§5 Schätztheorie pfad

▼

```
SetDirectory[$UserDocumentsDirectory] ;

<< StatisticalPlots`;

ScoreTest[stichprobe_, verteilungsgesetz_, options___] := Module[{z, n, p},
    z = Sort[stichprobe];
    n = Length[stichprobe];
    p = Table[{i, CDF[verteilungsgesetz, z[[i]]]}, {i, 1, n}]; ListPlot[p, options]]
```

Ein wichtiges Teilgebiet der Statistik ist die Schätztheorie. Wir werden in diesem Kapitel zunächst die Aufgabe der Schätztheorie formulieren und uns mit Gütekriterien für Schätzer befassen, Anschließend werden wir die wichtigsten Verfahren zur Konstruktion von Schätzern - nämlich die Momenten-Methode und die Maximum-Likelihood-Methode - besprechen.

5.1 Die Aufgabe der Schätztheorie

Wir beginnen mit einer für die Statistik zentralen Definition:

5.1.1 Definition: Ist von der Verteilung $\mathbb{P}_X$ der Zufallsvariablen X nur bekannt, dass sie Element einer Familie $\mathcal{P}_X$ von Wahrscheinlichkeitsmaßen auf $\mathbb{R}$ ist, so nennt man $\mathcal{P}_X$ ein a-priori Modell für die Verteilung von X. Lässt sich diese Familie $\mathcal{P}_X$ durch eine Menge $\Theta \subseteq \mathbb{R}^k$ parametrisieren, gilt also

$$\mathcal{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\} \in \Theta\}$$

so nennt man $\mathcal{P}_X$ ein k-parametrisches a-priori Modell für die Verteilung von X.

Beispiele für a-priori Modelle sind
- die Menge $\mathcal{P}_X$ aller diskreten Verteilungen auf $\mathbb{N}$;
- die Menge $\mathcal{P}_X$ aller stetigen Verteilungen auf $\mathbb{R}$ mit dem Träger $\mathbb{T}$.

Beispiele für ein-parametrische a-priori Modelle sind
- die Menge $\mathcal{P}_X = \{\mathcal{P}[\lambda] \mid \lambda \in]0, \infty[\}$ der Poissonverteilungen;
- die Menge $\mathcal{P}_X = \{\mathcal{E}[\lambda] \mid \lambda \in]0, \infty[\}$ der Exponentialverteilungen;
- die Menge $\mathcal{P}_X = \{\mathcal{B}[n_0, p] \mid p \in]0, 1[\}$ der Binomialverteilungen mit bekanntem Parameter n_0;
- die Menge $\mathcal{P}_X = \{\mathcal{N}[\mu, \sigma_0] \mid \mu \in \mathbb{R}\}$ der Normalverteilungen mit bekannter Streuung σ_0.

Beispiele für zwei-parametrische a-priori Modelle sind
- die Menge $\mathcal{P}_X = \{\mathcal{N}[\mu, \sigma] \mid \{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[\}$ der Normalverteilungen;
- die Menge $\mathcal{P}_X = \{\mathcal{E}xtrem[\mu, \beta] \mid \{\mu, \beta\} \in \mathbb{R} \times]0, \infty[\}$ der Extremwertverteilungen;
- die Menge $\mathcal{P}_X = \{\mathcal{G}amma[\alpha, \beta] \mid \{\alpha, \beta\} \in]0, \infty[\times]0, \infty[\}$ der Gammaverteilungen.

Der Begriff eines a-priori Modells bzw eines k-parametrischen a-priori Modells für die gemeinsame Verteilung der Zufallsvariablen $X, Y, \ldots$ wird analog definiert. Beispielsweise handelt es sich bei der Menge

$$\mathcal{P}_{X,Y} = \left\{\mathcal{MN}\left[\vec{\mu}, \Sigma\right] \mid \vec{\mu} = \{\mu_1, \mu_2\} \in \mathbb{R}^2, \Sigma = \begin{pmatrix} {\sigma_1}^2 & \varrho\,\sigma_1\,\sigma_2 \\ \varrho\,\sigma_1\,\sigma_2 & {\sigma_2}^2 \end{pmatrix} \in \mathbb{R}^2_2 \text{ mit } \sigma_1, \sigma_2 > 0 \text{ und } -1 < \varrho < 1\right\}$$

der zweidimensionalen Normalverteilungen um ein fünf-parametrisches a-priori Modell für die gemeinsame Verteilung der beiden Zufallsvariablen X und Y. Um unsere Ausführungen nicht unnötig zu komplizieren, beschränken wir uns bei theoretischen Untersuchungen stets auf den Fall einer einzigen Zufallsvariablen X.

Nach dieser vorbereitenden Begriffsbildung sind wir in der Lage, die Aufgabe der Schätztheorie zu formulieren:

5.1.2 Die Aufgabe der Schätztheorie: Sei $\mathcal{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\} \in \Theta\}$ ein k-parametrisches a-priori Modell für die Verteilung der Zufallsvariablen X. Von der Verteilung $\mathbb{P}_X$ dieser Zufallsvariablen ist also nur bekannt, dass sie ein Element der Menge $\mathcal{P}_X$ ist; es ist aber nicht bekannt, für welches spezielle $\vec{\vartheta} \in \Theta$ tatsächlich $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ gilt). Die Aufgabe der Schätztheorie besteht darin, aufgrund einer nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ ausgewählten konkreten Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ vom Umfang n den unbekannten Parameter $\theta = \theta[\vec{\vartheta}]$ möglichst gut zu schätzen. Dabei ist noch offen ist, was wir unter "möglichst gut" verstehen.

▼

Etwas respektlos lässt sich die Aufgabe der Schätztheorie folgendermaßen beschreiben: Der "Liebe Gott" kennt in seiner Allwissendheit natürlich den Parameter $\vec{\vartheta}$ mit dem experimentiert wurde und somit auch den uns nicht bekannten Wert $\theta = \theta[\vec{\vartheta}]$. Wir kennen nur eine nach dem Verteilungsgesetz $\mathbb{P}[\vec{\vartheta}]$ ausgewählte konkrete Stichprobe $\vec{x}$ vom Umfang n und versuchen, damit den Wert $\theta = \theta[\vec{\vartheta}]$ zu schätzen um so dem "Lieben Gott" auf die Schliche zu kommen.

Wir veranschaulichen die Aufgabe der Schätztheorie an zwei konkreten Beispielen, wobei sich diese beiden Beispiele dadurch auszeichnen, dass die zu schätzenden Parameter inhaltlich interpretierbar sind und damit ein vernünftiger Schätzwert offensichtlich ist:

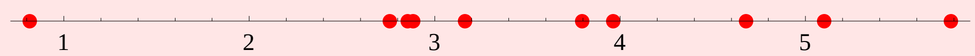
5.1.3 Beispiel: Man erzeuge durch Simulation eine Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ von $n = 10$ mit den Parametern $\mu = 3$ und $\sigma = 2$ normalverteilten Zufallszahlen und schätze die - nun als unbekannt anzusehenden - Parameter μ und σ.

▼

Lösung: Wir erzeugen in der üblichen Weise n nach dem Verteilungsgesetz $\mathcal{N}[\mu, \sigma]$ ausgewählte Zufallszahlen und veranschaulichen diese Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ mit Hilfe von ListPlot graphisch

```
n = 10; μ = 3; σ = 2;
x = RandomReal[NormalDistribution[μ, σ], n]
ListPlot[Thread[{x, 0}], PlotStyle → {PointSize[0.015], Red}, Axes → {True, None}, AspectRatio → 0.05]
```

```
{5.7843, 2.88342, 3.795, 4.67907, 2.75744, 5.10074, 0.81585, 2.85392, 3.16362, 3.96266}
```



Die Parameter μ und σ der Normalverteilung $\mathcal{N}[\mu, \sigma]$ besitzen eine inhaltliche Interpretation:

- Der Parameter μ entspricht dem Erwartungswert einer $\mathcal{N}[\mu, \sigma]$-verteilten Zufallsvariablen. Daher werden der Mittelwert $\bar{x}$ und aus Symmetriegründen auch der Median $\tilde{x}$ der Stichprobe $\vec{x}$ vernünftige Schätzwerte für den unbekannten Parameter μ sein.

- Der Parameter σ entspricht der Standardabweichung einer $\mathcal{N}[\mu, \sigma]$-verteilten Zufallsvariablen. Daher werden die Standardabweichung s_X sowie die Wurzel des zweiten zentralen Moments ${}^{(2)}\tilde{\bar{x}}$ der Stichprobe $\vec{x}$ vernünftige Schätzwerte für den unbekannten Parameter σ sein.

Wir berechnen diese Schätzwerte für die beiden Parameter μ und σ mit den Befehlen Mean bzw Median sowie StandardDeviation bzw CentralMoment

```
{Mean[x], Median[x]}
{StandardDeviation[x], Sqrt[CentralMoment[x, 2]]}
Clear[n, μ, σ, x]
```

```
{3.5796, 3.47931}

{1.41716, 1.34444}
```

und erkennen, dass sich mit einem derart kleinen Stichprobenumfang von $n = 10$ die unbekannten Parameter μ und σ nur grob schätzen lassen. Außerdem ist nicht klar, welchen der jeweiligen Schätzwerte man tatsächlich verwenden sollte.

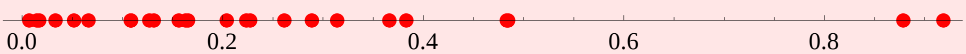
5.1.4 Beispiel: Man erzeuge durch Simulation eine Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ von $n = 25$ mit dem Parameter $\lambda = 3$ exponentialverteilten Zufallszahlen und schätze den (nun als unbekannt anzusehenden) Parameter λ.

▼

Lösung: Wir erzeugen in der üblichen Weise n nach dem Verteilungsgesetz $\mathcal{E}[\lambda]$ ausgewählte Zufallszahlen und veranschaulichen diese Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ wieder mit Hilfe von ListPlot graphisch

```
n = 25; λ = 3;
x = RandomReal[ExponentialDistribution[λ], n]
ListPlot[Thread[{x, 0}], PlotStyle → {PointSize[0.015], Red}, Axes → {True, None}, AspectRatio → 0.05]
```

```
{0.0171209, 0.484691, 0.108198, 0.0070829, 0.382997, 0.223553,
 0.918667, 0.165558, 0.227474, 0.163019, 0.878565, 0.126869,
 0.0332605, 0.131324, 0.0662562, 0.36625, 0.483164, 0.051625,
 0.288863, 0.314056, 0.108641, 0.261387, 0.156192, 0.0147141, 0.203975}
```



Der Parameter λ der Exponentialverteilung $\mathcal{E}[\lambda]$ besitzt eine inhaltliche Interpretation: $1/\lambda$ entspricht dem Erwartungswert einer $\mathcal{E}[\lambda]$-verteilten Zufallsvariablen; also wird der reziproke Wert des Mittelwerts $\bar{x}$ der Stichprobe $\vec{x}$ ein vernünftiger Schätzwert für den unbekannten Parameter λ sein. Wir berechnen diesen Schätzwert für den Parameter λ mit dem Befehl Mean

```
1/Mean[x]
Clear[n, λ, x]
```

```
4.04302
```

und erkennen, dass auch bei dem bereits relativ großen Stichprobenumfang von $n = 25$ unser Schätzwert noch immer ziemlich weit vom "richtigen" Wert λ entfernt ist.

