Symmetrie und Schiefe

Andreas Handl

Inhaltsverzeichnis

1 Was ist Symmetrie?

Wir gehen aus von einer stetigen Zufallsvariablen X mit Dichtefunktion $f_X(x)$ und Verteilungsfunktion $F_X(X)$. Schauen wir uns exemplarisch die Dichtefunktion einer mit den Parametern μ und σ^2 normalverteilten Zufallsvariablen an. Es gilt

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Die Dichtefunktion nimmt für $(x-\mu)^2 = t^2$ mit $t \geq 0$ identische Werte an. Es gilt

$$(x-\mu)^2 = t^2 \quad \Longleftrightarrow \quad |x-\mu| = t \quad \Longleftrightarrow \quad x-\mu = \pm t$$

Somit sind die Werte der Dichtefunktion in $x = \mu - t$ und $x = \mu + t$ für $t \geq 0$ identisch. Abbildung 1 verdeutlicht dies exemplarisch für $t = 1$.

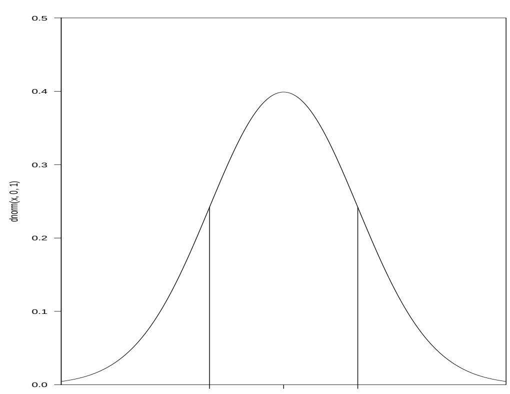


Abbildung 1: Die Dichtefunktion der Normalverteilung

Definition 1.1
Sei X eine stetige Zufallsvariable mit Dichtefunktion $f_X(x)$. Die Verteilung von X heißt **symmetrisch** bezüglich θ, wenn für alle $t \in \mathbb{R}$ gilt

$$f_X(\theta - t) = f_X(\theta + t) \tag{2}$$

Die Gleichverteilung, die logistische Verteilung, die Laplaceverteilung und die Cauchyverteilung sind symmetrische Verteilungen. Abbildung 2 zeigt die Dichtefunktionen dieser Verteilungen.

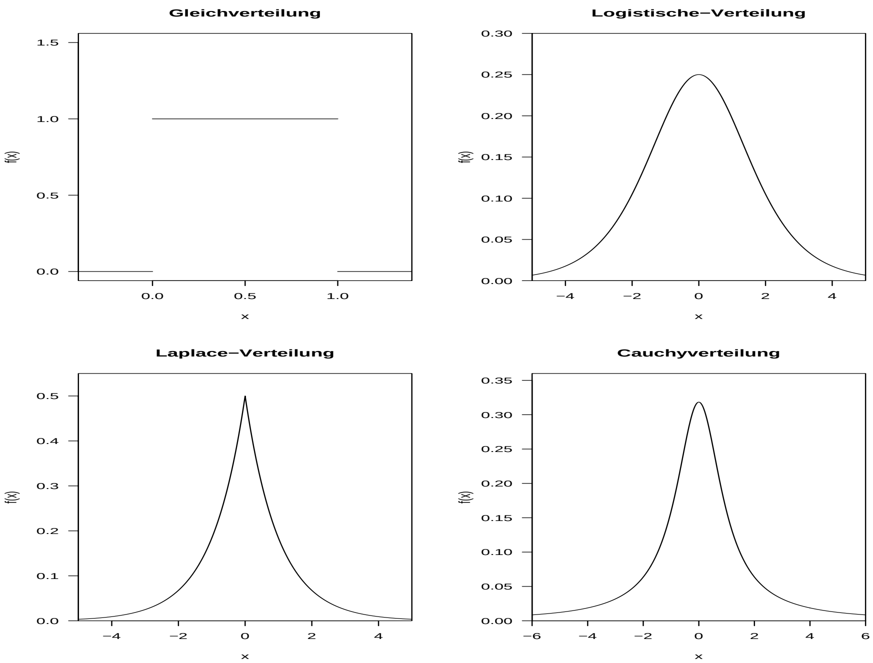


Abbildung 2: Symmetrische Verteilungen

Wir können Symmetrie auch über die Verteilungsfunktion definieren. Dies kann man sich an Abbildung 1 klarmachen. Die Fläche unterhalb von $\theta - t$ ist gleich der Fläche oberhalb von $\theta + t$. Dies führt zu folgender

Definition 1.2
Sei X eine stetige Zufallsvariable mit Verteilungsfunktion $F_X(x)$. Die Verteilung von X heißt symmetrisch bezüglich θ, wenn für alle $t \in \mathbb{R}$ gilt

$$F_X(\theta - t) = 1 - F_X(\theta + t) \tag{3}$$

Da die Fläche unter der Dichtefunktion durch die Gerade, die durch θ parallel zur Ordinate verläuft, halbiert wird, gilt $\theta = x_{0.5}$. Abbildung 3 verdeutlicht, dass wir die Symmetrie über Quantile auch folgendermaßen definieren können:

Definition 1.3
Sei X eine stetige Zufallsvariable mit Verteilungsfunktion $F_X(x)$. Die Verteilung von X heißt symmetrisch bezüglich θ, wenn für alle $p \in (0, 0.5)$ gilt

$$x_{0.5} - x_p = x_{1-p} - x_{0.5} \tag{4}$$

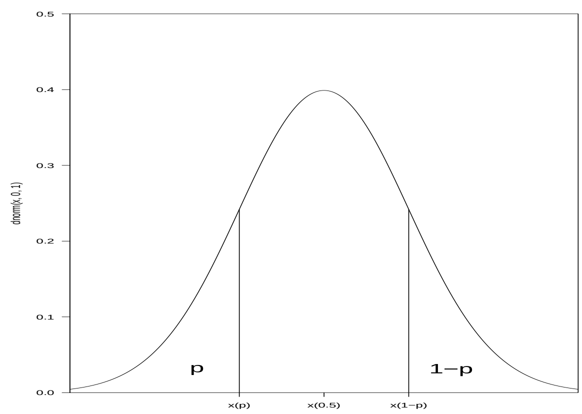


Abbildung 3: Definition von Symmetrie über Quantile

Eine Verteilung, die nicht symmetrisch ist, heißt **schief**. Ein Beispiel für eine schiefe Verteilung ist die Exponentialverteilung mit Parameter λ, deren Dichtefunktion folgendermaßen definiert ist:

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

Die Grafik links oben in Abbildung 4 zeigt die Dichtefunktion der Exponentialverteilung mit $\lambda = 1$.
Man unterscheidet rechtsschiefe und linksschiefe Verteilungen.
Eine Verteilung heißt **rechtsschief** bzw. **linkssteil**, wenn für alle $p \in (0, 1)$ gilt

$$x_{0.5} - x_p < x_{1-p} - x_{0.5} \tag{5}$$

Entsprechend heißt eine Verteilung **linksschief** bzw. **rechtssteil**, wenn für alle $p \in (0, 1)$ gilt

$$x_{0.5} - x_p > x_{1-p} - x_{0.5} \tag{6}$$

Die Exponentialverteilung ist rechtsschief. Abbildung 4 zeigt neben der Dichtefunktion der Exponentialverteilung die Dichtefunktion der Gammaverteilung, der Lognormal-Verteilung und der Weibull-Verteilung.

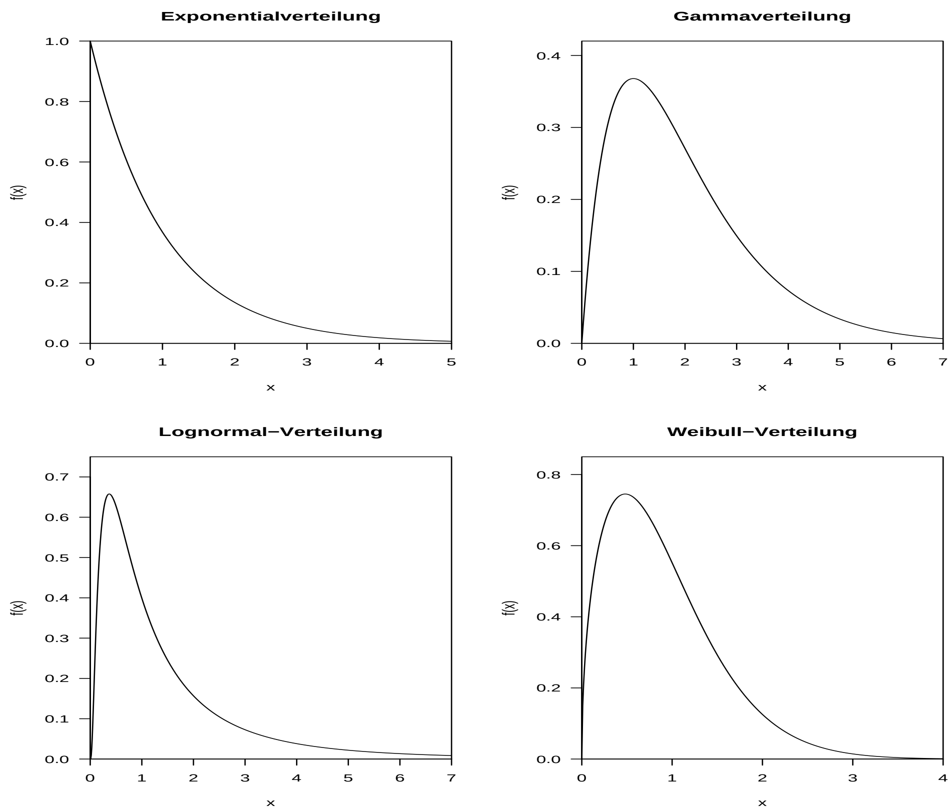


Abbildung 4: Schiefe Verteilungen

2 Wozu benötigt man Symmetrie?

Bei einer symmetrischen Verteilung ist der Lageparameter eindeutig. Man wählt das Symmetriezentrum. Existiert der Erwartungswert $E(X)$, so gilt bei einer symmetrischen Verteilung

$$E(X) = x_{0.5}$$

Bei einer schiefen Verteilung unterscheiden sich Erwartungswert und Median. Bei einer rechtsschiefen Verteilung gilt $x_{0.5} < E(X)$. Schauen wir uns exemplarisch die Exponentialverteilung an. Es gilt $x_{0.5} = \frac{1}{\lambda} \ln 2 = \frac{0.693}{\lambda}$. Wegen $E(X) = \frac{1}{\lambda}$ gilt $x_{0.5} < E(X)$.

