

Inhaltsverzeichnis

2	Likelihood	1
3	Frequentistische Eigenschaften der Likelihood	11
4	Likelihood-Inferenz bei vektorielltem Parameter	29
5	Bayes-Inferenz	49
6	Numerische Methoden zur Bayes-Inferenz	61
7	Modellwahl	77
8	Prognose	87
	Literaturverzeichnis	95
	Index	97

2 Likelihood

1. Formulieren Sie für die folgenden Problemstellungen jeweils die Likelihood.

- a) In einer Untersuchung eines wurzelinfizierenden Pilzes bei Getreide wird eine Saat von 250 Samen gepflanzt. Aus technischen Gründen kann nur beobachtet werden, dass $x \leq 25$ Samen gekeimt sind. Sei θ die Wahrscheinlichkeit, dass ein Samen keimt. Geben Sie einen Ausdruck für die Likelihood von θ basierend auf der Information des obigen Experiments an.

► *Da die Samen unabhängig voneinander mit Wahrscheinlichkeit θ keimen, ist die Anzahl der insgesamt gekeimten Samen binomialverteilt, $X \sim \text{Bin}(250, \theta)$. Die Likelihood von θ ist demnach die Verteilungsfunktion von X an $x = 25$,*

$$\begin{aligned} L(\theta) &= P(X \leq 25) \\ &= \sum_{x=0}^{25} \binom{250}{x} \theta^x (1-\theta)^{250-x}. \end{aligned}$$

- b) Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer $N(\theta, 1)$ -Verteilung. Es wird aber nur der größte Wert der Stichprobe, $Y = \max(X_1, \dots, X_n)$, gemeldet. Zeigen Sie, dass die Dichte von Y folgende Form hat:

$$f_Y(y) = n (\Phi(y - \theta))^{n-1} \phi(y - \theta),$$

wobei $\Phi(\cdot)$ die Verteilungsfunktion und $\phi(\cdot)$ die Dichte der Standardnormalverteilung ist. Wie lautet somit die Likelihoodfunktion $L(\theta)$? *Hinweis:* Bestimmen Sie zuerst die Verteilungsfunktion von Y .

► *Gegeben sind also $X_1, \dots, X_n \stackrel{iid}{\sim} N(\theta, 1)$ und die Transformation $Y = \max(X_1, \dots, X_n)$, deren Verteilungsfunktion die Form*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\max(X_1, \dots, X_n) \leq y) \\ &= \prod_{i=1}^n P(X_i \leq y) \\ &= (\Phi(y - \theta))^n \end{aligned}$$

hat, da $X_i \leq y$ äquivalent zu $X_i - \theta \leq y - \theta$ ist und $X_i - \theta$ für alle $i = 1, \dots, n$ einer Standardnormalverteilung genügt. Die Dichte von Y ergibt sich daher zu

$$f_Y(y) = \frac{d}{dy} F_Y(y) = n (\Phi(y - \theta))^{n-1} \phi(y - \theta),$$

und die Likelihood ist genau die Dichte, aufgefasst als Funktion von θ :

$$L(\theta) = n(\Phi(y - \theta))^{n-1}\phi(y - \theta)$$

- c) Sei X_1, X_2, X_3 eine Zufallsstichprobe aus einer $C(\theta, 1)$ -Verteilung, wobei $\theta \in \mathbb{R}$ der Lageparameter der Cauchyverteilung mit Dichte

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad x \in \mathbb{R}$$

ist. Bestimmen Sie einen Ausdruck für die Likelihood von θ .

► Wiederum liegt eine iid-Stichprobe vor, sodass sich die Likelihood von θ als Produkt der Dichten $f(x_i)$ ergibt:

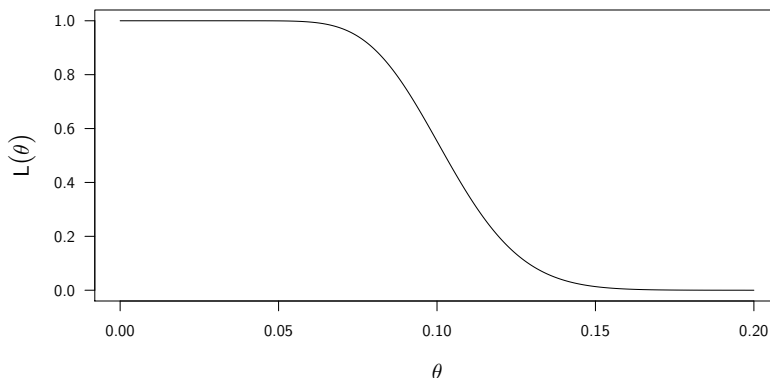
$$\begin{aligned} L(\theta) &= \prod_{i=1}^3 \frac{1}{\pi} \cdot \frac{1}{1 + (x_i - \theta)^2} \\ &= \frac{1}{\pi^3} \frac{1}{(1 + (x_1 - \theta)^2)(1 + (x_2 - \theta)^2)(1 + (x_3 - \theta)^2)} \end{aligned}$$

- d) Erstellen Sie in R einen Plot der Likelihood für

- i. $L(\theta)$ in 1a)

► Bei unserer Lösung können wir die in R bereits implementierte Verteilungsfunktion der Binomialverteilung, `pbinom`, nutzen:

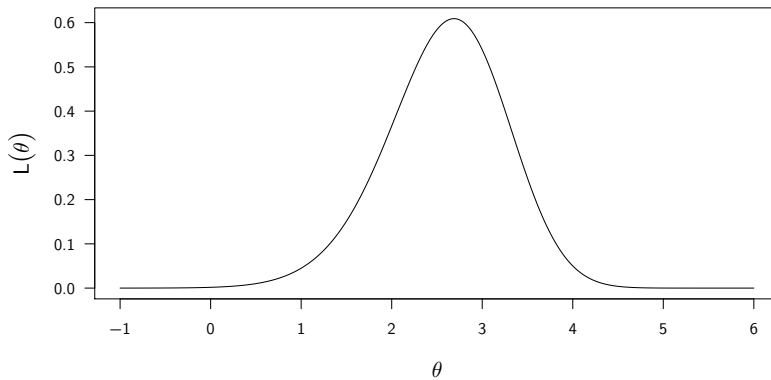
```
> likelihood1 <- function( # Likelihood für Wkeit im teilweise beobachteten
+                           # Binomialexperiment
+                           theta,      # Wkeit
+                           n,          # Stichprobenumfang
+                           x           # Anzahl in Intervall [0, x] wurde beobachtet
+ )
+ {
+   pbinom(x, size = n, p = theta) # eingebaute Verteilungsfunktion nutzen
+ }
> theta <- seq(0, 0.2,
+             length=1000
+             )
> plot(theta, likelihood1(theta, n = 250, x = 25),
+      type="l",
+      xlab = expression(theta), ylab = expression(L(theta)))
```



ii. $L(\theta)$ in 1b) bei Beobachtungen $x = (1.5, 0.25, 3.75, 3.0, 2.5)$.

► Auch hier können wir die implementierten Funktionen der Normalverteilung benutzen:

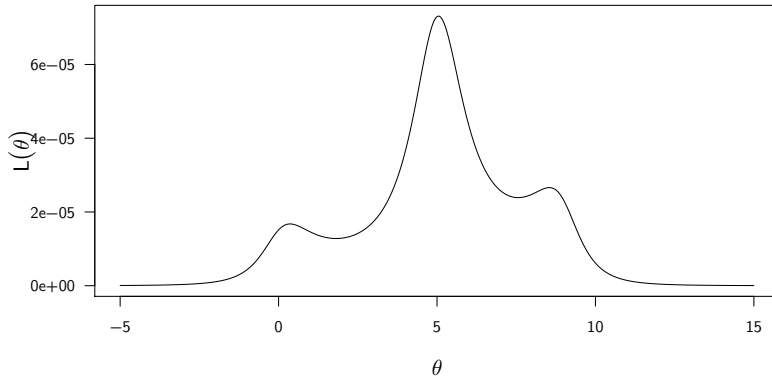
```
> likelihood2 <- function(      # Likelihood für Ewert bei beobachtetem NV-Maximum
+                             theta,      # Ewert der iid-Normalverteilungen
+                             y,         # beobachtetes Maximum
+                             n          # Anzahl der ZVn
+                             )
+ {
+   n * pnorm(y - theta)^(n-1) * dnorm(y - theta)
+ }
> x <- c(1.5, 0.25, 3.75, 3.0, 2.5)
> theta <- seq(-1, 6, length = 1000)
> plot(theta, likelihood2(theta, y = max(x), n = length(x)),
+       type = "l",
+       xlab = expression(theta), ylab = expression(L(theta)))
```



iii. $L(\theta)$ für 1c) bei Beobachtungen $x = (0, 5, 9)$.

► Die vektorwertige Berechnung von R wird hier deutlich:

```
> likelihood3 <- function(      # Likelihood für Lageparameter bei iid-Cauchy-Daten
+                             theta,      # Lageparameter
+                             x          # beobachteter Datenvektor
+                             )
+ {
+   1/pi^3 / sapply(theta,
+                   function(theta) prod((1+(x-theta)^2))
+                   )
+ }
> x <- c(0, 5, 9)
> theta <- seq(-5, 15, length = 1000)
> plot(theta, likelihood3(theta, x = x),
+       type = "l",
+       xlab = expression(theta), ylab = expression(L(theta)))
```



2. Ein autoregressiver Prozess 1. Ordnung X_0, X_1, \dots, X_n ist durch die bedingten Verteilungen

$$X_i | X_{i-1}, \dots, X_0 \sim N(\alpha X_{i-1}, 1), \quad i = 1, 2, \dots, n,$$

und $X_0 \sim N(0, 1)$ gegeben.

- a) Zeigen Sie, dass der Kern der Log-Likelihood folgende Form hat:

$$l(\alpha) = -\frac{1}{2} \sum_{i=1}^n (x_i - \alpha x_{i-1})^2.$$

► Die Likelihood ist gegeben durch

$$\begin{aligned} L(\alpha) &= f(x_0, \dots, x_n; \alpha) \\ &= f(x_n | x_{n-1}, \dots, x_0; \alpha) f(x_{n-1} | x_{n-2}, \dots, x_0; \alpha) \cdots f(x_1 | x_0; \alpha) f(x_0; \alpha) \\ &= f(x_0; \alpha) \cdot \prod_{i=1}^n f(x_i | x_{i-1}, \dots, x_0; \alpha) \\ &\propto \prod_{i=1}^n \exp \left\{ -\frac{1}{2} (x_i - \alpha x_{i-1})^2 \right\}, \end{aligned}$$

da $f(x_0; \alpha) = \phi(x_0)$ unabhängig von α ist. Die Log-Likelihood ist also

$$l(\alpha) = \log L(\alpha) = -\frac{1}{2} \sum_{i=1}^n (x_i - \alpha x_{i-1})^2.$$

- b) Bestimmen Sie $\hat{\alpha}_{ML}$ und überprüfen Sie dabei, dass es sich wirklich um ein Maximum von $l(\alpha)$ handelt. Bestimmen Sie auch die beobachtete Fisher-Information.

► Wir stellen zunächst die Score-Funktion für α auf,

$$\begin{aligned} S(\alpha) &= \frac{d}{d\alpha} l(\alpha) = -\frac{1}{2} \sum_{i=1}^n 2(x_i - \alpha x_{i-1}) \cdot (-x_{i-1}) \\ &= \sum_{i=1}^n x_i x_{i-1} - \alpha x_{i-1}^2 \\ &= \sum_{i=1}^n x_i x_{i-1} - \alpha \sum_{i=1}^n x_{i-1}^2, \end{aligned}$$

sodass die Score-Gleichung $S(\alpha) = 0$ durch

$$\hat{\alpha}_{ML} = \frac{\sum_{i=1}^n x_i x_{i-1}}{\sum_{i=1}^n x_{i-1}^2}$$

gelöst wird. Dies ist wirklich ein lokales Maximum der Log-Likelihood, da diese konkav ist, wie man durch

$$\frac{d}{d\alpha} S(\alpha) = -\sum_{i=1}^n x_{i-1}^2 < 0$$

leicht sieht. Die Fisher-Information

$$I(\alpha) = -\frac{d}{d\alpha} S(\alpha) = \sum_{i=1}^n x_{i-1}^2 > 0$$

ist hier unabhängig vom Parameter und somit gleich der beobachteten Fisher-Information $I(\hat{\alpha}_{ML})$. Da es keine anderen Maxima oder Parameterbeschränkungen gibt ($\alpha \in \mathbb{R}$), ist $\hat{\alpha}_{ML}$ wirklich das globale Maximum von $l(\alpha)$.

- c) Erstellen Sie in R einen Plot von $l(\alpha)$ und berechnen Sie $\hat{\alpha}_{ML}$ bei den folgenden Daten:

$$(x_0, \dots, x_6) = (-0.560, -0.510, 1.304, 0.722, 0.490, 1.960, 1.441).$$

► Im R-Code entspricht nun $x[1]$ dem x_0 aus dem Text, da die Indizierung erst bei 1 beginnt.

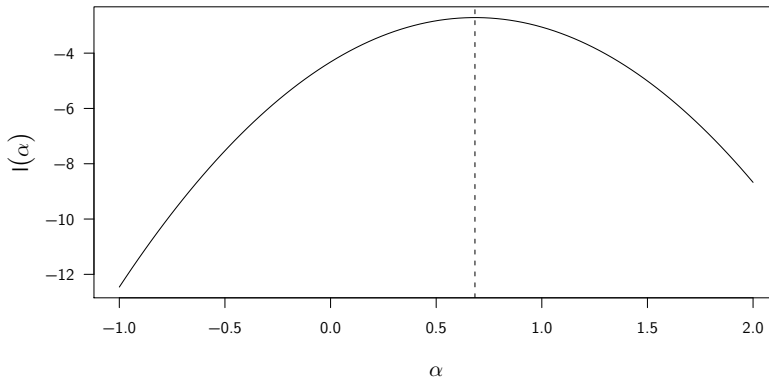
```
> ## Log-Likelihood implementieren
> loglik <- function(                                     # Log-Likelihood im AR(1)-Prozess
+                                                             # mit standardnormalverteilten Störgrößen
+               alpha,                                     # Parameter für den Term erster Ordnung
+               x                                           # beobachtete Daten
+             )
+ {
+   i <- 2:length(x)
+   - 1/2 * sum((x[i] - alpha * x[i-1])^2)
+ }
```



```

> ## und plotten für gegebene Daten
> x <- c(-0.560, -0.510, 1.304, 0.722, 0.490, 1.960, 1.441)
> alpha <- seq(-1, 2, length = 100)
> plot(alpha, sapply(alpha, function(alpha) loglik(alpha, x)),
+       type = "l",
+       xlab = expression(alpha), ylab = expression(l(alpha)))
> ## dann ML-Schätzer berechnen und einzeichnen
> i <- seq(along = x)[-1] # wieder Index-Vektor nötig (wie in loglik)
> alphaML <- sum(x[i] * x[i-1]) / sum(x[i-1]^2)
> alphaML
[1] 0.6835131
> abline(v = alphaML, lty = 2)

```



3. Zeigen Sie, dass im Beispiel 2.2 die Likelihoodfunktion $L(N)$ an der Stelle $\hat{N} = \lfloor \frac{M-n}{x} \rfloor$ maximiert wird, hierbei bezeichnet $\lfloor x \rfloor$ die abgerundete Zahl x . Studieren Sie hierzu Monotonieeigenschaften des Quotienten $L(N)/L(N-1)$. In welchen Fällen ist der Maximum-Likelihood-Schätzer nicht eindeutig? Geben Sie ein Zahlenbeispiel an.

► Die Likelihoodfunktion ist für $N \in \Theta = \{\max(n, M+n-x), \max(n, M+n-x) + 1, \dots\}$

$$L(N) \propto \frac{\binom{N-M}{n-x}}{\binom{N}{n}}.$$

Der Quotient $R(N) = L(N)/L(N-1)$ ist in diesem Bereich

$$\begin{aligned}
 R(N) &= \frac{\binom{N-M}{n-x}}{\binom{N}{n}} \cdot \frac{\binom{N-1}{n}}{\binom{N-1-M}{n-x}} \\
 &= \frac{(N-M)!n!(N-n)!}{(n-x)!(N-M-n+x)!N!} \cdot \frac{(N-1)!(n-x)!(N-1-M-n+x)!}{n!(N-1-n)!(N-1-M)!} \\
 &= \frac{(N-M)(N-n)}{(N-M-n+x)N}.
 \end{aligned}$$

Für den ML-Schätzer \hat{N}_{ML} muss als lokales Maximum von $L(N)$ sowohl $L(\hat{N}_{ML}) \geq L(\hat{N}_{ML}-1)$, also $R(\hat{N}_{ML}) \geq 1$, als auch $L(\hat{N}_{ML}) \geq L(\hat{N}_{ML}+1)$, also $R(\hat{N}_{ML}+1) \leq 1$, gelten.

Aus obiger Umformung erhält man leicht die Äquivalenz von $R(N) \geq 1$ und $N \leq Mn/x$, aus der wiederum die Äquivalenz von $R(N+1) \leq 1$ und $N+1 \geq Mn/x$ bzw. $N \geq Mn/x - 1$ folgt. Der ML-Schätzer ist also jede ganze Zahl im Intervall $[Mn/x - 1, Mn/x]$. Falls die rechte Grenze Mn/x bereits ganzzahlig ist, gibt es somit zwei Lösungen und der ML-Schätzer ist nicht eindeutig. Hätte im Zahlenbeispiel aus Abbildung 2.2 der Stichprobenumfang nicht $n = 63$ sondern $n = 65$ betragen, hätten sowohl $26 \cdot 65/5 = 1690$ als auch 1689 die höchste Likelihood gehabt. Falls Mn/x nicht ganzzahlig ist, ist dagegen $\hat{N}_{ML} = \lfloor Mn/x \rfloor$ eindeutig.

4. Berechnen Sie den ML-Schätzer von π bei Vorliegen einer geometrisch verteilten Beobachtung $X \sim \text{Geom}(\pi)$. Wie lautet der ML-Schätzer von π , wenn eine Zufallsstichprobe X_1, \dots, X_n aus einer geometrischen Verteilung vorliegt?

► Sei zunächst $X \sim \text{Geom}(\pi)$. Aus der Wahrscheinlichkeitsfunktion

$$f(x; \pi) = \pi(1 - \pi)^{x-1} \quad (x \in \mathbb{N})$$

erhalten wir die Log-Likelihood-Funktion

$$l(\pi) = \log(\pi) + (x - 1) \log(1 - \pi)$$

und die Score-Funktion

$$S(\pi) = \frac{1}{\pi} + \frac{x - 1}{\pi - 1}.$$

Die Lösung der Score-Gleichung $S(\pi) = 0$ errechnet man über die Umformungen

$$\begin{aligned} \frac{1}{\pi} &= \frac{1 - x}{\pi - 1} \\ \pi - 1 &= \pi(1 - x) \\ \pi \cdot x &= 1 \end{aligned}$$

zu $\hat{\pi}_{ML} = 1/x$.

Liegt nun eine Zufallsstichprobe $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Geom}(\pi)$ vor, so addieren sich die eben berechneten Score-Beiträge der x_i zu

$$S(\pi) = \frac{n}{\pi} + \frac{n(\bar{x} - 1)}{\pi - 1} = n \left\{ \frac{1}{\pi} + \frac{\bar{x} - 1}{\pi - 1} \right\},$$

wobei \bar{x} das arithmetische Mittel $\sum_{i=1}^n x_i/n$ bezeichnet. Man sieht, dass die Score-Gleichung analog zum Fall mit einer Beobachtung gelöst wird, wobei nur x durch \bar{x} zu ersetzen ist. Der ML-Schätzer ist also nun $\hat{\pi}_{ML} = 1/\bar{x}$.

5. Man zeige, dass in Beispiel 2.15 $T(X) = \max_i(X_i)$ suffizient für θ ist.

► In Beispiel 2.15 wurde hergeleitet, dass die Likelihood von θ bzw. die Dichte der beobachteten Daten an θ die Form

$$f(x; \theta) = \begin{cases} \prod_{i=1}^n f(x_i; \theta) = \frac{1}{\theta^n} & \text{für } \theta \geq \max_i(x_i), \\ 0 & \text{sonst} \end{cases}$$

hat. Dies können wir umschreiben zu

$$f(x; \theta) = \frac{1}{\theta^n} \mathbb{I}_{[0, \theta]}(\max_i(x_i)),$$

wobei die Indikatorfunktion $\mathbb{I}_A(x)$ angibt, ob x in A liegt, d. h. für $x \in A$ gleich 1 und sonst gleich 0 ist. Dies zeigt nach dem Faktorisierungssatz 2.2 mit $g(T(x); \theta) = \frac{1}{\theta^n} \mathbb{I}_{[0, \theta]}(T(x))$ und $h(x) = 1$ die Suffizienz von $T(X) = \max_i(X_i)$ für θ .

6. a) Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit Dichte

$$f_{X_i}(x; \theta) = \begin{cases} e^{i\theta - x} & x \geq i\theta \\ 0 & x < i\theta. \end{cases}$$

Zeigen Sie, dass $T(X) = \min_i(X_i/i)$ eine suffiziente Statistik für θ ist.

► Wegen der Äquivalenz von $x \geq i\theta$ und $x/i \geq \theta$ können wir die Dichte der i -ten Beobachtung als

$$f_{X_i}(x_i) = \exp(i\theta - x_i) \mathbb{I}_{[\theta, \infty)}(x_i/i)$$

schreiben. Die Dichte der gesamten Beobachtung $x = (x_1, \dots, x_n)$ lässt sich aufgrund der Unabhängigkeit von X_1, \dots, X_n als Produkt

$$\begin{aligned} f(x) &= \prod_{i=1}^n f_{X_i}(x_i) \\ &= \exp \left\{ \theta \left(\sum_{i=1}^n i \right) - n\bar{x} \right\} \prod_{i=1}^n \mathbb{I}_{[\theta, \infty)}(x_i) \\ &= h(x) \cdot g(T(x); \theta) \end{aligned}$$

mit $h(x) = \exp(-n\bar{x})$, $T(x) = \min_i(x_i/i)$ und

$$g(T(x); \theta) = \exp(\theta n(n+1)/2) \mathbb{I}_{[\theta, \infty)}(T(x))$$

darstellen. Mit Satz 2.2 ist $\min_i(X_i/i)$ also suffizient für θ .

b) Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit Dichte

$$f(x; \theta) = e^{-(x-\theta)}, \quad \theta < x < \infty, \quad -\infty < \theta < \infty.$$

Bestimmen Sie eine minimal suffiziente Statistik für θ .

► Im Wesentlichen liegt also nun eine Zufallsstichprobe aus der X_1 -Verteilung in 6a) vor. Man sieht damit sofort, dass $T(X) = \min_i(X_i)$ eine suffiziente Statistik für θ ist. Da die Likelihood nicht weiter vereinfacht werden kann und von den Daten nur über $T(x)$ abhängt, sowie der Minimalsuffizienz der Likelihood, folgt dann die Minimalsuffizienz von $T(X)$ für θ .

7. Sei $T(X)$ eine suffiziente Statistik für θ , $r(\cdot)$ eine eindeutige Funktion und $T^*(X) = r(T(X))$. Zeigen Sie, dass $T^*(X)$ suffizient für θ ist.

► Nach Satz 2.2 folgt aus der Suffizienz von $T(X)$ für θ , dass es Funktionen g und h gibt, sodass

$$f(x; \theta) = g(T(x); \theta) \cdot h(x)$$

ist. Setzen wir nun $\tilde{g} := g \circ r^{-1}$, so können wir

$$f(x; \theta) = g(r^{-1}(r(T(x))); \theta) \cdot h(x) = \tilde{g}(T^*(x); \theta) \cdot h(x)$$

schreiben, was hinreichend für die Suffizienz von $T^*(X)$ für θ ist.

8. Seien $X_1 \sim \text{Exp}(\lambda)$ und $X_2 \sim \text{Exp}(\lambda)$ unabhängige Zufallsvariablen, $\lambda > 0$. Zeigen Sie, dass $T(X_1, X_2) = X_1 + X_2$ suffizient für λ ist.

► Die Likelihood von λ lässt sich faktorisieren in

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^2 \lambda \exp(-\lambda x_i) \\ &= \lambda^2 \exp(-\lambda(x_1 + x_2)) \\ &= g(T(x); \lambda) \cdot h(x) \end{aligned}$$

mit $g(t; \lambda) = \lambda^2 \exp(-\lambda t)$ und $h(x) = 1$. Mit dem Faktorisierungssatz 2.2 folgt die Suffizienz von $T(X) = X_1 + X_2$ für λ . Alternativ könnte man dies auch über die Definition zeigen, da $T(X) \sim G(2, \lambda)$ gilt.

3 Frequentistische Eigenschaften der Likelihood

1. Skizzieren Sie, warum der ML-Schätzer

$$\hat{N}_{ML} = \left\lfloor \frac{M \cdot n}{X} \right\rfloor$$

im Rückfangexperiment (vgl. Beispiel 2.2) nicht erwartungstreu sein kann. Zeigen Sie, dass der alternative Schätzer

$$\hat{N} = \frac{(M+1) \cdot (n+1)}{(X+1)} - 1$$

erwartungstreu ist, wenn $N \leq M + n$.

► Der ML-Schätzer

$$\hat{N}_{ML} = \left\lfloor \frac{M \cdot n}{X} \right\rfloor$$

ist für $N \geq n + M$ mit echt positiver Wahrscheinlichkeit gleich unendlich, da dann X mit echt positiver Wahrscheinlichkeit gleich Null sein kann. Folglich ist der Erwartungswert des ML-Schätzers in diesem Fällen gleich unendlich, und somit ungleich dem wahren Parameterwert N . Zusammenfassend ist der ML-Schätzer nicht erwartungstreu, da dessen Erwartungswert für alle $N \geq n + M$ nicht gleich dem unbekannten Parameter N ist.

Der alternative Schätzer ist hingegen erwartungstreu für den Fall $N \leq M + n$, d. h. der Träger \mathcal{T} der hypergeometrischen Verteilung beginnt bei $\max\{0, n - (N - M)\} = n - (N - M)$. Der Erwartungswert des Bruches $g(X) = (M + 1)(n + 1)/(X + 1)$ ist nämlich

$$\begin{aligned} \mathbb{E}(g(X)) &= \sum_{x \in \mathcal{T}} g(x) \mathbb{P}(X = x) \\ &= \sum_{x=n-(N-M)}^{\min\{n, M\}} \frac{(M+1)(n+1)}{x+1} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \\ &= (N+1) \sum_{x=n-(N-M)}^{\min\{n, M\}} \frac{\binom{M+1}{x+1} \binom{(N+1)-(M+1)}{n-x}}{\binom{N+1}{n+1}}, \end{aligned}$$

durch geschicktes Zusammenfassen der Binomialkoeffizienten mit den Termen aus $g(x)$. Eine Indexverschiebung (bei Summanden Ersetzen von x durch $x-1$, bei den Indexgrenzen dagegen 1 dazuzugaddieren) zeigt nun, dass die Summe als Summe über den Träger einer $\text{HypGeom}(n+1, N+1, M+1)$ -Verteilung gleich 1 ist:

$$\begin{aligned}
 & \sum_{x=n-(N-M)}^{\min\{n,M\}} \frac{\binom{M+1}{x+1} \binom{(N+1)-(M+1)}{n-x}}{\binom{N+1}{n+1}} \\
 &= \sum_{x=(n+1)-((N+1)-(M+1))}^{\min\{n+1,M+1\}} \frac{\binom{M+1}{x} \binom{(N+1)-(M+1)}{(n+1)-x}}{\binom{N+1}{n+1}} \\
 &= \sum_{x \in T^*} \mathbb{P}(X^* = x) \\
 &= 1
 \end{aligned}$$

mit $X^* \sim \text{HypGeom}(n+1, N+1, M+1)$. Insgesamt also

$$\mathbb{E}(\hat{N}) = \mathbb{E}(g(X)) - 1 = N + 1 - 1 = N.$$

Der alternative Schätzer hinterlässt aber einen faden Beigeschmack, da er einerseits für $N > M + n$ ebenfalls nicht mehr erwartungstreu ist, andererseits im Allgemeinen nicht Element des Parameterraums der ganzen Zahlen ist. Natürlich könnte man ihn wieder geeignet auf- oder abrunden, dann würde er aber auch für $N \geq M + n$ die Eigenschaft der Erwartungstreue verlieren.

2. Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer Verteilung mit Erwartungswert μ und Varianz $\sigma^2 > 0$. Zeigen Sie, dass dann gilt:

$$\mathbb{E}(\bar{X}) = \mu \quad \text{und} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

► Das arithmetische Mittel \bar{X} ist erwartungstreu für den Erwartungswert μ , denn wegen der Linearität des Operators \mathbb{E} gilt

$$\mathbb{E}(\bar{X}) = n^{-1} \sum_{i=1}^n \mathbb{E}(X_i) = n^{-1} n \cdot \mu = \mu.$$

Die Varianz einer Summe von paarweise unkorrelierten Zufallsvariablen ist die Summe der Einzelvarianzen, sodass

$$\text{Var}(\bar{X}) = n^{-2} \sum_{i=1}^n \text{Var}(X_i) = n^{-2} n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

3. Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer Normalverteilung mit Erwartungswert μ und Varianz $\sigma^2 > 0$. Zeigen Sie, dass der Schätzer

$$\hat{\sigma} = \sqrt{\frac{n-1}{2} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}} S$$

erwartungstreu für σ ist, wobei S die Wurzel aus der Stichprobenvarianz S^2 in (3.1) ist.

► Seien $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$. Das bekannte Resultat

$$Y := \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

das z. B. in Davison (2003, S. 75) bewiesen wird, erlaubt nun die Bestimmung des Erwartungswertes von zunächst $\sqrt{Y} = \sqrt{n-1}S/\sigma$:

$$\begin{aligned} E(\sqrt{Y}) &= \int_0^\infty \left(\frac{1}{2}\right)^{\frac{n-1}{2}} \frac{y^{\frac{n-1}{2}-1} \exp(-y/2)}{\Gamma(\frac{n-1}{2})} y^{\frac{1}{2}} dy \\ &= \left(\frac{1}{2}\right)^{-\frac{1}{2}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \int_0^\infty \frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} \exp(-y/2) dy \\ &= \sqrt{2} \Gamma\left(\frac{n}{2}\right) / \Gamma\left(\frac{n-1}{2}\right), \end{aligned}$$

da sich das letzte Integral über die Dichte der $\chi^2(n)$ -Verteilung zu 1 ergibt. Somit

$$E(\hat{\sigma}) = E\left\{\sigma \frac{\sqrt{Y}}{\sqrt{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})}\right\} = \sigma.$$

4. Zeigen Sie, dass sich die Stichprobenvarianz S^2 zu

$$S^2 = \frac{1}{2n(n-1)} \sum_{i,j=1}^n (X_i - X_j)^2$$

umformen lässt. Verwenden Sie diese Umformung, um

$$\text{Var}(S^2) = \frac{1}{n} \left(\mu_4 - \left(\frac{n-3}{n-1}\right) \sigma^4 \right)$$

zu zeigen, wobei $\mu_4 = E\left\{(X - E(X))^4\right\}$ das 4. zentrale Moment von X bezeichnet.

► Zunächst lässt sich der Schätzer S^2 umschreiben zu $T := \frac{1}{2n(n-1)} \sum_{i,j=1}^n (X_i - X_j)^2$, denn

$$\begin{aligned} (n-1)T &= \frac{1}{2n} \cdot \sum_{i,j=1}^n (X_i^2 - 2X_iX_j + X_j^2) = \\ &= \frac{1}{2n} \cdot \left(n \sum_{i=1}^n X_i^2 - 2 \sum_{i=1}^n X_i \sum_{j=1}^n X_j + n \sum_{j=1}^n X_j^2 \right) = \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 = (n-1)S^2. \end{aligned}$$

Somit können wir die Varianz von S^2 über die paarweisen Kovarianzen aller Terme $(X_i - X_j)^2$, $i, j = 1, \dots, n$ berechnen:

$$\begin{aligned} \text{Var}(S^2) &= (2n(n-1))^{-2} \text{Var} \sum_{i,j} (X_i - X_j)^2 = \\ &= (2n(n-1))^{-2} \sum_{i,j,k,l} \text{Cov}((X_i - X_j)^2, (X_k - X_l)^2). \end{aligned} \quad (3.1)$$

Dazu unterscheiden wir drei Typen von Kovarianztermen in der Summe aus (3.1):

- $\text{Cov}((X_i - X_j)^2, (X_k - X_l)^2) = 0$, falls $i = j$ oder $k = l$ ist (denn dann ist das erste bzw. zweite Argument von Cov gleich Null) oder i, j, k, l paarweise verschieden sind (wegen Unabhängigkeit der X_i).
- Die Kovarianzen $\text{Cov}((X_i - X_j)^2, (X_i - X_j)^2)$ haben für $i \neq j$ den Wert $2\mu_4 + 2\sigma^4$. Dazu werden zwei Zwischenrechnungen benötigt; es sei $\mu := E(X_1)$.

$$\begin{aligned} E((X_i - X_j)^2) &= E((X_i - \mu)^2) + E((X_j - \mu)^2) - 2E((X_i - \mu)(X_j - \mu)) = \\ &= 2\sigma^2 - 2(E(X_i) - \mu)(E(X_j) - \mu) = 2\sigma^2 \end{aligned}$$

wegen der Unabhängigkeit von X_i und X_j , analog (durch Hinzufügen von μ und Ausmultiplizieren) zeigt man

$$\begin{aligned} E((X_i - X_j)^4) &= E((X_i - \mu + \mu - X_j)^4) = \\ &= E((X_i - \mu)^4) - 4E((X_i - \mu)^3(X_j - \mu)) + 6E((X_i - \mu)^2(X_j - \mu)^2) \\ &\quad - 4E((X_i - \mu)(X_j - \mu)^3) + E((X_j - \mu)^4) = \\ &= \mu_4 - 4 \cdot 0 + 6 \cdot (\sigma^2)^2 - 4 \cdot 0 + \mu_4 = 2\mu_4 + 6\sigma^4. \end{aligned}$$

Damit

$$\begin{aligned} \text{Cov}((X_i - X_j)^2, (X_i - X_j)^2) &= \text{Var}((X_i - X_j)^2) = \\ &= E((X_i - X_j)^4) - \{E((X_i - X_j)^2)\}^2 = \\ &= 2\mu_4 + 6\sigma^4 - (2\sigma^2)^2 = 2\mu_4 + 2\sigma^4. \end{aligned}$$

Da $(X_i - X_j)^2 = (X_j - X_i)^2$, gibt es $2 \cdot n(n-1)$ solcher Terme in (3.1).

- Analog lässt sich zeigen, dass $\text{Cov}((X_i - X_j)^2, (X_k - X_j)^2) = \mu_4 - \sigma^4$, falls i, j, k paarweise verschieden sind. Es gibt $n(n-1)(n-2)$ verschiedene Tupel (i, j, k) , für die i, j, k paarweise verschieden und Elemente der Menge $\{1, \dots, n\}$ sind. Zu jedem dieser Tupel gibt es zusätzlich zu obigem Kovarianzterm 3 weitere, verschiedene Kovarianzterme $\text{Cov}((X_i - X_j)^2, (X_j - X_k)^2)$, $\text{Cov}((X_j - X_i)^2, (X_j - X_k)^2)$ und $\text{Cov}((X_j - X_i)^2, (X_k - X_j)^2)$, die den gleichen Wert haben, sodass insgesamt $4 \cdot n(n-1)(n-2)$ solcher Terme in (3.1) vorkommen.

Insgesamt also

$$\begin{aligned} \text{Var}(S^2) &= \frac{1}{(2n(n-1))^2} \{2n(n-1)(2\mu_4 + 2\sigma^4) + 4n(n-1)(n-2)(\mu_4 - \sigma^4)\} = \\ &= \frac{1}{n(n-1)} \{\mu_4 + \sigma^4 + (n-2)(\mu_4 - \sigma^4)\} = \\ &= \frac{1}{n} \left(\mu_4 - \left(\frac{n-3}{n-1} \right) \sigma^4 \right). \end{aligned}$$

5. Zeigen Sie, dass das in Beispiel 3.5 definierte Konfidenzintervall tatsächlich Überdeckungswahrscheinlichkeit 50% für alle Werte $\theta \in \Theta$ besitzt.

► Sei $\theta \in \Theta = \mathbb{R}$. Die einfache Rechnung

$$\begin{aligned} P(\min\{X_1, X_2\} \leq \theta \leq \max\{X_1, X_2\}) &= P(X_1 \leq \theta \leq X_2) + P(X_2 \leq \theta \leq X_1) \\ &= P(X_1 \leq \theta) P(X_2 \geq \theta) + P(X_2 \leq \theta) P(X_1 \geq \theta) \\ &= 0.5 \cdot 0.5 + 0.5 \cdot 0.5 = 0.5 \end{aligned}$$

zeigt die Behauptung.

6. Berechnen Sie für $r = 0.7$ und $n = 20$ ein approximatives 95%-Konfidenzintervall für die wahre Korrelation ρ unter Verwendung von Fisher's z-Transformation.

► Die transformierte Korrelation ist nach Beispiel 3.25

$$z = 0.5 \log \left(\frac{1 + 0.7}{1 - 0.7} \right) = 0.867.$$

Auf dieser Skala beträgt der Standardfehler $1/\sqrt{n-3} = 0.243$. Das 95%-Wald-Konfidenzintervall für $\tanh^{-1}(\rho)$ ist demnach

$$[z \pm 1.96 \cdot \text{se}(z)] = [0.392, 1.343]$$

und kann mittels der streng monoton wachsenden Umkehrfunktion zum Konfidenzintervall

$$[\tanh(0.392), \tanh(1.343)] = [0.373, 0.872]$$

für ρ zurücktransformiert werden.

7. Bestimmen Sie allgemeingültige Formeln für die Score-Konfidenzintervalle bei Poisson-Verteilung basierend auf der Fisher-Information, vgl. Beispiel 3.18.

► Nach Beispiel 3.17 lautet die Score-Teststatistik für die Nullhypothese, dass die wahre Rate gleich λ ist,

$$T_2(\lambda) = \frac{S(\lambda)}{\sqrt{I(\lambda)}} = \sqrt{n} \cdot \frac{\bar{x} - \bar{e}\lambda}{\sqrt{\bar{x}}},$$

falls man die Berechnung auf der Fisher-Information $I(\lambda)$ basiert. Wir bestimmen nun die Werte von λ , die ein Signifikanztest mit Niveau α aufgrund der asymptotischen H_0 -Standardnormalverteilung von T_2 nicht ablehnen würde, d. h. für die $|T_2(\lambda)| \leq q := z_{1-\alpha/2}$ gilt:

$$\begin{aligned} \left| \sqrt{n} \cdot \frac{\bar{x} - \bar{e}\lambda}{\sqrt{\bar{x}}} \right| &\leq q \\ \left| \frac{\bar{x}}{\bar{e}} - \lambda \right| &\leq \frac{q}{\bar{e}} \sqrt{\frac{\bar{x}}{n}} \\ \lambda &\in \left[\frac{\bar{x}}{\bar{e}} \pm \frac{q}{\bar{e}} \sqrt{\frac{\bar{x}}{n}} \right] \end{aligned}$$

Somit haben wir ein approximatives $(1 - \alpha)$ -Score-Konfidenzintervall für λ erhalten. Im Gegensatz zu dem auf der erwarteten Fisher-Information basierenden Konfidenzintervall in Beispiel 3.18 ist es um den ML-Schätzer \bar{x}/\bar{e} zentriert.

8. Eine Grundgesamtheit besitze den Mittelwert μ und die Varianz σ^2 . Die Stichprobenvariablen X_1, \dots, X_5 seien unabhängige Ziehungen aus dieser Grundgesamtheit. Folgende Schätzfunktionen für μ werden betrachtet:

$$T_1 = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5),$$

$$T_2 = \frac{1}{3}(X_1 + X_2 + X_3),$$

$$T_3 = \frac{1}{8}(X_1 + X_2 + X_3 + X_4) + \frac{1}{2}X_5,$$

$$T_4 = X_1 + X_2$$

$$\text{und } T_5 = X_1.$$

- a) Welche der Schätzfunktionen sind erwartungstreu für μ ?

► T_1, T_2 und T_5 sind als arithmetische Mittel erwartungstreu für μ , vgl. Aufgabe 2. Weiterhin ist T_3 wegen

$$E(T_3) = \frac{1}{8} \cdot 4\mu + \frac{1}{2}\mu = \mu$$

ebenfalls erwartungstreu, im Gegensatz zu T_4 , denn

$$E(T_4) = E(X_1) + E(X_2) = 2\mu.$$

b) Berechnen Sie den MSE für jede Schätzfunktion.

► Für die erwartungstreuen Schätzer ist der MSE gleich der Varianz, da der Bias gleich Null ist, siehe Gleichung 3.3. Nach Aufgabe 2 haben wir somit für die arithmetischen Mittel

$$MSE(T_1) = \frac{\sigma^2}{5}, \quad MSE(T_2) = \frac{\sigma^2}{3} \quad \text{und} \quad MSE(T_5) = \sigma^2.$$

Der MSE von T_3 ergibt sich zu

$$MSE(T_3) = \text{Var}(T_3) = \frac{1}{8^2} \cdot 4\sigma^2 + \frac{1}{2^2}\sigma^2 = \frac{5}{16}\sigma^2.$$

Der MSE von T_4 schließlich setzt sich aus dem quadrierten Bias μ^2 und der Varianz $2\sigma^2$ zu

$$MSE(T_4) = \mu^2 + 2\sigma^2$$

zusammen.

9. Bei der Abfüllung von Mineralwasser in 1l-Flaschen soll der Sollwert von 1l eingehalten werden. Für die verwendete Abfüllanlage gilt nach Herstellerangaben, dass die Abfüllungen normalverteilt sind mit $\mu = 1000$ ml und $\sigma^2 = 100$ ml². Zur Überprüfung der Abfüllmenge werden zu bestimmten Zeitpunkten Stichproben vom Umfang $n = 10$ erhoben. Dabei ergab sich beim letzten Zeitpunkt eine durchschnittliche Abfüllmenge von 1020 ml.

a) Befindet sich der Abfüllprozess nicht mehr unter statistischer Kontrolle ($\alpha = 0.01$)? Wie lässt sich diese Frage als statistisches Testproblem erfassen?

► Als statistisches Testproblem können wir die Fragestellung über die Hypothesen $H_0 : \mu = 1000$ und $H_1 : \mu \neq 1000$ formulieren. Dabei soll H_0 überprüft werden. Laut Hersteller gilt für die Abfüllmenge X

$$X \sim N(\mu, \sigma^2 = 100),$$

d. h. die Werte streuen mit normalverteilten Abweichungen um die Zielmenge μ . Nun liegt eine Zufallsstichprobe vom Umfang $n = 10$ mit $\bar{x} = 1020$ vor. Da das arithmetische Mittel \bar{X} als Linearkombination von normalverteilten Zufallsvariablen wieder normalverteilt ist, erhalten wir mit Aufgabe 2

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

Nimmt man die Gültigkeit von H_0 an, so errechnet sich die Wahrscheinlichkeit, dass in Richtung von H_1 extremere Werte als \bar{x} der durchschnittlichen Abfüllmenge beobachtet werden, zu

$$P(|\bar{X} - 1000| > \bar{x} - 1000) = P\left(\left|\frac{\bar{X} - 1000}{\sigma^2/n}\right| > 20n/\sigma^2\right) = 2\Phi(-20n/\sigma^2),$$

weil $(\bar{X} - 1000)/(\sigma^2/n)$ unter H_0 standardnormalverteilt ist. Setzt man die konkreten Zahlen ein, ergibt sich

$$P(|\bar{X} - 1000| > 20) = 0.046.$$

Dieser p -Wert liegt also noch nicht unter dem Signifikanzniveau $\alpha = 0.01$, sodass sich der Prozess noch unter statistischer Kontrolle befindet.

b) Was sagt der Fehler 1. Art hier aus?

► Der Fehler 1. Art des Signifikanztests ist die Wahrscheinlichkeit, die Nullhypothese fälschlicherweise abzulehnen, d. h. einen Prozess mit $\mu = 1000$ als falsch eingestellt zu beurteilen. Unter der Normalverteilungsannahme wird diese Wahrscheinlichkeit immer kleiner als $\alpha = 0.01$ sein.

10. Gegeben sei eine Zufallsstichprobe X_1, \dots, X_n eines $\text{Exp}(\lambda)$ -verteilten Untersuchungsmerkmals ($\lambda > 0$).

a) Bestimmen Sie die Score-Funktion von λ und lösen Sie die Score-Gleichung.

► Aus der (Log-)Likelihood

$$L(\lambda) = \lambda^n \cdot \exp\left(-\lambda \sum_{i=1}^n x_i\right)$$

$$\text{bzw. } l(\lambda) = n \cdot \log(\lambda) - \lambda n \bar{x}$$

ergibt sich die Score-Funktion

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

die die Nullstelle $\hat{\lambda}_{ML} = 1/\bar{x}$ hat.

b) Bestimmen Sie die beobachtete Fisher-Information, den Standardfehler und ein 95%-Wald-Intervall für λ .

► Die Fisher-Information ist somit

$$I(\lambda) = -\frac{d}{d\lambda} S(\lambda) = -\left(-\frac{n}{\lambda^2}\right) = \frac{n}{\lambda^2}$$

und die beobachtete Fisher-Information ergibt sich durch Einsetzen des ML-Schätzers zu $I(\hat{\lambda}_{ML}) = n\bar{x}^2$. Der Standardfehler von $\hat{\lambda}_{ML}$ ist also

$$\text{se}(\hat{\lambda}_{ML}) = 1/\sqrt{I(\hat{\lambda}_{ML})} = \frac{1}{\bar{x}\sqrt{n}},$$

und liefert das 95%-Wald-Konfidenzintervall

$$[\hat{\lambda}_{ML} \pm 1.96 \text{se}(\hat{\lambda}_{ML})] = [1/\bar{x} \pm 1.96 \cdot (\bar{x}\sqrt{n})^{-1}]$$

für λ .

c) Bestimmen Sie die erwartete Fisher-Information $J(\lambda)$.

► Da die Fisher-Information $I(\lambda)$ nicht von X abhängt, gilt hier

$$J(\lambda) = \mathbb{E}(I(\lambda)) = I(\lambda) = \frac{n}{\lambda^2}.$$

d) Bestimmen Sie den ML-Schätzer für $\psi = \log \lambda$ und geben Sie das 95%-Wald-Intervall für ψ an. Vergleichen Sie die Wald-Intervalle für ψ und λ .

► Aufgrund der Invarianz der ML-Schätzung bzgl. eindeutigen Transformationen ist

$$\hat{\psi}_{ML} = \log \hat{\lambda}_{ML} = -\log \bar{x}.$$

Nach Satz 2.1 erhalten wir die beobachtete Fisher-Information des transformierten Parameters mittels

$$\tilde{I}(\hat{\psi}_{ML}) = I(\hat{\lambda}_{ML})\hat{\lambda}_{ML}^2 = n\bar{x}^2/\bar{x}^2 = n,$$

sodass das 95%-Wald-Konfidenzintervall für ψ

$$[-\log \bar{x} \pm 1.96 \cdot n^{-1/2}]$$

lautet. Das auf die ursprüngliche Skala rücktransformierte Konfidenzintervall ist

$$[\bar{x}^{-1} / \exp(1.96/\sqrt{n}), \bar{x}^{-1} \cdot \exp(1.96/\sqrt{n})]$$

und damit im Gegensatz zum Intervall aus 10b) nicht mehr um den ML-Schätzer zentriert.

e) Bestimmen Sie die Cramér-Rao-Schranke für erwartungstreue Schätzer von λ .

► Nach Satz 3.8 gilt für alle erwartungstreuen Schätzer $T(X)$

$$\text{Var}(T(X)) \geq \frac{1}{J(\lambda)} = \frac{\lambda^2}{n}.$$

f) Konstruieren Sie einen unverzerrten Schätzer für λ und berechnen Sie dessen Varianz.

► Aufgrund der Additionseigenschaft der Gamma-Verteilung gilt $n\bar{X} = \sum_{i=1}^n X_i \sim G(n, \lambda)$ und somit für die Transformation

$$\hat{\lambda}_{ML} = 1/\bar{X} \sim \text{IG}(n, n\lambda),$$

da der zweite Parameter der Gamma-Verteilung ein inverser Skalenparameter ist (d. h. aus $Y \sim G(a, b)$ folgt $cY \sim G(a, b/c)$). Nach Tabelle A.2 haben wir somit

$$\mathbb{E}(\hat{\lambda}_{ML}) = \frac{n\lambda}{n-1},$$

sodass der ML-Schätzer verzerrt ist. Der korrigierte Schätzer

$$\hat{\lambda} := \frac{n-1}{n\bar{X}}$$

ist demnach unverzerrt für λ . Er hat die Varianz

$$\text{Var}(\hat{\lambda}) = \frac{(n-1)^2}{n^2} \text{Var}(1/\bar{X}) = \frac{(n-1)^2}{n^2} \frac{n^2 \lambda^2}{(n-1)^2 (n-2)} = \frac{\lambda^2}{n-2}$$

und erreicht erst asymptotisch die Cramér-Rao-Schranke λ^2/n .

11. Eine andere Parametrisierung der Exponentialverteilung ist

$$f_X(x) = \frac{1}{\theta} \exp \left\{ -\frac{x}{\theta} \right\} I_{\{x \geq 0\}}, \quad \theta > 0.$$

Sei X_1, \dots, X_n eine Zufallsstichprobe mit Parameter θ exponentialverteilter Zufallsvariablen. Die Hypothese $H_0 : \theta = \theta_0$ ist gegen $H_1 : \theta \neq \theta_0$ zu testen.

a) Bestimmen Sie die Teststatistik für beide Varianten des Score-Tests, d.h. T_1 und T_2 .

► Hier liegt die Erwartungswert-Parametrisierung der Exponentialverteilung vor, d. h. $E(X) = \theta = h(\lambda) = 1/\lambda$, wenn λ die Rate ist, mit der in der vorigen Aufgabe parametrisiert wurde. In der Notation von Beispiel 3.17 wollen wir

$$T_1(\theta_0) = \frac{S(\theta_0)}{\sqrt{J(\theta_0)}} \quad \text{und} \quad T_2(\theta_0) = \frac{S(\theta_0)}{\sqrt{I(\theta_0)}}$$

bestimmen. Die Score-Funktion können wir aus der in Aufgabe 10 erhaltenen Score-Funktion für λ erhalten:

$$S(\theta) = (n\theta - n\bar{x}) \cdot \left(-\frac{1}{\theta^2} \right) = \frac{n(\bar{x} - \theta)}{\theta^2},$$

vgl. die allgemeine Regel zur Score-Funktion von Transformationen (2.3) im Beweis zu Satz 2.1. Die Fisher-Information ist somit

$$I(\theta) = -\frac{d}{d\theta} S(\theta) = n \frac{2\bar{x} - \theta}{\theta^3}$$

und die erwartete Fisher-Information vereinfacht sich zu

$$J(\theta) = n \frac{2E(\bar{X}) - \theta}{\theta^3} = \frac{n}{\theta^2}.$$

Die Teststatistiken lauten demnach

$$T_1(\theta_0) = \frac{n(\bar{x} - \theta_0)}{\theta_0^2} \cdot \frac{\theta_0}{\sqrt{n}} = \sqrt{n} \frac{\bar{x} - \theta_0}{\theta_0}$$

und $T_2(\theta_0) = \frac{n(\bar{x} - \theta_0)}{\theta_0^2} \cdot \frac{\theta_0^{3/2}}{\sqrt{n(2\bar{x} - \theta_0)}} = T_1(\theta_0) \sqrt{\frac{\theta_0}{2\bar{x} - \theta_0}}.$

- b) Eine Stichprobe der Größe $n = 100$ hat $\bar{x} = 0.26142$. Bestimmen Sie für T_1 und T_2 , ob $H_0 : \theta_0 = 0.25$ bei einem Test zum Signifikanzniveau 0.05 akzeptiert wird.

► Setzt man die Zahlen ein, so ergeben sich die Werte

$$T_1 = 0.457 \quad \text{und} \quad T_2 = 0.437.$$

Asymptotisch folgen beide Score-Teststatistiken unter der Nullhypothese einer Standardnormalverteilung, deren kritisches $(1 - \alpha/2)$ -Quantil 1.96 hier nicht überschritten wird. Somit kann H_0 nicht abgelehnt werden.

12. Ein Zufallsvektor \mathbf{X} bzw. seine Verteilung gehört zu einer p -parametrischen Exponentialfamilie, wenn die (stetige oder diskrete) logarithmierte Dichte von der Form

$$\log f(\mathbf{x} | \boldsymbol{\tau}) = \sum_{i=1}^p \eta_i(\boldsymbol{\tau}) T_i(\mathbf{x}) - A(\boldsymbol{\tau}) + c(\mathbf{x}) \quad (3.2)$$

ist. Dabei ist $\boldsymbol{\tau}$ der Parametervektor und T_i, η_i, A und c sind reellwertige Funktionen. Es wird angenommen, dass keine lineare Abhängigkeiten zwischen den η_i 's existieren; ebenso für die T_i 's. Wir bezeichnen dann $\theta_1 = \eta_1(\tau_1), \dots, \theta_p = \eta_p(\tau_p)$ als *kanonische Parameter*, sodass mit $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ und $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), \dots, T_p(\mathbf{x}))'$ die Log-Dichte auf sogenannter *kanonischer Form* folgendes Aussehen hat:

$$\log f(\mathbf{x} | \boldsymbol{\theta}) = \boldsymbol{\theta}' \mathbf{T}(\mathbf{x}) - A(\boldsymbol{\theta}) + c(\mathbf{x}). \quad (3.3)$$

Interessant an Exponentialfamilien ist, dass viele der bekannten Verteilungen, z. B. Poisson, Geometrisch, Binomial, Normal, Gamma zu den Exponentialfamilien gehören. Daher lohnt es sich, die Theorie generell für Exponentialfamilien zu formulieren, damit viele Verteilungen auf einmal behandelt werden können. Zum Beispiel ist ein nützliches Resultat für die einparametrische Exponentialfamilie in kanonischer Form, dass $E(T(X)) = A'(\theta)$ und $\text{Var}(T(X)) = A''(\theta)$ gilt.

- a) Zeigen Sie, dass die Poisson-Verteilung mit Parameter λ auf die Form (3.2) bzw. (3.3) gebracht werden kann. Berechnen Sie den Erwartungswert und die Varianz der Verteilung.