An diesen Beispielen erkennen wir: Ein Schätzwert s für den Parameter $\theta = \theta[\vec{\vartheta}]$ hängt von der jeweiligen konkreten Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ ab, ist damit eine Funktion $s[\vec{x}]$ dieser Stichprobe und daher zufällig. Will man den zufälligen Charakter des Schätzwerts $s[\vec{x}]$ hervorheben, so verwendet man die Statistik $S = s[\vec{X}]$.

Wir definieren in diesem Zusammenhang:

5.1.5 Definition: Ist $\mathcal{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\} \in \Theta\}$ ein k-parametrisches a-priori Modell für die Verteilung der Zufallsvariablen X und ist $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$

ausgewählte mathematische Stichprobe vom Umfang n, so nennt man jede Statistik $S = s[\vec{X}]$ einen Schätzer für den unbekannten Parameter $\theta = \theta[\vec{\vartheta}]$.

▼

Man achte auf den Unterschied zwischen Schätzwert $s[\vec{x}]$ und Schätzer $s[\vec{X}_n]$. In Beispiel 5.1.3 ist beispielsweise die (von Stichprobe zu Stichprobe verschiedene) konkrete Zahl $\overline{x}$ ein Schätzwert und die Statistik $\overline{X}$ ein Schätzer für den Parameter μ der Normalverteilung $\mathcal{N}[\mu, \sigma]$.

5.2 Die Güte von Schätzern

Sei $\mathbb{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\} \in \Theta\}$ ein k-parametrisches a-priori Modell für die Verteilung der Zufallsvariablen X und $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ ausgewählte mathematische Stichprobe vom Umfang n. Im Prinzip könnte man natürlich jede beliebige Statistik $S = s[\vec{X}]$ als Schätzer für den unbekannten Parameter $\theta = \theta[\vec{\vartheta}]$ verwenden. Es ist aber klar, dass sich einige Statistiken dafür besser eignen als andere.

Wir werden uns daher zunächst mit der Frage befassen, wie sich die "Güte von Schätzern" beschreiben lässt:

5.2.1 Definition:

a) Unter dem mittleren quadratischen Fehler (mean square error) des Schätzers S für den Parameter $\theta = \theta[\vec{\vartheta}]$ versteht man den Wert

$$\mathbb{E}[(S - \theta[\vec{\vartheta}])^2; \mathbb{P}[\vec{\vartheta}]]$$

b) Der Schätzer S für den Parameter $\theta = \theta[\vec{\vartheta}]$ heißt erwartungstreu (unbiased), wenn der Erwartungswert des Schätzers stets mit dem zu schätzenden Parameter übereinstimmt, wenn also für alle $\vec{\vartheta} \in \Theta$

$$\mathbb{E}[S; \mathbb{P}[\vec{\vartheta}]] = \theta[\vec{\vartheta}]$$

gilt. Man beachte, dass bei erwartungstreuen Schätzern der mittlere quadratische Fehler $\mathbb{E}[(S - \theta[\vec{\vartheta}])^2; \mathbb{P}[\vec{\vartheta}]]$ und die Varianz $\mathbb{V}[S; \mathbb{P}[\vec{\vartheta}]]$ stets übereinstimmen.

c) Sind S_1 und S_2 zwei erwartungstreue Schätzer für den Parameter $\theta = \theta[\vec{\vartheta}]$, so heißt S_1 wirksamer (more effective) als der Schätzer S_2, wenn die Varianz des Schätzer S_1 stets höchstens so groß ist, wie die Varianz des Schätzers S_2, wenn also für alle $\vec{\vartheta} \in \Theta$ die folgende Beziehung gilt:

$$\mathbb{V}[S_1; \mathbb{P}[\vec{\vartheta}]] \le \mathbb{V}[S_2; \mathbb{P}[\vec{\vartheta}]]$$

d) Eine Folge $(S^{(n)})_{n\in\mathbb{N}}$ von Schätzern für den Parameter $\theta = \theta[\vec{\vartheta}]$ heißt konsistent, wenn die Folge $(S^{(n)})_{n\in\mathbb{N}}$ mit zunehmendem Stichprobenumfang n stets stochastisch gegen den zu schätzenden Parameter $\theta[\vec{\vartheta}]$ konvergiert, wenn also für alle $\vec{\vartheta} \in \Theta$ und alle $\varepsilon > 0$ die folgende Beziehung gilt:

$$\lim_{n\to\infty} \mathbb{P}[\{\, | S^{(n)} - \theta[\vec{\vartheta}] \,| \le \varepsilon\}; \mathbb{P}[\vec{\vartheta}]] = 1$$

Wir erläutern diese Begriffe an einigen Beispielen:

5.2.2 Beispiel: Für das ein-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{P}[\lambda] \mid \lambda \in]\, 0, \infty[\}$ für die Verteilung der Zufallsvariablen X untersuche man die beiden Schätzer $\overline{X}$ und V_X für den Parameter λ.

▼

Lösung: a) Beide Schätzer $\overline{X}$ und V_X sind erwartungstreu: Für alle $\lambda \in]\, 0, \infty[$ gilt wegen Satz 4.2.2 und unseren Ausführungen über die Poissonverteilung

$$\mathbb{E}[\overline{X}; \mathcal{P}[\lambda]] = \mathbb{E}[X] = \lambda \qquad \text{und} \qquad \mathbb{E}[V_X; \mathcal{P}[\lambda]] = \mathbb{V}[X] = \lambda$$

b) Der Schätzer $\overline{X}$ ist wirksamer als der Schätzer V_X: Für alle $\lambda \in]0, \infty[$ gilt wegen Satz 4.2.2 und unseren Ausführungen über die Poissonverteilung

$$\mathbb{V}[\overline{X}; \mathcal{P}[\lambda]] = \frac{1}{n}\, \mathbb{V}[X] = \frac{\lambda}{n}$$

$$\mathbb{V}[V_X; \mathcal{P}[\lambda]] = \frac{1}{n}\, \mathbb{E}[(X - \mathbb{E}[X])^4] - \frac{n-3}{n\,(n-1)}\, \mathbb{V}[X]^2 = \frac{\lambda}{n} + \frac{2\,\lambda^2}{n-1}$$

Dabei haben wir verwendet, dass $\mathbb{E}[(X - \mathbb{E}[X])^4] = \lambda + 3\,\lambda^2$ ist, was sich mit *Mathematica* leicht zeigen lässt:

```
ExpectedValue[(x - λ)^4, PoissonDistribution[λ], x]
```

$\lambda + 3\,\lambda^2$

c) Beide Folgen $(\overline{X}^{(n)})_{n\in\mathbb{N}}$ und $(V_X^{(n)})_{n\in\mathbb{N}}$ sind konsistent: Für alle $\lambda \in]0, \infty[$ und alle $\varepsilon > 0$ folgt nämlich aus der Ungleichung von Tschebyscheff zusammen mit den eben hergeleiteten Formeln

$$\lim_{n\to\infty} \mathbb{P}[\{\,|\,\overline{X}^{(n)} - \lambda\,| \geq \varepsilon\}; \mathcal{P}[\lambda]] \leq \lim_{n\to\infty} \frac{\mathbb{V}[\overline{X}^{(n)}; \mathcal{P}[\lambda]]}{\varepsilon^2} = \lim_{n\to\infty} \frac{\lambda}{n\,\varepsilon^2} = 0$$

$$\lim_{n\to\infty} \mathbb{P}[\{\,|\,V_X^{(n)} - \lambda\,| \geq \varepsilon\}; \mathcal{P}[\lambda]] \leq \lim_{n\to\infty} \frac{\mathbb{V}[V_X^{(n)}; \mathcal{P}[\lambda]]}{\varepsilon^2} = \lim_{n\to\infty} \left(\frac{\lambda}{n\,\varepsilon^2} + \frac{2\,\lambda^2}{(n-1)\,\varepsilon^2}\right) = 0$$

5.2.3 Beispiel: Für das ein-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{N}[\mu_0, \sigma] \mid \sigma \in]0, \infty[\}$ für die Verteilung der Zufallsvariablen X untersuche man die beiden Schätzer V_X und

$$V_{X,\mu_0} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2$$

für den Parameter σ^2.

▼

Lösung: a) Beide Schätzer V_X und V_{X,μ_0} sind erwartungstreu: Für alle $\sigma \in]0, \infty[$ gilt wegen Satz 4.2.2 und unseren Ausführungen über die Normalverteilung

$$\mathbb{E}[V_X; \mathcal{N}[\mu_0, \sigma]] = \mathbb{V}[X] = \sigma^2 \qquad \text{und} \qquad \mathbb{E}[V_{X,\mu_0}; \mathcal{N}[\mu_0, \sigma]] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(X_i - \mu_0)^2] = \mathbb{V}[X] = \sigma^2$$

(Man beachte, dass aus der Tatsache, dass der Schätzer V_X für die Varianz σ^2 der Normalverteilung erwartungstreu ist, keineswegs darauf geschlossen werden darf, dass seine Wurzel S_X ein erwartungstreuer Schätzer für die Standardabweichung σ der Normalverteilung ist!)

b) Der Schätzer V_{X,μ_0} ist wirksamer als der Schätzer V_X: Für alle $\sigma \in]0, \infty[$ gilt wegen Satz 4.2.2 und unseren Ausführungen über die Normalverteilung

$$\mathbb{V}[V_X; \mathcal{N}[\mu_0, \sigma]] = \frac{1}{n}\, \mathbb{E}[(X - \mathbb{E}[X])^4] - \frac{n-3}{n\,(n-1)}\, \mathbb{V}[X]^2 = \frac{2\,\sigma^4}{n-1}$$

$$\mathbb{V}[V_{X,\mu_0}; \mathcal{N}[\mu_0, \sigma]] = \frac{1}{n^2}\, \mathbb{V}\Big[\sum_{i=1}^{n} (X_i - \mu_0)^2\Big] = \frac{1}{n}\, (\mathbb{E}[(X - \mathbb{E}[X])^4] - \mathbb{V}[X]^2) = \frac{2\,\sigma^4}{n}$$

Dabei haben wir verwendet, dass $\mathbb{E}[(X - \mathbb{E}[X])^4] = 3\,\sigma^4$ ist, was sich mit *Mathematica* leicht zeigen lässt:

ExpectedValue[(x − μ)^4, NormalDistribution[μ, σ], x]

$3\,\sigma^4$

c) Beide Folgen $(V_X^{(n)})_{n\in\mathbb{N}}$ und $(V_{X,\mu_0}^{(n)})_{n\in\mathbb{N}}$ sind konsistent: Diese Tatsache folgt wieder unmittelbar aus der Ungleichung von Tschebyscheff zusammen mit den eben hergeleiteten Formeln für die Varianzen.