Bei einer linksschiefen Verteilung gilt $x_{0.5} > E(X)$.

Will man also die Lage einer schiefen Verteilung beschreiben, so muss man angeben, welchen Lageparameter man benutzt.

Bei einigen statistischen Tests wird unterstellt, dass die Verteilung der Grundgesamtheit symmetrisch ist. Schauen wir uns exemplarisch den Wilcoxon-Vorzeichen-Rangtest an. Dieser ist ein Test auf einen Lageparameter θ im Einstichprobenproblem. Wir gehen aus von einer Zufallsstichprobe vom Umfang n aus einer Grundgesamtheit, deren Verteilung stetig und symmetrisch bezüglich θ ist. Es soll getestet werden

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta > \theta_0 \,. \tag{7}$$

Wir setzen im Folgenden θ gleich 0. Wollen wir auf einen Wert $\theta_0 \neq 0$ testen, so betrachten wir die Beobachtungen $x_1 - \theta_0, \ldots, x_n - \theta_0$.

Beispiel 1
Gegeben seien die Beobachtungen

$$x_1 = -0.8 \quad x_2 = -0.5 \quad x_3 = 0.4 \quad x_4 = 0.9 \quad x_5 = 1.2 \quad x_6 = 1.7$$

Der Wilcoxon-Vorzeichen-Rangtest verwendet zwei Informationen:

1. die Vorzeichen s_i der Beobachtungen
2. die Abstände $|x_i|$ der Beobachtungen vom Nullpunkt

Dabei ist

$$s_i = \begin{cases} 1 & \text{falls } x_i > 0 \\ 0 & \text{sonst} \end{cases}$$

Beispiel 1 (fortgesetzt)
Es gilt

$$s_1 = 0 \quad s_2 = 0 \quad s_3 = 1 \quad s_4 = 1 \quad s_5 = 1 \quad s_6 = 1$$

und

$$|x_1| = 0.8 \quad |x_2| = 0.5 \quad |x_3| = 0.4 \quad |x_4| = 0.9 \quad |x_5| = 1.2 \quad |x_6| = 1.7$$

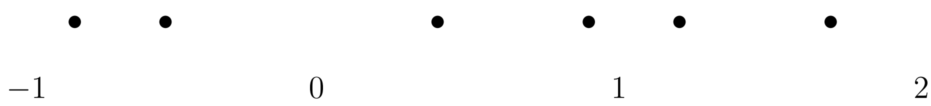
Beim Wilcoxon-Vorzeichen-Rangtest werden die Ränge R_i der $|x_i|$ betrachtet.

$$R_1 = 3 \quad R_2 = 2 \quad R_3 = 1 \quad R_4 = 4 \quad R_5 = 5 \quad R_6 = 6$$

Die Teststatistik des Wilcoxon-Vorzeichen-Rangtests ist gleich der Summe der Ränge der $|x_i|$.

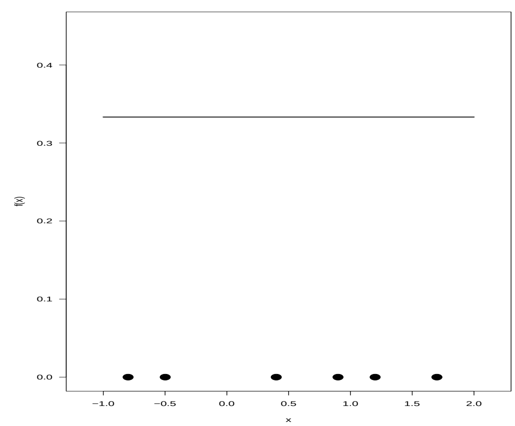
$$W^+ = \sum_{i=1}^{n} s_i \, R_i$$

Unter welchen Bedingungen ist der Wilcoxon-Vorzeichen-Rangtest ein geeigneter Test für die Hypothesen in Gleichung (7)? Zur Beantwortung dieser Frage schauen wir uns eine Abbildung der Daten an.



Diese Stichprobe kann aus unterschiedlichen Grundgesamtheiten stammen. Zum einen kann die Verteilung der Grundgesamtheit symmetrisch sein. Dies ist in Abbildung 5 der Fall.

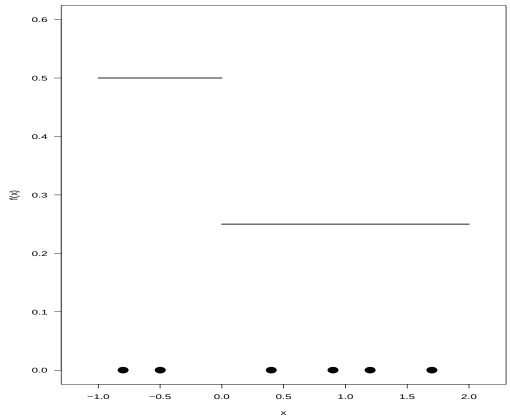
Abbildung 5: Daten mit Gleichverteilung



Kommen die Daten aus der Gleichverteilung, so ist der Wert des Lageparameters gleich 0.5. Somit ist H_1 erfüllt. Der Wilcoxon-Vorzeichen-Rangtest ist geeignet, diese Lagealternative aufzudecken, da die Mehrzahl der positiven Beobachtungen weiter vom Nullpunkt entfernt sind als die negativen Beobachtungen. Also sind auch die Ränge der positiven Bobachtungen größer. Dies führt zu einem großen Wert von W^+.

Ist die Verteilung der Grundgesamtheit aber schief, so ändert sich die Interpretation. Abbildung 6 zeigt eine schiefe Verteilung, die ein für die Daten angemessenes Modell darstellt.

Abbildung 6: Daten mit schiefer Verteilung



Wir sehen, dass der Median dieser Verteilung gleich 0 ist. Auf Grund der Konstellation der Beobachtungen würde der Wilcoxon-Vorzeichen-Rangtest die Hypothese H_0 aber ablehnen. Ist die Verteilung der Grundgesamtheit also schief, so muss eine andere Hypothese betrachtet werden, für die der Wilcoxon-Vorzeichen-Rangtest geeignet ist. Er ist in diesem Fall ein Test auf Symmetrie bezüglich eines bekannten Symmetriezentrums.

Um den Wilcoxon-Vorzeichen-Rangtest als Test auf einen Lageparameter auffassen zu können, benötigen wir also die Symmetrie der Verteilung der Grundgesamtheit.

3 Maßzahlen für die Schiefe einer Verteilung

Die klassischen Maßzahlen für die Lage und die Variabilität einer Verteilung sind der Erwartungswert $E(X)$ und die Varianz $Var(X) = E(X^2) - E(X)^2$. Die dabei auftretenden Größen $E(X)$ und $E(X^2)$ sind spezielle Momente der Zufallsvariablen X.

Definition 3.1
Sei X eine Zufallsvariable. Dann heißt

$$\mu_r = E(X^r) \tag{8}$$

r-tes Moment von X.

Neben den Momenten sind noch die zentralen Momente von Interesse.

Definition 3.2
Sei X eine Zufallsvariable mit Erwartungswert μ. Dann heißt

$$\mu'_r = E\left[(X - E(X))^r\right] \tag{9}$$

r-tes zentrales Moment von X.

