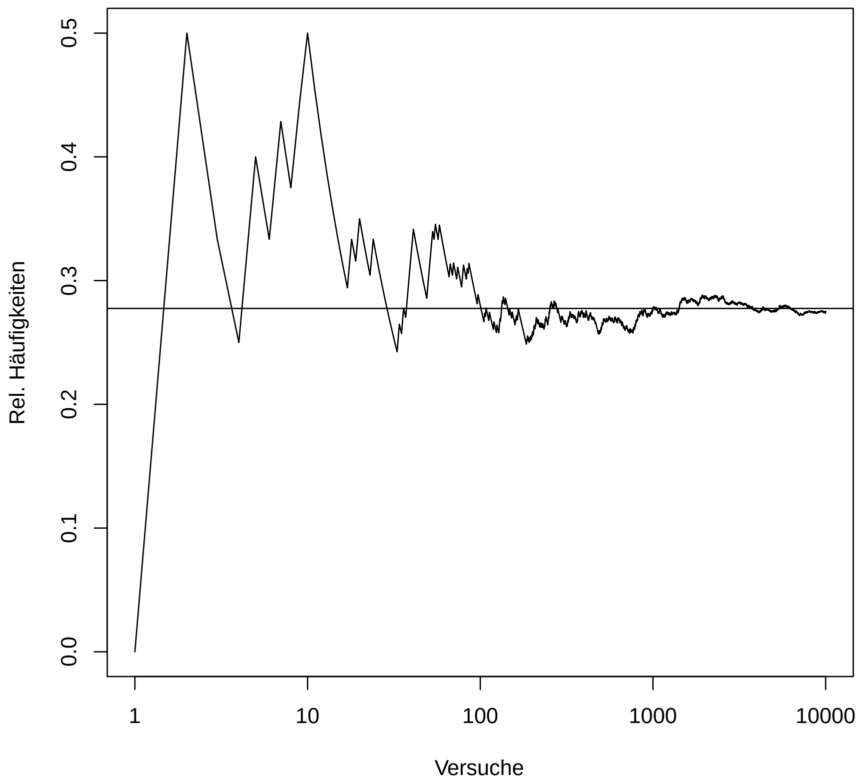
R: $N(n)$ läßt sich mittels einer rekursiven Funktion berechnen:

```
N.rek <- function(n) {
  if (n == 0) return(1)
  else if (n == 1) return(1)
  else { M <- 1
    for (i in 1:(n-1)) {
       M <- M + choose(n, i)*N.rek(i) }
    return(M) }
}
```

1.25. Der folgende R–Code simuliert das Experiment B Mal und stellt die kumulierten relativen Häufigkeiten für den Eintritt des fraglichen Ereignisses graphisch dar (*Empirisches Gesetz der Großen Zahlen*). Die waagrechte Linie entspricht der exakten Lösung.

```
B <- 10000
M <- matrix(rep(0,3*B), ncol=3)
for (i in 1:B) {
  u <- runif(2, min=0, max=200)
  x <- u[1]
  y <- u[2]
  M[i,1] <- ifelse(x < y, x, y)
  M[i,2] <- ifelse(x < y, y-x, x-y)
  M[i,3] <- ifelse(x < y, 200-y, 200-x)
  }
H <- apply(M, 1, function(x) ifelse(min(x) > 10, 0, 1))
H.rel.cum <- cumsum(H)/(1:B)
plot(1:B, H.rel.cum, type="l", lty=1, log="x",
  ylim=c(0,max(H.rel.cum)), xlab="Versuche",
  ylab="Rel. Häufigkeiten")
abline(h=1-(170/200)^2)
```

Ein Durchlauf mit $B = 10000$ ergibt:



2 Stochastische Grundbegriffe

2.1. [*DeMorgan'sche Regeln*] $A_1, A_2, \ldots$ seien Ereignisse aus einem Ereignisfeld $\mathcal{E}$. Zeigen Sie:

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c, \qquad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$$

2.2. [*Boole'sche Ungleichung*] $A_1, A_2, \ldots$ seien Ereignisse aus einem Ereignisfeld $\mathcal{E}$. Zeigen Sie:

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

(*Hinweis*: Stellen Sie $\bigcup_{i=1}^{\infty} A_i$ als Vereinigung von disjunkten Mengen dar.)

2.3. [*Bonferroni'sche Ungleichung*] $A_1, A_2, \ldots, A_n$ seien Ereignisse aus einem Ereignisfeld $\mathcal{E}$. Zeigen Sie:

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

(*Hinweis*: Führen Sie die Ungleichung auf die Boole'sche Ungleichung von Aufgabe 2.2 zurück.) Wie lautet die Ungleichung für eine unendliche Folge $A_1, A_2, \ldots$ von Ereignissen?

2.4. [*Borelmengen*] Das Ereignisfeld $\mathcal{B}$ der *Borelmengen* in $\mathbb{R}$ ist definiert als das kleinste Ereignisfeld, das alle halboffenen Intervall der Form $(a, b]$ ($a < b$, $a, b \in \mathbb{R}$) umfaßt. Zeigen Sie, daß (1) alle einpunktigen Mengen $\{a\}$, (2) alle offenen Intervalle (a, b) und (3) alle abgeschlossenen Intervalle $[a, b]$ Borelmengen sind.

2.5. [*Additionstheorem*] Formulieren und beweisen Sie das Additionstheorem für drei Ereignisse A, B und C. (*Bem.*: Geben Sie einen exakten Beweis, also nicht einen Beweis allein mittels Venndiagramm.)

2.6. [*Disjunkte Ereignisse*]

(a) Ein Würfelpaar wird geworfen. Mit welcher Wahrscheinlichkeit zeigt der zweite Würfel eine höhere Augenzahl als der erste Würfel? (*Bem.*: Einer der beiden Würfel sei der erste und der andere der zweite.)

(b) Ein Würfelpaar wird geworfen. Mit welcher Wahrscheinlichkeit beträgt die Augensumme i, $i = 2, 3, \ldots, 12$?

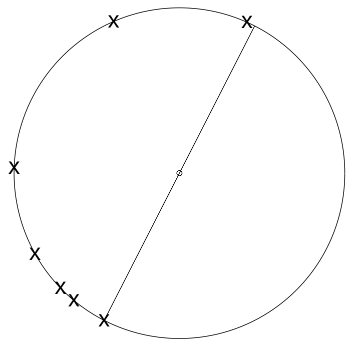
(c) Ein Würfelpaar wird solange geworfen, bis die Augensumme 5 oder 7 kommt. Mit welcher Wahrscheinlichkeit kommt 5 zuerst? (*Hinweis*: E_n sei das Ereignis, daß beim n–ten Wurf 5 kommt, aber weder 5 noch 7 bei den ersten $n - 1$ Würfen. Berechnen Sie $P(E_n)$ und argumentieren Sie, daß $\sum_{n=1}^{\infty} P(E_n)$ die gesuchte Wahrscheinlichkeit ist.)

2.7. [*Disjunkte Ereignisse*] Betrachten Sie zwei disjunkte Ereignisse A und B ($P(A) > 0$, $P(B) > 0$, $P(A) + P(B) \leq 1$), die bei einem Experiment eintreten können. Das Experiment werde solange unabhängig wiederholt, bis A oder B eintritt. Mit welcher Wahrscheinlichkeit kommt A vor B ? Zeigen Sie, daß letztere Wahrscheinlichkeit gegeben ist durch:

$$\frac{P(A)}{P(A)+P(B)}$$

(*Hinweis*: Dies läßt sich analog zu Aufgabe 2.6(c) zeigen. Als Alternative kann man aber auch durch das Ergebnis des *ersten* Experiments *bedingen* und den Satz von der vollständigen Wahrscheinlichkeit verwenden. Bearbeiten Sie die Aufgabe mit beiden Methoden.)

2.8. [*Disjunkte Ereignisse*] Angenommen, n Punkte werden zufällig und unabhängig auf einer Kreislinie markiert. Gesucht ist die Wahrscheinlichkeit, daß alle innerhalb eines Halbkreises liegen. Anders fomuliert, mit welcher Wahrscheinlichkeit gibt es eine Gerade durch den Mittelpunkt des Kreises, sodaß alle Punkte auf einer Seite der Geraden liegen?



$P_1, \ldots, P_n$ seien die n Punkte. A sei das Ereignis, daß alle Punkte in einem Halbkreis liegen, und A_i sei das Ereignis, daß alle Punkte im Halbkreis beginnend im Punkt P_i und weiter 180° im Uhrzeigersinn liegen, $i = 1, \ldots, n$. Drücken Sie A mit Hilfe der A_i aus und berechnen Sie $P(A)$.

2.9. [*Multiplikationstheorem*] Ein übliches Kartenpaket (52 Karten; 4 Farben: Kreuz, Herz, Pik, Karo; 13 Werte: 2–10, Bube (Jack), Dame (Queen), König, Ass) wird zufällig auf 4 Pakete zu je 13 Karten aufgeteilt. Mit welcher Wahrscheinlichkeit enthält jedes Paket ein Ass? (*Hinweis*: E_i sei das Ereignis, daß das i–te Paket genau ein Ass enthält. Berechnen Sie $P\big(\bigcap_{i=1}^4 E_i\big)$ mit Hilfe des Multiplikationstheorems.)

2.10. [*Bayes'sche Formel*] Ein Labortest entdeckt zu 95% eine bestimmte Erkrankung, wenn sie tatsächlich vorliegt. Der Test zeigt aber auch bei 1% der nicht erkrankten Personen ein „falsch positives“ Ergebnis. Wenn man vermutet, daß ca. 0.5% der Bevölkerung die Krankheit hat, wie groß ist die Wahrscheinlichkeit, daß eine zufällig ausgewählte Person, deren Test positiv ist, die Krankheit hat? (Geben Sie eine anschauliche Erklärung für die – unerwartet? – niedrige Wahrscheinlichkeit.)

2.11. [*Bayes'sche Formel*] Ein Flugzeug wird vermißt und man geht davon aus, daß es sich mit gleicher Wahrscheinlichkeit in einer von drei möglichen Regionen befindet. Sei $1-\beta_i$, $i = 1, 2, 3$, die Wahrscheinlichkeit, daß das Flugzeug bei einer Suche in der i–ten Region gefunden wird, wenn es sich dort befindet. (*Bem*: Abhängig von den geographischen und sonstigen Gegebenheiten ist β_i die Wahrscheinlichkeit, daß das Flugzeug übersehen wird.) Bestimmen Sie die bedingte Wahrscheinlichkeit, daß sich das Flugzeug in der i–ten Region, $i = 1, 2, 3$, befindet, wenn die Suche in Region 1 erfolglos war.

2.12. [*Bayes'sche Formel*] An einem bestimmten Punkt der Ermittlungen ist der Kommissar zu 60% davon überzeugt, daß der Hauptverdächtige der Täter ist. Ein *neues* Beweisstück zeigt, daß der Täter eine bestimmte Eigenart (Linkshänder, braune Haare, o. dgl.) hat. Wenn 20% der Bevölkerung diese Eigenart aufweist, wie überzeugt kann der Kommissar nun sein, wenn sich herausstellt, daß der Verdächtige diese Eigenart hat?

2.13. [*Chancen*] Die *Chance* (engl. *odds*) eines Ereignisses A ist definiert als:

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

Die Chance eines Ereignisses gibt an, um wieviel wahrscheinlicher der Eintritt gegenüber dem Nichteintritt des Ereignisses ist. Beträgt die Chance α, so sagt man üblicherweise, daß die Chancen $\alpha : 1$ zu Gunsten von A stehen.

(a) Angenommen, die Chancen für A stehen $2/3 : 1$. Wie groß ist $P(A)$?

(b) Eine Hypothese H treffe mit Wahrscheinlichkeit $P(H)$ zu. Nun wird ein neuer „Beweis“ E beobachtet. Zeigen Sie, daß sich die Chance von H wie folgt transformiert:

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \frac{P(E|H)}{P(E|H^c)}$$

Illustrieren Sie diese Formel an der Situation von Aufgabe 2.12.

(c) Vergleicht man die Chancen von zwei Ereignissen A und B spricht man vom *Chancenverhältnis* (engl. *odds ratio*):

$$\frac{P(A)/P(A^c)}{P(B)/P(B^c)} = \frac{P(A)[1 - P(B)]}{P(B)[1 - P(A)]}$$

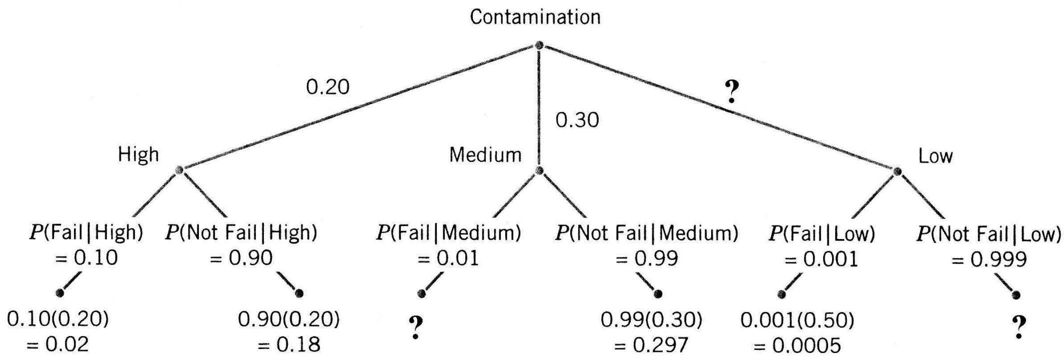
Dieses Verhältnis gibt an, um wieviel wahrscheinlicher der Eintritt von A gegenüber dem Eintritt von B ist. Analog zu (b) gilt nun:

$$\frac{P(H|E)}{P(G|E)} = \frac{P(H)}{P(G)} \frac{P(E|H)}{P(E|G)}$$

Wenn vor der Beobachtung von E die Hypothese H dreimal so wahrscheinlich wie die Hypothese G ist, E aber unter G zweimal so wahrscheinlich wie unter H ist, welche Hypothese ist nach der Beobachtung von E wahrscheinlicher?

2.14. [*Bayes'sches Netzwerk*] Ein (grob simplifiziertes) Beispiel für ein *Bayes'sches Netzwerk*: Ein Druckerhersteller entnimmt seiner Datenbasis, daß i.W. drei Fehlertypen auftreten: Hardwarefehler, Softwarefehler und „Sonstige“ (z.B. Anschlußfehler), mit den Wahrscheinlichkeiten 0.1, 0.6 und 0.3. Gibt es ein Hardwareproblem, ist der Drucker mit Wahrscheinlichkeit 0.9 defekt, bei einem Softwareproblem mit 0.2, und bei sonstigen Problemen mit 0.5. Wenn sich ein Kunde mit einem defekten Drucker an den Hersteller wendet, um welches Problem handelt es sich mit größter Wahrscheinlichkeit?

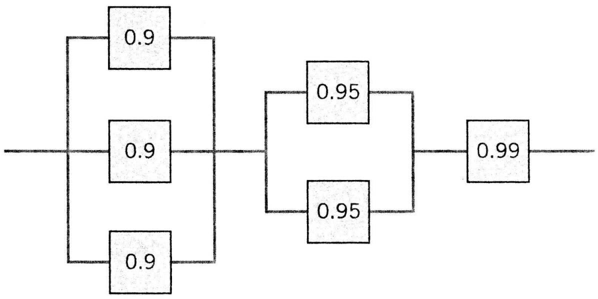
2.15. [*Wahrscheinlichkeitsbaum*] Halbleiterchips sind häufig durch Partikel kontaminiert. Je nach Grad der Kontaminierung (hoch, mittel, niedrig) hat dies unterschiedliche Auswirkungen auf die Funktionsfähigkeit der Chips. Der folgende *Wahrscheinlichkeitsbaum* zeigt die diesbezüglichen Wahrscheinlichkeiten.



(a) Ergänzen Sie die fehlenden Einträge (?) und bestimmen Sie die Wahrscheinlichkeit, daß ein zufällig ausgewählter Chip aus dieser Produktion defekt ist.

(b) Wenn einer dieser Chips defekt ist, mit welcher Wahrscheinlichkeit ist sein Kontaminierungsgrad hoch?

2.16. [*Unabhängige Ereignisse*] Das folgende System funktioniert nur, wenn es einen Pfad aus funktionierenden Komponenten von links nach rechts gibt. Die angegebenen Werte sind die Intaktwahrscheinlichkeiten der Komponenten. Man nehme an, daß die Komponenten unabhängig voneinander funktionieren/ausfallen.

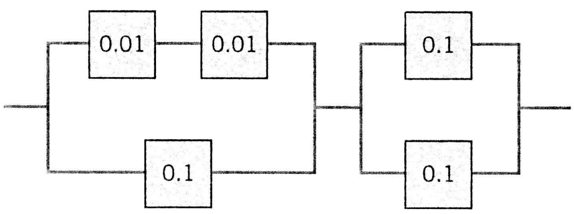


(a) Wie lautet ein passender Merkmalraum? Wieviele Elemente hat er? Wie lautet und aus wievielen Elementen besteht das zugehörige Ereignisfeld?

(b) Beschreiben Sie auf Basis von (a) die Ereignisse {Komponente i funktioniert} und {System funktioniert}.

(c) Berechnen Sie $P(\{\text{System funktioniert}\})$.

2.17. [*Unabhängige Ereignisse*] Das folgende System funktioniert nur, wenn es einen Pfad aus funktionierenden Komponenten von links nach rechts gibt. Die angegebenen Werte sind die Defektwahrscheinlichkeiten der Komponenten. Man nehme an, daß die Komponenten unabhängig voneinander funktionieren/ausfallen.



(a) Berechnen Sie $P(\{\text{System funktioniert}\})$.

(b) Um die Funktionssicherheit des Systems zu erhöhen, wird jede Komponente verdoppelt, d.h. durch ein Parallesystem aus zwei gleichartigen Komponenten ersetzt. Berechnen Sie für dieses System die Intaktwahrscheinlichkeit.

2.18. [*Unabhängige/Disjunkte Ereignisse*] A und B seien Ereignisse mit $P(A) > 0$ und $P(B) > 0$. Welche der folgenden Aussagen sind (i) wahr, (ii) falsch, oder (iii) möglicherweise wahr?

(a) Wenn A und B disjunkt sind, dann sind sie unabhängig.

(b) Wenn A und B unabhängig sind, dann sind sie disjunkt.

(c) $P(A) = P(B) = 0.6$, und A und B sind disjunkt.

(d) $P(A) = P(B) = 0.6$, und A und B sind unabhängig.

2.19. [*Bernoulli Versuche*] Unabhängige Versuche, die mit Wahrscheinlichkeit p zu einem Erfolg und mit Wahrscheinlichkeit $1-p$ zu einem Mißerfolg führen, nennt man *Bernouilli Versuche*. Sei P_n die Wahrscheinlichkeit, daß es bei n derartigen Versuchen eine gerade Anzahl von Erfolgen gibt (0 sei eine gerade Anzahl). Zeigen Sie:

$$P_n = p(1 - P_{n-1}) + (1 - p)P_{n-1}, \quad n \geq 1 \quad (P_0 := 1)$$

Verwenden Sie diese Gleichung um zu zeigen (Induktion), daß:

$$P_n = \frac{1 + (1 - 2p)^n}{2}$$

2.20. [*Unabhängige Versuche*] Ein Versuch habe m mögliche Ausgänge, wobei Ausgang i mit Wahrscheinlichkeit p_i vorkommt, $i = 1, 2, \ldots, m$, $\sum_{i=1}^{m} p_i = 1$. Wenn zwei unabhängige Versuche beobachtet werden, mit welcher Wahrscheinlichkeit ist der Ausgang des zweiten Versuchs größer als der des ersten?

2.21. [*Unabhängige Versuche*] Betrachten Sie die Situation von Aufgabe 2.20. Einfachheitshalber werde hier angenommen, daß $p_1 = p_2 = \cdots = p_m = 1/m$.

(a) Wenn m unabhängige Versuche durchgeführt werden, wie groß ist die Wahrscheinlichkeit, daß jeder mögliche Ausgang vorkommt?

(b) E_i sei das Ereignis, daß unter den ersten n Versuchen der Ausgang i nicht vorkommt. Bestimmen Sie $P(E_i)$.

(c) Bestimmen Sie $P\big(\bigcup_{i=1}^{m} E_i\big)$. (*Hinweis*: Verwenden Sie das Additionstheorem.)

(d) Wenn $m = 10$, wieviele Versuche sind nötig, sodaß mit einer Wahrscheinlichkeit von mindestens 0.9 alle möglichen Ausgänge vorkommen?

2.22. [*k-aus-n–Systeme*] Ein aus n Komponenten bestehendes System ist ein *k-aus-n–System* $(1 \leq k \leq n)$, wenn das System genau dann funktioniert, wenn zumindest k der n Komponenten funktionieren. Man nehme an, daß die Komponenten unabhängig voneinander funktionieren/ausfallen.

(a) Wenn die Komponenten mit Wahrscheinlichkeit p_i funktionieren, mit welcher Wahrscheinlichkeit funktioniert ein 2-aus-4–System?

(b) Wiederholen Sie (a) für ein 3-aus-5–System.

(c) Wie lauten die Wahrscheinlichkeiten von (a) und (b), wenn alle Intaktwahrscheinlichkeiten p_i identisch gleich p sind?

2.23. [*Bedingte Wahrscheinlichkeit*] Betrachten Sie das 2-aus-4–Systeme von Aufgabe 2.22. Nehmen Sie an, daß jede Komponente mit der identischen Wahrscheinlichkeit von $p = 1/2$ intakt ist. Bestimmen Sie die bedingte Wahrscheinlichkeit, daß Komponente 1 funktioniert, wenn das System funktioniert.

2.24. [*Diskrete stochastische Größe*] Ein Würfelpaar, wobei einer der Würfel der erste und der andere der zweite ist, wird geworfen. Dieses Experiment wird durch den Merkmalraum $M = \{\omega = (\omega_1, \omega_2) : \omega_1 = 1, \dots, 6;\ \omega_2 = 1, \dots, 6\}$ beschrieben. Ereignisfeld? Wahrscheinlichkeiten $P(\{\omega\})$, $\omega \in M$? Eine stochastische Größe X sei gegeben durch $X(\omega) = \omega_1\omega_2$. Bestimmen Sie M_X und $P_X\{X = k\}$, $k \in M_X$.

2.25. [*Kontinuierliche stochastische Größe*] Ein Wahrscheinlichkeitsraum $(M, \mathcal{E}, W)$ sei gegeben durch $M = \{\omega : 0 < \omega < 10\}$ und für $E \in \mathcal{E}$ sei $P(E) = \int_E \frac{1}{10}\, dx$. Begründen Sie, warum $X(\omega) = \omega^2$ eine stochastische Größe ist. Bestimmen Sie M_X und $P_X\{X \le x\}$ für $x \in \mathbb{R}$.

Anhang 2

2.1 Potenzmenge: Als *Potenzmenge* $\mathcal{P}(M)$ einer Menge M bezeichnet man die Menge aller Teilmengen von M:

$$\mathcal{P}(M) = \{E : E \subseteq M\}$$

Bezeichnet $|A|$ die *Mächtigkeit* der Menge A, so gilt (für $M \neq \emptyset$):

$$|\mathcal{P}(M)| = 2^{|M|} > |M|$$

Dies ist trivial für endliche Mengen, aber nicht für unendliche Mengen. Die *allgemeine Kontinuumshypothese* besagt, daß $|\mathcal{P}(M)|$ die nach $|M|$ nächstgrößere Mächtigkeit ist. Speziell heißt dies, daß es keine Menge gibt, deren Mächtigkeit zwischen der Mächtigkeit der natürlichen Zahlen $\mathbb{N}$ und der Mächtigkeit der rellen Zahlen $\mathbb{R}$ liegt. Anders ausgedrückt:

$$|\mathcal{P}(\mathbb{N})| = 2^{|\mathbb{N}|} = |\mathbb{R}|$$

Bem: Ein zentrales Resultat der Mengentheorie besagt, daß die Kontininuumshypothese im Rahmen der Standardaxiome der Mengentheorie weder beweisbar noch widerlegbar ist, also von den Standardaxiomen unabhängig ist (P. Cohen, 1963).

2.2 Ereignisfelder: Ist der Merkmalraum M *endlich* oder *abzählbar*, nimmt man als Ereignisfeld stets die Potenzmenge $\mathcal{P}(M)$. Letztere erfüllt trivialerweise alle Eigenschaften eines Ereignisfeldes. Ist M *überabzählbar* (z.B.: $(0, 1)$, $\mathbb{R}^+$, $\dots$) wird die Situation komplizierter. Vereinfacht ausgedrückt, die Potenzmenge umfaßt im überabzählbaren Fall „zu viele“ Elemente (vgl. Anhang 2.1) als daß auf $\mathcal{P}(M)$ widerspruchsfrei eine Wahrscheinlichkeitsverteilung definiert werden könnte. Aus diesem Grund beschränkt man sich auf eine (echte) Teilmenge von $\mathcal{P}(M)$, auf die *Borelmengen* (vgl. Aufgabe 2.4).

Bem: Für den/die Praktiker/in sind diese Überlegungen nur von untergeordneter Bedeutung. Alle praktisch interessanten Ereignisse (z.B. Intervalle aller Art) sind Borelmengen.

2.3 Arithmetische Reihe: Für $a_k = a_{k-1} + d$, $k = 2, 3, \ldots$, $d \neq 0$, gilt:

$$\sum_{k=1}^{n} a_k = \frac{n}{2}\,(a_1 + a_n)$$

2.4 Geometrische Reihe: Für $a_k = a_1\, q^{k-1}$, $k = 2, 3, \ldots$, $q \neq 0$, gilt:

$$\sum_{k=1}^{n} a_k = a_1\,\frac{1 - q^n}{1 - q}$$

$$|q| < 1: \quad \sum_{k=1}^{\infty} a_k = \frac{a_1}{1 - q}$$

2.5 Funktionen: Eine *Funktion* f ordnet jedem Element x einer *Definitionsmenge* D genau ein Element y einer *Zielmenge* Z zu:

$$f : x \in D \;\mapsto\; y \in Z$$

Das *Bild* (auch *Wertebereich*) einer Funktion ist die Menge der Bilder aller Elemente der Definitionsmenge D:

$$f(D) := \{f(x) : x \in D\} \subseteq Z$$

Das *Urbild* einer Teilmenge T der Zielmenge ist die Menge aller Elemente des Definitionsbereichs, deren Bild Element dieser Teilmenge ist:

$$f^{-1}(T) := \{x \in D : f(x) \in T\}$$

Eine Funktion f heißt *umkehrbar*, wenn es zu jedem y aus dem Bild von f genau ein $x \in D$ mit $y = f(x)$ gibt. Durch $f^{-1}(y) = x$ wird die *Umkehrfunktion* definiert.

Bem: Man verwechsle die Umkehrfunktion $f^{-1}(y)$ nicht mit $\big[f(x)\big]^{-1} = 1/f(x)$.

Bsp: Das Bild der Funktion $f(x) = x^2$ mit $D = \mathbb{R}$ ist $f(D) = \mathbb{R}_0^+$. Das Urbild von beispielsweise $T = (0, 2)$ ist $f^{-1}(T) = (-\sqrt{2}, \sqrt{2}) \cap \{0\}^c$. Auf dem eingeschränkten Definitionsbereich $D_1 = \mathbb{R}_0^+$ ist die Funktion umkehrbar: $f^{-1}(y) = \sqrt{y}$, $y \in \mathbb{R}_0^+$.

Lösungen/Bemerkungen zu ausgewählten Aufgaben

2.2. Die Vereinigung von beliebigen Ereignissen A_i, $i = 1, 2, \ldots$, läßt sich wie folgt als Vereinigung von paarweise disjunkten Ereignissen darstellen:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \cap A_1^c) \cup (A_3 \cap A_1^c \cap A_2^c) \cup \cdots$$

2.4. Daß z.B. die offenen Intervalle (a, b) $(a < b)$ Borelmengen sind, sieht man wie folgt. Für $a < b$ kann (a, b) als Vereinigung von halboffenen Intervallen dargestellt werden (dabei gelte $1/n_0 < b - a$):

$$(a, b) = \bigcup_{n=n_0}^{\infty} \underbrace{\left(a, b - \frac{1}{n}\right]}_{\in \mathcal{B}}$$

Nach Definition ist die abzählbare Vereinigung von Elementen aus einem Ereignisfeld aber wieder Element des Ereignisfeldes. Dies zeigt die Behauptung.

<u>Bem:</u> Die Behauptung gilt auch für offene Intervalle der Form $(-\infty, b)$ und (a, ∞).)

2.7. Ist C das fragliche Ereignis und E_1 das Ereignis, daß beim ersten Experiment A eintritt, E_2 das Ereignis, daß beim ersten Experiment B eintritt und E_3 das Ereignis, daß beim ersten Experiment weder A noch B eintritt, so gilt:

$$P(C) = P(C|E_1)P(E_1) + P(C|E_2)P(E_2) + P(C|E_3)P(E_3)$$

Offensichtlich gilt $P(C|E_1) = 1$ und $P(C|E_2) = 0$. Eine kleine Überlegung zeigt, daß $P(C|E_3) = P(C)$.

2.11. Ist R_i das Ereignis, daß sich das Flugzeug in Region i befindet, und E das Ereignis, daß die Suche in Region 1 erfolglos ist, so gilt:

$$P(R_1|E) = \frac{P(E|R_1)P(R_1)}{\displaystyle\sum_{i=1}^{3} P(E|R_i)P(R_i)} = \frac{(\beta_1)(\frac{1}{3})}{(\beta_1)(\frac{1}{3}) + (1)(\frac{1}{3}) + (1)(\frac{1}{3})} = \cdots$$

2.13. Die Gleichung (c) läßt sich wie folgt interpretieren:

$$\underbrace{\frac{P(H|E)}{P(G|E)}}_{\text{a-post. odds ratio}} = \underbrace{\frac{P(H)}{P(G)}}_{\text{a-priori odds ratio}} \times \underbrace{\frac{P(E|H)}{P(E|G)}}_{\text{Likelihood ratio}}$$

2.16. Ein passender Merkmalraum besteht beispielsweise aus allen 6–Tupeln $(x_1, \ldots, x_6)$, wobei x_i den Status (1 = intakt, 0 = defekt) der i–te Komponente angibt.

$$M = \big\{(x_1, \ldots, x_6) : x_i = 0 \text{ oder } 1\big\}, \qquad |M| = 2^6 = 64$$

Die Wahrscheinlichkeit eines Tupels ist gegeben durch (p_i = Intaktwahrscheinlichkeit der i–ten Komponente):

$$P(\{(x_1, \ldots, x_6)\}) = \prod_{i=1}^{6} p_i^{x_i}(1-p_i)^{1-x_i}$$

2.17. Wird eine Komponente mit Defektwahrscheinlichkeit q parallelisiert, verringert sich die Defektwahrscheinlichkeit auf q^2.

2.19. Bedingen Sie durch den Ausgang des 1. Versuches und verwenden Sie den *Satz von der vollständigen Wahrscheinlichkeit*.

2.20. Ist A (B) das Ereignis, daß der Ausgang des ersten (zweiten) Versuchs größer als der Ausgang des zweiten (ersten) Versuchs ist, und E das Ereignis, daß beide Ausgänge gleich sind, so gilt $P(A) + P(B) + P(E) = 1$.

2.21. Die Wahrscheinlichkeit, daß unter den ersten n Versuchen k bestimmte Ausgänge nicht vorkommen, ist gegeben durch:

$$\left(\frac{m-k}{m}\right)^n$$

3 Eindimensionale Verteilungen

3.1. [*Verteilungsfunktion*] Die Verteilungsfunktion einer stochastischen Größe X sei gegeben durch:

$$F(x) = \begin{cases} 0 & \text{für} \quad x < 0 \\ \dfrac{x}{2} & \text{für} \quad 0 \leq x < 1 \\ \dfrac{2}{3} & \text{für} \quad 1 \leq x < 2 \\ \dfrac{11}{12} & \text{für} \quad 2 \leq x < 3 \\ 1 & \text{für} \quad 3 \leq x \end{cases}$$

(a) Stellen Sie die Funktion graphisch dar.

(b) Überzeugen Sie sich davon, daß es sich um eine Verteilungsfunktion handelt. (Welche Eigenschaften müssen erfüllt sein?)

(c) Wie lautet der Merkmalraum M_X ?

(d) Ist die Verteilung diskret, stetig oder gemischt?

(e) Bestimmen Sie die folgenden Wahrscheinlichkeiten: (a) $P\{X < 3\}$, (b) $P\{X = 1\}$, (c) $P\{X > 1/2\}$, (d) $P\{2 < X \leq 4\}$.

3.2. [*Erwartungswert*] Eine Übung wird in vier Gruppen mit 20, 25, 35 bzw. 40 Student/inn/en abgehalten. Wenn von den insgesamt 120 Personen, die an der Übung teilnehmen, eine Person zufällig ausgewählt wird und X die Größe der Gruppe ist, aus der die Person stammt, berechnen Sie $\mathbb{E}(X)$. Geben Sie eine anschauliche Erklärung dafür, warum $\mathbb{E}(X)$ größer als die durchschnittliche Gruppengröße $(20 + 25 + 35 + 40)/4 = 30$ ist.

3.3. [*Geometrische Verteilung*] Ein Behälter enthalte N weiße und M schwarze Kugeln. Die Kugeln werden eine nach der anderen zufällig *mit* Zurücklegen solange gezogen, bis man eine schwarze Kugel bekommt. Mit welcher Wahrscheinlichkeit werden dazu:

(a) genau n Ziehungen benötigt?

(b) mindestens k Ziehungen benötigt?

3.4. [*Geometrische Verteilung*] Betrachten Sie einen Produktionsprozeß, bei dem jede Stunde zufällig 20 Elemente zur Prüfung entnommen werden. X sei die Zahl der Elemente, die sich dabei als defekt herausstellen. Man nehme an, daß die Elemente bezüglich dieser Eigenschaft (intakt/defekt) unabhängig sind.

(a) Wenn der Defektanteil der Produktion 1% beträgt, mit welcher Wahrscheinlichkeit ist die Stichprobe von Stunde 10 die erste Stichprobe mit $X > 1$?

(b) Wie (a), aber mit Defektanteil 4%.

(c) Wenn der Defektanteil 4% beträgt, wieviele Stichproben muß man im Mittel entnehmen, bis erstmals $X > 1$?

3.5. [*Geometrische Verteilung*] Berechnen Sie den Erwartungswert $\mathbb{E}(X)$ einer geometrisch verteilten stochastischen Größe $X \sim G_p$. (*Hinweis*: Schreiben Sie $\mathbb{E}(X) = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = \sum_{i=1}^{\infty}(i-1+1)(1-p)^{i-1}p$.)

3.6. [*Geometrische Verteilung*] Berechnen Sie die Varianz $\mathbb{E}(X)$ einer geometrisch verteilten stochastischen Größe $X \sim G_p$. (*Hinweis*: Schreiben Sie $\mathbb{E}(X^2) = \sum_{i=1}^{\infty} i^2(1-p)^{i-1}p = \sum_{i=1}^{\infty}(i-1+1)^2(1-p)^{i-1}p$.)

3.7. [*Negative Binomialverteilung*] Wenn unabhängige Versuche, wobei jeder mit Wahrscheinlichkeit p ein Erfolg und mit Wahrscheinlichkeit $1-p$ ein Mißerfolg ist, durchgeführt werden, mit welcher Wahrscheinlichkeit kommen r Erfolge vor m Mißerfolgen? (*Hinweis*: Man überlege sich zuerst, daß das fragliche Ereignis genau dann eintritt, wenn der r–te Erfolg spätestens beim $(r+m-1)$–ten Versuch auftritt.)

3.8. [*Poissonverteilung*] Anfragen erreichen einen Server gemäß einer Poissonverteilung mit einem Mittelwert von 10 pro Stunde. Bestimmen Sie die Länge eines Zeitintervalls (in Sekunden), sodaß mit einer Wahrscheinlichkeit von 0.90 während dieses Intervalls keine Anfrage eintrifft.

3.9. [*Poissonverteilung*] Angenommen, bei der Herstellung von optischen Speichermedien (CDs) treten Verunreinigungen durch Staubteilchen gemäß einer Poissonverteilung mit einem Mittelwert von 0.0002 Teilchen pro cm^2 auf. Die CDs haben eine Fläche von 100 cm^2.

(a) Wenn 50 CDs untersucht werden, wie groß ist die Wahrscheinlichkeit, daß keine Teilchen entdeckt werden?

(b) Wieviele CDs müssen im Mittel untersucht werden, bevor ein Teilchen entdeckt wird?

(c) Wenn 50 CDs untersucht werden, wie groß ist die Wahrscheinlichkeit, daß es darunter höchstens 2 CDs mit einem oder mehr Teilchen gibt?

3.10. [*Poissonverteilung*] Derzeit gibt es in Ö etwa 35000 Eheschließungen im Jahr. Berechnen Sie approximativ die Wahrscheinlichkeit dafür, daß bei zumindest einem der Paare:

(a) beide Partner am 30. April geboren sind.

(b) beide Partner am selben Tag geboren sind.

Welche Voraussetzungen liegen den Berechnungen zugrunde?

3.11. [*Binomialverteilung*] Eine Kommunikationssystem bestehe aus n Komponenten, wobei jede Komponente unabhängig mit Wahrscheinlichkeit p funktioniert. Das System funktioniert nur, wenn zumindest die Hälfte der Komponenten funktioniert. Für welche Werte von p ist ein 5–Komponentensystem einem 3–Komponentensystem vorzuziehen? (*Hinweis*: Die Lösung führt auf eine Gleichung 3. Grades. Falls Sie diese Gleichung nicht explizit lösen können, lösen Sie sie numerisch unter Verwendung der R–Funktion `polyroot`.)

3.12. [*Binomialverteilung*] Zwei Freunde A und B werfen je zehn Freiwürfe mit einem Basketball. A ist bei jedem Wurf mit Wahrscheinlichkeit 0.80 erfolgreich, B mit Wahrscheinlichkeit 0.85. Mit welcher Wahrscheinlichkeit gewinnt (a) A, (b) B, (c) keiner von beiden? Welche (Unabhängigkeits–) Voraussetzungen liegen den Berechnungen zugrunde?

3.13. [*Poissonverteilung*] Bestimmen Sie (a) den Mittelwert und (b) die Varianz einer poissonverteilten stochastischen Größe $X \sim P_\mu$. (*Hinweis*: Berechnen Sie für (b) zuerst $\mathbb{E}\big[X(X-1)\big]$ und verwenden Sie dann den Verschiebungssatz.)

3.14. [*Diskrete Gleichverteilung*] Bestimmen Sie für eine allgemeine diskrete Gleichverteilung auf $M_X = \{x_1, x_2, \ldots, x_m\}$ ($x_i \neq x_j$ für $i \neq j$) (a) den Mittelwert und (b) die Varianz. Speziell für $M_X = \{1, 2, \ldots, 10\}$.

3.15. [*Approximationen*] In der VO werden Bedingungen angegeben, unter denen die Approximation der hypergeometrische Verteilung $H_{N,A,n}$ durch die Binomialverteilung $B_{n,A/N}$ bzw. die Approximation der Binomialverteilung $B_{n,p}$ durch die Poissonverteilung P_{np} zulässig ist. Diskutieren Sie die Bedingungen und geben Sie Beispiele für Situationen mit grob verletzten, nur leicht verletzten bzw. erfüllten Bedingungen. Verwenden Sie dazu die (eigenen) Funktionen `hyper.binom` und `binom.pois`.

