# Anhang A

# Appendix: Stetige Verteilungen

In diesem Appendix werden – der üblichen statistischen Praxis folgend – Zufallsvariablen mit Großbuchstaben und Realisationen mit Kleinbuchstaben bezeichnet.

## A.1 Gleichverteilung

Die Gleichverteilung spielt für das Folgende zwar keine große Rolle, aber aufgrund ihrer Einfachheit ist sie aus didaktischen Gründen gut geeignet um die Grundkonzepte zu wiederholen.

Eine *diskrete Zufallsvariable* ist gleichverteilt (*uniformly distributed)*, wenn jede ihrer möglichen $m$ Ausprägungen mit gleicher Wahrscheinlichkeit eintritt:

$$f(x) = \Pr[X = x_i] = \frac{1}{m} \qquad (i = 1, \ldots, m)$$

Die Wahrscheinlichkeitsfunktion der Augenzahl eines Würfels ist z.B. $\Pr[X = x_i] = 1/6$ für $(i = 1, \ldots, 6)$.

Für eine *stetige Zufallsvariable* liegen die möglichen Realisationen zwischen den beiden endlichen Grenzen $a$ und $b$. Da für jede Dichtefunktion gilt

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

folgt für die **Dichtefunktion einer gleichverteilten stetigen Zufallsvariable** $X$

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

Die Wahrscheinlichkeit, dass eine Realisation von $X$ einen Wert zwischen $x_1$ und $x_2$ annimmt, ist

$$\int_{x_1}^{x_2} f(x)\, dx = \Pr[x_1 \leq x \leq x_2] = \frac{(x_2 - x_1)}{b - a}$$

Durch Integration der Dichtefunktion erhält man die **Verteilungsfunktion** $F(x)$

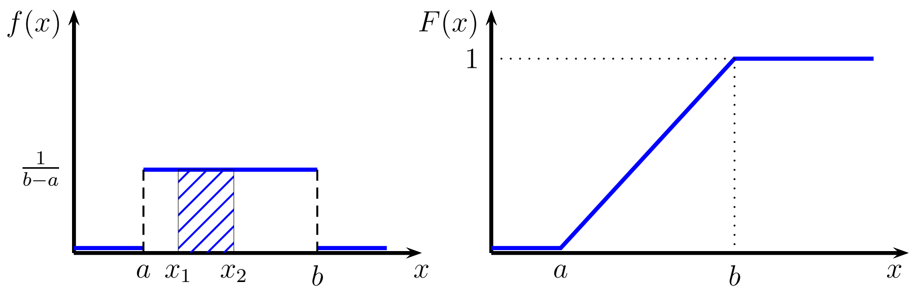



**Abbildung A.1:** Dichte- und Verteilungsfunktion einer stetigen gleichverteilten Zufallsvariable.

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

Abbildung A.1 zeigt die Dichte- und Verteilungsfunktion einer stetigen gleichverteilten Zufallsvariable.

Der Erwartungswert ist

$$\begin{aligned} \mathrm{E}(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{a} x \cdot 0 dx + \int_{a}^{b} x\, f(x) dx + \int_{b}^{+\infty} x \cdot 0 dx \\ &= \int_a^b x \frac{1}{b-a} dx \\ &= \left. \frac{x^2}{2(b-a)} \right|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

weil $(b^2 - a^2) = (a+b)(b-a)$.

Die Varianz $\mathrm{var}(X) = \mathrm{E}[X - \mathrm{E}(X)]^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ kann ebenfalls berechnet werden:

$$\mathrm{E}(X^2) = \int_a^b x^2 \left( \frac{1}{b-a} \right) dx = \left. \frac{x^3}{3(b-a)} \right]_a^b = \frac{b^3 - a^3}{3(b-a)}$$

Unter Verwendung von $b^3 - a^3 = (b-a)(a^2 + ab + b^2)$ erhalten wir schließlich

$$\mathrm{var}(X) = \frac{(a^2 + ab + b^2)}{3} - \left( \frac{a+b}{2} \right)^2 = \frac{(a^2 + b^2 - 2ab)}{12} = \frac{(a-b)^2}{12}$$

Die Gleichverteilung ist also durch die zwei Parameter $a$ und $b$ vollständig bestimmt.

## A.2 Normalverteilung

Die Normalverteilung spielt in der Ökonometrie eine besondere Rolle, und dies aus mehreren Gründen. Erstens sind viele Variablen in ihren Grundgesamtheiten