Offensichtlich ist das erste zentrale Moment gleich Null und das zweite zentrale Moment gleich der Varianz.
Höhere zentrale Moment beschreiben bestimmte Charakteristika einer Verteilung. So ist μ'_3 bei einer symmetrischen Verteilung gleich 0. Für eine bezüglich μ symmetrische Verteilung gilt nämlich für alle $t \in \mathbb{R}$:

$$f_X(\mu - t) = f_X(\mu + t) \tag{10}$$

Somit gilt

$$\begin{aligned}
\mu'_3 &= \int\limits_{-\infty}^{\infty} (x-\mu)^3\, f_X(x)\, dx \\
&= \int\limits_{-\infty}^{\mu} (x-\mu)^3\, f_X(x)\, dx + \int\limits_{\mu}^{\infty} (x-\mu)^3\, f_X(x)\, dx \\
&\overset{y=x-\mu}{=} \int\limits_{-\infty}^{0} y^3\, f_X(\mu + x)\, dy + \int\limits_{0}^{\infty} y^3\, f_X(\mu + x)\, dx
\end{aligned}$$

Wir substituieren $y = -t$ im ersten Summanden und erhalten

$$\begin{aligned}
\mu'_3 &= -\int\limits_{0}^{\infty} t^3\, f_X(\mu - t)\, dt + \int\limits_{0}^{\infty} y^3\, f_X(x + \mu)\, dx \\
&\overset{(10)}{=} -\int\limits_{0}^{\infty} y^3\, f_X(\mu + y)\, dy + \int\limits_{0}^{\infty} y^3\, f_X(x + \mu)\, dx \\
&= 0
\end{aligned}$$

Ist also $\mu'_3 \neq 0$, so ist die Verteilung schief. Auf Grund dieser Eigenschaft kann man μ'_3 als Maßzahl für die Schiefe auffassen. Diese hat jedoch den Nachteil, dass sie von der Skalierung der Daten abhängt. Es gilt nämlich

$$\begin{aligned}
E[(a\,X - E(a\,X))^3] &= E[(a\,X - a\,E(X))^3] = E[(a(X - E(X)))^3] \\
&= E[a^3 (X - E(X))^3] = a^3\, E[(X - E(X))^3]
\end{aligned}$$

Die folgende Definition gibt eine auf μ'_3 basierende Maßzahl für die Schiefe an, die skalenunabhängig ist.

Definition 3.3
Die Schiefe einer Zufallsvariablen X ist definiert durch

$$\gamma_1 = \frac{E((X-\mu)^3)}{\sigma^3}. \tag{11}$$

In Tabelle 1 sind die Werte von γ_1 für ausgewählte schiefe Verteilungen zu finden.

Verteilung	γ_1	QS	τ_3
Exponentialverteilung	2.00	0.26	0.33
Gammaverteilung (r=2)	0.71	0.17	0.16
Lognormalverteilung	6.19	0.33	0.46

Tabelle 1: Werte von γ_1 für ausgewählte schiefe Verteilungen

Positive Werte von γ_1 sprechen für eine rechtsschiefe Verteilung, während negative Werte auf eine linksschiefe Verteilung hindeuten. Ist γ_1 gleich 0, so ist die Verteilung nicht notwendigerweise symmetrisch.
Die Maßzahl γ_1 hat einige Nachteile. Sie ist schwer zu interpretieren. Außerdem muss sie nicht existieren. Dies ist bei der Cauchyverteilung der Fall. Da der Erwartungswert der Cauchyverteilung nicht existiert, existiert auch nicht γ_1.
Aus einer Stichprobe $x_1, \ldots, x_n$ schätzt man γ_1 durch

$$\hat{\gamma}_1 = \frac{\hat{\mu}'_3}{\hat{\mu}'^{1.5}_2}. \tag{12}$$

Dabei ist

$$\hat{\mu}'_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r \tag{13}$$

Beispiel 2
Studenten wurden in einer Vorlesung gefragt, wie viele CDs sie besitzen. Hier sind die Daten von 10 Studierenden:

```
10 20 30 40 60 70 90 150 200 300
```

Es gilt $\hat{\gamma}_1 = 1.15$.

Wie das folgende Beispiel zeigt, ist $\hat{\gamma}_1$ nicht robust. Ein Ausreißer hat einen starken Einfluss auf den Wert von $\hat{\gamma}_1$.

Beispiel 3
Für den Datensatz

```
-3 -1 0 1 3
```

ist $\hat{\gamma}_1$ gleich 0.
Nimmt die fünfte Beobachtung den Wert 8 an, so erhalten wir $\hat{\gamma}_1 = 1.03$.

Royston (1992) zeigt mit Simulationsstudien, dass $\hat{\gamma}_1$ für kleine Stichprobenumfänge einen großen Bias besitzt.
Da $\hat{\gamma}_1$ viele Nachteile besitzt, sollte man eine andere Maßzahl für die Schiefe berechnen. Einige dieser Maßzahlen basieren auf Quantilen. Ausgangspunkt ist Gleichung (4) auf Seite 3. Subtrahieren wir bei dieser Gleichung auf beiden Seiten $x_{0.5} - x_p$, so erhalten wir eine Größe, die als Maßzahl für die Schiefe aufgefasst werden kann:

$$(x_{1-p} - x_{0.5}) - (x_{0.5} - x_p) \tag{14}$$

Bowley (1920) setzt $p = 0.25$ und betrachtet

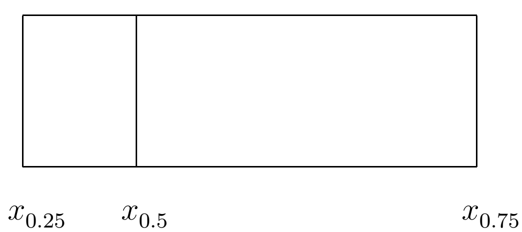
$$QS = \frac{x_{0.75} - x_{0.5} - (x_{0.5} - x_{0.25})}{x_{0.75} - x_{0.25}} = \frac{x_{0.75} + x_{0.25} - 2x_{0.5}}{x_{0.75} - x_{0.25}} \tag{15}$$

Die Division durch den Interquartilsabstand bewirkt, dass QS unabhängig von der Skalierung der Daten ist. Offensichtlich liegt QS zwischen -1 und 1. Werte von QS für ausgewählte Verteilungen sind in Tabelle 1 auf Seite 10 zu finden. Man erhält Schätzer für QS und die anderen quantilbasierten Schiefemaße, indem man die Quantile schätzt und in die Formeln einsetzt. Man schätzt QS also, indem man $\hat{x}_{0.25}$, $\hat{x}_{0.5}$ und $\hat{x}_{0.75}$ schätzt und in die Formel (15) einsetzt:

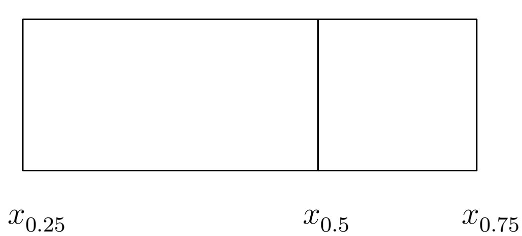
$$\widehat{QS} = \frac{\hat{x}_{0.75} - \hat{x}_{0.5} - (\hat{x}_{0.5} - \hat{x}_{0.25})}{\hat{x}_{0.75} - \hat{x}_{0.25}} = \frac{\hat{x}_{0.75} + \hat{x}_{0.25} - 2\hat{x}_{0.5}}{\hat{x}_{0.75} - \hat{x}_{0.25}} \tag{16}$$

Beim Boxplot zeichnet man ein Rechteck, das vom unteren Quartil $x_{0.25}$ bis zum oberen Quartil $x_{0.75}$ verläuft. Den Median markiert man im Rechteck als vertikale Linie. Der Median teilt das Rechteck in zwei kleinere Rechtecke. Der Nenner der Maßzahl von Bowley ist gleich der Länge des großen Rechtecks, während der Zähler gleich der Differenz aus der Länge des rechten Rechtecks und der Länge des linken Rechtecks ist. Bei einer symmetrischen Verteilung teilt der Median das Rechteck in zwei gleich große Rechtecke.

In diesem Fall ist QS gleich 0.
Bei einer rechtsschiefen Verteilung ist das linke Rechteck kleiner als das rechte.



In diesem Fall ist QS größer als 0.
Bei einer linksschiefen Verteilung ist das linke Rechteck größer als das rechte.



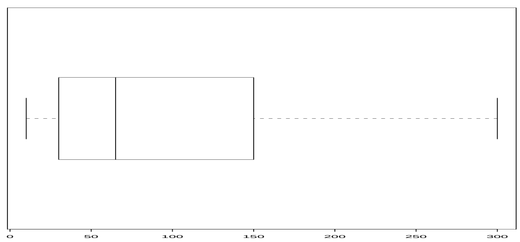
In diesem Fall ist QS kleiner als 0.

Beispiel 2 (fortgesetzt)
Wir bestimmen $x_{0.25}$ und $x_{0.75}$ so, wie es Tukey vorgeschlagen hat. Es gilt $x_{0.25} = 30$ und $x_{0.75} = 150$. Außerdem gilt $x_{0.5} = 65$.
Somit gilt

$$QS = \frac{150 - 65 - (65 - 30)}{150 - 30} = 0.42$$

Abbildung 7 zeigt den Boxplot.