5.2.4 Beispiel: Für das zwei-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{N}[\mu, \sigma] \mid \{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[\}$ für die Verteilung der Zufallsvariablen X untersuche man die beiden Schätzer V_X und ${}^{(2)}\tilde{\overline{X}}$ für den Parameter σ^2 (beim Schätzer ${}^{(2)}\tilde{\overline{X}}$ handelt es sich um das zweite zentrale Moment).

▼

Lösung: a) Der Schätzer V_X ist erwartungstreu: Für alle $\{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[$ gilt nämlich wegen Satz 4.2.2 und unseren Ausführungen über die Normalverteilung

$$\mathbb{E}[V_X; \mathcal{N}[\mu, \sigma]] = \mathbb{V}[X] = \sigma^2$$

Der Schätzer ${}^{(2)}\tilde{\overline{X}}$ ist hingegen nicht erwartungstreu: Für alle $\{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[$ gilt nämlich

$$\mathbb{E}[{}^{(2)}\tilde{\overline{X}}; \mathcal{N}[\mu, \sigma]] = \mathbb{E}[\frac{n-1}{n} V_X; \mathcal{N}[\mu, \sigma]] = \frac{n-1}{n} \mathbb{E}[V_X; \mathcal{N}[\mu, \sigma]] = \frac{n-1}{n} \sigma^2$$

b) Für alle $\{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[$ ergibt sich für die mittleren quadratischen Fehler der beiden Schätzer V_X und ${}^{(2)}\tilde{\overline{X}}$ wegen Satz 4.2.2 und unseren Ausführungen über die Normalverteilung

$$\mathbb{E}[(V_X - \sigma^2)^2; \mathcal{N}[\mu, \sigma]] = \mathbb{V}[V_X; \mathcal{N}[\mu, \sigma]] = \frac{1}{n} \mathbb{E}[(X - \mathbb{E}[X])^4] - \frac{n-3}{n(n-1)} \mathbb{V}[X]^2 = \frac{2}{n-1} \sigma^4$$

$$\mathbb{E}[({}^{(2)}\tilde{\overline{X}} - \sigma^2)^2; \mathcal{N}[\mu, \sigma]] = \mathbb{E}[(\frac{n-1}{n} (V_X - \sigma^2) - \frac{1}{n} \sigma^2)^2; \mathcal{N}[\mu, \sigma]] =$$

$$= \frac{(n-1)^2}{n^2} \mathbb{E}[((V_X - \sigma^2))^2; \mathcal{N}[\mu, \sigma]] + \frac{1}{n^2} \sigma^4 = \frac{2n-1}{n^2} \sigma^4$$

Man beachte, dass stets $(2n-1)/n^2 < 2/(n-1)$ ist. Damit besitzt der nicht erwartungstreue Schätzer ${}^{(2)}\tilde{\overline{X}}$ stets einen kleineren mittleren quadratischen Fehler als der erwartungstreue Schätzer V_X.

c) Beide Folgen $(V_X^{(n)})_{n\in\mathbb{N}}$ und $({}^{(2)}\tilde{\overline{X}}^{(n)})_{n\in\mathbb{N}}$ sind konsistent: Diese Tatsache folgt wieder direkt aus der Ungleichung von Tschebyscheff unter Berücksichtigung der folgenden Formeln für die Varianzen

$$\mathbb{V}[V_X^{(n)}; \mathcal{N}[\mu, \sigma]] = \mathbb{E}[(V_X - \sigma^2)^2; \mathcal{N}[\mu, \sigma]] = \frac{2}{n-1} \sigma^4$$

$$\mathbb{V}[{}^{(2)}\tilde{\overline{X}}^{(n)}; \mathcal{N}[\mu, \sigma]] = \mathbb{V}[\frac{n-1}{n} V_X^{(n)}; \mathcal{N}[\mu, \sigma]] = \frac{2(n-1)}{n^2} \sigma^4$$

5.2.5 Beispiel: Für das zwei-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{N}[\mu, \sigma] \mid \{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[\}$ für die Verteilung der Zufallsvariablen X untersuche man die beiden Schätzer $\overline{X}$ und $\tilde{X}$ für den Parameter μ (man vergleiche dazu Beispiel 5.1.3).

▼

Lösung: a) Der Schätzer $\overline{X}$ ist erwartungstreu: Für alle $\{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[$ gilt nämlich wegen Satz 4.2.2 und unseren Ausführungen über die Normalverteilung

$$\mathbb{E}[\overline{X}; \mathcal{N}[\mu, \sigma]] = \mathbb{E}[X] = \mu$$

b) Der Schätzer $\tilde{X}$ ist ebenfalls erwartungstreu: Im Fall $n = 2m + 1$ gilt für alle $\{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[$ wegen Satz 4.2.5 (die Verteilungsdichte bzw die Verteilungsfunktion der Normalverteilung mit den Parametern μ und σ bezeichnen wir mit $\mathbb{f}_{\mu,\sigma}[x]$ bzw $\mathbb{F}_{\mu,\sigma}[x]$)

$$\mathbb{E}[\tilde{X}; \mathcal{N}[\mu, \sigma]] = \mathbb{E}[X^*_{m+1}; \mathcal{N}[\mu, \sigma]] = \frac{(2m+1)!}{m!\, m!} \int_{-\infty}^{\infty} x\, \mathbb{f}_{\mu,\sigma}[x]\, \mathbb{F}_{\mu,\sigma}[x]^m\, (1 - \mathbb{F}_{\mu,\sigma}[x])^m\, dx = \mu$$

Wir haben dieses Integral dabei mit Hilfe von *Mathematica* berechnet: Dazu haben wir zuerst den Integranden unter Verwendung der Transformation $x \to \mu + \sqrt{2}\ \sigma\, t$ umgeformt, mit FullSimplify vereinfacht und anschließend das gesuchte Integral ermittelt (man vergesse dabei nicht auf den Faktor $\sqrt{2}\ \sigma$, welcher von der Variablensubstitution $x \to \mu + \sqrt{2}\ \sigma\, t$ herrührt):

```
f[x_] = PDF[NormalDistribution[μ, σ], x];
F[x_] = CDF[NormalDistribution[μ, σ], x];
integrand = FullSimplify[x f[x] F[x]^m (1 – F[x])^m /. x → μ + Sqrt[2] σ t];
Integrate[(2 m + 1)! / (m! m!) Sqrt[2] σ integrand, {t, –∞, ∞}, Assumptions –> {m ∈ Integers, m > 0}]
Clear[f, F, integrand]
```

```
μ
```

Analog dazu gilt im Fall $n = 2m$ für alle $\{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[$

$$\mathbb{E}[\tilde{X}; \mathcal{N}[\mu, \sigma]] = \frac{1}{2} (\mathbb{E}[X^*_m; \mathcal{N}[\mu, \sigma]] + \mathbb{E}[X^*_{m+1}; \mathcal{N}[\mu, \sigma]]) =$$

$$= \frac{(2m)!}{2\, m!\, (m-1)!} \int_{-\infty}^{\infty} x\, \mathbb{f}_{\mu,\sigma}[x]\, \mathbb{F}_{\mu,\sigma}[x]^{m-1}\, (1 - \mathbb{F}_{\mu,\sigma}[x])^{m-1}\, dx = \mu$$

Wir haben auch dieses Integral mit Hilfe von *Mathematica* berechnet: Dazu haben wir wieder zuerst den Integranden unter Verwendung der Transformation $x \to \mu + \sqrt{2}\ \sigma\, t$ umgeformt und vereinfacht und anschließend das gesuchte Integral ermittelt (man vergesse dabei ebenfalls nicht auf den Faktor $\sqrt{2}\ \sigma$):

```
f[x_] = PDF[NormalDistribution[μ, σ], x];
F[x_] = CDF[NormalDistribution[μ, σ], x];
integrand = FullSimplify[x f[x] F[x]^(m–1) (1 – F[x])^(m–1) /. x → μ + Sqrt[2] σ t];
Integrate[(2 m)! / (2 m! (m – 1)!) Sqrt[2] σ integrand, {t, –∞, ∞}, Assumptions –> {m ∈ Integers, m > 0}]
Clear[f, F, integrand]
```

```
μ
```

c) Die Varianz dieser beiden Schätzer lässt sich (wenn überhaupt) nur schwer berechnen. Daher sind wir derzeit nicht in der Lage, diese beiden Schätzer hinsichtlich ihrer Wirksamkeit miteinander zu vergleichen. Unter Verwendung tiefliegender Erkenntnisse über Schätzer lässt sich aber zeigen, dass der Schätzer $\overline{X}$ wirksamer ist als der Schätzer $\tilde{X}$.

5.2.6 Beispiel: Für das zwei-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{U}[\{a, b\}] \mid a < b \in \mathbb{R}\}$ für die Verteilung der Zufallsvariablen X untersuche man die beiden Schätzer $\overline{X}$ und $\left(X^*_1 + X^*_n\right)/2$ für den Parameter $(a + b)/2$.

▼

Lösung: a) Beide Schätzer $\overline{X}$ und $(X^*_1 + X^*_n)/2$ sind erwartungstreu: Für alle $a < b \in \mathbb{R}$ gilt wegen Satz 4.2.2 und

Beispiel 4.2.7

$$\mathbb{E}[\overline{X}; \mathcal{U}[\{a, b\}]] = \mathbb{E}[X] = (a+b)/2$$

$$\mathbb{E}[(X_1^* + X_n^*)/2; \mathcal{U}[\{a, b\}]] = (a + \frac{(b-a)}{n+1} + a + \frac{n\,(b-a)}{n+1})/2 = (a+b)/2$$

b) Der Schätzer $(X_1^* + X_n^*)/2$ ist wirksamer als der Schätzer $\overline{X}$: Für alle $a < b \in \mathbb{R}$ gilt wegen Satz 4.2.2 und Beispiel 4.2.10

$$\mathbb{V}[\overline{X}; \mathcal{U}[\{a, b\}]] = \frac{1}{n}\,\mathbb{V}[X] = \frac{1}{12\,n}\,(b-a)^2$$

$$\mathbb{V}[(X_1^* + X_n^*)/2; \mathcal{U}[\{a, b\}]] = \frac{1}{2\,(n+1)\,(n+2)}\,(b-a)^2$$

c) Beide Folgen $(\overline{X}^{(n)})_{n\in\mathbb{N}}$ und $((X_1^{*(n)} + X_n^{*(n)})/2)_{n\in\mathbb{N}}$ sind konsistent: Diese Tatsache folgt wieder direkt aus der Ungleichung von Tschebyscheff zusammen mit den eben hergeleiteten Formeln für die Varianz.