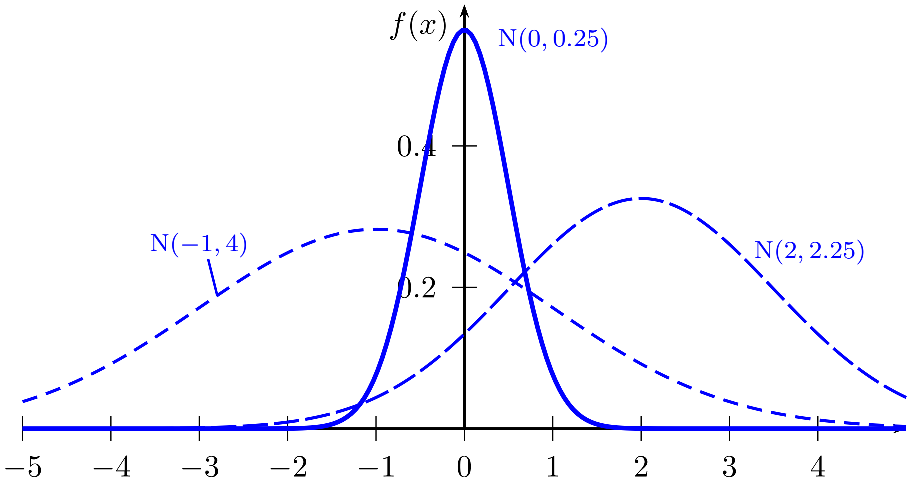



**Abbildung A.2:** Dichtefunktionen von normalverteilten Zufallsvariablen mit unterschiedlichem $\mu$ und $\sigma^2$.

'annähernd' normalverteilt (z.B. Körpergröße, Intelligenz, Meßfehler, etc.), zweitens können viele andere Verteilungen aus ihr hergeleitet werden (z.B. die $\chi^2$-, $t$- oder $F$-Verteilung), und drittens nähern sich bei sehr großen Beobachtungszahlen andere Verteilungen manchmal der Normalverteilung an (z.B. die Binomialverteilung). Außerdem werden wir später sehen, sind für große Beobachtungszahlen die *Stichprobenkennwertverteilungen* unabhängig von der Verteilung der Grundgesamtheit annähernd normalverteilt (*zentraler Grenzwertsatz*).

Die Normalverteilung ist eine stetige Verteilung mit der Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(1/2)[(x_i-\mu)/\sigma]^2}$$

mit $\mu$ = Erwartungswert, $\sigma$ = Standardabweichung von $X$, $\pi = 3.14159\ldots, e = 2.71828\ldots$.

Wenn eine Zufallsvariable $X$ normalverteilt ist mit Mittelwert $\mu$ und Varianz $\sigma^2$, d.h.

$$\begin{aligned} \mathrm{E}(X) &= \mu \\ \mathrm{var}(X) &= \sigma^2 \end{aligned}$$

wird dies üblicherweise geschrieben als

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

Abbildung A.2 zeigt drei Dichtefunktionen von normalverteilten Zufallsvariablen mit unterschiedlichem $\mu$ und $\sigma^2$.

**Eigenschaften der Normalverteilung:**

1. Die Verteilung ist stetig und symmetrisch um den Erwartungswert $\mu$.
2. Die Verteilung ist unbeschränkt und erstreckt sich von $-\infty$ bis $+\infty$.