3.16. [*Poissonverteilung*] Die Zahl der Erkältungen, die sich eine Person pro Jahr zuzieht, sei eine poissonverteilte stochastische Größe mit Parameter $\mu = 5$. Ein neues Wundermittel kommt auf den Markt, das den Poissonparameter für 75% der Bevölkerung auf $\mu = 3$ senkt, auf die anderen 25% hat das Mittel aber keine erkennbaren Auswirkungen. Wenn nun eine Person dieses Mittel ein Jahr lang ausprobiert und in dieser Zeit 2 Erkältungen hat, wie wahrscheinlich ist es, daß das Mittel einen positiven Effekt auf diese Person hat?

3.17. [*Hypergeometrische Verteilung*] Aus einer Gruppe bestehend aus 6 Männern und 9 Frauen soll ein Gremium aus 5 Personen gebildet werden. Das Gremium werde ganz zufällig gebildet und X sei die Zahl der Männer im Gremium. Wie ist X verteilt? Bestimmen Sie den Erwartungswert und die Varianz von X.

Zusatzfrage1: Wie ließe sich die zufällige Zusammenstellung des Gremiums mit Hilfe von R praktisch realisieren? (*Hinweis*: `sample`.)

Zusatzfrage2: Angenommen, im Gremium gibt es 4 Männer. Erfolgte die Auswahl rein zufällig? (*Hinweis*: Wie groß ist die Wahrscheinlichkeit, daß bei zufälliger Auswahl $X \geq 4$?)

3.18. [*Hypergeometrische/Binomial/Poissonverteilung*] Ein Produkt werde in Losen der Größe $N = 500$ geliefert. Zum Zwecke der Qualitätsprüfung werden dem Los willkürlich $n = 50$ Elemente ohne Zurücklegen entnommen und geprüft. Gibt es unter den geprüften Elementen mehr als ein defektes Element, wird das Los zurückgewiesen. Angenommen, das Los enthält (i) 0.8%, (ii) 9% defekte Elemente. Mit welcher Wahrscheinlichkeit wird das Los zurückgewiesen? Rechnen Sie mit (a) der (exakten) hypergeometrischen Verteilung, (b) einer passenden Binomialapproximation und (c) einer passenden Poissonapproximation. (Sind die Approximationen hier zulässig?)

Zusatz: Der Ausschußanteil betrage allgemein $100p\%$. Bestimmen Sie unter Verwendung aller drei Verteilungen die Wahrscheinlichkeit mit der das Los angenommen wird und stellen Sie die Wahrscheinlichkeiten als Funktion von p graphisch dar.

R: Die graphische Darstellung von mehreren Kurven auf einmal läßt sich mittels `matplot` (= Matrixplot) einfach realisieren.

3.19. [*Stetige Verteilung*] Die Dichte einer sG X sei gegeben durch:

$$f(x) = \begin{cases} a + bx^2 & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

(a) Wenn $\mathbb{E}(X) = 3/5$, bestimmen Sie a und b und stellen Sie f graphisch dar.

(b) Bestimmen Sie die Verteilungsfunktion F und stellen Sie sie graphisch dar.

(c) Bestimmen Sie die Varianz von X.

3.20. [*Logistische Verteilung*] Betrachten Sie die folgende Funktion:

$$F(x) = \frac{e^x}{1+e^x}, \quad x \in \mathbb{R}$$

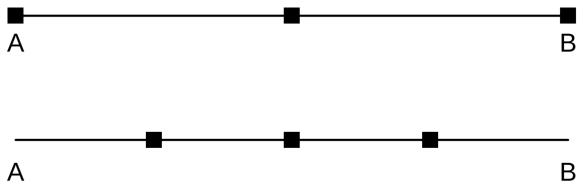
(a) Stellen Sie die Funktion graphisch dar.

(b) Zeigen Sie, daß F die Verteilungsfunktion einer (stetigen) sG X ist.

(c) Ermitteln Sie allgemein einen Ausdruck für das p–Quantile x_p und bestimmen Sie konkret die drei Quartile (25%, 50%, 75%) der Verteilung.

(d) Bestimmen Sie die zugehörige Dichte f und stellen Sie sie graphisch dar.

3.21. [*Uniforme Verteilung*] Ein Linienbus verkehrt zwischen zwei 100 km voneinander entfernten Orten A und B. Für die Reperatur von technischen Defekten gibt es in A, in B und in der Mitte zwischen A und B eine Werkstatt. Ein/e Mitarbeiter/in des Busunternehmens macht den Vorschlag, daß es effizienter wäre, die drei Werkstätten nach 25, 50 und 75 km einzurichten. Wie beurteilen Sie diesen Vorschlag? (*Hinweis*: Gehen Sie für die Stelle eines möglichen Defekts von einer uniformen Verteilung aus. Verteilung des Abstands zur nächsten Werkstätte? Mittlerer Abstand?)



3.22. [*Exponentialverteilung*] Die Kilometerleistung einer Autobatterie sei exponentialverteilt mit $\tau = 10000$ km.

(a) Mit welcher Wahrscheinlichkeit läßt sich eine 5000 km lange Reise ohne Ersetzung der Batterie absolvieren?

(b) Wie lang darf eine Reise höchstens sein, daß sie mit 90% Wahrscheinlichkeit ohne Ersetzung der Batterie beendet werden kann?

(c) Bestimmen Sie den Median, den Mittelwert und die Streuung der Kilometerleistung der Batterie.

3.23. [*Exponentialverteilung*] Zeigen Sie für eine exponentialverteilte sG $X \sim Ex_\tau$:

$$P\{X > x + y | X > x\} = P\{X > y\}, \quad x, y > 0$$

Was bedeutet diese Eigenschaft („Gedächtnislosigkeit“) in Worten? (*Hinweis*: X sei z.B. die Lebensdauer einer elektronischen Komponente, etwa eines Transistors.)

3.24. [*Poisson–/Exponentialverteilung*] Die Anzahl N_t von bestimmten Ereignissen (z.B. Telefonanrufe, Aufträge an einen Netzwerkdrucker, etc.) im Zeitintervall $(0, t]$ sei eine nach $P_{\lambda t}$ verteilte sG und T sei die Zeitspanne bis zum Auftreten des ersten Ereignisses. Bestimmen Sie die Verteilung von T. (*Hinweis*: Bestimmen Sie zunächst $P\{T > x\}$.)

3.25. [*Normalverteilung*] Die Lebensdauer eines Bildschirms sei eine normalverteilte sG mit Mittelwert $\mu = 8.2$ Jahre und Streuung $\sigma = 1.4$ Jahre.

(a) Welcher Anteil solcher Bildschirme funktioniert länger als 10 Jahre, nicht länger als 5 Jahre, zwischen 5 und 10 Jahren?

(b) Bestimmen Sie das 10% und das 90% Quantile der Lebensdauer. Wie sind diese Werte zu interpretieren?

(c) Sie kaufen einen 3 Jahre alten gebrauchten Bildschirm. Mit welcher Wahrscheinlichkeit funktioniert er noch länger als 5 Jahre?

3.26. [*Normalverteilung*] Angenommen, die Wegzeit von zu Hause zur TU ist normalverteilt mit Mittelwert 40 Minuten und Standardabweichung 7 Minuten. Wenn Sie um 13 Uhr eine Prüfung haben und mit Wahrscheinlichkeit 0.95 nicht zu spät kommen möchten, wann spätestens müssen Sie aufbrechen?

3.27. [*Logarithmische Normalverteilung*] Angenommen, die Lebensdauer eines Halbleiterlasers hat eine logarithmische Normalverteilung mit $\mu = 10$ Stunden und $\sigma = 1.5$ Stunden.

(a) Stellen Sie die Verteilungsfunktion und die Dichte der Lebensdauer graphisch dar.

(b) Mit welcher Wahrscheinlichkeit arbeitet der Laser länger als 10000 Stunden?

(c) Welche Lebensdauer wird von 99% der Laser überschritten?

(d) Bestimmen Sie den Median, den Mittelwert und die Streuung der Lebensdauer. (*Hinweis*: Vgl. Anhang A.2.4.)

3.28. [*Funktion einer sG*] Die sG X habe eine Verteilung mit Dichte $f(x) = e^{-x}$, $x > 0$. Bestimmen Sie unter Verwendung des Transformationssatzes die Dichte (+ Zeichnung) von:

$$\text{(a) } Y = X^2 \qquad \text{(b) } Y = \sqrt{X} \qquad \text{(c) } Y = \ln X \qquad \text{(d) } Y = e^{-X}$$

3.29. [*Satz vom unbewußten Statistiker*] Der Radius X eines Kreises sei eine sG mit der Dichte $f(x) = e^{-x} I_{(0,\infty)}(x)$.

(a) Zeigen Sie, daß $\mathbb{E}(X^k) = k!$ für $k \in \mathbb{N}$.

(b) Berechnen Sie den Erwartungswert der Kreisfläche $A = X^2\pi$.

(c) Berechnen Sie die Varianz der Kreisfläche.

3.30. [*Funktion einer sG*] Die sG X habe eine Standardnormalverteilung $N(0, 1)$.

(a) Bestimmen Sie $\mathbb{E}(|X|)$. (*Hinweis*: Satz vom unbewußten Statistiker.)

(b) Bestimmen Sie die Verteilungsfunktion und (durch Ableiten) die Dichte von $Y = |X|$.

(*Bem*: Die Verteilung von $|X|$ heißt auch *Halbnormalverteilung*. Warum?)

3.31. [*Gemischte Verteilung*] Die Grünphase (einschließlich Blinkphase) bei einer Fußgängerampel betrage 25 Sekunden, die Rotphase 65 Sekunden. Sie kommen zu einem zufälligen Zeitpunkt zu dieser Ampel und X sei die Wartezeit. Bestimmen Sie:

(a) die Verteilungsfunktion von X (+ Zeichnung).

(b) die Wahrscheinlichkeit, mit der Sie länger als 20 Sekunden warten.

(c) die (bedingte) Wahrscheinlichkeit, mit der Sie noch mindestens weitere 20 Sekunden warten, wenn Sie bereits 20 Sekunden gewartet haben.

(d) den Median, den Mittelwert und die Streuung von X.

3.32. [*Gemischte Verteilung*] Angenommen, Sie haben wiederholt mit einer Servicestelle zu tun, bei der die Wartezeit eine exponentialverteilte sG X mit Mittelwert 20 Minuten ist. Sie warten allerdings nicht unbegrenzt, nie länger als 30 Minuten.

(a) Bestimmen und zeichnen Sie die Verteilungsfunktion der Wartezeit.

(b) Wie lange warten Sie im Mittel?

3.33. [*Zufallszahlen/Inversionsmethode*] Entwickeln Sie – ausgehend von auf $(0, 1)$ uniform verteilten Zufallszahlen – einen Algorithmus zur Simulation von Beobachtungen der sG X von Aufgabe 3.19. Schreiben Sie dazu eine R–Funktion. Generieren Sie mit Hilfe dieser Funktion 500 Beobachtungen von X und stellen Sie das Ergebnis in Form eines (Dichte–) Histogramms graphisch dar.

R: Bei Verwendung der Inversionsmethode ist eine Gleichung 3. Grades zu lösen. Verwenden Sie dazu die Funktion `uniroot`.

3.34. [*Zufallszahlen/Inversionsmethode*] Entwickeln Sie – ausgehend von auf $(0, 1)$ uniform verteilten Zufallszahlen – einen Algorithmus zur Simulation von Beobachtungen der sG X von Aufgabe 3.20. Schreiben Sie dazu eine R–Funktion. Generieren Sie mit Hilfe dieser Funktion 500 Beobachtungen von X und stellen Sie das Ergebnis in Form eines (Dichte–) Histogramms graphisch dar.

3.35. [*Zufallszahlen/Inversionsmethode*] Entwickeln Sie – ausgehend von auf $(0, 1)$ uniform verteilten Zufallszahlen – einen Algorithmus zur Simulation der Wartezeiten bei der Ampel von Aufgabe 3.31. Schreiben Sie dazu eine R–Funktion. (*Bem*: Die Funktion sollte Wartezeiten für eine beliebige Ampel mit r = Länge der Rotphase und g = Länge der Grünphase generieren.)

Anhang 3

3.1 Erwartungswert: Ist X eine diskrete sG auf $M_X = \mathbb{N}_0$, so kann $\mathbb{E}(X)$ auch wie folgt berechnet werden:

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} P\{X > i\}$$

Bsp: Für eine geometrisch verteilte sG X gilt $P\{X > i\} = (1 - p)^i$; daher:

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

3.2 Erwartungswert: Ist X eine stetige sG auf $M_X = \mathbb{R}^+$, so kann $\mathbb{E}(X)$ auch wie folgt berechnet werden:

$$\mathbb{E}(X) = \int_0^\infty P\{X > x\}\,dx = \int_0^\infty \big[1 - F_X(x)\big]\,dx$$

Bsp: Für eine exponentiell verteilte sG X gilt $F_X(x) = 1 - e^{-x/\tau}$; daher:

$$\mathbb{E}(X) = \int_0^\infty e^{-x/\tau}\,dx = -\tau e^{-x/\tau}\Big|_0^\infty = \tau$$

3.3 Erwartungswert: Nach Definition existiert allgemein der Erwartungswert von $Y = g(X)$ nur dann, wenn $\int_{-\infty}^{\infty} |g(x)| f_X(x)\,dx < \infty$. Dies ist nicht immer der Fall. Ein typisches Beispiel ist die *Cauchy–Verteilung* mit Dichte $f_X(x) = 1/[\pi(1+x^2)]$ für $-\infty < x < \infty$:

$$\int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} = \frac{2}{\pi}\int_0^\infty \frac{x}{1+x^2}\,dx = \frac{1}{\pi}\int_1^\infty \frac{1}{u}\,du = \frac{1}{\pi}\ln(u)\Big|_1^\infty = \infty$$

(Dabei wurde die Substitution $u = 1 + x^2$ verwendet.) D.h., $\mathbb{E}(X)$ existiert nicht. (Daraus folgt, daß auch alle höheren Momente – $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$, etc. – nicht existieren.) Man beachte allerdings, daß der Median existiert: $x_{0.5} = 0$.

3.4 Mittelwert/Median: Die beiden wichtigsten Lageparameter erfüllen eine Minimumseigenschaft:

$$\mathbb{E}(X) = \operatorname*{Argmin}_c \mathbb{E}\big[(X-c)^2\big], \qquad \mathrm{Med}(X) = \operatorname*{Argmin}_d \mathbb{E}\big[|X-d|\big]$$

3.5 Momentenerzeugende Funktion: Eine Funktion mit vielfältigen Anwendungen ist die *momentenerzeugende Funktion* einer sG X. Sie ist definiert durch:

$$m_X(t) = \mathbb{E}(e^{tX})$$

Sie existiert zwar immer für $t = 0$ ($m_X(0) = 1$), von Interesse ist sie aber nur, wenn sie für $t \in (-h, h)$, für ein $h > 0$, existiert. In diesem Fall ist die Verteilung von X eindeutig durch $m_X(t)$ festgelegt. Weiters gilt, daß die k–te Ableitung im Nullpunkt (sofern sie existiert) gleich dem k–ten Moment von X ist:

$$m_X^{(k)}(t)\Big|_{t=0} = \mathbb{E}(X^k)$$

Bsp: Vgl. Anhang A (Diskrete/Stetige Verteilungen) für Beispiele von momentenerzeugenden Funktionen. Beispielsweise ist die momentenerzeugende Funktion der Ex_τ–Verteilung gegeben durch:

$$m(t) = \frac{1}{1-\tau t}, \quad t < \frac{1}{\tau}$$

Für die 1.Ableitung gilt:

$$m'(t) = \frac{\tau}{(1-\tau t)^2} \quad \Longrightarrow \quad \mathbb{E}(X) = m'(0) = \tau$$

3.6 Wichtige Integrationsregeln:

$$\text{Partielle Integration:} \quad \int_a^b f'(x)g(x)\,dx = f(x)g(x)\Big|_a^b - \int_a^b f(x)g'(x)\,dx$$

$$\text{Substitution:} \quad \int_a^b f\big(g(x)\big)g'(x)\,dx = \int_{g(a)}^{g(b)} f(y)\,dy \quad (y = g(x))$$

3.7 Gammafunktion: Die Gammafunktion ist eine Erweiterung der Fakultätsfunktion auf nicht-ganzzahlige Argumente:

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt, \quad x > 0$$

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma(x+1) = x\Gamma(x) \quad \text{(Fakultätseigenschaft)}$$

3.8 Exponentialfunktion:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!},\ x \in \mathbb{R}, \qquad \lim_{x\to\infty}\left(1 + \frac{z}{x}\right)^x = e^z,\ z \in \mathbb{R}$$

Lösungen/Bemerkungen zu ausgewählten Aufgaben

3.3. Man benötigt mindestens k Ziehungen, wenn die ersten $k-1$ Ziehungen alle weiß sind.

3.4. X ist binomialverteilt mit $n = 20$ und $p =$ Defektanteil.

3.6. Mit $q = 1 - p$ gilt:

$$
\begin{aligned}
\mathbb{E}(X^2) &= \sum_{i=1}^{\infty} i^2 q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i-1+1)^2 q^{i-1} p \\
&= \sum_{i=1}^{\infty} (i-1)^2 q^{i-1} p + \sum_{i=1}^{\infty} 2(i-1) q^{i-1} p + \sum_{i=1}^{\infty} q^{i-1} p \\
&= \sum_{j=0}^{\infty} j^2 q^j p + 2 \sum_{j=0}^{\infty} j q^j p + 1 \\
&= q\mathbb{E}(X^2) + 2q\mathbb{E}(X) + 1
\end{aligned}
$$

3.7 Ist X die Nummer des Versuchs mit dem r–ten Erfolg, so gilt:

$$
P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \ldots
$$

3.10. Vorausgesetzt wird, daß (für jede Person) jeder Tag des Jahres mit gleicher Wahrscheinlichkeit als Geburtstag in Frage kommt, und daß die Geburtstage (bei Hochzeitspaaren) unabhängig sind.

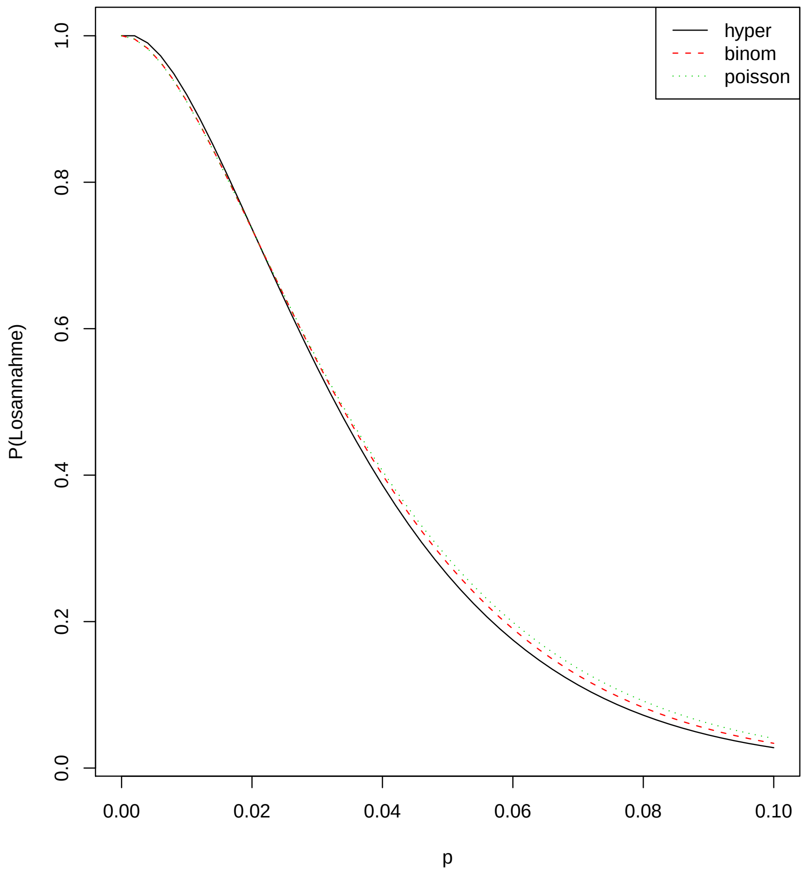
3.13. $\mathbb{E}(X^2) = \mathbb{E}\big[X(X-1)\big] + \mathbb{E}(X)$

3.15. Vgl. für die Approximation von diskreten Verteilungen auch die Abbildung in Anhang A (Diskrete Verteilungen).

3.16. Ist E das Ereignis, daß die Person zu der Bevölkerungsgruppe gehört, für die das Mittel eine positive Wirkung hat, und ist X die Zahl der Erkältungen pro Jahr (und Person), so ist $P\{E|X = 2\}$ zu berechnen. Nehmen Sie dazu die Bayes'sche Formel.

3.17. Zusatzfrage2: Die Wahrscheinlichkeit $P\{X \geq 4\}$ (berechnet unter der Annahme einer zufälligen Auswahl des Gremiums) nennt man den p–*Wert* der (statistischen) *Hypothese* „Zufällige Auswahl des Gremiums“. Kleine p–Werte sprechen gegen die Gültigkeit der Hypothese. (Warum?)

3.18. Abbildung für die Zusatzfrage:



3.19. Verwenden Sie für die Varianzberechnung den *Verschiebungssatz*:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

$\mathbb{E}(X^2)$ berechnet man wie folgt:

$$\mathbb{E}(X^2) = \int_0^1 x^2(a + bx^2)\, dx$$

3.23. Diese Eigenschaft charakterisiert die Exponentialverteilung: Sei X eine sG auf $\mathbb{R}^+$ mit dieser Eigenschaft. Dann gilt mit $G(x) := P\{X > x\}$:

$$G(x + y) = G(x)G(y), \quad x, y \in \mathbb{R}^+$$

Für $a \in \mathbb{N}$ folgt daraus:

$$G(a) = G\left(\sum_{i=1}^{a} 1\right) = G(1)^a$$

Weiter gilt:

$$G(1) = G\left(\sum_{i=1}^{b} \frac{1}{b}\right) = G\left(\frac{1}{b}\right)^{b} \quad \Longrightarrow \quad G\left(\frac{1}{b}\right) = G(1)^{1/b}$$

Für rationale Zahlen $q = a/b$ folgt daher:

$$G(q) = G\left(\frac{a}{b}\right) = G\left(\sum_{i=1}^{a} \frac{1}{b}\right) = G\left(\frac{1}{b}\right)^{a} = G(1)^{a/b} = G(1)^{q}$$

Jede reelle Zahl $x > 0$ kann von rechts durch rationale Zahlen $q_n > 0$ angenähert werden: $q_n \to x$. Wegen der Rechtsstetigkeit von $G(x) = 1 - F(x)$ folgt:

$$G(x) = \lim_{n\to\infty} G(q_n) = \lim_{n\to\infty} G(1)^{q_n} = G(1)^{x}$$

Setzt man $\tau := -1/\ln G(1)$ (d.h. $G(1) = e^{-1/\tau}$), so gilt:

$$G(x) = P\{X > x\} = e^{-x/\tau}$$

D.h., X ist exponentialverteilt:

$$F(x) = 1 - G(x) = 1 - e^{-x/\tau}, \quad x > 0$$

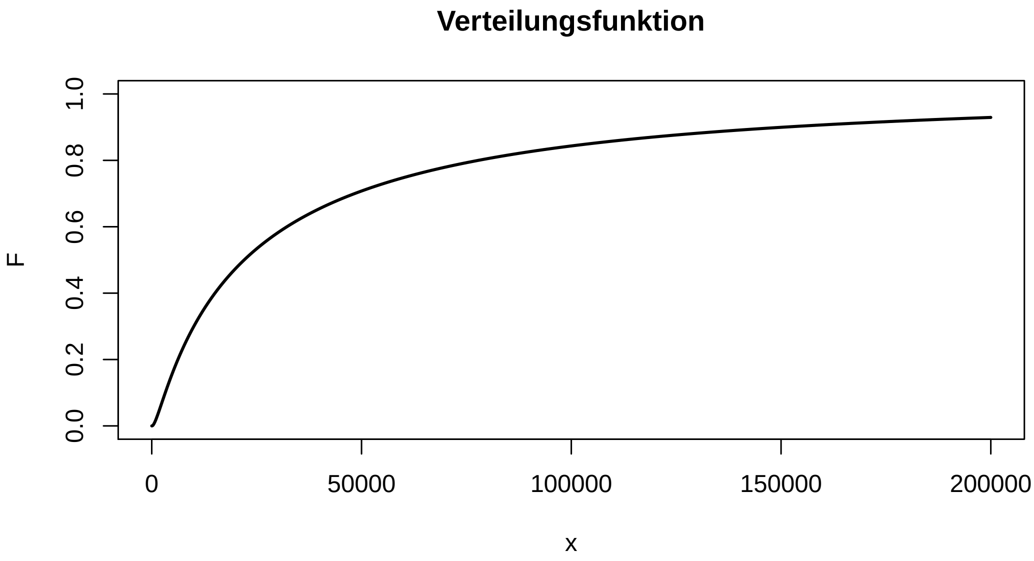
3.24. Das Ereignis $\{T > x\}$ ist äquivalent zu $\{N_x = 0\}$.

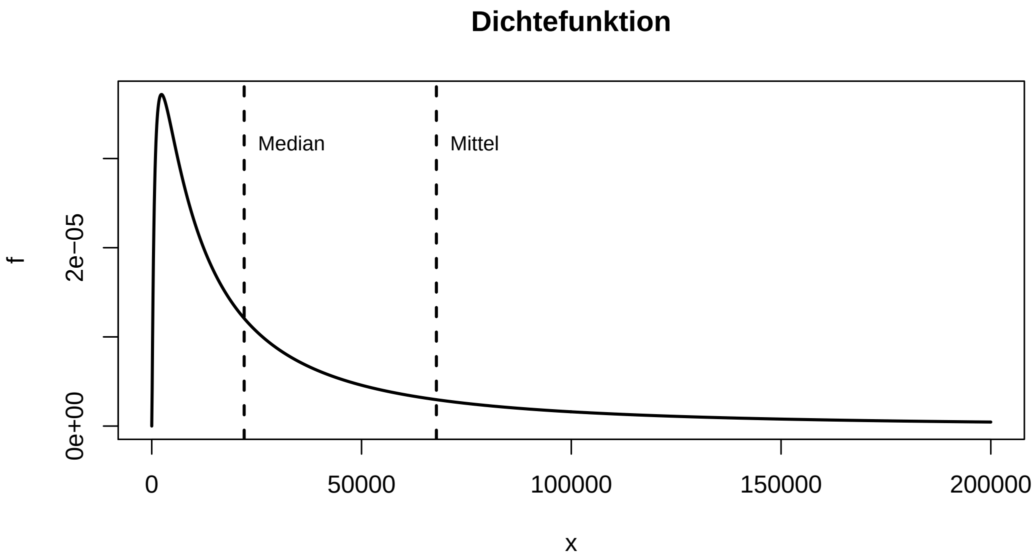
3.27. Die Verteilungsfunktion ist gegeben durch

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \quad x > 0$$

Die Dichte bekommt man durch Ableiten:

$$f(x) = F'(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0$$





3.28. Lösung für (a):

$$y = x^2,\ x = \sqrt{y},\ \frac{dx}{dy} = \frac{1}{2\sqrt{y}}\ :\quad f_Y(y) = \frac{1}{2\sqrt{y}}\, e^{-\sqrt{y}},\quad y > 0$$

3.29. Punkt (a) zeigt man mittels (wiederholter) partieller Integration oder – einfacher – mit Hilfe von Anhang 3.7.

3.30.

$$\int \frac{x}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = -\frac{2}{\sqrt{2\pi}}\, e^{-x^2/2} + C$$

3.32. Lösung für (b):

$$\begin{aligned}
\mathbb{E}(X) &= 30P\{X = 30\} + \int_0^{30} \frac{x}{20}\, e^{-x/20}\, dx \\
&= 30e^{-3/2} + \left[-xe^{-x/20} \Big|_0^{30} + \int_0^{30} e^{-x/20}\, dx \right] \\
&= 30e^{-3/2} + \left[-30e^{-3/2} + \left(-20e^{-x/20} \right) \Big|_0^{30} \right] \\
&= -20e^{-3/2} + 20 \\
&\doteq 15.54 \text{ [min]}
\end{aligned}$$

Bem: Nach Anhang 3.2 läßt sich der Mittelwert auch wie folgt berechnen:

$$\mathbb{E}(X) = \int_0^{30} e^{-x/20}\, dx = -20\, e^{-x/20} \Big|_0^{30} = -20(e^{-3/2} - 1) \doteq 15.54$$

3.33. Bei der *Inversionsmethode* zur Generierung von Zufallszahlen ist die Verteilungsfunktion zu intvertieren. Ist F_X die Verteilungsfunktion der sG X und sind u uniform $U(0, 1)$ verteilte Zufallszahlen (mit `runif` erzeugt), so bekommt man durch:

$$u \longmapsto F_X^{-1}(u)$$

simulierte Beobachtungen („Realisationen“) von X.

Bem: Dies gilt für stetige sGn, aber auch – nach entsprechender Erweiterung der Definition der Inversen einer im strengen Sinn nicht invertierbaren Verteilungsfunktion – für diskrete oder gemischte sGn. (Vgl. Aufgabe 3.35.)

4 Mehrdimensionale Verteilungen

4.1. [*Zweidimensionale Intervalle*] Der Merkmalraum des stochastischen Vektors $(X,Y)'$ sei $\mathbb{R}^2$. Betrachten Sie die folgenden Ereignisse und ihre Wahrscheinlichkeiten:

$$\begin{aligned}
A_1 &= \{(x,y) : x \le 2, y \le 4\}, \quad P(A_1) = 7/8 \\
A_2 &= \{(x,y) : x \le 2, y \le 1\}, \quad P(A_2) = 4/8 \\
A_3 &= \{(x,y) : x \le 0, y \le 4\}, \quad P(A_3) = 3/8 \\
A_4 &= \{(x,y) : x \le 0, y \le 1\}, \quad P(A_4) = 2/8
\end{aligned}$$

Bestimmen Sie die Wahrscheinlichkeit von $A_5 = \{(x,y) : 0 < x \le 2, 1 < y \le 4\}$.

4.2. [*Bivariate diskrete Verteilung*] Die gemeinsamen Punktwahrscheinlichkeiten von X und Y seien gegeben wie folgt:

$$\begin{aligned}
p(1,1) &= \frac{1}{8} \qquad p(1,2) = \frac{1}{4} \\
p(2,1) &= \frac{1}{8} \qquad p(2,2) = \frac{1}{2}
\end{aligned}$$

(a) Bestimmen Sie die Randverteilung von X und von Y.

(b) Bestimmen Sie die bedingte Verteilung von X gegeben $Y = i$, $i = 1, 2$.

(c) Sind X und Y unabhängig?

(d) Berechnen Sie $P\{XY \le 3\}$, $P\{X + Y > 2\}$, $P\{X/Y > 1\}$.

4.3. [*Bivariate diskrete Verteilung*] Drei Kugeln werden zufällig und ohne Zurücklegen aus einem Behälter bestehend aus 3 roten, 4 weißen und 5 blauen Kugel entnommen, und X bzw. Y sei die Zahl der roten bzw. weißen Kugeln in der Stichprobe.

(a) Bestimmen Sie die gemeinsame Verteilung von X und Y.

(b) Bestimmen Sie die Randverteilung von X und von Y.

(c) Sind X und Y unabhängig?

(d) Bestimmen Sie den Korrelationskoeffizienten von X und Y.

4.4. [*Multivariate diskrete Verteilung*] Die gemeinsamen Punktwahrscheinlichkeiten von X, Y und Z seien gegeben wie folgt:

$$p(1,2,3) = p(2,1,1) = p(2,2,1) = p(2,3,2) = \frac{1}{4}$$

Berechnen Sie (a) $\mathbb{E}(XYZ)$ und (b) $\mathbb{E}(XY + XZ + YZ)$.

4.5. [*Bivariate uniforme Verteilung*] Angenommen, C macht sich zwischen 8:00 und 8:30 auf den Weg ins Büro und benötigt dazu zwischen 40 und 50 Minuten. X sei der Zeitpunkt des Aufbruchs und Y die benötigte Zeitspanne. Wenn diese sGn unabhängig und uniform verteilt sind, bestimmen Sie die Wahrscheinlichkeit, daß C vor 9:00 im Büro eintrifft.

4.6. [*Bivariate stetige Verteilung*] Die gemeinsame Dichte von X und Y sei gegeben durch:

$$f(x,y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \quad 0 < x < 1,\ 0 < y < 2$$

(a) Bestätigen Sie, daß es sich um eine Dichtefunktion handelt.
(b) Bestimmen Sie die Randdichte von X.
(c) Bestimmen Sie die Randdichte von Y.
(d) Berechnen Sie $P\{X > Y\}$.
(e) Bestimmen Sie $\mathbb{E}(X)$.
(f) Bestimmen Sie $\mathbb{E}(Y)$.

4.7. [*Bivariate stetige Verteilung*] Die gemeinsame Dichte von (X,Y) sei $f(x,y) = C(x+y)$ für $(x,y) \in (0,1) \times (0,1)$ und $f(x,y) = 0$ sonst.

(a) Welchen Wert hat die Konstante C ?
(b) Bestimmen Sie die Randdichte von X und von Y.
(c) Bestimmen Sie die Kovarianz von X und Y.
(d) Bestimmen Sie den Korrelationskoeffizienten von X und Y.
(e) Sind X und Y unkorreliert/unabhängig?
(f) Bestimmen Sie die bedingte Dichte von $X|Y = y$ und von $Y|X = x$.
(g) Bestimmen Sie die Regressionsfunktion von X bezüglich Y und von Y bezüglich X, d.h., bestimmen Sie $\mathbb{E}(X|Y = y)$ und $\mathbb{E}(Y|X = x)$.

4.8. [*Bivariate uniforme Verteilung*] Ein Punkt (X,Y) wird zufällig in einem Kreis (in Nullpunktslage) mit Radius 1 gewählt.

(a) Wie lautet die gemeinsame Dichte von (X,Y) ?
(b) Bestimmen (und zeichnen) Sie die Randdichte von X und von Y.
(c) Sind X und Y unabhängig?
(d) Zeigen Sie, daß die Kovarianz (und daher auch der Korrelationskoeffizient) von X und Y gleich Null ist.
(e) $D = \sqrt{X^2 + Y^2}$ sei der Abstand des Zufallspunktes (X,Y) vom Nullpunkt $(0,0)$. Bestimmen Sie die Verteilungsfunktion/Dichte von D und berechnen Sie $\mathbb{E}(D)$. (*Hinweis*: Bestimmen Sie die Verteilungsfunktion mit Hilfe einer geometrischen Überlegung.)

4.9. [*Bivariate stetige Verteilung*] Die gemeinsame Dichte von X und Y sei gegeben durch:

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{sonst} \end{cases}$$

Bestimmen Sie die Dichte von $Z = X/Y$. (*Hinweis*: Bestimmen Sie zuerst die Verteilungsfunktion von Z.)

4.10. [*Bivariate stetige Verteilung*] Der Input eines Programms sei eine stochastische Größe X mit Dichte $f_X(x) = e^{-x} I_{(0,\infty)}(x)$ (Ex_1–Verteilung). Bedingt durch $X = x$ sei die Ausführungszeit des Programms eine exponentialverteilte sG mit Mittelwert $1/x$. Bestimmen Sie die Dichte der Ausführungszeit Y des Programms. (*Hinweis*: Bestimmen Sie zuerst die gemeinsame Dichte von (X,Y) und anschließend die Randdichte von Y.

4.11. [*Bivariate Normalverteilung*] Zeigen Sie, daß bei einem bivariat normalverteilten stochastischen Vektor (X, Y) die Unabhängigkeit von X und Y äquivalent zur Unkorreliertheit von X und Y ist.

4.12. [*Bivariate Normalverteilung*] X und Y seien bivariat normalverteilt mit den Parametern $\mu_x = 5$, $\mu_y = 10$, $\sigma_x^2 = 1$, $\sigma_y^2 = 25$ und $\rho > 0$. Wenn $P\{4 < Y < 16|X = 5\} = 0.954$, bestimmen Sie ρ.

4.13. [*Bivariate Normalverteilung*] In einem (amerikanischen) Lehrbuch findet sich die folgende Aufgabe: Angenommen, der Korrelationskoeffizient zwischen der Köpergröße des Mannes und der Frau von verheirateten Paaren beträgt 0.70, und die mittlere Köpergröße des Mannes beträgt 5 ft. 10 in. mit der Standardabweichung 2 in., und die mittlere Köpergröße der Frau beträgt 5 ft. 4 in. mit der Standardabweichung $1\frac{1}{2}$ in. Wenn man von einer bivariaten Normalverteilung ausgeht:

(a) Wie lautet die gemeinsame Verteilung der Körpergrößen in der Einheit cm? (*Hinweis*: 1 ft. = 12 in. = 30.48 cm, 1 in. = 2.54 cm)

(b) Was ist der beste Prognosewert für die Größe einer Frau, deren Mann 6 ft. groß ist? *Zusatz*: Bestimmen Sie für die Größe der Frau ein 95%–Prognoseintervall.

4.14. [*Multivariate Normalverteilung*] Der stochastische Vektor $\underline{X} = (X_1, X_2, X_3)'$ sei normalverteilt $N_3(\underline{0}, \Sigma)$, wobei:

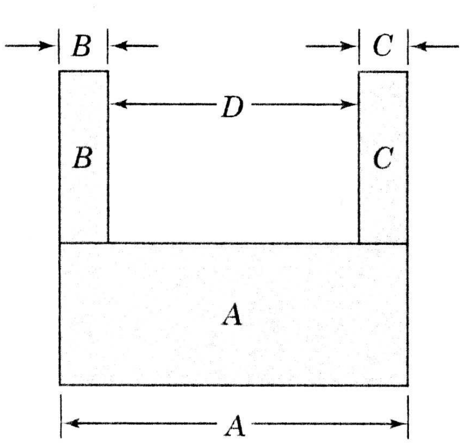
$$\Sigma = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

Bestimmen Sie die Verteilung von $Y = X_1 - 2X_2 + X_3$ und berechnen Sie $P\{Y^2 > 15.36\}$.

4.15. [*Erwartungswert einer Funktion von sGn*] Die Kantenlängen X, Y, Z eines Quaders seien unabhängige $U(0, 1)$ verteilte sGn. Bestimmen Sie den Erwartungswert und die Varianz des Volumens $V = XYZ$. (*Hinweis*: Verwenden Sie VO/Satz 18.7 und nehmen Sie für die Varianzberechnung den Verschiebungssatz.)

4.16. [*Erwartungswert einer Funktion von sGn*] X und Y seien unabhängige $N(0, 1)$–verteilte sGn. Wenn man (X, Y) als Zufallspunkt in der Ebene betrachtet, bestimmen Sie den mittleren Abstand des Punktes vom Nullpunkt, d.h. bestimmen Sie $\mathbb{E}\big(\sqrt{X^2 + Y^2}\big)$. (*Hinweis*: Transformieren Sie auf Polarkoordinaten; vgl. Anhang 4.5.)

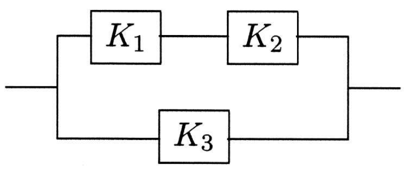
4.17. [*Additionstheorem/Normalverteilung*] Ein U–förmiges Werkstück bestehe aus den Teilen A, B und C. Für die angegebenen Dimensionen gelte: A ist normalverteilt mit Mittelwert 10 mm und Standardabweichung 0.1 mm, B und C sind normalverteilt mit Mittelwert 2 mm und Standardabweichung 0.05 mm. Alle Dimensionen seien unabhängig.