► Sei $X \sim \text{Po}(\lambda)$. Dann ist

$$\log f(x | \lambda) = \log \left(\frac{\lambda^x}{x!} \exp(-\lambda) \right) = \log(\lambda)x - \lambda + (-\log(x!)),$$

d. h. $p = 1$, $\theta = \log(\lambda)$, $T(x) = x$ und

$$A(\theta) = \lambda = \exp(\theta).$$

Erwartungswert und Varianz von X sind somit beide gleich $\exp(\theta)$.

b) Bearbeiten Sie Aufgabe 12a) für die Exponentialverteilung.

► Für eine exponentialverteilte Zufallsvariable $X \sim \text{Exp}(\lambda)$ können wir

$$\log f(x | \lambda) = \log(\lambda \exp(-\lambda x)) = (-\lambda)x - (-\log(\lambda))$$

schreiben, sodass $\theta = -\lambda$ der kanonische Parameter ist und

$$A(\theta) = -\log(\lambda) = -\log(-\theta).$$

Erwartungswert und Varianz lauten demnach bekanntermaßen

$$E(X) = A'(\theta) = -\frac{1}{-\theta}(-1) = -\frac{1}{\theta} = \frac{1}{\lambda}$$

$$\text{und } \text{Var}(X) = A''(\theta) = \frac{1}{\theta^2} = \frac{1}{\lambda^2}.$$

c) Zeigen Sie, dass die Normalverteilung mit Parameter μ und bekanntem $\sigma^2 = \sigma_0^2$ auf die Form (3.2) bzw. (3.3) gebracht werden kann.

► Man erhält

$$\log f(x | \mu) = \frac{\mu}{\sigma_0^2}x - \frac{\mu^2}{2\sigma_0^2} + \left(-\frac{x^2}{2\sigma_0^2} - \log(\sqrt{2\pi}\sigma_0) \right).$$

d) Zeigen Sie für die einparametrische Exponentialfamilie, dass $T(x)$ suffizient für θ ist.

► Die Suffizienz folgt mit dem Faktorisierungssatz 2.2, denn

$$f(x | \theta) = \exp(\theta T(x) - A(\theta) + c(x)) = g(T(x); \theta)h(x)$$

mit $g(T(x); \theta) = \exp(\theta T(x) - A(\theta))$ und $h(x) = \exp(c(x))$. ($T(x)$ ist sogar minimal suffizient für θ .)

13. Zeigen Sie, dass für die einparametrische Exponentialfamilie $I(\hat{\tau}_{ML}) = J(\hat{\tau}_{ML})$ gilt. Zeigen Sie auch, dass falls θ ein kanonischer Parameter ist, $I(\theta) = J(\theta)$ gilt.

► Es liege also eine Zufallsvariable X mit Dichte $f(x | \tau)$, die (3.2) ($p = 1$) erfüllt, vor. Die Score-Funktion ergibt sich durch Ableitung der Log-Dichte zu

$$S(\tau) = \eta'(\tau)T(x) - A'(\tau),$$

sodass für den ML-Schätzer $\hat{\tau}_{ML}$

$$T(x) = \frac{A'(\hat{\tau}_{ML})}{\eta'(\hat{\tau}_{ML})}$$

gilt. Die beobachtete Fisher-Information erhalten wir aus der Fisher-Information

$$I(\tau) = A''(\tau) - \eta''(\tau)T(x)$$

deshalb als

$$I(\hat{\tau}_{ML}) = A''(\hat{\tau}_{ML}) - \eta''(\hat{\tau}_{ML}) \frac{A'(\hat{\tau}_{ML})}{\eta'(\hat{\tau}_{ML})}.$$

Andererseits folgt aus

$$\mathbb{E}(T(X)) = (A \circ \eta^{-1})'(\theta) = A'(\eta^{-1}(\theta)) \cdot (\eta^{-1})'(\theta) = \frac{A'(\tau)}{\eta'(\tau)},$$

wobei $\theta = \eta(\tau)$ der kanonische Parameter ist, dass

$$J(\tau) = A''(\tau) - \eta''(\tau) \frac{A'(\tau)}{\eta'(\tau)}$$

gilt. Setzt man nun $\hat{\tau}_{ML}$ ein, erhält man die gleiche Form wie für $I(\hat{\tau}_{ML})$.

Falls bereits die kanonische Parametrisierung (3.3) vorliegt, ist

$$S(\theta) = T(x) - A'(\theta)$$

$$\text{und } I(\theta) = A''(\theta),$$

unabhängig von x , sodass $I(\theta) = J(\theta)$ gilt.

14. Angenommen X_1, \dots, X_n ist eine Zufallsstichprobe aus einer einparametrischen Exponentialfamilie mit kanonischem Parameter θ . Bestimmen Sie einen Ausdruck für $l(\theta)$.

► Die einzelnen Log-Likelihood-Beiträge

$$\log f(x_i | \theta) = \theta T(x_i) - A(\theta) + c(x_i)$$

addieren sich zur gesamten Log-Likelihood

$$l(\theta) = \sum_{i=1}^n \log f(x_i | \theta) = \theta \sum_{i=1}^n T(x_i) - nA(\theta) + \sum_{i=1}^n c(x_i).$$

15. Alle Taxis in einer Großstadt seien von $1, \dots, N$ durchnummeriert. Ein Besucher sieht an einem Taxistand n Taxis, die eine zufällige Auswahl aller Taxis darstellen. Deren Nummern seien X_1, \dots, X_n mit $X_i \in \{1, \dots, N\}$. Sei weiterhin $Y = \max(X_1, \dots, X_n)$.

- a) Zeigen Sie, dass Y suffizient für N ist und leiten Sie die Likelihoodfunktion $L(N)$ her.

► Die Wahrscheinlichkeitsfunktion ist (bei Annahme von unsortierten Taxinummern)

$$P(X = x; N) = \left\{ \binom{N}{n} n! \right\}^{-1} I_{\{n, \dots, N\}}(\max_i(x_i)),$$

und hängt somit von den Daten x nur über $y = \max_i(x_i)$ ab, sodass Y suffizient für N ist. Die Likelihoodfunktion vereinfacht sich zu

$$L(N) = \frac{(N-n)!}{N!} I_{\{y, y+1, \dots\}}(N). \quad (3.4)$$

b) Zeigen Sie, dass

$$\hat{N} = \frac{n+1}{n}Y - 1$$

ein erwartungstreuer Schätzer von N ist.

► Die Wahrscheinlichkeitsfunktion von Y ist

$$P(Y = y; N) = \frac{\binom{y-1}{n-1}}{\binom{N}{n}} I_{\{n, \dots, N\}}(y).$$

Somit ergibt sich der Erwartungswert zu

$$\begin{aligned} E(Y) &= \binom{N}{n}^{-1} \sum_{y=n}^N y \cdot \binom{y-1}{n-1} \\ &= \binom{N}{n}^{-1} \sum_{y=n}^N n \cdot \binom{y}{n} \\ &= \binom{N}{n}^{-1} n \cdot \sum_{y=n}^N \binom{y+1-1}{n+1-1} \\ &= \binom{N}{n}^{-1} n \cdot \sum_{y=n+1}^{N+1} \binom{y-1}{(n+1)-1}. \end{aligned}$$

Es gilt $\sum_{y=n}^N \binom{y-1}{n-1} = \binom{N}{n}$ und somit

$$\begin{aligned} E(Y) &= \binom{N}{n}^{-1} n \cdot \binom{N+1}{n+1} \\ &= \frac{n!(N-n)!}{N!} n \cdot \frac{(N+1)!}{(n+1)!(N-n)!} \\ &= \frac{n}{n+1}(N+1). \end{aligned}$$

Insgesamt also

$$\begin{aligned} E(\hat{N}) &= \frac{n+1}{n}E(Y) - 1 \\ &= \frac{n+1}{n} \frac{n}{n+1}(N+1) - 1 \\ &= N. \end{aligned}$$

c) Bestimmen Sie auch den ML-Schätzer von N und vergleichen Sie diesen mit dem erwartungstreuen Schätzer.

► Der Quotient der Likelihood-Werte von zwei aufeinanderfolgenden Werten N , $N+1$ ($N \geq y$) ist

$$\frac{L(N+1)}{L(N)} = \frac{(N+1-n)!}{(N+1)!} \frac{N!}{(N-n)!} = \frac{N+1-n}{N+1} < 1,$$

sodass der ML-Schätzer $\hat{N}_{ML} = Y$ ist. Im Gegensatz zu \hat{N} ist er verzerrt und unterschätzt N systematisch.

16. Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer geometrischen Verteilung mit Träger $\mathcal{T} = \{1, 2, \dots\}$.

a) Bestimmen Sie die Log-Likelihoodfunktion $l(\pi)$ und den ML-Schätzer $\hat{\pi}_{ML}$.

► Die Likelihoodfunktion ist

$$\begin{aligned} L(\pi) &= \prod_{i=1}^n \pi(1-\pi)^{x_i-1} \\ &= \pi^n (1-\pi)^{\sum x_i - n} \\ &= \pi^n (1-\pi)^{n(\bar{x}-1)}, \end{aligned}$$

sodass die Log-Likelihoodfunktion

$$l(\pi) = n \log(\pi) + n(\bar{x} - 1) \log(1 - \pi)$$

ist. Die Score-Funktion ist folglich

$$S(\pi) = \frac{d}{d\pi} l(\pi) = \frac{n}{\pi} - \frac{n(\bar{x} - 1)}{1 - \pi}$$

und die Lösung der Score-Gleichung $S(\pi) = 0$ ist der ML-Schätzwert

$$\hat{\pi}_{ML} = \frac{n}{\sum_{i=1}^n x_i} = 1/\bar{x}.$$

b) Berechnen Sie ein 95%-Wald-Konfidenzintervall für π bei Daten $x = (4, 20, 10)$.

► Für das 95%-Wald-Konfidenzintervall

$$[\hat{\pi}_{ML} \pm 1.96 \cdot I(\hat{\pi}_{ML})^{-1/2}]$$

für π benötigen wir noch die beobachtete Fisher-Information:

$$\begin{aligned} I(\pi) &= -\frac{d}{d\pi} S(\pi) = \frac{n}{\pi^2} + \frac{n(\bar{x} - 1)}{(1 - \pi)^2} \\ \Rightarrow I(\hat{\pi}_{ML}) &= n \left(\bar{x}^2 + \frac{\bar{x} - 1}{(\frac{\bar{x}-1}{\bar{x}})^2} \right) = \frac{n\bar{x}^3}{\bar{x} - 1} \end{aligned}$$

Bei den beobachteten Daten ergibt sich $\bar{x} = 11.333$ und damit $\hat{\pi}_{ML} = 0.088$ sowie $I(\hat{\pi}_{ML}) = 422.560$, sodass das Konfidenzintervall

$$[0.088 \pm 1.96 \cdot 0.049] = [-0.008, 0.184]$$

Werte außerhalb des Parameterraums $(0, 1)$ enthält.

- c) Betrachtet wird der Parameter $\phi = \text{logit}(\pi) = \log(\pi/(1-\pi))$. Berechnen Sie $\hat{\phi}_{ML}$ für die Daten aus 16b) und ein 95%-Wald-Konfidenzintervall für ϕ . Transformieren Sie dieses Intervall auf die π -Skala zurück.

► Um eine bessere quadratische Approximation zu erreichen, wird oft der ursprüngliche Parameter (hier $\pi \in [0, 1]$) in einen unbeschränkten Parameter (hier $h(\pi) = \phi \in \mathbb{R}$) transformiert.

Aus der Invarianz der ML-Schätzung folgt sofort

$$\hat{\phi}_{ML} = \text{logit}(\hat{\pi}_{ML}) = -\log(\bar{x} - 1) = -2.335.$$

Mit Satz 2.1 folgt außerdem

$$I(\hat{\phi}_{ML}) = \frac{n\bar{x}^3}{\bar{x} - 1} \cdot \left(\frac{1}{\hat{\pi}_{ML}(1 - \hat{\pi}_{ML})} \right)^{-2} = \frac{n(\bar{x} - 1)}{\bar{x}},$$

sodass $\text{se}(\hat{\pi}_{ML}) = \sqrt{\frac{\bar{x}}{n(\bar{x}-1)}} = 0.605$. Das gleiche Ergebnis hätte man übrigens auch durch die Anwendung der Delta-Regel

$$\text{se}(\hat{\phi}_{ML}) = \text{se}(\hat{\pi}_{ML}) \cdot \left| \frac{dh(\hat{\pi}_{ML})}{d\pi} \right|$$

erhalten. Das 95%-Wald-Konfidenzintervall für ϕ ist also

$$\left[-\log(\bar{x} - 1) \pm 1.96 \cdot \sqrt{\frac{\bar{x}}{n(\bar{x} - 1)}} \right] = [-3.521, -1.149]$$

und wird mit der Verteilungsfunktion der logistischen Verteilung,

$$h^{-1}(\phi) = \text{logit}^{-1}(\phi) = \exp(\phi)/(1 + \exp(\phi)),$$

zurücktransformiert zum 95%-Konfidenzintervall $[0.029, 0.241]$ für π .

- d) Erstellen Sie in R einen Plot der relativen Log-Likelihoodfunktion $\tilde{l}(\pi)$ und deren quadratischer Approximation. Bestimmen Sie numerisch ein 95%-Likelihood-Konfidenzintervall für π . Benutzen Sie hierzu die `uniroot`-Funktion.

► Zunächst implementieren wir die (normierte) Log-Likelihoodfunktion $l(\pi)$ bzw. $\tilde{l}(\pi) = l(\pi) - l(\hat{\pi}_{ML})$:

```

> geomLogLikelihood <- function(          # gibt die Log-Likelihood von
+                                     pi,    # Parameter pi
+                                     x      # bei iid geometrisch verteilten Daten x
+                                     )      # zurück
+ {
+   length(x) * (log(pi) + (mean(x) - 1) * log(1 - pi))
+ }
> geomNormLogLikelihood <- function(      # gibt die normierte Log-Likelihood
+                                     # zurück
+                                     pi,    # wie oben
+                                     x      # wie oben
+                                     )
+ {
+   geomLogLikelihood(pi, x) - geomLogLikelihood(1 / mean(x), x)
+ }
> ## beobachtete Daten waren
> x <- c(4, 20, 10)

```

Die quadratische Taylor-Approximation um $\hat{\pi}_{ML}$ ist

$$\begin{aligned}
 \tilde{l}(\pi) &\approx \tilde{l}(\hat{\pi}_{ML}) + \tilde{l}'(\hat{\pi}_{ML})(\pi - \hat{\pi}_{ML}) + \frac{1}{2}\tilde{l}''(\hat{\pi}_{ML})(\pi - \hat{\pi}_{ML})^2 \\
 &= 0 + 0 - \frac{1}{2}I(\hat{\pi}_{ML})(\pi - \hat{\pi}_{ML})^2,
 \end{aligned}$$

da $\tilde{l}'(\hat{\pi}_{ML}) = S(\hat{\pi}_{ML}) = 0$ gilt.

```

> geomApproxNormLogLikelihood <- function( # gibt die quadratisch approximierte
+                                     # normierte Log-Likelihood zurück
+                                     pi,    # wie oben
+                                     x      # wie oben
+                                     )
+ {
+   n <- length(x)
+   mittel <- mean(x)
+
+   beobFisher <- n * mittel^3 / (mittel - 1)
+
+   - 1 / 2 * beobFisher * (pi - 1 / mittel)^2
+ }

```

Die Nullstellen der Funktion

$$\tilde{l}(\pi) + \frac{1}{2} \cdot \chi_{1-\alpha}^2(1)$$

sind die Grenzen des Likelihood-Konfidenzintervalls zum Niveau $(1 - \alpha)$:

```

> ## allgemeine Funktion, die beliebige Likelihood entgegennimmt:
> likelihoodCi <- function(
+   alpha = 0.05, # 1-alpha ist das Niveau des Intervalls
+   loglik,       # unnormierte Log-Likelihoodfunktion
+   thetaML,      # ML-Schätzer
+   lower,        # untere Grenze des Parameterraums
+   upper,        # obere Grenze des Parameterraums
+   ...           # weitere Argumente für loglik (z.B.
+               # Daten)
+ )

```

```

+ {
+   ## Zielfunktion
+   f <- function(theta, ...)
+     loglik(theta, ...) - loglik(thetaMl, ...) + 1/2*qchisq(1-alpha, df=1)
+
+   ## Grenzen des Likelihood-Intervalls bestimmen:
+   eps <- sqrt(.Machine$double.eps) # von Grenzen etwas wegbleiben
+   lowerBound <- uniroot(f, interval = c(lower + eps, thetaMl), ...)$root
+   upperBound <- uniroot(f, interval = c(thetaMl, upper - eps), ...)$root
+
+   return(c(lower = lowerBound, upper = upperBound))
+ }
> ## damit das 95%-Intervall berechnen:
> likCi <- likelihoodCi(loglik = geomLogLikelihood,
+   thetaMl= 1 / mean(x),
+   lower = 0, upper=1,
+   x = x)
> likCi
      lower      upper
0.02268733 0.21311094

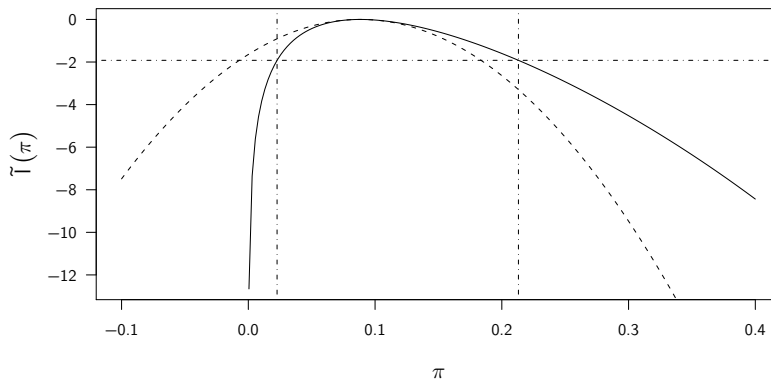
```

Insgesamt können wir die drei Schritte wie folgt darstellen:

```

> ## normierte Log-Likelihood:
> pi <- seq(-0.1, 0.4, length = 200)
> plot(pi, geomNormLogLikelihood(pi, x),
+   type = "l",
+   xlab = expression(pi), ylab = expression(tilde(l)(pi)))
> ## quadratische Approximation:
> lines(pi, geomApproxNormLogLikelihood(pi, x),
+   lty = 2)
> ## Likelihood-Intervall:
> abline(h = -1/2 * qchisq(0.95, 1), lty = 4)
> abline(v = likCi, lty = 4)

```



Wie in Abbildung 3.6 verdeutlicht wird, entsprechen die Grenzen des Wald-Konfidenzintervalls aus Aufgabe 16b) genau den Abszissen der Schnittpunkte der quadratischen Approximation mit der Gerade $c = -\frac{1}{2}\chi^2_{1-\alpha}(1)$.

4 Likelihood-Inferenz bei vektoriellem Parameter

1. Eine Kohortenstudie zur Untersuchung der Inzidenz der ischämischen Herzkrankheit (IHK) wurde anhand von 337 männlichen Probanden durchgeführt. Dabei wurde jeder Proband als nicht-exponiert (Gruppe 1, tägliche Energiezufuhr ≥ 2750 kcal) oder exponiert (Gruppe 2, tägliche Energiezufuhr < 2750 kcal) eingestuft um ein Bild seiner körperlichen Aktivität zu bekommen. Es wurde die Anzahl von Personenjahren ($Y_1 = 2768.9$ und $Y_2 = 1857.5$) und die Anzahl von IHK-Erkrankungen ($D_1 = 17$ und $D_2 = 28$) registriert. Es wird angenommen, dass $D_i \sim \text{Po}(\lambda_i Y_i)$, $i = 1, 2$, wobei $\lambda_i > 0$ die Rate von Neuerkrankungen ist.
a) Bestimmen Sie für jede Kategorie den ML-Schätzer $\hat{\lambda}_i$ und ein zugehöriges 95%-Wald-Intervall für $\log(\lambda_i)$ mit anschließender Rücktransformation.
► Die Rate der Poissonverteilung ist also jeweils durch die Transformation $\theta_i = \lambda_i Y_i$ von λ_i gegeben, bzw. wir berechnen λ_i aus θ_i mittels

$$\lambda_i = \frac{\theta_i}{Y_i} = g(\theta_i).$$

Durch die Invarianz des ML-Schätzers erhalten wir

$$\hat{\lambda}_i = g(\hat{\theta}_i) = \frac{D_i}{Y_i},$$

da der ML-Schätzer von θ_i gleich D_i ist, vgl. etwa Beispiel 2.4. Bei den vorliegenden Daten haben wir somit $\hat{\lambda}_1 = D_1/Y_1 = 6.14 \cdot 10^{-3}$ und $\hat{\lambda}_2 = D_2/Y_2 = 1.51 \cdot 10^{-2}$. Mittels der Δ -Regel folgt

$$\text{se}(\hat{\lambda}_i) = \text{se}(\hat{\theta}_i) \cdot \left| g'(\hat{\theta}_i) \right|,$$

wobei $g'(\theta_i) = 1/Y_i$ und

$$\text{se}(\hat{\theta}_i) = \left(\frac{1}{D_i} \right)^{-\frac{1}{2}} = \sqrt{D_i},$$

da man leicht zeigt, dass die beobachtete Fisher-Information $I(\hat{\theta}_i) = 1/D_i$ ist. Insgesamt also $\text{se}(\hat{\lambda}_i) = \sqrt{D_i}/Y_i$.

Die weitere Transformation $\psi_i = \log(\lambda_i) = f(\lambda_i)$ hat den ML-Schätzer

$$\hat{\psi}_i = \log \left(\frac{D_i}{Y_i} \right)$$

und den Standardfehler

$$\begin{aligned} \text{se}(\hat{\psi}_i) &= \text{se}(\hat{\lambda}_i) \cdot \left| f'(\hat{\lambda}_i) \right| \\ &= \frac{\sqrt{D_i}}{Y_i} \cdot \left| \frac{1}{\hat{\lambda}_i} \right| \\ &= \frac{1}{\sqrt{D_i}}. \end{aligned}$$

Damit ist ein 95%-Wald-Intervall mit Log-Transformation für λ_i gleich

$$\exp \left(\hat{\psi}_i \pm 1.96 \cdot \frac{1}{\sqrt{D_i}} \right);$$

für die konkreten Daten ergeben sich die Konfidenzintervalle

$$[3.82 \cdot 10^{-3}, 9.88 \cdot 10^{-3}] \quad \text{und} \quad [1.04 \cdot 10^{-2}, 2.18 \cdot 10^{-2}].$$

- b) Um zu untersuchen, ob $\lambda_1 = \lambda_2$ gilt, wird das Modell reparametrisiert mit Parametern $\lambda = \lambda_1$ und $\theta = \lambda_2/\lambda_1$. Zeigen Sie, dass die gemeinsame Log-Likelihood von λ und θ folgende Form hat ($D := D_1 + D_2$):

$$l(\lambda, \theta) = D \log(\lambda) + D_2 \log(\theta) - \lambda Y_1 - \theta \lambda Y_2.$$

► Die alten Parameter können wir aus den neuen mit

$$\lambda_1 = \lambda \quad \text{und} \quad \lambda_2 = \lambda \theta$$

erhalten. Setzen wir dies in die gemeinsame Log-Likelihood

$$l(\lambda_1, \lambda_2) = D_1 \log(\lambda_1) - \lambda_1 Y_1 + D_2 \log(\lambda_2) - \lambda_2 Y_2$$

von λ_1 und λ_2 ein, so erhalten wir die Log-Likelihood von λ und θ als

$$\begin{aligned} l(\lambda, \theta) &= D_1 \log(\lambda) - \lambda Y_1 + D_2 \log(\lambda) + D_2 \log(\theta) - \lambda \theta Y_2 \\ &= D \log(\lambda) + D_2 \log(\theta) - \lambda Y_1 - \lambda \theta Y_2. \end{aligned}$$

- c) Bestimmen Sie den ML-Schätzer $(\hat{\lambda}, \hat{\theta})$, die beobachtete Fisher-Information $I(\hat{\lambda}, \hat{\theta})$ und einen Ausdruck für die beiden Profil-Log-Likelihoodfunktionen $l_p(\lambda) = l(\lambda, \hat{\theta}(\lambda))$ und $l_p(\theta) = l(\hat{\lambda}(\theta), \theta)$.

► Zunächst erhalten wir über die (multivariate) Invarianz der ML-Schätzung und den Ergebnissen aus 1a)

$$(\hat{\lambda}, \hat{\theta}) = \left(\frac{D_1}{Y_1}, \frac{D_2 Y_1}{D_1 Y_2} \right).$$

Die Score-Funktion ist

$$\mathbf{S}(\lambda, \theta) = \begin{pmatrix} \frac{d}{d\lambda} l(\lambda, \theta) \\ \frac{d}{d\theta} l(\lambda, \theta) \end{pmatrix} = \begin{pmatrix} \frac{D}{\lambda} - Y_1 - \theta Y_2 \\ \frac{D_2}{\theta} - \lambda Y_2 \end{pmatrix},$$

sodass sich die Fisher-Information zu

$$\mathbf{I}(\lambda, \theta) = - \begin{pmatrix} \frac{d^2}{d\lambda^2} l(\lambda, \theta) & \frac{d^2}{d\lambda d\theta} l(\lambda, \theta) \\ \frac{d^2}{d\lambda d\theta} l(\lambda, \theta) & \frac{d^2}{d\theta^2} l(\lambda, \theta) \end{pmatrix} = \begin{pmatrix} \frac{D}{\lambda^2} & Y_2 \\ Y_2 & \frac{D_2}{\theta^2} \end{pmatrix}$$

ergibt. Einsetzen der ML-Schätzer liefert die beobachtete Fisher-Information

$$\mathbf{I}(\hat{\lambda}, \hat{\theta}) = \begin{pmatrix} \frac{DY_1^2}{D_1^2} & Y_2 \\ Y_2 & \frac{D_2^2 Y_2^2}{D_2 Y_1^2} \end{pmatrix}$$

Für die Profil-Log-Likelihoodfunktionen berechnet man zunächst die Maxima der (Log-)Likelihoodfunktion bei festgehaltenem λ bzw. θ , $\hat{\theta}(\lambda)$ und $\hat{\lambda}(\theta)$. In diesem Fall geht das am einfachsten über das (separate) Lösen der zwei Score-Gleichungen $\frac{d}{d\theta} l(\lambda, \theta) = 0$ und $\frac{d}{d\lambda} l(\lambda, \theta) = 0$ nach θ bzw. λ ; die Lösungen sind

$$\hat{\theta}(\lambda) = \frac{D_2}{\lambda Y_2} \quad \text{und} \quad \hat{\lambda}(\theta) = \frac{D}{Y_1 + \theta Y_2}.$$

Dass diese lokalen Extrema wirklich Maxima der Funktionen sind, erkennt man an den stets positiven Diagonaleinträgen der Fisher-Information. Nun können wir die Profil-Log-Likelihoodfunktionen durch Einsetzen von $\hat{\theta}(\lambda)$ und $\hat{\lambda}(\theta)$ in die Log-Likelihood-Funktion bekommen. Das Ergebnis ist

$$l_p(\lambda) = D \log(\lambda) + D_2 \log(D_2) - D_2 \log(\lambda Y_2) - \lambda Y_1 - D_2$$

bzw. $l_p(\theta) = D \log(D) - D \log(Y_1 + \theta Y_2) + D_2 \log(\theta) - D.$

- d) Skizzieren Sie die beiden Funktionen $l_p(\lambda)$ und $l_p(\theta)$ und erstellen Sie in R einen Konturplot der relativen Log-Likelihood $\tilde{l}(\lambda, \theta)$ mit der Funktion `contour`. Fügen Sie zusammengehörige Werte $(\lambda, \hat{\theta}(\lambda))$ und $(\hat{\lambda}(\theta), \theta)$ dem Konturplot hinzu.