5.2.7 Beispiel: Für das ein-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{E}[\lambda] \mid \lambda > 0\}$ für die Verteilung der Zufallsvariablen X prüfe man, ob der Schätzer

$$(\overline{X})^{-1} = \frac{n}{X_1 + X_2 + \ldots + X_n}$$

für den Parameter λ erwartungstreu ist (man vergleiche dazu Beispiel 5.1.4).

▼

Lösung: Unter Berücksichtigung der Tatsache, dass die Summe $Z = X_1 + X_2 + \ldots + X_n$ von n vollständig unabhängigen, $\mathcal{E}[\lambda]$-verteilten Zufallsvariablen $X_1, X_2, \ldots, X_n$ bekanntlich $\mathcal{E}[n, \lambda]$-verteilt ist, gilt für alle $\lambda > 0$

$$\mathbb{E}[(\overline{X})^{-1}; \mathcal{E}[\lambda]] = \mathbb{E}[\frac{n}{X_1 + X_2 + \ldots + X_n}; \mathcal{E}[\lambda]] = \mathbb{E}[\frac{n}{Z}; \mathcal{E}[n, \lambda]] = \frac{n}{n-1}\,\lambda$$

wobei wir den Erwartungswert mit Hilfe von *Mathematica* und dem Befehl ExpectedValue berechnet haben

```
ExpectedValue[n/z, GammaDistribution[n, 1/λ], z, Assumptions -> {λ > 0, n > 1}]
```

```
 n λ
------
-1 + n
```

Der Schätzer $(\overline{X})^{-1}$ für den Parameter λ ist somit nicht erwartungstreu, wohl aber der modifizierte Schätzer

$$(\tilde{\overline{X}})^{-1} = (\frac{n}{n-1}\,\overline{X})^{-1} = \frac{n-1}{X_1 + X_2 + \ldots + X_n}$$

Oft geht es nur darum, den Erwartungswert $\mathbb{E}[X]$ bzw die Varianz $\mathbb{V}$[X] von X zu schätzen. Diese Fragestellung wird mit der folgenden Bemerkung beantwortet:

5.2.8 Bemerkung: Wegen Satz 4.2.2 gilt generell

a) Der Mittelwert $\overline{X}$ ist stets ein erwartungstreuer Schätzer für den Erwartungswert $\mathbb{E}[X]$ von X;

b) Die Varianz V_X ist stets ein erwartungstreuer Schätzer für die Varianz $\mathbb{V}[X]$ von X.

c) Die Standardabweichung S_X ist ein guter (nicht erwartungstreuer) Schätzer für die Streuung $\mathbb{S}[X]$ von X.

5.3 Die Momenten-Methode

Sind die zu schätzenden Parameter inhaltlich nicht interpretierbar, so benötigt man Methoden, mit denen sich vernünftige Schätzer (bzw Schätzwerte) für diese Parameter erzeugen lassen. Wir befassen uns in diesem Abschnitt mit der sogenannten Momenten-Methode:

Sei $\mathcal{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\} \in \Theta\}$ ein k-parametrisches a-priori Modell für die Verteilung der Zufallsvariablen X und sei $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ ausgewählte mathematische Stichprobe vom Umfang n. Gesucht sind Schätzer für die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$.

5.3.1 Momenten-Methode: Für jedes $r \in \{1, 2, \ldots, k\}$ schätze man einerseits das r-te Moment $\mathbb{E}[X^r; \mathbb{P}[\vec{\vartheta}]]$ von X durch das r-te Moment ${}^{(r)}\overline{X}$ (man überzeugt sich mühelos davon, dass ${}^{(r)}\overline{X}$ ein erwartungstreuer Schätzer für $\mathbb{E}[X^r; \mathbb{P}[\vec{\vartheta}]]$ ist) und drücke andererseits dieses r-te Moment $\mathbb{E}[X^r]$ durch die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ in der Form $m_r[\vartheta_1, \vartheta_2, \ldots, \vartheta_k]$ aus. Wird das auf diese Weise entstehende Gleichungssystem

$$(*)\qquad \left.\begin{array}{rcl} m_1[\vartheta_1, \vartheta_2, \ldots, \vartheta_k] & = & {}^{(1)}\overline{X} \\ m_2[\vartheta_1, \vartheta_2, \ldots, \vartheta_k] & = & {}^{(2)}\overline{X} \\ \ldots & & \ldots \\ m_k[\vartheta_1, \vartheta_2, \ldots, \vartheta_k] & = & {}^{(k)}\overline{X} \end{array}\right\}$$

nach den Parametern $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ aufgelöst, so erhält man die Schätzer

$$S_1 = \hat{\vartheta}_1[{}^{(1)}\overline{X}, {}^{(2)}\overline{X}, \ldots, {}^{(k)}\overline{X}], S_2 = \hat{\vartheta}_2[{}^{(1)}\overline{X}, {}^{(2)}\overline{X}, \ldots, {}^{(k)}\overline{X}], \ldots, S_k = \hat{\vartheta}_k[{}^{(1)}\overline{X}, {}^{(2)}\overline{X}, \ldots, {}^{(k)}\overline{X}]$$

für die unbekannten Parametern $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$.

Zur Momenten-Methode sind einige Ergänzungen angebracht:

■ Das Gleichungssystem (*) ist in der Regel nicht linear und lässt sich in vielen Fällen nur numerisch lösen. Man gewinnt dann mit Hilfe der Momentenmethode nicht Schätzer $S_1, S_2, \ldots, S_k$ sondern nur (von der jeweiligen Stichprobe abhängige) Schätzwerte $s_1, s_2, \ldots, s_k$ für die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$.

■ Ist das Gleichungssystem (*) linear, so sind die Schätzer $S_1, S_2, \ldots, S_k$ für die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ erwartungstreu.

■ Ist $\mathcal{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2\} \in \Theta\}$ ein zwei-parametrisches a-priori Modell für die Verteilung von X, so gelangt man zu den gleichen Schätzern für die beiden Parameter ϑ_1, ϑ_2, wenn man statt (*) das Gleichungssystem

$$(**)\qquad \left.\begin{array}{rl} \mathbb{E}[X; \mathbb{P}[\{\vartheta_1, \vartheta_2\}]] & = \overline{X} \\ \mathbb{V}[X; \mathbb{P}[\{\vartheta_1, \vartheta_2\}]] & = V_X \end{array}\right\}$$

nach den Parametern ϑ_1, ϑ_2 auflöst (man vergleiche dazu etwa Beispiel 24.2.3).

■ Unter dem mit der Momenten-Methode gewonnenen Schätzer für den Parameter $\theta = \theta[\vec{\vartheta}]$ versteht man den Schätzer $S = \theta[\{S_1, S_2, \ldots, S_k\}]$, wobei $S_1, S_2, \ldots, S_k$ die mit der Momenten-Methode gewonnenen Schätzer für die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ sind.

■ Über den mittleren quadratischen Fehler, die Wirksamkeit bzw die Konsistenz der mit der Momenten-Methode gewonnenen Schätzer kann keine allgemeine Aussage gemacht werden.

Mit den folgenden Beispielen wird die Momenten-Methode näher erläutert:

5.3.2 Beispiel: Für das zwei-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{LN}[\mu, \sigma] \mid \{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[\}$ für die Verteilung der Zufallsvariablen X sollen Schätzer für die Parameter μ und σ ermittelt werden.

▼

Lösung: Wir verwenden die Momenten-Methode und haben dazu das Gleichungssystem (Bemerkung 24.1.3)

$$\left.\begin{array}{lll} m_1[\mu, \sigma] = \mathbb{E}[X; \mathcal{LN}[\mu, \sigma]] & = \operatorname{Exp}[\mu + \sigma^2/2] & = {}^{(1)}\overline{X} \\ m_2[\mu, \sigma] = \mathbb{E}[X^2; \mathcal{LN}[\mu, \sigma]] = & \operatorname{Exp}[2\mu + 2\sigma^2] & = {}^{(2)}\overline{X} \end{array}\right\}$$

nach den beiden unbekannten Parametern μ und σ aufzulösen. Wir logarithmieren dazu beide Gleichungen und erhalten das zum ursprünglichen Gleichungssystem äquivalente Gleichungssystem

$$\left.\begin{array}{ll} 2\mu + \sigma^2 & = 2\operatorname{Log}[{}^{(1)}\overline{X}] \\ 2\mu + 2\sigma^2 & = \operatorname{Log}[{}^{(2)}\overline{X}] \end{array}\right\}$$

welches unter Verwendung von *Mathematica*

```
Solve[{2 μ + σ^2 == 2 Log[x1], 2 μ + 2 σ^2 == Log[x2]}, {μ, σ}]
```

$$\left\{\left\{\mu \to \frac{1}{2}(4\operatorname{Log}[x1] - \operatorname{Log}[x2]),\ \sigma \to -\sqrt{-2\operatorname{Log}[x1] + \operatorname{Log}[x2]}\right\},\right.$$
$$\left.\left\{\mu \to \frac{1}{2}(4\operatorname{Log}[x1] - \operatorname{Log}[x2]),\ \sigma \to \sqrt{-2\operatorname{Log}[x1] + \operatorname{Log}[x2]}\right\}\right\}$$

auf die Schätzer

$$S_1 = 2\operatorname{Log}[{}^{(1)}\overline{X}] - \operatorname{Log}[{}^{(2)}\overline{X}]/2 \qquad \text{bzw} \qquad S_2 = \sqrt{\operatorname{Log}[{}^{(2)}\overline{X}] - 2\operatorname{Log}[{}^{(1)}\overline{X}]}$$

für die beiden Parametern μ und σ führt (da σ positiv sein muss, kommt nur die zweite der von *Mathematica* ausgegebenen Lösungen in Frage).

5.3.3 Beispiel: Für das zwei-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{E}xtrem[\mu, \beta] \mid \{\mu, \beta\} \in \mathbb{R} \times]0, \infty[\}$ für die Verteilung der Zufallsvariablen X sollen Schätzer für die Parameter μ und β ermittelt werden.