Bestimmen Sie die Verteilung des Abstands D. Mit welcher Wahrscheinlichkeit ist D kleiner als 5.9 mm ?

4.18. [*Minimum von sGn*] Ein Seriensystem bestehe aus drei Komponenten mit unabhängigen exponentialverteilten Lebensdauern mit den Mittelwerten 100, 200 bzw. 300 Stunden. Bestimmen Sie die Verteilungsfunktion und die Dichte der Lebensdauer des Systems sowie den Mittelwert und die Streuung. *Zusatz*: Betrachten Sie allgemein ein Seriensystem aus k unabhängigen Komponenten mit exponentialverteilten Lebensdauern mit den Mittelwerten τ_i, $i = 1, \dots, k$.

4.19. [*Minimum/Maximum von sGn*] Die logische Struktur eines Systems bestehend aus drei Komponenten sei gegeben wie folgt:



Die Lebensdauern der Komponenten seien unabhängig und identisch verteilt mit Dichte $f(x) = e^{-x}\, I_{(0,\infty)}(x)$. Bestimmen Sie die Verteilungsfunktion und die Dichte der Lebensdauer des Systems sowie den Mittelwert.

4.20. [*Minimum/Maximum von sGn*] Für die Komponenten des Systems von Aufgabe 2.16 gelte: Die Lebensdauern der Komponenten der ersten Parallegruppe sind exponentialverteilt mit Mittelwert 1000 Stunden, die der zweiten Parallelgruppe sind exponentialverteilt mit Mittelwert 3000 Stunden, und die Lebensdauer der letzten Serienkomponente ist exponentialverteilt mit Mittelwert 5000 Stunden. Alle Lebensdauern seien unabhängig.

(a) Bestimmen Sie einen Ausdruck für die Verteilungsfunktion der Systemlebensdauer.

(b) Simulieren Sie die Systemlebensdauer mehrere tausend Mal und stellen Sie das Ergebnis in Form eines (Dichte–) Histogramms dar.

R: *Eine* simulierte Lebensdauer für beispielsweise die erste Parallelgruppe läßt sich mittels `max(rexp(3, rate=1/1000))` erzeugen. Nehmen Sie eine `for` Schleife.

4.21. [*Faltung/diskret*] Drei Würfel werden geworfen. Die geworfenen Augenzahlen seien X_1, X_2 und X_3. Bestimmen Sie mit Hilfe der Faltungsformel die Verteilung (d.h. die Punktwahrscheinlichkeiten) von X_1+X_2 und $X_1+X_2+X_3$. Stellen Sie die Punktwahrscheinlichkeiten graphisch dar.

R: `convolve` (mit `type="open"`).

4.22. [*Faltung/diskret*] Ein fehlertolerantes System für die Ausführung von Finanztransaktionen besteht aus drei Computern. Fällt der arbeitende Computer aus, springt der zweite sofort ein. Fällt auch dieser aus, springt der dritte Computer unverzüglich ein. Man nehme an, daß die Wahrscheinlichkeit für einen Computerausfall während einer Transaktion 10^{-8} beträgt und daß die Transaktionen als unabhängige Ereignisse betrachtet werden können. Wieviele Transaktionen können im Mittel vor dem Ausfall aller Computer durchgeführt werden? Varianz?

4.23. [*Faltung/stetig*] Zwei Punkte X und Y werden zufällig auf einer Geraden der Länge L so gewählt, daß X in der ersten Hälfte und Y in der zweiten Hälfte der Geraden liegt. (Mit anderen Worten, X ist auf $(0, L/2)$ uniform verteilt, Y auf $(L/2, L)$.) Bestimmen Sie die Dichte des Abstands $Z = Y - X$ der beiden Punkte. Mit welcher Wahrscheinlichkeit ist dieser Abstand größer als $L/3$? (*Hinweis*: Schreiben Sie $Z = (-X) + Y$. Wie ist $X' = -X$ verteilt?)

4.24. [*Faltung/stetig*] Die Lebensdauer X einer Komponente folge einer Exponentialverteilung mit Mittelwert 5. Fällt die Komponente aus, wird sie sofort durch eine gleichartige Reservekomponente (Lebensdauer Y) ersetzt.

(a) Bestimmen (und zeichnen) Sie die Dichte der Gesamtlebensdauer, d.h. bestimmen Sie die Dichte von $S = X + Y$.

(b) Bestimmen Sie den Mittelwert und die Streuung von S.

(c) Wiederholen Sie (a) und (b) unter der Annahme, daß die Reservekomponente eine exponentialverteilte Lebensdauer mit Mittelwert 10 hat.

4.25. [*Additionstheorem/Exponentialverteilung*] An einem Schalter folgen die Servicezeiten einer Exponentialverteilung mit Mittelwert 10 Minuten. Wie ist Ihre Wartezeit verteilt, wenn bei Ihrem Eintreffen drei Personen vor dem Schalter warten und eine Person bedient wird? Mittelwert? Streuung? (*Hinweis*: Nützen Sie die Gedächtnislosigkeit der Exponentialverteilung; vgl. Aufgabe 3.23.)

4.26. [*Streudiagramm/Bivariate Normalverteilung*] Mit Hilfe der (eigenen) Funktion `biv.rnorm` lassen sich bivariat normalverteilte Beobachtungen simulieren. Erzeugen Sie mit dieser Funktion $n = 500$ Beobachtungen einer (a) $N(0, 0, 1, 1, 0)$, (b) $N(100, 200, 25, 36, 0.8)$ und einer (c) $N(100, 200, 25, 36, -0.6)$ Verteilung und und stellen Sie die Beobachtungen mittels *Streudiagramm* (oder *Scatterplot*) graphisch dar. Verwenden Sie auch die Funktionen `scatter.with.hist` und `scatter.with.box`.

R: Beim Scatterplot interpretiert man die zusammengehörigen Datenwerte als (kartesische) Koordinaten von Punkten in der Ebene. Sind `x` und `y` die Datenvektoren, so kann man den Scatterplot etwa wie folgt zeichnen:

```
plot(x, y, type="p", pch=19, xlab="Data x", ylab="Data y")
```

4.27. [*Mehrdimensionale Daten*] Der Datensatz `normtemp` (Package: `UsingR`), umfaßt Angaben zu Körpertemperatur (`temperature` [°F]) und Ruhepuls (`hr` [Schläge pro Minute]) von 130 (gesunden) Personen beiderlei Geschlechts (`gender`).

R: Das Package `UsingR` muß zuerst installiert und anschließend geladen werden. Unter Windows läßt sich dies u.a. über das Menü 'Pakete' erledigen. Eine andere Möglichkeit geht über die Funktion `install.packages`.

(a) Rechnen Sie zunächst die Einheit °F in die Einheit °C um. (*Bem*: $x\,$°F $= 5(x-32)/9\,$°C)

R: Nützen Sie dazu (z.B.) die Möglichkeiten der Funktion `within`.

(b) Bestimmen Sie für die beiden numerischen Variablen Lage– und Streuungsparameter. Unterscheiden Sie dabei nach Geschlecht.

R: Eine Möglichkeit ist `by(normtemp, gender, summary)`.

(c) Bereiten Sie den Datensatz graphisch auf (Histogramme, Boxplots, ...). Unterscheiden Sie dabei nach Geschlecht.

(d) Stellen Sie die Variablen `temperature` und `hr` in Form eines Streudiagramms (Scatterplots) graphisch dar. Unterscheiden Sie dabei nach Geschlecht (durch unterschiedliche Farben, Symbole, ...). Gibt es einen Zusammenhang? Wenn ja, welcher Art?

Anhang 4

4.1 Multinomialverteilung: Ein Experiment bestehe aus einer Folge von n identischen und unabhängigen Versuchen, wobei jeder Versuch mit den (konstanten) Wahrscheinlichkeiten $p_1, \ldots, p_k$ ($\sum_{i=1}^k p_i = 1$) auf k Arten ausgehen kann. Ist X_i, $i = 1, \ldots, k$, die Zahl der Versuche, die auf die i–te Art ausgehen, so gilt:

$$p(x_1, \ldots, x_k) = P\{X_1 = x_1, \ldots, X_k = x_k\} = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad \sum_{i=1}^k x_i = n$$

Man beachte, daß es sich – entgegen der Schreibweise – nur um eine $(k-1)$–dimensionale Verteilung handelt. Die X_i sind binomialverteilt, $X_i \sim B_{n,p_i}$, $i = 1, \ldots, k$, aber nicht unabhängig.

Eigenschaften:

$$\begin{aligned}
\mathbb{E}(X_i) &= np_i \\
\mathrm{Var}(X_i) &= np_i(1-p_i) \\
\mathrm{Cov}(X_i, X_j) &= -np_ip_j, \quad i \neq j \\
\rho_{X_i,X_j} &= -\sqrt{\frac{p_i}{1-p_i}}\sqrt{\frac{p_j}{1-p_j}}, \quad i \neq j
\end{aligned}$$

Spezialfälle: Für $k = 2$ ergibt sich die *Binomialverteilung*, für $k = 3$ spricht man von einer *Trinomialverteilung*.

Bsp: Erfolgen die Ziehungen in Aufgabe 4.3 *mit* Zurücklegen, so sind X, Y und $Z = 3 - X - Y$ gemeinsam trinomialverteilt, wobei:

$$n = 3,\ p_1 = \frac{3}{12},\ p_2 = \frac{4}{12},\ p_3 = \frac{5}{12}$$

Der Korrelationskoeffizient von X und Y ist gegeben durch:

$$\rho_{X,Y} = -\sqrt{\frac{3/12}{1-3/12}}\sqrt{\frac{4/12}{1-4/12}} = -\frac{1}{\sqrt{6}}$$

(Etwas überraschend ergibt sich dieselbe Korrelation wie bei Ziehungen *ohne* Zurücklegen. Vgl. dazu den folgenden Punkt.)

4.2 Polyhypergeometrische Verteilung: Die Verallgemeinerung der Situation von Aufgabe 4.3 führt zur *polyhypergeometrischen Verteilung*. Unter N (gleichartigen) Objekten gebe es A_i Objekte der i–ten Art, $i = 1, \ldots, k$, $\sum_{i=1}^k A_i = N$. Werden zufällig n Objekte *ohne* Zurücklegen gezogen und ist X_i, $i = 1, \ldots, k$, die Zahl der dabei erhaltenen Objekte der i–ten Art, so gilt:

$$p(x_1, \ldots, x_k) = P\{X_1 = x_1, \ldots, X_k = x_k\} = \frac{\binom{A_1}{x_1}\binom{A_2}{x_2}\cdots\binom{A_k}{x_k}}{\binom{N}{n}}, \quad \sum_{i=1}^k x_i = n$$

Man beachte, daß es sich – entgegen der Schreibweise – nur um eine $(k-1)$–dimensionale Verteilung handelt. Die X_i sind hypergeometrisch verteilt, $X_i \sim H_{N,A_i,n}$, $i = 1, \ldots, k$, aber nicht unabhängig.

Eigenschaften:

$$\begin{aligned}
\mathbb{E}(X_i) &= n\,\frac{A_i}{N} \\
\mathrm{Var}(X_i) &= n\,\frac{A_i}{N}\left(1 - \frac{A_i}{N}\right)\frac{N-n}{N-1} \\
\mathrm{Cov}(X_i, X_j) &= -n\,\frac{A_i}{N}\,\frac{A_j}{N}\,\frac{N-n}{N-1}, \quad i \neq j \\
\rho_{X_i,X_j} &= -\sqrt{\frac{A_i}{N-A_i}}\sqrt{\frac{A_j}{N-A_j}}, \quad i \neq j
\end{aligned}$$

Bem: Der *Korrekturfaktor* (für endliche Grundgesamtheiten) $(N-n)/(N-1)$ kürzt sich im Ausdruck für ρ heraus.

Bsp: In Aufgabe 4.3 sind X, Y und $Z = 3 - X - Y$ gemeinsam polyhypergeometrisch verteilt, wobei:

$$N = 12,\ A_1 = 3,\ A_2 = 4,\ A_3 = 5,\ n = 3$$

Der Korrelationskoeffizient von X und Y ist also gegeben durch:

$$\rho_{X,Y} = -\sqrt{\frac{3}{12-3}}\sqrt{\frac{4}{12-4}} = -\frac{1}{\sqrt{6}}$$

4.3 Kovarianzmatrix: Die Kovarianzmatrix Σ eines stochastischen Vektors $\underline{X} = (X_1, \ldots, X_n)$:

$$\Sigma = \mathrm{VCov}(\underline{X}) = \big[\sigma_{ij}\big]_{i,j=1,\ldots,n} = \big[\mathrm{Cov}(X_i, X_j)\big]_{i,j=1,\ldots,n}$$

ist *symmetrisch* und *positiv definit*, d.h.:

$$\Sigma^T = \Sigma, \quad \underline{x}^T \Sigma\, \underline{x} > 0 \quad \text{für alle} \quad \underline{x} \neq \underline{0},\ \underline{x} \in \mathbb{R}^n$$

4.4 Zweidimensionale Integrale: Ist $f : D \longrightarrow \mathbb{R}$ eine (stückweise) stetige Funktion und A ein flächenhafter Integrationsbereich, dann existiert das Integral:

$$I = \int\limits_A f(x,y)\, dA = \int\limits_{x_1}^{x_2} \int\limits_{y_1(x)}^{y_2(x)} f(x,y)\, dy dx = \int\limits_{y_1}^{y_2} \int\limits_{x_1(y)}^{x_2(y)} f(x,y)\, dx dy$$

Die schrittweise Integration erfolgt von innen nach außen. Die Reihenfolge der Variablen, nach denen integriert wird, darf vertauscht werden. (*Bem*: Analoges gilt für mehrfache Bereichsintegrale.)

4.5 Variablentransformation bei Doppelintegralen: Ist A der Integrationsbereich in der (x,y)–Ebene und B der Bereich in der (u,v)–Ebene, so gilt:

$$\int\limits_A f(x,y)\, dx\, dy = \int\limits_B f\big(x(u,v), y(u,v)\big) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du\, dv$$

Dabei ist:

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x(u,v)}{\partial u} & \dfrac{\partial x(u,v)}{\partial v} \\[2ex] \dfrac{\partial y(u,v)}{\partial u} & \dfrac{\partial y(u,v)}{\partial v} \end{vmatrix}$$

die *Funktionaldetermimante* (oder *Jacobi–Determinante*).

Transformation in Polarkoordinaten: Ein wichtiger Spezialfall ist die Transformation in Polarkoordinaten:

$$x = r\cos(\theta),\ \ y = r\sin(\theta), \quad 0 \leq r < \infty,\ 0 \leq \theta < 2\pi$$

Funktionaldeterminante:

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} = r$$

(Letzteres gilt wegen $\sin^2(\theta) + \cos^2(\theta) = 1$.) Somit gilt:

$$\int\limits_A f(x,y)\, dx\, dy = \int\limits_{\theta_1}^{\theta_2} \int\limits_{r_1(\theta)}^{r_2(\theta)} f(r,\theta) r\, dr\, d\theta$$

Lösungen/Bemerkungen zu ausgewählten Aufgaben

4.3. Vgl. für die gemeinsame Verteilung von X, Y und $Z = 3 - X - Y$ Anhang 4.2. Aus der Art der Stichprobenentnahme folgt, daß X und Y nicht unabhängig sind; überprüfen Sie dies aber auch formal. Verwenden Sie für die Kovarianzberechnung den *Verschiebungssatz*:

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

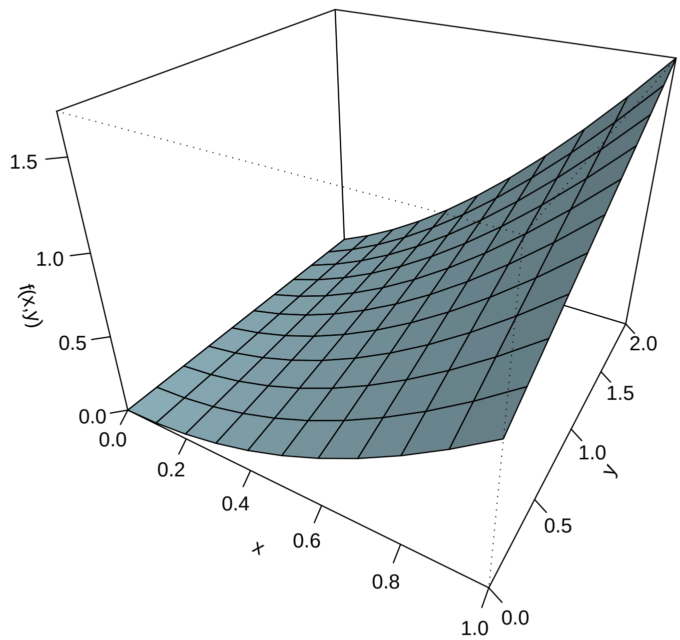
$\mathbb{E}(XY)$ berechnet man wie folgt:

$$\mathbb{E}(XY) = \sum_{x,y} xy\, p(x, y)$$

4.5. Rechnet man in Minuten, so gilt $X \sim U(0, 30)$ und $Y \sim U(40, 50)$. Zu berechnen ist $P\{X + Y \leq 60\}$. Wenn $X < 10$, trifft die Person sicher vor $9:00$ ein, wenn $X > 20$, sicher nicht. Somit:

$$P\{X + Y \leq 60\} = P\{X \leq 10\} + \int_{10}^{20} \int_{40}^{60-x} \frac{1}{300}\, dy\, dx$$

4.6. Die folgende Abbildung zeigt die gemeinsame Dichte $f(x, y)$:



4.7. Verwenden Sie für die Kovarianzberechnung den *Verschiebungssatz*:

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

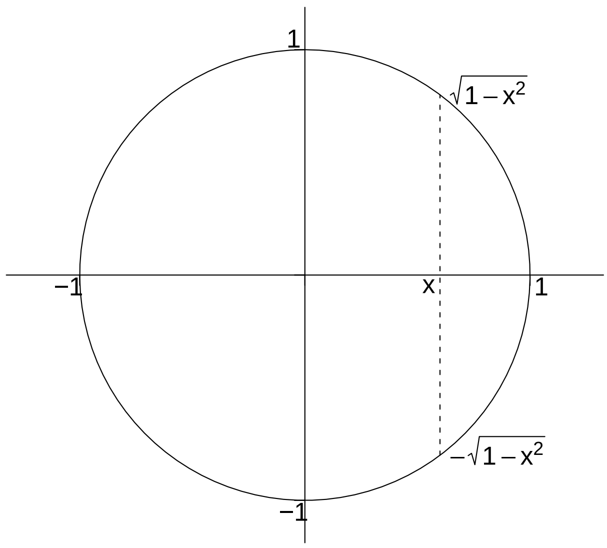
$\mathbb{E}(XY)$ berechnet man wie folgt:

$$\mathbb{E}(XY) = C \int_0^1 \int_0^1 xy\,(x+y)\,dx\,dy$$

4.8. Die Dichte einer uniformen Verteilung ist auf ihrem *Träger* (= Bereich auf dem die Dichte größer als Null ist) konstant. Somit gilt:

$$f(x, y) = \begin{cases} \dfrac{1}{\pi} & x^2 + y^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

Um beispielsweise die Randdichte von X zu bestimmen, ist nach y zu integrieren. Vgl. für die geometrischen Verhältnisse die folgende Abbildung:



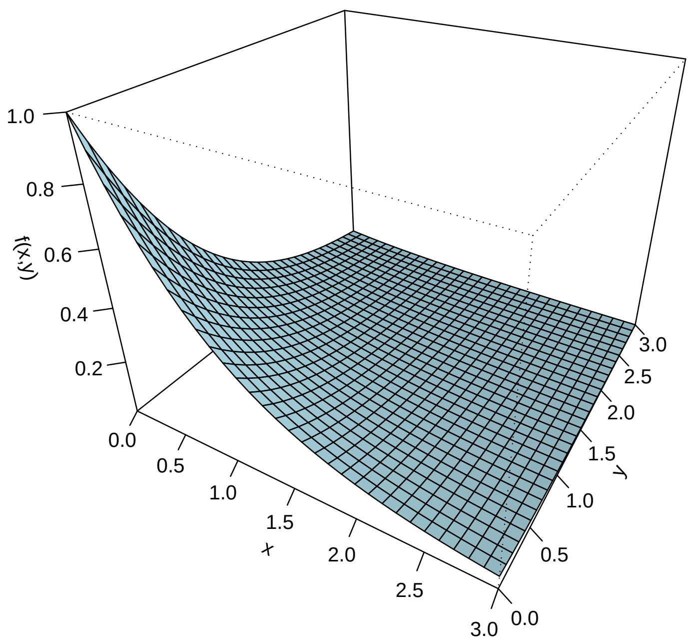
Berechnung von $\mathbb{E}(D)$ mittels SvuStat:

$$\mathbb{E}(D) = \iint\limits_{x^2+y^2 \le 1} \sqrt{x^2 + y^2}\,\frac{1}{\pi}\,dx\,dy$$

Transformation auf Polarkoordinaten (vgl. Anhang 4.5):

$$\mathbb{E}(D) = \iint\limits_{x^2+y^2 \le 1} \sqrt{x^2 + y^2}\,\frac{1}{\pi}\,dx\,dy = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 r^2\,dr\,d\theta = \frac{1}{3\pi}\int_0^{2\pi} d\theta = \frac{2}{3}$$

4.9. Die folgende Abbildung zeigt die gemeinsame Dichte $f(x, y)$:



Für die Verteilungsfunktion von $Z = X/Y$ gilt:

$$F_Z(z) = P\left\{\frac{X}{Y} \leq z\right\} = \iint\limits_{x/y \leq z} e^{-(x+y)}\, dx\, dy$$

4.10. Bedingte Dichte von $Y|X = x$:

$$f(y|x) = xe^{-xy}, \quad y > 0$$

Gemeinsame Dichte von (X, Y):

$$f(x, y) = f(y|x)\, f_X(x) = xe^{-x(y+1)}, \quad x, y > 0$$

4.11. Zeigen Sie, daß für $\rho = 0$ die Dichte der bivariaten Normalverteilung in einen nur von x und in einen nur von y abhängigen Faktor zerfällt.

4.13. Sind X, Y die Maße in in., und X', Y' die Maße in cm, so gilt:

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{bmatrix} 2.54 & 0 \\ 0 & 2.54 \end{bmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

4.14. Y ist eine Linearkombination der X_i, $i = 1,2,3$:

$$Y = (1, -2, 1)\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

4.15. Nach VO/Satz 18.7 gilt:

$$\mathbb{E}(V) = \mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z), \quad \mathbb{E}(V^2) = \mathbb{E}(X^2)\mathbb{E}(Y^2)\mathbb{E}(Z^2)$$

4.16. Satz vom unbewußten Statistiker:

$$\mathbb{E}(R) = \mathbb{E}\big(\sqrt{X^2+Y^2}\big) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sqrt{x^2+y^2}\, e^{-(x^2+y^2)/2}\, dx\, dy$$

Transformation auf Polarkoodinaten (vgl. Anhang 4.5):

$$\mathbb{E}(R) = \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} r\, e^{-r^2/2}\, r\, d\theta\, dr = \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} r^2\, e^{-r^2/2}\, d\theta\, dr = \int_0^{\infty} r^2\, e^{-r^2/2}\, dr$$

Letzteres Integral läßt sich auf die Varianz einer $N(0,1)$–Verteilung zurückführen:

$$\mathbb{E}(R) = \int_0^{\infty} r^2\, e^{-r^2/2}\, dr = \frac{\sqrt{2\pi}}{2}\underbrace{\int_{-\infty}^{\infty} r^2\, \frac{1}{\sqrt{2\pi}}\, e^{-r^2/2}\, dr}_{1} = \frac{\sqrt{2\pi}}{2}$$

<u>Bem:</u> Als Nebenprodukt der obigen Überlegungen entnimmt man, daß die Dichte von R gegeben ist durch (*Rayleighverteilung*):

$$f(r) = r\, e^{-r^2/2}, \quad r > 0$$

4.18. Allgemein gilt:

$$F_{\min}(x) = 1 - \prod_{i=1}^{k}\big[1 - F_i(x)\big] = 1 - e^{-x\sum_{i=1}^{k} 1/\tau_i}, \quad x > 0$$

Mit (harmonisches Mittel der τ_i):

$$\tau_h = \frac{k}{\displaystyle\sum_{i=1}^{k}\frac{1}{\tau_i}}$$

läßt sich die Verteilungsfunktion/Dichte wie folgt schreiben:

$$F_{\min}(x) = 1 - e^{-kx/\tau_h}, \quad x > 0$$
$$f_{\min}(x) = F'_{\min}(x) = \frac{k}{\tau_h}\, e^{-kx/\tau_h}, \quad x > 0$$

Dies entspricht einer Exponentialverteilung mit Mittelwert/Streuung τ_h/k.

Speziell für $\tau_1 = \ldots = \tau_k = \tau$ ist $\tau_h = \tau$ und es gilt:

$$F_{\min}(x) = 1 - e^{-kx/\tau}, \quad x > 0$$
$$f_{\min}(x) = \frac{k}{\tau}\, e^{-kx/\tau}, \quad x > 0$$
$$\mathbb{E}(X_{\min}) = \sqrt{\mathrm{Var}(X_{\min})} = \frac{\tau}{k}$$

4.19. Für die Verteilungsfunktion der Lebensdauer X des Systems gilt:

$$F_X(x) = P\{X \le x\} = P\Big\{\max\big\{\min\{X_1, X_2\}, X_3\big\} \le x\Big\}$$

4.22. Die Zahl der Transaktionen bis zum Ausfall des 1. Computers ist geometrisch verteilt mit $p = 10^{-8}$; ebenso für den 2. und 3. Computer.

4.23. Orientieren Sie sich am entsprechenden Beispiel der VO (Faltung zweier unabhängiger $U(0,1)$–Verteilungen).

4.24. Lösung für (c):

$$\begin{aligned}
f_S(s) &= \int_0^s \frac{1}{5}\, e^{-(s-y)/5}\, \frac{1}{10}\, e^{-y/10}\, dy \\
&= \frac{1}{50}\, e^{-s/5} \int_0^s e^{y(1/5-1/10)}\, dy \\
&= \frac{1}{50}\, e^{-s/5} \int_0^s e^{y/10}\, dy \\
&= \frac{1}{50}\, e^{-s/5} \big[10\, e^{y/10}\big]\Big|_0^s \\
&= \frac{1}{5}\, e^{-s/5}\big(e^{s/10} - 1\big) \\
&= \frac{1}{5}\, e^{-s/10} - \frac{1}{5}\, e^{-s/5}, \quad s > 0
\end{aligned}$$

4.25. Die Wartezeit X folgt einer Erlangverteilung $Er_{4,10}$. (Beachten Sie, daß es als Folge der Gedächtnislosigkeit der Exponentialverteilung keine Rolle spielt, wie lange die Person am Schalter bereits bedient wird, wenn Sie eintreffen. Die restliche Servicezeit dieser Person ist unverändert nach Ex_{10} verteilt.)

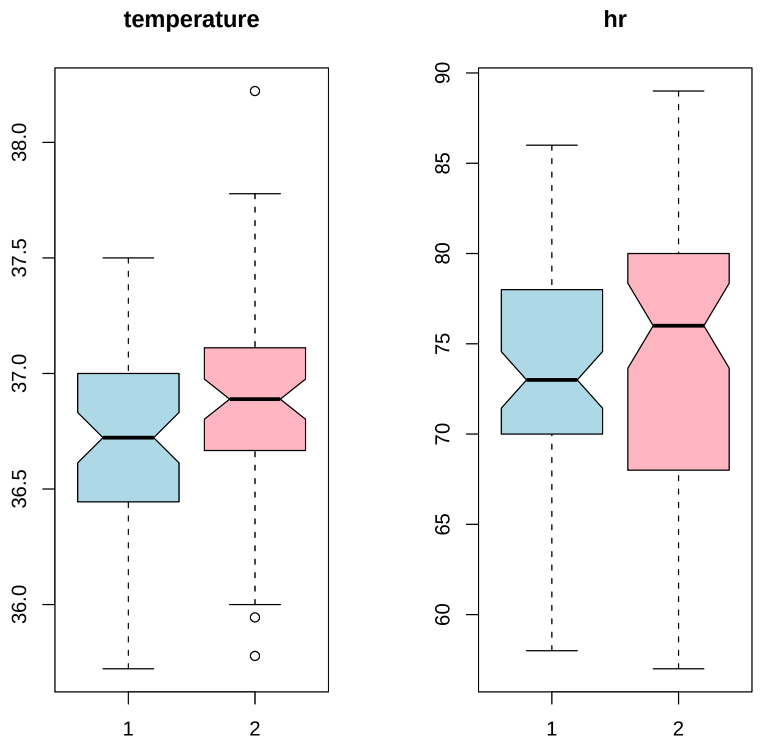
4.27. (a) Laden und umrechnen:

```
require(UsingR)
data(normtemp)
normtemp2 <- within(normtemp, {
  temperature <- 5*(temperature - 32)/9 })

head(normtemp2)
  temperature gender hr
1    35.72222      1 70
2    35.94444      1 71
3    36.05556      1 74
4    36.11111      1 80
5    36.16667      1 73
6    36.16667      1 75
```

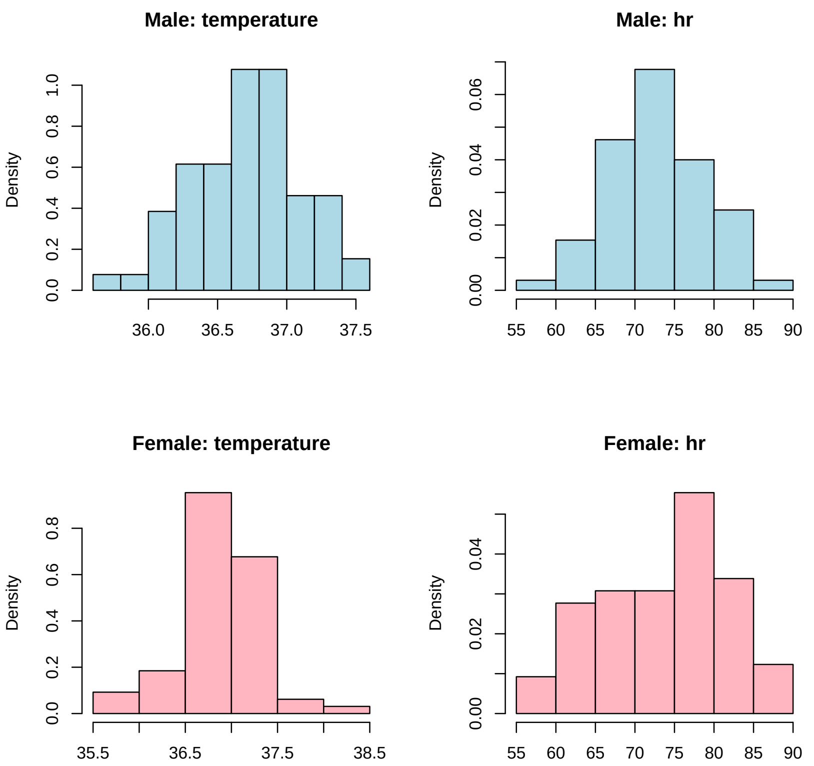
(b) Lage-/Streuungsparameter:

```
summary(normtemp2)
  temperature        gender           hr
 Min.   :35.72   Min.   :1.0   Min.   :57.00
 1st Qu.:36.56   1st Qu.:1.0   1st Qu.:69.00
 Median :36.83   Median :1.5   Median :74.00
 Mean   :36.81   Mean   :1.5   Mean   :73.76
 3rd Qu.:37.06   3rd Qu.:2.0   3rd Qu.:79.00
 Max.   :38.22   Max.   :2.0   Max.   :89.00


attach(normtemp2)
by(normtemp2, gender, summary)
gender: 1
  temperature        gender         hr
 Min.   :35.72   Min.   :1   Min.   :58.00
 1st Qu.:36.44   1st Qu.:1   1st Qu.:70.00
 Median :36.72   Median :1   Median :73.00
 Mean   :36.72   Mean   :1   Mean   :73.37
 3rd Qu.:37.00   3rd Qu.:1   3rd Qu.:78.00
 Max.   :37.50   Max.   :1   Max.   :86.00
---------------------------------------------------------
gender: 2
  temperature        gender         hr
 Min.   :35.78   Min.   :2   Min.   :57.00
 1st Qu.:36.67   1st Qu.:2   1st Qu.:68.00
 Median :36.89   Median :2   Median :76.00
 Mean   :36.89   Mean   :2   Mean   :74.15
 3rd Qu.:37.11   3rd Qu.:2   3rd Qu.:80.00
 Max.   :38.22   Max.   :2   Max.   :89.00
```

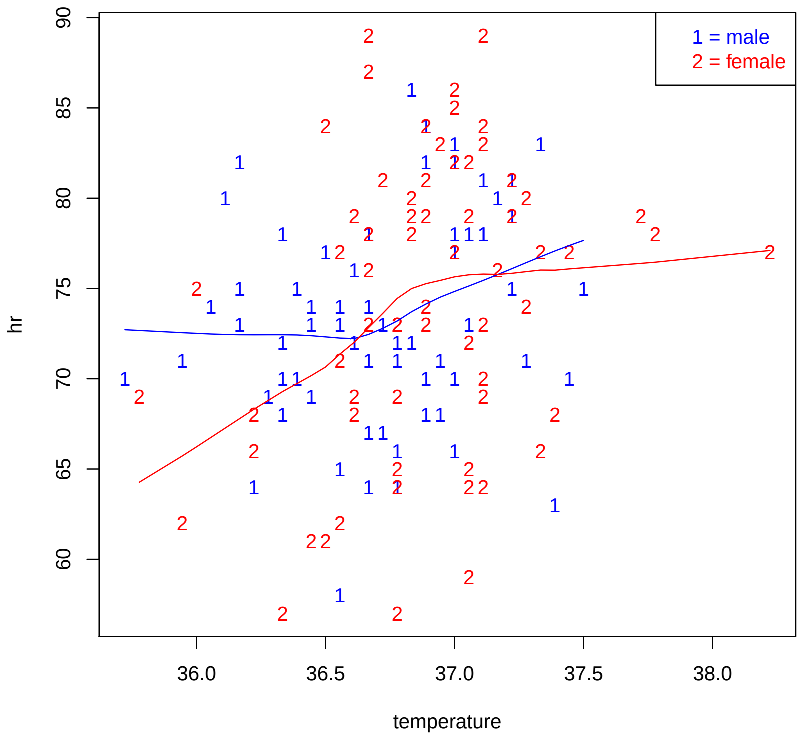
(c) Boxplots:



Histogramme:



(d) Scatterplot:



5 Folgen stochastischer Größen

5.1. [*Starkes GgZ*] X sei eine diskrete sG mit den folgenden Punktwahrscheinlichkeiten:

x	0	1	2	3	4
$P\{X = x\}$	0.1	0.2	0.3	0.35	0.05

Wenn $(X_n; n \in \mathbb{N})$ eine UIV–Folge mit der obigen Verteilung ist, wie lautet das starke GgZ, d.h., wogegen konvergiert $\overline{X}_n = \sum_{i=1}^{n} X_i/n$ fast sicher (für $\to \infty$) ?

5.2. [*Starkes GgZ*] Das starke GgZ besagt, daß – unter bestimmten Voraussetzungen – die sukzessiven arithmetischen Mittelwerte $\overline{X}_n$ einer UIV–Folge von sGn fast sicher zum gemeinsamen Mittelwert μ konvergieren. Wogegen konvergieren aber die sukzessiven *geometrischen* Mittelwerte fast sicher?

$$\left(\prod_{i=1}^{n} X_i\right)^{1/n} \xrightarrow{\text{f.s.}} \ ? \qquad (\text{Vs.: } X_i \geq 0)$$

5.3. [*Starkes GgZ*] Die Zeit bis zum Ausfall einer bestimmten Komponente sei eine stochastische Größe mit Dichte:

$$f(x) = 2x, \quad 0 < x < 1$$

Fällt eine Komponente aus, wird sie sofort durch eine gleichartige neue Komponente ersetzt. Bezeichnet X_i die Lebensdauer der i–ten Komponente, so ist $S_n = \sum_{i=1}^{n} X_i$ der Zeitpunkt des n–ten Ausfalls. Die *Rate* mit der auf lange Sicht Ausfälle auftreten, ist definiert durch:

$$r = \lim_{n\to\infty} \frac{n}{S_n}$$

Wenn die sGn X_i, $i \geq 1$, unabhängig sind, bestimmen Sie r.

5.4. [*Starkes GgZ*] Betrachten Sie ein Quadrat der Seitenlänge 2 (in Nullpunktslage) und den eingeschriebenen Kreis. Wählt man zufällig einen Punkt (V_1, V_2) im Quadrat, so ist die Wahrscheinlichkeit, daß der Punkt innerhalb des Kreises liegt, gleich $\pi/4$. (Warum?) Simuliert man eine Folge von Punkten und definiert:

$$X_i = \begin{cases} 1 & \text{wenn der } i\text{–te Punkt innerhalb des Kreises liegt} \\ 0 & \text{sonst} \end{cases}$$

so folgt, daß X_i, $i \geq 1$, eine UIV–Folge mit $\mathbb{E}(X_i) = \pi/4$ ist. Nach dem starken GgZ gilt:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{f.s.}} \frac{\pi}{4}$$

D.h., durch Simulation einer großen Zahl von Punkten (V_1, V_2) läßt sich der Wert von π approximieren. Erzeugen Sie auf diese Weise $n = 10000$ Punkte und ermitteln Sie einen Schätzwert für π. (Streuung des Schätzwerts?) Verwenden Sie auch die Funktion `simpi3`.

5.5. [*ZGVS/Binomialverteilung*] Ein symmetrischer Würfel wird 1000 Mal geworfen. Berechnen Sie approximativ die Wahrscheinlichkeit, daß die Augenzahl 6 zwischen 150 und 200 Mal inklusive geworfen wird. Wenn die Augenzahl 6 exakt 200 Mal geworfen wird, berechnen Sie approximativ die Wahrscheinlichkeit, daß die Augenzahl 5 weniger als 150 Mal geworfen wird. (Rechnen Sie mit Stetigkeitskorrektur.)

5.6. [*ZGVS/Binomialverteilung*] Beim französischen Roulette gibt es 37 Felder, nummeriert mit 0, 1, ..., 36. Wenn Sie 1€ auf eine bestimmte Zahl setzen, so gewinnen Sie entweder 35€, wenn diese Zahl kommt, oder Sie verlieren den Einsatz, wenn die Zahl nicht kommt. Wenn Sie kontinuierlich auf diese Weise spielen, mit welcher approximativen Wahrscheinlichkeit sind Sie (a) nach 34 Spielen, (b) nach 1000 Spielen, (c) nach 100000 Spielen im Plus? (Rechnen Sie mit Stetigkeitskorrektur.) Beim amerikanischen Roulette gibt es 38 Felder, nummeriert mit 0, 00, 1, ..., 36. Wenn sonst alles gleich bleibt, sind die obigen Wahrscheinlichkeiten kleiner oder größer?

5.7. [*ZGVS/Poissonverteilung*] Ein Programm bestehe aus $n = 100$ Seiten Code und X_i sei die Zahl der Fehler auf der i–ten Seite. Wenn die X_i's unabhängig und identisch poissonverteilt mit Mittel $\mu = 0.8$ sind, bestimmen Sie für die Gesamtzahl $Y = \sum_{i=1}^{n} X_i$ der Fehler einen approximativen Wert für $P\{75 < Y < 85\}$. (Rechnen Sie mit Stetigkeitskorrektur.)