Abbildung 7: Boxplot der Anzahl Cds



Hinley (1975) hat diesen Schätzer verallgemeinert

$$H = \frac{x_{1-p} - x_{0.5} - (x_{0.5} - x_p)}{x_{1-p} - x_p} \tag{17}$$

Brys et al. (2003) setzen $p = 0.125$.

$$H = \frac{x_{0.875} - x_{0.5} - (x_{0.5} - x_{0.125})}{x_{0.875} - x_{0.125}} \tag{18}$$

Handl (1985) betrachtet nicht die Differenz sondern den Quotienten aus $x_{1-p} - x_{0.5}$ und $x_{0.5} - x_p$:

$$Q = \frac{x_{1-p} - x_{0.5}}{x_{0.5} - x_p} \tag{19}$$

Sinnvolle Werte für p sind hier 0.25 oder 0.125.
Hosking (1990) wählte einen anderen Zugang zu einer Maßzahl für die Schiefe. Er geht aus von den geordneten Beobachtungen einer Stichprobe vom Umfang n, die er mit $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ bezeichnet. Er definiert für $r = 1, 2, \ldots$ sogenannte L-Momente

$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E\left(X_{r-k:r}\right) \tag{20}$$

Für $r = 1, 2, 3$ gilt

$$\lambda_1 = E(X_{1:1}) \tag{21}$$

$$\lambda_2 = \frac{1}{2}[E(X_{2:2}) - E(X_{1:2})] \tag{22}$$

$$\lambda_3 = \frac{1}{3}[E(X_{3:3}) - 2\,E(X_{2:3}) + E(X_{1:3})] \tag{23}$$

Man kann λ_3 folgendermaßen umformen:

$$\begin{aligned}
\lambda_3 &= \frac{1}{3}[E(X_{3:3}) - 2\,E(X_{2:3}) + E(X_{1:3})] \\
&= \frac{1}{3}[E(X_{3:3}) - E(X_{2:3}) - E(X_{2:3}) + E(X_{1:3})] \\
&= \frac{1}{3}[E(X_{3:3}) - E(X_{2:3}) - (E(X_{2:3}) - E(X_{1:3}))])
\end{aligned}$$

Somit ist λ_3 bei einem Stichprobenumfang von $n = 3$ gleich der Differenz aus der erwarteten Distanz zwischen Maximum und Median und der erwarteten Distanz aus Median und Minimum. Bei einer symmetrischen Verteilung nimmt diese Differenz gleich den Wert 0 an, während sie bei einer schiefen Verteilung ungleich 0 ist. Ein Schätzer von λ_3 ist

$$l_3 = \frac{1}{3}\ [x_{3:3} - 2x_{2:3} + x_{1:3}] \tag{24}$$

Um λ_3 auf Basis einer Zufallsstichprobe $x_1, \ldots, x_n$ zu schätzen, bestimmt er für jede geordnete Teilstichprobe $x_{(i)}, x_{(j)}, x_{(k)}$ aus $x_1, \ldots, x_n$ den Schätzwert

$$\frac{1}{3}\ \left[x_{(3)} - 2x_{(2)} + x_{(1)}\right] \tag{25}$$

Als Schätzer l_3 für λ_3 auf Basis der n Beobachtungen dient dann der Mittelwert der Schätzwerte der Teilstichproben:

$$l_3 = \frac{1}{3}\binom{n}{3}^{-1} \sum_{i<j<k} (x_{(k)} - 2x_{(j)} + x_{(i)}) \tag{26}$$

Beispiel 2 (fortgesetzt)
Wir betrachten aus Gründen der Übersichtlichkeit die letzten 5 Beobachtungen

```
70   90  150  200  300
```

Mit $\binom{5}{3} = 10$ gilt

$$\begin{aligned}
l_3 &= \frac{1}{30}[(150 - 2 \cdot 90 + 70) + (200 - 2 \cdot 90 + 70) + (300 - 2 \cdot 90 + 70) \\
&+ (200 - 2 \cdot 150 + 70) + (300 - 2 \cdot 150 + 70) + (300 - 2 \cdot 200 + 70) \\
&+ (200 - 2 \cdot 150 + 90) + (300 - 2 \cdot 150 + 90) + (300 - 2 \cdot 200 + 90) \\
&+ (300 - 2 \cdot 200 + 150)] \\
&= 15
\end{aligned}$$

Hosking (1990) betrachtet folgende Maßzahl für die Schiefe

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \tag{27}$$

Einen Schätzer für τ_3 erhält man, indem man λ_3 und λ_2 schätzt:

$$\hat{\tau}_3 = \frac{l_3}{l_2} \tag{28}$$

Den Schätzer von λ_3 kennen wir bereits. Wenn wir beim Schätzer l_2 von λ_2 das gleiche Prinzip anwenden, erhalten wir

$$l_2 = \frac{1}{2}\binom{n}{2}^{-1} \sum_{i<j}(x_{(j)} - x_{(i)}) \tag{29}$$

Beispiel 2 (fortgesetzt)
Mit $\binom{5}{2} = 10$ gilt

$$\begin{aligned}
l_2 &= \frac{1}{20}[90 - 70 + 150 - 70 + 200 - 70 + 300 - 70 + 150 - 90 \\
&+ 200 - 90 + 300 - 90 + 200 - 150 + 300 - 150 + 300 - 200] \\
&= 57
\end{aligned}$$

Somit gilt

$$\hat{\tau}_3 = \frac{l_3}{l_2} = \frac{15}{57} = 0.26$$

Hosking (1990) zeigt, dass man l_2 und l_3 folgendermaßen in Abhängigkeit von den geordneten Beobachtungen $x_{(1)}, \ldots, x_{(n)}$ und $\bar{x}$ darstellen kann:

$$\begin{aligned}
l_2 &= 2w_2 - \bar{x} \\
l_3 &= 6w_3 - 6w_2 + \bar{x}
\end{aligned}$$

mit

$$w_2 = \frac{1}{n(n-1)} \sum_{i=2}^{n} (i-1)\, x_{(i)}$$

und

$$w_3 = \frac{1}{n(n-1)(n-2)} \sum_{i=2}^{n} (i-1)(i-2)\, x_{(i)}$$

Beispiel 2 (fortgesetzt)
Wir betrachten wieder die letzten 5 Beobachtungen

```
70   90  150  200  300
```

Es gilt $\bar{x} = 162$. Außerdem gilt

$$w_2 = \frac{1}{5 \cdot 4}\, [90 + 2 \cdot 150 + 3 \cdot 200 + 4 \cdot 300] = 109.5$$

und

$$w_3 = \frac{1}{5 \cdot 4 \cdot 3}\, [2 \cdot 150 + 6 \cdot 200 + 12 \cdot 300] = 85$$

Also gilt

$$\begin{aligned} l_2 &= 2 \cdot 109.5 - 162 = 57 \\ l_3 &= 6 \cdot 85 - 6 \cdot 109.5 + 162 = 15 \end{aligned}$$

4 Ein Test auf Symmetrie

Die von Hosking (1990) vorgeschlagene Maßzahl t_3 für die Schiefe einer Verteilung beruht für eine geordnete Stichprobe $x_{(1)}, x_{(2)}, x_{(3)}$ vom Umfang 3 auf folgender Größe

$$x_{(1)} + x_{(3)} - 2\,x_{(2)}$$

Bereits 1980 wurde diese Größe von Randles et al. (1980) benutzt, um einen Test auf Symmetrie zu konstruieren. Sie betrachten also folgende Hypothesen

H_0 : Die Verteilung der Grundgesamtheit ist symmetrisch
H_1 : Die Verteilung der Grundgesamtheit ist schief

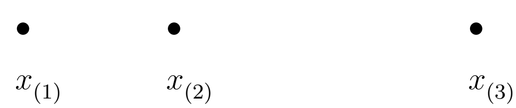
Randles et al. (1980) nennen das Tripel $x_{(1)}, x_{(2)}, x_{(3)}$ ein **rechtes Tripel**, wenn gilt

$$x_{(1)} + x_{(3)} - 2\,x_{(2)} > 0$$

Für ein rechtes Tripel gilt also

$$x_{(3)} - x_{(2)} > x_{(2)} - x_{(1)}$$

Die folgende Abbildung veranschaulicht den Sachverhalt:



Ein Tripel $x_{(1)}, x_{(2)}, x_{(3)}$ mit

$$x_{(1)} + x_{(3)} - 2\,x_{(2)} < 0$$

heißt **linkes Tripel**. Für ein linkes Tripel gilt also

$$x_{(3)} - x_{(2)} < x_{(2)} - x_{(1)}$$

Die folgende Abbildung veranschaulicht den Sachverhalt:

Um dieses Konzept auf eine Stichprobe $x_1, \ldots, x_n$ zu übertragen, betrachten Randles et al. (1980) für $1 \leq i < j < k \leq n$ alle geordneten Stichproben $x_{(i)}, x_{(j)}, x_{(k)}$ aus der Stichprobe $x_1, \ldots, x_n$ und bestimmen für jede dieser Stichproben den Wert der folgenden Funktion:

$$f^*(x_{(i)}, x_{(j)}, x_{(k)}) = \begin{cases} 1 & \text{für } x_{(i)} + x_{(k)} - 2\,x_{(j)} > 0 \\ -1 & \text{für } x_{(i)} + x_{(k)} - 2\,x_{(j)} < 0 \\ 0 & \text{sonst} \end{cases}$$

Beispiel 3
Studierende wurden in einer Vorlesung gefragt, wie viele CDs sie besitzen. Hier sind die Daten von 5 Studierenden:

```
30   40   60   100  150
```

Es gilt

$$\begin{array}{rr} f^*(x_{(1)}, x_{(2)}, x_{(3)}) = 1 & f^*(x_{(1)}, x_{(2)}, x_{(4)}) = 1 \\ f^*(x_{(1)}, x_{(2)}, x_{(5)}) = 1 & f^*(x_{(1)}, x_{(3)}, x_{(4)}) = 1 \\ f^*(x_{(1)}, x_{(3)}, x_{(5)}) = 1 & f^*(x_{(1)}, x_{(4)}, x_{(5)}) = -1 \\ f^*(x_{(2)}, x_{(3)}, x_{(4)}) = 1 & f^*(x_{(2)}, x_{(3)}, x_{(5)}) = 1 \\ f^*(x_{(2)}, x_{(4)}, x_{(5)}) = -1 & f^*(x_{(3)}, x_{(4)}, x_{(5)}) = 1 \end{array}$$

Die Teststatistik des Tests von Randles et al. (1980) ist

$$T = \sum_{1 \leq i < j < k \leq n} f^*(x_{(i)}, x_{(j)}, x_{(k)}) \tag{30}$$

Dies ist die Differenz aus der Anzahl der rechten Tripel und der Anzahl der linken Tripel in allen geordneten Stichproben vom Umfang 3 aus der Stichprobe $x_1, \ldots, x_n$.