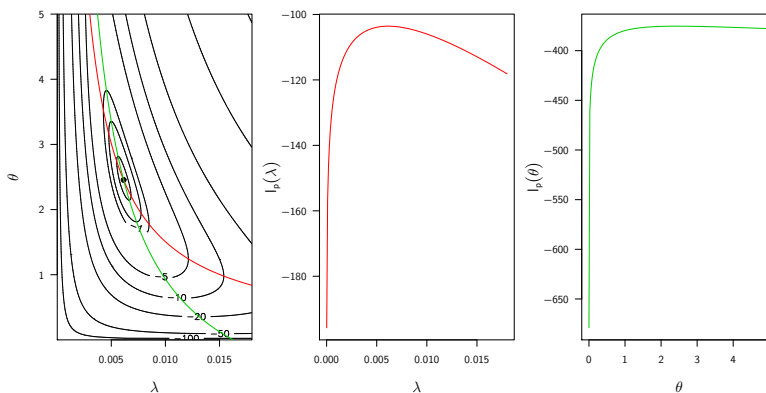
```
► ## Daten sind
> Y1 <- 2768.9
> D1 <- 17
> Y2 <- 1857.5
> D2 <- 28
> ## Log-Likelihood ist
> loglik <- function(param){
+   lambda <- param[1]
+   theta <- param[2]
+   D <- D1 + D2
+   return(D * log(lambda) + D2 * log(theta) - lambda * Y1 - theta * lambda * Y2)
+ }
```

```

> ## ML-Schätzer ist
> (ml <- c(D1/Y1, (Y1*D2)/(Y2*D1)))
> ## Normierte Log-Likelihood ist
> norm.loglik <- function(param){
+   return(loglik(param) - loglik(ml))
+ }
> ## Plot der normierten Log-Likelihood:
> lambda.grid <- seq(1e-5, 0.018, length = 200)
> theta.grid <- seq(1e-5, 5, length = 200)
> grid <- expand.grid(lambda = lambda.grid, theta = theta.grid)
> werte <- matrix(data = apply(grid, 1, norm.loglik), nrow = length(lambda.grid))

> par(mfrow = c(1,3))
> ## Konturplot der normierten Log-Likelihood:
> contour(lambda.grid, theta.grid, werte,
+   xlab = expression(lambda), ylab = expression(theta),
+   levels = -c(0.1, 0.5, 1, 5, 10, 20, 50, 100, 500, 1000, 1500, 2000),
+   xaxs = "i", yaxs = "i")
> points(ml[1], ml[2], pch = 19)
> ## Verlauf der Profile-LLs in Konturplot einzeichnen:
> lines(lambda.grid, D2 / (lambda.grid * Y2), col = 2)
> lines((D1 + D2)/(Y1 + theta.grid*Y2), theta.grid, col = 3)
> ## Profile-Log-Likelihoods:
> prof.lambda <- function(lambda){
+   return(D1 * log(lambda) - lambda * Y1)
+ }
> prof.theta <- function(theta){
+   return(-(D1 + D2) * log(Y1 + theta*Y2) + D2 * log(theta))
+ }
> ## separat plotten:
> plot(lambda.grid, prof.lambda(lambda.grid), xlab = expression(lambda),
+   ylab = expression(l[p](lambda)), col = 2, type = "l")
> plot(theta.grid, prof.theta(theta.grid), xlab = expression(theta),
+   ylab = expression(l[p](theta)), col = 3, type = "l")

```



- e) Berechnen Sie ein 95%-Wald-Intervall für $\log \theta$ basierend auf der Profil-Log-Likelihood und folgern Sie, ob $\lambda_1 = \lambda_2$ verworfen werden kann.

► Zunächst leiten wir den Standardfehler für $\hat{\theta}$ her. Leiten wir die negative Profil-Log-Likelihood $-l_p(\theta)$ zweimal ab, so erhalten wir die negative Krümmung von $l_p(\theta)$,

$$I_p(\theta) = -\frac{d}{d\theta} \left(-\frac{DY_2}{Y_1 + \theta Y_2} + \frac{D_2}{\theta} \right) = -\frac{DY_2^2}{(Y_1 + \theta Y_2)^2} + \frac{D_2}{\theta^2}.$$

Am ML-Schätzer $\hat{\theta} = D_2 Y_1 / (D_1 Y_2)$ (der natürlich auch die Profil-Log-Likelihood maximiert) beträgt die negative Krümmung also

$$I_p(\hat{\theta}) = \frac{D_1^3 Y_2^2}{D Y_1^2 D_2}.$$

(Nach Satz 4.1 erhielten wir die gleiche Formel, wenn wir die beobachtete Fisher-Information $\mathbf{I}(\hat{\lambda}, \hat{\theta})$ invertieren und anschließend den Kehrwert des zweiten Diagonalelements nehmen würden.) Der Standardfehler von $\hat{\theta}$ ist somit

$$\text{se}(\hat{\theta}) = (I_p(\hat{\theta}))^{-\frac{1}{2}} = \frac{Y_1 \sqrt{D_2 D}}{Y_2 D_1 \sqrt{D_1}}$$

und wiederum mit der Delta-Regel bekommen wir den Standardfehler für $\hat{\phi} = \log(\hat{\theta})$:

$$\text{se}(\hat{\phi}) = \text{se}(\hat{\theta}) \cdot (\hat{\theta})^{-1} = \frac{Y_1 \sqrt{D_2 D}}{Y_2 D_1 \sqrt{D_1}} \cdot \frac{D_1 Y_2}{D_2 Y_1} = \sqrt{\frac{D}{D_1 D_2}}$$

Das 95%-Wald-Konfidenzintervall für ϕ ist somit

$$\left[\log(D_2 Y_1 / (D_1 Y_2)) \pm 1.96 \cdot \sqrt{\frac{D}{D_1 D_2}} \right]$$

und ergibt sich für die konkreten Daten zu $[0.296, 1.501]$. Da ϕ die log-relative Rate ist und die Null nicht im Konfidenzintervall enthalten ist, wird der entsprechende Test mit Signifikanzniveau 5% die Hypothese der Gleichheit $\lambda_1 = \lambda_2$ ablehnen.

2. Sei $(x_1, y_1), \dots, (x_n, y_n)$ eine Zufallsstichprobe aus einer bivariaten Normalverteilung mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

- a) Bestimmen Sie den ML-Schätzer $(\hat{\sigma}_{ML}^2, \hat{\rho}_{ML})$.

► Nach Einsetzen von $|\Sigma| = \sigma^4(1 - \rho^2)$ und

$$\Sigma^{-1} = \frac{1}{\sigma^2(1 - \rho^2)} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

in die Log-Likelihoodfunktion mit allgemeiner Kovarianz,

$$\sum_{i=1}^n -\frac{1}{2} \left\{ \log |\Sigma| + (x_i, y_i) \Sigma^{-1} \begin{pmatrix} x_i \\ y_i \end{pmatrix} \right\},$$

erhält man die Log-Likelihood von σ^2 und ρ als

$$l(\sigma^2, \rho) = -\frac{n}{2} \log(\sigma^4(1 - \rho^2)) - \frac{1}{2\sigma^2(1 - \rho^2)} Q(\rho),$$

wobei die quadratische Form $\sum_{i=1}^n x_i^2 - 2\rho x_i y_i + y_i^2$ als $Q(\rho)$ abgekürzt wird. Die Komponenten der Score-Funktion sind

$$\begin{aligned} \frac{d}{d\sigma^2} l(\sigma^2, \rho) &= -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4(1 - \rho^2)} Q(\rho) \\ \text{und} \quad \frac{d}{d\rho} l(\sigma^2, \rho) &= \frac{n\rho}{1 - \rho^2} + \frac{1}{\sigma^2(1 - \rho^2)} \left\{ \sum_{i=1}^n x_i y_i - \frac{\rho}{1 - \rho^2} Q(\rho) \right\}. \end{aligned}$$

Die erste Score-Gleichung lässt sich zunächst umformen zu

$$\sigma^2 = \frac{1}{2n(1 - \rho^2)} Q(\rho),$$

was nach Einsetzen in die zweite Score-Gleichung auf

$$\frac{2 \sum_{i=1}^n x_i y_i}{Q(\rho)} = \frac{\rho}{1 - \rho^2}$$

führt. So gelangt man zu den ML-Schätzern

$$\hat{\rho}_{ML} = \frac{\sum_{i=1}^n x_i y_i}{\frac{1}{2} \sum_{i=1}^n x_i^2 + y_i^2}$$

$$\text{und schließlich} \quad \hat{\sigma}_{ML}^2 = \frac{1}{2n} \sum_{i=1}^n x_i^2 + y_i^2.$$

b) Zeigen Sie, dass

$$\mathbf{I}(\hat{\sigma}_{ML}^2, \hat{\rho}_{ML}) = \begin{pmatrix} \frac{n}{\hat{\sigma}_{ML}^4} & -\frac{n\hat{\rho}_{ML}}{\hat{\sigma}_{ML}^2(1 - \hat{\rho}_{ML}^2)} \\ -\frac{n\hat{\rho}_{ML}}{\hat{\sigma}_{ML}^2(1 - \hat{\rho}_{ML}^2)} & \frac{n(1 + \hat{\rho}_{ML}^2)}{(1 - \hat{\rho}_{ML}^2)^2} \end{pmatrix}.$$

► Die Einträge der Fisher-Information $\mathbf{I}(\sigma^2, \rho)$ ergeben sich (durch mühsame Anwendung der bekannten Ableitungsregeln oder aber Benutzung eines Computer-Algebra-Systems wie etwa *Maxima* (?) zu

$$\begin{aligned} -\frac{d^2}{d(\sigma^2)^2} l(\sigma^2, \rho) &= \frac{Q(\rho) - n\sigma^2(1 - \rho^2)}{\sigma^6(1 - \rho^2)}, \\ -\frac{d^2}{d\sigma^2 d\rho} l(\sigma^2, \rho) &= \frac{\sum_{i=1}^n x_i y_i (1 - \rho^2) - \rho Q(\rho)}{\sigma^4(1 - \rho^2)^2} \\ \text{und} \quad -\frac{d^2}{d\rho^2} l(\sigma^2, \rho) &= \frac{(1 - \rho^2)Q(\rho) - n\sigma^2(1 - \rho^4) - 4\rho \sum_{i=1}^n (\rho y_i - x_i)(\rho x_i - y_i)}{\sigma^2(1 - \rho^2)^3}. \end{aligned}$$

Setzt man in diese Formeln die ML-Schätzer ein, erreicht man nach einigen Umformungen die gewünschte Form der beobachteten Fisher-Information $\mathbf{I}(\hat{\sigma}_{ML}^2, \hat{\rho}_{ML})$. Dabei könnte z. B. die Identität

$$\sum_{i=1}^n (\hat{\rho}_{ML} y_i - x_i)(\hat{\rho}_{ML} x_i - y_i) = n \hat{\rho}_{ML} \hat{\sigma}_{ML}^2 (\hat{\rho}_{ML}^2 - 1)$$

hilfreich sein.

- c) Zeigen Sie, dass der Standardfehler von $\hat{\rho}_{ML}$ gleich

$$\text{se}(\hat{\rho}_{ML}) = \frac{1 - \hat{\rho}_{ML}^2}{\sqrt{n}}$$

ist.

- Mittels der Rechenregel zur Inversion von (2×2) -Matrizen

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

sehen wir, dass das interessierende Diagonalelement I^{22} der inversen beobachteten Fisher-Information $\mathbf{I}(\hat{\sigma}_{ML}^2, \hat{\rho}_{ML})^{-1}$ gleich

$$\left(\frac{n^2(1 + \hat{\rho}_{ML}^2)}{\hat{\sigma}_{ML}^4(1 - \hat{\rho}_{ML}^2)^2} - \frac{n^2\hat{\rho}_{ML}^2}{\hat{\sigma}_{ML}^4(1 - \hat{\rho}_{ML}^2)^2} \right)^{-1} \cdot \frac{n}{\hat{\sigma}_{ML}^4} = \frac{(1 - \hat{\rho}_{ML}^2)^2}{n}$$

ist. Der Standardfehler ist die Wurzel daraus, was gezeigt werden sollte.

3. Seien $X_1 \sim \text{Bin}(n_1, \pi_1)$ und $X_2 \sim \text{Bin}(n_2, \pi_2)$ unabhängige Zufallsvariablen. Zur Untersuchung der Nullhypothese $H_0 : \pi_1 = \pi_2$ betrachtet man häufig das *relative Risiko* („Relative risk“) $\theta = \pi_1/\pi_2$ oder das *logarithmierte relative Risiko* $\psi = \log(\theta)$.

- a) Berechnen Sie den Standardfehler des ML-Schätzers des logarithmierten relativen Risikos. Gehen Sie dabei wie in Beispiel 4.8 vor.

- Nach Beispiel 2.9 ist der Standardfehler des ML-Schätzers $\hat{\pi}_i = X_i/n_i$ in der i -ten Gruppe ($i = 1, 2$) gleich

$$\text{se}(\hat{\pi}_i) = I(\hat{\pi}_i)^{-\frac{1}{2}} = \sqrt{\frac{\hat{\pi}_i(1 - \hat{\pi}_i)}{n_i}}.$$

Da X_1 und X_2 unabhängig sind, sind auch die ML-Schätzer unabhängig. Asymptotisch gilt also

$$\begin{pmatrix} \hat{\pi}_1 \\ \hat{\pi}_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}, \begin{pmatrix} \frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} & 0 \\ 0 & \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2} \end{pmatrix} \right).$$

Zur Berechnung des Standardfehlers des ML-Schätzers des logarithmierten relativen Risikos

$$\hat{\psi}_{ML} = g(\hat{\pi}_1, \hat{\pi}_2) = \log(\hat{\pi}_1) - \log(\hat{\pi}_2)$$

können wir daher die multivariate Delta-Regel A.2.5 anwenden. Als Varianzschätzung für $\hat{\psi}_{ML}$ ergibt sich damit

$$\text{se}(\hat{\psi}_{ML})^2 = \left(1/\hat{\pi}_1, 1/\hat{\pi}_2\right) \begin{pmatrix} \frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} & 0 \\ 0 & \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2} \end{pmatrix} \begin{pmatrix} 1/\hat{\pi}_1 \\ 1/\hat{\pi}_2 \end{pmatrix} = \frac{1-\hat{\pi}_1}{\hat{\pi}_1 n_1} + \frac{1-\hat{\pi}_2}{\hat{\pi}_2 n_2},$$

und die Wurzel hieraus ist der gesuchte Standardfehler.

- b) Berechnen Sie für die Daten aus Tabelle 4.1 im Buch ein 95%-Konfidenzintervall für das relative Risiko.

► In der Diuretika-Gruppe wird der Anteil mit $\hat{\pi}_1 = 6/108 = 0.056$, in der Kontroll-Gruppe mit $\hat{\pi}_2 = 2/103 = 0.019$ geschätzt. Die Schätzung für das log-relative Risiko ist demnach

$$\hat{\psi}_{ML} = \log(6/108) - \log(2/103) = 1.051$$

mit 95%-Wald-Konfidenzintervall

$$[1.051 \pm 1.96 \cdot 0.648] = [-0.218, 2.321].$$

Rücktransformation mit exp auf die Originalskala liefert uns dann das Konfidenzintervall $[0.804, 10.183]$ für das relative Risiko $\theta = \pi_1/\pi_2$.

- c) Berechnen Sie auch die Profil-Likelihood und das zugehörige 95%-Konfidenzintervall für das relative Risiko.

► Die Log-Likelihood von $\theta = \pi_1/\pi_2$ und π_2 ist

$$l(\theta, \pi_2) = x_1 \log(\theta) + (n_1 - x_1) \log(1 - \theta\pi_2) + (x_1 + x_2) \log(\pi_2) + (n_2 - x_2) \log(1 - \pi_2).$$

Für die Profil-Log-Likelihood brauchen wir das Maximum bzgl. π_2 bei festem θ . Dazu suchen wir die Nullstelle von

$$\frac{d}{d\pi_2} l(\theta, \pi_2) = \frac{\theta(n_1 - x_1)}{\theta\pi_2 - 1} + \frac{x_1 + x_2}{\pi_2} + \frac{n_2 - x_2}{\pi_2 - 1},$$

wozu wir letztendlich die quadratische Gleichung

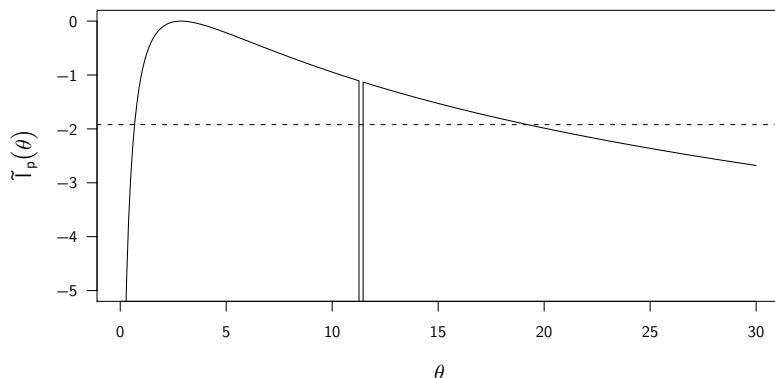
$$\pi_2^2 \theta (n_1 + n_2) - \pi_2 (\theta (n_1 + x_2) + n_2 + x_1) + x_1 + x_2 = 0$$

nach π_2 lösen müssen. Einfacher ist es aber in diesem Fall, die Maximierung numerisch vorzunehmen. Da wir die Ableitung angeben können, bietet sich ein Gradientenverfahren an:

```

> ## Daten sind
> x <- c(6, 2)
> n <- c(108, 103)
> ## ML-Schätzer
> piMl <- x/n
> thetaMl <- piMl[1] / piMl[2]
> ## Log-Likelihood von theta und pi2
> loglik <- function(theta, pi2)
+ {
+   pi <- c(theta * pi2, pi2)
+   sum(dbinom(x, n, pi, log = TRUE))
+ }
> ## Gradient in pi2 implementieren
> grad <- function(theta, pi2)
+ {
+   (theta * (n[1] - x[1])) / (theta * pi2 - 1) +
+     sum(x) / pi2 +
+     (n[2] - x[2]) / (pi2 - 1)
+ }
> ## damit Profil-Log-Likelihood von theta:
> profilLoglik <- function(theta)
+ {
+   res <- theta
+   eps <- sqrt(.Machine$double.eps)
+
+   for(i in seq_along(theta)){ # Funktion soll mit Vektoren umgehen können
+     optimResult <-
+       optim(par = 0.5,
+             fn = function(pi2) loglik(theta[i], pi2),
+             gr = function(pi2) grad(theta[i], pi2),
+             method = "L-BFGS-B", lower = eps, upper = 1/theta[i] - eps,
+             control = list(fnscale = -1))
+     if(optimResult$convergence == 0){ # ist der Algorithmus konvergiert?
+       res[i] <- optimResult$value # nur dann Wert (nicht Parameter!) speichern
+     } else {
+       res[i] <- NA # sonst NA ausgeben
+     }
+   }
+
+   return(res)
+ }
> ## Normierte Profil-Log-Likelihood plotten:
> thetaGrid <- seq(from = 0.1, to = 30, length = 309)
> normProfVals <- profilLoglik(thetaGrid) - profilLoglik(thetaMl)
> plot(thetaGrid, normProfVals,
+       type = "l", ylim = c(-5, 0),
+       xlab = expression(theta), ylab = expression(tilde(l)[p](theta))
+       )
> ## Cutpoint einzeichnen:
> abline(h = - 1/2 * qchisq(0.95, 1), lty = 2)

```

Die Ausreißer nach unten sind Artefakte der numerischen Optimierung. Mit der Funktion aus Aufgabe 16d) in Kapitel 3 können wir nun das Profil-Likelihood-Intervall erhalten, dessen Lage man schon ungefähr durch das Einzeichnen des Schwellenwertes erkennen kann.

```
> thetaProfilCi <-
+   likelihoodCi(alpha = 0.05, loglik = profilLoglik, thetaMl = thetaMl,
+               lower = 0.01, upper = 20)
> thetaProfilCi
      lower      upper
0.6766635 19.2217991
```

Im Vergleich zum Wald-Konfidenzintervall mit Log-Transformation ist es fast doppelt so breit und drückt damit eine stärkere Unsicherheit in der Schätzung von θ aus. Auch dieses Mal wird die Eins überdeckt, sodass kein signifikant erhöhtes Risiko (auf dem 5%-Signifikanzniveau) bei Einnahme von Diuretika vorliegt.

4. Das AB0-System wurde 1901 von KARL LANDSTEINER beschrieben, wofür er 1930 den Nobelpreis für Medizin bekam. Es ist das wichtigste Blutgruppenmerkmal bei der Bluttransfusion und umfasst vier verschiedene Hauptgruppen: A, B, AB und 0.

Die Blutgruppenfaktoren A und B sind dominant gegenüber dem Blutgruppenfaktor 0. Die Blutgruppenfaktoren A und B verhalten sich kodominant zueinander. Der Blutgruppenfaktor 0 verhält sich rezessiv gegenüber den Blutgruppenfaktoren A und B. Hierdurch ergibt sich für die Blutgruppe A ein Genotyp von AA oder A0, für Blutgruppe B ein Genotyp von BB oder B0, für Blutgruppe AB ein Genotyp von AB und für Blutgruppe 0 ein Genotyp von 00.

Seien p , q und r die Anteile der Allele A, B und 0 in einer Population, es gilt also die Restriktion $p + q + r = 1$. Dann gilt unter bestimmten Annahmen das Schema in Tabelle 4.1, in der auch die Ergebnisse einer Stichprobe vom Umfang $n = 435$ enthalten sind.

Tab. 4.1: Schema der Blutgruppen mit Wahrscheinlichkeiten und Stichprobenergebnis ($n = 435$).

Blutgruppe	Wahrscheinlichkeit	Beobachtung
A={AA,A0}	$\pi_1 = p^2 + 2pr$	$x_1 = 182$
B={BB,B0}	$\pi_2 = q^2 + 2qr$	$x_2 = 60$
AB={AB}	$\pi_3 = 2pq$	$x_3 = 17$
0={00}	$\pi_4 = r^2$	$x_4 = 176$

- a) Stellen Sie die Likelihoodfunktion von $\boldsymbol{\theta} = (p, q)^T$ auf. Nehmen Sie dafür an, dass $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ eine Realisation aus einer Multinomialverteilung mit Parametern $n = 435$ und $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)^T$ ist.

► *Wir nehmen also an, dass*

$$\mathbf{X} = (X_1, X_2, X_3, X_4)^T \sim M_4 \left(n = 435, \boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{pmatrix} \right)$$

gilt. Die Wahrscheinlichkeitsfunktion von \mathbf{X} ist demnach

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!} \prod_{i=1}^4 \pi_i^{x_i}$$

und durch Einsetzen der Parametrisierung der Klassenwahrscheinlichkeiten π_i , $i = 1, \dots, 4$, aus Tabelle 4.1 und Logarithmieren gelangen wir zur Log-Likelihood von p und q ,

$$\begin{aligned} l(p, q) &= \sum_{i=1}^4 x_i \log(\pi_i) \\ &= x_1 \log(p^2 + 2p(1-p-q)) + x_2 \log(q^2 + 2q(1-p-q)) \\ &\quad + x_3 \log(2pq) + x_4 \log((1-p-q)^2). \end{aligned}$$

- b) Berechnen Sie numerisch die ML-Schätzer von p und q mit zugehörigen Standardfehlern. Berechnen Sie daraus den ML-Schätzer von r mit zugehörigem Standardfehler.

```
► ## beobachtete Realisation
> data <- c(182, 60, 17, 176)
> n <- sum(data)
> ## Loglikelihood ist
> loglik <- function(theta, data) {
+   p <- theta[1]
+   q <- theta[2]
+   r <- 1-p-q
+
+   ## Prüfen, ob Parameter zulässig sind:
+   if ((p>0) && (p<1) && (r>0) && (r<1) && (q>0) && (q<1)) {
```

```

+       probs <- c(p^2+2*p*r,q^2+2*q*r, 2*p*q, r^2)
+       return(dmultinom(data,prob=probs,size=sum(data),log=T))
+     } else {
+       return(NA)
+     }
+ }
> ## Log-Likelihood numerisch optimieren
> optimResult <- optim(c(0.1,0.3), loglik,
+       control = list(fnscale=-1),
+       hessian = TRUE,
+       data = data)
> optimResult[["convergence"]]
[1] 0
> ## und ML-Schätzer sowie Standardfehler extrahieren
> (thetaML <- optimResult$par)
[1] 0.26442773 0.09317313
> thetaCov <- solve(- optimResult$hessian)
> (thetaSe <- sqrt(diag(thetaCov)))
[1] 0.01624791 0.01011839
> rML <- 1 - sum(thetaML)

```

Die numerische Analyse liefert also $\hat{p}_{ML} = 0.264$ und $\hat{q}_{ML} = 0.093$ mit zugehörigen Standardfehlern $se(\hat{p}_{ML}) = 0.016$ und $se(\hat{q}_{ML}) = 0.01$. Aus der Invarianz des ML-Schätzers folgt sofort $\hat{r}_{ML} = 1 - \hat{p}_{ML} - \hat{q}_{ML} = 0.642$. Wie aber lautet $se(\hat{r}_{ML})$? Durch die Restriktion

$$r = 1 - p - q = 1 + \begin{pmatrix} -1 \\ -1 \end{pmatrix}^T \begin{pmatrix} p \\ q \end{pmatrix} = g(p, q)$$

können wir die multivariate Delta-Regel im Spezialfall der linearen Transformation der Parameterkomponenten $g(\boldsymbol{\theta}) = \mathbf{a}^T \cdot \boldsymbol{\theta} + b$ anwenden: Dann ist $D(\hat{\boldsymbol{\theta}}_{ML}) = \mathbf{a}^T$, somit

$$se(g(\hat{\boldsymbol{\theta}}_{ML})) = \sqrt{\mathbf{a}^T \mathbf{I}(\hat{\boldsymbol{\theta}}_{ML})^{-1} \mathbf{a}}.$$

In diesem Fall also mit $\boldsymbol{\theta} = (p, q)^T$, $\mathbf{a}^T = (-1, -1)$ und $b = 1$:

$$se(\hat{r}_{ML}) = se(g(p, q)) = \sqrt{(-1, -1) \mathbf{I}(\hat{p}_{ML}, \hat{q}_{ML})^{-1} \begin{pmatrix} -1 \\ -1 \end{pmatrix}} = \sqrt{\sum_{i,j=1}^2 (\mathbf{I}(\hat{p}_{ML}, \hat{q}_{ML})^{-1})_{ij}}$$

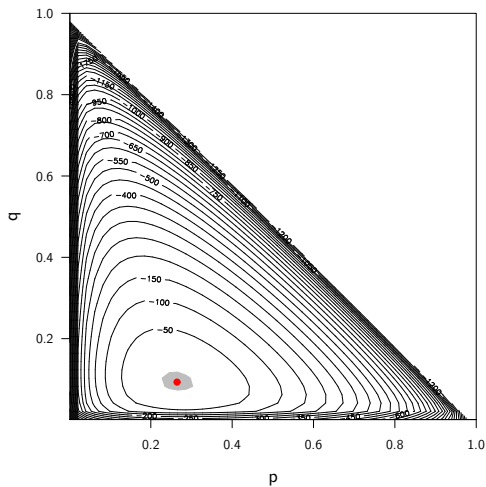
Es ergibt sich $se(\hat{r}_{ML}) = 0.018$.