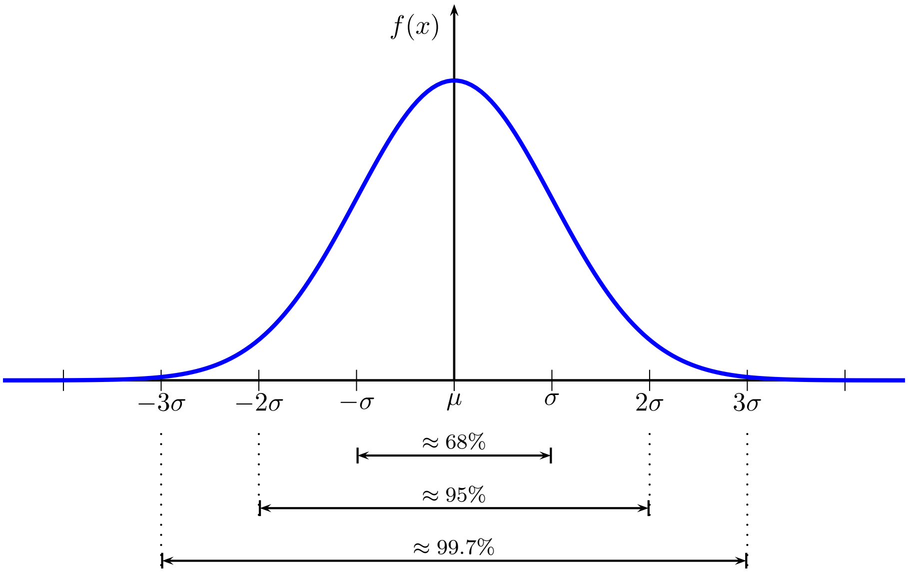



**Abbildung A.3:** Dichtefunktion einer normalverteilten Zufallsvariable und % der Fläche darunter.

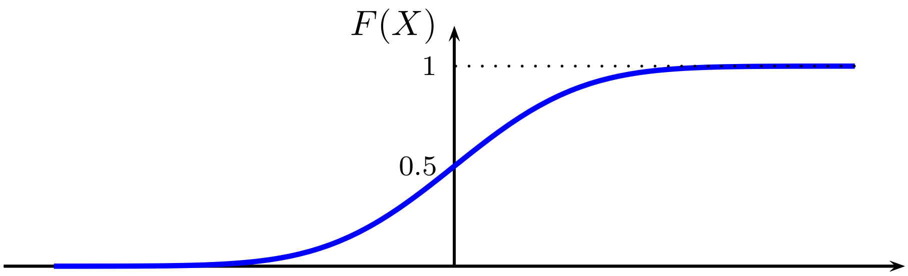



**Abbildung A.4:** Verteilungsfunktion einer normalverteilten Zufallsvariable.

3. Das Maximum der Verteilung liegt beim Mittelwert $\mu$ und der Wendepunkt bei $x = \mu \pm \sigma$. Außerdem gilt:

$$
\begin{aligned}
\Pr(\mu - 1.96\sigma < x_i < \mu + 1.96\sigma) &\approx 0.95 \\
\Pr(\mu - 2.57\sigma < x_i < \mu + 2.57\sigma) &\approx 0.99
\end{aligned}
$$

4. Die Normalverteilung ist durch die zwei Parameter $\mu$ und $\sigma^2$ vollständig spezifiziert, wir schreiben $X \sim \mathrm{N}(\mu, \sigma^2)$.

Abbildung A.3 zeigt die Dichtefunktion einer Normalverteilung.

Die Verteilungsfunktion einer normalverteilten Zufallsvariable ist in Abbildung A.4 dargestellt.

**Theorem 1** *Wenn $X, Y, \ldots, Z$ normal und unabhängig verteilte Zufallsvariablen und $a, b, \ldots, c$ beliebige Konstanten sind, dann ist die Linearkombination $aX + bY + \cdots + cZ$ auch normalverteilt.*

Beispiel: Wenn $X_i \sim \mathrm{N}(\mu, \sigma^2)$ mit $i = 1,\ 2,\ 3$ drei *unabhängig* verteilte Zufallsvariablen sind, und dann sind Erwartungswert und Varianz der Zufallsvariable

$W = 2X_1 + X_2 - 4X_3$

$$\begin{aligned}
\mathrm{E}(W) &= 2\,\mathrm{E}(X_1) + \mathrm{E}(X_2) - 4\,\mathrm{E}(X_3) = 2\mu + \mu - 4\mu = -1\mu \\
\mathrm{var}(W) &= 2^2\,\mathrm{var}(X_1) + 1^2\,\mathrm{var}(X_2) + 4^2\,\mathrm{var}(X_3) = 21\sigma^2
\end{aligned}$$

d.h. $W \sim \mathrm{N}(-\mu, 21\sigma^2)$.

Oder, wenn $X \sim \mathrm{N}(\mu, \sigma^2)$ dann ist $Y = a + bX \sim \mathrm{N}(a + b\mu, b^2\sigma^2)$, z.B. $X \sim \mathrm{N}(3, 4)$ und $Y = 2X - 5$ dann ist $Y \sim \mathrm{N}(1, 16)$.

Dieses Theorem gestattet eine Standardisierung der Normalverteilungen auf einen Mittelwert von Null und eine Standardabweichung von Eins (z–Transformation).

$$z_i = \frac{(x_i - \mu)}{\sigma} \quad \text{bzw.} \quad Z = \frac{(X - \mu)}{\sigma}$$

mit $Z \sim \mathrm{N}(0, 1)$, man sagt $Z$ ist *standardnormalverteilt*. Die Tabelle der Standardnormalverteilung findet sich in jedem Statistiklehrbuch.

Wenn $X$ und $Y$ gemeinsam normalverteilt sind, dann sind sie unabhängig wenn und nur wenn $\mathrm{cov}(X, Y) = 0$. Achtung, dies gilt nur für die Normalverteilung! Sind zwei Zufallsvariablen nicht gemeinsam normalverteilt ist $\mathrm{cov}(X, Y) = 0$ in der Regel nicht hinreichend um Unabhängigkeit zu garantieren.

**Bivariate Normalverteilung:** $(X, Y) \sim \mathrm{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, $|\rho| \leq 1$ oder in Matrixschreibweise

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}\right]$$

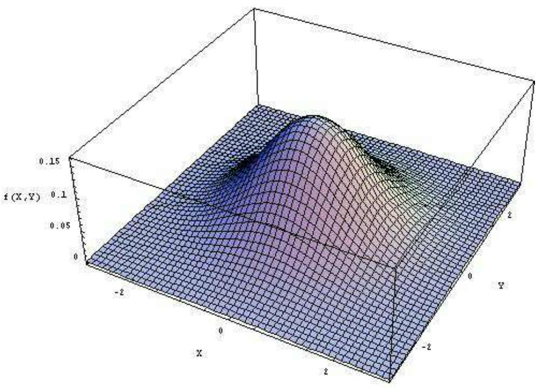
mit der Dichtefunktion

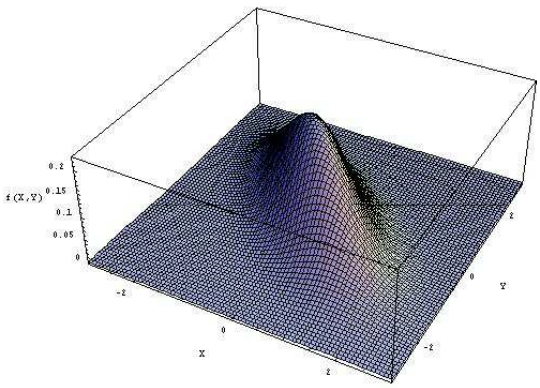
$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}\left[z_X^2 - z_X z_Y + z_Y^2\right]\right\}$$

mit

$$z_X = \frac{x - \mu_X}{\sigma_X}, \qquad z_Y = \frac{y - \mu_Y}{\sigma_Y}$$

siehe Abbildung A.5.