5.8. [*ZGVS/Poissonverteilung*] Die Zahl X der Zugriffe auf eine Webseite folge einer Poissonverteilung mit einem Mittelwert von 10000 pro Tag. Bestimmen Sie approximativ:

(a) Die Wahrscheinlichkit von mehr als 20000 Zugriffen pro Tag.
(b) Die Wahrscheinlichkeit von weniger als 9900 Zugriffen pro Tag.
(c) Einen Wert c so, daß $P\{X > c\} \approx 0.01$.
(d) Die zu erwartende Anzahl von Tagen in einem Jahr (365 Tage), an denen es mehr als 10200 Zugriffe gibt.
(e) Die Wahrscheinlichkeit, daß es in einem Jahr (365 Tage) mehr als 15 Tage mit jeweils mehr als 10200 Zugriffen gibt.

5.9. [*ZGVS*] Wenn 10 symmetrische Würfel geworfen werden, mit welcher (approximativen) Wahrscheinlichkeit liegt die Augensumme zwischen 30 und 40 inklusive? (Rechnen Sie mit Stetigkeitskorrektur.)

5.10. [*ZGVS*] Angenommen, eine bestimmte Komponente ist kritisch für die Funktionsfähigkeit eines Systems, und muß nach Ausfall sofort ausgetauscht werden. Wenn die mittlere Lebensdauer dieser Komponente 10 [h] und die Standardabweichung 30 [h] beträgt, wieviele derartige Komponenten müssen vorrätig sein, sodaß die Funktion des Systems für die nächsten 2000 Stunden mit einer Mindestwahrscheinlichkeit von 0.95 gewährleistet ist?

5.11. [*ZGVS*] A hat 20 Jobs zu erledigen, wobei die für die Erledigung der Jobs benötigten Zeitspannen unabhängige sGn mit Mittelwert 50 [min] und Standardabweichung 10 [min] sind. B hat ebenfalls 20 Jobs zu erledigen, wobei die für die Erledigung der Jobs benötigten Zeitspannen unabhängige sGn mit Mittelwert 52 [min] und Standardabweichung 15 [min] sind. Mit welcher (approximativen) Wahrscheinlichkeit ist A vor B fertig?

5.12. [*Empirische Verteilungsfunktion*] Die folgenden 20 Beobachtungen (`randex.dat`) sind simulierte Beobachtungen einer Ex_{10}–Verteilung:

3.1	0.7	23.5	6.5	3.5	1.9	35.0	4.3	1.9	8.9
4.5	12.6	22.7	11.2	7.3	0.1	25.7	0.6	9.1	3.4

Zeichnen Sie die empirische Verteilungsfunktion F_{20}^* und bestimmen Sie das Supremum des Abstands zwischen F_{20}^* und der Verteilungsfunktion F der Ex_{10}–Verteilung. Bestimmen Sie dieses Supremum „mit der Hand“, mit Hilfe der Funktion `ks.test`, und mit Hilfe der (eigenen) Funktion `dist.exp`.

R: Ist `x` der Datenvektor, so lauten die R–Commands `ks.test(x, pexp, rate=1/10)` bzw. `dist.exp(x, tau=10)`.

5.13. [*Empirische Verteilungsfunktion*] Zeichnen Sie für die beiden Datensätze (`euroweight4.dat`, `euroweight6.dat`) von Aufgabe 1.6 die empirische Verteilungsfunktion und bestimmen Sie das Supremum des Abstands von der Verteilungsfunktion einer Normalverteilung $N(\mu, \sigma^2)$. Ersetzen Sie die (unbekannten) Parameter μ und σ^2 durch die empirischen Gegenstücke $\overline{x}$ und s^2. (Vgl. auch Aufgabe 1.12.) Nehmen Sie dazu die Funktion `ks.test` und die (eigene) Funktion `dist.norm`.

R: Ist `x` der Datenvektor, so lauten die R–Commands `ks.test(x, pnorm, mean=mean(x), sd=sd(x))` bzw. `dist.norm(x)`.

5.14. [*Statistiken*] $X_1, \ldots, X_n$ sei eine Stichprobe aus einer Verteilung F mit Mittelwert μ und Varianz σ^2. Welche der folgenden Größen sind Statistiken?

(1) $X_{(i)},\ i = 1, \ldots, n$ (2) $\overline{X}_n$ (3) S_n^2

(4) $\sqrt{n}\,\dfrac{\overline{X}_n - \mu}{S_n}$ (5) $F_n^*(x),\ x \in \mathbb{R}$ (6) $\sup_x |F_n^*(x) - F(x)|$

5.15. [*Stichprobe*] Eine exponentialverteilte stochastische Größe, $X \sim Ex_\tau$, wird n Mal unabhängig beobachtet, $X_1, \ldots, X_n$.

(a) Wie lautet der Merkmalraum, der Parameterraum, der Stichprobenraum?
(b) Ermitteln Sie die gemeinsame Dichtefunktion der Stichprobe.
(c) Wie lautet eine passende Schätzfunktion für τ ?

5.16. [*Stichprobe*] $X_1, \ldots, X_n$ sei eine Stichprobe aus einer stetigen Verteilung mit Dichte:

$$f(x; \theta) = \theta x^{\theta - 1} I_{(0,1)}(x), \quad \theta > 0$$

(a) Wie lautet der Merkmalraum, der Parameterraum, der Stichprobenraum?
(b) Ermitteln Sie die gemeinsame Dichtefunktion der Stichprobe.
(c) Wie lautet eine passende Schätzfunktion für θ ?

Anhang 5

5.1 Markoff'sche Ungleichung: Ist X eine nichtnegative sG, deren Mittelwert $\mathbb{E}(X)$ existiert, so gilt für jedes $a > 0$:

$$P\{X \geq a\} \leq \frac{\mathbb{E}(X)}{a}$$

Bsp: Angenommen, die Zahl der Einheiten, die in einer Fabrik in einer Woche erzeugt werden, ist eine sG mit Mittelwert 50. Was läßt sich über die Wahrscheinlichkeit sagen, mit der die Wochenproduktion 75 Einheiten übersteigt?

$$P\{X > 75\} \leq \frac{\mathbb{E}(X)}{75} = \frac{50}{75} = \frac{2}{3}$$

5.2 Tschebyscheff'sche Ungleichung: X sei eine sG mit existierendem Mittelwert μ und existierender Varianz σ^2, dann gilt für jedes $k > 0$:

$$P\big\{|X - \mu| \geq k\big\} \leq \frac{\sigma^2}{k^2}$$

Bem1: Die Tschebyscheff'sche Ungleichung folgt aus der Markoff'schen Ungleichung, denn $(X - \mu)^2$ ist eine nichtnegative sG mit $\mathbb{E}\big[(X - \mu)^2\big] = \mathrm{Var}(X)$.

Bem2: Als Folge der nur geringen Voraussetzungen sind Abschätzungen auf Basis der Tschebyscheff'sche Ungleichung meist nur grob. An Hand von Beispielen kann man aber demonstrieren, daß die Ungleichung *scharf* ist, womit gemeint ist, daß sie ohne zusätzliche Voraussetzungen nicht verbessert („verschärft“) werden kann.

Bsp: Wenn die Streuung der Wochenproduktion im Bsp von Anhang 5.1 gleich 5 ist, was läßt sich über die Wahrscheinlichkeit sagen, mit der die Wochenproduktion zwischen 40 und 60 Einheiten liegt?

$$P\big\{|X - 50| \geq 10\big\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Damit folgt:

$$P\big\{|X - 50| < 10\big\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

5.3 Schwaches GgZ: $X_1, X_2, \ldots$ sei eine UIV–Folge von sGn mit existierendem Mittelwert $\mathbb{E}(X_i) = \mu$. Dann gilt für jedes $\varepsilon > 0$:

$$\lim_{n\to\infty} P\big\{|\overline{X}_n - \mu| \geq \varepsilon\big\} = 0$$

Diese Form der Konvergenz nennt man *Konvergenz in der Wahrscheinlichkeit* (oder *stochastische Konvergenz*) und schreibt:

$$\overline{X}_n \xrightarrow{P} \mu$$

Bem1: Der Beweis des schwachen GgZ in der obigen Form ist nicht ganz einfach. Unter der zusätzlichen Voraussetzung einer existierenden endlichen Varianz σ^2 ist das Gesetz aber eine unmittelbare Folgerung aus der Tschebyscheff'schen Ungleichung. Darüberhinaus folgt das schwache GgZ aus dem starken Gesetz.

Bem2: Worin besteht der wesentliche Unterschied zwischen dem starken und dem schwachen Gesetz? Das schwache Gesetz besagt, daß sich für großes n das arithmetische Mittel $\overline{X}_n$ mit hoher Wahrscheinlichkeit in der Nähe von μ aufhält, nicht aber, daß es dort auch bleibt. Das starke Gesetz hingegen besagt, daß $|\overline{X}_n - \mu|$ mit Wahrscheinlichkeit 1 nur endlich oft größer als ein beliebiges $\varepsilon > 0$ sein kann.

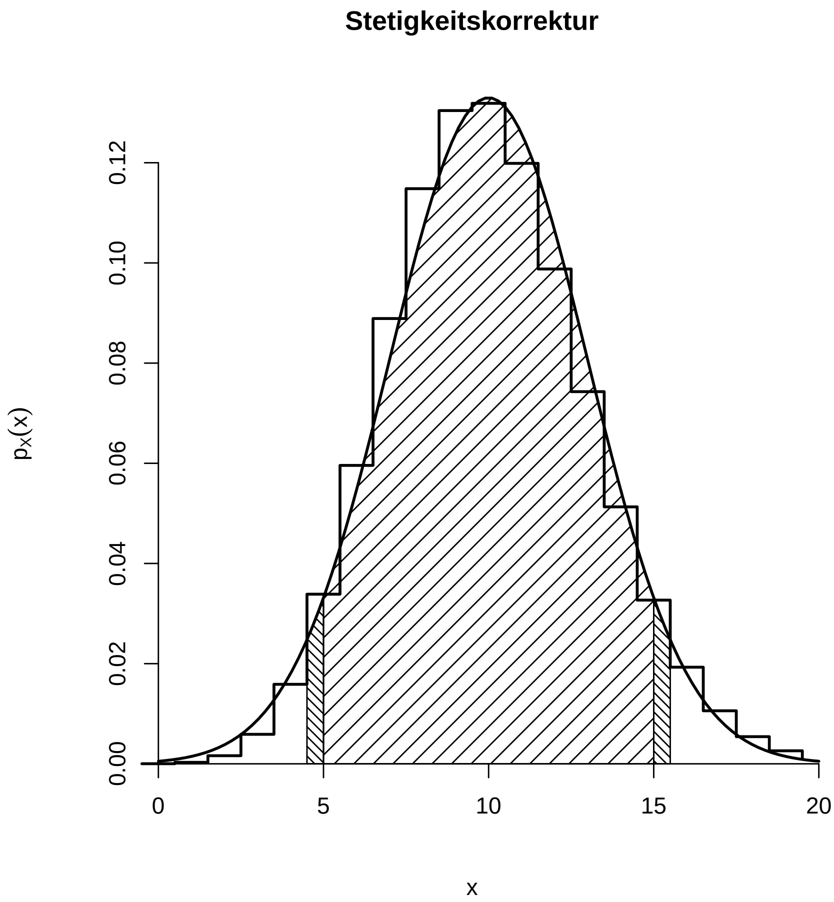
5.4 Stetigkeitskorrektur: Die Approximation einer diskreten Verteilung durch eine stetige Verteilung (insbesondere Normalverteilung) läßt sich meist durch die *Stetigkeitskorrektur* verbessern. (Gelegentlich kann es auch zu einer Verschlechterung kommen.) Ist X die diskrete und Y die stetige Größe, und besteht der Merkmalraum von X aus einer kontinuierlichen Folge von ganzen Zahlen, lautet die Approximation unter Anwendung der Stetigkeitskorrektur wie folgt:

$$P\{a \le X \le b\} \approx P\left\{a - \frac{1}{2} \le Y \le b + \frac{1}{2}\right\}$$

Bsp: Angenommen, man möchte für $X \sim B_{100,0.1}$ die Wahrscheinlichkeit von $5 \le X \le 15$ mit Hilfe der Normalapproximation $X \approx N\big(np, np(1-p)\big) = N(10, 9)$ berechnen:

$$\begin{aligned}
\text{ohne Korrektur:} \quad & P\{5 \le X \le 15\} \approx \Phi\left(\frac{15-10}{3}\right) - \Phi\left(\frac{5-10}{3}\right) \doteq 0.9044 \\
\text{mit Korrektur:} \quad & P\{5 \le X \le 15\} \approx \Phi\left(\frac{15.5-10}{3}\right) - \Phi\left(\frac{4.5-10}{3}\right) \doteq 0.9332 \\
\text{exakt:} \quad & P\{5 \le X \le 15\} \doteq 0.9364
\end{aligned}$$

Die folgende Abbildung zeigt die geometrischen Verhältnisse (die stärker schraffierten Bereiche entsprechen der Stetigkeitskorrektur):



Lösungen/Bemerkungen zu ausgewählten Aufgaben

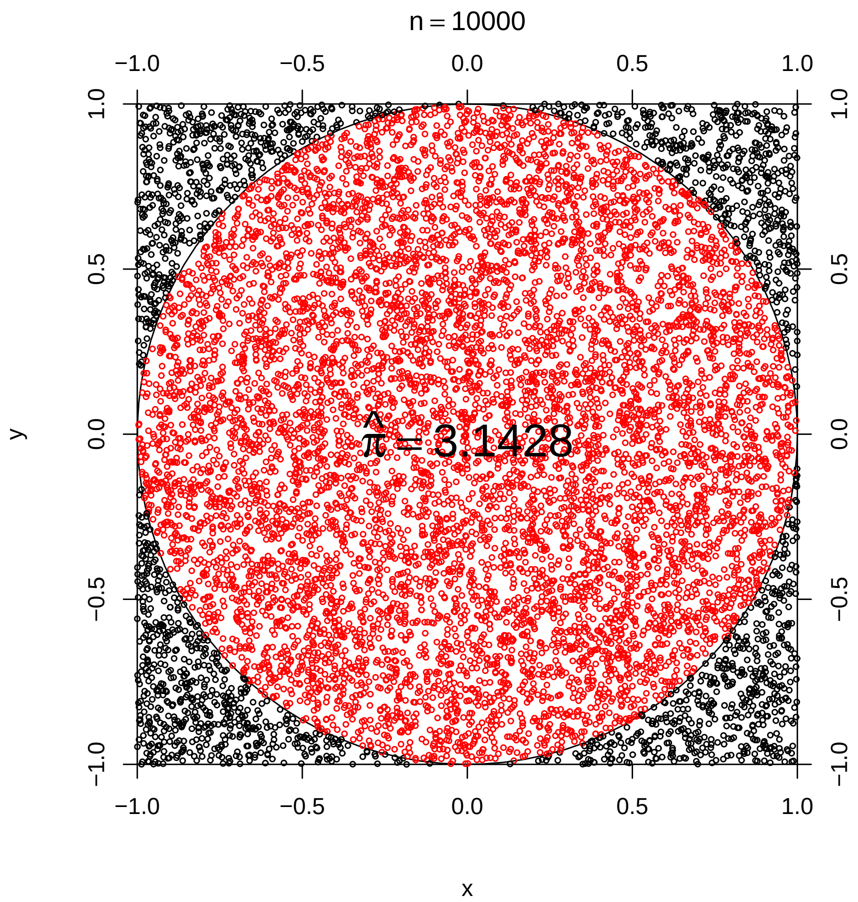
5.2. Nach dem starken GgZ gilt für den Logarithmus:

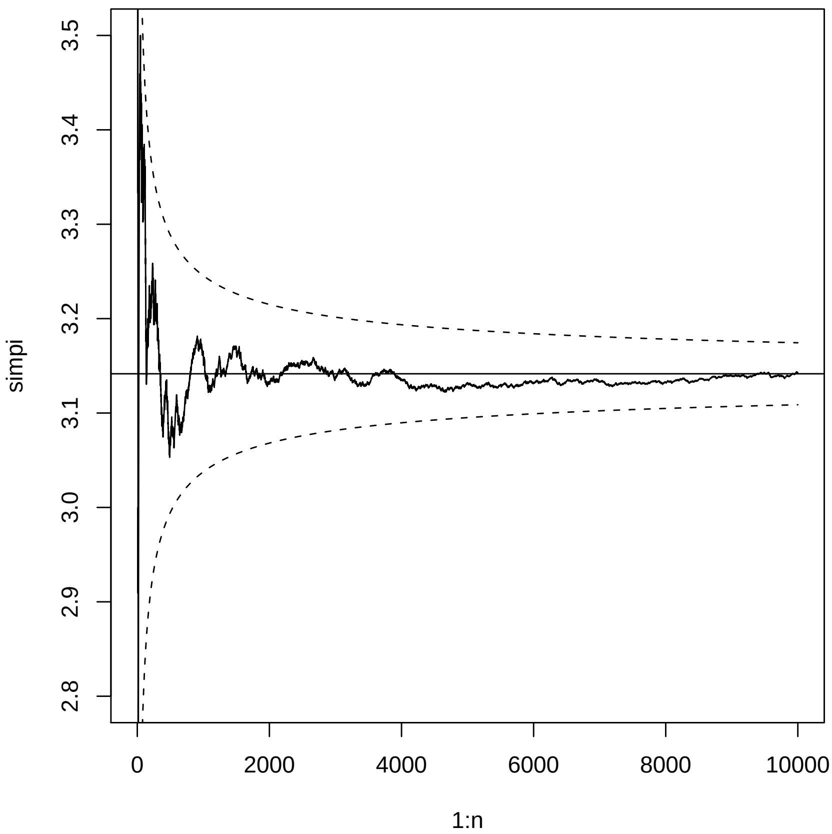
$$\ln\left[\left(\prod_{i=1}^{n} X_i\right)^{1/n}\right] = \frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{\text{f.s.}} \mathbb{E}[\ln(X)]$$

Daher:

$$\left(\prod_{i=1}^{n} X_i\right)^{1/n} \xrightarrow{\text{f.s.}} e^{\mathbb{E}[\ln(X)]}$$

5.4. Ein Aufruf von `simpi3` ergab das folgende Bild:





5.5. Normalapproxiamtion der Binomialverteilung (n groß):

$$B_{n,p} \approx N\big(np, np(1-p)\big)$$

Die Approximation erfolgt so, daß die beiden Mittelwerte und Varianzen übereinstimmen. Sie gilt als zulässig, wenn $np(1-p) > 9$ („Faustregel").

5.6. Ist X die Zahl der gewonnenen Spiele bei insgesamt n Spielen, so beträgt der Nettogewinn nach n Spielen $G = 35X - (n - X) = 36X - n$. Gesucht ist $P\{G > 0\}$.

5.7. Normalapproxiamtion der Poissonverteilung (n groß):

$$P_\mu \approx N(\mu, \mu)$$

Die Approximation erfolgt so, daß die beiden Mittelwerte und Varianzen übereinstimmen. Sie gilt als zulässig, wenn $\mu > 9$ („Faustregel").

5.9. X_i = Augenzahl des i–ten Würfels, $i = 1, \ldots, 10$:

$$\mathbb{E}(X_i) = \frac{7}{2}, \quad \mathrm{Var}(X_i) = \frac{35}{12}, \quad \sum_{i=1}^{10} X_i \mathrel{\dot\sim} N\left(\frac{70}{2}, \frac{350}{12}\right)$$

5.10. Die Lösung führt auf eine quadratische (Un–) Gleichung für n. Letztere läßt sich mit Hilfe von R auch einfach durch Probieren lösen.

5.12. Erstellen Sie für die rechnerische Bestimmung von $D = \sup_x |F^*(x) - F(x)|$, mit $F(x) = 1 - e^{-x/10}$, $x > 0$, eine Tabelle der folgenden Form:

i	$x_{(i)}$	i/n	$(i-1)/n$	$F(x_{(i)})$	$\lvert F(x_{(i)}) - i/n\rvert$	$\lvert F(x_{(i)}) - (i-1)/n\rvert$
1	0.1	0.05	0.00	0.0100	0.0400	0.0100
2	0.6	0.10	0.05	0.0582	0.0418	0.0082
3	0.7	0.15	0.10	0.0676	0.0824	0.0324
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Bem: Die Funktion `ks.test` implementiert den – in der VO/UE nicht behandelten – *Kolmogoroff-Smirnoff–Test*. Dieser verwendet D als *Testgröße*. (Näheres zu statistischen Tests im folgenden Kapitel.)

5.14. Statistiken sind („meßbare“) Funktionen der Stichprobe (allein).

5.16. Wie man sofort zeigt, gilt $\mathbb{E}(X) = \theta/(\theta+1)$. Einen vernünftigen Schätzer für θ könnte man also wie folgt bestimmen:

$$\frac{\theta}{\theta+1} = \overline{X}_n \quad \Longrightarrow \quad \widehat{\theta} = \frac{\overline{X}_n}{1 - \overline{X}_n}$$

Einen auf diese Weise bestimmten Schätzer nennt man einen *Momentenschätzer*. Es gibt allerdings noch andere Möglichkeiten (vgl. das folgende Kapitel).

6 Klassische schließende Statistik

6.1. [*Unverzerrtheit*] Zeigen Sie, daß für eine Stichprobe $X_1, \dots, X_n$ aus einer Verteilung mit dem Mittelwert μ und der Varianz σ^2 der Stichprobenmittelwert $\overline{X}_n$ und die Stichprobenvarianz S_n^2:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

unverzerrte Schätzer für μ bzw. σ^2 sind. (*Hinweis*: Verwenden Sie für den Nachweis der Unverzerrtheit von S_n^2 für σ^2 die Darstellung von Aufgabe 1.11.)

6.2. [*Unverzerrtheit/Effizienz*] $X_1, \dots, X_7$ sei eine Stichprobe aus einer Verteilung mit Mittelwert μ und Varianz σ^2. Betrachten Sie die folgenden (linearen) Schätzer von μ:

$$\widehat{\mu}_1 = \frac{X_1 + \cdots + X_7}{7}, \qquad \widehat{\mu}_2 = \frac{2X_1 - X_6 + X_4}{2}$$

Sind die Schätzer unverzerrt? Welcher ist effizienter (d.h., hat die kleinere Varianz)? Wie lautet der lineare effiziente Schätzer von μ?

6.3. [*Unverzerrtheit/Effizienz*] $X_1, \dots, X_n$ sei eine Stichprobe aus einer uniformen $U(0, \theta)$–Verteilung, wobei $\theta > 0$. Zwei Schätzer stehen zur Auswahl:

$$\widehat{\theta}_1 = 2\overline{X}_n, \qquad \widehat{\theta}_2 = \max_i\{X_i\}$$

(a) Zeigen Sie, daß $\widehat{\theta}_1$ ein unverzerrter Schätzer für θ ist. Varianz des Schätzers?

(b) Zeigen Sie, daß $\widehat{\theta}_2$ *nicht* unverzerrt für θ ist. (*Hinweis*: Bestimmen Sie zuerst die Dichte – vgl. Kapitel 5 – und dann den Erwartungswert von $\widehat{\theta}_2$.)

(c) Bestimmen Sie die Konstante c so, daß $c\widehat{\theta}_2$ ein unverzerrter Schätzer für θ ist.

(d) Wie lautet der lineare effiziente Schätzer von θ?

6.4. [*Unverzerrtheit/Effizienz*] Ein Parameter μ wird nach zwei Methoden gemessen: $X_1, \dots, X_{n_1}$ sind (ua.) Messungen nach der ersten, $Y_1, \dots, Y_{n_2}$ sind (ua.) Messungen nach der zweiten Methode. Beide Meßmethoden sind unverzerrt, für die Varianzen gilt aber $\sigma_2^2 = a\sigma_1^2$. Betrachten Sie den folgenden kombinierten (oder gewichteten) Schätzer für μ:

$$\widehat{\mu} = \alpha\overline{X}_{n_1} + (1-\alpha)\overline{Y}_{n_2}, \quad 0 \le \alpha \le 1$$

(a) Zeigen Sie, daß $\widehat{\mu}$ für jedes α ein unverzerrter Schätzer für μ ist. Varianz/Standardabweichung des Schätzers?

(b) Wie ist α zu wählen, sodaß die Varianz des Schätzers minimal ist? Wie groß ist dann die minimale Varianz?

(c) Angenommen, $a = 4$ und $n_1 = 2n_2$. Welcher Wert von α ist optimal?

6.5. [*Konsistenz*] Zeigen Sie, daß für eine Stichprobe $X_1, \dots, X_n$ aus einer Verteilung mit Mittelwert μ und Varianz σ^2 der Stichprobenmittelwert $\overline{X}_n$ und die Stichprobenvarianz S_n^2 (stark) konsistente Schätzer für μ bzw. σ^2 sind.

6.6. [*Plausible Schätzung/Alternativverteilung*] Angenommen, bei der Herstellung von ICs mittels Photolithographie stellt sich heraus, daß von 300 zufällig ausgewählten ICs 13 defekt sind. Bestimmen Sie – inklusive Herleitung – den plausiblen Schätzer/Schätzwert des Defektanteils p bei dieser Produktionsmethode.

6.7. [*Plausible Schätzung/Poissonverteilung*] Bei der optischen Prüfung von 20 zufällig herausgegriffenen Autoblechen wurden die folgenden Anzahlen von Lackfehlern pro Blech gefunden:

$$1 \quad 7 \quad 1 \quad 3 \quad 2 \quad 5 \quad 2 \quad 8 \quad 5 \quad 4 \quad 6 \quad 5 \quad 4 \quad 6 \quad 2 \quad 4 \quad 5 \quad 2 \quad 3 \quad 6$$

Bestimmen Sie – inklusive Herleitung – den plausiblen Schätzer/Schätzwert für die mittlere Lackfehleranzahl μ pro Blech.

6.8. [*Plausible Schätzung/Normalverteilung*] Einer größeren Lieferung von Kondensatoren werden probeweise 15 Stück zufällig entnommen und ihre Kapazität gemessen. Dabei ergaben sich die folgenden Werte (in pF):

492	512	502	487	500	483	490	498	489	503
497	494	508	506	497					

Wenn man davon ausgeht, daß es sich um Beobachtungen aus einer $N(\mu, \sigma^2)$–Verteilung handelt, bestimmen Sie – inklusive Herleitung (vgl. dazu die VO) – die plausiblen Schätzer/Schätzwerte von (a) μ, (b) σ^2 und (c) σ.

6.9. [*Plausible Schätzung/Exponentialverteilung*] Werden (herkömmliche) Glühlampen (60 W) unter normalen Bedingungen (230 V, Glühfadentemperatur: 2700 K) vom Einschalten bis zum Ausfall beobachtet, so folgt die Brenndauer näherungsweise einer Exponentialverteilung. Angenommen, bei 25 Glühlampen ergibt sich eine mittlere Brenndauer von 976 Stunden. Bestimmen Sie – inklusive Herleitung – den plausiblen Schätzer/Schätzwert (a) für den Mittelwert und (b) für den Median der Brenndauer, sowie (c) für die Wahrscheinlichkeit, daß eine Glühlampe länger als 2000 Stunden brennt. (*Hinweis:* Beachten Sie für (b) und (c) Anhang 6.1.)

6.10. [*Plausible Schätzung/Uniforme Verteilung*] Zeigen Sie, daß für die Situation von Aufgabe 6.3 der Schätzer $\widehat{\theta}_2$ der plausible Schätzer von θ ist.

6.11. [*Plausible Schätzung*] Die folgenden Beobachtungen stammen aus einer Verteilung mit der Dichte $f(x; \theta)$ von Aufgabe 5.16:

$$0.94 \quad 0.74 \quad 0.49 \quad 0.76 \quad 0.59 \quad 0.75 \quad 0.60 \quad 0.58$$

Bestimmen Sie den plausiblen Schätzer/Schätzwert von θ.

6.12. [*Konfidenzintervall/Normalverteilung*] Bestimmen Sie ein (zweiseitiges) (a) 90%–, (b) 95%– und (c) 99%–Konfidenzintervall für für die mittlere Kapazität μ von Aufgabe 6.8.

6.13. [*Konfidenzintervall/Normalverteilung*] Bestimmen Sie ein (zweiseitiges) 95%–Konfidenzintervall für (a) die Varianz σ^2 und für (b) die Streuung σ der Kapazitäten von Aufgabe 6.8.

6.14. [*Konfidenzintervall/Exponentialverteilung*] Entwickeln Sie unter Berufung auf den ZGVS ein (approximatives) 95%–Konfidenzintervall für die mittlere Brenndauer der Glühlampen von Aufgabe 6.9.

6.15. [*Konfidenzintervall/Alternativverteilung*] $X_1, \ldots, X_n$ sei eine Stichprobe aus einer Alternativverteilung A_p. Begründen Sie, warum die folgende Größe eine approximative Pivot–Größe ist:

$$Z_n = \frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1 - \overline{X}_n)/n}} \mathrel{\dot\sim} N(0,1)$$

(a) Entwickeln Sie auf Basis von Z_n ein approximatives $100(1-\alpha)\%$–Konfidenzintervall für den Parameter p. (Vgl. auch Anhang B.4.)

(b) Ermitteln Sie ein 95%–Konfidenzintervall für den Defektanteil p von Aufgabe 6.6.

6.16. [*Konfidenzintervall/Poissonverteilung*] $X_1, \ldots, X_n$ sei eine Stichprobe aus einer Poissonverteilung P_μ. Begründen Sie, warum die folgende Größe eine approximative Pivot–Größe ist:

$$Z_n = \frac{\overline{X}_n - \mu}{\sqrt{\overline{X}_n/n}} \mathrel{\dot\sim} N(0,1)$$

(a) Entwickeln Sie auf Basis von Z_n ein approximatives $100(1-\alpha)\%$–Konfidenzintervall für den Parameter μ. (Vgl. auch Anhang B.6.)

(b) Ermitteln Sie ein 95%–Konfidenzintervall für die mittlere Fehlerzahl μ pro Blech von Aufgabe 6.7.

6.17. [*Normalverteilungsnetz*] Prüfen Sie mittels Wahrscheinlichkeitsnetz, ob die Daten von Aufgabe 6.8 aus einer Normalverteilung stammen. Nehmen Sie dazu das vorgefertigte Netz aus Anhang D und zeichnen Sie die Punkte $\big(x_{(i)}, 100(i-1/2)/n\%\big)$, $i = 1, \ldots, n$, ein. Wie kann man dem Netz gegebenenfalls Schätzwerte für μ und σ entnehmen?

R: Ist `x` der Datenvektor, so bekommt man die Punkte mittels `qqnorm(x, datax=TRUE)` und die „Ausgleichsgerade" mittels `qqline(x, datax=TRUE)`. (Wie wird die Gerade gezeichnet?)

6.18. [*Normalverteilungsnetz*] Prüfen Sie mittels Wahrscheinlichkeitsnetz, ob die beiden Datensätze von Aufgabe 1.6 (`euroweight4.dat`, `euroweight6.dat`) aus einer Normalverteilung stammen. Nehmen Sie dazu `qqnorm` und `qqline` (vgl. Aufgabe 6.17) oder die (eigene) Funktion `net.normal2`. Bei letzterer müssen die Daten als Objekt vom Typ `list` übergeben werden, beispielsweise `net.normal2(list(dat1, dat2))`.

6.19. [*Testproblem*] Ein Produzent behauptet, daß höchstens 1% seiner Produkte fehlerhaft ist. Zur Prüfung dieser Behauptung entnehmen Sie – ohne Zurücklegen – aus einem Los der Größe $N = 1000$ zufällig 55 Einheiten, und beschließen, das Los nur dann zu akzeptieren, wenn die Stichprobe nicht mehr als 1 fehlerhafte Einheit enthält.

(a) Formulieren Sie die Null– und Alternativhypothese.

(b) Wie groß ist bei diesem Test die Wahrscheinlichkeit eines Fehlers 1. Art?

(c) Wie groß ist die Wahrscheinlichkeit eines Fehlers 2. Art, wenn der Defektanteil tatsächlich 5% (10%) beträgt?

(*Hinweis*: Rechnen Sie mit der Binomialverteilung.)

6.20. [*Test für den Mittelwert/Normalverteilung/Varianz bekannt*] Für eine normalverteilte stochastische Größe X, deren Varianz mit $\sigma^2 = 4$ bekannt ist, möchten wir $\mathcal{H}_0 : \mu = 100$ gegen $\mathcal{H}_1 : \mu \neq 100$ auf Basis einer Stichprobe der Größe $n = 9$ testen.

(a) Wenn der Annahmeraum durch $98.5 \leq \overline{x} \leq 101.5$ gegeben ist, wie groß ist die Wahrscheinlichkeit eines Fehlers 1. Art?

(b) Wenn der tatsächliche Mittelwert gleich 103 ist, wie groß ist die Wahrscheinlichkeit eines Fehlers 2. Art?

6.21. [*Test für den Mittelwert/Normalverteilung/Varianz unbekannt*] Testen Sie auf Basis der Daten von Aufgabe 6.8 die Hypothese $\mathcal{H}_0 : \mu = 500$ gegen $\mathcal{H}_1 : \mu \neq 500$. Die Fehlerwahrscheinlichkeit 1. Art des Test soll $\alpha = 0.05$ betragen. (Gibt es einen Zusammenhang mit den in Aufgabe 6.12 bestimmten Konfidenzintervallen? Vgl. Anhang 6.3.)

6.22. [*Test für die Varianz/Normalverteilung*] Bei 15 unabhängigen Messungen des Gewichts von einem Blatt Papier ergibt sich eine Stichprobenstreuung von $s = 0.0083$ g. Wenn die Meßwerte normalverteilt sind, testen Sie $\mathcal{H}_0 : \sigma = 0.01$ gegen $\mathcal{H}_1 : \sigma \neq 0.01$ mit $\alpha = 5\%$. Wie groß ist der p–Wert? (Vgl. Anhang 6.2.)

6.23. [*Zwei–Stichproben–Tests/Normalverteilung*] Die Zeit für die Ausführung eines (standardisierten) Programms wurde auf zwei verschiedenen Computersystemen gemessen. Die Stichprobengrößen waren $n_x = 10$ bzw. $n_y = 20$; für die Stichprobenmittel und die Stichprobenvarianzen ergab sich:

$$\overline{x} = 104 \text{ [s]}, \quad \overline{y} = 114 \text{ [s]}$$
$$s_x^2 = 290 \text{ [s}^2\text{]}, \quad s_y^2 = 510 \text{ [s}^2\text{]}$$

(a) Testen Sie unter der Annahme normalverteilter Beobachtungen (mit gleicher Varianz), ob die mittleren Ausführungszeiten auf beiden Systemen gleich sind, d.h. testen Sie $\mathcal{H}_0 : \mu_x = \mu_y$ gegen $\mathcal{H}_1 : \mu_x \neq \mu_y$. ($\alpha = 5\%$)

(b) Ermitteln Sie ein 95%–Konfidenzintervall für die Differenz $\delta = \mu_y - \mu_x$ der Mittelwerte. Gibt es einen Zusammenhang mit (a)? (*Hinweis*: Vgl. für eine passende Pivotgröße VO/Abschnitt 28.3.)

(b) Testen Sie, ob die beiden Varianzen als gleich angesehen werden können, d.h. testen Sie $\mathcal{H}_0 : \sigma_x^2 = \sigma_y^2$ gegen $\mathcal{H}_1 : \sigma_x^2 \neq \sigma_y^2$. ($\alpha = 10\%$)

6.24. [*Chiquadrat–Anpassungstest/einfach*] Ein Würfel wird 100 Mal geworfen, mit dem Ergebnis:

Augenzahl	1	2	3	4	5	6
Häufigkeit	13	17	9	17	18	26

Ist der Würfel ausbalanciert? ($\alpha = 5\%$).

6.25. [*Chiquadrat–Anpassungstest/einfach*] Die folgenden 30 Zufallszahlen wurden mittels des R–Commands `round(sort(runif(30)),4)` erzeugt:

0.0920	0.1469	0.1696	0.1903	0.2304	0.2415	0.2550	0.2917	0.2949	0.3201
0.3300	0.3474	0.3690	0.4259	0.4725	0.4749	0.5155	0.5820	0.5959	0.6509
0.6829	0.6950	0.7144	0.7415	0.8392	0.8459	0.8678	0.8853	0.9005	0.9640

Prufen Sie mit $\alpha = 5\%$, ob die Werte als Beobachtungen einer nach $U(0, 1)$ verteilten sG angesehen werden können. Nehmen Sie dazu die Klasseneinteilung $[0, 0.2), [0.2, 0.4), \ldots , [0.8, 1]$.

6.26. [*Chiquadrat–Anpassungstest/zusammengesetzt*] Eine sG X wurde 100 Mal beobachtet, mit dem folgenden Ergebnis:

Wert	0	1	2	3	4
Häufigkeit	24	30	31	11	4

Ist eine Poissonverteilung ein geeignetes Modell ($\alpha = 5\%$)? Wie groß ist der p–Wert? (*Hinweis*: Bestimmen Sie zuerst den plausiblen Schätzwert für den Parameter μ der Poissonverteilung. Achten Sie auf die Einhaltung der Faustregel $n\widehat{w} \geq 5$.)

6.27. [*Chiquadrat–Anpassungstest/zusammengesetzt*] Der Datensatz `lifetimes.dat` umfaßt 24 Beobachtungen der Ausfallzeit einer elektronischen Komponente. Prüfen Sie mit $\alpha = 5\%$, ob die Exponentialverteilung ein geeignetes Modell ist. Wie groß ist der p–Wert? (*Hinweis*: Bestimmen Sie zuerst den plausiblen Schätzwert für den Parameter τ der Exponentialverteilung. Klassieren Sie die Daten z.B. wie folgt: $(0, 44], (44, 106], (106, 212], (212, \infty)$.)

R: Nehmen Sie die (eigene) Funktion `chi2.exp` mit `nk=4` (Klassen). (Wie erfolgt bei `chi2.exp` die Klasseneinteilung?)

6.28. [*Chiquadrat–Anpassungstest/zusammengesetzt*] Prüfen Sie, ob die beiden Datensätze von Aufgabe 1.6 (`euroweight4.dat`, `euroweight6.dat`) aus einer Normalverteilung stammen. Dazu müssen zuerst die beiden Parameter μ und σ (plausibel) geschätzt werden. Anschließend sind die Daten zu klassieren. Nicht unüblich ist es, die Klassen so zu wählen, daß die zu erwartenden Klassenhäufigkeiten $e_i = n\widehat{w}_i$ alle gleich sind. (Dadurch läßt sich auch die Einhaltung der Faustregel $e_i \geq 5$ auf einfache Weise kontrollieren.)

R: Nehmen Sie die (eigene) Funktion `chi2.normal` mit `nk=20` (Klassen).

6.29. [*Einfache lineare Regression*] Betrachten Sie die folgenden Datenpaare (`xy.dat`):

Beob.	y	x
1	16.68	7
2	11.50	3
3	12.03	3
4	14.88	4
5	13.75	6
6	18.11	7
7	8.00	2
8	17.83	7
9	79.24	30
10	21.50	5
11	40.33	16
12	21.00	10
13	13.50	4

Beob.	y	x
14	19.75	6
15	24.00	9
16	29.00	10
17	15.35	6
18	19.00	7
19	9.50	3
20	35.10	17
21	17.90	10
22	52.32	26
23	18.75	9
24	19.83	8
25	10.75	4

(a) Stellen Sie die Datenpaare graphisch dar (Streudiagramm). Besteht ein (annähernder) linearer Zusammenhang?