▼

Lösung: Wir verwenden die Momenten-Methode und haben dazu das Gleichungssystem (Bemerkung 24.2.2)

$$\left.\begin{array}{llll} m_1[\mu, \sigma] & = \mathbb{E}[X; \mathcal{E}xtrem[\mu, \beta]] & = \gamma\beta + \mu & = {}^{(1)}\overline{X} \\ m_2[\mu, \sigma] & = \mathbb{E}\left[X^2; \mathcal{E}xtrem[\mu, \beta]\right] & = (\gamma\beta + \mu)^2 + \beta^2\pi^2/6 & = {}^{(2)}\overline{X} \end{array}\right\}$$

nach den beiden unbekannten Parametern μ und β aufzulösen ($\gamma = 0.577216$ bezeichnet dabei die sogenannte Euler'sche Konstante). Unter Verwendung von *Mathematica*

```
Solve[{γ β + μ == x1, (γ β + μ)^2 + β^2 π^2/6 == x2}, {μ, β}]
```

$$\left\{\left\{\mu \to x1 - \frac{\sqrt{6}\sqrt{-x1^2 + x2}\,\gamma}{\pi},\ \beta \to \frac{\sqrt{6}\sqrt{-x1^2 + x2}}{\pi}\right\},\right.$$
$$\left.\left\{\mu \to x1 + \frac{\sqrt{6}\sqrt{-x1^2 + x2}\,\gamma}{\pi},\ \beta \to -\frac{\sqrt{6}\sqrt{-x1^2 + x2}}{\pi}\right\}\right\}$$

ergeben sich damit die Schätzer

$$S_1 = {}^{(1)}\overline{X} - \frac{\gamma\sqrt{6}}{\pi}\sqrt{{}^{(2)}\overline{X} - ({}^{(1)}\overline{X})^2} \qquad \text{bzw} \qquad S_2 = \frac{\sqrt{6}}{\pi}\sqrt{{}^{(2)}\overline{X} - ({}^{(1)}\overline{X})^2}$$

für die beiden Parametern μ und β (da β positiv sein muss, kommt nur die erste der von *Mathematica* ausgegebenen Lösungen in Frage).

5.3.4 Beispiel: Für das zwei-parametrische a-priori Modell $\mathcal{P}_X = \{\mathcal{U}[\{a, b\}] \mid a, b \in \mathbb{R} \text{ mit } a < b\}$ für die Verteilung der Zufallsvariablen X sollen Schätzer für die Parameter a und b ermittelt werden.

▼

Lösung: Wir verwenden die Momenten-Methode und haben dazu das Gleichungssystem

$$\left.\begin{array}{lll} m_1[a, b] = \mathbb{E}[X; \mathcal{U}[\{a, b\}]] & = (a+b)/2 & = {}^{(1)}\overline{X} \\ m_2[\mu, \sigma] = \mathbb{E}[X^2; \mathcal{U}[\{a, b\}]] & = (a^2 + a\,b + b^2)/3 & = {}^{(2)}\overline{X} \end{array}\right\}$$

nach den beiden unbekannten Parametern a und b aufzulösen. Unter Verwendung von *Mathematica*

```
Solve[{(a + b)/2 == x1, (a^2 + a b + b^2)/3 == x2}, {a, b}]
```

$$\left\{\left\{a \to x1 - \sqrt{3}\sqrt{-x1^2 + x2},\ b \to x1 + \sqrt{3}\sqrt{-x1^2 + x2}\right\},\ \left\{a \to x1 + \sqrt{3}\sqrt{-x1^2 + x2},\ b \to x1 - \sqrt{3}\sqrt{-x1^2 + x2}\right\}\right\}$$

ergeben sich damit die Schätzer

$$S_1 = {}^{(1)}\overline{X} - \sqrt{3}\sqrt{{}^{(2)}\overline{X} - ({}^{(1)}\overline{X})^2} \qquad \text{bzw} \qquad S_2 = {}^{(1)}\overline{X} + \sqrt{3}\sqrt{{}^{(2)}\overline{X} - ({}^{(1)}\overline{X})^2}$$

für die beiden Parameter a und b (da $a < b$ sein muss, kommt nur die erste der von *Mathematica* ausgegebenen Lösungen in Frage).

5.4 Die Maximum-Likelihood-Methode

Eine weitere Methode zur Erzeugung von Schätzern ist die sogenannte Maximum-Likelihood-Methode.

Sei $\mathcal{P}_X = \{\mathbb{P}[\vec{\vartheta}] \mid \vec{\vartheta} = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\} \in \Theta\}$ ein k-parametrisches a-priori Modell für die Verteilung der Zufallsvariablen X, wobei wir voraussetzen, dass alle Verteilungen entweder diskret oder stetig sind und die Verteilungsdichte von $\mathbb{P}[\vec{\vartheta}]$ mit $\mathbb{f}_{\vec{\vartheta}}$ bezeichnet wird, und sei $\vec{X} = \{X_1, X_2, \ldots, X_n\}$ eine nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ ausgewählte mathematische Stichprobe vom Umfang n. Gesucht sind Schätzer für die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$.

5.4.1 Maximum-Likelihood-Methode: Für jede nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ ausgewählte konkrete Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ vom Umfang n bestimmt man die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ so, dass die sogenannte Likelihood-Funktion

$$L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}] = \mathbb{f}_{\vec{\vartheta}}[x_1]\, \mathbb{f}_{\vec{\vartheta}}[x_2] \ldots \mathbb{f}_{\vec{\vartheta}}[x_n]$$

maximal wird. Auf diese Weise erhält man die Schätzwerte

$$s_1 = \hat{\vartheta}_1[\{x_1, x_2, \ldots, x_n\}],\ s_2 = \hat{\vartheta}_2[\{x_1, x_2, \ldots, x_n\}], \ldots, s_k = \hat{\vartheta}_k[\{x_1, x_2, \ldots, x_n\}]$$

und damit die Schätzer

$$S_1 = \hat{\vartheta}_1[\{X_1, X_2, \ldots, X_n\}],\ S_2 = \hat{\vartheta}_2[\{X_1, X_2, \ldots, X_n\}], \ldots, S_k = \hat{\vartheta}_k[\{X_1, X_2, \ldots, X_n\}]$$

für die Parametern $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$.

Zur Maximum-Likelihood-Methode und den damit erzeugten Schätzern sind einige Ergänzungen angebracht:

■ Ist $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ diskret, so gilt

$$L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}] = \mathbb{P}[\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \ldots \cap \{X_n = x_n\}]$$

Ist $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ stetig, so gilt

$$L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}] = \frac{\mathbb{P}[\{X_1 \in [x_1, x_1 + d\,x_1]\} \cap \{X_2 \in [x_2, x_2 + d\,x_2]\} \cap \ldots \cap \{X_n \in [x_n, x_n + d\,x_n]\}]}{d\,x_1\, d\,x_2 \ldots d\,x_n}$$

Die Likelihood-Funktion $L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}]$ entspricht damit der Wahrscheinlichkeit dafür, die konkrete Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ zu erhalten, falls X nach dem Verteilungsgesetz $\mathbb{P}_X = \mathbb{P}[\vec{\vartheta}]$ verteilt ist. Das Prinzip der Maximum-Likelihood-Methode besteht somit darin, die unbekannten Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ so zu bestimmen, dass diese Wahrscheinlichkeit $L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}]$ maximal wird.

■ Das Aufsuchen von Schätzern mit der Maximum-Likelihood-Methode läuft auf das Lösen der Aufgabe

$$L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}] = \text{Max!}$$

hinaus. Da der Logarithmus eine monoton zunehmende Funktion ist, ist diese Aufgabe äquivalent zur oft einfacheren Aufgabe

$$\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \vec{\vartheta}]] = \text{Max!}$$

■ Unter dem Maximum-Likelihood-Schätzer für den Parameter $\theta = \theta[\{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\}]$ versteht man den Schätzer $S = \theta[\{S_1, S_2, \ldots, S_k\}]$, wobei $S_1, S_2, \ldots, S_k$ die mit der Maximum-Likelihood-Methode gewonnenen Schätzer für die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ sind.

■ Maximum-Likelihood-Schätzer sind nicht notwendig erwartungstreu. Unter schwachen Voraussetzungen lässt sich aber zeigen, dass diese Schätzer asymptotisch (also für den Fall, dass der Stichprobenumfang n gegen ∞ geht) erwartungstreu und normalverteilt sind und obendrein wirksamer sind, als alle anderen erwartungstreuen Schätzer. Man sagt in diesem Zusammenhang, dass Maximum-Likelihood-Schätzer ban (best asymptotic normal) sind.

Mit den folgenden Beispielen wird die Maximum-Likelihood-Methode näher erläutert:

5.4.2 Beispiel: Für das ein-parametrische a-priori Modell $\mathcal{P}_X = \{\mathcal{E}[\lambda] \mid \lambda \in]0, \infty[\}$ für die Verteilung der Zufallsvariablen X ist der Maximum-Likelihood-Schätzer für den Parameter λ gesucht.

▼

Lösung: Wir haben den Parameter λ so zu bestimmen, dass der Logarithmus der Likelihood-Funktion, also

$$\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \lambda]] = \text{Log}[\prod_{i=1}^{n} \lambda\, e^{-\lambda x_i}] = n\, \text{Log}[\lambda] - \lambda \sum_{i=1}^{n} x_i$$

maximal wird. Dazu differenzieren wir $\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \lambda]]$ nach λ, setzen diese Ableitung gleich 0, lösen die dabei entstehende Gleichung

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

nach λ auf und erhalten auf diese Weise den Maximum-Likelihood-Schätzer

$$S = \frac{n}{\sum_{i=1}^{n} X_i} = (\overline{X})^{-1}$$

für den Parameter λ (man vergleiche dazu Beispiel 5.1.4 und Beispiel 5.2.7, wobei aus dem zuletzt genannten Beispiel hervorgeht, dass der Maximum-Likelihood-Schätzer für den Parameter λ nicht erwartungstreu ist).

5.4.3 Beispiel: Für das zwei-parametrische a-priori Modell $\mathbb{P}_X = \{\mathcal{N}[\mu, \sigma] \mid \{\mu, \sigma\} \in \mathbb{R} \times]0, \infty[\}$ für die Verteilung der Zufallsvariablen X sind Maximum-Likelihood-Schätzer für die Parameter μ und σ^2 gesucht.

▼

Lösung: Wir haben zuerst die beiden Parameter μ und σ so zu bestimmen, dass der Logarithmus der Likelihood-Funktion, also

$$\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \mu, \sigma]] = -n\, \text{Log}[\sqrt{2\pi}\ \sigma] - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

maximal wird. Dazu differenzieren wir $\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \mu, \sigma]]$ partiell nach μ bzw σ, setzen diese partiellen Ableitungen gleich 0, lösen das dabei entstehende Gleichungssystem

$$\left.\begin{array}{lll} \frac{\partial}{\partial \mu}: & -\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) & = 0 \\ \frac{\partial}{\partial \sigma}: & -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 & = 0 \end{array}\right\}$$

nach μ und σ auf und erhalten auf diese Weise die Maximum-Likelihood-Schätzer

$$S_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} \qquad \text{und} \qquad S_2 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2} = \sqrt{{}^{(2)}\tilde{\overline{X}}}$$

für die Parameter μ und σ und damit den Maximum-Likelihood-Schätzer ${}^{(2)}\tilde{\overline{X}}$ für den Parameter σ^2 (aus Beispiel 5.2.3 geht hervor, dass dieser Schätzer für den Parameter σ^2 nicht erwartungstreu ist).