$\mu_X = \mu_Y = 0,\ \sigma_X^2 = \sigma_Y^2 = 1;\ \rho = 0$ $\quad$ $\mu_X = \mu_Y = 0,\ \sigma_X^2 = \sigma_Y^2 = 1;\ \rho = -0.7$

**Abbildung A.5:** Bivariate Normalverteilungen

**Übungsbeispiele:**

1. Zwei Zufallsvariablen $X$ und $Y$ seien bivariat normalverteilt mit $X \sim \mathrm{N}(2,3)$, $Y \sim \mathrm{N}(1,2)$, und $\mathrm{cov}(X,Y) = 0.5$. bzw. in Matrixschreibweise

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right]$$

Wie ist dann eine Zufallsvariable $V = 2X + 3Y$ verteilt?

$V$ ist univariat normalverteilt mit

$$\begin{aligned} \mathrm{E}(V) &= \mathrm{E}(2X + 3Y) = 2\,\mathrm{E}(X) + 3\,\mathrm{E}(Y) \\ &= 2 \times 2 + 3 \times 1 \\ &= 7 \end{aligned}$$

und

$$\begin{aligned} \mathrm{var}(V) &= \mathrm{var}(2X + 3Y) \\ &= \mathrm{var}(2X) + \mathrm{var}(3Y) + 2\,\mathrm{cov}(2X, 3Y) \\ &= 2^2 \underbrace{\mathrm{var}(X)}_{3} + 3^2 \underbrace{\mathrm{var}(Y)}_{2} + 2 \times 2 \times 3 \underbrace{\mathrm{cov}(X,Y)}_{0.5} \\ &= 12 + 18 + 6 = 36 \end{aligned}$$

Also ist $V \sim \mathrm{N}(7, 36)$.

Dies kann einfach mit R simuliert werden. Der Befehl `mvrnorm` aus dem MASS package erzeugt Zufallszahlen (Realisationen) aus einer multivariaten Normalverteilung. Der folgende Code demonstriert das obige Resultat anhand von 10000 Replikationen.

```
library(MASS)
mu = c(2,1)
Sigma <- matrix(c(3,0.5,0.5,2), nrow=2, ncol=2)
set.seed(1234567)
XY <- mvrnorm(n=10000, mu, Sigma)
V <- 2*XY[,1] + 3*XY[,2]
mean(V)  #  -> 7.05138
var(V)   #  -> 36.11783
```

2. Erzeugen Sie in einem Computerprogramm zwei unabhängige Zufallsvariablen $X \sim \mathrm{N}(2,3)$, $Y \sim \mathrm{N}(1,2)$, und $\mathrm{cov}(X,Y) = 0$, d.h.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\right]$$

mit je 100,000 Beobachtungen, und erzeugen Sie eine Grafik mit den Histogrammen.

Berechnen Sie $V = 2X + 3Y$ und zeigen Sie das Histogramm. Welchen Mittelwert und welche Varianz von $V$ erwarten Sie? Stimmen die erwarteten Werte näherungsweise mit den empirischen Werten überein?

Wenn $Z \sim \mathrm{N}(0,1)$ ist $X = 2 + \sqrt{3}\,Z \sim \mathrm{N}(2,3)$, da $\mathrm{var}(X) = \mathrm{var}(2 + \sqrt{3}\,Z) = \sqrt{3}^2\,\mathrm{var}(Z) = 3$.

In R wird eine standardnormalverteilte Zufallszahl z.B. mit der Funktion `rnorm()` erzeugt.

Der folgende R Code erzeugt zuerst 10000 Realisationen zweier unabhängiger Zufallsvariablen $X \sim \mathrm{N}(\mu = 2, \sigma^2 = 3)$ und $Y \sim \mathrm{N}(\mu = 1, \sigma^2 = 2)$ mit $\mathrm{cov}(X,Y) = 0$ und daraus eine dritte Zufallsvariable $Z = 2 * X + 3 * Y$.

Für $Z$ erwarten wir $\mathrm{E}(Z) = 2 * \mathrm{E}(X) + 3 * \mathrm{E}(Y) = 7$ und für die Varianz

$$\begin{aligned}
\mathrm{var}(Z) &= \mathrm{var}(2X + 3Y) \\
&= \mathrm{var}(2X) + \mathrm{var}(3Y) + 2\,\mathrm{cov}(2X, 3Y) \\
&= 2^2\,\underbrace{\mathrm{var}(X)}_{3} + 3^2\,\underbrace{\mathrm{var}(Y)}_{2} + 2 \times 2 \times 3\,\underbrace{\mathrm{cov}(X,Y)}_{0} \\
&= 12 + 18 + 0 = 30 \\
\text{Std.Dev.}(Z) = \sqrt{\mathrm{var}(Z)} &= \sqrt{30} = 5.477 \approx 5.424
\end{aligned}$$

Wenig überraschend bestätigt die folgende einfache Simulation diese Vermutung.