(b) Bestimmen Sie nach den in der VO angegebenen Formeln Schätzwerte für die Koeffizienten α und β einer Regressionsgeraden:

$$Y_x = \alpha + \beta x + U_x$$

(c) Wenn $\mathrm{Var}(U_x) = \sigma^2$ (konstant), bestimmen Sie einen Schätzwert für σ^2.

(d) Führen Sie die Berechnungen auf Basis von R durch. (*Hinweis*: Die zentrale Funktion für die Anpassung von *linearen Modellen* verschiedenster Art ist `lm`. Im vorliegenden Fall lauten entsprechende Commands `mod <- lm(y ~ x)` und `summary(mod)`.)

6.30. [*Polynomiale Regression*] Der Datensatz `bremsweg.dat` umfaßt 62 Messungen des Bremsweges eines PKW bei verschiedenen Geschwindigkeiten (unter gleichen Bedingungen hinsichtlich Straßenbelag, Witterung, ...).

(a) Stellen Sie die Daten graphisch dar (Streudiagramm von `weg` gegen `v`). Besteht ein linearer Zusammenhang?

(b) Bestimmen Sie die Koeffizienten des folgenden quadratischen Modells:

$$\texttt{weg} = \alpha + \beta\,\texttt{v} + \gamma\,\texttt{v}^2 + \text{Fehler}$$

R: `mod <- lm(weg ~ v + I(v^2), data=bremsweg)` und `summary(mod)`.

Anhang 6

6.1 Plausible Schätzung: Der plausible Schätzer hat eine bemerkenswerte *Invarianzeigenschaft*: Sei $X_1, \dots, X_n$ eine Stichprobe von $X \sim f(x;\theta)$, $\theta \in \Theta$, und $\eta = g(\theta)$ eine Funktion des Parameters. Ist $\widehat{\theta}$ der plausible Schätzer von θ, dann ist $\widehat{\eta} = g(\widehat{\theta})$ der plausible Schätzer von η. (Dies gilt für ein– und mehrdimensionale Parameter.)

Bsp: Die Invarianz der plausiblen Schätzung erspart eine Menge an Rechenarbeit. Beispielsweise ist der plausible Schätzer von σ^2 auf Basis einer Stichprobe $X_1, \dots, X_n$ aus einer Normalverteilung $N(\mu, \sigma^2)$ gegeben durch (vgl. die VO und Aufgabe 6.8):

$$\widehat{\sigma^2} = \frac{n-1}{n} S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Ohne weitere Rechnung gilt auf Grund der Invarianz, daß der plausible Schätzer beispielsweise für die Streuung σ $(= \sqrt{\sigma^2})$ gegeben ist durch:

$$\widehat{\sigma} = \sqrt{\widehat{\sigma^2}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

6.2 p-Wert: Praktisch alle Statistikpakete (auch R) verfolgen beim Testen von Hypothesen nicht die in der VO dargestellte „klassische“ Vorgangsweise, sondern berechnen statt dessen den p–Wert. Der *p–Wert* (auch *beobachtetes Signifikanzniveau*, engl. *p–value*) einer $\mathcal{H}_0$ entspricht der Wahrscheinlichkeit, bei Zutreffen von $\mathcal{H}_0$ den beobachteten Wert der Teststatistik oder einen extremeren zu bekommen. Was unter „extremer“ zu verstehen ist, hängt von der Gegenhypothese (oder dem kritischen Bereich) ab. Testet man beispielsweise:

$$\mathcal{H}_0 : \theta \le \theta_0 \quad \text{gegen} \quad \mathcal{H}_1 : \theta > \theta_0$$

und ist $T = t$ der Wert der Teststatistik, so ist der p–Wert wie folgt zu berechnen:

$$p\text{–Wert} = P_{\theta_0}(T \geq t)$$

Die Beziehung zum klassischen Testen ergibt sich dadurch, daß eine $\mathcal{H}_0$, deren p–Wert kleiner als α ist, auf dem Niveau α verworfen wird. Die Beurteilung von Hypothesen mittels p–Wert hat (u.a.) den Vorteil, daß man auf Basis *einer* Zahl für *alle* Werte von α die Testentscheidung unmittelbar ablesen kann. Anders ausgedrückt:

Der p–Wert der $\mathcal{H}_0$ ist der *größte* Wert von α, für den die $\mathcal{H}_0$ *nicht* verworfen wird.

Bei der Beurteilung von p–Werten hält man sich meist an das folgende Schema:

p–Wert	Signifikanz	
< 0.01	sehr hoch	(sehr starke Einwände gegen $\mathcal{H}_0$)
$0.01 - 0.05$	hoch	(starke Einwände gegen $\mathcal{H}_0$)
$0.05 - 0.10$	schwach	(schwache Einwände gegen $\mathcal{H}_0$)
> 0.10	keine	(sehr schwache/keine Einwände gegen $\mathcal{H}_0$)

Bem: Die obige Sprechweise von der „Signifikanz“ eines Tests ist zwar weit verbreitet aber mit einer gewissen Vorsicht zu gebrauchen. Ein Test ist „signifikant“, wenn er die Nullhypothese verwirft. Das ist eine formale Aussage, die von den Hypothesen, vom verwendeten Test, von der Stichprobengröße und von α abhängt. Diese *statistische Signifikanz* sollte nicht mit der *praktischen Signifikanz* verwechselt werden. Abhängig vom sachlichen Hintergund mag ein formal signifikantes Ergebnis von großer oder nur von geringer praktischer Bedeutung sein.

6.3 Dualität von Tests und Konfidenzintervallen: Es gibt eine enge Beziehung zwischen Tests und Konfidenzintervallen. Dies soll an einem Beispiel verdeutlicht werden. Angenommen, man möchte für eine Stichprobe $X_1, \ldots, X_n$ aus einer $N(\mu, \sigma^2)$–Verteilung die folgenden Hypothesen auf dem Niveau α gegeneinander testen:

$$\mathcal{H}_0 : \mu = \mu_0 \quad \text{gegen} \quad \mathcal{H}_1 : \mu \neq \mu_0$$

Dazu kann man den üblichen t–Test nehmen und $\mathcal{H}_0$ verwerfen, falls:

$$|T| = \left| \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} \right| > t_{n-1;1-\alpha/2}$$

Oder man bestimmt zuerst ein $100(1-\alpha)\%$–Konfidenzintervall für μ:

$$\big(U(\underline{X}), O(\underline{X})\big) = \left(\overline{X}_n - t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}}, \overline{X}_n + t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \right)$$

und entscheidet sich wie folgt:

$$\begin{aligned} \mu_0 &\in \big(U(\underline{X}), O(\underline{X})\big) \quad \longrightarrow \quad \mathcal{H}_0 \text{ nicht verwerfen} \\ \mu_0 &\notin \big(U(\underline{X}), O(\underline{X})\big) \quad \longrightarrow \quad \mathcal{H}_0 \text{ verwerfen} \end{aligned}$$

Bem: Die Dualität besteht in analoger Weise auch zwischen einseitigen Tests und einseitigen Konfidenzintervallen. (Vgl. Anhang B und C.)

Lösungen/Bemerkungen zu ausgewählten Beispielen

6.1. Nach dem (theoretischen) Verschiebungssatz gilt:

$$\mathbb{E}\big(X_i^2\big) = \mathrm{Var}(X_i) + \mathbb{E}^2(X_i) = \sigma^2 + \mu^2$$
$$\mathbb{E}\big(\overline{X}_n^2\big) = \mathrm{Var}(\overline{X}_n) + \mathbb{E}^2(\overline{X}_n) = \frac{\sigma^2}{n} + \mu^2$$

6.3. Ein *anschauliches* Argument dafür, warum $\widehat{\theta}_2$ nicht unverzerrt für θ ist, lautet wie folgt: Der Erwartungswert einer sG ist ein Durchschnittswert. Da es aber keine Beobachtungen größer als θ geben kann, muß der Erwartungswert des Maximums von n Beoachtungen (echt) kleiner als θ sein.

<u>Formal:</u> Die Verteilungsfunktion/Dichte von $\widehat{\theta}_2$ ist gegeben durch:

$$F_{\max}(x) = \left(\frac{x}{\theta}\right)^n \quad \Longrightarrow \quad f_{\max}(x) = F_{\max}^{'}(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}, \quad 0 < x < \theta$$

Damit läßt sich $\mathbb{E}(\widehat{\theta}_2)$ einfach berechnen.

6.5. Die (starke) Konsistenz von $\overline{X}_n$ ist äquivalent zum starken GgZ:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{f.s.}} \mathbb{E}(X) = \mu$$

Ebenfalls nach dem starken GgZ gilt:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{\text{f.s.}} \mathbb{E}(X^2) = \sigma^2 + \mu^2$$

Damit folgt:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2\right]$$
$$\xrightarrow{\text{f.s.}} (1)\big[\mathbb{E}(X^2) - \mu^2\big] = \sigma^2$$

Dies zeigt die (starke) Konsistenz von S_n^2.

6.9. (a) Plausibilitätsfunktion:

$$l(\tau;\underline{x}) = \prod_{i=1}^{n} f(x_i;\tau) = \prod_{i=1}^{n} \frac{1}{\tau}\, e^{-x_i/\tau} = \frac{1}{\tau^n}\, e^{-n\overline{x}_n/\tau}$$

Log–Plausibilitätsfunktion:

$$l^*(\tau;\underline{x}) = -n\ln(\tau) - \frac{n\overline{x}_n}{\tau}$$

Ableiten und Nullsetzen:

$$\frac{\partial l^*(\tau;\underline{x})}{\partial \tau} = -\frac{n}{\tau} + \frac{n\overline{x}_n}{\tau^2} = 0 \quad \Longrightarrow \quad \widehat{\tau} = \overline{x}_n$$

Plausibler Schätzer:

$$\widehat{\tau} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Plausibler Schätzwert:

$$\widehat{\tau} = \overline{x}_{25} = 976 \text{ [Stunden]}$$

(b) Plausibler Schätzer:

$$\widehat{x}_{0.5} = -\overline{\tau}\ln(0.5) = -\overline{X}_n \ln(0.5)$$

Plausibler Schätzwert:

$$\widehat{x}_{0.5} = -976\ln(0.5) \doteq 676.5 \text{ [Stunden]}$$

(c) Plausibler Schätzer:

$$\widehat{P}\{X > 2000\} = e^{-2000/\widehat{\tau}} = e^{-2000/\overline{X}_n}$$

Plausibler Schätzwert:

$$\widehat{P}\{X > 2000\} = e^{-2000/976} \doteq 0.1288$$

6.10. Zeigen Sie zunächst, daß sich die Plausibilitätsfunktion wie folgt schreiben läßt:

$$l(\theta;\underline{x}) = \frac{1}{\theta^n} I_{(\max\{x_i\},\infty)}(\theta)$$

6.14. Nach dem ZGVS gilt für eine (große) Stichprobe $X_1, \ldots, X_n$ aus einer Ex_τ–Verteilung:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\tau}{\sqrt{n\tau^2}} = \frac{\overline{X}_n - \tau}{\tau/\sqrt{n}} \overset{\cdot}{\sim} N(0,1)$$

D.h., Z_n ist eine approximative Pivot–Größe.

<u>Bem:</u> Vgl. für ein *exaktes* $100(1-\alpha)\%$–Intervall für τ Anhang B.3.

6.15. Die Begründung dafür, daß Z_n eine approximative Pivotgröße ist, liegt im ZGVS, und darin, daß $\overline{X}_n(1-\overline{X}_n)$ ein konsistenter Schätzer von $p(1-p)$ ist.

6.16. Die Begründung dafür, daß Z_n eine approximative Pivotgröße ist, liegt im ZGVS, und darin, daß $\overline{X}_n$ ein konsistenter Schätzer von μ ist.

6.17. Vorbereitungen:

```
x <- c(492,512,502,487,500,483,490,498,489,503,
  497,494,508,506,497)
sort(x)
 [1] 483 487 489 490 492 494 497 497 498 500 502 503 506 508 512
round((1:15 - 1/2)/15*100)
 [1]  3 10 17 23 30 37 43 50 57 63 70 77 83 90 97
```

6.19. Als Nullhypothese wählt man in der Regel diejenige Behauptung, die den „Normalfall“ (oder den „Status quo“) repäsentiert. Als Alternativhypothese wählt man diejenige Behauptung, deren Zutreffen ein bestimmtes Handeln erfordert oder die gravierenderen Konsequenzen hat. (Ein hoher Defektanteil führt zu höheren Kosten, ...).

6.22. Ist c der Wert der Teststatistik, so ist der p–Wert gegeben durch:

$$p\text{–Wert } = 2P\{\chi^2_{14} \le c\}$$

6.23. Zum F–Test: Da es keine Rolle spielt, welche Stichprobe X und welche Y ist, wählt man die Reihenfolge zweckmäßigerweise so, daß $T = S_x^2/S_y^2$ größer als 1 ist. In diesem Fall genügt der Vergleich mit der oberen Grenze des Annahmebereichs (die untere Grenze ist kleiner als 1 und muß dann nicht mehr überprüft werden).

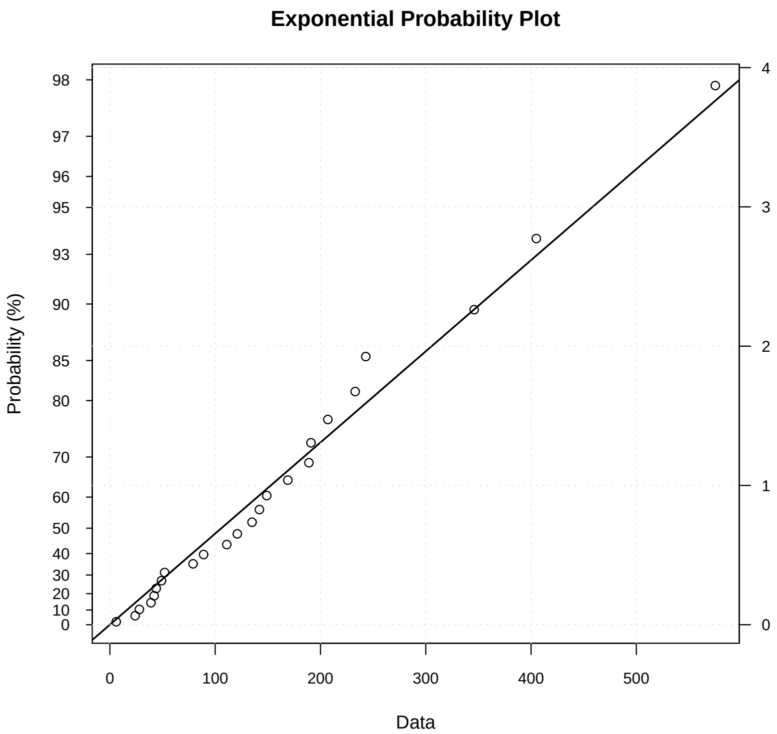
6.25. Erstellen Sie eine Tabelle der folgenden Form:

Klasse	H	w	$e = 30\,w$	$(H-e)^2/e$
[0, 0.2)	4	0.2	6	4/6
[0.2, 0.4)	9	0.2	6	9/6
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

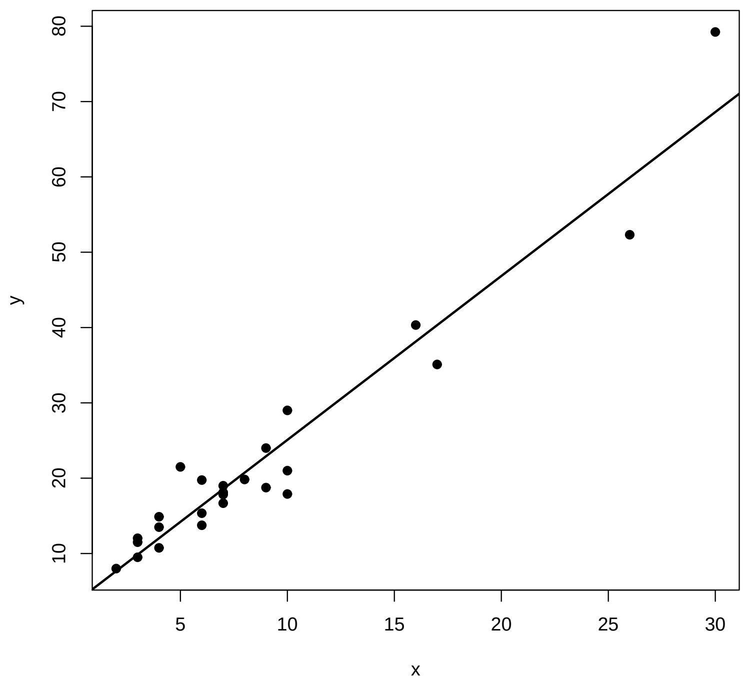
6.26. Ist $\widehat{\mu}$ der plausible Schätzwert von μ (vgl. Aufgabe 6.7), so werden die Punktwahrscheinlichkeiten der Poissonverteilung wie folgt geschätzt:

$$\widehat{p}_x = \widehat{P}\{X = x\} = \frac{\widehat{\mu}^x e^{-\widehat{\mu}}}{x!}, \quad x = 0, 1, \ldots$$

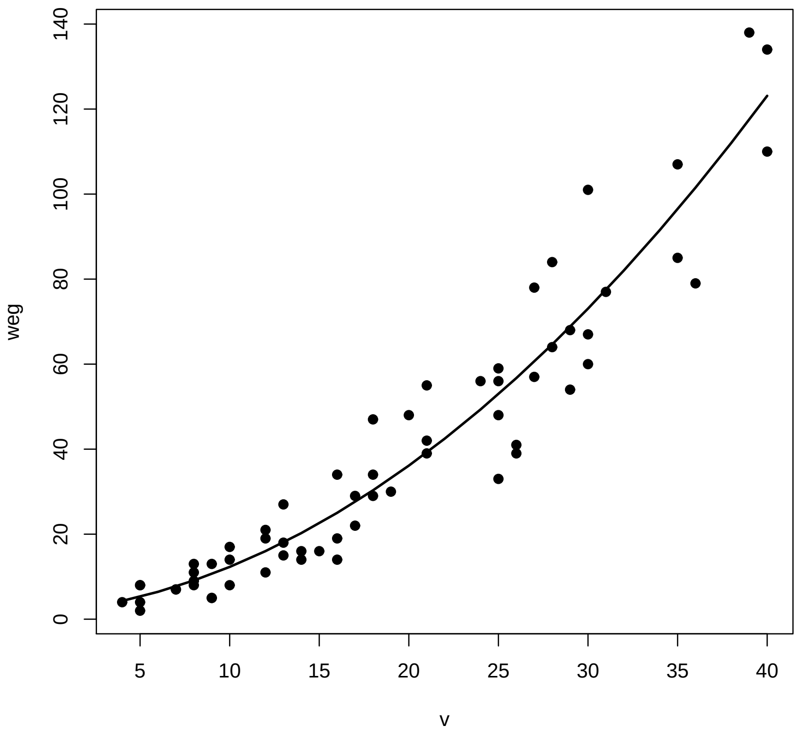
6.27. Zur Prüfung auf Exponentialverteilung kann man auch ein entsprechend konstruiertes W–Netz verwenden (vgl. Anhang D). Dazu zeichnet man die Punkte $\big(x_{(i)}, 100(i-1/2)/n\%\big)$, $i = 1, \ldots, n$, im Netz ein und versucht, sie durch eine Gerade (durch den Nullpunkt) auszugleichen.



6.29. Die folgende Abbildung zeigt das Streudiagramm und die geschätzte Kleinste-Quadrate–Gerade:



6.30. Die folgende Abbildung zeigt das Streudiagramm und die geschätzte Kleinste-Quadrate–Parabel:



7 Elemente der Bayes–Statistik

7.1. [*A-posteriori–Verteilung*] Man nehme an, daß der Defektanteil θ in einem großen Los entweder 0.1 oder 0.2 ist, und daß a-priori gilt:

$$\pi(0.1) = 0.7, \quad \pi(0.2) = 0.3$$

Wenn 8 Einheiten zufällig aus dem Los entnommen werden und davon genau 2 defekt sind, wie lautet die A-posteriori–Verteilung von θ?

7.2. [*A-posteriori–Verteilung*] Die Zahl der (unerwünschten) Bläschen auf einer Glasscheibe folge einer Poissonverteilung, deren Mittelwert μ entweder gleich 1.0 oder gleich 1.5 ist. Wenn a-priori gilt:

$$\pi(1.0) = 0.4, \quad \pi(1.5) = 0.6$$

und bei einer zufällig ausgewählten Glasscheibe 3 Bläschen gefunden werden, wie lautet die A-posteriori–Verteilung von μ?

7.3. [*A-priori–Verteilung*] Die A-priori–Verteilung eines Parameters θ sei eine Gammaverteilung mit dem Mittelwert 10 und der Varianz 5. Bestimmen Sie die A-priori–Dichte von θ. (*Hinweis*: Anhang A.2.5.)

7.4. [*A-priori–Verteilung*] Die A-priori–Verteilung eines Parameters θ sei eine Betaverteilung mit dem Mittelwert 1/3 und der Varianz 1/45. Bestimmen Sie die A-priori–Dichte von θ. (*Hinweis*: Anhang A.2.9.)

7.5. [*A-posteriori–Verteilung/Bayes–Schätzer*] Der Defektanteil θ in einem großen Los sei unbekannt. A-priori gelte (a) $\pi(\theta) = I_{(0,1)}(\theta)$, (b) $\pi(\theta) = 2(1-\theta)I_{(0,1)}(\theta)$. Wenn von 8 zufällig ausgewählten Einheiten genau 3 defekt sind, wie lautet die A-posteriori–Verteilung? Bayes–Schätzer?

7.6. [*A-posteriori–Verteilung/Bayes–Schätzer/HPD–Bereich*] Die Zeit (in Minuten), die eine Person in der Früh auf den Bus warten muß, sei auf dem Intervall $(0,\theta)$ uniform verteilt, wobei $\theta > 0$ unbekannt ist. Die A-priori–Verteilung sei gegeben wie folgt:

$$\pi(\theta) = \begin{cases} \dfrac{192}{\theta^4} & \text{für } \theta \geq 4 \\[2ex] 0 & \text{sonst} \end{cases}$$

Wenn an drei aufeinanderfolgenden Tagen die Wartezeiten 5, 3 und 8 Minuten betragen, bestimmen Sie (a) die A-posteriori–Verteilung, (b) den Bayes–Schätzer und (c) den 95% HPD–Bereich für θ.

7.7. [*Bayes–Schätzer/Bayes–Intervall/Bayes–Test*] Die folgende Stichprobe der Größe 20 stamme aus einer Poissonverteilung mit Parameter μ:

11	7	11	6	5	9	14	10	9	5
8	10	8	10	12	9	3	12	14	4

Wir vermuten, daß μ etwa 12 ist, aber wir sind nicht sicher. Daher wählen wir eine $Gam(10, 1.2)$ als A-priori–Verteilung für μ.

(a) Bestimmen Sie die A-posteriori–Verteilung von μ.

(b) Wie lautet der Bayes–Schätzer von μ ?

(c) Wie lautet der Bayes–Schätzer von μ bezüglich der Verlustfunktion $L(\mu, \widehat{\mu}) = |\mu - \widehat{\mu}|$?

(d) Bestimmen Sie ein 95% (a-posteriori) Bayes–Intervall für μ, d.h., bestimmen Sie ein Intervall (a, b), sodaß:

$$P\{a < \widetilde{\mu} < b|D\} = 0.95$$

(e) Wir möchten die folgenden Hypothesen testen:

$$\mathcal{H}_0 : \theta \leq 10 \quad \text{gegen} \quad \mathcal{H}_1 : \theta > 10$$

Bestimmen Sie die relative Plausibilität α_1/α_0 der Hypothesen.

(*Hinweis*: Anhang A.2.5.)

7.8. [*Bayes–Schätzer/HPD–Intervall*] X sei normalverteilt mit unbekanntem Mittelwert μ und bekannter Varianz $\sigma^2 = 9$. A-priori sei μ normalverteilt mit $\mu_0 = 4$ und $\sigma_0^2 = 1$. Eine Stichprobe des Umfangs $n = 25$ ergibt einen Stichprobenmittelwert von $\overline{x} = 4.85$.

(a) Bestimmen Sie die A-posteriori–Verteilung von μ.

(b) Wie lautet der Bayes–Schätzer von μ ? (Vergleichen Sie mit dem plausiblen Schätzer.)

(c) Wie lautet der Bayes–Schätzer von μ bezüglich der Verlustfunktion $L(\mu, \widehat{\mu}) = |\mu - \widehat{\mu}|$?

(d) Bestimmen Sie das 95% HPD–Intervall für μ.

(e) Beantworten Sie die vorhergehenden Fragen, wenn für μ eine nichtinformative A-priori–Verteilung der Form $\pi(\mu) \propto c$ gewählt wird.

(*Hinweis*: Anhang 7.2.)

7.9. [*A-posteriori–Varianz*] Eine Normalverteilung mit unbekanntem Mittelwert μ und bekannter Varianz $\sigma^2 = 2$ wird n Mal beobachtet. A-priori sei μ normalverteilt mit $\sigma_0^2 = 4$. Wie groß muß n mindestens sein, sodaß die A-posteriori–Varianz nicht größer als 0.01 ist? (*Hinweis*: Anhang 7.2.)

7.10. [*Bayes'sche Entscheidung*] Angenommen, bei einer größeren Lieferung von Früchten gibt es für den Anteil θ an beschädigten Früchten nur die drei Möglichkeiten 0.1, 0.3 und 0.5, und drei mögliche Entscheidungen d_1, d_2 und d_3, deren Verluste sich wie folgt beziffern lassen:

	d_1	d_2	d_3
$\theta = 0.1$	0	1	3
$\theta = 0.3$	2	0	2
$\theta = 0.5$	3	1	0

Wenn a-priori $\pi(0.1) = 0.5$, $\pi(0.3) = 0.3$ und $\pi(0.5) = 0.2$, und die Zahl Y von beschädigten Früchten in einer Stichprobe der Größe 20 als Grundlage für die Entscheidung herangezogen wird, wie lautet die Bayes–Entscheidung?

Anhang 7

7.1 A-priori–Verteilung: In den meisten Fällen hat man ein mehr oder weniger genaues Vorwissen über den in Frage stehenden Parameter. So kann man beispielsweise bei einer Meinungsbefragung zu einem Thema („dafür“/ „dagegen“) auf ähnlich gelagerte (frühere) Umfragen oder auf die eigene Erfahrung zurückgreifen. Das gesamte Vorwissen wird (idealerweise) in der A-priori–Verteilung modelliert. Für die Praxis sind insbesondere zwei Klassen von A-priori–Verteilungen von Bedeutung.

Konjugierte A-priori–Verteilungen: Eine Klasse von A-priori–Verteilungen für eine Familie von Verteilungen mit Dichten $f(x|\theta)$, $\theta \in \Theta$, ist eine *konjugierte Familie* von Verteilungen, wenn die A-posteriori–Verteilung von θ zur gleichen Verteilungsfamilie gehört wie die A-priori–Verteilung.

Bsp: Vgl. die Aufgaben 7.5 (konj. Fam.: Betaverteilungen), 7.6 (konj. Fam.: *Paretoverteilungen*; vgl. Anhang 7.3) und 7.7 (konj. Fam.: Gammaverteilungen).

Nichtinformative A-priori–Verteilungen: Wenn man sich „objektiv“ verhalten möchte, oder wenn man keine Vorstellung von der A-priori–Verteilung hat, greift man auf *nichtinformative* A-priori–Verteilungen zurück. Dabei handelt es sich um Verteilungen, die jeden möglichen Parameterwert gleich gewichten. Im Falle des obigen Beispiels wäre dies für den Anteil der Befürworter etwa eine uniforme Verteilung auf $(0, 1)$. Häufig führt die Forderung nach Objektivität zu *uneigentlichen* Dichten, d.h. zu Dichten, deren Integral nicht endlich ist. Als Beispiel denke man an den Mittelwert μ einer Normalverteilung $N(\mu, \sigma^2)$. Eine nichtinformative A-priori–Verteilung ist hier durch eine konstante Dichte auf ganz $\mathbb{R}$, $\pi(\mu) \propto c$, gegeben (vgl. Anhang 7.2). (*Bem*: Die Verwendung derartiger Verteilungen verursacht in weiterer Folge keine Probleme, solange die A-posteriori–Verteilung wieder eigentlich ist.)

Bem: Die Wahl einer nichtinformativen A–priori–Verteilung ist nicht ganz so einfach, wie man auf den ersten Blick vermuten könnte. Im obigen Beispiel der Meinungsbefragung scheint eine $U(0, 1)$–Verteilung für den Anteil θ der Befürworter die richtige Wahl zu sein. Was aber, wenn man an Stelle von θ beispielsweise $\eta = \theta^2$ als Parameter nimmt? Mit dem Transformationssatz für Dichten folgt, daß die A-priori–Verteilung von η *nicht* uniform ist:

$$\pi(\theta) = I_{(0,1)}(\theta) \quad \Longrightarrow \quad \pi(\eta) = I_{(0,1)}(\sqrt{\eta})\,\frac{1}{2\sqrt{\eta}} = \frac{1}{2\sqrt{\eta}}\,I_{(0,1)}(\eta)$$

Über θ gibt es keine Information, über θ^2 aber schon? Zur Überwindung dieser Problematik gibt es mehrere Vorschläge, auf die hier aber nicht eingegangen werden kann.

7.2 Normalverteilung: $X_1, \ldots, X_n$ sei eine Stichprobe aus einer $N(\mu, \sigma^2)$–Verteilung, wobei μ unbekannt und σ^2 bekannt sei. A-priori sei μ nach $N(\mu_0, \sigma_0^2)$ verteilt. (Die *Hyperparameter* μ_0 und σ_0^2 seien bekannt.) Dann ist die A-posteriori–Verteilung von μ wieder eine Normalverteilung mit den folgenden Parametern ($D = \{x_1, \ldots, x_n\}$):

$$\mu_{\text{a-post}} = \mathbb{E}(\widetilde{\mu}|D) = \frac{\sigma^2\mu_0 + n\sigma_0^2\overline{x}_n}{n\sigma_0^2 + \sigma^2}$$

$$\sigma^2_{\text{a-post}} = \mathrm{Var}(\widetilde{\mu}|D) = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}$$

Man beachte, daß $\mu_{\text{a-post}}$ (= A-posteriori-Bayes–Schätzer) ein gewichteter Mittelwert aus dem A-priori–Mittelwert μ_0 und dem Datenmittelwert $\overline{x}_n$ ist:

$$\mu_{\text{a-post}} = \frac{\sigma^2}{n\sigma_0^2 + \sigma^2}\,\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\,\bar{x}_n$$

Für große Stichproben gilt $\mu_{\text{a-post}} \approx \overline{x}_n$ (= plausibler Schätzwert). Für eine nichtinformative (uneigentliche) A-priori–Verteilung der Form $\pi(\mu) \propto c$ gilt a-posteriori:

$$\widetilde{\mu}|D \sim N\left(\overline{x}_n, \frac{\sigma^2}{n}\right)$$

Dieses Ergebnis erhält man auch, wenn man in der A-priori–Verteilung $N(\mu_0, \sigma_0^2)$ die Varianz σ_0^2 gegen ∞ gehen läßt, d.h. wenn die A-priori–Verteilung immer flacher wird.

Bsp: Vgl. die Aufgaben 7.8 und 7.9.

7.3 Paretoverteilung: Eine für viele Anwendungen wichtige Verteilung ist die *Paretoverteilung*. Ihre Dichte ist gegeben durch:

$$f(x|x_0, \alpha) = \begin{cases} \dfrac{\alpha x_0^\alpha}{x^{\alpha+1}} & x > x_0 \\ 0 & \text{sonst} \end{cases}$$

Verteilungsfunktion:

$$F(x|x_0, \alpha) = \begin{cases} 1 - \left(\dfrac{x_0}{x}\right)^\alpha & x > x_0 \\ 0 & \text{sonst} \end{cases}$$

Mittelwert und Varianz:

$$\mathbb{E}(X) = \frac{\alpha x_0}{\alpha - 1} \quad (\text{Vs.: } \alpha > 1) \qquad \text{Var}(X) = \frac{\alpha x_0^2}{(\alpha-1)^2(\alpha-2)} \quad (\text{Vs.: } \alpha > 2)$$

Bem: Vgl. Aufgabe 7.6 für eine Anwendung der Paretoverteilung in der Bayes–Statistik .

Lösungen/Bemerkungen zu ausgewählten Aufgaben

7.3. Nach Anhang A.2.5 gilt für $X \sim Gam(\alpha, \beta)$:

$$\left.\begin{aligned} \mathbb{E}(X) &= \alpha\beta = 10 \\ \mathrm{Var}(X) &= \alpha\beta^2 = 5 \end{aligned}\right\} \quad \Longrightarrow \quad \alpha = 20,\ \beta = \frac{1}{2}$$

7.5. Lösung für (a):

$$\begin{aligned} \pi(\theta|X=3) &\propto \theta^3(1-\theta)^5\, I_{(0,1)}(\theta) \ \ldots\ Be(4,6) \\ \pi(\theta|X=3) &= 504\,\theta^3(1-\theta)^5\, I_{(0,1)}(\theta) \\ \mathbb{E}\big(\widetilde{\theta}\big|X=3\big) &= \frac{\alpha}{\alpha+\beta} = \frac{2}{5} \quad \text{(Bayes–Schätzer)} \end{aligned}$$

7.6. (a) A-posteriori muß θ größer als 5, 3, 8 (Beobachtungen) und 4 (A-priori–Verteilung) sein, also größer als 8:

$$\pi(\theta|D) \ \propto\ \frac{1}{\theta^7}\, I_{(8,\infty)}(\theta)$$

Die normierende Konstante c ergibt sich wie folgt:

$$1 = c\int_8^\infty \frac{1}{\theta^7}\,d\theta = -\frac{c}{6\,\theta^6}\bigg|_8^\infty = \frac{c}{(6)(8^6)} \quad \Longrightarrow \quad c = (6)(8^6)$$

Die A-posteriori–Dichte lautet also wie folgt:

$$\pi(\theta|D) = \begin{cases} \dfrac{(6)(8^6)}{\theta^7} & \text{für } \theta > 8 \\[2ex] 0 & \text{sonst} \end{cases}$$

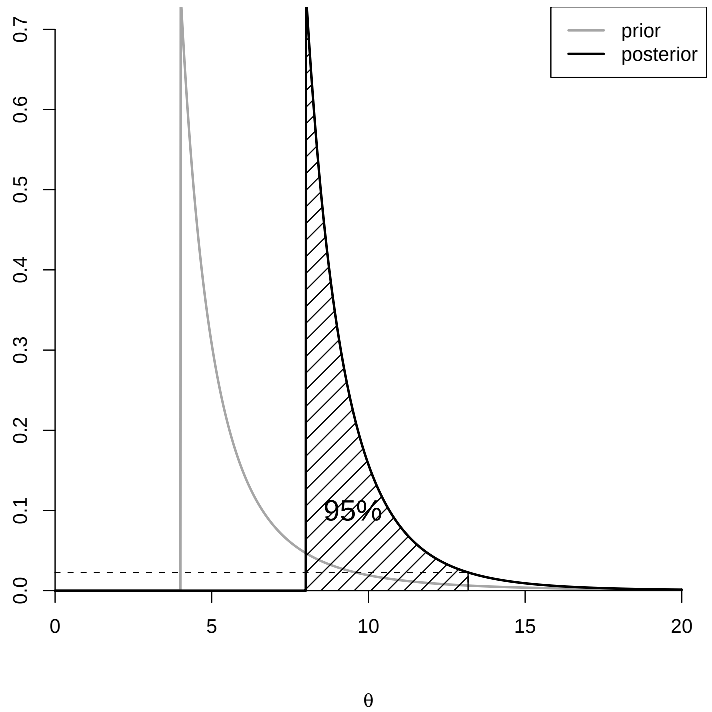
(b) Bayes–Schätzer:

$$\mathbb{E}\big(\widetilde{\theta}\big|D\big) = \int_8^\infty \frac{(6)(8^6)}{\theta^6}\,d\theta = \frac{48}{5} = 9.6 \text{ [Minuten]}$$

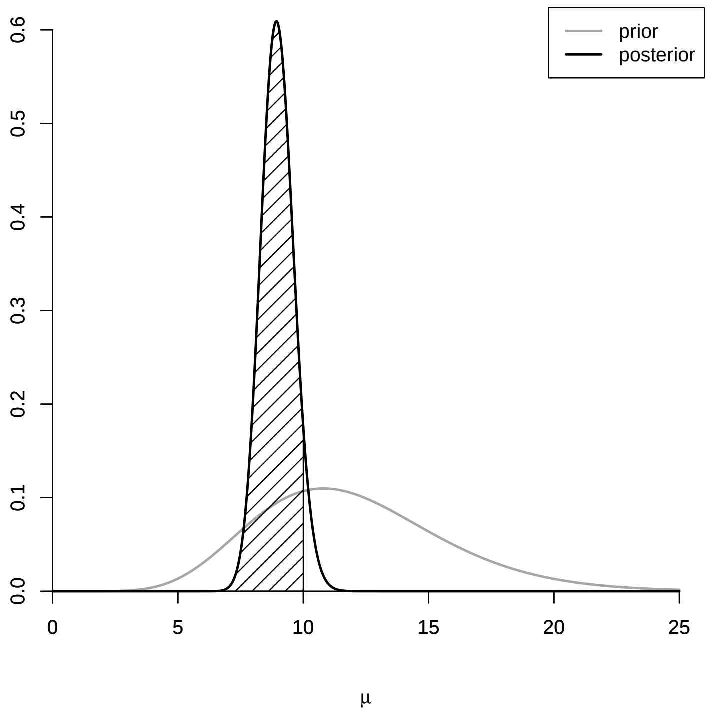
(c) Der HPD–Bereich lautet $(8,b)$, wobei b so zu bestimmen ist, daß:

$$\int_8^b \frac{(6)(8^6)}{\theta^7}\,d\theta = 1 - \left(\frac{8}{b}\right)^6 = 0.95 \quad \Longrightarrow \quad b = \frac{8}{(0.05)^{1/6}} \doteq 13.18$$

(M.a.W., b ist das 95%–Quantil der A-posteriori–Verteilung.)