Beispiel 3 (fortgesetzt)
Es gilt $T = 6$.

Ist die Anzahl der rechten Tripel und linken Tripel ungefähr gleich, so deutet dies auf Symmetrie hin. In diesem Fall nimmt T einen Wert in der Nähe von 0 an. Gibt es in der Stichprobe aber viel mehr rechte als linke Tripel, so spricht dies für eine rechtsschiefe Verteilung. In diesem Fall nimmt T einen großen Wert an. Ein kleiner Wert von T ist ein Indikator für eine linksschiefe Verteilung. Wir lehnen also H_0 ab, wenn T zu groß oder zu klein ist.

Zur Testentscheidung benötigen wir die Verteilung von T, wenn die Nullhypothese der Symmetrie der Verteilung zutrifft. Randles et al. (1980) zeigen, dass

$$\frac{T - E(T)}{\sqrt{Var(T)}}$$

approximativ standardnormalverteilt ist, wenn H_0 zutrifft. Wir lehnen H_0 zum Signifikanzniveau α ab, wenn gilt

$$\left| \frac{T - E(T)}{\sqrt{Var(T)}} \right| \geq z_{1-\alpha/2}$$

Dabei ist $z_{1-\alpha/2}$ das $1 - \alpha/2$-Quantil der Standardnormalverteilung.
Trifft H_0 zu, so gilt

$$E(T) = 0$$

und

$$\begin{aligned} Var(T) &= \frac{(n-3)(n-4)}{(n-1)(n-2)} \sum_{t=1}^{n} B_t^2 + \frac{n-3}{n-4} \sum_{s=1}^{n-1} \sum_{t=s+1}^{n} B_{s,t}^2 \\ &\quad + \frac{n(n-1)(n-2)}{6} - \left[1 - \frac{(n-3)(n-4)(n-5)}{n(n-1)(n-2)} \right] T^2 \end{aligned}$$

Der Beweis ist bei Randles et al. (1980) zu finden. Dabei ist B_t gleich der Differenz aus der Anzahl der rechten Tripel und der Anzahl der linken Tripel, in denen $x_{(t)}$ vorkommt, und $B_{s,t}$ gleich der Differenz aus der Anzahl der rechten Tripel und der Anzahl der linken Tripel, in denen $x_{(s)}$ und $x_{(t)}$ vorkommen.

Beispiel 3 (fortgesetzt)
Es gilt

$$B_1 = 4 \quad B_2 = 4 \quad B_3 = 6 \quad B_4 = 2 \quad B_5 = 2\,.$$

Außerdem gilt

$$\begin{array}{lllll} B_{1,2} = 3 & B_{1,3} = 3 & B_{1,4} = 1 & B_{1,5} = 1 & B_{2,3} = 3 \\ B_{2,4} = 1 & B_{2,5} = 1 & B_{3,4} = 3 & B_{3,5} = 3 & B_{4,5} = -1 \end{array}$$

Also gilt

$$\sum_{t=1}^{5} B_t^2 = 4^2 + 4^2 + 6^2 + 2^2 + 2^2 = 76$$

und

$$\sum_{s=1}^{4}\sum_{t=s+1}^{5} B_{s,t}^2 = 3^2+3^2+1^2+1^2+3^2+1^2+1^2+3^2+3^2+(-1)^2 = 50$$

Somit gilt

$$Var(T) \quad = \quad \frac{2\cdot 1}{4\cdot 3}\cdot 76 + \frac{2}{1}\cdot 50 + \frac{5\cdot 4\cdot 3}{6} - \left[1 - \frac{2\cdot 1\cdot 0}{4\cdot 3\cdot 2}\right] 6^2 = 86.67$$

Wir bestimme den Wert der Teststatistik

$$T = \frac{6}{\sqrt{86.67}} = 0.64$$

Wegen $z_{0.975} = 1.96$ lehnen wir H_0 zum Niveau 0.05 nicht ab.

Beispiel 4
Studierende wurden in einer Vorlesung gefragt, wie viele CDs sie besitzen. Hier sind die Daten von 10 Studierenden:

```
10 20 30 40 60 70 90 150 200 300
```

Wir testen auf Symmetrie. Es gilt $T = 58$ und $Var(T) = 786.83$. Somit gilt

$$T = \frac{58}{\sqrt{786.83}} = 2.07$$

Wegen $z_{0.975} = 1.96$ lehnen wir H_0 zum Niveau 0.05 ab.

5 Transformation auf Symmetrie

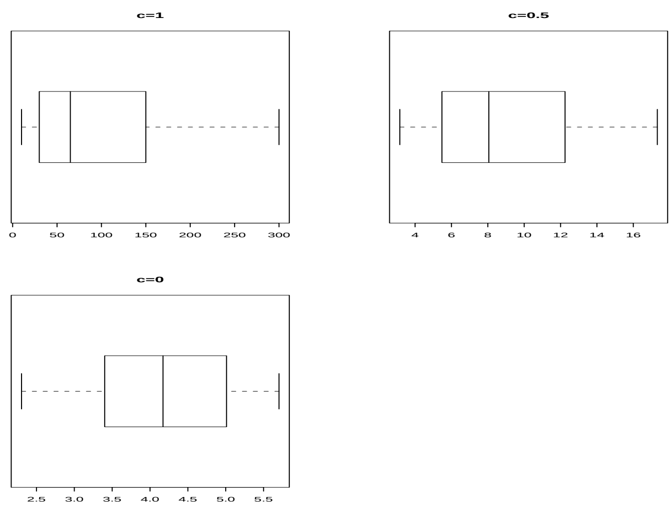
Ist die Verteilung der Grundgesamtheit symmetrisch, so können Parameter leichter interpretiert werden. Deshalb versucht man die Daten so zu transformieren, dass die Verteilung symmetrisch ist. Da lineare Transformationen die Schiefe einer Verteilung nicht verändern, betrachtet man nichtlineare Transformationen. Am einfachsten zu interpretieren ist die Power-Transformation:

$$g(x) = \begin{cases} x^c & \text{für } c > 0 \\ \ln x & \text{für } c = 0 \\ -x^c & \text{für } c < 0 \end{cases} \tag{31}$$

Das negative Vorzeichen für $c < 0$ ist notwendig, da sonst die Ordnung der Daten zerstört würde.

Beispiel 4 (fortgesetzt)
Abbildung 8 zeigt die Boxplots für $c = 1, 0.5, 0$.

Abbildung 8: Boxplots der Originaldaten und der transformierten Daten



Wir sehen, dass die Verteilung durch Logarithmieren der Daten nahezu symmetrisch ist. Aber auch Verteilung der Quadratwurzel der Daten sieht symmetrisch aus.

Emerson and Stoto (1982) zeigen, wie man den Transformationsparameter systematisch schätzen kann. Sie gehen von der bekannten Gleichung

$$x_{0.5} - x_p = x_{1-p} - x_{0.5}$$

aus, die für alle $p \in (0, 0.5)$ gilt, wenn die Verteilung symmetrisch ist. Man kann diese Gleichung umformen zu

$$\frac{x_p + x_{1-p}}{2} = x_{0.5}$$

Transformiert man die Beobachtungen $x_1, \ldots, x_n$ mit einem Wert von c , so muss bei Symmetrie für alle $p \in (0, 0.5)$ gelten:

$$\frac{x_p^c + x_{1-p}^c}{2} = x_{0.5}^c \tag{32}$$

bzw.

$$\frac{\ln x_p + \ln x_{1-p}}{2} = \ln x_{0.5} \tag{33}$$

Man kann c schätzen, indem man einen Wert für p vorgibt und den Wert von c wählt, für den die Gleichung (32) bzw. (33) erfüllt ist.