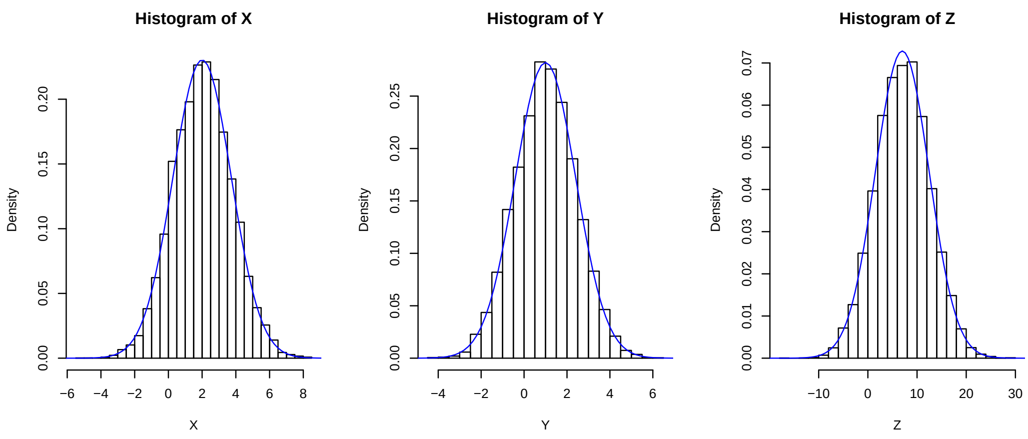
```
library(stargazer)
rm(list = ls())

set.seed(123456)
X <- rnorm(n = 10000, mean = 2, sd = sqrt(3))
Y <- rnorm(n = 10000, mean = 1, sd = sqrt(2))
Z <- 2*X + 3*Y

stargazer(data.frame(X, Y, Z), type = "text",
          summary.stat = c('n', 'mean','sd','min', 'median', 'max'))

x11(width = 9, height = 4)
par(mfrow = c(1,3))
hist(X, freq = FALSE, breaks = 25)
curve(dnorm(x,mean=2,sd=sqrt(3)),-10,12, add=TRUE, col="blue")
hist(Y, freq = FALSE, breaks = 25)
curve(dnorm(x,mean=1,sd=sqrt(2)),-10,10, add=TRUE, col="blue")
hist(Z, freq = FALSE, breaks = 25)
curve(dnorm(x,mean=7,sd=sqrt(30)),-20,40, add=TRUE, col="blue")
```

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|---|---|---|---|---|---|---|
| X | 10,000 | 2.029 | 1.729 | −5.354 | 2.032 | 8.272 |
| Y | 10,000 | 1.008 | 1.415 | −4.298 | 1.009 | 6.359 |
| Z | 10,000 | 7.083 | 5.424 | −17.634 | 7.094 | 28.320 |




### A.2.1 R Funktionen

Die Dichte- (d-), Verteilungs- (p-) und Quantilfunktion (q-) der Normalverteilung erhalten Sie mit

```
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)

Arguments:
x, q vector of quantiles.
p  vector of probabilities.
n    number of observations.
mean vector of means.
sd   vector of standard deviations.
```

`rnorm()` erzeugt normalverteilte Zufallszahlen.

Details finden Sie in der Hilfefunktion von R.

## A.3 Die Chi–Quadrat Verteilung:

Die Chi–Quadrat oder $\chi^2$ Verteilung geht auf den Astronomen F.R. Helmert (1875) zurück.

**Theorem 2** *Sind $Z_1, Z_2, \ldots, Z_\nu$ unabhängig standardnormalverteilte Zufallsvariablen (d.h. normalverteilte Zufallsvariable mit Mittelwert 0 und Varianz 1), so folgt die Quadratsumme $\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2$ einer $\chi^2$ Verteilung mit $\nu$ Freiheitsgraden.*

$$Z_i \sim \mathrm{N}(0,1) \quad \Rightarrow \quad \sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2$$