7.7. Die Abbildung zeigt die A-priori– und A-posteriori–Verteilung (und die Wahrscheinlichkeit von $\mathcal{H}_0$):



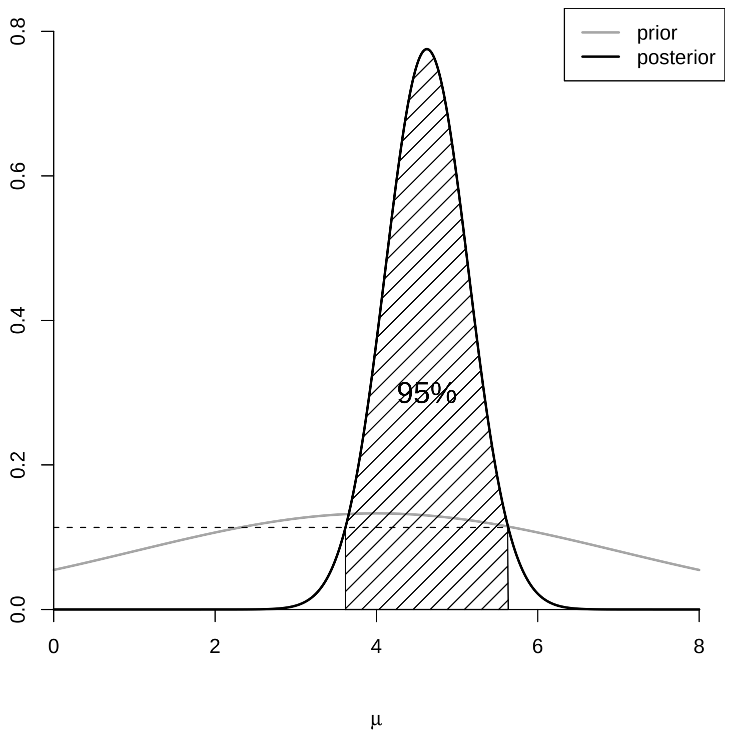
Die A-posteriori–Verteilung findet man wie folgt:

$$\begin{aligned}\pi(\mu|D) &\propto \mu^{\sum_{i=1}^{20} x_i} e^{-20\mu}\, \mu^9 e^{-\mu/1.2} \\ &= \mu^{\sum_{i=1}^{20} x_i + 9} e^{-\mu(20+1/1.2)}\end{aligned}$$

Dies entspricht einer $Gam(\alpha^*, \beta^*)$–Verteilung mit:

$$\alpha^* = \sum_{i=1}^{20} x_i + 10 = 177 + 10 = 187, \quad \beta^* = \frac{1}{20 + 1/1.2} = \frac{1.2}{25} = 0.048$$

7.8. Die Abbildung zeigt die A-priori– und A-posteriori–Verteilung (und das 95% HPD–Intervall):



7.10. $Y|\widetilde{\theta} = \theta$ ist binomialverteilt:

$$p(y|\theta) = \binom{20}{y} \theta^y (1-\theta)^{20-y}, \quad y = 0, 1, \ldots, 20$$

A-posteriori–Wahrscheinlichkeiten:

$$\begin{aligned}\pi(0.1|y) &= \frac{p(y|\theta = 0.1)(0.5)}{p(y|\theta = 0.1)(0.5) + p(y|\theta = 0.3)(0.3) + p(y|\theta = 0.5)(0.2)} \\ \pi(0.3|y) &= \frac{p(y|\theta = 0.3)(0.3)}{p(y|\theta = 0.1)(0.5) + p(y|\theta = 0.3)(0.3) + p(y|\theta = 0.5)(0.2)} \\ \pi(0.5|y) &= 1 - \pi(0.1|y) - \pi(0.3|y)\end{aligned}$$

A-posteriori zu erwartende Verluste (*Risiken*):

$$\begin{aligned} d_1: &\quad 2\pi(0.3|y) + 3\pi(0.5|y) \\ d_2: &\quad \pi(0.1|y) + \pi(0.5|y) \\ d_3: &\quad 3\pi(0.1|y) + 2\pi(0.3|y) \end{aligned}$$

Alle diese Größen lassen sich mit R einfach berechnen:

```
n <- 20
thet <- c(0.1,0.3,0.5)
prior <- c(0.5,0.3,0.2)
nen <- dbinom(0:n, n, thet[1])*prior[1] +
       dbinom(0:n, n, thet[2])*prior[2] +
       dbinom(0:n, n, thet[3])*prior[3]
p0.1y <- dbinom(0:n, n, thet[1])*prior[1] /nen
p0.3y <- dbinom(0:n, n, thet[2])*prior[2] /nen
p0.5y <- dbinom(0:n, n, thet[3])*prior[3] /nen
post <- data.frame(p0.1y=p0.1y, p0.3y=p0.3y, p0.5y=p0.5y)
risk <- data.frame(rho1=2*p0.3y+3*p0.5y, rho2=1*p0.1y+1*p0.5y,
  rho3=3*p0.1y+2*p0.3y)
rownames(post) <- rownames(risk) <- 0:20

round(post,5)
     p0.1y   p0.3y   p0.5y
0  0.99607 0.00392 0.00000
1  0.98501 0.01496 0.00003
2  0.94443 0.05533 0.00024
3  0.81416 0.18398 0.00186
4  0.52848 0.46064 0.01088
5  0.21991 0.73934 0.04075
6  0.06396 0.82939 0.10665
7  0.01515 0.75755 0.22730
8  0.00304 0.58641 0.41055
9  0.00051 0.37952 0.61997
10 0.00007 0.20781 0.79211
11 0.00001 0.10107 0.89892
12 0.00000 0.04597 0.95403
13 0.00000 0.02023 0.97977
14 0.00000 0.00877 0.99123
15 0.00000 0.00378 0.99622
16 0.00000 0.00162 0.99838
17 0.00000 0.00070 0.99930
18 0.00000 0.00030 0.99970
19 0.00000 0.00013 0.99987
20 0.00000 0.00005 0.99995
```

```
round(risk, 5)
      rho1    rho2    rho3
0  0.00785 0.99608 2.99607
1  0.03001 0.98504 2.98496
2  0.11138 0.94467 2.94395
3  0.37354 0.81602 2.81043
4  0.95391 0.53936 2.50673
5  1.60092 0.26066 2.13842
6  1.97873 0.17061 1.85066
7  2.19701 0.24245 1.56055
8  2.40447 0.41359 1.18194
9  2.61895 0.62048 0.76057
10 2.79197 0.79219 0.41584
11 2.89890 0.89893 0.20217
12 2.95402 0.95403 0.09195
13 2.97977 0.97977 0.04047
14 2.99123 0.99123 0.01755
15 2.99622 0.99622 0.00756
16 2.99838 0.99838 0.00325
17 2.99930 0.99930 0.00139
18 2.99970 0.99970 0.00060
19 2.99987 0.99987 0.00026
20 2.99995 0.99995 0.00011
```

Aus der letzten Tabelle kann man die Bayes–Entscheidung ablesen:

$$\delta(y) = \begin{cases} d_1 & \text{wenn } y = 0, 1, 2, 3 \\ d_2 & \text{wenn } y = 4, 5, 6, 7, 8, 9 \\ d_3 & \text{wenn } y = 10, 11, \ldots, 20 \end{cases}$$

A Verteilungen

A.1 Diskrete Verteilungen

A.1.1 Diracverteilung (Kausalverteilung)

Bezeichnung: $X \sim \delta_c$

Parameter: $c \in \mathbb{R}$ (Lageparameter)

Merkmalraum: $M_X = \{c\}$

Wahrscheinlichkeitsfunktion:

$$P\{X = x\} = I_{\{c\}}(x)$$

Mittelwert:

$$\mathbb{E}(X) = c$$

Varianz:

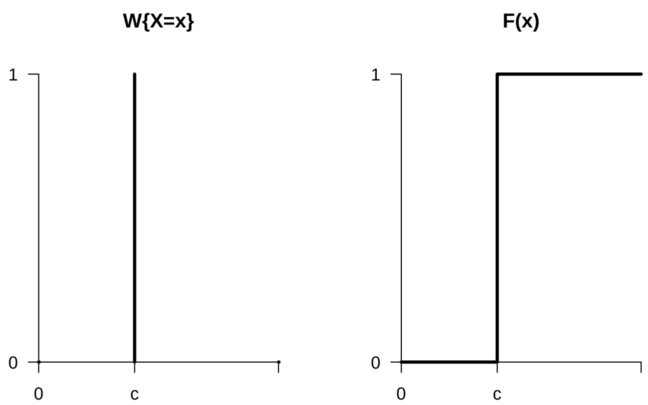
$$\mathrm{Var}(X) = 0$$

Momente:

$$\mathbb{E}(X^k) = c^k, \quad k = 1, 2, \ldots$$

Momentenerzeugende Funktion:

$$m(t) = e^{tc}, \quad t \in \mathbb{R}$$



A.1.2 Uniforme Verteilung (Gleichverteilung)

Bezeichnung: $X \sim D_N$, $D(N)$

Parameter: $N \in \mathbb{N}$

Merkmalraum: $M_X = \{1, 2, \ldots, N\}$

Wahrscheinlichkeitsfunktion:

$$P\{X = x\} = \frac{1}{N} I_{\{1,2,\ldots,N\}}(x)$$

Mittelwert:

$$\mathbb{E}(X) = \frac{N+1}{2}$$

Varianz:

$$\mathrm{Var}(X) = \frac{N^2 - 1}{12}$$

Höhere Momente:

$$\begin{aligned}
\mathbb{E}(X^3) &= \frac{N(N+1)^2}{4} \\
\mathbb{E}(X^4) &= \frac{(N+1)(2N+1)(3N^2+3N-1)}{30}
\end{aligned}$$

Momentenerzeugende Funktion:

$$m(t) = \frac{1}{N} \sum_{k=1}^{N} e^{tk}, \quad t \in \mathbb{R}$$

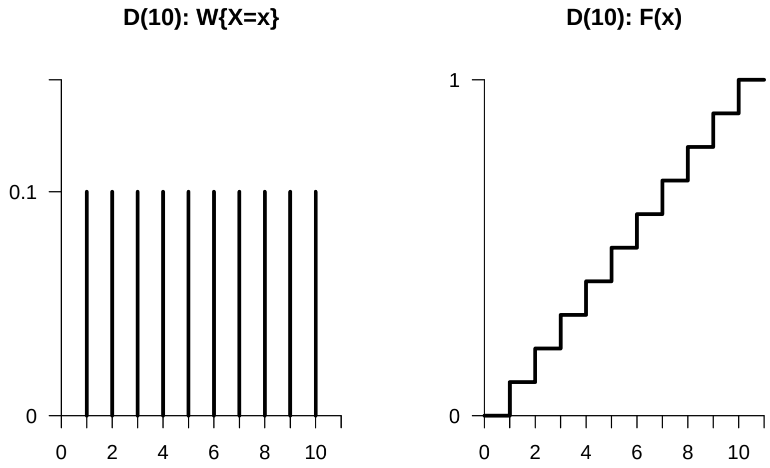
Verallgemeinerung: Diskrete uniforme Verteilung auf $M_X = \{x_1, x_2, \ldots, x_N\}$: $X \sim D_{M_X}$

$$P\{X = x\} = \frac{1}{N} I_{\{x_1,x_2,\ldots,x_N\}}(x)$$

$$\mathbb{E}(X) = \frac{1}{N} \sum_{k=1}^{N} x_k = \overline{x}, \qquad \mathrm{Var}(X) = \frac{1}{N} \sum_{k=1}^{N} (x_k - \overline{x})^2$$

R–Funktionen: Zwei Beispiele zur Erzeugung von diskret uniform verteilten Zufallszahlen:

```
x <- sample(x=10, size=5, replace=TRUE) # N=10
x <- sample(x=c(-1,0,1,2,3,4,5,10,20), size=5, replace=TRUE) # Verallg.
```


D(10): W{X=x}
0.1
0
0 2 4 6 8 10
D(10): F(x)
1
0
0 2 4 6 8 10

A.1.3 Alternativverteilung (Bernoulliverteilung)

Bezeichnung: $X \sim A_p,\ A(p)$

Parameter: $p \in (0,1)$ (Formparameter), $q := 1-p$

Merkmalraum: $M_X = \{0,1\}$

Wahrscheinlichkeitsfunktion:

$$P\{X=x\} = p^x(1-p)^{1-x} = p^x q^{1-x}, \quad x = 0,1$$

Mittelwert:

$$\mathbb{E}(X) = p$$

Varianz:

$$\mathrm{Var}(X) = p(1-p) = pq$$

Momente:

$$\mathbb{E}(X^k) = p, \quad k = 1,2,\ldots$$

Momentenerzeugende Funktion:

$$m(t) = 1-p+pe^t = q+pe^t, \quad t \in \mathbb{R}$$

Modus:

$$x_{\text{mod}} = \begin{cases} 0 & p \leq 0.5 \\ 1 & p \geq 0.5 \end{cases}$$

Additionstheorem: $X_i \sim A_p,\ i = 1,\ldots,n,$ ua.:

$$\sum_{i=1}^n X_i \ \sim\ B_{n,p}$$

ZGVS: $X_i \sim A_p,\ i = 1,\ldots,n,$ ua.; für $np(1-p) > 9$ gilt in guter Näherung:

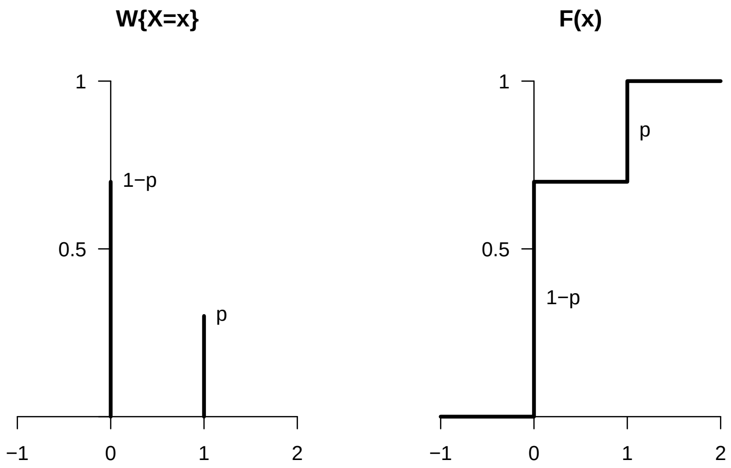
$$\sum_{i=1}^n X_i \ \dot{\sim}\ N\big(np, np(1-p)\big)$$

<u>R–Funktionen</u>: `prob` $\widehat{=}\, p$

```
dbinom(x, size=1, prob, log = FALSE)
pbinom(q, size=1, prob, lower.tail = TRUE, log.p = FALSE)
qbinom(p, size=1, prob, lower.tail = TRUE, log.p = FALSE)
rbinom(n, size=1, prob)
```

Alternative zu `rbinom`:

```
x <- sample(0:1, size, replace=TRUE, prob)
```



A.1.4 Binomialverteilung

Bezeichnung: $X \sim B_{n,p}$, $B(n,p)$

Parameter: $n \in \mathbb{N}$, $p \in (0,1)$ (Formparameter), $q := 1 - p$

Merkmalraum: $M_X = \{0, 1, \ldots, n\}$

Wahrscheinlichkeitsfunktion:

$$P\{X = x\} = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

Mittelwert:

$$\mathbb{E}(X) = np$$

Varianz:

$$\mathrm{Var}(X) = np(1-p) = npq$$

Höhere zentrale Momente:

$$\mathbb{E}\big[X - \mathbb{E}(X)\big]^3 = npq(q-p)$$
$$\mathbb{E}\big[X - \mathbb{E}(X)\big]^4 = 3n^2p^2q^2 + npq(1 - 6pq)$$

Momentenerzeugende Funktion:

$$m(t) = (1 - p + pe^t)^n = (q + pe^t)^n, \quad t \in \mathbb{R}$$

Modus:

$$x_{\mathrm{mod}} = \begin{cases} \lfloor (n+1)p \rfloor & (n+1)p \notin \mathbb{N} \\ (n+1)p - 1,\ (n+1)p & (n+1)p \in \mathbb{N} \end{cases}$$

Spezialfall: $B_{1,p} \equiv A_p$

Additionstheorem: $X_k \sim B_{n_k,p}$, $k = 1, 2, \ldots, K$, ua.:

$$\sum_{k=1}^{K} X_k \sim B_{n,p}, \quad n = \sum_{k=1}^{K} n_k$$

Poissonapproximation: Für $n \geq 50$, $p \leq 1/10$, $np \leq 10$ gilt in guter Näherung:

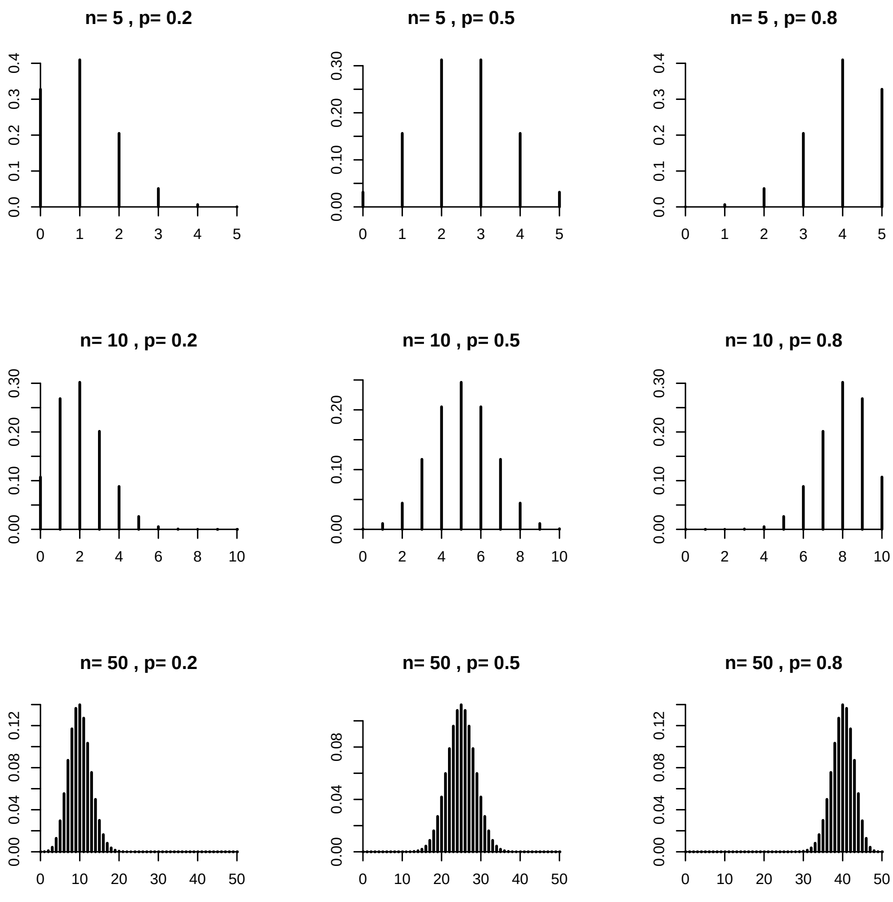
$$B_{n,p} \;\approx\; P_\mu,\; \mu = np: \quad P\{X = x\} \;\approx\; \frac{\mu^x e^{-\mu}}{x!}$$

ZGVS: Für $np(1-p) > 9$ gilt in guter Näherung ($a, b, x \in \{0, 1, \ldots, n\}$, $a < b$):

$$P\{X = x\} \;\approx\; \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$
$$P\{a \leq X \leq b\} \;\approx\; \Phi\left(\frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

R–Funktionen: `size` $\widehat{=}\, n$, `prob` $\widehat{=}\, p$

```
choose(n, k)  # Binomialkoeff.
dbinom(x, size, prob, log = FALSE)
pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
rbinom(n, size, prob)
```



A.1.5 Hypergeometrische Verteilung

Bezeichnung: $X \sim H_{N,A,n}$, $H(N,A,n)$

Parameter: $N, A, n \in \mathbb{N}$ (Formparameter), $A \leq N$, $n \leq N$, $p := A/N$, $q := 1-p$

Merkmalraum: $M_X = \{a_1, \ldots, a_2\}$, $a_1 = \max\{0, n-(N-A)\}$, $a_2 = \min\{n, A\}$

Wahrscheinlichkeitsfunktion:

$$P\{X = x\} = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}, \quad x = a_1, \ldots, a_2$$

Mittelwert:

$$\mathbb{E}(X) = n\,\frac{A}{N}$$

Varianz:

$$\mathrm{Var}(X) = n\,\frac{A}{N}\left(1 - \frac{A}{N}\right)\frac{N-n}{N-1}$$

Faktorielle Momente:

$$\mathbb{E}\big[X(X-1)\cdots(X-k+1)\big] = k!\,\frac{\binom{A}{k}\binom{n}{k}}{\binom{N}{k}}, \quad k \in \mathbb{N}$$

Modus:

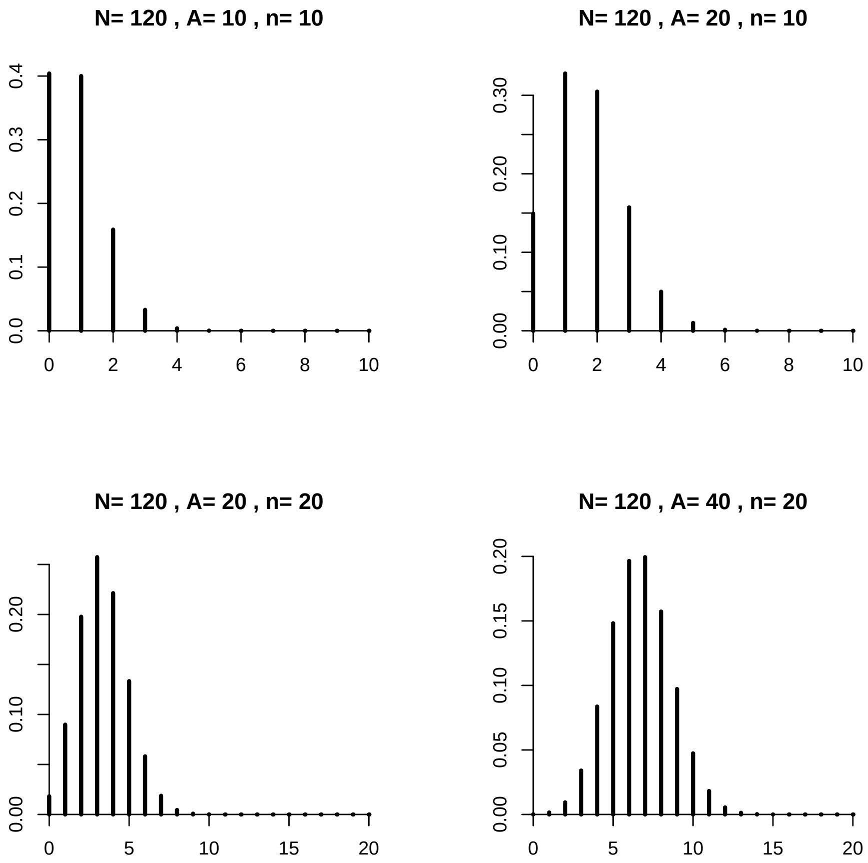
$$x_{\mathrm{mod}} = \begin{cases} \left\lfloor (n+1)\dfrac{A+1}{N+2} \right\rfloor & (n+1)\dfrac{A+1}{N+2} \notin \mathbb{N} \\[2ex] (n+1)\dfrac{A+1}{N+2} - 1,\ (n+1)\dfrac{A+1}{N+2} & (n+1)\dfrac{A+1}{N+2} \in \mathbb{N} \end{cases}$$

Binomialapproximation: Für A, $N-n$ groß und $n/N \leq 0.05$ gilt in guter Näherung:

$$H_{N,A,n} \;\approx\; B_{n,p}: \quad P\{X = x\} \;\approx\; \binom{n}{x}\left(\frac{A}{N}\right)^x \left(1 - \frac{A}{N}\right)^{n-x}$$

<u>R–Funktionen</u>: m $\widehat{=} A$, n $\widehat{=} N - A$, k $\widehat{=} n$

```
dhyper(x, m, n, k, log = FALSE)
phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)
qhyper(p, m, n, k, lower.tail = TRUE, log.p = FALSE)
rhyper(nn, m, n, k)
```



A.1.6 Poissonverteilung

Bezeichnung: $X \sim P_\mu,\ P(\mu)$

Parameter: $\mu \in \mathbb{R}^+$ (Formparameter)

Merkmalraum: $M_X = \mathbb{N}_0 = \{0, 1, 2, \dots\}$

Wahrscheinlichkeitsfunktion:

$$P\{X = x\} = \frac{\mu^x e^{-\mu}}{x!}, \quad x = 0, 1, 2, \dots$$

Mittelwert:

$$\mathbb{E}(X) = \mu$$

Varianz:

$$\mathrm{Var}(X) = \mu$$

Höhere zentrale Momente:

$$\mathbb{E}\big[X - \mathbb{E}(X)\big]^3 = \mu$$
$$\mathbb{E}\big[X - \mathbb{E}(X)\big]^4 = \mu + 3\mu^2$$

Momentenerzeugende Funktion:

$$m(t) = \exp\big[\mu(e^t - 1)\big], \quad t \in \mathbb{R}$$

Modus:

$$x_{\mathrm{mod}} = \begin{cases} \lfloor \mu \rfloor & \mu \notin \mathbb{N} \\ \mu - 1,\ \mu & \mu \in \mathbb{N} \end{cases}$$

Additionstheorem: $X_k \sim P_{\mu_k},\ k = 1, 2, \dots, K,$ ua.:

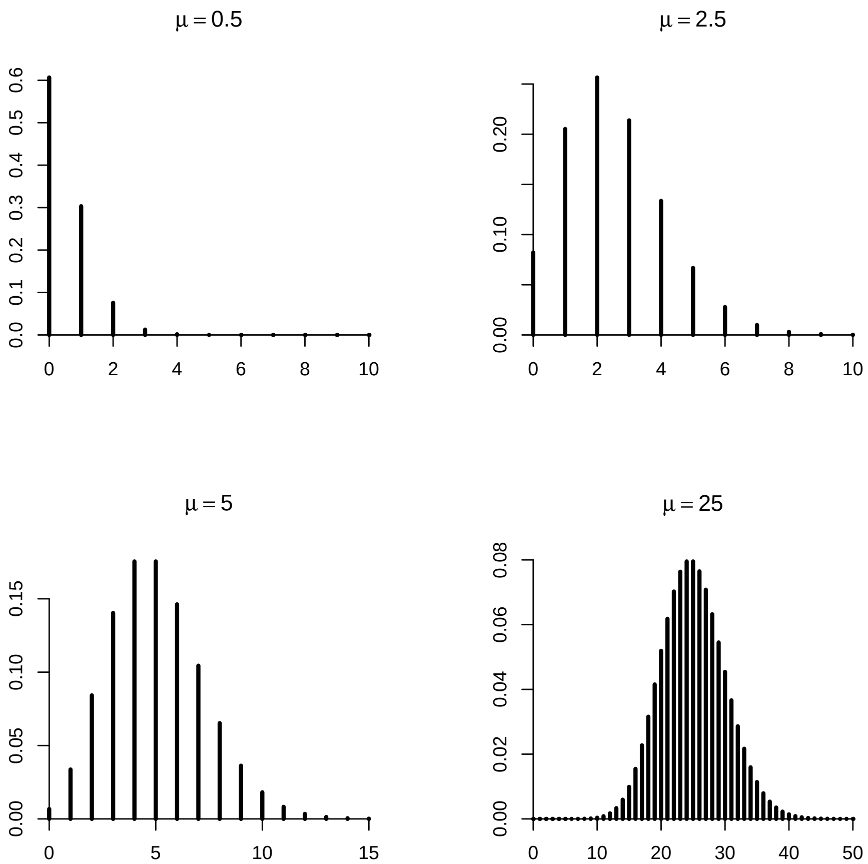
$$\sum_{k=1}^{K} X_k \sim P_\mu, \quad \mu = \sum_{k=1}^{K} \mu_k$$

<u>ZGVS</u>: Für $\mu > 9$ gilt in guter Näherung ($a, b, x \in \mathbb{N}_0$, $a < b$):

$$P\{X = x\} \approx \Phi\left(\frac{x + 0.5 - \mu}{\sqrt{\mu}}\right) - \Phi\left(\frac{x - 0.5 - \mu}{\sqrt{\mu}}\right)$$
$$P\{a \le X \le b\} \approx \Phi\left(\frac{b + 0.5 - \mu}{\sqrt{\mu}}\right) - \Phi\left(\frac{a - 0.5 - \mu}{\sqrt{\mu}}\right)$$

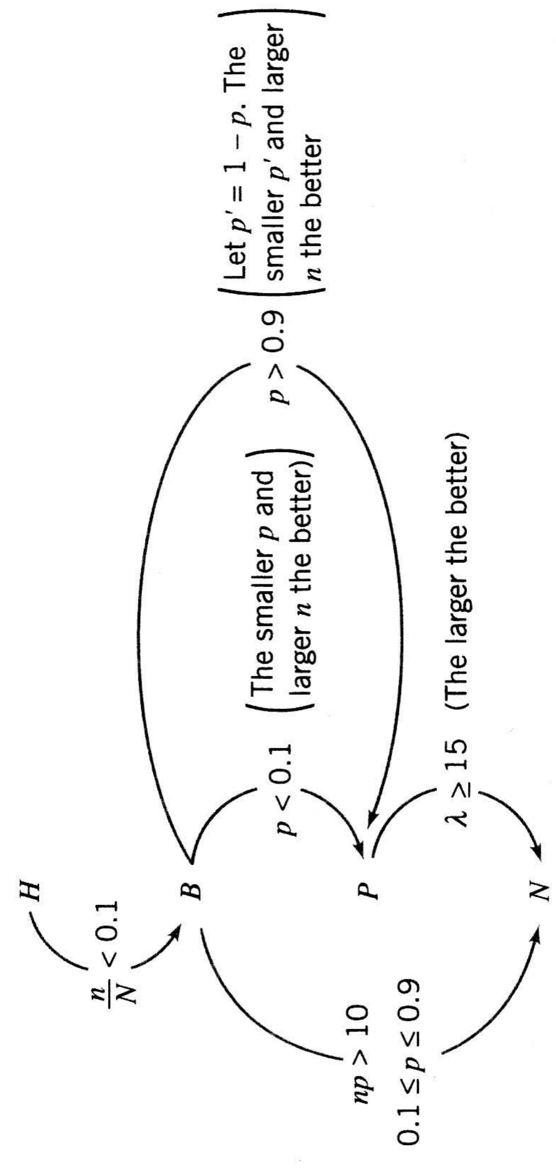
<u>R–Funktionen</u>: `lambda` $\widehat{=}\, \mu$

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```



Zusatz: Die folgende Übersicht (aus D.C. Montgomery: *Introduction to Statistical Quality Control*, 5e, 2005) zeigt einige wichtige Approximationen von diskreten Verteilungen und ihre Bedingungen. Dabei werden die folgenden Bezeichnungen verwendet:

$$H = \text{Hypergeometrische Verteilung}$$
$$B = \text{Binomialverteilung}$$
$$P = \text{Poissonverteilung } (\lambda \equiv \mu)$$
$$N = \text{Normalverteilung}$$



A.1.7 Geometrische Verteilung (Pascalverteilung)

Bezeichnung: $X \sim G_p$, $G(p)$

Parameter: $p \in (0,1)$, $q := 1-p$

Merkmalraum: $M_X = \mathbb{N}$ (1. Version); $M_{X'} = \mathbb{N}_0$ (2. Version)

Wahrscheinlichkeitsfunktion:

1. Version: X = Zahl der Versuche bis zum ersten Erfolg

$$P\{X = x\} = p(1-p)^{x-1} = pq^{x-1}, \quad x = 1, 2, \ldots$$

2. Version: $X^{'} = X - 1$ = Zahl der Mißerfolge vor dem ersten Erfolg

$$P\{X^{'} = x\} = p(1-p)^{x} = pq^{x}, \quad x = 0, 1, 2, \ldots$$

Mittelwert:

$$\mathbb{E}(X) = \frac{1}{p}, \qquad \mathbb{E}(X^{'}) = \mathbb{E}(X) - 1 = \frac{1-p}{p} = \frac{q}{p}$$

Varianz:

$$\mathrm{Var}(X) = \mathrm{Var}(X^{'}) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Momentenerzeugende Funktion:

$$m_X(t) = \frac{p\,e^t}{1-qe^t}, \quad t < -\ln q$$
$$m_{X'}(t) = \frac{p}{1-qe^t}, \quad t < -\ln q$$

Modus:

$$X: \; x_{\mathrm{mod}} = 1, \qquad X^{'}: \; x^{'}_{\mathrm{mod}} = 0$$

Additionstheorem: $X_k \sim G_p, \; k = 1, 2, \ldots, K,$ ua.:

$$\sum_{k=1}^{K} X_k \sim NB_{K,p}$$

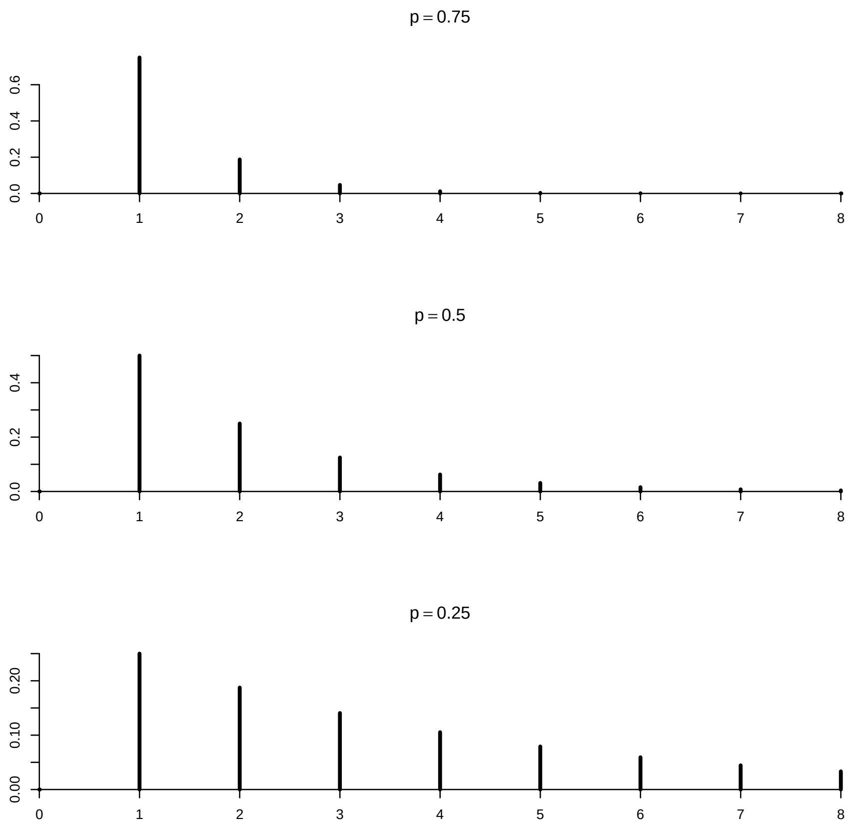
Gedächtnislosigkeit:

$$P\{X > a + b|X > a\} = P\{X > b\}, \quad a, b \in \mathbb{N}$$

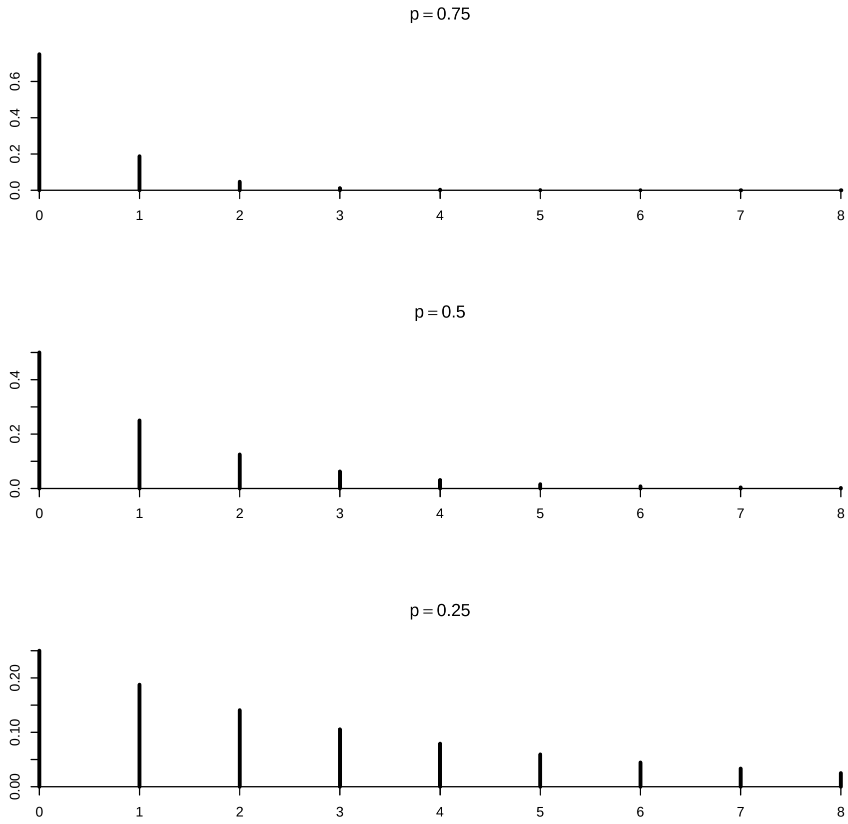
R–Funktionen: Die Funktionengruppe `geom` entspricht der 2. Version der geometrischen Verteilung.

```
dgeom(x, prob, log = FALSE)
pgeom(q, prob, lower.tail = TRUE, log.p = FALSE)
qgeom(p, prob, lower.tail = TRUE, log.p = FALSE)
rgeom(n, prob)
```

1. Version



2. Version


p = 0.75
0.0
0.2
0.4
0.6
0
1
2
3
4
5
6
7
8
p = 0.5
0.0
0.2
0.4
0
1
2
3
4
5
6
7
8
p = 0.25
0.00
0.10
0.20
0
1
2
3
4
5
6
7
8

A.1.8 Negative Binomialverteilung

Bezeichnung: $X \sim NB_{r,p}$, $NB(r,p)$

Parameter: $p \in (0,1)$, $q := 1-p$, $r \in \mathbb{N}$ (Formparameter)

Merkmalraum: $M_X = \{r, r+1, \ldots\}$ (1. Version); $M_{X'} = \mathbb{N}_0$ (2. Version)

Wahrscheinlichkeitsfunktion:

1. Version: X = Zahl der Versuche bis zum r–ten Erfolg

$$P\{X = x\} = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, \ldots$$

2. Version: $X^{'} = X - r$ = Zahl der Mißerfolge vor dem r–ten Erfolg

$$P\{X^{'} = x\} = \binom{r+x-1}{r-1} p^r q^x, \quad x = 0, 1, \ldots$$

Mittelwert:

$$\mathbb{E}(X) = \frac{r}{p}, \qquad \mathbb{E}(X^{'}) = \mathbb{E}(X) - r = \frac{r(1-p)}{p} = \frac{rq}{p}$$

Varianz:

$$\operatorname{Var}(X) = \operatorname{Var}(X^{'}) = \frac{r(1-p)}{p^2} = \frac{rq}{p^2}$$

Momentenerzeugende Funktion:

$$m_X(t) = \left(\frac{p\,e^t}{1-qe^t}\right)^r, \quad t < -\ln q$$

$$m_{X'}(t) = \left(\frac{p}{1-qe^t}\right)^r, \quad t < -\ln q$$

Modus:

$$X: \; x_{\text{mod}} = \begin{cases} \left\lfloor \dfrac{r-1}{p} \right\rfloor + 1 & \dfrac{r-1}{p} \notin \mathbb{N} \\[2ex] \dfrac{r-1}{p}, \; \dfrac{r-1}{p} + 1 & \dfrac{r-1}{p} \in \mathbb{N} \end{cases} \qquad (r > 1)$$

$$X^{'}: \; x^{'}_{\text{mod}} = x_{\text{mod}} - r$$

Spezialfall: $NB_{1,p} \equiv G_p$

Additionstheorem: $X_k \sim NB_{r_k,p},\ k = 1, 2, \ldots, K,$ ua.:

$$\sum_{k=1}^{K} X_k \sim NB_{r,p}, \quad r = \sum_{k=1}^{K} r_k$$

Beziehung zur Binomialverteilung: Für $X \sim NB_{r,p}$ (1. Version) und $Y \sim B_{x,p}$, $x \in \mathbb{N}$, gilt:

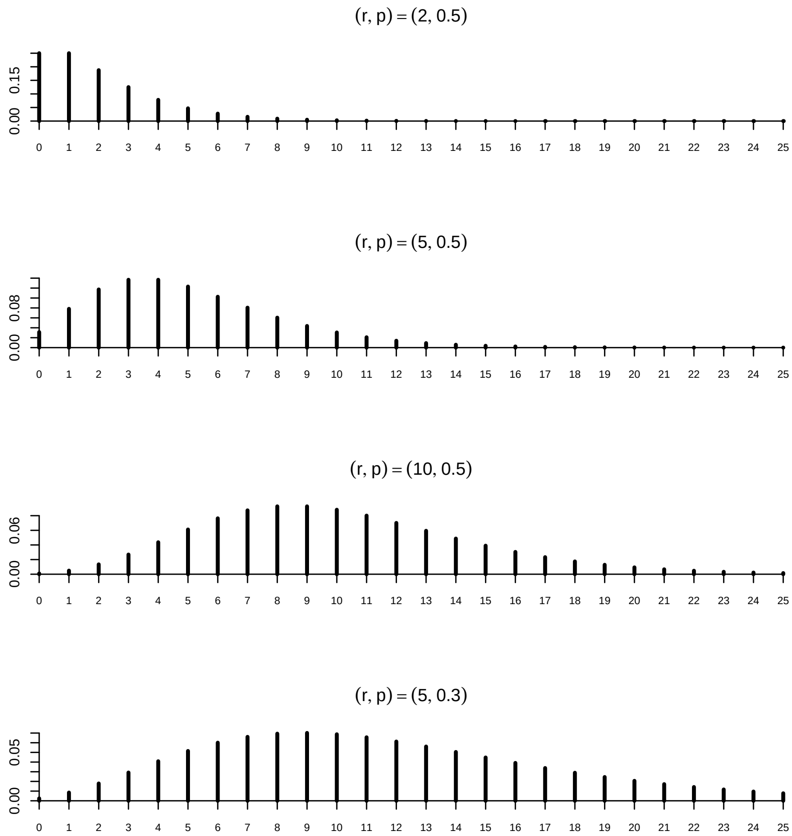
$$P\{X > x\} = P\{Y < r\}$$

R–Funktionen: Die Funktionengruppe `nbinom` entspricht der 2. Version der negativen Binomialverteilung.