5.4.4 Beispiel: Für das fünf-parametrische a-priori Modell

$$\mathbb{P}_{X,Y} = \{\mathcal{MN}[\vec{\mu}, \Sigma] \mid \vec{\mu} \in \mathbb{R}^2, \Sigma \in \mathbb{R}_2^2\}$$

mit $\vec{\mu} = \{\mu_1, \mu_2\}$ und $\Sigma = \{\{\sigma_1{}^2, \varrho\, \sigma_1\, \sigma_2\}, \{\varrho\, \sigma_1\, \sigma_2, \sigma_2{}^2\}\}$ für die gemeinsame Verteilung der Zufallsvariablen X und Y sind die Maximum-Likelihood-Schätzer für die Parameter $\mu_1, \mu_2, \sigma_1, \sigma_2$ und ρ gesucht.

▼

Lösung: Wir haben die Parameter μ_1, μ_2, σ_1, σ_2 und ϱ so zu bestimmen, dass der Logarithmus der Likelihood-Funktion, also

$$\mathrm{Log}[L[\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\} \mid \vec{\mu}, \mathbf{\Sigma}]] = \mathrm{Log}[\prod_{i=1}^{n} \frac{1}{2\,\pi\,\sigma_1\,\sigma_2\,\sqrt{1-\rho^2}}\ \mathrm{Exp}[-\frac{1}{2}(\vec{x}_i - \vec{\mu})\,\mathbf{\Sigma}^{-1}(\vec{x}_i - \vec{\mu})^t] =$$

$$= -n\,\mathrm{Log}[2\,\pi\,\sigma_1\,\sigma_2\,\sqrt{1-\rho^2}\ ] - \frac{1}{2\,(1-\rho^2)} \sum_{i=1}^{n} \left(\frac{(x_i-\mu_1)^2}{\sigma_1{}^2} - 2\,\rho\,\frac{(x_i-\mu_1)\,(y_i-\mu_2)}{\sigma_1\,\sigma_1} + \frac{(y_i-\mu_2)^2}{\sigma_2{}^2}\right)$$

maximal wird. Dazu differenzieren wir die Funktion $\mathrm{Log}[L[\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n \mid \vec{\mu}, \mathbf{\Sigma}]]$ mit Hilfe von *Mathematica* partiell nach μ_1, μ_2, σ_1, σ_2 und ρ (wir zeigen diese Rechnung exemplarisch für μ_1)

LogLikelihoodfunktion = $-$n Log[2 $\pi\,\sigma_1\,\sigma_2\,\sqrt{1-\rho^2}$] $- \frac{1}{2\left(1-\rho^2\right)} \sum_{i=1}^{n}\left(\frac{(x_i-\mu_1)^2}{\sigma_1{}^2} - 2\,\rho\,\frac{(x_i-\mu_1)\,(y_i-\mu_2)}{\sigma_1\,\sigma_2} + \frac{(y_i -}{\sigma}\right.$

D[LogLikelihoodfunktion, μ_1]

Clear[LogLikelihoodfunktion]

$$-\frac{\sum_{i=1}^{n}\left(-\frac{2\,(x_i-\mu_1)}{\sigma_1^2} + \frac{2\,\rho\,(y_i-\mu_2)}{\sigma_1\,\sigma_2}\right)}{2\,(1-\rho^2)}$$

setzen diese partiellen Ableitungen gleich 0, lösen das dabei entstehende Gleichungssystem

$$\left.\begin{array}{ll}
\frac{\partial}{\partial\mu_1}: \frac{1}{\sigma_1\,(1-\rho^2)}\left[\sum_{i=1}^{n}\frac{x_i-\mu_1}{\sigma_1} - \rho\sum_{i=1}^{n}\frac{y_i-\mu_2}{\sigma_2}\right] & = 0 \\
\frac{\partial}{\partial\mu_2}: \frac{1}{\sigma_2\,(1-\rho^2)}\left[\sum_{i=1}^{n}\frac{y_i-\mu_2}{\sigma_2} - \rho\sum_{i=1}^{n}\frac{x_i-\mu_1}{\sigma_1}\right] & = 0 \\
\frac{\partial}{\partial\sigma_1}: -\frac{n}{\sigma_1} + \frac{1}{\sigma_1\,(1-\rho^2)}\left[\sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{\sigma_1{}^2} - \rho\sum_{i=1}^{n}\frac{x_i-\mu_1}{\sigma_1}\,\frac{y_i-\mu_2}{\sigma_2}\right] & = 0 \\
\frac{\partial}{\partial\sigma_2}: -\frac{n}{\sigma_2} + \frac{1}{\sigma_2\,(1-\rho^2)}\left[\sum_{i=1}^{n}\frac{(y_i-\mu_2)^2}{\sigma_2{}^2} - \rho\sum_{i=1}^{n}\frac{x_i-\mu_1}{\sigma_1}\,\frac{y_i-\mu_2}{\sigma_2}\right] & = 0 \\
\frac{\partial}{\partial\rho}: \frac{n\,\rho}{(1-\rho^2)} - \frac{1}{(1-\rho^2)^2}\left[\rho\sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{\sigma_1{}^2} + \rho\sum_{i=1}^{n}\frac{(y_i-\mu_2)^2}{\sigma_2{}^2} - (1+\rho^2)\sum_{i=1}^{n}\frac{x_i-\mu_1}{\sigma_1}\,\frac{y_i-\mu_2}{\sigma_2}\right] & = 0
\end{array}\right\}$$

nach μ_1, μ_2, σ_1, σ_2 und ρ auf und erhalten auf diese Weise die Maximum-Likelihood-Schätzer

$$S_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$S_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}$$

$$S_3 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sqrt{{}^{(2)}\tilde{\overline{X}}}$$

$$S_4 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = \sqrt{{}^{(2)}\tilde{\overline{Y}}}$$

$$S_5 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}} = R_{X,Y}$$

für die Parameter μ_1, μ_2, σ_1, σ_2 und ρ (hinsichtlich der Statistiken ${}^{(2)}\tilde{\overline{X}}$, ${}^{(2)}\tilde{\overline{Y}}$ und $R_{X,Y}$ vergleiche man Definition 3.1.9 und Definition 3.2.1).

5.4.5 Beispiel: Für das zwei-parametrische a-priori Modell $\mathcal{P}_X = \{\mathcal{U}[\{a, b\}] \mid a < b \in \mathbb{R}\}$ für die Verteilung der Zufallsvariablen X bestimme man die Maximum-Likelihood-Schätzer für die beiden Parameter a und b (man vergleiche dazu Beispiel 5.3.4).

▼

Lösung: Wir haben die beiden Parameter a und b so zu bestimmen, dass die Likelihood-Funktion

$$L[\{x_1, x_2, \ldots, x_n\} \mid a, b] = \begin{cases} (b-a)^{-n} & \text{für } x_1, x_2, \ldots, x_n \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

maximal wird. Diese Extremwertaufgabe führt offensichtlich auf die beiden (wegen Beispiel 4.2.7 nicht erwartungstreuen) Maximum-Likelihood-Schätzer X_1^* und X_n^* für die beiden Parameter a und b. Aus dem eben erwähnten Beispiel entnimmt man aber auch, dass die modifizierten Schätzer

$$S_1 = X_1^* - \frac{1}{n}(X_n^* - X_1^*) \qquad \text{und} \qquad S_2 = X_n^* + \frac{1}{n}(X_n^* - X_1^*)$$

erwartungstreue Schätzer für die beiden Parameter a und b sind.

5.4.6 Beispiel: Für das ein-parametrische a-priori Modell $\mathcal{P}_X = \{\mathcal{R}[\sigma] \mid \sigma > 0\}$ für die Verteilung der Zufallsvariablen X bestimme man den Maximum-Likelihood-Schätzer für den Parameter σ.

▼

Lösung: Wir haben den Parameter σ so zu bestimmen, dass der Logarithmus der Likelihood-Funktion, also

$$\mathrm{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \sigma]] = -2\, n\, \mathrm{Log}[\sigma] + \sum_{i=1}^{n} \mathrm{Log}[x_i] - \frac{1}{2\,\sigma^2} \sum_{i=1}^{n} x_i$$

maximal wird. Diese Extremwertaufgabe führt offenbar auf den Maximum-Likelihood-Schätzer $S = \sqrt{\overline{X}/2}$ für den Parameter σ.

Nach diesen eher theoretischen Beispielen folgen einige Beispiele, mit denen gezeigt wird, wie sich die Maximum-Likelihood-Methode in der Praxis einsetzen lässt:

5.4.7 Beispiel: Eine biologische Theorie besagt, dass bei einer bestimmten Art von Tieren die Merkmale *AB* bzw *Ab* bzw *aB* bzw *ab* mit den Wahrscheinlichkeiten $(2+\vartheta)/4$ bzw $(1-\vartheta)/4$ bzw $(1-\vartheta)/4$ bzw $\vartheta/4$ auftreten, wobei der Parameter $\vartheta \in]0, 1[$ unbekannt ist. Es wurden n zufällig ausgewählte Tiere dieser Art untersucht und dabei festgestellt, dass davon n_1 das Merkmal *AB*, n_2 das Merkmal *Ab*, n_3 das Merkmal *aB* und n_4 das Merkmal *ab* trugen. Man schätze den Parameter ϑ.

▼

Lösung: Wir haben den Parameter ϑ so zu bestimmen, dass für eine Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ vom Umfang n, bei der n_1 der x_i gleich *AB*, n_2 der x_i gleich *Ab*, n_3 der x_i gleich *aB* und n_4 der x_i gleich *ab* sind, der Logarithmus der Likelihood-Funktion, also

$$\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \vartheta]] = n_1 \text{Log}[2+\vartheta] + (n_2+n_3)\text{Log}[1-\vartheta] + n_4 \text{Log}[\vartheta] - n\,\text{Log}[4]$$

maximal wird. Diese Extremwertaufgabe lässt sich unter Verwendung von *Mathematica* leicht lösen

```
Solve[D[n1 Log[2 + ϑ] + (n2 + n3) Log[1 − ϑ] + n4 Log[ϑ] − n Log[4], ϑ] == 0, ϑ]
```

```
{{ϑ → (n1 - 2 n2 - 2 n3 - n4 - √(8 n4 (n1 + n2 + n3 + n4) + (-n1 + 2 n2 + 2 n3 + n4)^2)) /
    (2 (n1 + n2 + n3 + n4))},
 {ϑ → (n1 - 2 n2 - 2 n3 - n4 + √(8 n4 (n1 + n2 + n3 + n4) + (-n1 + 2 n2 + 2 n3 + n4)^2)) /
    (2 (n1 + n2 + n3 + n4))}}
```

und führt auf den Schätzwert (da $\vartheta > 0$ sein muss, kommt offenbar nur die zweite Lösung in Frage)

$$\hat{\vartheta}[\{x_1, x_2, \ldots, x_n\}] = \frac{n_1 - 2n_2 - 2n_3 - n_4 + \sqrt{8 n_4 n + (n_1 - 2n_2 - 2n_3 - n_4)^2}}{2n}$$

für den Parameter ϑ.