- c) Stellen Sie die relative Log-Likelihoodfunktion graphisch dar und skizzieren sie die 95%-Konfidenzregion für $\boldsymbol{\theta}$.

```

► ## Grid mit Werten der relativen Log-Likelihood auffüllen
> gridSize <- 50
> eps <- 1e-3
> loglikgrid <- matrix(NA,gridSize,gridSize)
> p <- seq(eps,1,length=gridSize)
> q <- seq(eps,1,length=gridSize)
> for (i in 1:length(p)) {
+   for (j in 1:length(q)) {
+     loglikgrid[i,j] <-
+       loglik(c(p[i],q[j]),data=data) - loglik(thetaML, data = data)
+   }
+ }
> ## Plotten
> contour(p,q,
+         loglikgrid,
+         nlevels = 50,
+         xlab=expression (p),ylab= expression (q), xaxs = "i", yaxs = "i")
> ## Konfidenzregion und ML-Schätzer hinzufügen:
> region <- contourLines(x = p, y = q, z = loglikgrid, levels = log (0.05))[[1]]
> polygon(region$x, region$y, density = NA, col = "gray")
> points(thetaML[1], thetaML[2], pch = 19, col = 2)

```



- d) Verwenden Sie die χ^2 - und G^2 -Teststatistik um die Plausibilität der getroffenen Modellannahmen zu untersuchen.

► Die im Modell erwarteten Anzahlen e_i sind die Erwartungswerte nach Einsetzen der ML-Schätzer, d. h.

$$e_1 = n(\hat{p}_{ML}^2 + 2\hat{p}_{ML}\hat{r}_{ML}) = 178.201,$$

$$e_2 = n(\hat{q}_{ML}^2 + 2\hat{q}_{ML}\hat{r}_{ML}) = 55.85,$$

$$e_3 = 2n\hat{p}_{ML}\hat{q}_{ML} = 21.435$$

$$\text{und } e_4 = n\hat{r}_{ML}^2 = 179.514.$$

In R können wir nun einfach die Anpassungsstatistiken berechnen:

```

> ## Werte
> (x <- data)
[1] 182 60 17 176
> (e <- n * c(thetaM1[1]^2 + 2 * thetaM1[1] * rM1,
+             thetaM1[2]^2 + 2 * thetaM1[2] * rM1,
+             2 * prod(thetaM1),
+             rM1^2
+             ))
[1] 178.20137 55.84961 21.43468 179.51434
> ## Statistiken
> (G2 <- 2 * sum(x * log(x / e)))
[1] 1.438987
> (Chi2 <- sum((x - e)^2 / x))
[1] 1.593398

```

Bei $k = 4$ Kategorien, also 3 freien Wahrscheinlichkeiten, und $r = 2$ freien Parametern (p und q) sind beide Statistiken unter der Nullhypothese, dass das Modell stimmt, asymptotisch $\chi^2(1)$ -verteilt. Das zugehörige 95%-Quantil 3.84 wird von beiden Teststatistiken nicht überschritten, sodass H_0 auf dem 5%-Signifikanzniveau nicht abgelehnt werden kann – die Modellanpassung ist also recht gut.

5. Sei $T \sim t(n-1)$.

a) Berechnen Sie analytisch die Dichtefunktion der Zufallsvariable

$$W = n \log \left(1 + \frac{T^2}{n-1} \right),$$

vergleiche Beispiel 4.15, und vergleichen Sie diese graphisch mit der Dichtefunktion der $\chi^2(1)$ -Verteilung für verschiedene Werte von n .

► $W = g(T^2)$ ist eine eindeutige Transformation von T^2 , wobei die Umkehrfunktion

$$g^{-1}(w) = (n-1)(\exp(w/n) - 1)$$

ist. Für die Anwendung des Dichtetransformationssatzes fehlt uns somit nur noch die Dichte von T^2 . Da dies keine eindeutige Transformation von T ist, leiten wir zunächst die Verteilungsfunktion von T^2 her. Sei $x \geq 0$, dann ist

$$\begin{aligned}
 F_{T^2}(x) &= P(T^2 \leq x) \\
 &= P(|T| \leq \sqrt{x}) \\
 &= P(-\sqrt{x} \leq T \leq \sqrt{x}) \\
 &= F_T(\sqrt{x}) - F_T(-\sqrt{x}) \\
 &= 2F_T(\sqrt{x}) - 1,
 \end{aligned}$$

wobei die letzte Umformung aus der Symmetrie der Standard- t -Dichte f_T um $x = 0$ folgt. Die Dichte von T^2 ist somit

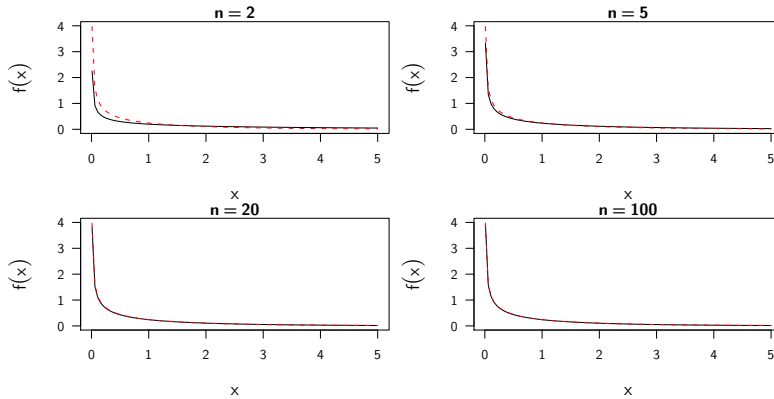
$$f_{T^2}(x) = \frac{d}{dx} F_{T^2}(x) = 2f_T(\sqrt{x}) \cdot \frac{1}{2}x^{-\frac{1}{2}} = \frac{f_T(\sqrt{x})}{\sqrt{x}},$$

sodass die Dichte von W gleich

$$\begin{aligned} f_W(x) &= f_{T^2}(g^{-1}(x)) \left| \frac{d}{dx} g^{-1}(x) \right| \\ &= \frac{f_T\left(\sqrt{(n-1)(\exp(x/n) - 1)}\right)}{\sqrt{(n-1)(\exp(x/n) - 1)}} \frac{n-1}{n} \exp(x/n) \end{aligned}$$

ist. Dies reicht für das Zeichnen der Dichte bereits aus:

```
> ## Dichte implementieren
> densityW <- function(x, n)
+ {
+   y <- sqrt((n - 1) * (exp(x / n) - 1))
+   dt(y, df = n - 1) / y * (n - 1) / n * exp(x / n)
+ }
> ## Test durch Vergleich mit Histogramm:
> ## set.seed(1234)
>
> ## n <- 10
> ## m <- 1e+5
> ## T2 <- rt(m, df = n - 1)^2
> ## W <- n * log(1 + T2 / (n - 1))
> ## hist(W[W < 5], breaks = 100, prob = TRUE)
> grid <- seq(from = 0.01, to = 5, length = 101)
> ## lines (grid, densityW(grid, n = n), col = 2)
>
> ## und für verschiedene Werte von n plotten
> par(mfrow = c(2, 2))
> for(n in c(2, 5, 20, 100)){
+   ## Dichte von W
+   plot(grid, densityW(grid, n = n), type = "l",
+        xlab = expression(x), ylab = expression(f(x)),
+        ylim = c(0, 4), main = paste("n =", n)
+   )
+   ## Dichte von chi^2(1)
+   lines(grid, dchisq(grid, df = 1), lty = 2, col = 2)
+ }
```



Bereits bei $n = 20$ sind kaum noch Unterschiede der Dichten erkennbar!

b) Zeigen Sie, dass W für $n \rightarrow \infty$ tatsächlich einer $\chi^2(1)$ -Verteilung folgt.

► Im Folgenden wird eine Beweisskizze gegeben. Ausgehend von $T \xrightarrow{D} N(0, 1)$ für $n \rightarrow \infty$ folgt $T^2 \xrightarrow{D} \chi^2(1)$. Weiterhin wird die Transformation g für großes n der Identität näher kommen, da

$$g(x) = \log \left(\left(1 + \frac{x}{n-1} \right)^n \right)$$

und $(1 + x/(n-1))^n \xrightarrow[n \rightarrow \infty]{} \exp(x)$ gilt. Insgesamt damit $W = g(T^2) \xrightarrow{D} \chi^2(1)$.

6. Betrachten Sie den χ^2 -Anpassungstest für k Kategorien bei n Beobachtungen. Seien

$$D_n = \sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}} \quad \text{und} \quad W_n = 2 \sum_{i=1}^k n_i \log \left(\frac{n_i}{np_{i0}} \right).$$

Zeigen Sie, dass $W_n - D_n \xrightarrow{P} 0$ für $n \rightarrow \infty$.

► Es soll also gezeigt werden, dass asymptotisch die G^2 - oder Likelihood-Quotienten-Statistik (W_n) und die χ^2 -Statistik (D_n) äquivalent sind, in dem Sinne, dass ihre Differenz nach Wahrscheinlichkeit gegen 0 konvergiert. (In der Notation aus dem Buch ist $x_i = n_i$ und $e_i = np_{i0}$.)

Zunächst muss dazu angenommen werden, dass die Differenzen aus beobachteten Werten n_i und erwarteten Werten np_{i0} für wachsendes n immer kleiner werden, d. h. $n_i/n \xrightarrow[n \rightarrow \infty]{} p_{i0}$. Unter der Nullhypothese $H_0 : P(X = i) = p_{i0}$ wird das Gesetz der großen Zahlen dies sicherstellen (im Sinne der schwachen Konvergenz). Dann gilt

$$W_n \approx 2 \sum_{i=1}^k n_i \log \left(\frac{n_i}{np_{i0}} \right) + 2 \sum_{i=1}^k (np_{i0} - n_i) = 2n \sum_{i=1}^k \psi(a_i + 1) p_{i0}$$

mit $\psi(s) = 1 - s + s \log(s)$ und $a_i = n_i/(np_{i0}) - 1$. Eine Taylor-Approximation zweiter Ordnung von ψ um $s = 1$ zeigt weiterhin

$$\psi(s) = \frac{1}{2}(s-1)^2 + o(|s-1|^2),$$

vgl. zur Landau-Notation Anhang B.9 und zur Taylor-Approximation Anhang B.6. Setzen wir dieses Ergebnis oben ein, erhalten wir

$$W_n \doteq 2n \sum_{i=1}^k \left(\frac{1}{2}a_i^2 + o(a_i^2) \right) p_{i0} = D_n + 2n \sum_{i=1}^k o(a_i^2) p_{i0}.$$

Die letzte Summe wird für $n \rightarrow \infty$ vernachlässigbar sein, da $a_i = n_i/(np_{i0}) - 1$ wegen $n_i/n \xrightarrow{P} p_{i0}$ nach Wahrscheinlichkeit gegen 0 konvergiert. Insgesamt folgt dann, dass $W_n - D_n \xrightarrow{P} 0$ falls $n_i/n \xrightarrow{P} p_{i0}$, $i = 1, \dots, k$, für $n \rightarrow \infty$.

7. Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer $\text{LN}(\mu, \sigma^2)$ -Verteilung.
- a) Bestimmen Sie den ML-Schätzer von μ und σ^2 . Benutzen Sie dabei den Zusammenhang zwischen den Dichten der Normalverteilung und der Lognormalverteilung. Berechnen Sie auch die zugehörigen Standardfehler.
- Sei X_1, \dots, X_n eine Stichprobe aus einer $\text{LN}(\mu, \sigma^2)$ -Verteilung und Y_1, \dots, Y_n eine Stichprobe aus einer $\text{N}(\mu, \sigma^2)$ -Verteilung. Die zugehörigen Dichtefunktionen lauten

$$\begin{aligned} f_x(x; \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(\frac{-(\log x - \mu)^2}{2\sigma^2}\right) \\ f_y(y; \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y - \mu)^2}{2\sigma^2}\right) \end{aligned}$$

Die Dichtefunktionen erfüllen

$$f_x(x; \mu, \sigma) = \frac{1}{x} f_y(\log x; \mu, \sigma),$$

was für die zugehörigen Log-Likelihoodfunktionen l_x und l_y impliziert dass

$$l_x(\mu, \sigma^2; x_1, \dots, x_n) = - \sum_{i=1}^n \log x_i + l_y(\mu, \sigma^2; \log x_1, \dots, \log x_n).$$

Da der erste Summand auf der rechten Seite nicht von μ und σ^2 abhängt, ergeben sich die ML-Schätzer

$$\begin{aligned} \hat{\mu}_{ML} &= n^{-1} \sum_{i=1}^n \log x_i \\ \hat{\sigma}_{ML}^2 &= n^{-1} \sum_{i=1}^n (\log x_i - \hat{\mu}_{ML})^2. \end{aligned}$$

Analoge Überlegungen ergeben für die beobachtete Fisher-Informationsmatrix:

$$\mathbf{I}(\hat{\mu}_{ML}, \hat{\sigma}_{ML}^2) = \begin{pmatrix} n/\hat{\sigma}_{ML}^2 & 0 \\ 0 & n/(2\hat{\sigma}_{ML}^4) \end{pmatrix}.$$

Gemäss Abschnitt 4.2 erhalten wir die Standardfehler als

$$\begin{aligned} \text{se}(\hat{\mu}_{ML}) &= \sqrt{[\mathbf{I}(\hat{\mu}_{ML}, \hat{\sigma}_{ML}^2)]_{11}^{-1}} \\ &= \hat{\sigma}_{ML}/\sqrt{n} \\ \text{se}(\hat{\sigma}_{ML}^2) &= \sqrt{[\mathbf{I}(\hat{\mu}_{ML}, \hat{\sigma}_{ML}^2)]_{22}^{-1}} \\ &= \hat{\sigma}_{ML}^2 \sqrt{2}/\sqrt{n}. \end{aligned}$$

- b) Bestimmen Sie die Profil-Log-Likelihood von μ bzw. σ^2 und stellen Sie diese graphisch für folgende Daten dar:

$$x = (225, 171, 198, 189, 189, 135, 162, 136, 117, 162).$$

Vergleichen Sie die Profil-Log-Likelihood mit ihrer quadratischen Approximation.

► Die gemeinsame Log-Likelihood Funktion von μ und σ^2 lautet

$$l_x(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - (2\sigma^2)^{-1} \sum_{i=1}^n (\log x_i - \mu)^2,$$

wobei additive Konstanten ignoriert worden sind. Einsetzen von $\hat{\sigma}^2(\mu) = n^{-1} \sum_{i=1}^n (\log x_i - \mu)^2$ und Ignorieren der nicht von μ abhängigen Terme ergibt die Profil-Log-Likelihood Funktion von μ :

$$l_p(\mu) = l_x(\mu, \hat{\sigma}^2(\mu)) = -\frac{n}{2} \log \left(n^{-1} \sum_{i=1}^n (\log x_i - \mu)^2 \right).$$

Die Profil-Log-Likelihood Funktion von σ^2 ist analog

$$l_p(\sigma^2) = l_x(\hat{\mu}(\sigma^2), \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{n \hat{\sigma}_{ML}^2}{2 \sigma^2}$$

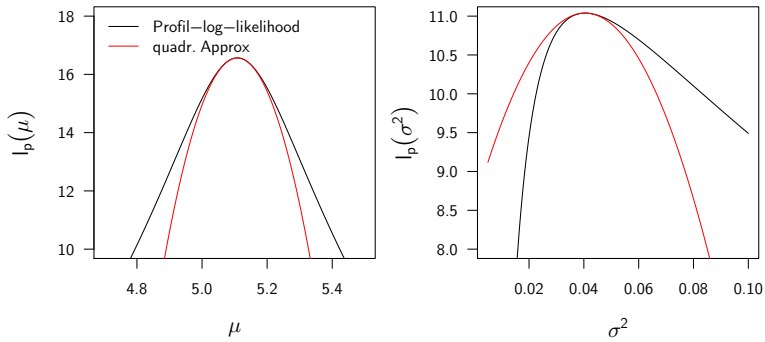
wobei wir $\hat{\mu}(\sigma^2) = \hat{\mu}_{ML} = n^{-1} \sum_{i=1}^n \log x_i$ in $l_x(\mu, \sigma^2)$ einsetzen ($\hat{\mu}_{ML}$ ist unabhängig von σ^2).

Die quadratische Approximation von $l_p(\mu)$ an der Stelle $\hat{\mu}_{ML}$ berechnet sich gemäss Satz 4.1 als

$$\begin{aligned} l_p(\mu) &\approx l_p(\hat{\mu}_{ML}) - \frac{1}{2} \frac{n}{\hat{\sigma}_{ML}^2} (\mu - \hat{\mu}_{ML})^2 \\ &= -\frac{n}{2} \left(\log(\hat{\sigma}_{ML}^2) + \frac{(\mu - \hat{\mu}_{ML})^2}{\hat{\sigma}_{ML}^2} \right). \end{aligned}$$

Die quadratischen Approximationen von $l_p(\sigma^2)$ an der Stelle $\hat{\sigma}_{ML}^2$ ergibt sich analog zu

$$\begin{aligned} l_p(\sigma^2) &\approx l_p(\hat{\sigma}_{ML}^2) - \frac{1}{2} \frac{n}{2\hat{\sigma}_{ML}^4} (\sigma^2 - \hat{\sigma}_{ML}^2)^2 \\ &= -\frac{n}{2} \left(\log \hat{\sigma}_{ML}^2 + 1 + \frac{(\sigma^2 - \hat{\sigma}_{ML}^2)^2}{2\hat{\sigma}_{ML}^4} \right). \end{aligned}$$



5 Bayes-Inferenz

1. Betrachten Sie die Fragestellung aus Aufgabe 15 in Kapitel 3. Ziel sei es nun, Bayes-Inferenz für die unbekannte Anzahl N der Taxis unter Annahme einer uneigentlichen diskreten Priori-Verteilung für N durchzuführen:

$$f(N) \propto 1 \quad \text{für } N = 0, 1, \dots, \infty$$

- a) Zeigen Sie, dass der Posteriori-Modus gleich dem ML-Schätzer ist.

► *Der Posteriori-Modus $\text{Mod}(N | y)$ maximiert die Posteriori*

$$f(N | y) = \frac{f(y | N)f(N)}{f(y)} \propto f(y | N),$$

da (hier) die Priori $f(N)$ und (allgemein) die marginale Likelihood $f(y)$ nicht von N abhängen. Aber $f(y | N)$ ist die Likelihood von N , sodass $\text{Mod}(N | y)$ mit \hat{N}_{ML} zusammenfällt, der in Aufgabe 15c), Kapitel 3 als $\hat{N}_{ML} = Y$ hergeleitet wurde.

- b) Zeigen Sie, dass für $n > 1$ die Posteriori-Wahrscheinlichkeitsfunktion gleich

$$f(N | y) = \frac{n-1}{y} \binom{y}{n} \binom{N}{n}^{-1}, \quad \text{für } N \geq y$$

ist.

► *Die Posteriori $f(N | y)$ ist nach 1a) proportional zur Likelihood $L(N)$, die bereits in (3.4) gegeben war als*

$$L(N) = \frac{(N-n)!}{N!} = \frac{1}{n!} \binom{N}{n}^{-1} \quad \text{für } N \geq y.$$

Zur vollständigen Angabe fehlt somit nur noch die marginale Likelihood

$$\begin{aligned} f(y) &= \sum_{N=0}^{\infty} L(N) \\ &= \lim_{k \rightarrow \infty} \sum_{N=y}^k \frac{(N-n)!}{N!}. \end{aligned}$$

Um auf die angegebene Form für $f(N | y)$ zu kommen, müssen wir also zeigen, dass sich diese Reihe zu

$$f(y) = \left\{ \frac{n-1}{y} \binom{y}{n} n! \right\}^{-1} = \frac{(y-n)!}{(n-1)(y-1)!}$$

vereinfachen lässt. Dies gelingt uns mittels der Umformung

$$\sum_{N=y}^k \frac{(N-n)!}{N!} = \frac{(y-n)!}{(n-1)(y-1)!} - \frac{(k-(n-1))!}{k!(n-1)!},$$

die man leicht über vollständige Induktion nach $k \geq y$ nachweist (und auf die man z. B. mit **Maxima** kommt). Der erste Term hat bereits die gewünschte Form, und der zweite Term strebt wegen

$$\frac{(k-(n-1))!}{k!(n-1)!} = \frac{1}{k(k-1) \cdots (k-(n-1)+1)} \cdot \frac{1}{n-1}$$

für $k \rightarrow \infty$ gegen Null, was den Beweis abschließt.

- c) Leiten Sie die folgende Formel für den Posteriori-Erwartungswert her:

$$E(N|y) = \frac{n-1}{n-2} \cdot (y-1) \quad \text{für } n > 2$$

► Der Ansatz ist

$$\begin{aligned} E(N|y) &= \sum_{N=0}^{\infty} N f(N|y) \\ &= \frac{n-1}{y} \binom{y}{n} n! \sum_{N=y}^{\infty} \frac{(N-n)!}{(N-1)!} \end{aligned}$$

und mittels der Reihe aus 1b) erhalten wir

$$\sum_{N=y}^{\infty} \frac{(N-n)!}{(N-1)!} = \sum_{N=y-1}^{\infty} \frac{(N-(n-1))!}{N!} = \frac{(y-n)!}{(n-2)(y-2)!}.$$

Setzen wir dies oben ein, kürzen sich die Terme zur gewünschten Formel.

- d) Vergleichen Sie die frequentistischen Schätzer aus Aufgabe 15 in Kapitel 3 mit dem Posteriori-Modus und Erwartungswert für $n = 48$ und $y = 1812$. Berechnen Sie numerisch das zugehörige 95%-HPD-Kredibilitätsintervall für N .

► Der erwartungstreue Schätzer aus Aufgabe 15 liefert in diesem Fall

$$\hat{N} = \frac{n+1}{n}y - 1 = 1848.75,$$

was deutlich größer als der ML-Schätzer und Posteriori-Modus $y = 1812$ ist. Der Posteriori-Erwartungswert ist mit

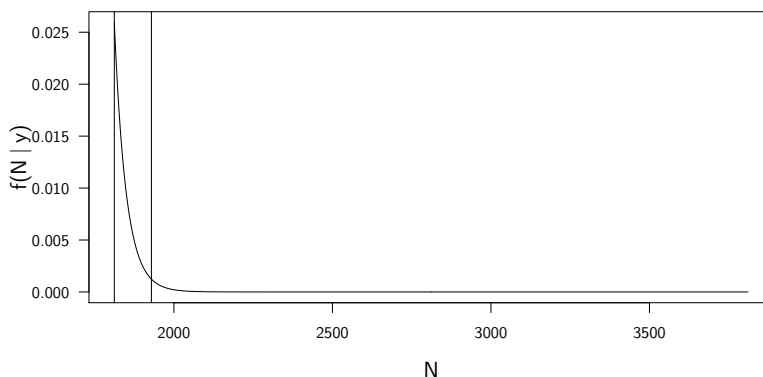
$$E(N|y) = \frac{n-1}{n-2} \cdot (y-1) = 1850.37$$

sogar noch größer. Das HPD-Intervall für N berechnen wir in R:

```

> ## Daten sind
> n <- 48
> y <- 1812
> ## Posteriori-Werte in genügend großem Bereich berechnen:
> N <- seq(from = y, length = 2000)
> posterior <- exp(log(n - 1) - log(y) + lchoose(y, n) - lchoose(N, n))
> plot(N, posterior, type = "l",
+      ylab = "f(N | y)" # offensichtlich groß genug
> ## HPD insofern hier speziell, weil Modus am Rand des Parameterraums liegt.
> ## Somit ist nur die Frage, wie weit "nach oben" das HPD reicht.
> level <- 0.95
> hpdLower <- y
> ## streng monoton fallende Posteriori ausnutzen, d.h. kleinstes Intervall ist
> ## das mit kleinster rechter Grenze:
> cumulatedPosterior <- cumsum(posterior)
> hpdUpper <- min(N[cumulatedPosterior >= level])
> ## und HPD einzeichnen
> abline(v = c(hpdLower, hpdUpper))

```



Das 95%-HPD-Intervall ist hier $[1812, 1929]$.

2. Gegeben sei die Zufallsstichprobe X_1, \dots, X_n eines auf dem Intervall $[0, \theta]$ gleichverteilten Untersuchungsmerkmals. Als *a priori* Verteilung für θ legen wir die *Pareto-Verteilung* $\text{Par}(a, b)$ mit den Parametern $a > 0$ und $b > 0$ zugrunde. Diese Verteilung besitzt die Dichte

$$f(\theta) = \begin{cases} ab^a \theta^{-(a+1)} & \text{falls } \theta \geq b, \\ 0 & \text{sonst} \end{cases}$$

und hat Modus b . Der Erwartungswert der $\text{Par}(a, b)$ -Verteilung existiert für $a > 1$ und lautet $ab/(a - 1)$.

- a) Zeigen Sie, dass $T(X) = \max\{X_1, \dots, X_n\}$ suffizient für θ ist.

► Siehe Aufgabe 5 in Kapitel 2.

- b) Bestimmen Sie die *a posteriori* Verteilung von θ – um welche Verteilung handelt es sich?

► Die *a posteriori* Verteilung von θ ist wieder eine Pareto-Verteilung, denn mit $t = \max\{x_1, \dots, x_n\}$ ist

$$\begin{aligned} f(\theta | x) &\propto f(x | \theta) f(\theta) \\ &\propto \frac{1}{\theta^n} \mathbb{I}_{[0, \theta]}(t) \cdot \frac{1}{\theta^{a+1}} \mathbb{I}_{[b, \infty)}(\theta) \\ &= \frac{1}{\theta^{(a+n)+1}} \mathbb{I}_{[\max\{b, t\}, \infty)}(\theta), \end{aligned}$$

d. h. $\theta | x \sim \text{Par}(a + n, \max\{b, t\})$. Die Pareto-Verteilung ist also konjugiert zur Gleichverteilung.

c) Bestimmen Sie den Posteriori-Modus $\text{Mod}(\theta | x)$.

► Der Angabe zufolge ist der Modus der $\text{Par}(a, b)$ -Verteilung gleich $ab/(a - 1)$ (für $a > 1$), sodass der Posteriori-Modus

$$\text{Mod}(\theta | x) = \frac{(a + n) \max\{b, t\}}{a + n - 1}$$

ist für $a + n > 1$, was wegen $a > 0$ immer gesichert ist.

3. Die ersten zuverlässigen Datierungen der Vulkan-Gesteinsart Granophyr aus dem Ennerdale in der britischen Region West Cumbria wurden in den 1960ern mit der sogenannten K/Ar-Methode vorgenommen. Mit dieser Methode wurde das Alter des Gesteins auf 370 ± 20 Millionen Jahren geschätzt. Die in den späten 1970ern entwickelte präzisere Rb/Sr-Methode ergab eine Schätzung des Alters von 421 ± 8 Millionen Jahren.

a) Angenommen, die obigen Messfehler sind normalverteilt, sodass die angegebenen Unsicherheiten den Standardabweichungen entsprechen: Verwenden Sie die Resultate der K/Ar-Methode als Priori-Verteilung für die mit der Rb/Sr-Methode erhobenen Daten und bestimmen Sie die Posteriori-Verteilung für das Alter des Gesteins.