Beispiel 4 (fortgesetzt)
Setzt man zum Beispiel $p = 0.25$, so erhält man folgende Schätzer

$$x_{0.25} = 30 \qquad x_{0.5} = 65 \qquad x_{0.25} = 150$$

Wir suchen also den Wert von c, für die folgende Gleichung erfüllt

$$\frac{30^c + 150^c}{2} = 65^c$$

Da die Bestimmung der Nullstelle der Gleichung (32) nicht einfach ist, wählen Emerson and Stoto (1982) ein approximatives Verfahren. Sie entwickeln die Funktionen x_p^c und x_{1-p}^c in eine Taylorreihe um den Median $x_{0.5}$ und brechen sie nach dem quadratischen Glied ab. Sie bilden also

$$x_p^c \approx x_{0.5}^c + c\, x_{0.5}^{c-1}\,(x_p - x_{0.5}) + \frac{c\,(c-1)}{2}\, x_{0.5}^{c-2}\,(x_p - x_{0.5})^2 \tag{34}$$

und

$$x_{1-p}^c \approx x_{0.5}^c + c\, x_{0.5}^{c-1}\,(x_{1-p} - x_{0.5}) + \frac{c\,(c-1)}{2}\, x_{0.5}^{c-2}\,(x_{1-p} - x_{0.5})^2 \tag{35}$$

Setzen wir die Gleichungen (34) und (35) in die die Gleichung (32) ein, so erhalten wir

$$\begin{aligned} x_{0.5}^c &= x_{0.5}^c + \frac{c}{2}\, x_{0.5}^{c-1}\ (x_p + x_{1-p} - 2x_{0.5}) \\ &\quad + \frac{c\,(c-1)}{4}\, x_{0.5}^{c-2}\left[(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2\right] \end{aligned}$$

Subtrahieren wir $x_{0.5}^c$ von beiden seiten dieser Gleichung, so erhalten wir die äquivalente Gleichung

$$\frac{c}{2}\, x_{0.5}^{c-1}\ (x_p + x_{1-p} - 2x_{0.5}) + \frac{c\,(c-1)}{4}\, x_{0.5}^{c-2}\left[(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2\right] = 0$$

Wir klammern aus der Summe auf der linken Seite der Gleichung den Ausdruck $\frac{c}{2}\, x_{0.5}^{c-2}$ aus

$$\frac{c}{2}\, x_{0.5}^{c-2}\left(x_{0.5}\ (x_p + x_{1-p} - 2x_{0.5}) + \frac{c-1}{2}\ \left[(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2\right]\right) = 0$$

Dividieren wir beide Seiten dieser Gleichung durch $\frac{c}{2}\,x_{0.5}^{c-2}$, so erhalten wir

$$x_{0.5}\,(x_p + x_{1-p} - 2x_{0.5}) + \frac{c-1}{2}\,\left[(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2\right] = 0$$

Wir dividieren beide Seiten dieser Gleichung durch $2\,x_{0.5}$, so erhalten wir

$$\frac{x_p + x_{1-p}}{2} - x_{0.5} + (c-1)\,\frac{\left[(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2\right]}{4\,x_{0.5}} = 0$$

Nun lösen wir diese Gleichung nach $\frac{x_p+x_{1-p}}{2} - x_{0.5}$ auf:

$$\frac{x_p + x_{1-p}}{2} - x_{0.5} = (1-c)\,\frac{(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2}{4\,x_{0.5}} \tag{36}$$

Es gibt nun zwei Möglichkeiten, mit Gleichung (36) c zu bestimmen.
Man kann die Gleichung nach c auflösen:

$$c = 1 - \frac{2\,x_{0.5}\,(x_p + x_{1-p} - 2\,x_{0.5})}{(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2} \tag{37}$$

Dann gibt man einen Wert von p vor, schätzt die Quantile x_p, $x_{0.5}$ und x_{1-p} und setzt diese in die Gleichung (37) ein.

Beispiel 4 (fortgesetzt)
Wir setzen $p = 0.25$ und erhalten die Schätzer

$$x_{0.25} = 30 \qquad x_{0.5} = 65 \qquad x_{0.25} = 150$$

Setzen wir dies in Gleichung (37) ein, so erhalten wir

$$c = 1 - \frac{2 \cdot 65\,(30 + 150 - 2 \cdot 65)}{(30-65)^2 + (150-65)^2} = 0.23$$

Dies spricht dafür, dass man die Daten logarithmieren sollte.

Emerson and Stoto (1982) schätzen für $p = 1/2, 1/4, 1/8, 1/16, \ldots$ die Quantile x_p und x_{1-p}. Für diese geschätzten Werte kann man die Gleichung (36) als Regressionsgleichung in der erklärenden Variablen $\frac{(x_p-x_{0.5})^2+(x_{1-p}-x_{0.5})^2}{4\,x_{0.5}}$ und der zu erklärenden Variablen $\frac{x_p+x_{1-p}}{2} - x_{0.5}$ auffassen. Man kann durch die Punktewolke die Regressionsgerade legen und erhält so einen Schätzer für $1-c$ und auch für c. Hierbei ist zu beachten, dass die Regressionsgerade durch den Ursprung gilt, da für $p = 0.5$

$$\frac{(x_{0.5} - x_{0.5})^2 + (x_{0.5} - x_{0.5})^2}{4\,x_{0.5}} = 0$$

und

$$\frac{x_p + x_{1-p}}{2} - x_{0.5} = 0$$

Emerson and Stoto (1982) schlagen vor, den Steigungsparameter robust zu schätzen. Hierzu bestimmen sie die Steigungen aller Geraden durch den Nullpunkt und die Punkte

$$\left(\frac{(x_p - x_{0.5})^2 + (x_{1-p} - x_{0.5})^2}{4\, x_{0.5}}, \frac{x_p + x_{1-p}}{2} - x_{0.5}\right)$$

für $p = 1/2, 1/4, 1/8, 1/16, \ldots$ und den Punkt $\left(x_{(1)}, x_{(n)}\right)$. Der Schätzer des Steigungsparameters ist der Median der Steigungen.

Beispiel 4 (fortgesetzt)
Emerson and Stoto (1982) schlagen vor, x_p für $p = 1/2, 1/4, 1/8, 1/16, \ldots$ nach der Methode von Tukey zu schätzen. Somit ist $x_{0.25}$ der Median der unteren Hälfte des geordneten Datensatzes, also von

```
10 20 30 40 60
```

und $x_{0.75}$ der Median der oberen Hälfte des geordneten Datensatzes, also von

```
70 90 150 200 300
```

Als Schätzer für $x_{0.125}$ wählen wir den Median der unteren Hälfte der unteren Hälfte des geordneten Datensatzes, also von

```
10 20 30
```

Als Schätzer für $x_{0.875}$ wählen wir den Median der oberen Hälfte der oberen Hälfte des geordneten Datensatzes, also von

```
150 200 300
```

Der geordnete Datensatz wird so oft geteilt, bis nur noch eine Beobachtung übrig bleibt. Für das Beispiel erhalten wir die Schätzer

$$\begin{aligned} x_{0.25} &= 30 & x_{0.75} &= 150 \\ x_{0.125} &= 20 & x_{0.875} &= 200 \\ x_{0.0625} &= 15 & x_{0.9375} &= 250 \end{aligned}$$

In Tabelle 2 werden die relevanten Größen bestimmt, wobei zu $p = 0$ das Minimum bzw. Maximum der Daten gehört.

Tabelle 2: Hilfstabelle zur Berechnung der Schätzer

p	x_p	x_{1-p}	$\frac{x_p+x_{1-p}}{2}$	$\frac{x_p+x_{1-p}}{2} - x_{0.5}$	$\frac{(x_p-x_{0.5})^2+(x_{1-p}-x_{0.5})^2}{4\,x_{0.5}}$	$\hat{c}$
0.2500	30	150	90.0	25.0	32.50	0.23
0.1250	20	200	110.0	45.0	77.88	0.42
0.0625	15	250	132.5	67.5	141.25	0.52
	10	300	155.0	90.0	224.04	0.60

In der letzten Spalte der Tabelle stehen die Steigungen der Geraden. Der Schätzer $\hat{c}$ ist gleich dem Median der Werte in der letzten Spalte, also $\hat{c} = 0.47$. Man sollte also die Quadratwurzel der Daten bilden.

6 Wie man eine Funktion durch eine lineare bzw. quadratische Funktion approximiert

Wir gehen aus von einer Funktion $f : D \subset \mathbb{R} \to \mathbb{R}$, die in $x = a$ existiert. Außerdem mögen die erste Ableitung $f'(x)$ und zweite Ableitung $f''(x)$ in einer Umgebung von $x = a$ existieren.

Beispiel 5
$f : \mathbb{R} \to \mathbb{R}$ mit $x \mapsto y = f(x) = e^x$. Unser Ziel ist es, die Funktion f durch eine einfache Funktion zu approximieren.

Es liegt nahe, die Funktion f durch eine Gerade zu approximieren. Wir setzen an

$$g(x) = a_0 + a_1\,(x - a) \tag{38}$$

Wie sollen wir a_0 und a_1 wählen? Wir fordern, dass g mit f an der Stelle $x = a$ sehr gut übereinstimmt. Es sollte also gelten

$$g(a) = f(a) \tag{39}$$

Außerdem fordern wir noch, dass die erste Ableitung von g mit der ersten Ableitung von f an der Stelle $x = a$ übereinstimmt:

$$g'(a) = f'(a) \tag{40}$$

gilt. Unter diesen Bedingungen können wir die Konstanten a_0 und a_1 bestimmen. Es gilt

$$g(a) = a_0 + a_1\,(a - a) = a_0$$

Mit Gleichung (39) auf Seite 25 folgt also

$$a_0 = f(a)$$

Außerdem gilt

$$g'(a) = a_1$$

Mit Gleichung (40) auf Seite 25 folgt also

$$a_1 = f'(a)$$

Also ist die gesucht Lösung gleich

$$g(x) = f(a) + f'(a)\,(x - a) \tag{41}$$

Beispiel 5 (fortgesetzt)
Wir betrachten $f(x) = e^x$ und $a = 0$. Es gilt

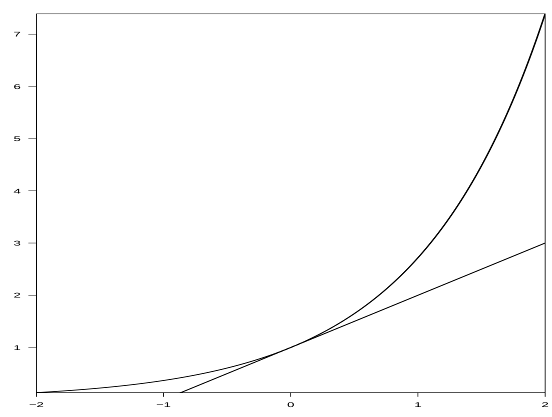
$$f'(x) = e^x$$

Aus $f(0) = 1$ und $f'(0) = 1$ folgt

$$g(x) = 1 + x \tag{42}$$

Abbildung 9 zeigt die Approximation.