5.4.8 Beispiel: Um den Prozentsatz p der defekten Stücke in einer großen Lieferung zu schätzen, entschließt man sich, so lange zu prüfen, bis n defekte Stücke gefunden wurden. Man ermittle einen Schätzwert für p, wenn dabei k Stücke geprüft werden mussten.

▼

Lösung: Die Anzahl X der geprüften und für gut befundenen Stücke ist bekanntlich $\mathcal{NB}[n, p]$ verteilt. Gemäß dem Prinzip der Maximum-Likelihood-Methode haben wir daher die Aufgabe, den Parameter p so zu bestimmen, dass die Wahrscheinlichkeit

$$\mathbb{P}[\{X = k-n\}] = \binom{k-1}{n-1} p^n (1-p)^{k-n}$$

maximal wird, was offenbar auf den (naheliegenden) Schätzwert $\hat{p} = n/k$ führt.

5.4.9 Beispiel: Um den Parameter λ eines zweidimensionalen Poissonprozesses zu schätzen, wird für n zufällig herausgegriffene Kreise (die sich nicht überschneiden) mit den Radien $r_1, r_2, \ldots, r_n$ die Anzahl der zufälligen Punkte innerhalb der einzelnen Kreise bestimmt. Aus dieser Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ bestimme man den Maximum-Likelihood-Schätzer für den Parameter λ.

▼

Lösung: Die Anzahl X_i der zufälligen Punkte in einem Kreis mit Radius r_i ist bekanntlich $\mathcal{P}[r_i^2 \pi \lambda]$-verteilt. Gemäß dem Prinzip der Maximum-Likelihood-Methode haben wir die Aufgabe, den Parameter λ so zu bestimmen, dass die Wahrscheinlichkeit

$$\mathbb{P}[\{X_1 = x_1\} \cap \{X_2 = x_2\} \cap \ldots \cap \{X_n = x_n\}] = \prod_{i=1}^{n} \text{Exp}[-r_i^2 \pi \lambda] \frac{(r_i^2 \pi \lambda)^{x_i}}{x_i!}$$

maximal wird, was offenbar auf den (naheliegenden) Schätzer

$$S = \sum_{i=1}^{n} X_i / \sum_{i=1}^{n} r_i^2 \pi$$

für den Parameter λ führt.

5.4.10 Beispiel: Das Datenmaterial *nerven* enthält eine Stichprobe für die Längen der Zeitintervalle zwischen zwei aufeinander folgenden Impulsen, welche von einer Nervenzelle ausgehen. Man prüfe, ob dieses Datenmaterial einer Gammaverteilung genügt.

▼

Lösung: Es bezeichne X die Länge der Intervalle zwischen zwei aufeinander folgenden Impulsen, welche von dieser Nervenzelle ausgehen. Wir nehmen an, dass die Zufallsvariable X einer Gammaverteilung genügt, dass also das zwei-parametrische a-priori Modell $\mathcal{P}_X = \{\mathcal{G}amma[\alpha, \lambda] \mid \{\alpha, \lambda\} \in]0, \infty[\times]0, \infty[\}$ vorliegt.

a) Unsere erste Aufgabe besteht darin, auf der Basis der vorliegenden Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ Schätz-werte für die beiden Parameter α und λ zu ermitteln. Wir verwenden dazu die Maximum-Likelihood-Methode, haben also die Parameter α und λ so zu bestimmen, dass der Logarithmus der Likelihood-Funktion

$$\text{Log}[L[\{x_1, x_2, \ldots, x_n\} \mid \alpha, \lambda]] = -n\,\alpha\,\text{Log}[\lambda] + (\alpha - 1) \sum_{i=1}^{n} \text{Log}[x_i] - \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n\,\text{Log}[\Gamma[\alpha]]$$

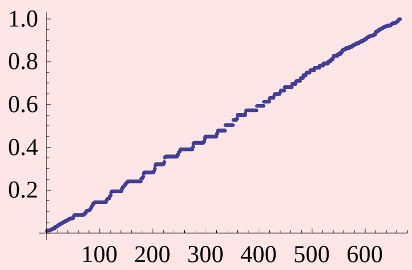
maximal wird. Wir lesen dazu das im Datenordner abgelegte File *nervenfile* ein und ermitteln für die zugehörige Stichprobe "nerven" mit Hilfe von FindMaximum das Maximum von $\text{Log}[L[x_1, x_2, \ldots, x_n \mid \alpha, \lambda]]$, wobei wir die (weitgehend willkürlichen) Startwerte $\{\alpha, 1, 2\}$ und $\{\lambda, 0.1, 1\}$ verwenden:

```
nerven = Rest[<< nervenfile];
n = Length[nerven];
a = Apply[Plus, Log[nerven]];
b = Apply[Plus, nerven];
FindMaximum[−n α Log[λ] + (α − 1) a − b / λ − n Log[Gamma[α]], {α, 1, 2}, {λ, 0.1, 1}]
α = %[[2, 1, 2]];
λ = %%[[2, 2, 2]];
```

```
{454.401, {α → 0.803072, λ → 0.235555}}
```

b) Unsere zweite Aufgabe besteht in der Überprüfung, ob die Länge X der Intervalle zwischen zwei Impulsen dieser Nervenzelle tatsächlich einer Gammaverteilung mit den Parametern $\alpha = 0.803072$ und $\lambda = 0.235555$ genügt. Wir verwenden dazu den Score Test (später werden wir Methoden kennen lernen, mit denen diese Überprüfung auch rechnerisch erfolgen kann)

```
ScoreTest[nerven, GammaDistribution[α, λ], PlotStyle → PointSize[0.01], ImageSize → {150, 100}]
```



und erkennen an Hand dieser Zeichnung (der Punktschwarm liegt weitgehend auf einer Geraden), dass nichts gegen die Annahme spricht, dass die Länge X der Intervalle zwischen zwei aufeinander folgenden Impulsen dieser Nervenzelle mit den Parametern $\alpha = 0.803072$ und $\lambda = 0.235555$ gammaverteilt ist.

```
Clear[nerven, n, a, b, α, λ]
```

5.4.11 Beispiel: Die Lebensdauer von Transistoren ist bekanntlich annähernd $\mathcal{E}[\lambda]$-verteilt. Zur Bestimmung des unbekannten Parameters λ sollte von $n = 60$ Transistoren die Lebensdauer ermittelt werden. Nach $T = 600$ Stunden wurde der Versuch abgebrochen. Das Datenmaterial *transistor* enthält die Lebensdauern der bis dahin ausgefallenen $m = 48$ Transistoren. Man verwende dieses Datenmaterial (man spricht von zensurierten Daten) dazu, den unbekannten Parameter λ zu schätzen.

▼

Lösung: Es bezeichne X die $\mathcal{E}[\lambda]$-verteilte Lebensdauer der Transistoren.

a) Ignoriert man die Tatsache, dass zum Zeitpunkt T noch $n - m$ Transistoren funktionieren, so läuft unser Problem auf die folgende Fragestellung hinaus: Gegeben ist eine $\mathcal{E}[\lambda]$-verteilte Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_m\}$ vom Umfang m. Gesucht ist ein Schätzwert für den Parameter λ. Dieses Problem haben wir bereits behandelt und dafür den Maximum-Likelihood-Schätzer $S_1 = (\overline{X})^{-1}$ gefunden. Als Schätzwert für den unbekannten Para-meter λ erhält man damit wegen

```
Mean[Rest[<< transistorfile]]^-1
```

```
0.00324186
```

den Wert $\hat{\lambda}_1 = 0.00324186$.

b) In der Tatsache, dass zum Zeitpunkt T noch $n - m$ Transistoren einwandfrei funktionieren, ist aber auch eine gewisse Information über den Parameter λ enthalten, die genützt werden sollte:

Bei der von uns festgestellten Lebensdauer handelt es sich nämlich nicht um die Lebensdauer X dieser Transistoren schlechthin, sondern um die beobachtete Lebensdauer $Y = \mathrm{Min}[T, X]$. Diese beobachtete Lebensdauer Y besitzt offenbar die Verteilungsfunktion

$$\mathbb{F}_Y[x] = \begin{cases} 0 & \text{für } x < 0 \\ 1 - e^{-\lambda x} & \text{für } 0 \le x < T \\ 1 & \text{für } x \ge T \end{cases}$$

Wir haben nun die Aufgabe, auf der Basis unserer nach dem Verteilungsgesetz $\mathbb{P}_Y$ ausgewählten Stichprobe

$$\vec{y} = \{x_1, x_2, \ldots, x_m, T, T, \ldots, T\}$$

vom Umfang n, den Parameter λ zu schätzen. Wir verwenden dazu die Maximum-Likelihood-Idee, bestimmen also den Parameter λ so, dass die Wahrscheinlichkeit

$$L[\{x_1, x_2, \ldots, x_m, T, T, \ldots, T\} \mid \lambda] = \lambda^m \,\mathrm{Exp}[-\lambda \sum_{i=1}^{m} x_i]\, \mathrm{Exp}[-\lambda\, T]^{n-m}$$

für das Auftreten dieser Stichprobe maximal wird. Wir setzen dazu die Ableitung des Logarithmus der Likelihood-Funktion $L[x_1, x_2, \ldots, x_m, T, T, \ldots, T \mid \lambda]$ gleich 0 und erhalten auf diese Weise den Schätzer

$$S_2 = \frac{m}{X_1 + X_2 + \ldots + X_m + (n - m)\, T}$$

Speziell ergibt sich für unser Datenmaterial wegen

```
n = 60; m = 48; T = 600;
m/(Apply[Plus, Rest[<< transistorfile]] + (n - m) T)
Clear[n, m, T]
```

```
0.0022122
```

der Schätzwert $\hat{\lambda}_2 = 0.0022122$. Die Berücksichtigung der Tatsache, dass von $n = 60$ Transistoren zum Zeitpunkt $T = 600$ noch $n - m = 12$ Transistoren einwandfrei arbeiten, hat also eine deutliche Verringerung des Schätzwertes des unbekannten Parameters λ zur Folge.