► Wir beobachten also die Rb/Sr-Messung $x = 421$ aus der $N(\mu, \sigma^2)$ -Verteilung mit $\sigma = 8$ und haben Vorwissen $\mu \sim N(\nu, \tau^2)$ aus der K/Ar-Methode, die $\nu = 370$ mit Standardabweichung $\tau = 20$ lieferte. Dies ist der Spezialfall von Beispiel 5.5 mit $n = 1$, sodass die Posteriori-Verteilung von μ wieder eine Normalverteilung ist mit Präzision

$$\lambda = \frac{1}{\sigma^2} + \frac{1}{\tau^2} = 0.0181$$

und Erwartungswert

$$\eta = \lambda^{-1} \left(\frac{x}{\sigma^2} + \frac{\nu}{\tau^2} \right) = 413.966.$$

- b) Nun nehmen wir an, dass Ihnen die Resultate der K/Ar-Methode nicht zur Verfügung stehen. Beruhend auf Messungen von ähnlichen Gesteinsarten vermuten Sie jedoch, dass das Alter 400 ± 50 Millionen Jahre ist. Berechnen Sie die Posteriori-Verteilung für das Alter nach der Rb/Sr-Messung.

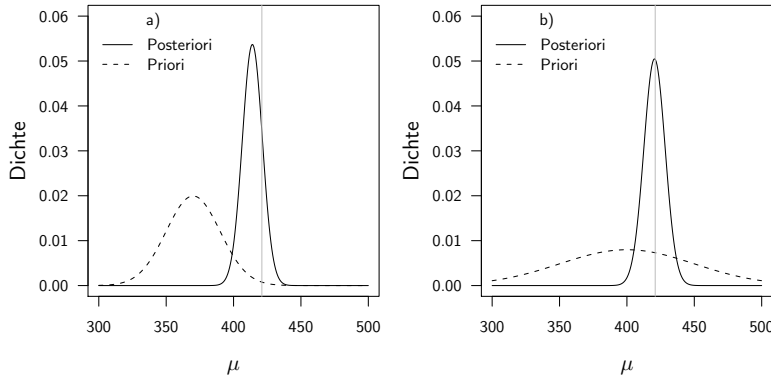
► *Das Vorwissen ist also nun ungenauer, was mit der höheren Priori-Standardabweichung $\tau = 50$ repräsentiert wird. Analog zu Aufgabe 3a) ergibt sich a posteriori eine Normalverteilung für μ mit Präzision $\lambda = 0.016$ und Erwartungswert $\eta = 420.476$.*

- c) Stellen Sie für beide Szenarien die zusammengehörigen Priori- und Posteriori-verteilungen grafisch dar.

```

► ## Dichte der Posteriori von mu im (allgemeinen) Normal-Normal-Modell
> ## bei beobachteter Zufallsstichprobe x aus einer N(mu, sigma2)-Verteilung und
> ## angenommener Priori-Verteilung N(nu, tau2) für mu:
> nn.posterior <- function(mu, x, sigma2, nu, tau2)
+ {
+   n <- length(x)
+   post.var <- 1/(n/sigma2 + 1/tau2)
+   post.mean <- (sum(x)/sigma2 + nu/tau2) * post.var
+   return(dnorm(mu, mean = post.mean, sd = sqrt(post.var)))
+ }
> ## Plotten:
> par(mfrow = c(1, 2))
> grid <- seq(300, 500, length = 300)
> ## a)
> plot(grid, nn.posterior(grid, x = 421, sigma2 = 8^2, nu = 370, tau2 = 20^2),
+       xlab = expression(mu), ylab = "Dichte", type = "l",
+       ylim = c(0, 0.06))
> lines(grid, dnorm(grid, mean = 370, sd = 20), lty = 2)
> abline(v = 421, col = "gray")
> legend("topleft", lty = c(1, 2),
+       legend = c("Posteriori", "Priori"), bty = "n", title = "a)")
> ## b)
> plot(grid, nn.posterior(grid, x = 421, sigma2 = 8^2, nu = 400, tau2 = 50^2),
+       xlab = expression(mu), ylab = "Dichte", type = "l",
+       ylim = c(0, 0.06))
> lines(grid, dnorm(grid, mean = 400, sd = 50), lty = 2)
> abline(v = 421, col = "gray")
> legend("topleft", lty = c(1, 2),
+       legend = c("Posteriori", "Priori"), bty = "n", title = "b)")

```

Bei ungenauem Vorwissen ist die Posteriori fast genau um x zentriert (mit grauem Strich markiert), und etwas flacher als bei genauerem Vorwissen, was der geringeren Präzision entspricht.

4. Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer $B(\pi)$ -Verteilung. Die Priori-Verteilung für π sei gegeben durch die Mischung zweier Betaverteilungen:

$$f(\pi) = w f_1(\pi | \alpha_1, \beta_1) + (1 - w) f_2(\pi | \alpha_2, \beta_2),$$

wobei $f_i(\pi | \alpha_i, \beta_i)$ für $i = 1, 2$ jeweils die Dichte der Betaverteilung mit Parametern α_i und β_i ist und $w \in (0, 1)$ bekannt sei.

- a) Bestimmen Sie einen Ausdruck für die Posteriori-Verteilung $f(\pi | x)$.

► Die Posteriori-Verteilung ist wieder eine Mischung aus zwei Beta-Verteilungen:

$$\begin{aligned} f(\pi | x) &\propto f(x | \pi) f(\pi) \\ &\propto \pi^{n\bar{x}} (1 - \pi)^{n - n\bar{x}} \left(w \frac{\pi^{\alpha_1 - 1} (1 - \pi)^{\beta_1 - 1}}{B(\alpha_1, \beta_1)} + (1 - w) \frac{\pi^{\alpha_2 - 1} (1 - \pi)^{\beta_2 - 1}}{B(\alpha_2, \beta_2)} \right) \\ &= \frac{w}{B(\alpha_1, \beta_1)} \pi^{\alpha_1 + n\bar{x} - 1} (1 - \pi)^{\beta_1 + n - n\bar{x} - 1} + \frac{1 - w}{B(\alpha_2, \beta_2)} \pi^{\alpha_2 + n\bar{x} - 1} (1 - \pi)^{\beta_2 + n - n\bar{x} - 1} \\ &= \frac{w B(\alpha_1^*, \beta_1^*)}{B(\alpha_1, \beta_1)} f_1(\pi | \alpha_1^*, \beta_1^*) + \frac{(1 - w) B(\alpha_2^*, \beta_2^*)}{B(\alpha_2, \beta_2)} f_2(\pi | \alpha_2^*, \beta_2^*) \end{aligned}$$

mit $\alpha_i^* = \alpha_i + n\bar{x}$ und $\beta_i^* = \beta_i + n - n\bar{x}$, $i = 1, 2$ als aufdatierten Parametern der Mischungskomponenten und Mischungsanteilen, die proportional zu

$$\gamma_1 = \frac{w B(\alpha_1^*, \beta_1^*)}{B(\alpha_1, \beta_1)} \quad \text{bzw.} \quad \gamma_2 = \frac{(1 - w) B(\alpha_2^*, \beta_2^*)}{B(\alpha_2, \beta_2)}$$

sind (d. h. um die tatsächlichen Anteile zu erhalten muss man $\gamma_i^* = \gamma_i / (\gamma_1 + \gamma_2)$ setzen).

- b) Geben Sie einen Ausdruck für den Posteriori-Erwartungswert an.

► *Aus der Form der Posteriori-Dichte*

$$f(\pi | x) = \gamma_1^* f_1(\pi | \alpha_1^*, \beta_1^*) + (1 - \gamma_1^*) f_2(\pi | \alpha_2^*, \beta_2^*),$$

die die Beta-Dichten in einer Linearkombination verbindet, sieht man leicht mit der Linearität der Integration, dass

$$\begin{aligned} E(\pi | x) &= \gamma_1^* E(\pi | \alpha_1^*, \beta_1^*) + (1 - \gamma_1^*) E(\pi | \alpha_2^*, \beta_2^*) \\ &= \gamma_1^* \frac{\alpha_1^*}{\alpha_1^* + \beta_1^*} + (1 - \gamma_1^*) \frac{\alpha_2^*}{\alpha_2^* + \beta_2^*} \end{aligned}$$

gilt.

- c) Bestimmen Sie ein gleichendiges $(1 - \alpha)\%$ -Vertrauensintervall, indem Sie die $(\alpha/2)$ - und $(1 - \alpha/2)$ -Quantile der Posteriori-Verteilung in R numerisch ermitteln.

► *Zunächst brauchen wir einen Ausdruck für die Verteilungsfunktion der Posteriori. Ebenfalls aus der Form der Posteriori-Dichte folgt, dass sich die Posteriori-Verteilungsfunktion aus den Verteilungsfunktionen der Mischungskomponenten ergibt, d. h.*

$$F(\pi | x) = \gamma_1^* F(\pi | \alpha_1^*, \beta_1^*) + (1 - \gamma_1^*) F(\pi | \alpha_2^*, \beta_2^*).$$

Das gleichendige $(1 - \alpha)$ -Kreditivitätsintervall ist dann

$$[F^{-1}(\alpha/2 | x), F^{-1}(1 - \alpha/2 | x)];$$

wir suchen also die Abszissen, bei denen die Verteilungsfunktion die Werte $\alpha/2$ bzw. $(1 - \alpha/2)$ annimmt:

```
> ## Die Verteilungsfunktion einer Betamischung mit
> ## Anteil gamma1 für die erste Komponente und Parametervektoren
> ## alpha und beta, die die zwei Komponenten bestimmen:
> pbetamix <- function(pi, gamma1, alpha, beta)
+ {
+   gamma1 * pbeta(pi, alpha[1], beta[1]) +
+   (1 - gamma1) * pbeta(pi, alpha[2], beta[2])
+ }
> ## und die numerisch bestimmte Quantilsfunktion
> qbetamix <- function(q, gamma1, alpha, beta)
+ {
+   f <- function(pi)
+   {
+     pbetamix(pi, gamma1, alpha, beta) - q
+   }
+   unirootResult <- uniroot(f, lower=0, upper=1)
+   if(unirootResult$iter < 0)
+     return(NA)
+   else
+     return(unirootResult$root)
+ }
```

```

> ## Kredibilitätsregion mit Niveau level:
> credBetamix <- function(level, gamma1, alpha, beta)
+ {
+   halfa <- (1 - level)/2
+   ret <- c (qbetamix(halfa, gamma1, alpha, beta),
+             qbetamix(1 - halfa, gamma1, alpha, beta)
+           )
+   ret
+ }

```

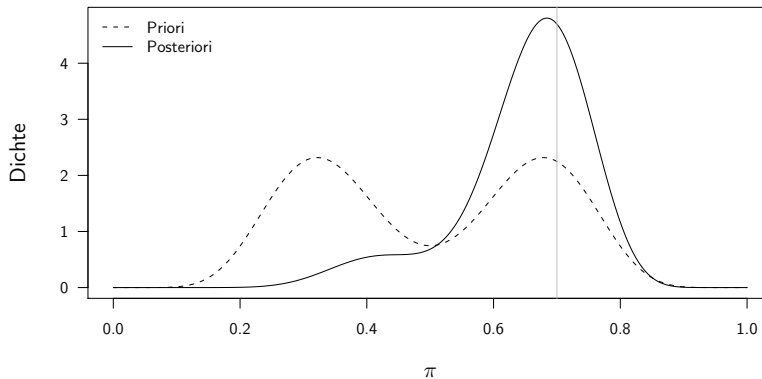
- d) Testen Sie ihre Resultate wenn $w = 0.5$, $\alpha_1 = 10$, $\beta_1 = 20$, $\alpha_2 = 20$, $\beta_2 = 10$ und bei $n = 10$ Experimenten insgesamt dreimal die Null beobachtet wurde.

► *Zunächst zeichnen wir die Priori- und Posteriori-Verteilung:*

```

> ## Dichtefunktion
> dbetamix <- function(pi, gamma1, alpha, beta)
+ {
+   gamma1 * dbeta(pi, alpha[1], beta[1]) +
+   (1 - gamma1) * dbeta(pi, alpha[2], beta[2])
+ }
> ## Priori-Parameter
> alpha <- c(10, 20)
> beta <- c(20, 10)
> gamma1 <- 0.5
> ## Daten
> n <- 10
> xq <- (10 - 3)/10
> ## Posteriori-Parameter
> (alphaStar <- alpha + n * xq)
[1] 17 27
> (betaStar <- beta + n - n * xq)
[1] 23 13
> gamma <- c(gamma1, 1 - gamma1) * beta(alphaStar, betaStar) / beta(alpha, beta)
> (gammaStar <- gamma / sum(gamma))
[1] 0.1085271 0.8914729
> ## Plotten
> piGrid <- seq(0, 1, length = 1000)
> prior <- dbetamix(piGrid, gamma1, alpha, beta)
> post <- dbetamix(piGrid, gammaStar[1], alphaStar, betaStar)
> plot(piGrid, prior, type = "l", ylim=c(0,max(prior,post)), lty = 2,
+      ylab = "Dichte", xlab = expression(pi))
> lines(piGrid, post, lty = 1)
> legend("topleft", c("Priori", "Posteriori"), lty=c(2, 1), bty = "n")
> abline(v = xq, col = "gray")

```



Die Formel für den Posteriori-Erwartungswert können wir per Simulation von Zufallszahlen aus der Posteriori testen:

```
> ## zum Samplen aus einer Betamischung: latente Variable which mischt die
> ## beiden Betaverteilungen!
> rbbmix <- function(n, gamma1, alpha, beta)
+ {
+   which <- sample(1:2, rep = T, size = n, prob = c(gamma1, 1 - gamma1))
+   ret <- rbeta(n, alpha[which], beta[which])
+   ret
+ }
> ## analytische Formel war
> sum( gammaStar * (alphaStar/(alphaStar + betaStar)) )
[1] 0.6478682
> ## Mittelwert in der Simulation
> set.seed(12)
> posteriorSample <- rbbmix(1e+6, gammaStar[1], alphaStar, betaStar)
> mean(posteriorSample)
[1] 0.6479378
```

Bis auf 3 Dezimalstellen stimmt in diesem Fall der simulierte Mittelwert mit dem analytisch berechneten Erwartungswert überein.

Zuletzt berechnen wir noch ein gleichendiges 95%-Vertrauensintervall für π :

```
> ## numerisch:
> credBetamix(0.95, gammaStar[1], alphaStar, betaStar)
[1] 0.3665792 0.8062333
> ## durch Simulation:
> quantile(posteriorSample, prob = c(0.025, 0.975))
      2.5%      97.5%
0.3669860 0.8063795
```

Hier bestätigt wiederum die Simulation das (numerische) Ergebnis.

5. Angenommen die Lebensdauern X_1, \dots, X_n von n Glühlampen sind eine Zufallsstichprobe eines exponentialverteilten Untersuchungsmerkmals mit Parameter λ .
 - a) Bestimmen Sie Jeffreys' Priori für λ und zeigen Sie, dass die Priori uneigentlich ist.

► Aus Aufgabe 10a) in Kapitel 3 haben wir die Score-Funktion von λ ,

$$S(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i,$$

die als Zufallsvariable (als Transformation von X_1, \dots, X_n) die Varianz

$$J(\lambda) = \text{Var}(S(\lambda)) = n \cdot \text{Var}(X_1) = \frac{n}{\lambda^2}$$

besitzt. Jeffreys' Priori ist also

$$f(\lambda) \propto \sqrt{J(\lambda)} \propto \lambda^{-1},$$

und dieser Kern lässt sich wegen

$$\int_0^{\infty} \lambda^{-1} d\lambda = [\log(\lambda)]_0^{\infty} = \log(\infty) - \log(0) = \infty - (-\infty) = \infty$$

nicht zu einer Dichte normieren, sodass $f(\lambda)$ uneigentlich ist.

- b) Angenommen, die Glühbirnen werden gleichzeitig eingeschaltet und bis zum r -ten Ausfall beobachtet. Bestimmen Sie die Likelihoodfunktion für dieses Szenario. Berücksichtigen Sie, dass die Beobachtungen von $(n - r)$ Glühbirnen zensiert werden.

► Seien $X_{(1)}, \dots, X_{(n)}$ die aufsteigend sortierten und teilweise unbeobachteten Lebensdauern der Glühbirnen. Nur $X_{(1)}, \dots, X_{(r)}$ werden beobachtet, alle anderen Zufallsvariablen sind rechtszensiert. Als Likelihood ergibt sich deshalb ein Spezialfall von Beispiel 2.7 mit $\delta_{(i)} = \mathbf{1}_{\{1, \dots, r\}}(i)$ und somit $\sum_{i=1}^n \delta_i = r$:

$$L(\lambda) = \lambda^r \exp \left\{ -\lambda \left(\sum_{i=1}^r x_{(i)} + (n - r)x_{(r)} \right) \right\},$$

da für $X_{(r+1)}, \dots, X_{(n)}$ jeweils der Wert $X_{(r)}$ als zensierte Zufallsvariable verwendet wird.

- c) Angenommen, das System wird nur bis zu einem vorher spezifizierten Zeitpunkt $c > 0$ beobachtet. Zeigen Sie, dass falls keine der Birnen vor dem Zeitpunkt c ausfällt auch die Posteriori uneigentlich ist.

► Falls keine der Glühbirnen vor der Zeit c ausgefallen ist, ist die Likelihood von λ gleich

$$L(\lambda) = \exp(-n\lambda c).$$

Verwenden wir Jeffreys' Priori $f(\lambda) \propto \lambda^{-1}$, so ergibt sich als Posteriori

$$f(\lambda | x) \propto \frac{1}{\lambda} \exp(-nc\lambda).$$

Dies ist der Kern der uneigentlichen $G(0, nc)$ -Verteilung.

Dass diese Verteilung uneigentlich ist, kann man mit der Grenzbetrachtung von $Z \sim G(a, b)$ für $a \searrow 0$ bei $b > 0$ sehen. Für jedes $a > 0$ integriert sich der Kern der Verteilung zu

$$\int_0^{\infty} z^{a-1} \exp(-bz) dz = \frac{\Gamma(a)}{b^a}.$$

Da aber $b^a \xrightarrow{a \searrow 0} 1$ und $\Gamma(a) \xrightarrow{a \searrow 0} \infty$ gilt, folgt, dass das Integral über $z^{-1} \exp(-bz)$ nicht endlich ist, sodass dieser Kern nicht zu einer Dichtefunktion mit $\int_0^{\infty} f(z) dz = 1$ normiert werden kann.

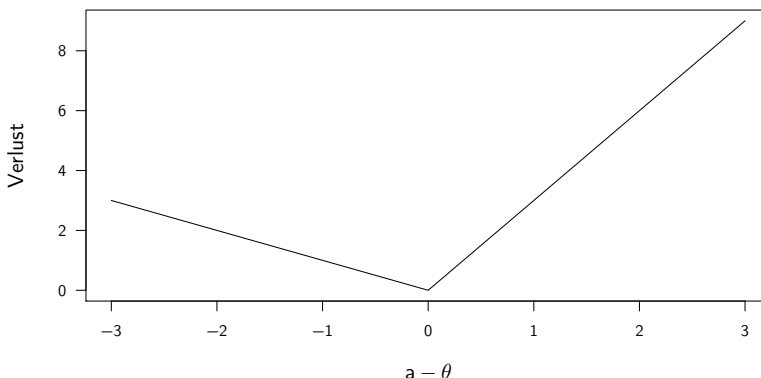
6. Betrachten Sie folgende Verlustfunktion, die durch $c, d > 0$ parametrisiert ist:

$$l(a, \theta) = \begin{cases} -c(a - \theta) & \text{falls } a - \theta \leq 0 \\ d(a - \theta) & \text{falls } a - \theta > 0 \end{cases}.$$

- a) Skizzieren Sie $l(a, \theta)$ als Funktion von $a - \theta$ für $c = 1$ und $d = 3$.

► Die Verlustfunktion ist eine Verallgemeinerung der linearen Verlustfunktion $l(a, \theta) = |a - \theta|$, die sich mit $c = d = 1$ ergibt. Eine Skizze für $c = 1$, aber $d = 3$, als Funktion der Differenz von a und θ (von der die Verlustfunktion letztendlich abhängt):

```
> ## Verlustfunktion als Funktion von a - theta:
> loss <- function(aMinusTheta, c, d)
+ {
+   ifelse(aMinusTheta <= 0, - c * aMinusTheta, d * aMinusTheta)
+ }
> aMinusTheta <- seq(-3, 3, length = 101)
> plot(aMinusTheta, loss(aMinusTheta, c = 1, d = 3),
+      type = "l", xlab = expression(a - theta), ylab = "Verlust")
```



Zu große Werte von a bringen also einen höheren Verlust als zu kleine Werte.

- b) Bestimmen Sie den Bayes-Schätzer bezüglich der Verlustfunktion $l(a, \theta)$.

► Der *a posteriori* erwartete Verlust ist

$$\begin{aligned} E(l(a, \theta) | x) &= \int l(a, \theta) f(\theta | x) d\theta \\ &= \int_{-\infty}^a d(a - \theta) f(\theta | x) d\theta + \int_a^{\infty} c(\theta - a) f(\theta | x) d\theta. \end{aligned}$$

Um den Bayes-Schätzer zu finden, d. h. die Minimumsstelle

$$\hat{a} = \arg \min_a E(l(a, \theta) | x),$$

leiten wir den Ausdruck zunächst nach a ab. Dabei müssen wir die Leibnizregel aus Anhang B.7 benutzen, da die Integrationsgrenzen von a abhängen. Wir identifizieren die Variablen t mit a und x mit θ . Beim ersten Integral sind $a(t) = -\infty$, $b(t) = a$ und $f(x, t)$ aus der Formel gleich $d(a - \theta)f(\theta | x)$, beim zweiten Integral sind $a(t) = a$ und $b(t) = \infty$ und $f(x, t)$ aus der Formel gleich $c(\theta - a)f(\theta | x)$. Mit der Konvention $\infty \cdot 0 = 0$ ergibt sich dann

$$\begin{aligned} \frac{d}{da} E(l(a, \theta) | x) &= d \int_{-\infty}^a f(\theta | x) d\theta - c \int_a^{\infty} f(\theta | x) d\theta \\ &= dF(a | x) - c(1 - F(a | x)) \\ &= (c + d)F(a | x) - c. \end{aligned}$$

Die Nullstelle dieser Funktion in a ist also

$$\hat{a} = F^{-1}(c/(c + d) | x),$$

d. h. der Bayes-Schätzer \hat{a} ist das $c/(c + d) \cdot 100\%$ -Quantil der Posteriori-Verteilung von θ . Bei $c = d$ ergibt sich somit der Median, der beim Spezialfall der linearen Verlustfunktion als Bayes-Schätzer resultierte. Im Zahlenbeispiel aus 6a) ergibt sich dagegen als Bayes-Schätzer das 25%-Quantil der Posteriori-Verteilung.

6 Numerische Methoden zur Bayes-Inferenz

1. Sei $X \sim \text{Po}(e\lambda)$ bei bekanntem e eine Poissonverteilte Zufallsvariable und $\lambda \sim G(\alpha, \beta)$ a priori.

a) Berechnen Sie den Posteriori-Erwartungswert von λ .

► Die Posteriori-Dichte ist

$$\begin{aligned} f(\lambda | x) &\propto f(x | \lambda) f(\lambda) \\ &\propto \lambda^x \exp(-e\lambda) \cdot \lambda^{\alpha-1} \exp(-\beta\lambda) \\ &= \lambda^{(\alpha+x)-1} \exp(-(\beta+e)\lambda), \end{aligned}$$

was wir als Kern der $G(\alpha+x, \beta+e)$ -Verteilung identifizieren können. Somit ist der Posteriori-Erwartungswert

$$E(\lambda | x) = \frac{\alpha+x}{\beta+e}.$$

b) Berechnen Sie nun die Laplace-Approximation des Posteriori-Erwartungswertes.

► Wir setzen $g(\lambda) = \lambda$ und $n = 1$ zur Anwendung der Approximation 6.3. Die logarithmierten Integranden sind

$$\begin{aligned} -h(\lambda) &= (\alpha+x-1) \log(\lambda) - (\beta+e)\lambda \\ \text{und} \quad -h^*(\lambda) &= \log(\lambda) - h(\lambda). \end{aligned}$$

Deren Ableitungen

$$\begin{aligned} \frac{d(-h(\lambda))}{d\lambda} &= \frac{\alpha+x-1}{\lambda} - (\beta+e) \\ \text{bzw.} \quad \frac{d(-h^*(\lambda))}{d\lambda} &= \frac{\alpha+x}{\lambda} - (\beta+e) \end{aligned}$$

haben Nullstellen

$$\hat{\lambda} = \frac{\alpha+x-1}{\beta+e} \quad \text{bzw.} \quad \hat{\lambda}^* = \frac{\alpha+x}{\beta+e}.$$

Die negativen Krümmungen an diesen Maximumsstellen von $-h(\lambda)$ bzw. $-h^*(\lambda)$ ergeben sich zu

$$\hat{\kappa} = \frac{(\beta+e)^2}{\alpha+x-1} \quad \text{und} \quad \hat{\kappa}^* = \frac{(\beta+e)^2}{\alpha+x},$$

sodass die Laplace-Approximation des Posteriori-Erwartungswertes

$$\begin{aligned}\hat{E}(\lambda | x) &= \sqrt{\frac{\alpha + x}{\alpha + x - 1}} \exp \left\{ \log(\hat{\lambda}^*) + (\alpha + x - 1) \log(\hat{\lambda}^*) - (\beta + e) \hat{\lambda}^* \right. \\ &\quad \left. - (\alpha + x - 1) \log(\hat{\lambda}) + (\beta + e) \hat{\lambda} \right\} \\ &= \sqrt{\frac{\alpha + x}{\alpha + x - 1}} \exp \left\{ \log(\hat{\lambda}^*) + (\alpha + x - 1) \log(\hat{\lambda}^* / \hat{\lambda}) + (\beta + e)(\hat{\lambda} - \hat{\lambda}^*) \right\}\end{aligned}$$

ist, was sich weiter vereinfachen lässt zu

$$\hat{E}(\lambda | x) = \exp \{ (\alpha + x + 0.5) \log(\alpha + x) - (\alpha + x - 0.5) \log(\alpha + x - 1) - \log(\beta + e) - 1 \}.$$

- c) Vergleichen Sie für $\alpha = 0.5$ und $\beta = 0$ die Laplace-Approximation mit dem exakten Erwartungswert bei der Beobachtung $x = 11$ und $e = 3.04$ bzw. $x = 110$ und $e = 30.4$. Berechnen Sie auch den relativen Fehler der Approximation.