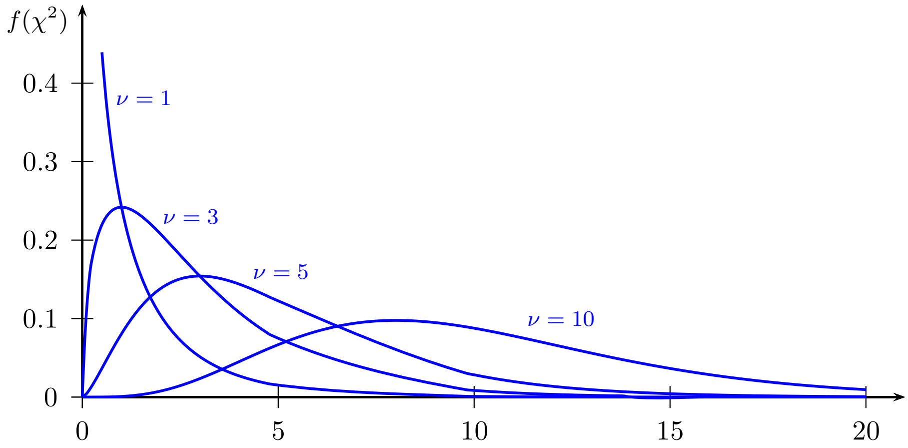



**Abbildung A.6:** Dichtefunktionen von $\chi^2$ verteilten Zufallsvariablen mit $\nu = 1, 3, 5$ und 10 Freiheitsgraden.

Die $\chi^2$ Verteilung ist nicht symmetrisch und abhängig von der Anzahl der Freiheitsgrade $\nu$. Da sie eine Quadratsumme ist kann eine $\chi^2$-verteilte Zufallsvariable natürlich nie negativ sein. Erwartungswert und Varianz sind

$$\begin{aligned} \mathrm{E}(\chi^2) &= \nu \\ \mathrm{var}(\chi^2) &= 2\nu \end{aligned}$$

Wenn $X_1$ und $X_2$ zwei unabhängig $\chi^2$-verteilte Zufallsvariablen mit $\nu_1$ bzw. $\nu_2$ Freiheitsgraden sind, so ist die Summe $X_1 + X_2$ ebenfalls $\chi^2$-verteilt mit $\nu_1 + \nu_2$ Freiheitsgraden.

Diese Verteilung wird v.a. für Tests benötigt, die Varianzen von Zufallsvariablen betreffen. Abbildung A.6 zeigt Dichtefunktionen von $\chi^2$-verteilten Zufallsvariablen mit unterschiedlichen Freiheitsgraden.

Dies kann einfach mit einer einfachen Simulation demonstriert werden. Dazu erzeugen wir mit R Matrix M mit 10 Spalten, und in jede dieser Spalte schreiben wir 10000 Realisationen einer *quadrierten* standardnormalverteilten Zufallsvariable. Die 10 Zufallsvariablen sind unabhängig.

Durch Summieren über eine entsprechende Anzahl von Spalten erhalten wir $\chi^2$-verteilte Zufallsvariablen (oder genauer, deren Realisationen).

Das folgende kleine Programm führt dies aus, und Abbildung A.7 zeigt das Ergebnis, die Simulation der theoretischen Verteilungen in Abbildung A.6 (achten Sie auf die Skala der Ordinate).

```
###############################################################
## Chi^2
set.seed(123456)
M <- rnorm(n = 10000, mean = 0, sd = 1)^2
for (i in 2:10) {
  M <- cbind(M, (rnorm(n = 10000, mean = 0, sd = 1)^2))
}
```