5.4.12 Beispiel: Die Anzahl X der Eier in einem Nest kann in erster Näherung als $\mathcal{P}[\lambda]$-verteilt angenommen werden, wobei der Parameter λ unbekannt ist. Ein Nest kann aber nur dann als solches identifiziert werden, wenn es mindestens ein Ei enthält. Das Datenmaterial *eier* enthält die Anzahl der Eier in $n = 92$ zufällig gefundenen Nestern. Man verwende dieses Datenmaterial dazu, den unbekannten Parameter λ zu schätzen.

▼

Lösung: Es bezeichne X die $\mathcal{P}[\lambda]$-verteilte Anzahl von Eiern ein einem Nest. Bei der uns vorliegenden Stichprobe $\vec{y} = \{y_1, y_2, \ldots, y_n\}$ handelt es sich aber nicht! um eine nach dem Verteilungsgesetz $\mathbb{P}_X = \mathcal{P}[\lambda]$ ausgewählte Stichprobe, sondern um eine nach dem Verteilungsgesetz $\mathbb{P}_Y$ ausgewählte Stichprobe, wobei Y die Anzahl der Eier in einem identifizierbaren Nest bezeichnet und offenbar die Verteilungsdichte

$$\mathbb{f}_Y[y] = \begin{cases} \mathbb{P}[\{X = y\} \mid \{X > 0\}] = \dfrac{e^{-\lambda}\,\lambda^y}{(1 - e^{-\lambda})\, y!} & \text{für } y \in \mathbb{N} \\ 0 & \text{sonst} \end{cases}$$

besitzt. Wir haben die Aufgabe, den Parameter λ so zu bestimmen, dass die Wahrscheinlichkeit

$$L[\{y_1, y_2, \ldots, y_n\} \mid \lambda] = \prod_{i=1}^{n} \mathbb{f}_Y[y_i] = \frac{e^{-n\lambda}\,\lambda^{y_1 + y_2 + \cdots + y_n}}{(1 - e^{-\lambda})^n\, y_1!\, y_2! \ldots y_n!}$$

für das Auftreten dieser Stichprobe $\{y_1, y_2, \ldots, y_n\}$ maximal ist. Wir lösen diese Aufgabe unter Verwendung des Befehls FindMaximum mit Hilfe von *Mathematica* (wobei wir den Faktor $y_1!\, y_2! \ldots y_n!$ weglassen und den weitgehend willkürlichen Startwert $\{\lambda, 2, 4\}$ verwenden)

```
eier = Rest[<< eierfile];
n = Length[eier];
a = Apply[Plus, eier];
FindMaximum[(ⅇ^(-n λ) λ^a/((1 - ⅇ^(-λ))^n)), {λ, 2, 4}]
Clear[eier, n, a]
```

```
{1.04984×10^34, {λ → 3.32124}}
```

und erhalten als Schätzwert für den unbekannten Parameter λ den Wert $\hat{\lambda} = 3.32124$.

5.4.13 Beispiel: Wir betrachten das folgende Modell für die Abhängigkeit der Krebssterbewahrscheinlichkeit vom Alter: Für einen Mann der Altersklasse K_i (das entspricht einem Alter zwischen $25 + 5\,i$ und $30 + 5\,i$ Jahren) beträgt die Wahrscheinlichkeit an Krebs zu sterben

$$p_i = \mathrm{Exp}[\alpha + i\,\beta] / (1 + \mathrm{Exp}[\alpha + i\,\beta])$$

Das Datenmaterial *krebs* enthält für eine europäische Großstadt die Anzahl n_i der Männer sowie die Anzahl x_i der Krebssterbefälle in der Altersklasse K_i im Jahr 1986. Man schätze die unbekannten Parameter α und β.

▼

Das Prinzip der Maximum-Likelihood-Methode besteht darin, die Parameter $\vartheta_1, \vartheta_2, \ldots, \vartheta_k$ so zu bestimmen, dass die Wahrscheinlichkeit dafür, mit diesen Parametern die Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ zu erhalten, maximal ist. Dieses Prinzip kann aber auch dann zum Schätzen von Parametern verwendet werden, wenn die Zufallsvariablen $X_1, X_2, \ldots, X_n$, von denen eine Stichprobe $\{x_1, x_2, \ldots, x_n\}$ vorliegt, nicht unabhängig sind. Wir erläutern diese Vorgangsweise an folgendem Beispiel:

▼

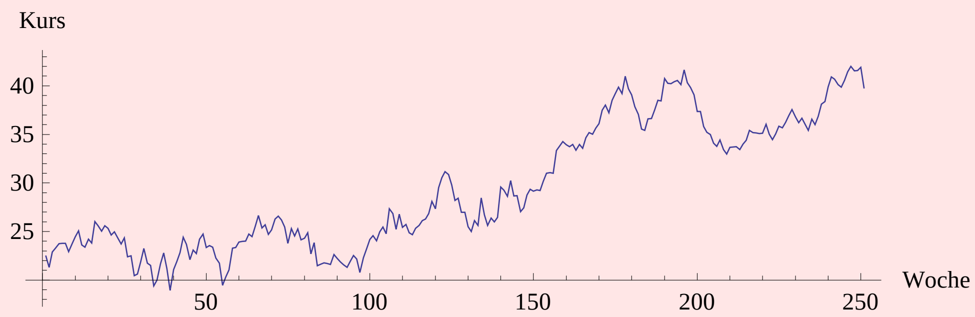
Achtung! Wir haben vorausgesetzt, dass es sich bei einer mathematischen Stichprobe $X_1, X_2, \ldots, X_n$ stets um vollständig unabhängige, identisch verteilte Zufallsvariablen handelt. Diese Voraussetzung ist auch weiterhin gültig. Lediglich bei dem folgenden Beispiel machen wir davon eine Ausnahme.

5.4.14 Beispiel: Der zufällige Verlauf des Kurses einer Aktie lässt sich in erster Näherung durch einen sogenannten autoregressiven Prozess modellieren. Darunter versteht man eine Folge $(X_i)_{i\in\mathbb{N}}$ von Zufallsvariablen, welche dem Bildungsgesetz $X_i = X_{i-1} + U_i$ mit konstantem $X_0 = x_0$ und vollständig unabhängigen und identisch $\mathcal{N}[\mu, \sigma]$-verteilten Zuwächsen $U_1, U_2, \ldots$ genügen. Unser Datenmaterial *aktienkurs* enthält die Schlusskurse einer Aktie über einen Bereich von $n = 250$ Wochen. Man veranschauliche dieses Datenmaterial zuerst graphisch und verwende dieses Datenmaterial anschließend dazu, die beiden unbekannten Parameter μ und σ zu schätzen.

▼

Lösung: a) Das Datenmaterial *aktienkurs* enthält die beiden Variablen Woche und Kurs. Wir rufen die zur Variablen Kurs (zweite Spalte) gehörige Stichprobe auf und veranschaulichen diese Stichprobe mit Hilfe von ListPlot graphisch:

```
aktienkurs = Rest[Part[<< aktienkursfile, All, 2]];
ListPlot[aktienkurs, Joined → True, AxesLabel → {"Woche", "Kurs"}, AspectRatio → 0.3]
```



b) Für den zufälligen Verlauf $X_1, X_2, \ldots$ des Kurses dieser Aktie gilt auf Grund unserer Modellbildung

$$\mathbb{P}_{X_1} = \mathcal{N}[x_0 + \mu, \sigma]$$
$$\mathbb{P}_{X_2 \mid \{X_1 = x_1\}} = \mathcal{N}[x_1 + \mu, \sigma]$$
$$\mathbb{P}_{X_3 \mid \{X_2 = x_2\} \cap \{X_1 = x_1\}} = \mathbb{P}_{X_3 \mid \{X_2 = x_2\}} = \mathcal{N}[x_2 + \mu, \sigma]$$
$$\mathbb{P}_{X_4 \mid \{X_3 = x_3\} \cap \{X_2 = x_2\} \cap \{X_1 = x_1\}} = \mathbb{P}_{X_4 \mid \{X_3 = x_3\}} = \mathcal{N}[x_3 + \mu, \sigma]$$
$$\ldots$$

Zum Schätzen der beiden unbekannten Parameter μ und σ verwenden wir die Idee der Maximum-Likelihood-Methode; wir bestimmen also diese beiden Parameter so, dass die Wahrscheinlichkeit für das Auftreten der Stichprobe $\vec{x} = \{x_1, x_2, \ldots, x_n\}$ maximal ist. Wegen des Multiplikationssatzes gilt für diese Wahrscheinlichkeit

$$L[\{x_1, x_2, \ldots, x_n\} \mid \mu, \sigma] = \mathbb{f}_{X_1}[x_1]\, \mathbb{f}_{X_2 \mid \{X_1 = x_1\}}[x_2] \ldots \mathbb{f}_{X_n \mid \{X_{n-1} = x_{n-1}\} \cap \ldots \cap \{X_1 = x_1\}}[x_n] =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \operatorname{Exp}[-\frac{(x_1 - x_0 - \mu)^2}{2\sigma^2}] \frac{1}{\sqrt{2\pi}\,\sigma} \operatorname{Exp}[-\frac{(x_2 - x_1 - \mu)^2}{2\sigma^2}]$$

$$\ldots \frac{1}{\sqrt{2\pi}\,\sigma} \operatorname{Exp}[-\frac{(x_n - x_{n-1} - \mu)^2}{2\sigma^2}] = \mathbb{f}_{\mu,\sigma}[x_1 - x_0]\, \mathbb{f}_{\mu,\sigma}[x_2 - x_1] \ldots \mathbb{f}_{\mu,\sigma}[x_n - x_{n-1}]$$

wobei wir mit $\mathbb{f}_{\mu,\sigma}$ die Verteilungsdichte der $\mathcal{N}[\mu, \sigma]$-Verteilung bezeichnet haben. Unter Verwendung von FindMaximum (und den weitgehend willkürlich gewählten Startwerten $\{\mu, 0, 1\}$ und $\{\sigma, 1, 2\}$) ermitteln wir das Maximum von $\operatorname{Log}[L[x_1, x_2, \ldots, x_n \mid \mu, \sigma]]$ und erhalten wegen

```
dif = Table[aktienkurs[[i + 1]] - aktienkurs[[i]], {i, 1, Length[aktienkurs] - 1}];
FindMaximum[Apply[Plus, Table[Log[PDF[NormalDistribution[μ, σ], dif[[i]]]], {i, 1, Length[dif]}]],
  {μ, 0, 1}, {σ, 1, 2}]
Clear[aktienkurs, dif]
```

```
{-356.167, {μ → 0.0693847, σ → 1.00575}}
```

für die beiden unbekannten Parameter μ und σ die Schätzwerte $\hat{\mu} = 0.0693847$ und $\hat{\sigma} = 1.00575$.