► Zunächst implementieren wir die Laplace-Approximation sowie die exakte Formel:

```
> ## Laplace-Approximation des Posteriori-Erwartungswertes
> ## bei beobachteten Daten x, Offset e und Priori-Parametern alpha, beta:
> laplaceApprox <- function(x, e, alpha, beta)
+ {
+   logRet <- (alpha + x + 0.5) * log(alpha + x) -
+             (alpha + x - 0.5) * log(alpha + x - 1) -
+             log(beta + e) - 1
+   exp(logRet)
+ }
> ## exakte Berechnung des Posteriori-Erwartungswertes mit
> exact <- function(x, e, alpha, beta)
+   (alpha + x) / (beta + e)
```

Im Zahlenbeispiel ergeben sich damit folgende Werte:

```
> (small <- c(exact = exact(11, 3.04, 0.5, 0),
+             approx = laplaceApprox(11, 3.04, 0.5, 0)))
      exact approx
3.782895 3.785504
> (large <- c(exact = exact(110, 30.4, 0.5, 0),
+             approx = laplaceApprox(110, 30.4, 0.5, 0)))
      exact approx
3.634868 3.634893
> ## relative Fehler:
> diff(small) / small["exact"]
      approx
0.000689798
> diff(large) / large["exact"]
      approx
6.887162e-06
```

Bei festem Verhältnis von Beobachtung x zu Offset e wird die Approximation für größere Zahlen somit offensichtlich besser. Betrachtet man das Verhältnis von Approximation zum wahren Wert,

$$\begin{aligned}\frac{\hat{E}(\lambda|x)}{E(\lambda|x)} &= \exp\{(\alpha+x-0.5)(\log(\alpha+x)-\log(\alpha+x-1))-1\} \\ &= \left(1 + \frac{1}{\alpha+x-1}\right)^{\alpha+x-0.5} / \exp(1),\end{aligned}$$

so kann man erkennen, dass dieses für $x \rightarrow \infty$ gegen Eins geht.

- d) Betrachten Sie nun $\theta = \log(\lambda)$. Bestimmen Sie zunächst die Posteriori-Dichtefunktion mittels des Transformationssatzes für Dichten. Berechnen Sie daraus die Laplace-Approximation des Posteriori-Erwartungswertes von $\exp(\theta)$ und vergleichen Sie diese wieder mit dem wahren Erwartungswert, den Sie diesmal auch numerisch mit der R-Funktion `integrate` für obige Daten bestimmen.

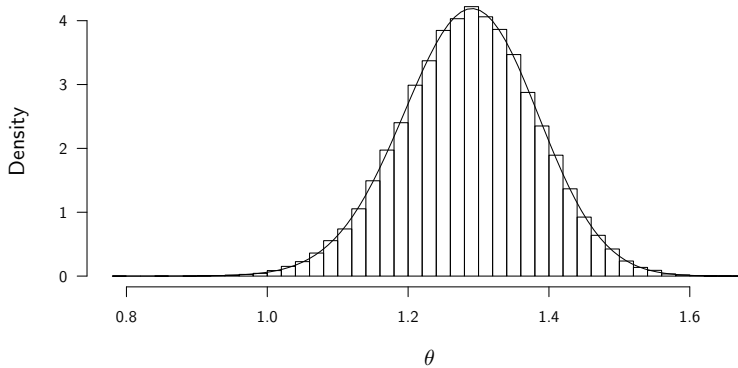
► Der Parameter $\theta = g(\lambda) = \log(\lambda)$ hat nun die gesamte reelle Achse als Träger, wodurch evtl. eine bessere Laplace-Approximation erzielt werden kann. Die Posteriori-Dichte ist

$$\begin{aligned}f_{\theta}(\theta|x) &= f_{\lambda}(g^{-1}(\theta)|x) \cdot \left| \frac{d}{d\theta} g^{-1}(\theta) \right| \\ &= \frac{(\beta+e)^{\alpha+x}}{\Gamma(\alpha+x)} \exp(\theta)^{\alpha+x-1} \exp(-(\beta+e)\exp(\theta)) \cdot \exp(\theta) \\ &= \frac{(\beta+e)^{\alpha+x}}{\Gamma(\alpha+x)} \exp\{(\alpha+x)\theta - (\beta+e)\exp(\theta)\}\end{aligned}$$

und entspricht keiner bekannten Verteilung.

```
> ## Posteriori-Dichte von theta = log(lambda):
> thetaDens <- function(theta, x, e, alpha, beta, log = FALSE)
+ {
+   logRet <- (alpha + x) * (theta + log(beta + e)) -
+     (beta + e) * exp(theta) - lgamma(alpha + x)
+   if(log)
+     return(logRet)
+   else
+     return(exp(logRet))
+ }
> ## Test durch Simulation:
> x <- 110
> e <- 30.4
> alpha <- 0.5
> beta <- 0
> ## Histogramm
> set.seed(59)
> thetaSamples <- log(rgamma(1e+5, alpha + x, beta + e))
> histResult <- hist(thetaSamples, prob = TRUE, breaks = 50,
+   xlab = expression(theta), main = "")
> ## berechnete Dichte
```

```
> thetaGrid <- seq(from = min(histResult$breaks), to = max(histResult$breaks),
+                 length = 101)
> lines(thetaGrid, thetaDens(thetaGrid, x, e, alpha, beta))
```



Die Laplace-Approximation von $E(\exp(\theta) | x)$ geht somit von

$$-h(\theta) = (\alpha + x)\theta - (\beta + e)\exp(\theta)$$

und $-h^*(\theta) = \theta - h(\theta)$

aus, da die Normierungskonstante der Posteriori-Dichte weggelassen werden kann. Die Maxima dieser Funktionen ergeben sich zu

$$\hat{\theta} = \log\left(\frac{\alpha + x}{\beta + e}\right) \quad \text{und} \quad \hat{\theta}^* = \log\left(\frac{\alpha + x + 1}{\beta + e}\right)$$

mit negativen Krümmungen

$$\hat{\kappa} = \alpha + x \quad \text{bzw.} \quad \hat{\kappa}^* = \alpha + x + 1.$$

Als Laplace-Approximation erhält man daraus mit

$$\hat{E}(\exp(\theta) | x) = \exp\{(\alpha + x + 0.5) \log(\alpha + x + 1) - (\alpha + x - 0.5) \log(\alpha + x) - \log(\beta + e) - 1\}$$

eine erstaunlich ähnliche Formel wie in 1b). Man beachte, dass der gleiche Erwartungswert approximiert wird, da $\exp(\theta) = \lambda$ gilt. Das Verhältnis von Approximation zum wahren Wert ist hier

$$\frac{\hat{E}(\exp(\theta) | x)}{E(\exp(\theta) | x)} = \left(1 + \frac{1}{\alpha + x}\right)^{\alpha + x + 0.5} / \exp(1),$$

und damit um einen Schritt in x dem Grenzwert 1 näher als die Approximation aus 1b). Die Laplace-Approximation funktioniert also hier tatsächlich etwas besser! Im Zahlenbeispiel wollen wir beide Laplace-Approximationen mit der numerischen vergleichen:

```

> ## Laplace-Approximation für E(exp(theta))
> laplaceApprox2 <- function(x, e, alpha, beta)
+ {
+   logRet <- (alpha + x + 0.5) * log(alpha + x + 1) -
+             (alpha + x - 0.5) * log(alpha + x) -
+             log(beta + e) - 1
+   exp(logRet)
+ }
> ## numerische Approximation
> numApprox <- function(x, e, alpha, beta)
+ {
+   integrand <- function(theta)
+     exp(theta + thetaDens(theta, x, e, alpha, beta, log = TRUE))
+   ## wichtig, hier erst auf log-Ebene addieren und dann exp nehmen!
+   intRes <- integrate(integrand, lower = -Inf, upper = Inf,
+                       rel.tol = sqrt(.Machine$double.eps))
+   if(intRes$message == "OK")
+     return(intRes$value)
+   else
+     return(NA)
+ }
> ## Vergleich aller Methoden:
> (small <- c(exact = exact(11, 3.04, 0.5, 0),
+             approx1 = laplaceApprox(11, 3.04, 0.5, 0),
+             approx2 = laplaceApprox2(11, 3.04, 0.5, 0),
+             approx3 = numApprox(11, 3.04, 0.5, 0)))
      exact approx1 approx2 approx3
3.782895 3.785504 3.785087 3.782895
> (large <- c(exact = exact(110, 30.4, 0.5, 0),
+             approx = laplaceApprox(110, 30.4, 0.5, 0),
+             approx2 = laplaceApprox2(110, 30.4, 0.5, 0),
+             approx3 = numApprox(110, 30.4, 0.5, 0)))
      exact approx approx2 approx3
3.634868 3.634893 3.634893 3.634868
> ## relative Fehler:
> (small[2:4] - small["exact"]) / small["exact"]
      approx1 approx2 approx3
6.897981e-04 5.794751e-04 2.230486e-15
> (large[2:4] - large["exact"]) / large["exact"]
      approx approx2 approx3
6.887162e-06 6.763625e-06 -3.384241e-14

```

Die numerische Approximation mit *integrate* kann in diesem Fall also noch bessere Ergebnisse liefern als die beiden Laplace-Approximationen.

2. Zur Schätzung der relativen Chance θ aus Beispiel 4.8 sollen nun Bayesianische Methoden verwendet werden. Hierzu nehmen wir *a priori* zwei unabhängige $\text{Be}(0.5, 0.5)$ -Verteilungen als Priori-Verteilungen für π_1 und π_2 an.

- a) Berechnen Sie die Posteriori-Verteilung von π_1 und π_2 für die in Tabelle 4.1 angegebenen Daten. Simulieren Sie aus diesen Posteriori-Verteilungen und berechnen Sie daraus Zufallszahlen aus der Posteriori-Verteilung von θ und

$\psi = \log(\theta)$. Berechnen Sie damit ausgewählte Monte-Carlo-Schätzer (Posteriori-Erwartungswert, Posteriori-Median, gleichendiges Kredititätsintervall, HPD-Intervall) für θ und ψ und vergleichen Sie diese mit den Likelihood-Ergebnissen aus Beispiel 4.8.

► *Nach Beispiel 5.1 sind a posteriori*

$$\pi_1 | x_1 \sim \text{Be}(0.5 + x_1, 0.5 + n_1 - x_1) \quad \text{und} \quad \pi_2 | x_2 \sim \text{Be}(0.5 + x_2, 0.5 + n_2 - x_2)$$

verteilt. Für die gegebenen Daten können wir Zufallszahlen aus diesen Verteilungen generieren und jeweils zu

$$\theta = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)}$$

bzw. $\psi = \log(\theta)$ transformieren, um anschließend Monte-Carlo-Schätzer zu gewinnen:

```
> ## Daten aus Tabelle 4.1 waren
> x <- c(6, 2)
> n <- c(108, 103)
> ## aus Posteriori von pi1 und pi2 simulieren
> size <- 1e+5
> set.seed(89)
> pi1 <- rbeta(size, 0.5 + x[1], 0.5 + n[1] - x[1])
> pi2 <- rbeta(size, 0.5 + x[2], 0.5 + n[2] - x[2])
> ## transformieren
> theta <- (pi1 / (1 - pi1)) / (pi2 / (1 - pi2))
> psi <- log(theta)
> ## und Monte-Carlo-Schätzer berechnen:
> ## dazu HPD-Funktion schreiben (nach Code aus Beispiel 6.7)
> hpd <- function(                                # liefert HPD-Intervall-
+           samples,                                # MC-Schätzung aus samples
+           prob = 0.95                             # zur Wkeit prob
+           )                                         # Funktion nimmt ungefähre Unimodalität an!
+                                                    # (es sind keine zerstückelten
+                                                    # HPD-Intervalle möglich)
+ {
+   ## samples sortieren
+   M <- length(samples)
+   level <- 1 - prob
+   samplesorder <- samples[order(samples)]
+
+   ## und schauen welches Intervall das mindestens 95% enthält
+   ## die kleinste Länge hat
+   max.size <- round(M * level)
+   size <- rep(NA, max.size)
+   for(i in 1:max.size){
+     lower <- samplesorder[i]
+     upper <- samplesorder[M-max.size+i]
+     size[i] <- upper - lower
+   }
+   size.min <- which.min(size)
+
+   ## dieses zurückgeben
```

```

+   HPD.lower <- samplesorder[size.min]
+   HPD.upper <- samplesorder[M-max.size+size.min]
+
+   return(c(lower = HPD.lower, upper = HPD.upper))
+ }
> ## Schätzungen für theta
> quantile(theta, prob = c(0.025, 0.5, 0.975)) # gleichendiges KI mit Median
      2.5%      50%      97.5%
0.6597731 2.8296129 16.8312942
> mean(theta)                                     # Erwartungswert
[1] 4.338513
> hpd(theta)                                     # HPD-Intervall
      lower      upper
0.2461746 12.2797998
> ## Schätzungen für psi
> quantile(psi, prob = c(0.025, 0.5, 0.975)) # gleichendiges KI mit Median
      2.5%      50%      97.5%
-0.4158593 1.0401399 2.8232399
> mean(psi)                                     # Erwartungswert
[1] 1.082216
> (psiHpd <- hpd(psi))                         # HPD-Intervall
      lower      upper
-0.5111745 2.7162842

```

Die Bayesianische Punktschätzung $E(\psi|x) \approx 1.082$ ist somit nur etwas kleiner als der ML-Schätzer $\hat{\psi}_{ML} = 1.089$. Das HPD-Intervall $[-0.511, 2.716]$ ist vergleichbar mit dem Wald-Intervall, das gleichendige Kreditabilitätsintervall $[-0.416, 2.823]$ mit dem Profil-Likelihood-Intervall aus Beispiel 4.8. (Vergleichen Sie auch Aufgabe 3 in Kapitel 4.)

- b) Versuchen Sie, die Dichte der Posteriori-Verteilung von θ und ψ analytisch zu berechnen. Berechnen Sie daraus mit numerischen Methoden den Posteriori-Erwartungswert und ein 95%-HPD-Kreditabilitätsintervall und vergleichen Sie die Ergebnisse mit den Monte-Carlo-Schätzern.

► Das Log Odds Ratio ψ ist die Differenz der beiden unabhängigen Log Odds ϕ_i , $i = 1, 2$,

$$\psi = \log \left(\frac{\pi_1/(1 - \pi_1)}{\pi_2/(1 - \pi_2)} \right) = \text{logit}(\pi_1) - \text{logit}(\pi_2) = \phi_1 - \phi_2.$$

Um die Dichte von ψ zu berechnen, liegt es deshalb nahe, zunächst die Verteilung von $\phi = g(\pi) = \text{logit}(\pi)$ zu bestimmen, wenn $\pi \sim \text{Be}(a, b)$ ist. Wegen

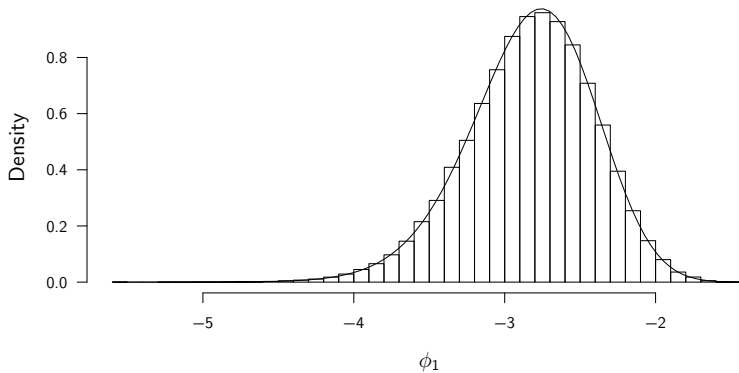
$$g^{-1}(\phi) = \frac{\exp(\phi)}{1 + \exp(\phi)} \quad \text{und} \quad \frac{d}{d\phi} g^{-1}(\phi) = g^{-1}(\phi)(1 - g^{-1}(\phi))$$

ergibt sich mittels des Dichtetransformationssatzes

$$f(\phi) = \frac{1}{B(a, b)} \frac{\exp(\phi)^a}{(1 + \exp(\phi))^{a+b}}.$$

Für ϕ_1 können wir das beispielhaft anhand unserer simulierten Zufallszahlen überprüfen:

```
> ## Dichte von phi = logit(pi), wenn pi ~ Be(a, b)
> phiDens <- function(phi, a, b)
+ {
+   pi <- plogis(phi)
+   logRet <- a * log(pi) + b * log(1 - pi) - lbeta(a, b)
+   return(exp(logRet))
+ }
> ## simuliert:
> histRes <- hist(qlogis(pi1), prob = TRUE, breaks = 50,
+               xlab = expression(phi[1]), main = "")
> ## analytisch:
> phiGrid <- seq(from = min(histRes$breaks), to = max(histRes$breaks),
+               length = 101)
> lines(phiGrid, phiDens(phiGrid, 0.5 + x[1], 0.5 + n[1] - x[1]))
```



Seien nun $\phi_i = \text{logit}(\pi_i)$ mit $\pi_i \sim \text{Be}(a_i, b_i)$, $i = 1, 2$. Der Faltungssatz liefert die Dichte von $\psi = \phi_1 - \phi_2$,

$$f_\psi(\psi) = \int f_1(\psi + \phi_2) f_2(\phi_2) d\phi_2$$

$$= \frac{1}{B(a_1, b_1) B(a_2, b_2)} \int_{-\infty}^{\infty} \frac{\exp(\psi + \phi_2)^{a_1} \exp(\phi_2)^{a_2}}{(1 + \exp(\psi + \phi_2))^{a_1 + b_1} (1 + \exp(\phi_2))^{a_2 + b_2}} d\phi_2.$$

Das Integral lässt sich durch die Substitution $\pi = g^{-1}(\phi_2)$ zu

$$\int_0^1 g^{-1}(g(\pi) + \psi)^{a_1} \{1 - g^{-1}(g(\pi) + \psi)\}^{b_1} \pi^{a_2 - 1} (1 - \pi)^{b_2 - 1} d\pi$$

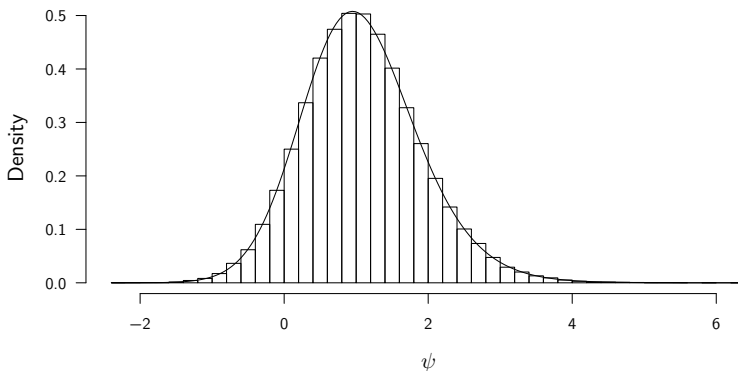
umformen, und kann leider nicht weiter vereinfacht werden. Die Dichtefunktion von ψ kann somit nur numerisch berechnet werden:

```
> ## Dichte von psi = logit(pi1) - logit(pi2), wenn pij ~ Be(a[j], b[j])
> psiDens <- function(psi, a, b, log = FALSE)
+ {
+   ## für jedes Element in psi separat
```

```

+   logRet <- psi
+   ## den Wert der Dichtefunktion berechnen
+   integrand <- function(pi, psi){
+     ## möglichst eingebaute Verteilungs-Funktionen benutzen
+     logRet <-
+       a[1] * plogis(qlogis(pi) + psi, log.p = TRUE) +
+       b[1] * plogis(qlogis(pi) + psi, lower = FALSE, log.p = TRUE) -
+       lbeta(a[1], b[1]) +
+       dbeta(pi, a[2], b[2], log = TRUE)
+     exp(logRet)
+   }
+   for(i in seq_along(psi)){
+     ## wenn Integration funktioniert hat, Wert speichern
+     intRes <- integrate(integrand, lower = 0, upper = 1, psi = psi[i],
+       stop.on.error = FALSE, rel.tol = 1e-6,
+       subdivisions = 200)
+     if(intRes$message == "OK")
+       logRet[i] <- log(intRes$value)
+     else
+       logRet[i] <- NA
+   }
+   ## gesamten Vektor zurückgeben
+   if(log)
+     return(logRet)
+   else
+     return(exp(logRet))
+ }
> ## Test wieder an simulierten Daten:
> histRes <- hist(psi, prob = TRUE, breaks = 50,
+   main = "", xlab = expression(psi))
> psiGrid <- seq(from = min(histRes$breaks), to = max(histRes$breaks),
+   length = 201)
> lines(psiGrid, psiDens(psiGrid, 0.5 + x, 0.5 + n - x))

```



Ein Problem ist hier, dass in einem Bereich um $\psi \approx 4.7$ die Integrationsroutine bei Standardeinstellungen nicht konvergiert. Nach dem Einstellen der relativen Genauigkeit (*rel.tol*) und der Zahl der Teilintervalle (*subdivisions*) auf andere Werte trat dieses Problem dann nicht mehr auf.

Wiederum mit numerischer Integration können wir den Posteriori-Erwartungswert von ψ berechnen – für diesen Schritt ist es notwendig, dass für jeden Wert ψ auch die Dichte $f(\psi|x)$ berechnet werden kann!

```
> ## numerisch:
> integrate(function(psi) psi * psiDens(psi, 0.5 + x, 0.5 + n - x),
+           lower = -Inf, upper = Inf)
1.079639 with absolute error < 0.00012
> ## Monte-Carlo war
> mean(psi)
[1] 1.082216
```

Der Monte-Carlo-Schätzwert liegt also recht nahe am numerisch ermittelten Erwartungswert. Relativ aufwändig ist auch die numerische Berechnung des HPD-Intervalls:

```
> ## für HPD-Intervall:
> ## Funktion outerdens liefert für den Wert h die Wahrscheinlichkeit aller
> ## Werte psi, deren Dichte kleiner als h ist.
> ## (a, b wie bei Funktion psiDens)
> outerdens <- function(h)
+ {
+   ## hier speziell für die Daten!
+   modus <- 1 # aus Grafik geschätzt
+   a <- 0.5 + x
+   b <- 0.5 + n - x
+   ## Schnittpunkte von h mit Dichte ermitteln
+   schnitt.links <-
+     uniroot(function(x) psiDens(x, a, b) - h, interval = c(-5, modus))$root
+   schnitt.rechts <-
+     uniroot(function(x) psiDens(x, a, b) - h, interval = c(modus, 5))$root
+   ## und Wahrscheinlichkeitsmasse außerhalb dieser Schnittpunkte
+   p.links <- integrate(psiDens, lower = -Inf, upper = schnitt.links,
+                       a = a, b = b)$value
+   p.rechts <- integrate(psiDens, lower = schnitt.rechts, upper = Inf,
+                       a = a, b = b)$value
+   ## diese mit Schnittpunkten zurückgeben
+   return(c(prob = p.links + p.rechts,
+            lower = schnitt.links, upper = schnitt.rechts))
+ }
> ## jetzt optimales h ermitteln: außerhalb sollen 5% liegen
> result <- uniroot(function(h) outerdens(h)[1] - 0.05,
+                   interval = c(0.001, 0.4))
> height <- result[["root"]]
> ## schließlich erhalten wir numerisch
> (psiHpdNum <- outerdens(height)[c(2,3)])
      lower      upper
-0.4898296  2.733044
> ## mit MC hatten wir vorher bekommen
> psiHpd
      lower      upper
-0.5111745  2.7162842
```

Auch das HPD-Intervall wurde also schon sehr gut aus den simulierten Daten geschätzt. Man sieht, dass die Monte-Carlo-Schätzung aber deutlich einfacher

ist (z. B. müssen schon keine Dichten berechnet werden!) und im Gegensatz zum numerischen Weg kein Tuning von Integrationsroutinen nötig ist.

3. In dieser Aufgabe wollen wir ein Bayesianisches hierarchisches Modell mit MCMC schätzen. Dazu betrachten wir unter Verwendung der Notation aus Beispiel 5.26 das Modell

$$\begin{aligned}\hat{\eta}_i | \eta_i &\sim N(\eta_i, \sigma_i^2), \\ \eta_i | \nu, \tau^2 &\sim N(\nu, \tau^2), \\ \nu &\sim N(0, 10) \\ \text{und } \tau^2 &\sim \text{IG}(1, 1),\end{aligned}$$

wobei die logarithmierten empirischen relativen Chancen $\hat{\eta}_i$ und die zugehörigen Varianzen $\sigma_i^2 := 1/a_i + 1/b_i + 1/c_i + 1/d_i$ bekannt seien.

- a) Leiten Sie zunächst die vollständig bedingten Verteilungen der unbekannten Parameter η_i ($i = 1, \dots, n$), ν und τ^2 her.

► Sei $i \in \{1, \dots, n = 9\}$. Die bedingte Dichte von η_i , gegeben alle anderen Parameter sowie die Daten \mathcal{D} (die man auf die empirischen log-relativen Chancen $\{\hat{\eta}_i\}$ sowie deren asymptotische Varianzen $\{\sigma_i^2\}$ reduzieren kann), ist

$$\begin{aligned}f(\eta_i | \{\eta_j\}_{j \neq i}, \nu, \tau^2, \mathcal{D}) &\propto f(\boldsymbol{\eta}, \nu, \tau^2, \mathcal{D}) \\ &\propto f(\hat{\eta}_i | \eta_i) f(\eta_i | \nu, \tau^2).\end{aligned}$$

Dies entspricht dem Normal-Normal-Modell in Beispiel 5.5: μ ist hier η_i sowie σ^2 hier σ_i^2 ist und x entspricht $\hat{\eta}_i$, sodass sich die vollständig bedingte Verteilung

$$\eta_i | \{\eta_j\}_{j \neq i}, \nu, \tau^2, \mathcal{D} \sim N\left(\left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{\hat{\eta}_i}{\sigma_i^2} + \frac{\nu}{\tau^2}\right), \left(\frac{1}{\sigma_i^2} + \frac{1}{\tau^2}\right)^{-1}\right)$$

ergibt. Eine Stufe höher in der Hierarchie, bei der vollständig bedingten Verteilung des Populationsmittels ν , liegt ebenfalls ein Normal-Normal-Modell vor:

$$f(\nu | \boldsymbol{\eta}, \tau^2, \mathcal{D}) \propto \prod_{i=1}^n f(\eta_i | \nu, \tau^2) \cdot f(\nu)$$

Hier nimmt beispielsweise η_i die Rolle der Zufallsvariable X_i aus Beispiel 5.5 ein. Mit Formel (5.7) ergibt sich

$$\nu | \boldsymbol{\eta}, \tau^2, \mathcal{D} \sim N\left(\left(\frac{n}{\tau^2} + \frac{1}{10}\right)^{-1} \left(\frac{n\bar{\eta}}{\tau^2} + \frac{0}{10}\right), \left(\frac{n}{\tau^2} + \frac{1}{10}\right)^{-1}\right).$$

Weiterhin haben wir

$$\begin{aligned} f(\tau^2 | \boldsymbol{\eta}, \nu, \mathcal{D}) &\propto \prod_{i=1}^n f(\eta_i | \nu, \tau^2) \cdot f(\tau^2) \\ &\propto \prod_{i=1}^n (\tau^2)^{-\frac{1}{2}} \exp\left(-\frac{(\eta_i - \nu)^2}{2\tau^2}\right) \cdot (\tau^2)^{-(1+1)} \exp(-1/\tau^2) \\ &= (\tau^2)^{-(\frac{n+2}{2}+1)} \exp\left(-\frac{\sum_{i=1}^n (\eta_i - \nu)^2 + 2}{2\tau^2}\right), \end{aligned}$$

was wir als

$$\tau^2 | \boldsymbol{\eta}, \nu, \mathcal{D} \sim \text{IG}\left(\frac{n+2}{2}, \frac{\sum_{i=1}^n (\eta_i - \nu)^2 + 2}{2}\right)$$

identifizieren können.

- b) Implementieren Sie nun einen Gibbs Sampler zur Simulation aus der zugehörigen Posteriori-Verteilung für die Daten aus Tabelle 1.1.