`size` $\widehat{=}\, r$, `mu` $\widehat{=}\, r(1-p)/p =$ Mittelwert (alternative Parametrisierung)

```
dnbinom(x, size, prob, mu, log = FALSE)
pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
rnbinom(n, size, prob, mu)
```

2. Version



A.2 Stetige Verteilungen

A.2.1 Uniforme Verteilung (Gleichverteilung)

Bezeichnung: $X \sim U_{a,b}$, $U(a,b)$

Parameter: $a, b \in \mathbb{R}$, $a < b$

Merkmalraum: $M_X = (a,b)$, $M_X = [a,b]$

Dichte:

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x)$$

Verteilungsfunktion:

$$F(x) = \frac{x-a}{b-a} I_{(a,b)}(x) + I_{[b,\infty)}(x)$$

Mittelwert:

$$\mathbb{E}(X) = \frac{a+b}{2}$$

Varianz:

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Zentrale Momente:

$$\mathbb{E}\big[X - \mathbb{E}(X)\big]^k = \begin{cases} 0 & k = 1, 3, \ldots \\ \dfrac{(b-a)^k}{2^k(k+1)} & k = 2, 4, \ldots \end{cases}$$

Momentenerzeugende Funktion:

$$m(t) = \frac{e^{bt} - e^{at}}{(b-a)t}, \quad t \in \mathbb{R} \quad [m(0) = 1]$$

Quantile:

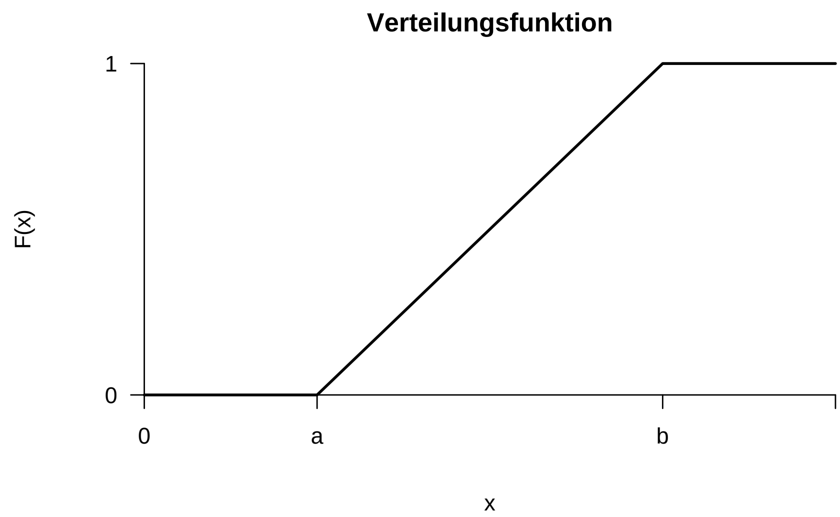
$$x_p = a + (b-a)p, \quad 0 < p < 1$$

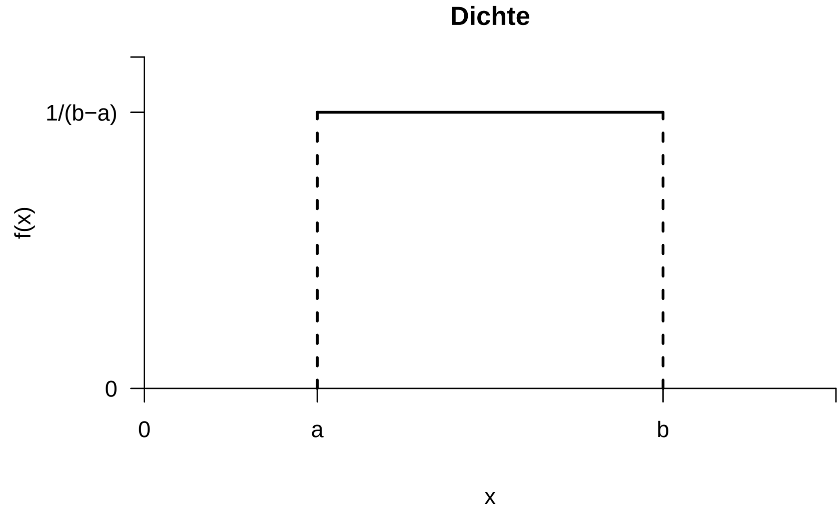
Integraltransformation: Ist F_X die Verteilungsfunktion einer stetigen sG X, so gilt:

$$U = F_X(X) \sim U_{0,1}$$

R–Funktionen: `min` $\widehat{=}$ a, `max` $\widehat{=}$ b

```
dunif(x, min=0, max=1, log = FALSE)
punif(q, min=0, max=1, lower.tail = TRUE, log.p = FALSE)
qunif(p, min=0, max=1, lower.tail = TRUE, log.p = FALSE)
runif(n, min=0, max=1)
```





A.2.2 Exponentialverteilung

Bezeichnung: $X \sim Ex_\tau$, $Ex(\tau)$, Ex_λ, $Ex(\lambda)$

Parameter: $\tau > 0$ (Skalierungsparameter), $\lambda := 1/\tau$ (Ausfallrate)

Merkmalraum: $M_X = (0, \infty)$, $M_X = [0, \infty)$

Dichte:

$$f(x) = \frac{1}{\tau} e^{-x/\tau} I_{(0,\infty)}(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$$

Verteilungsfunktion:

$$F(x) = \big(1 - e^{-x/\tau}\big) I_{(0,\infty)}(x) = \big(1 - e^{-\lambda x}\big) I_{(0,\infty)}(x)$$

Mittelwert:

$$\mathbb{E}(X) = \tau = \frac{1}{\lambda}$$

Varianz:

$$\mathrm{Var}(X) = \tau^2 = \frac{1}{\lambda^2}$$

Momente:

$$\mathbb{E}\big(X^k\big) = \tau^k k! = \frac{k!}{\lambda^k}, \quad k = 1, 2, \ldots$$

Momentenerzeugende Funktion:

$$m(t) = \frac{1}{1 - \tau t} = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

Quantile:

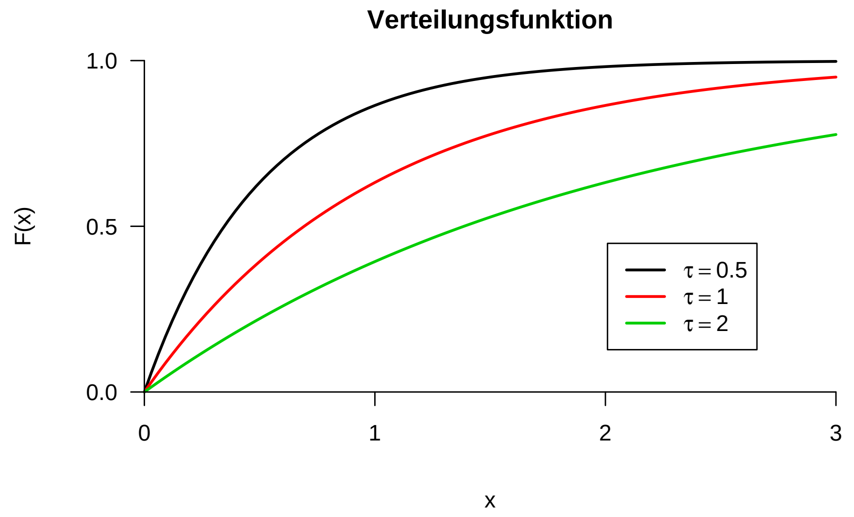
$$x_p = -\tau \ln(1 - p) = -\frac{\ln(1 - p)}{\lambda}, \quad 0 < p < 1$$

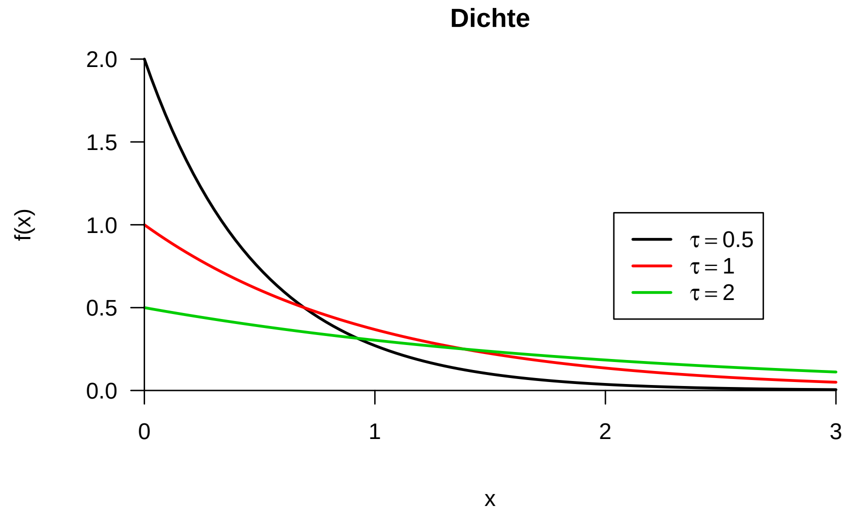
Additionsthorem: $X_i \sim Ex_\tau$, $i = 1, 2, \ldots, n$, ua.:

$$\sum_{i=1}^{n} X_i \sim Er_{n,\tau} \quad \text{(Erlangverteilung)}$$

<u>R–Funktionen</u>: `rate` $\widehat{=}\ \lambda$

```
dexp(x, rate = 1, log = FALSE)
pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE)
qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)
rexp(n, rate = 1)
```





A.2.3 Normalverteilung (Gaußverteilung)

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

Parameter: $\mu \in \mathbb{R}$ (Lageparameter), $\sigma > 0$ (Skalierungsparameter)

Merkmalraum: $M_X = \mathbb{R}$

Dichte:

$$N(0,1): \quad \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \quad x \in \mathbb{R}$$

$$N(\mu,\sigma^2): \quad f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$$

Verteilungsfunktion:

$$N(0,1): \quad \Phi(x) = \int_{-\infty}^{x} \varphi(u)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du, \quad x \in \mathbb{R}$$

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

Mittelwert:

$$\mathbb{E}(X) = \mu$$

Varianz:

$$\mathrm{Var}(X) = \sigma^2$$

Zentrale Momente:

$$\mathbb{E}\big[X - \mathbb{E}(X)\big]^k = \begin{cases} 0 & k = 1,3,\ldots \\ \dfrac{k!}{(k/2)!}\,\dfrac{\sigma^k}{2^{k/2}} & k = 2,4,\ldots \end{cases}$$

Momentenerzeugende Funktion:

$$m(t) = e^{\mu t + \sigma^2 t^2/2}, \quad t \in \mathbb{R}$$

Quantile:

$$N(0,1): \quad u_p \quad \left[\text{häufig auch } z_p\right]$$

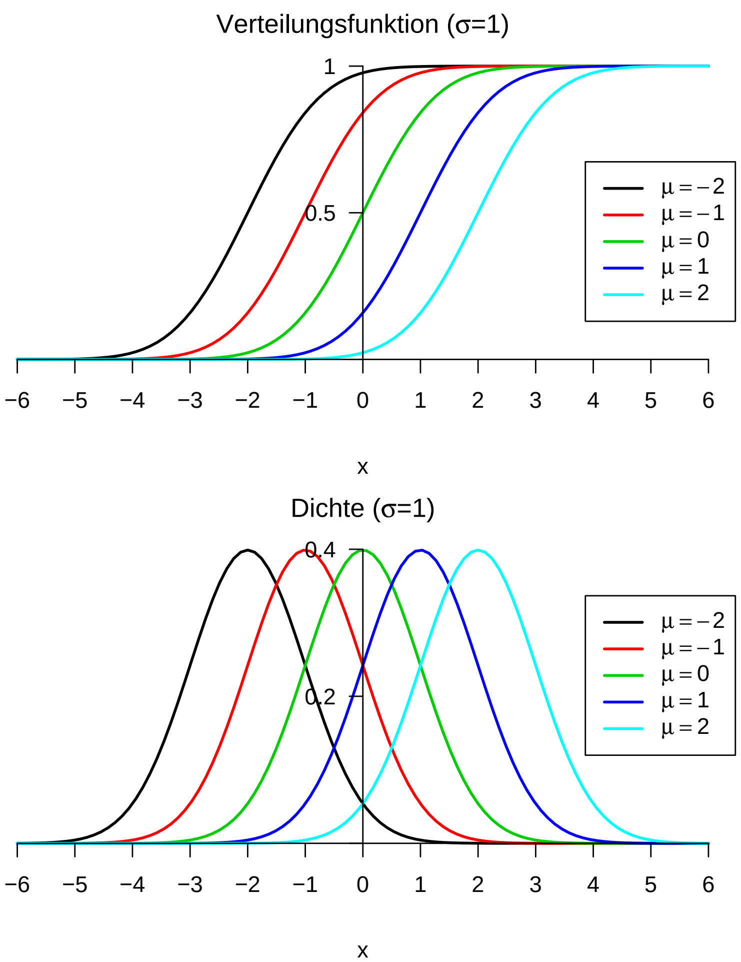
$$N(\mu,\sigma^2): \quad x_p = \mu + \sigma u_p$$

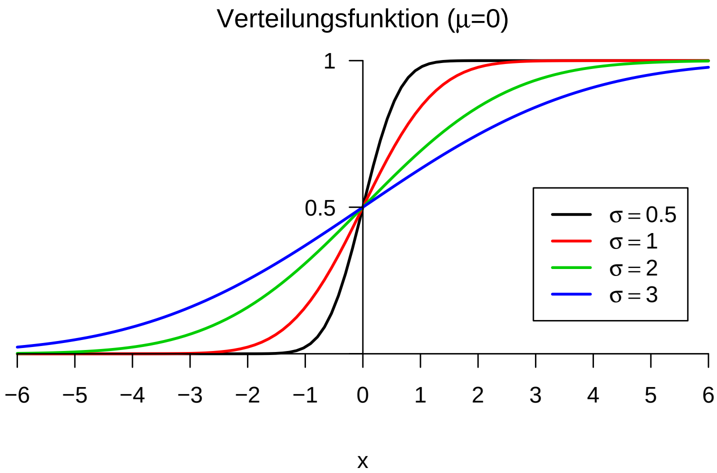
Additionstheorem: $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, ua.; $c_i \in \mathbb{R}$:

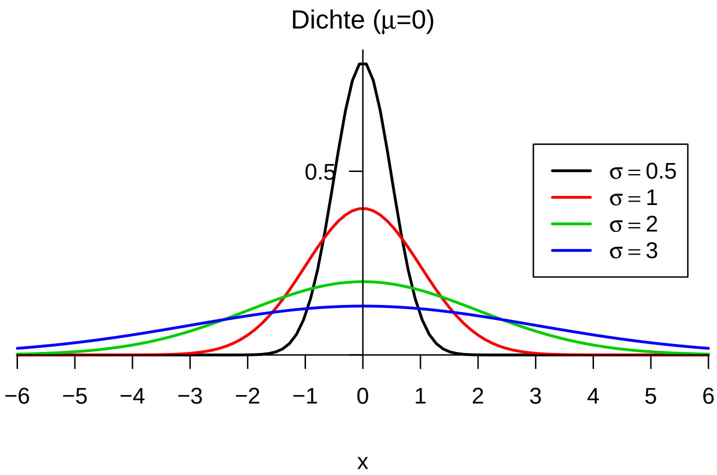
$$\sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

R–Funktionen: `mean` $\widehat{=}\ \mu$, `sd` $\widehat{=}\ \sigma$

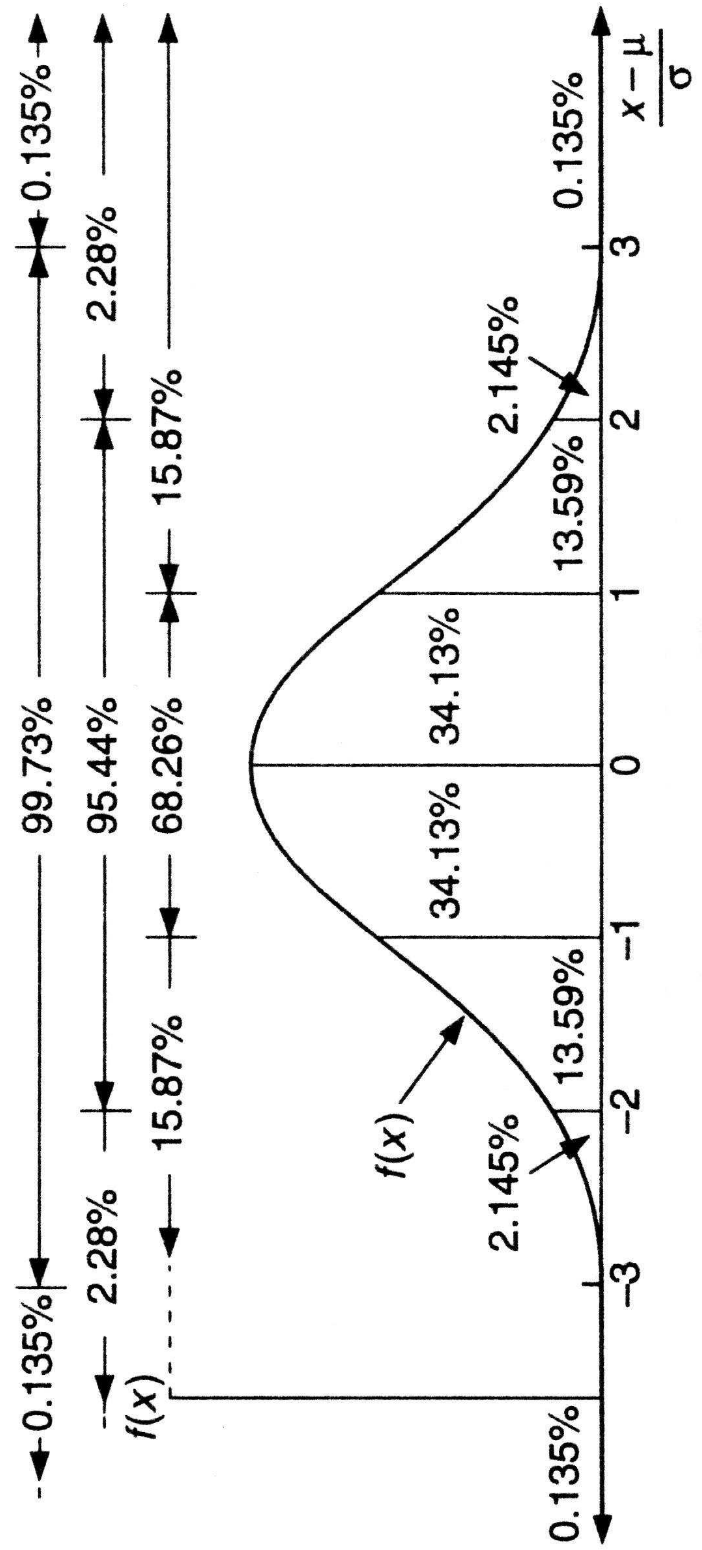
```
dnorm(x, mean=0, sd=1, log = FALSE)
pnorm(q, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean=0, sd=1)
```




Verteilungsfunktion (μ=0)
1
0.5
σ = 0.5
σ = 1
σ = 2
σ = 3
−6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6
x


Dichte (μ=0)
0.5
σ = 0.5
σ = 1
σ = 2
σ = 3
−6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6
x

Zusatz: Die folgende Übersicht (aus D. Bissell: *Statistical Methods for SPC and TQM*, 1994) zeigt eine Reihe von häufig benötigten Wahrscheinlichkeiten im Zusammenhang mit der Normalverteilung.



A.2.4 Logarithmische Normalverteilung (Log–Normalverteilung)

Bezeichnung: $X \sim LN(\mu, \sigma^2)$, $L(\mu, \sigma^2)$

Parameter: $\mu \in \mathbb{R}$ (e^μ Skalierungsparameter), $\sigma > 0$ (Formparameter)

Merkmalraum: $M_X = \mathbb{R}^+$

Dichte:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-(\ln x - \mu)^2/2\sigma^2}, \quad x > 0$$

Verteilungsfunktion:

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \quad x > 0$$

Mittelwert:

$$\mathbb{E}(X) = e^{\mu + \sigma^2/2}$$

Varianz:

$$\mathrm{Var}(X) = e^{2\mu + \sigma^2}\left[e^{\sigma^2} - 1\right]$$

Momente:

$$\mathbb{E}(X^k) = e^{k\mu + k^2\sigma^2/2}, \quad k = 1, 2, \ldots$$

Quantile:

$$x_p = e^{\mu + \sigma u_p}, \quad 0 < p < 1$$

Modus:

$$x_{\mathrm{mod}} = e^{\mu - \sigma^2}$$

Produkttheorem: $X_i \sim LN(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, ua.:

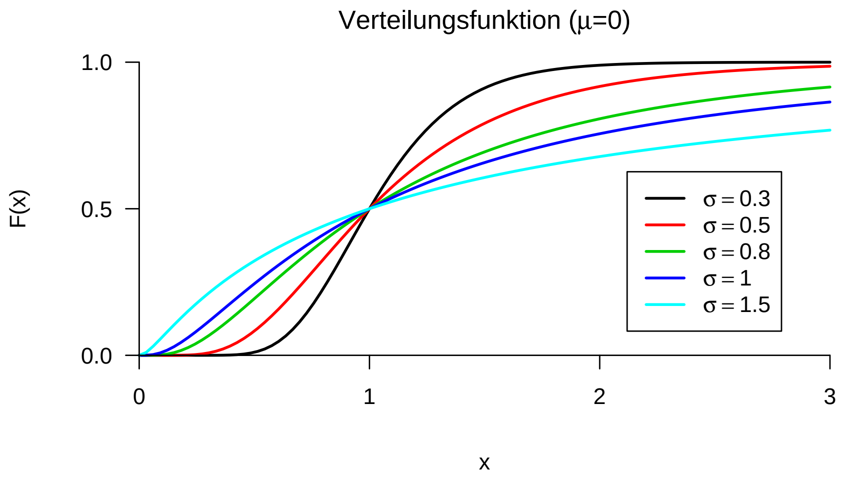
$$\prod_{i=1}^n X_i \sim LN\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

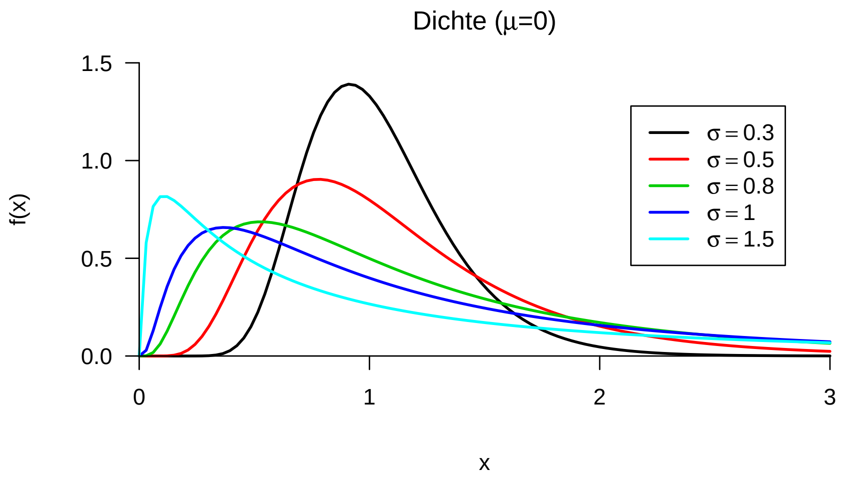
Beziehung zur Normalverteilung:

$$X \sim LN(\mu, \sigma^2) \quad \Longrightarrow \quad Y = \ln X \sim N(\mu, \sigma^2)$$
$$Y \sim N(\mu, \sigma^2) \quad \Longrightarrow \quad X = e^Y \sim LN(\mu, \sigma^2)$$

R–Funktionen: `meanlog` $\widehat{=}$ μ, `sdlog` $\widehat{=}$ σ

```
dlnorm(x, meanlog = 0, sdlog = 1, log = FALSE)
plnorm(q, meanlog = 0, sdlog = 1, lower.tail = TRUE, log.p = FALSE)
qlnorm(p, meanlog = 0, sdlog = 1, lower.tail = TRUE, log.p = FALSE)
rlnorm(n, meanlog = 0, sdlog = 1)
```





A.2.5 Gammaverteilung

Bezeichnung: $X \sim \gamma(\alpha, \beta)$, $Gam(\alpha, \beta)$

Parameter: $\alpha > 0$ (Formparameter), $\beta > 0$ (Skalierungsparameter)

Merkmalraum: $M_X = \mathbb{R}^+$

Gammafunktion:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt, \quad x > 0$$

$$\Gamma(x+1) = x\Gamma(x), \quad n \in \mathbb{N} : \Gamma(n) = (n-1)!, \quad \Gamma(1/2) = \sqrt{\pi}$$

Dichte:

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\, \beta^\alpha}, \quad x > 0$$

Verteilungsfunktion:

$$F(x) = \frac{1}{\Gamma(\alpha)\, \beta^\alpha} \int_0^x t^{\alpha-1} e^{-t/\beta}\, dt, \quad x > 0 \qquad \text{[unvollständige Gammafunktion]}$$

Mittelwert:

$$\mathbb{E}(X) = \alpha\beta$$

Varianz:

$$\mathrm{Var}(X) = \alpha\beta^2$$

Momente:

$$\mathbb{E}(X^k) = \frac{\beta^k \Gamma(\alpha+k)}{\Gamma(\alpha)}, \quad k = 1, 2, \ldots$$

Momentenerzeugende Funktion:

$$m(t) = \left(\frac{1}{1-\beta t} \right)^\alpha, \quad t < \frac{1}{\beta}$$

Modus:

$$x_{\text{mod}} = (\alpha - 1)\beta, \quad \alpha \geq 1$$

Additionstheorem: $X_i \sim \gamma(\alpha_i, \beta)$, $i = 1, 2, \ldots, n$, ua.:

$$\sum_{i=1}^{n} X_i \sim \gamma\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$$

Spezialfälle:

$$\begin{aligned} \alpha = 1 :\ & \gamma(1, \beta) \equiv Ex_\beta \text{ (Exponential)} \\ \alpha = k \in \mathbb{N} :\ & \gamma(k, \beta) \equiv Er_{k,\beta} \text{ (Erlang)} \\ \alpha = n/2\ (n \in \mathbb{N}),\ \beta = 2 :\ & \gamma(n/2, 2) \equiv \chi^2_n \text{ (Chiquadrat)} \end{aligned}$$

Beziehung zur Betaverteilung:

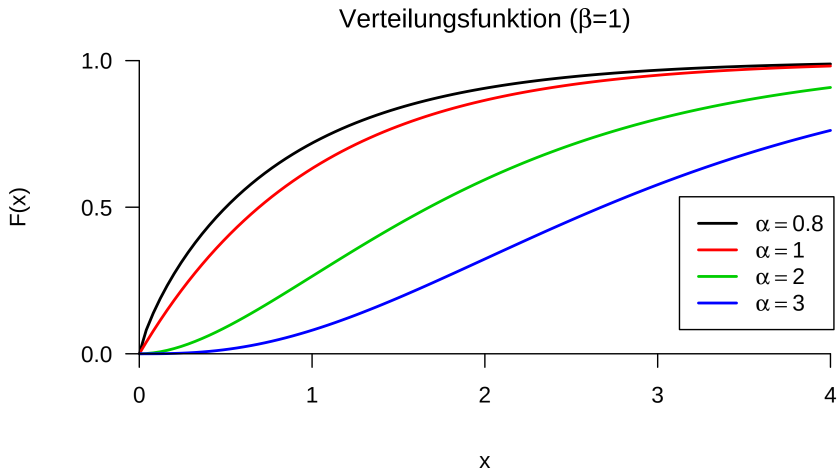
$$X \sim \gamma(\alpha_1, \beta),\ Y \sim \gamma(\alpha_2, \beta),\ \text{ua.} \quad \Longrightarrow \frac{X}{X + Y} \sim Be(\alpha_1, \alpha_2)$$

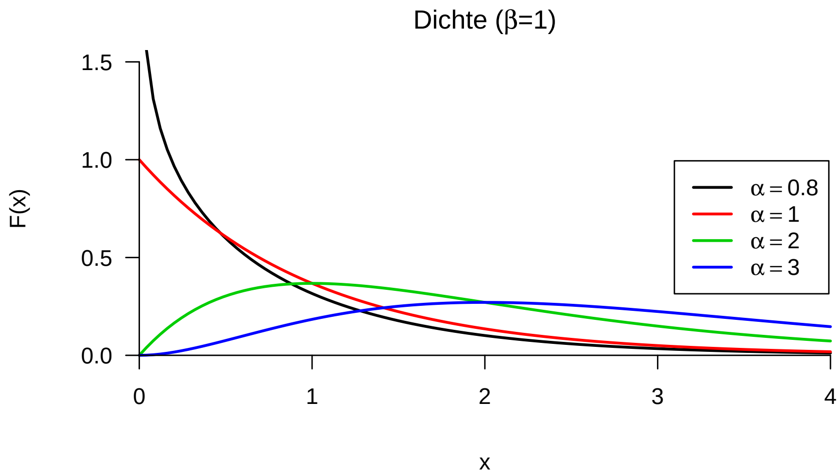
Beziehung zur Poissonverteilung: Für $X \sim \gamma(k, \beta)$, $k \in \mathbb{N}$, und $Y \sim P_\mu$ gilt:

$$P\{X > \mu\beta\} = P\{Y \leq k - 1\}$$

R–Funktionen: `shape` $\widehat{=}$ α, `rate` $\widehat{=}$ $1/\beta$, `scale` $\widehat{=}$ β

```
gamma(x)  # Gammafunktion
dgamma(x, shape, rate = 1, scale = 1/rate, log = FALSE)
pgamma(q, shape, rate = 1, scale = 1/rate, lower.tail = TRUE,log.p = FALSE)
qgamma(p, shape, rate = 1, scale = 1/rate, lower.tail = TRUE,log.p = FALSE)
rgamma(n, shape, rate = 1, scale = 1/rate)
```


Verteilungsfunktion (β=1)
F(x)
1.0
0.5
0.0
0
1
2
3
4
x
α = 0.8
α = 1
α = 2
α = 3


Dichte (β=1)
F(x)
1.5
1.0
0.5
0.0
0
1
2
3
4
x
α = 0.8
α = 1
α = 2
α = 3

A.2.6 Chiquadratverteilung (χ^2–Verteilung)

Bezeichnung: $X \sim \chi^2_n$, $\chi^2(n)$

Parameter: $n \in \mathbb{N}$ (Freiheitsgrade, Formparameter)

Merkmalraum: $M_X = \mathbb{R}^+$

Dichte:

$$f(x) = \frac{x^{n/2-1} e^{-x/2}}{\Gamma(n/2)\, 2^{n/2}}, \quad x > 0$$

Mittelwert:

$$\mathbb{E}(X) = n$$

Varianz:

$$\mathrm{Var}(X) = 2n$$

Momente:

$$\mathbb{E}(X^k) = \frac{2^k \Gamma(n/2 + k)}{\Gamma(n/2)}, \quad k = 1, 2, \ldots$$

Momentenerzeugende Funktion:

$$m(t) = \left(\frac{1}{1 - 2t} \right)^{n/2}, \quad t < \frac{1}{2}$$

Modus:

$$x_{\text{mod}} = n - 2, \quad n \geq 2$$

Spezialfall: $\chi^2_2 \equiv Ex_2$

Additionstheorem: $X_i \sim \chi^2_{n_i}$, $i = 1, 2, \ldots, K$, ua.:

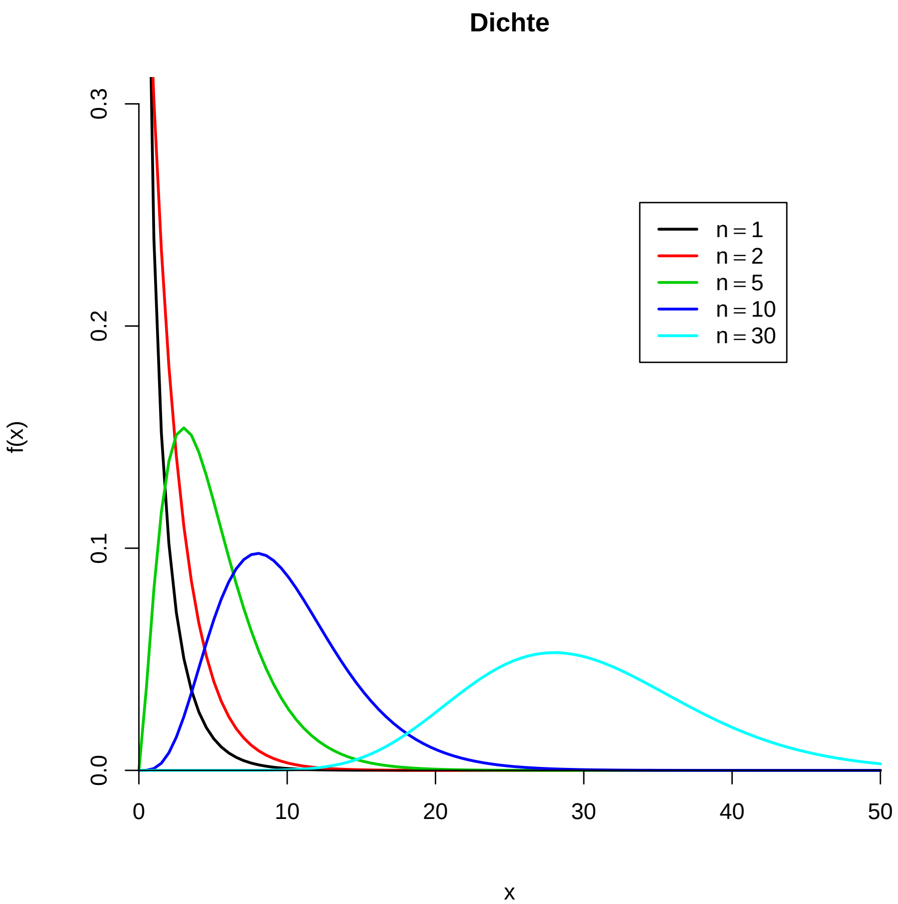
$$\sum_{i=1}^{K} X_i \sim \chi^2_n, \; n = \sum_{i=1}^{K} n_i$$

Operative Definition:

$$X \sim N(0,1) \quad \Longrightarrow \quad X^2 \sim \chi^2_1$$

R–Funktionen: `df` $\widehat{=}\ n$

```
dchisq(x, df, ncp=0, log = FALSE)
pchisq(q, df, ncp=0, lower.tail = TRUE, log.p = FALSE)
qchisq(p, df, ncp=0, lower.tail = TRUE, log.p = FALSE)
rchisq(n, df, ncp=0)
```



A.2.7 t–Verteilung (Studentverteilung)

Bezeichnung: $X \sim t_n$, $t(n)$

Parameter: $n \in \mathbb{N}$ (Freiheitsgrade, Formparameter)

Merkmalraum: $M_X = \mathbb{R}$

Dichte:

$$f(x) = \frac{\Gamma\big((n+1)/2\big)}{\sqrt{n\pi}\,\Gamma(n/2)(1+x^2/n)^{(n+1)/2}}, \quad x \in \mathbb{R}$$

Mittelwert:

$$\mathbb{E}(X) = 0, \quad n > 1$$

Varianz:

$$\mathrm{Var}(X) = \frac{n}{n-2}, \quad n > 2$$

(Zentrale) Momente:

$$\mathbb{E}(X^k) = \begin{cases} 0 & n > k,\ k \text{ ungerade} \\ \dfrac{n^{k/2} B\big((k+1)/2, (n-k)/2\big)}{B(1/2, n/2)} & n > k,\ k \text{ gerade} \end{cases}$$

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \quad \text{(Betafunktion)}$$

Quantile:

$$t_{n;\,p} = -t_{n;\,1-p}, \quad 0 < p < 1$$

Spezialfall: $t_1 \equiv C(0,1)$ (Cauchy–Verteilung)

Beziehung zur Normalverteilung: $f(x|n)$ = Dichte der t_n–Verteilung

$$\lim_{n\to\infty} f(x|n) = \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

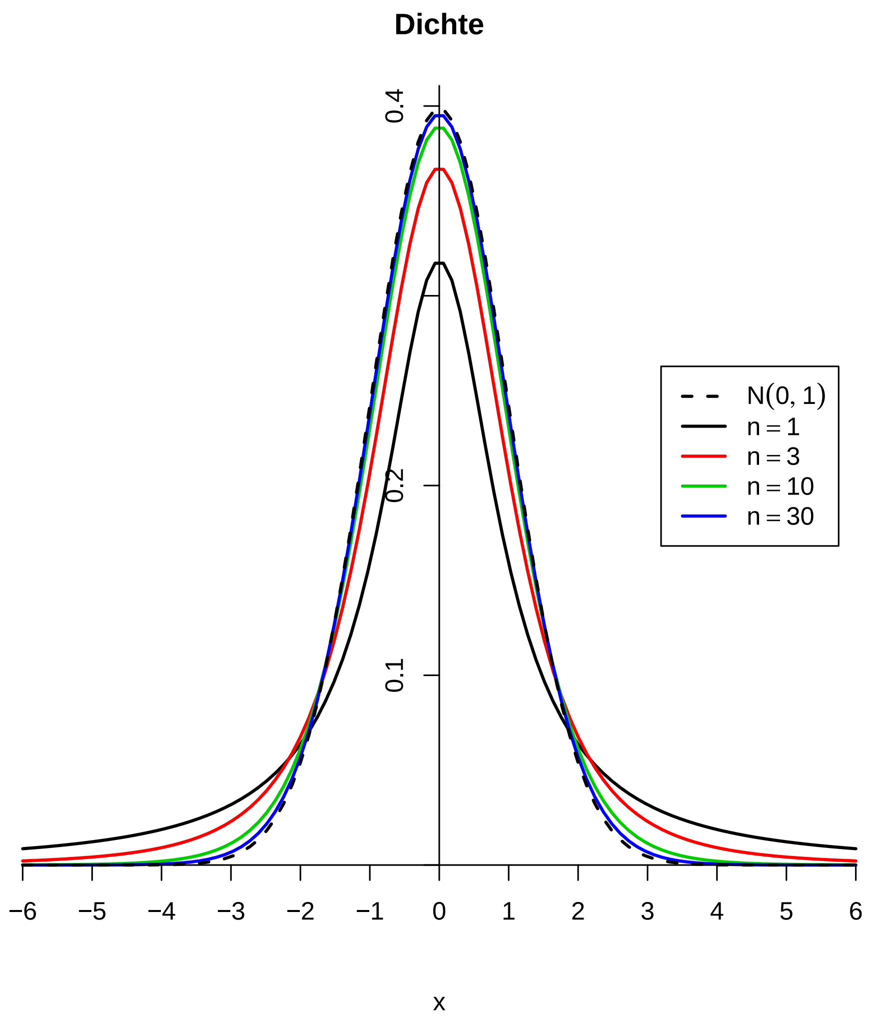
$$\lim_{n\to\infty} t_{n;\,p} = u_p \quad (0 < p < 1)$$