Abbildung 9: Approximation von $f(x) = e^x$ durch eine Gerade



Wir können f aber auch durch eine quadratische Funktion approximieren. Gesucht ist eine Funktion

$$g(x) = a_0 + a_1\,(x-a) + a_2\,(x-a)^2 \tag{43}$$

In Analogie zur linearen Approximation fordern wir zusätzlich zur Gültigkeit der Gleichungen (39) und (40) auf Seite 25 noch die Gültigkeit von

$$g''(a) = f''(a) \tag{44}$$

Unter diesen Bedingungen können wir die Konstanten a_0, a_1 und a_2 bestimmen. Es gilt

$$g(a) = a_0 + a_1\,(a-a) + a_2\,(a-a)^2 = a_0$$

Mit Gleichung (39) auf Seite 25 folgt also

$$a_0 = f(a)$$

Außerdem gilt

$$g'(x) = a_1 + 2\,a_2\,(x-a)$$

Somit gilt

$$g'(a) = a_1$$

Mit Gleichung (40) auf Seite 25 folgt also

$$a_1 = f'(a)$$

Außerdem gilt

$$g''(x) = 2\,a_2$$

Also gilt speziell

$$g''(a) = 2\,a_2$$

Mit Gleichung (44) auf Seite 27 folgt also

$$a_2 = \frac{1}{2}\,f''(a)$$

Somit gilt

$$g(x) = f(a) + f'(a)\,(x-a) + \frac{1}{2}\,f''(a)\,(x-a)^2 \tag{45}$$

Beispiel 5 (fortgesetzt)
Es gilt

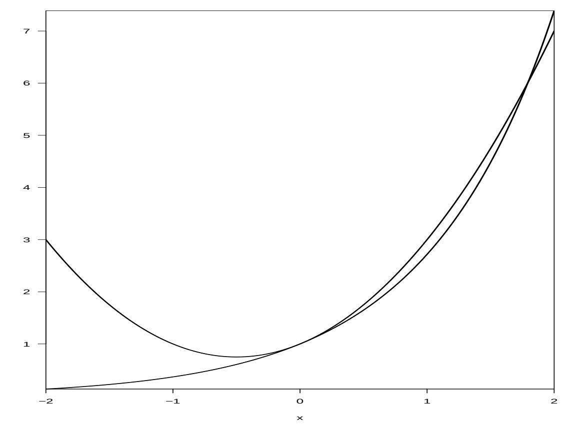
$$f''(x) = e^x$$

Mit $f''(0) = 1$ folgt

$$g(x) = 1 + x + x^2$$

Abbildung 10 zeigt die Approximation.

Abbildung 10: Approximation von $f(x) = e^x$ durch eine quadratische Funktion



7 Eine Anwendung der Approximation einer Funktion durch eine quadratische Funktion

Wir gehen im Folgenden von einer Zufallsstichprobe $x_1, \ldots, x_n$ aus einer Grundgesamtheit mit Verteilungsfunktion F aus. Somit sind die Beobachtungen $x_1, \ldots, x_n$ also Realisationen der unabhängigen und identisch mit Verteilungsfunktion F verteilten Zufallsvariablen $X_1, \ldots, X_n$. Die Verteilungsfunktion hänge von einem Parameter θ ab. Diesen wollen wir schätzen. Beim **Maximum-Likelihood-Verfahren** stellt man die gemeinsame Dichtefunktion $f(x_1, \ldots, x_n)$ bzw. Wahrscheinlichkeitsfunktion $P(X_1 = x_1, \ldots, X_n = x_n)$ auf und fasst diese als Funktion des Parameters θ gegeben die Daten $x_1, \ldots, x_n$ auf. Man bezeichnet sie als Likelihoodfunktion $L(\theta)$. Der **Maximum-Likelihood-Schätzer** ist nun der Wert von θ, für den die Likelihoodfunktion ihr Maximum annimmt. Ist die Likelihoodfunktion differenzierbar, so können wir die Verfahren der klassischen Analysis anwenden, um das Maximum zu bestimmen. Da die Likelihoodfunktion gleich dem Produkt der Randdichten bzw. Randwahrscheinlichkeitsfunktionen ist, erleichtert Logarithmieren die Bestimmung des Maximums beträchtlich. Man erhält also die sogenannte **Loglikelihoodfunktion**:

$$l(\theta) = \ln L(\theta)$$

Da der Logarithmus eine monotone Transformation ist, nimmt die Loglikelihoodfunktion ihr Maximum an der gleichen Stelle an wie die Likelihoodfunktion. Manchmal wird das Maximum schnell gefunden.

Beispiel 6
Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und identisch mit Parameter λ poissonverteilt. Somit gilt

$$P(X_i = x_i) = \frac{\lambda^{x_i}}{x_i!}\, e^{-\lambda}$$

Die Likelihoodfunktion ist somit

$$L(\lambda) = \frac{\lambda^{x_1}}{x_1!}\, e^{-\lambda} \cdots \frac{\lambda^{x_n}}{x_n!}\, e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\, e^{-n\,\lambda} = \frac{\lambda^{n\,\bar{x}}}{\prod_{i=1}^n x_i!}\, e^{-n\,\lambda}$$

Die Loglikelihoodfunktion ist

$$l(\lambda) = n\,\bar{x}\,\ln\lambda - \sum_{i=1}^n \ln x_i! - n\,\lambda$$

Eine notwendige Bedingung für einen Extremwert ist, dass die erste Ableitung $l'(\theta)$ gleich 0 ist.
Es gilt

$$l'(\lambda) = \frac{n\,\bar{x}}{\lambda} - n$$

Offensichtlich erfüllt $\hat{\lambda} = \bar{x}$ die Gleichung

$$l'(\hat{\lambda}) = 0$$

Weiterhin gilt

$$l''(\lambda) = -\frac{n\,\bar{x}}{\lambda^2}$$

Es gilt

$$l''(\bar{x}) = -\frac{n}{\bar{x}}$$

Sind nicht alls x_i gleich 0, so handelt es sich um ein Maximum.

Wie das folgende Beispiel zeigt, kann man das Maximum nicht immer so leicht bestimmen.

Beispiel 7
In einer Vorlesung wurden die Studierenden gefragt, wie viele Kinder ihre Eltern haben. Da die befragte Person ein Kind seiner Eltern ist, war die bei der Antwort gegebene Zahl mindestens 1. Tabelle 3 zeigt die Häufigkeitsverteilung.

Tabelle 3: Häufigkeitstabelle des Merkmals Anzahl Kinder

Anzahl Kinder	absolute Häufigkeit	relative Häufigkeit
1	15	0.170
2	48	0.545
3	17	0.193
4	7	0.080
5	1	0.011

Hier kann man die Poissonverteilung nicht als Modell verwenden, da der Wert 0 nicht annehmen wird. Bei der Poissonverteilung ist diese Wahrscheinlichkit aber $e^{-\lambda}$ und somit größer als 0. Ein Modell für eine Zählvariable, die bei 1 zu zählen beginnt, ist die abgeschnittene Poissonverteilung. Deren Wahrscheinlichkeitsfunktion ist für $x = 1, 2, \ldots$ gegeben durch

$$P(X = x) = \frac{\lambda^x}{x!} \frac{e^{-\lambda}}{1 - e^{-\lambda}}$$

Um den Maximum-Likelihood-Schätzer aus einer Stichprobe $x_1, \ldots, x_n$ vom Umfang n zu bestimmen, stellen wir die Likelihood auf:

$$L(\lambda) = \frac{\lambda^{n\,\bar{x}}}{\prod_{i=1}^{n} x_i!} \frac{e^{-n\,\lambda}}{(1 - e^{-\lambda})^n}$$

Die Loglikelihoodfunktion ist

$$l(\lambda) = n\,\bar{x}\,\ln\lambda - \sum_{i=1}^{n} \ln x_i! - n\,\lambda - n\,\ln\left(1 - e^{-\lambda}\right) \tag{46}$$

Eine notwendige Bedingung für einen Extremwert ist, dass die erste Ableitung $l'(\theta)$ gleich 0 ist.

Es gilt

$$l'(\lambda) = \frac{n\,\bar{x}}{\lambda} - n - \frac{n\,e^{-\lambda}}{(1-e^{-\lambda})} \tag{47}$$

Die Gleichung

$$\frac{n\,\bar{x}}{\hat{\lambda}} - n - \frac{n\,e^{-\hat{\lambda}}}{\left(1-e^{-\hat{\lambda}}\right)}$$

kann man nicht explizit nach $\hat{\lambda}$ auflösen. Hat man aber den Schätzwert gefunden, so kann man schnell nachprüfen, ob es sich um ein Maximum handelt, da die zweite Ableitung der Loglikelihoodfunktion lautet:

$$l''(\lambda) = -\frac{n\,\bar{x}}{\lambda^2} + \frac{n\,e^{-\lambda}}{(1-e^{-\lambda})^2} \tag{48}$$

Kann man das Maximum der Loglikelihoodfunktion nicht explizit bestimmen, so sollte man ein iteratives Verfahren verwenden. Wir schauen uns das **Newton-Raphson-Verfahren** an, das von Everitt (1987) sehr anschaulich beschrieben wird.