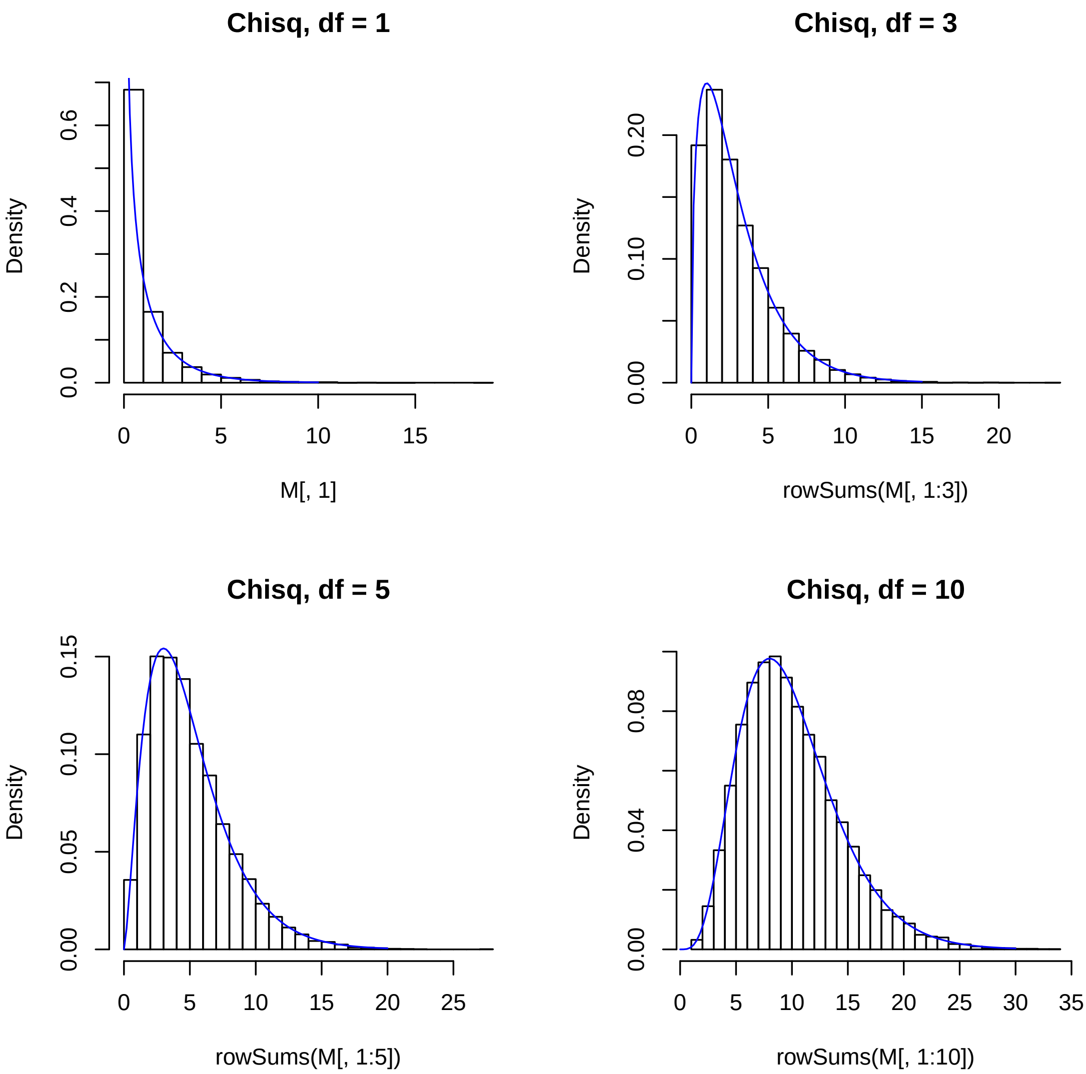



**Abbildung A.7:** Histogramme der Summe von quadrierten standardnormalverteilten Zufallsvariablen.

```
x11()
par(mfrow = c(2,2))
hist(M[ ,1], freq = FALSE, breaks = 25, main = "Chisq, df = 1")
curve(dchisq(x, df = 1), 0, 10, add=TRUE, col="blue")
hist(rowSums(M[, 1:3]), freq = FALSE, breaks = 25, main = "Chisq, df = 3")
curve(dchisq(x, df = 3), 0, 15, add=TRUE, col="blue")
hist(rowSums(M[, 1:5]), freq = FALSE, breaks = 25, main = "Chisq, df = 5")
curve(dchisq(x, df = 5), 0, 20, add=TRUE, col="blue")
hist(rowSums(M[, 1:10]), freq = FALSE, breaks = 25, main = "Chisq, df = 10")
curve(dchisq(x, df = 10), 0, 30, add=TRUE, col="blue")
```

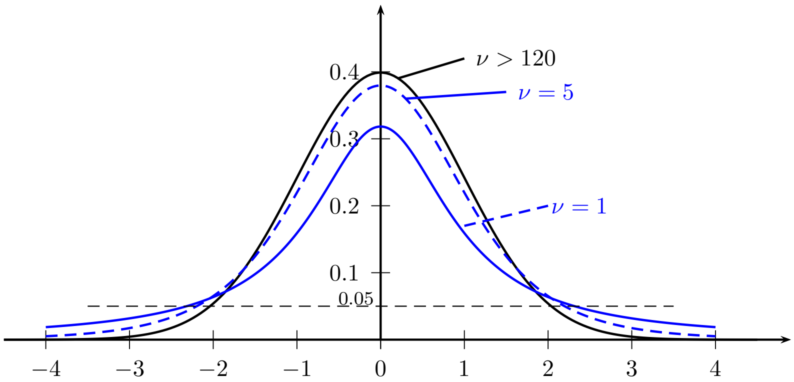



**Abbildung A.8:** Dichtefunktionen für $t$-verteilte Zufallsvariablen mit $\nu = 1,\ 5$ und $\infty$ Freiheitsgraden.

### A.3.1 R Funktionen

Die Dichte- (d-), Verteilungs- (p-) und Quantilfunktion (q-) der $\chi^2$ Verteilung erhalten Sie mit

```
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)
```

Details finden Sie in der Hilfefunktion von R.

## A.4 Die t-Verteilung:

Die t- bzw. Studentverteilung verdankt ihren Namen W.S. Gosset, der deren Ableitung 1908 unter dem Pseudonym ‘Student’ veröffentlichte.

**Theorem 3** *Sei $Z$ eine standardnormalverteilte Zufallsvariable [$Z \sim \mathrm{N}(0,1)$] und $V$ eine $\chi^2$-verteilte Zufallsvariable mit $\nu$ Freiheitsgraden [$V \sim \chi^2(\nu)$], wobei $Z$ und $V$ unabhängig voneinander verteilt sind, dann ist die Zufallsvariable $T$*

$$t = \frac{Z}{\sqrt{(V/\nu)}} \sim \mathrm{t}_\nu$$

t-*verteilt mit $\nu$ Freiheitsgraden.*

Wie aus Abbildung A.8 ersichtlich ist die t Verteilung symmetrisch und ‘flacher’ als die Standardnormalverteilung. Sie nähert sich mit steigender Zahl von Freiheitsgraden der Standardnormalverteilung an und ist für $\nu \geq 120$ nahezu mit dieser identisch.
Weiters ist

$$\begin{aligned} \mathrm{E}(t) &= 0 \qquad \text{für } \nu > 1 \\ \mathrm{var}(t) &= \frac{\nu}{\nu - 2} \quad \text{für } \nu > 2 \end{aligned}$$

Wir werden die $t$-Verteilung v.a. benötigen, wenn die Varianz der Grundgesamtheit nicht bekannt ist, sondern aus der Stichprobe geschätzt werden muss.