Operative Definition:

$$Z \sim N(0,1),\ V \sim \chi^2_\nu,\ \text{ua.} \quad \Longrightarrow \quad \frac{Z}{\sqrt{V/\nu}} \sim t_\nu$$

R–Funktionen: `df` $\widehat{=}\ n$

```
beta(a, b)  # Betafunktion
dt(x, df, ncp = 0, log = FALSE)
pt(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
qt(p, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
rt(n, df, ncp = 0)
```



A.2.8 F–Verteilung (Fisher–Verteilung)

Bezeichnung: $X \sim F_{m,n}$, $F(m,n)$

Parameter: $m, n \in \mathbb{N}$ (Freiheitsgrade, Formparameter)

Merkmalraum: $M_X = \mathbb{R}^+$

Dichte:

$$f(x) = \frac{\Gamma\big((m+n)/2\big)\, m^{m/2} n^{n/2} x^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)(mx+n)^{(m+n)/2}}, \quad x > 0$$

Mittelwert:

$$\mathbb{E}(X) = \frac{n}{n-2}, \quad n > 2$$

Varianz:

$$\mathrm{Var}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad n > 4$$

Momente:

$$\mathbb{E}(X^k) = \left(\frac{n}{m}\right)^k \frac{\Gamma(m/2+k)\Gamma(n/2-k)}{\Gamma(m/2)\Gamma(n/2)}, \quad k < \frac{n}{2}$$

Quantile:

$$F_{m,n;\,p} = \frac{1}{F_{n,m;\,1-p}}, \quad 0 < p < 1$$

Modus:

$$x_{\text{mod}} = \frac{n(m-2)}{m(n+2)}, \quad m \geq 2$$

Symmetrie:

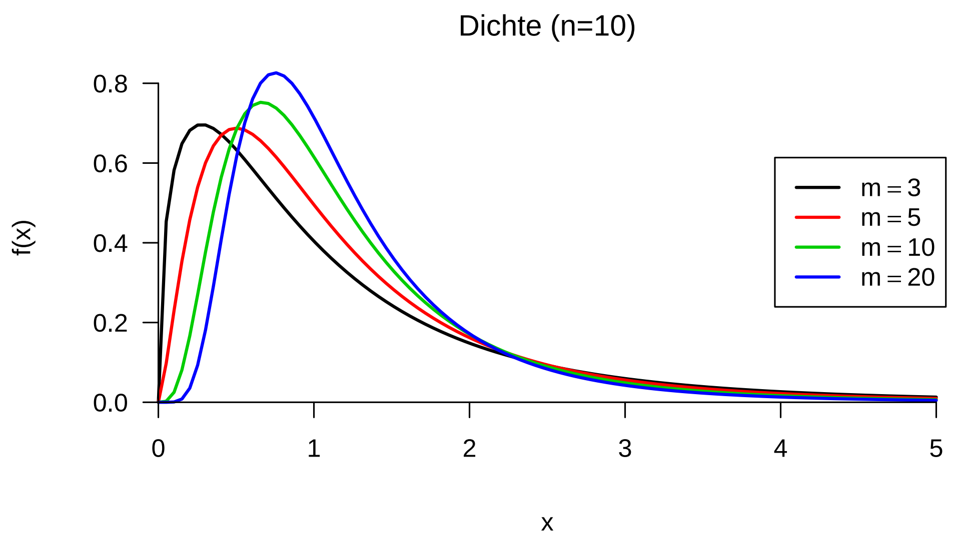
$$X \sim F_{m,n} \quad \Longleftrightarrow \quad \frac{1}{X} \sim F_{n,m}$$

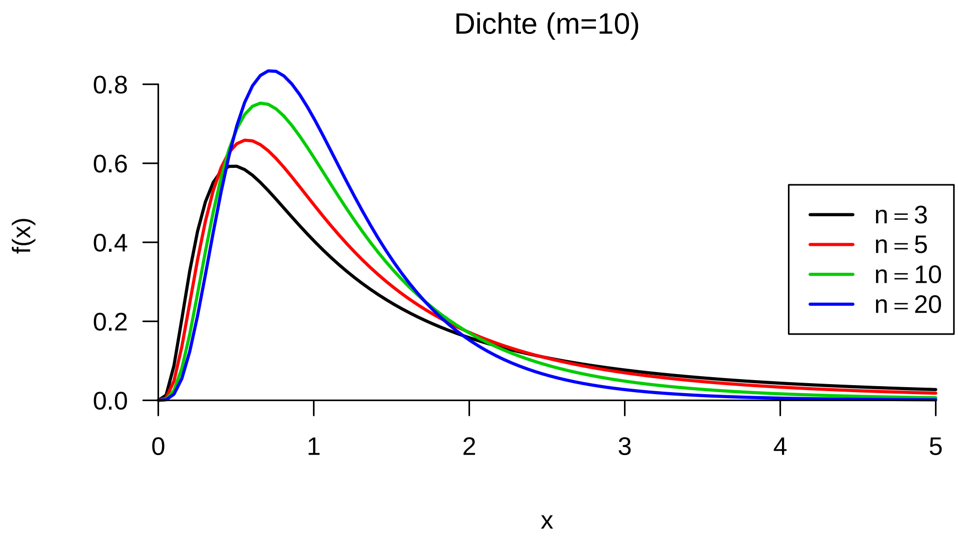
Operative Definition:

$$V_1 \sim \chi^2_{\nu_1},\ V_2 \sim \chi^2_{\nu_2},\ \text{ua.} \quad \Longrightarrow \quad \frac{V_1/\nu_1}{V_2/\nu_2} \sim F_{\nu_1,\nu_2}$$

R–Funktionen: `df1` $\widehat{=}\ m$, `df2` $\widehat{=}\ n$

```
df(x, df1, df2, log = FALSE)
pf(q, df1, df2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
qf(p, df1, df2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
rf(n, df1, df2, ncp = 0)
```





A.2.9 Betaverteilung

Bezeichnung: $X \sim Be(a,b),\ \beta(a,b)$

Parameter: $a, b > 0$ (Formparameter)

Merkmalraum: $M_X = (0,1)$

Betafunktion:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt, \quad x, y > 0$$

$$B(x,y) = B(y,x), \quad B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

Dichte:

$$f(x) = \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1} I_{(0,1)}(x)$$

$$f(x|a,b) = f(1-x|b,a)$$

Verteilungsfunktion:

$$F(x) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1}\,dt, \quad x \in (0,1) \qquad \text{[unvollständige Betafunktion]}$$

$$F(x|a,b) = 1 - F(1-x|b,a)$$

Mittelwert:

$$\mathbb{E}(X) = \frac{a}{a+b}$$

Varianz:

$$\mathrm{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}$$

Momente:

$$\mathbb{E}(X^k) = \frac{B(a+k,b)}{B(a,b)}, \quad k \in \mathbb{N}$$

Quantile:

$$Be(a,b;\, p) = 1 - Be(b,a;\, 1-p), \quad 0 < p < 1$$

Modus:

$$x_{\text{mod}} = \frac{a-1}{a+b-2}, \quad a,b > 1$$

Symmetrie:

$$X \sim Be(a,b) \quad \Longleftrightarrow \quad 1 - X \sim Be(b.a)$$

Spezialfall: $Be(1,1) \equiv U_{0,1}$

Beziehung zur F–Verteilung: Für $m,n \in \mathbb{N}$ gilt:

$$X \sim Be(m,n) \quad \Longleftrightarrow \quad \frac{X/m}{(1-X)/n} \sim F_{2m,2n}$$

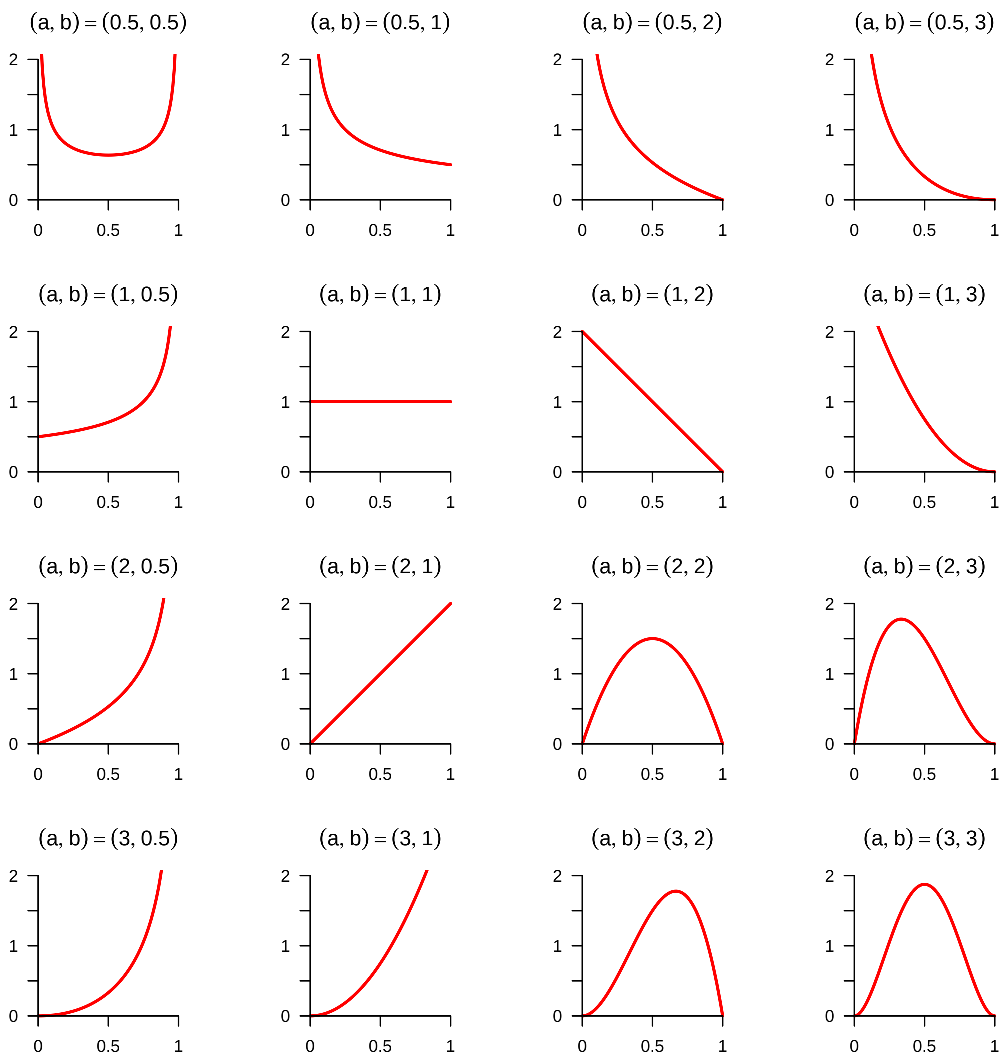
Beziehung zur Binomialverteilung: Für $X \sim Be(k, n-k+1)$, $k = 1,2,\ldots,n$, mit $n \in \mathbb{N}$, und $Y \sim B_{n,p}$ gilt:

$$P\{X > p\} = P\{Y \leq k-1\}$$

R–Funktionen: `shape1` $\widehat{=}\, a$, `shape2` $\widehat{=}\, b$

```
beta(a, b)  # Betafunktion
dbeta(x, shape1, shape2, ncp = 0, log = FALSE)
pbeta(q, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
qbeta(p, shape1, shape2, ncp = 0, lower.tail = TRUE, log.p = FALSE)
rbeta(n, shape1, shape2, ncp = 0)
```

Dichte


(a, b) = (0.5, 0.5)
(a, b) = (0.5, 1)
(a, b) = (0.5, 2)
(a, b) = (0.5, 3)
(a, b) = (1, 0.5)
(a, b) = (1, 1)
(a, b) = (1, 2)
(a, b) = (1, 3)
(a, b) = (2, 0.5)
(a, b) = (2, 1)
(a, b) = (2, 2)
(a, b) = (2, 3)
(a, b) = (3, 0.5)
(a, b) = (3, 1)
(a, b) = (3, 2)
(a, b) = (3, 3)
0
0.5
1
2

B Konfidenzintervalle

B.1 $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, ua.

Parameter	Voraussetzung	Typ	$100(1-\alpha)\%$ Intervall
μ	σ bekannt	zweiseitig	$\overline{X} \pm u_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
		oberes	$\left(-\infty, \overline{X} + u_{1-\alpha} \dfrac{\sigma}{\sqrt{n}}\right)$
		unteres	$\left(\overline{X} - u_{1-\alpha} \dfrac{\sigma}{\sqrt{n}}, \infty\right)$
	σ unbekannt	zweiseitig	$\overline{X} \pm t_{n-1;\,1-\alpha/2} \dfrac{S}{\sqrt{n}}$
		oberes	$\left(-\infty, \overline{X} + t_{n-1;\,1-\alpha} \dfrac{S}{\sqrt{n}}\right)$
		unteres	$\left(\overline{X} - t_{n-1;\,1-\alpha} \dfrac{S}{\sqrt{n}}, \infty\right)$

Parameter	Voraussetzung	Typ	$100(1-\alpha)\%$ Intervall
σ^2	μ unbekannt	zweiseitig	$\left(\dfrac{(n-1)S^2}{\chi^2_{n-1;\,1-\alpha/2}},\ \dfrac{(n-1)S^2}{\chi^2_{n-1;\,\alpha/2}}\right)$
		oberes	$\left(0,\ \dfrac{(n-1)S^2}{\chi^2_{n-1;\,\alpha}}\right)$
		unteres	$\left(\dfrac{(n-1)S^2}{\chi^2_{n-1;\,1-\alpha}},\ \infty\right)$
σ	μ unbekannt	zweiseitig	$\left(\dfrac{\sqrt{n-1}\ S}{\sqrt{\chi^2_{n-1;\,1-\alpha/2}}},\ \dfrac{\sqrt{n-1}\ S}{\sqrt{\chi^2_{n-1;\,\alpha/2}}}\right)$
		oberes	$\left(0,\ \dfrac{\sqrt{n-1}\ S}{\sqrt{\chi^2_{n-1;\,\alpha}}}\right)$
		unteres	$\left(\dfrac{\sqrt{n-1}\ S}{\sqrt{\chi^2_{n-1;\,1-\alpha}}},\ \infty\right)$

B.2 $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$, $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$, ua.

Parameter	Voraussetzung	Typ	$100(1-\alpha)\%$ Intervall
$\mu_1 - \mu_2$	σ_1, σ_2 bekannt	zweiseitig	$\overline{X} - \overline{Y} \pm u_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
		oberes	$\left(-\infty,\, \overline{X} - \overline{Y} + u_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
		unteres	$\left(\overline{X} - \overline{Y} - u_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\, \infty\right)$
	σ_1, σ_2 unbekannt $\sigma_1 = \sigma_2$	zweiseitig	$\overline{X} - \overline{Y} \pm t_{n_1+n_2-2;\,1-\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ mit $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
		oberes	$\left(-\infty,\, \overline{X} - \overline{Y} + t_{n_1+n_2-2;\,1-\alpha}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$
		unteres	$\left(\overline{X} - \overline{Y} - t_{n_1+n_2-2;\,1-\alpha}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\, \infty\right)$
$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1, μ_2 unbekannt	zweiseitig	$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1,n_2-1;\,1-\alpha/2}},\, \frac{S_1^2}{S_2^2} F_{n_2-1,n_1-1;\,1-\alpha/2}\right)$
		oberes	$\left(0,\, \frac{S_1^2}{S_2^2} F_{n_2-1,n_1-1;\,1-\alpha}\right)$
		unteres	$\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{n_1-1,n_2-1;\,1-\alpha}},\, \infty\right)$

B.3 $X_1, \dots, X_n \sim Ex_\tau$, ua.

Parameter	Typ	$100(1-\alpha)\%$ Intervall
τ	zweiseitig	$\left(\dfrac{2n\,\overline{X}}{\chi^2_{2n;\,1-\alpha/2}},\ \dfrac{2n\,\overline{X}}{\chi^2_{2n;\,\alpha/2}} \right)$
	oberes	$\left(0,\ \dfrac{2n\,\overline{X}}{\chi^2_{2n;\,\alpha}} \right)$
	unteres	$\left(\dfrac{2n\,\overline{X}}{\chi^2_{2n;\,1-\alpha}},\ \infty \right)$

B.4 $X_1, \dots, X_n \sim A_p$, ua.

Parameter	Typ	(approx.) $100(1-\alpha)\%$ Intervall (n groß)
p	zweiseitig	$\widehat{p} \pm u_{1-\alpha/2} \sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{n}}$ mit $\quad \widehat{p} = \overline{X}$
	oberes	$\left(0,\ \widehat{p} + u_{1-\alpha} \sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{n}} \right)$
	unteres	$\left(\widehat{p} - u_{1-\alpha} \sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{n}},\ 1 \right)$

B.5 $X_1, \ldots, X_{n_1} \sim A_{p_1}$, $Y_1, \ldots, Y_{n_2} \sim A_{p_2}$ ua.

Parameter	Typ	(approx.) $100(1-\alpha)\%$ Intervall (n_1, n_2 groß)
$p_1 - p_2$	zweiseitig	$\widehat{p}_1 - \widehat{p}_2 \pm u_{1-\alpha/2} \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$ mit $\widehat{p}_1 = \overline{X}, \quad \widehat{p}_2 = \overline{Y}$
	oberes	$\left(-1,\, \widehat{p}_1 - \widehat{p}_2 + u_{1-\alpha} \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}\right)$
	unteres	$\left(\widehat{p}_1 - \widehat{p}_2 - u_{1-\alpha} \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}},\, 1\right)$

B.6 $X_1, \ldots, X_n \sim P_\mu$, ua.

Parameter	Typ	(approx.) $100(1-\alpha)\%$ Intervall (n groß)
μ	zweiseitig	$\widehat{\mu} \pm u_{1-\alpha/2} \sqrt{\frac{\widehat{\mu}}{n}}$ mit $\widehat{\mu} = \overline{X}$
	oberes	$\left(0,\, \widehat{\mu} + u_{1-\alpha} \sqrt{\frac{\widehat{\mu}}{n}}\right)$
	unteres	$\left(\widehat{\mu} - u_{1-\alpha} \sqrt{\frac{\widehat{\mu}}{n}},\, \infty\right)$

C Parametertests

C.1 $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, ua.

$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_0$ verwerfen, falls:
	σ bekannt	
$\mu = \mu_0$	$\mu \neq \mu_0$	$\dfrac{\lvert\overline{X} - \mu_0\rvert}{\sigma/\sqrt{n}} > z_{1-\alpha/2}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha}$
	σ unbekannt	
$\mu = \mu_0$	$\mu \neq \mu_0$	$\dfrac{\lvert\overline{X} - \mu_0\rvert}{S/\sqrt{n}} > t_{n-1;\,1-\alpha/2}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} > t_{n-1;\,1-\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} < -t_{n-1;\,1-\alpha}$
	μ unbekannt	
$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$\dfrac{(n-1)S^2}{\sigma_0^2} \notin \left[\chi^2_{n-1;\,\alpha/2}, \chi^2_{n-1;\,1-\alpha/2}\right]$
$\sigma \leq \sigma_0$	$\sigma > \sigma_0$	$\dfrac{(n-1)S^2}{\sigma_0^2} > \chi^2_{n-1;\,1-\alpha}$
$\sigma \geq \sigma_0$	$\sigma < \sigma_0$	$\dfrac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1;\,\alpha}$

C.2 $X_1, \ldots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$, $Y_1, \ldots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$, ua.

$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_0$ verwerfen, falls:
σ_1, σ_2 bekannt		
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\dfrac{\lvert\overline{X} - \overline{Y}\rvert}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} > z_{1-\alpha/2}$
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} > z_{1-\alpha}$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < -z_{1-\alpha}$
σ_1, σ_2 unbekannt: $\sigma_1 = \sigma_2$		
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\dfrac{\lvert\overline{X} - \overline{Y}\rvert}{S_p\sqrt{1/n_1 + 1/n_2}} > t_{n_1+n_2-2;\,1-\alpha/2}$
		mit $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{1/n_1 + 1/n_2}} > t_{n_1+n_2-2;\,1-\alpha}$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{1/n_1 + 1/n_2}} < -t_{n_1+n_2-2;\,1-\alpha}$

$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_0$ verwerfen, falls:
σ_1, σ_2 unbekannt: $\sigma_1 \neq \sigma_2$ (approx. Tests)		
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$\dfrac{\lvert\overline{X} - \overline{Y}\rvert}{\sqrt{S_1^2/n_1 + S_2/n_2}} > t_{\nu;\,1-\alpha/2}$
		mit $\nu = \left\lfloor \dfrac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\left(S_1^2/n_1\right)^2/(n_1-1) + \left(S_2^2/n_2\right)^2/(n_2-1)} \right\rfloor$
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/n_1 + S_2/n_2}} > t_{\nu;\,1-\alpha}$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$\dfrac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/n_1 + S_2/n_2}} < -t_{\nu;\,1-\alpha}$
μ_1, μ_2 unbekannt		
$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$	$\dfrac{S_1^2}{S_2^2} \notin \left[\dfrac{1}{F_{n_2-1,n_1-1;\,1-\alpha/2}},\, F_{n_1-1,n_2-1;\,1-\alpha/2}\right]$
$\sigma_1 \leq \sigma_2$	$\sigma_1 > \sigma_2$	$\dfrac{S_1^2}{S_2^2} > F_{n_1-1,n_2-1;\,1-\alpha}$
$\sigma_1 \geq \sigma_2$	$\sigma_1 < \sigma_2$	$\dfrac{S_1^2}{S_2^2} < \dfrac{1}{F_{n_2-1,n_1-1;\,1-\alpha}}$

C.3 $X_1, \ldots, X_n \sim A_p$, ua.

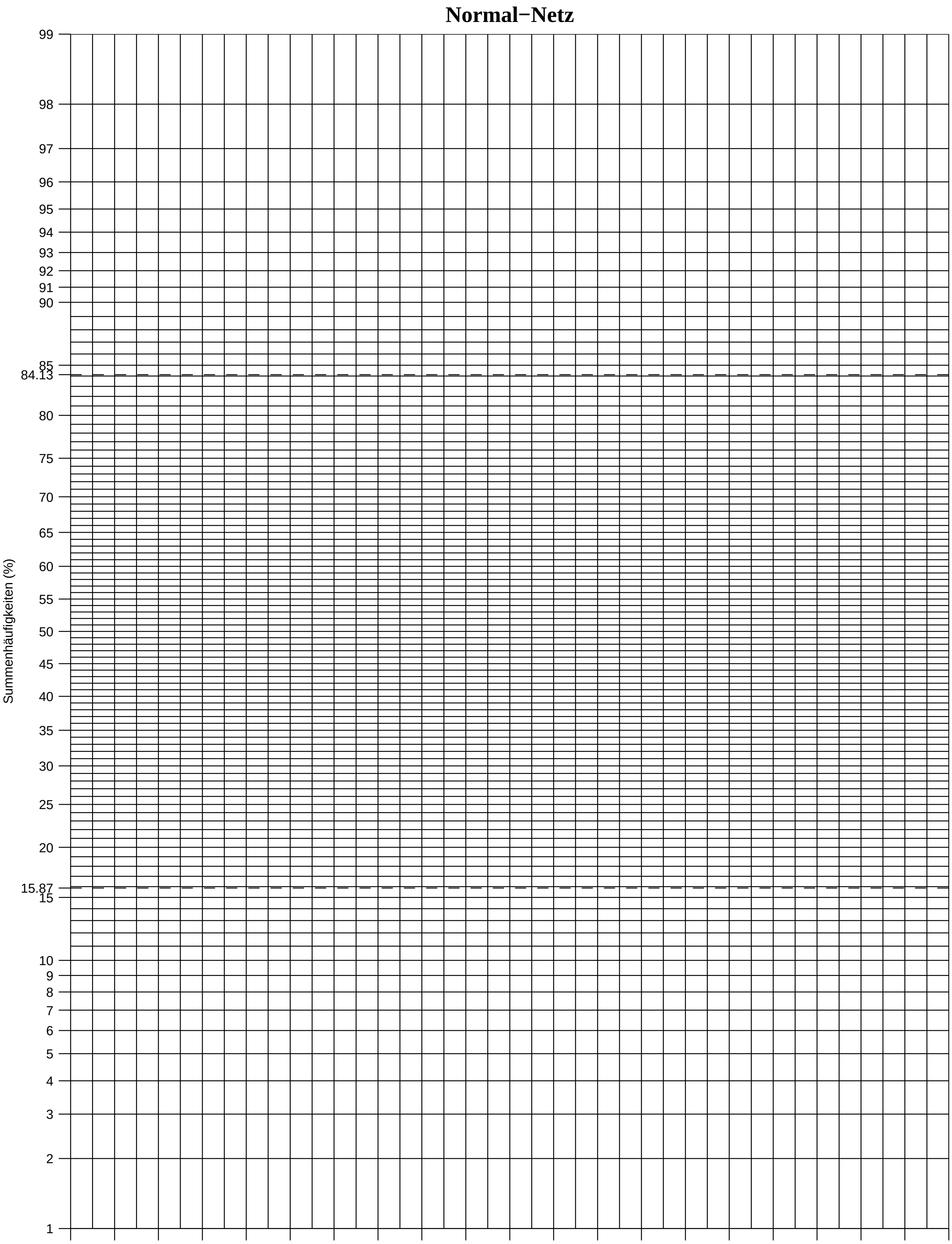
$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_0$ verwerfen, falls:
	n groß	
$p = p_0$	$p \neq p_0$	$\dfrac{\lvert\widehat{p} - p_0\rvert}{\sqrt{p_0(1-p_0)/n}} > z_{1-\alpha/2}$
		mit $\widehat{p} = \overline{X}$
$p \leq p_0$	$p > p_0$	$\dfrac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} > z_{1-\alpha}$
$p \geq p_0$	$p < p_0$	$\dfrac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} < -z_{1-\alpha}$

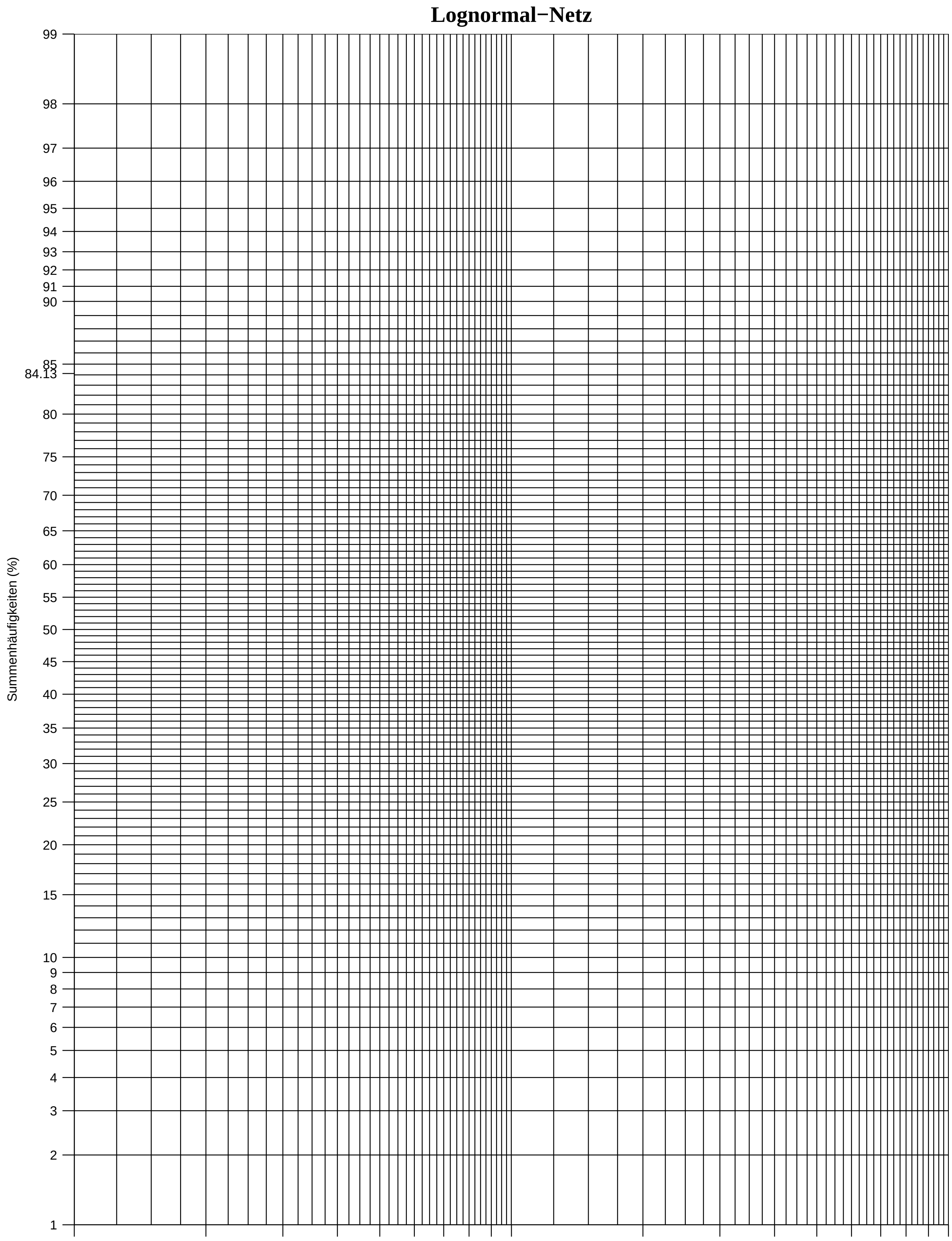
C.4 $X_1, \ldots, X_{n_1} \sim A_{p_1}$, $Y_1, \ldots, Y_{n_2} \sim A_{p_2}$ ua.

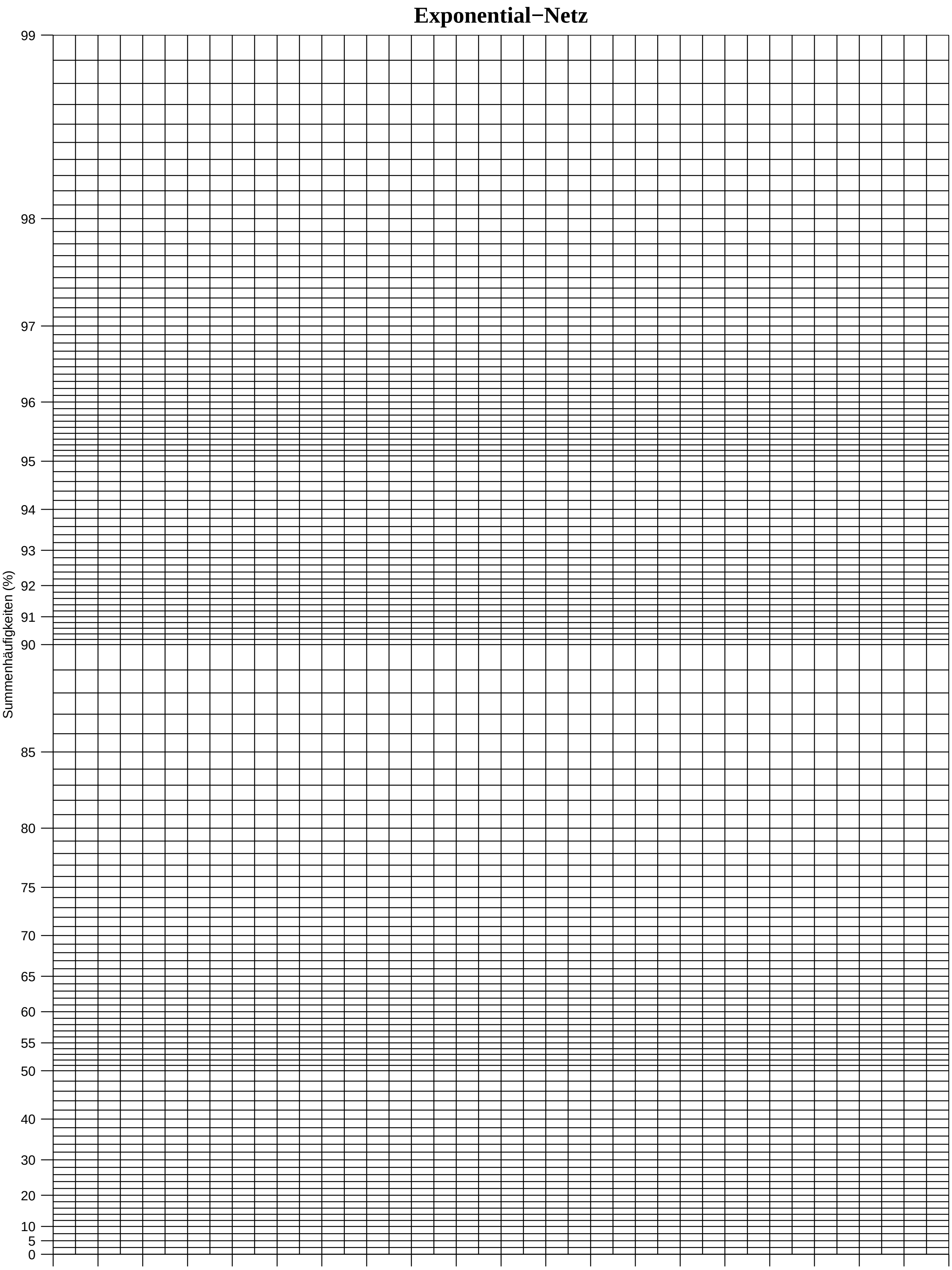
$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_0$ verwerfen, falls:
	n, m groß	
$p_1 = p_2$	$p_1 \neq p_2$	$\dfrac{\lvert\widehat{p}_1 - \widehat{p}_2\rvert}{\sqrt{\widehat{p}(1-\widehat{p})\big(1/n_1 + 1/n_2\big)}} > z_{1-\alpha/2}$
		mit $\widehat{p}_1 = \overline{X}$, $\widehat{p}_2 = \overline{Y}$, $\widehat{p} = \dfrac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2}$
$p_1 \leq p_2$	$p_1 > p_2$	$\dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\big(1/n_1 + 1/n_2\big)}} > z_{1-\alpha}$
$p_1 \geq p_2$	$p_1 < p_2$	$\dfrac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\big(1/n_1 + 1/n_2\big)}} < -z_{1-\alpha}$

D Wahrscheinlichkeitsnetze

- Normalnetz
- Lognormalnetz
- Exponentialnetz


Normal−Netz
Summenhäufigkeiten (%)
99
98
97
96
95
94
93
92
91
90
85
84.13
80
75
70
65
60
55
50
45
40
35
30
25
20
15.87
15
10
9
8
7
6
5
4
3
2
1


Lognormal−Netz
Summenhäufigkeiten (%)
99
98
97
96
95
94
93
92
91
90
85
84.13
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
9
8
7
6
5
4
3
2
1


Exponential−Netz
Summenhäufigkeiten (%)
99
98
97
96
95
94
93
92
91
90
85
80
75
70
65
60
55
50
40
30
20
10
5
0

E Tabellen

Tabelle 1: Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung $N(0,1)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Tabelle 2: Quantile z_p der $N(0,1)$

p	0.60	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995	0.999
z_p	0.2533	0.6745	0.8416	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

Tabelle 3: Quantile $t_{\nu;\,p}$ der t–Verteilung

	p								
ν	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995	0.999
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	318.309
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385
31	0.682	0.853	1.054	1.309	1.696	2.040	2.453	2.744	3.375
32	0.682	0.853	1.054	1.309	1.694	2.037	2.449	2.738	3.365
33	0.682	0.853	1.053	1.308	1.692	2.035	2.445	2.733	3.356
34	0.682	0.852	1.052	1.307	1.691	2.032	2.441	2.728	3.348
35	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.340
36	0.681	0.852	1.052	1.306	1.688	2.028	2.434	2.719	3.333
37	0.681	0.851	1.051	1.305	1.687	2.026	2.431	2.715	3.326
38	0.681	0.851	1.051	1.304	1.686	2.024	2.429	2.712	3.319
39	0.681	0.851	1.050	1.304	1.685	2.023	2.426	2.708	3.313
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232
70	0.678	0.847	1.044	1.294	1.667	1.994	2.381	2.648	3.211
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195
90	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090

Bem: In der letzten Zeile stehen die entsprechenden Quantile der $N(0,1)$ (vgl. Tabelle 2).

Tabelle 4: Quantile $\chi^2_{\nu;\,p}$ der Chiquadratverteilung

	p									
ν	0.005	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.892	62.883
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.476
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
41	21.421	22.906	25.215	27.326	29.907	52.949	56.942	60.561	64.950	68.053
42	22.138	23.650	25.999	28.144	30.765	54.090	58.124	61.777	66.206	69.336
43	22.859	24.398	26.785	28.965	31.625	55.230	59.304	62.990	67.459	70.616
44	23.584	25.148	27.575	29.787	32.487	56.369	60.481	64.201	68.710	71.893
45	24.311	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957	73.166
46	25.041	26.657	29.160	31.439	34.215	58.641	62.830	66.617	71.201	74.437
47	25.775	27.416	29.956	32.268	35.081	59.774	64.001	67.821	72.443	75.704
48	26.511	28.177	30.755	33.098	35.949	60.907	65.171	69.023	73.683	76.969
49	27.249	28.941	31.555	33.930	36.818	62.038	66.339	70.222	74.919	78.231
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490

Tabelle 5: Quantile $F_{\nu_1,\nu_2;p}$ der F–Verteilung

		ν_1											
p	ν_2	1	2	3	4	5	6	7	8	9	10	12	15
0.95	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9
0.975		647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9
0.99		4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157
0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43
0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.42	39.43
0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43
0.95	3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703
0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25
0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87
0.95	4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858
0.975		12.22	10.65	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844	8.751	8.657
0.99		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20
0.95	5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619
0.975		10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.428
0.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.722
0.95	6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938
0.975		8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366	5.269
0.99		13.75	10.92	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559
0.95	7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511
0.975		8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.568
0.99		12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314
0.95	8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218
0.975		7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.200	4.101
0.99		11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515
0.95	9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006
0.975		7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868	3.769
0.99		10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962
0.95	10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845
0.975		6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621	3.522
0.99		10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558
0.95	12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617
0.975		6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277	3.177
0.99		9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010
0.95	15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403
0.975		6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963	2.862
0.99		8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522