Beim Newton-Raphson-Verfahren gibt man einen Startwert θ_0 für θ vor und approximiert die Loglikelihood $l(\theta)$ durch die quadratische Funktion $g(\theta)$, für die gilt

$$\begin{aligned} g(\theta_0) &= l(\theta_0) \\ g'(\theta_0) &= l'(\theta_0) \\ g''(\theta_0) &= l''(\theta_0) \end{aligned}$$

Wie wir in Kapitel 6 auf Seite 27 gesehen haben, ist die Funktion $g(x)$ gegeben durch

$$g(\theta) = l(\theta_0) + l'(\theta_0)\,(\theta-\theta_0) + \frac{1}{2}\,l''(\theta_0)\,(\theta-\theta_0)^2 \tag{49}$$

Das Maximum der Funktion $g(x)$ in Gleichung (49) ist leicht zu bestimmen. Wir bilden die erste Ableitung

$$g'(\theta) = l'(\theta_0) + l''(\theta_0)\,(\theta-\theta_0)$$

setzen sie gleich 0

$$l'(\theta_0) + l''(\theta_0)\,(\theta-\theta_0) = 0 \tag{50}$$

und lösen die Gleichung nach θ auf

$$\theta = \theta_0 - \frac{l'(\theta_0)}{l''(\theta_0)} \tag{51}$$

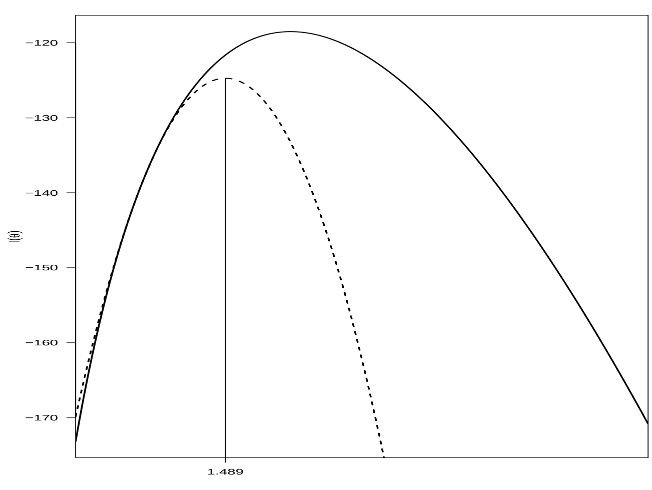
Hierdurch erhalten wir einen neuen Wert für das Maximum, der oft näher am Maximum der Loglikelihoodfunktion liegt als der Startwert θ_0.

Beispiel 7 (fortgesetzt)
Es gilt $n = 88$ und $\bar{x} = 2.22$. Wir wählen $\lambda_0 = 1$. Es gilt $l(1) = -138.04$, $l'(1) = 55.79$ und $l''(1) = -113.98$. Somit approximieren wir $l(\theta)$ durch

$$g(x) = -138.04 + 55.79\,(\lambda - 1) - \frac{1}{2}\,113.98\,(\lambda - 1)^2 \tag{52}$$

Abbildung 11 zeigt die Likelihoodfunktion zusammen mit der approximierenden quadratischen Funktion und dem Maximum der approximierenden Funktion.

Abbildung 11: Likelihoodfunktion zusammen mit der approximierenden quadratischen Funktion



Als neuen Schätzwert für das Maximum erhalten wir

$$\lambda = 1 - \frac{55.79}{-113.98} = 1.489$$

Wir sehen in Abbildung 11, dass das Maximum der approximierenden quadratischen Funktion näher am Maximum der Loglikelihood liegt als der Startwert. Wir approximieren nun die Loglikelihood durch die quadratische Funktion, die am neuen Maximum mit der Loglikelihood übereinstimmt und bestimmen das Maximum dieser Funktion. Dieses Verfahren iterieren wir so lange, bis sich der Wert nicht mehr ändert. Diese Vorgehensweise nennt man das Newton-Raphson-Verfahren. Hier ist der Algorithmus:

1. Wähle einen Startwert θ_0, eine vorgegeben Genauigkeit ϵ und setze i auf den Wert 0.

2. Berechne $\theta_{i+1} = \theta_i - l'(\theta_i)/l''(\theta_i)$.

3. Gehe zu Schritt 4, wenn $|\theta_i - \theta_{i+1}| < \epsilon$ gilt. Gilt aber $|\theta_i - \theta_{i+1}| \geq \epsilon$, so erhöhe i um 1 und gehe zu Schritt 2.

4. Wähle θ_i als Wert des Maximums.

Man kann im Schritt 2. des Algorithmus auch die Änderung der Loglikelihoodfunktion als Kriterium wählen. In diesem Fall bestimmt man also $|l(\theta_i) - l(\theta_{i+1})|$

Beispiel 7 (fortgesetzt)
Wir wählen $\epsilon = 0.01$.
Mit $\lambda_1 = 1.489$, $l'(1.489) = 17.32$ und $l''(1.489) = -54.85$ erhalten wir

$$\lambda_2 = 1 - \frac{17.32}{-54.85} = 1.805$$

Wegen $\lambda_2 - \lambda_1| = 0.316$ sind wir noch nicht am Ende angelangt.
Mit $\lambda_2 = 1.805$, $l'(1.805) = 2.71$ und $l''(1.805) = -39.12$ erhalten wir

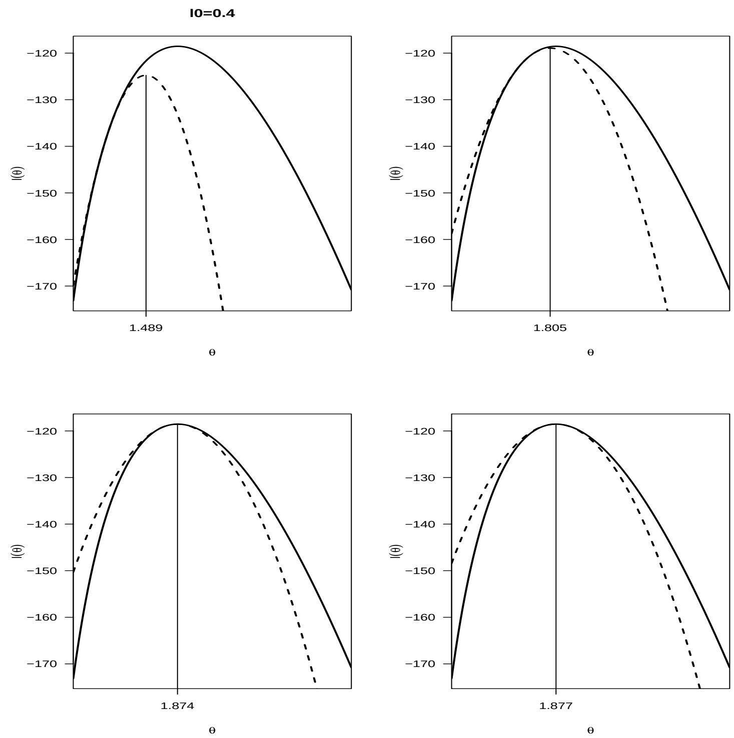
$$\lambda_3 = 1.805 - \frac{2.71}{-39.12} = 1.874$$

Wegen $|\lambda_3 - \lambda_2| = 0.069$ sind wir noch nicht am Ende angelangt.
Mit $\lambda_3 = 1.874$, $l'(1.874) = 2.71$ und $l''(1.874) = -39.12$ erhalten wir

$$\lambda_4 = 1.874 - \frac{0.097}{-36.67} = 1.877$$

Wegen $|\lambda_4 - \lambda_3| = 0.003$ sind wir am Ende angelangt. Als M-L-Schätzer erhalten wir $\hat{\lambda} = 1.877$.
Abbildung 12 zeigt die Iteration.

Abbildung 12: Likelihoodfunktion zusammen mit der approximierenden quadratischen Funktion



Literatur

Bowley, A. L. (1920). *Elements of statistics*. Charles Scribner's sons, New York.

Brys, G., Hubert, M., and Struyf, A. (2003). A comparison of some new measures of skewness. In Dutter, R., Filzmoser, P., Gather, U., and Rousseeuw, P., editors, *Developments in Robust Statistics (ICORS 2001)*, pages 98–113, Heidelberg. Physika.

Emerson, J. D. and Stoto, M. A. (1982). Exploratory methods for choosing power transformation. *Journal of the American Statistical Association*, 77:103–108.

Everitt, B. S. (1987). *Introduction to optimization methods and their application in statistics*. Chapman and Hall.

Handl, A. (1985). Maßzahlen zur Klassifizierung von Verteilungen bei der Konstruktion adaptiver Tests im unverbundenen Zweistichprobenproblem. Unveröffentlichte Dissertation, Freie Universität Berlin.

Hinley, D. V. (1975). Pm power transformations to symmetry. *Bimoetrika*, 62:101–111.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52:105–124.

Randles, R. H., Fligner, M. A., Policello, G. E., and Wolfe, D. A. (1980). An asymptotically distribution-free test for symmetry vs. asymmertry. *Journal of the American Statistical Association*, 75:168–172.

Royston, P. (1992). Which measures of skewness and kurtosis are best? *Statistics in Medicine*, 11:333–343.