### A.4.1 R Funktionen

Die Dichte- (d-), Verteilungs- (p-) und Quantilfunktion (q-) der t Verteilung erhalten Sie mit

```
dt(x, df, ncp, log = FALSE)
pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)
qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)
rt(n, df, ncp)
```

Mit `rt()` können Sie Zufallszahlen erzeugen.
Details finden Sie in der Hilfefunktion von R.

## A.5 Die F-Verteilung:

Die F Verteilung werden wir später häufig benötigen, z.B. für Tests, die mehrere Regressionskoeffizienten betreffen, oder um die Gleichheit zweier Varianzen zu testen. Sie ist nach R.A. Fisher benannt.

**Theorem 4** *Wenn $V_1$ und $V_2$ zwei unabhängig $\chi^2$-verteilte Zufallsvariablen mit $\nu_1$ bzw. $\nu_2$ Freiheitsgraden sind, dann ist*

$$F = \frac{V_1/\nu_1}{V_2/\nu_2} \sim \mathrm{F}_{\nu_1,\nu_2}$$

F-*verteilt mit $\nu_1$ Zähler- und $\nu_2$ Nennerfreiheitsgraden.*

Wie aus Abbildung A.9 ersichtlich ist die F Verteilung schief. Für große $\nu_1$ und $\nu_2$ nähert sie sich einer Normalverteilung an.
Man kann zeigen, dass

$$\begin{aligned} \mathrm{E}(F_{\nu_1,\nu_2}) &= \frac{\nu_2}{\nu_2 - 2} \qquad \text{für } \nu_2 > 2 \\ \mathrm{var}(F_{\nu_1,\nu_2}) &= \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \quad \text{für } \nu_2 > 4 \end{aligned}$$

Das Quadrat einer t-verteilten Zufallsvariable mit $\nu$ Freiheitsgraden folgt einer $F$ Verteilung mit 1 Zähler- und $\nu$ Nennerfreiheitsgraden, d.h.

$$\mathrm{t}_\nu^2 \sim \mathrm{F}_{1,\nu}$$

### A.5.1 R Funktionen

Die Dichte- (d-), Verteilungs- (p-) und Quantilfunktion (q-) der F Verteilung erhalten Sie mit

```
df(x, df1, df2, ncp, log = FALSE)
pf(q, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
qf(p, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
rf(n, df1, df2, ncp)
```

Mit `rf()` können Sie Zufallszahlen erzeugen.
Details finden Sie in der Hilfefunktion von R.

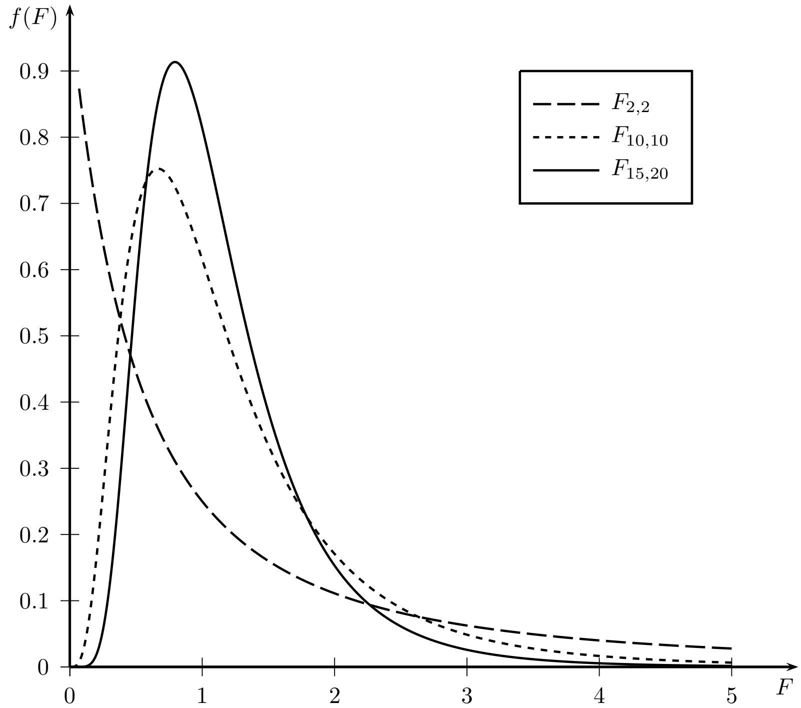



**Abbildung A.9:** Dichtefunktionen für $F$-verteilte Zufallsvariablen mit unterschiedlichen Freiheitsgraden.