► Mit folgendem R-Code simulieren wir abwechselnd aus den vollständig bedingten Verteilungen von ν , τ^2 , η_1, \dots, η_n :

```
> ## Der Datensatz ist
> preeclampsia <- read.table("../Daten/preeclampsia.txt", header = TRUE)
> preeclampsia
```

	Trial	Diuretic	Control	Preeclampsia
1	Weseley	14	14	yes
2	Weseley	131	136	no
3	Flowers	21	17	yes
4	Flowers	385	134	no
5	Menzies	14	24	yes
6	Menzies	57	48	no
7	Fallis	6	18	yes
8	Fallis	38	40	no
9	Cuadros	12	35	yes
10	Cuadros	1011	760	no
11	Landesman	138	175	yes
12	Landesman	1370	1336	no
13	Krans	15	20	yes
14	Krans	506	524	no
15	Tervila	6	2	yes
16	Tervila	108	103	no
17	Campbell	65	40	yes
18	Campbell	153	102	no

```
> ## mit folgenden Funktionen für eine 2x2-Tafel square
> oddsRatio <- function (square)
+   (square[1,1] * square[2,2]) / (square[1,2] * square[2,1])
> variance <- function (square)
+   sum (1 / square)
> ## werden daraus die Daten extrahiert
> groups <- split (subset (preeclampsia, select = c (Diuretic, Control)),
+   preeclampsia[["Trial"]]) # Liste von 2x2-Tafeln
> (logOddsRatios <- log (sapply (groups, oddsRatio))) # etaDach-Vektor
```

```

      Campbell      Cuadros      Fallis      Flowers      Krans      Landesman
0.08004271 -1.35581820 -1.04731899 -0.84409444 -0.25272706 -0.26266295
      Menzies      Tervila      Weseley
-0.71084676  1.05121005  0.03745756
> (variances <- sapply (groups, variance))      # sigma^2-Vektor
      Campbell      Cuadros      Fallis      Flowers      Krans      Landesman      Menzies
0.05672448 0.11420967 0.27353801 0.11650267 0.12055135 0.01443909 0.15147243
      Tervila      Weseley
0.68563466 0.15784367
> n <- length(groups)      # Anzahl Studien
> ## Gibbs-Sampler für Inferenz in voll-bayesianischem Modell:
> niter <- 1e+5
> s <- matrix(nrow = 2 + n, ncol = niter)
> rownames(s) <- c("nu", "tau2", paste("eta", 1:n, sep = ""))
> etaIndices <- 3:(n + 2)
> ## Startwerte setzen: könnten auch andere Werte (im zulässigen Bereich) sein!
> s[, 1] <- c(nu = mean(logOddsRatios), tau2 = var(logOddsRatios), logOddsRatios)
> set.seed(59)
> ## jetzt reihum updaten:
> for(j in 2:niter){
+   ## erst nu
+   nuPrecision <- n / s["tau2",j-1] + 1 / 10
+
+   etaSum <- sum(s[etaIndices,j-1])
+   nuMean <- (etaSum / s["tau2",j-1]) / nuPrecision
+
+   s["nu",j] <- rnorm(1, mean = nuMean, sd = 1 / sqrt(nuPrecision))
+
+   ## dann tau2
+   sumEtaNuSquared <- sum( (s[etaIndices,j-1] - s["nu",j])^2 )
+
+   tau2a <- (n + 2) / 2
+   tau2b <- (sumEtaNuSquared + 2) / 2
+
+   s["tau2",j] <- 1 / rgamma(1, shape = tau2a, rate = tau2b)
+
+   ## zuletzt etai, ..., etan
+   for(i in 1:n){
+     etaiPrecision <- 1 / variances[i] + 1 / s["tau2",j]
+     etaiMean <- (logOddsRatios[i] / variances[i] + s["nu",j] / s["tau2",j]) /
+       etaiPrecision
+
+     s[etaIndices[i],j] <- rnorm(1, mean = etaiMean, sd = 1 / sqrt(etaiPrecision))
+   }
+ }

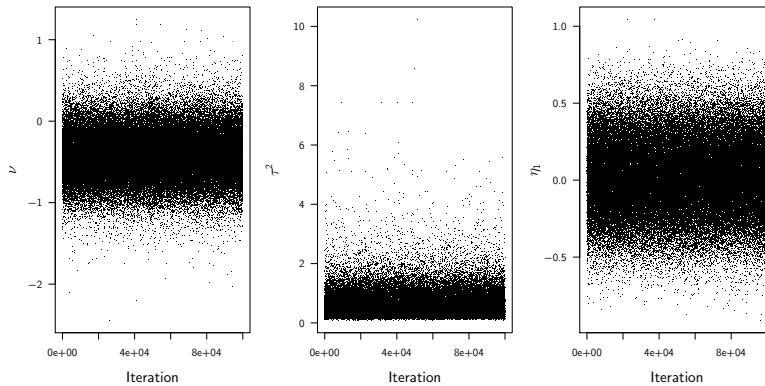
```

Nach der Erzeugung der Markov-Ketten sollte man einen Blick auf die Verläufe der Zufallszahlen werfen. Exemplarisch greifen wir die Variablen ν , τ^2 und η_1 heraus.

```

> par(mfrow = c(1, 3))
> for(j in 1:3){
+   plot(s[j, ], pch = ".",
+       xlab = "Iteration", ylab = expression(nu, tau^2, eta[1])[j])
+ }

```



Die Konvergenz scheint schnell erreicht worden zu sein, sodass ein Burn-In der Länge 1000 ausreichend scheint.

- c) Berechnen Sie 95%-Monte-Carlo-Kreditibilitätsintervalle für $\eta_i, i = 1, \dots, n$ und ν und stellen Sie diese graphisch dar. Vergleichen Sie die Ergebnisse mit denen aus Abbildung 5.13.

► Mit der Funktion zur Monte-Carlo-Schätzung von HPD-Intervallen aus Ziehungen der Posteriori (Aufgabe 2a)) berechnen wir empirische 95%-HPD-Intervalle und erstellen eine zu Abbildung 5.13 analoge Grafik:

```
> ## Burn-In entfernen
> s <- s[, -(1:1000)]
> ## HPD-Intervalle und Posteriori-Erwartungswerte schätzen
> (mcmcHpds <- apply(s, 1, hpd))

      nu      tau2      eta1      eta2      eta3      eta4
lower -0.9754692 0.1268147 -0.4291020 -1.7957179 -1.67182930 -1.3752161
upper 0.1505892 1.2740573 0.4670055 -0.5576167 0.02481017 -0.1557725
      eta5      eta6      eta7      eta8      eta9
lower -0.8982984 -0.49913721 -1.32713223 -0.9550535 -0.7640214
upper 0.3291375 -0.03449616 0.02959004 1.3667635 0.6166187

> (mcmcExpectations <- rowMeans(s))

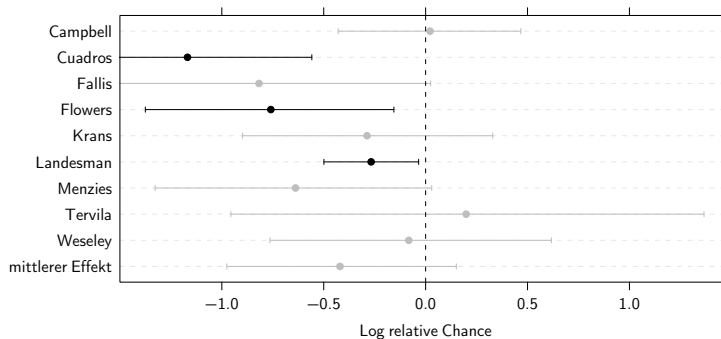
      nu      tau2      eta1      eta2      eta3      eta4
-0.42011446 0.57488918 0.02165663 -1.16823529 -0.81742313 -0.75950804
      eta5      eta6      eta7      eta8      eta9
-0.28747888 -0.26747976 -0.63833610 0.19910561 -0.08254766

> ## und Grafik erstellen
> library (lattice)
> panel.ci <- function(                                # eigene Panel-Funktion
+                               x,                        # Punktschätzung
+                               y,                        # Höhe des "Balkens"
+                               lx,                       # lower, untere Grenze des Intervalls
+                               ux,                       # upper, obere Grenze des Intervalls
+                               subscripts,               # Indexvektor
+                               ...                       # weitere Grafik-Argumente
+ )
+ {
+   x <- as.numeric(x)
+   y <- as.numeric(y)
```

```

+
+   lx <- as.numeric(lx[subscripts])
+   ux <- as.numeric(ux[subscripts])
+
+   panel.dotplot(x, y, lty = 2, ...)           # normal dotplot for point estimates
+   panel.arrows(lx, y, ux, y,                  # draw intervals
+               length = 0.1, unit = "native",
+               angle = 90,                      # deviation from line
+               code = 3,                        # left and right whisker
+               ...)
+   panel.abline (v = 0, lty = 2)              # reference line
+ }
> ## für Beschriftung
> studyNames <- c (names(groups), "mittlerer Effekt")
> studyNames <- ordered (studyNames, levels = rev (studyNames)) # levels important for order!
> indices <- c(etaIndices, nu = 1)
> ## Daten in data.frame zusammenfassen
> ciData <- data.frame (low = mcmcHpds["lower", indices],
+                       up = mcmcHpds["upper", indices],
+                       mid = mcmcExpectations[indices],
+                       names = studyNames
+                       )
> ciData[["signif"]] <- with (ciData,
+                             up < 0 | low > 0)
> ciData[["farbe"]] <- with (ciData,
+                             ifelse (signif, "black", "gray"))
> randomEffectsCiPlot <- with (ciData,
+                               dotplot (names ~ mid,
+                                       panel = panel.ci,
+                                       lx = low, ux = up,
+                                       pch = 19, col = farbe,
+                                       xlim = c (-1.5, 1.5), xlab = "Log relative Chance",
+                                       scales = list (cex = 1)
+                                       )
+                               )
> print (randomEffectsCiPlot)

```



Im Vergleich mit den Ergebnissen der empirischen Bayes-Analyse fällt auf, dass die Kreditabilitätsintervalle bei dieser voll-Bayesianischen Analyse deutlich breiter sind. Die Punktschätzung für den mittleren Effekt ν ist mit $E(\nu | \mathcal{D}) = -0.42$

recht ähnlich zur vorherigen ($\hat{\nu}_{ML} = -0.43$). Allerdings schließt das zugehörige Intervall nun die Null mit ein, was auch bei den Studien Fallis und Menzies zu beobachten ist. Die Bayesianischen Punktschätzungen der einzelnen Studien sind hier außerdem weniger gegen den mittleren Effekt geschrumpft, vgl. beispielsweise die Studie Tervila.

7 Modellwahl

1. Wir möchten nun zum einführenden Beispiel zurückkehren und ein adäquates Modell für die Überlebenszeitdaten der PBC-Patienten aus Beispiel 1.2.8 finden. Dabei sollen auch die rechtszensierten Beobachtungen berücksichtigt werden.

a) Benutzen Sie den Code aus Beispiel 4.9, um die maximale Log-Likelihood im Weibull-Modell zu ermitteln (Ergebnis: -424.0043).

```
► placebo <- read.table("Daten/placebo.txt", header = TRUE)
> ## Weibull: numerisch
> weibullLik <- function(mualpha, log = TRUE){
+   mu <- mualpha[1]
+   alpha <- mualpha[2]
+   loglik <-
+     with (placebo,
+           sum (d * dweibull (time, alpha, mu, log = TRUE) +
+                 (1 - d) * pweibull (time, alpha, mu,
+                                     lower.tail = FALSE, log.p = TRUE)
+             )
+   )
+   if (log)
+     return (loglik)
+   else
+     return (exp (loglik))
+ }
> start <- c (1000, 1)
> resultWeibull <- optim(start, weibullLik, control=list(fnscale=-1), hessian=TRUE)
> stopifnot (resultWeibull$convergence == 0)
> ## Ergebnis
> resultWeibull$value
[1] -424.0043
```

b) Berechnen Sie für das Gamma-Modell, das für unzensierte Daten bereits in Beispiel 2.3 beschrieben wurde, in der Parametrisierung mit μ und ϕ in analoger Vorgehensweise zu 1a) die maximierte Log-Likelihood (Ergebnis: -424.0047).

```
► ## Gamma: numerisch
> gammaLik <- function(muphi, log = TRUE){
+   mu <- muphi[1]
+   phi <- muphi[2]
+   loglik <-
+     with (placebo,
+           sum (d * dgamma (time, shape = mu / phi, scale = phi, log = TRUE) +
+                 (1 - d) * pgamma (time, shape = mu / phi, scale = phi,
+                                   lower.tail = FALSE, log.p = TRUE)
+             )
+   )
+   if (log)
+     return (loglik)
+ }
```



```

+     else
+         return (exp (loglik))
+ }
> start <- c (1000, 1000)
> resultGamma <- optim(start, gammaLik, control=list(fnscale=-1), hessian=TRUE)
> stopifnot (resultGamma$convergence == 0)
> ## Ergebnis:
> resultGamma$value
[1] -424.0047

```

- c) Im Exponential-Modell aus Beispiel 2.7 kann der ML-Schätzer analytisch gefunden werden. Setzen Sie diesen in die Log-Likelihoodfunktion ein, um schließlich $l(\hat{\lambda}_{ML}) = -424.0243$ zu erhalten.

```

> ## Exponential: analytisch
> lambdaMl <- with (placebo, sum (d) / sum (time))
> expLik <- function(lambda, log = TRUE){
+     loglik <-
+         with (placebo,
+             sum (d * dexp (time, lambda, log = TRUE) +
+                 (1 - d) * pexp (time, lambda,
+                                 lower.tail = FALSE, log.p = TRUE)
+             )
+         )
+     if (log)
+         return (loglik)
+     else
+         return (exp (loglik))
+ }
> expMaxLogLik <- expLik (lambdaMl)
> ## Ergebnis
> expMaxLogLik
[1] -424.0243

```

- d) Vergleichen Sie die drei Modelle mittels *AIC* und *BIC*.

► Im Weibull- und Gamma-Modell gibt es $p = 2$ Parameter, während das Exponential-Modell mit einem Parameter auskommt. Für das *BIC* wird außerdem die Zahl der Beobachtungen benötigt, die hier der Anzahl der Zeilen im *placebo*-Datensatz entspricht.

```

> ## AIC und BIC-Werte berechnen:
> aicWeibull <- 2*resultWeibull$value - 2 * 2
> bicWeibull <- 2*resultWeibull$value - 2 * log (nrow (placebo))
> aicGamma <- 2 * resultGamma$value - 2 * 2
> bicGamma <- 2 * resultGamma$value - 2 * log (nrow (placebo))
> aicExp <- 2 * expMaxLogLik - 1 * 2
> bicExp <- 2 * expMaxLogLik - 1 * log (nrow (placebo))
> ## und zusammenfassen
> result <- data.frame (
+     model = c ("Exp", "Weibull", "Gamma"),
+     AIC = c (aicExp, aicWeibull, aicGamma),
+     BIC = c (bicExp, bicWeibull, bicGamma)
+ )
> result

```

	model	AIC	BIC
1	Exp	-850.0486	-852.5919
2	Weibull	-852.0087	-857.0953
3	Gamma	-852.0094	-857.0960

Sowohl AIC als auch BIC sprechen für das sparsamere Exponential-Modell.

2. Wiederholen Sie die in Beispiel 7.3 geschilderte Analyse der Fischstudie mit nun unbekannter Varianz κ^{-1} . Verwenden Sie hierzu die in Beispiel 5.17 beschriebene konjugierte Normal-Gamma-Verteilung als Priori-Verteilung für κ und μ .

- a) Berechnen Sie zunächst durch die in (7.2) beschriebene Umstellung des Satzes von Bayes die marginale Likelihood in diesem Modell.

► Betrachten wir zunächst wieder Modell M_1 , das x_1, \dots, x_n als unabhängige Realisationen aus der $N(\mu, \kappa^{-1})$ -Verteilung ansieht. Als Priori-Verteilung für $\theta = (\mu, \kappa)^T$ nehmen wir die konjugierte Normal-Gamma-Verteilung $\theta \sim \text{NG}(\nu, \lambda, \alpha, \beta)$, sodass sich als Posteriori die Normal-Gamma-Verteilung aus (5.17) ergibt. Diese notieren wir im Folgenden als

$$\theta | x \sim \text{NG}(\nu^*, \lambda^*, \alpha^*, \beta^*).$$

Die explizite Form der Likelihood $f(x|\theta)$, der Priori-Dichte $f(\theta)$ und der Posteriori-Dichte $f(\theta|x)$ sind uns also bekannt, sodass wir die marginale Likelihood im Modell M_1 über

$$f(x|M_1) = \frac{f(x|\theta, M_1)f(\theta, M_1)}{f(\theta|x, M_1)}$$

herleiten können. Wichtig ist dabei, alle Normierungskonstanten der drei Dichtefunktionen mit einzubeziehen. Übrig bleiben schließlich genau diese, d. h.

$$\begin{aligned} f(x|M_1) &= \frac{(2\pi)^{-\frac{n}{2}} \frac{\beta^\alpha}{\Gamma(\alpha)} (2\pi)^{-\frac{1}{2}} \lambda^{\frac{1}{2}}}{\frac{(\beta^*)^{\alpha^*}}{\Gamma(\alpha^*)} (2\pi)^{-\frac{1}{2}} (\lambda^*)^{\frac{1}{2}}} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{\lambda}{\lambda+n}\right)^{\frac{1}{2}} \frac{\Gamma(\alpha+n/2)\beta^\alpha}{\Gamma(\alpha)} \left\{ \beta + \frac{n\hat{\sigma}_{ML}^2 + (\lambda+n)^{-1}n\lambda(\nu-\bar{x})^2}{2} \right\}^{-(\alpha+\frac{n}{2})}, \end{aligned} \quad (7.1)$$

wobei im letzten Schritt die Posteriori-Parameter eingesetzt wurden.

- b) Berechnen Sie nun explizit die Posteriori-Wahrscheinlichkeiten der (*a priori* gleich wahrscheinlichen) vier Modelle M_1 bis M_4 , wobei Sie als Priori-Verteilung für κ und μ eine $\text{NG}(25, 5, 1, 1000)$ -Verteilung verwenden.

► Wie in Beispiel 7.3 nehmen wir für die Modelle M_2 bis M_4 unabhängige Priori-Verteilungen für die Parameter θ_i der einzelnen Gruppen an, sodass sich die Werte der marginalen Likelihood durch Multiplikation von Termen der Form (7.1) ergeben. Wir implementieren also zunächst diese grundlegende Funktion:

```

> ## marginale (Log-)Likelihood im Normal-Normal-Gamma-Modell,
> ## bei n Realisationen mit Mittelwert mean und ML-Varianzschätzung var
> marginalLikelihood <- function(n, mean, var,          # Daten
+                               nu, lambda, alpha, beta, # Priori-Parameter
+                               log = FALSE           # soll log(f(x))
+                               # statt f(x) ausgegeben werden?
+                               )
+ {
+   betaStar <- beta +
+     (n * var + n * lambda * (nu - mean)^2 / (lambda + n)) / 2
+   logRet <- - n/2 * log(2 * pi) + 1/2 * log(lambda / (lambda + n)) +
+     lgamma(alpha + n/2) - lgamma(alpha) + alpha * log(beta) -
+     (alpha + n/2) * log(betaStar)
+   if(log)
+     return(logRet)
+   else
+     return(exp(logRet))
+ }

```

Anschließend geben wir Tabelle 1.3 in R ein und können mit den in der Aufgabenstellung angegebenen Priori-Parametern die marginalen Likelihood-Werte berechnen:

```

> ## Daten eingeben
> (fischdaten <-
+   data.frame(Gewicht = rep(c("Normalgewichtig", "Übergewichtig"), each = 2),
+     Behandlung = rep(c("Fischlos", "Fisch"), 2),
+     Anzahl = c(97, 98, 50, 63),
+     Mittel = c(23.2, 21.0, 40.8, 34.6),
+     Standardabweichung = c(32.05, 30.43, 34.23, 37.04))
+ )

```

	Gewicht	Behandlung	Anzahl	Mittel	Standardabweichung
1	Normalgewichtig	Fischlos	97	23.2	32.05
2	Normalgewichtig	Fisch	98	21.0	30.43
3	Übergewichtig	Fischlos	50	40.8	34.23
4	Übergewichtig	Fisch	63	34.6	37.04

```

> attach(fischdaten)
> ## Priori-Parameter:
> nu <- 25
> lambda <- 5
> alpha <- 1
> beta <- 1000
> ## hier kommen log-marginale Likelihoodwerte rein
> logMargLik <- numeric(4)
> ## Modell M1: zunächst Daten aus Gruppen-Daten rekonstruieren
> n <- sum(Anzahl)
> (mittel <- weighted.mean(Mittel, w = Anzahl))
[1] 27.68896
> (var <- sum ((Anzahl - 1) * Standardabweichung^2) / n)
[1] 1074.372
> ## dann log-marginale Likelihood bestimmen
> (logMargLik[1] <- marginalLikelihood(n = n, mean = mittel, var = var,
+                                     nu, lambda, alpha, beta,
+                                     log = TRUE))

```

```

[1] -1516.564
> ## Modell M2: Normal vs. Übergewichtige
> (n <- tapply(Anzahl, Gewicht, sum))
Normalgewichtig  Übergewichtig
             195             113
> (mittel <- tapply(Anzahl * Mittel, Gewicht, sum) / n)
Normalgewichtig  Übergewichtig
             22.09436             37.34336
> (var <- tapply((Anzahl - 1) * Standardabweichung^2, Gewicht, sum) / n)
Normalgewichtig  Übergewichtig
             966.3178             1260.8369
> ## zunächst log-Marginale Likelihoodwerte in den Gruppen berechnen
> einzelLogMargLiks <-
+   marginalLikelihood(n = n, mean = mittel, var = var,
+                       nu, lambda, alpha, beta,
+                       log = TRUE)
> ## dann summieren
> (logMargLik[2] <- sum(einzelLogMargLiks))
[1] -1518.824
> ## Modell M3: Fischlos vs. Fisch
> ## hier wird einfach "Gewicht" durch "Behandlung" ersetzt!
> (n <- tapply(Anzahl, Behandlung, sum))
Fisch Fischlos
             161             147
> (mittel <- tapply(Anzahl * Mittel, Behandlung, sum) / n)
Fisch Fischlos
26.32174 29.18639
> (var <- tapply((Anzahl - 1) * Standardabweichung^2, Behandlung, sum) / n)
Fisch Fischlos
1086.225 1061.390
> ## zunächst log-Marginale Likelihoodwerte in den Gruppen berechnen
> einzelLogMargLiks <-
+   marginalLikelihood(n = n, mean = mittel, var = var,
+                       nu, lambda, alpha, beta,
+                       log = TRUE)
> ## dann summieren
> (logMargLik[3] <- sum(einzelLogMargLiks))
[1] -1519.888
> ## Modell M4: alle Gruppen separat
> ## zunächst log-Marginale Likelihoodwerte in den Gruppen berechnen
> einzelLogMargLiks <-
+   marginalLikelihood(n = Anzahl, mean = Mittel, var =
+                       (Standardabweichung^2 * (Anzahl - 1) / Anzahl),
+                       nu, lambda, alpha, beta,
+                       log = TRUE)
> ## dann summieren
> (logMargLik[4] <- sum(einzelLogMargLiks))
[1] -1524.138

```

Nimmt man die Präzision κ als unbekannt an, sind die Modelle bzgl. ihrer marginalen Likelihood also gerade nach ihrer Nummerierung geordnet: Modell M_1

hat die höchste marginale Likelihood, Modell M_4 die niedrigste. Setzen wir die Priori-Wahrscheinlichkeiten der Modelle fest auf beispielsweise

$$P(M_1) = \dots = P(M_4) = \frac{1}{4},$$

so können wir die Posteriori-Wahrscheinlichkeiten berechnen. Bei wie hier gleich wahrscheinlichen Modellen gilt

$$\begin{aligned} P(M_i | x) &= \frac{f(x | M_i)}{\sum_{j=1}^4 f(x | M_j)} \\ &= \frac{f(x | M_i)/c}{\sum_{j=1}^4 f(x | M_j)/c}, \end{aligned}$$

wobei die Erweiterung mit einer Konstante c^{-1} nötig ist, um zu vermeiden, dass die implementierte Exponentialfunktion im vorliegenden Bereich um -1520 den Wert Null zurückgibt.

```
> const <- min(logMargLik)
> posterioriWkeiten <- exp(logMargLik - const)
> (posterioriWkeiten <- posterioriWkeiten / sum(posterioriWkeiten))
[1] 0.8764567399 0.0915041616 0.0315889001 0.0004501984
```

Die Ordnung der Modelle bleibt hierbei unverändert bei der bisherigen durch die marginale Likelihood.

- c) Studieren Sie das Verhalten der Posteriori-Wahrscheinlichkeiten für unterschiedliche Parameter der NG-Priori-Verteilung.

► Den R-Code aus 2b) kann man in eine Funktion *modellvergleich* verpacken, die die Priori-Parameter entgegennimmt und die Posteriori-Wahrscheinlichkeiten der vier Modelle ausgibt. Im Folgenden werden wir jeweils einen Parameter variieren.

```
> ## alte Parameter
> modellvergleich(nu, lambda, alpha, beta)
[1] 0.8764567399 0.0915041616 0.0315889001 0.0004501984

> ## nu variieren
> for(nuNeu in c(0, 15, 35, 50))
+   print(modellvergleich(nuNeu, lambda, alpha, beta))
[1] 9.762386e-01 1.718033e-02 6.577915e-03 3.160164e-06
[1] 0.9131827285 0.0634198371 0.0232485500 0.0001488844
[1] 0.8800473606 0.0907087116 0.0288313046 0.0004126231
[1] 9.446636e-01 4.351264e-02 1.178557e-02 3.823971e-05

> ## lambda
> for(lambdaNeu in c(1e-100, 1e-10, 100, 1e+10))
+   print(modellvergleich(nu, lambdaNeu, alpha, beta))
[1] 1.000000e+00 6.433633e-52 1.696585e-52 1.033910e-154
[1] 9.999992e-01 6.433627e-07 1.696583e-07 1.033910e-19
[1] 0.9028800159 0.0167216966 0.0802654665 0.0001328209
[1] 8.999925e-01 9.101361e-04 9.907153e-02 2.583428e-05
```

```

> ## alpha
> for(alphaNeu in c(1e-10, 5, 100, 1e+10))
+   print(modellvergleich(nu, lambda, alphaNeu, beta))
[1] 1.000000e+00 1.212249e-11 3.940076e-12 7.773370e-34
[1] 9.957950e-01 2.937570e-03 1.267414e-03 1.946857e-08
[1] NaN 0 0 0
[1] NaN NaN NaN 0
> ## beta
> for(betaNeu in c(1e-10, 0.001, 1, 1e+10))
+   print(modellvergleich(nu, lambda, alpha, betaNeu))
[1] 1.000000e+00 2.576444e-14 9.237843e-15 8.320705e-42
[1] 9.999996e-01 2.576441e-07 9.237831e-08 8.320679e-21
[1] 9.996504e-01 2.573197e-04 9.225847e-05 8.294308e-12
[1] 1.000000e+00 4.250718e-44 1.992545e-46 3.877156e-89

```

Während die Wahl von ν die Modell-Wahrscheinlichkeiten kaum beeinflusst, führen kleinere Werte von λ zu einer stärkeren Bevorzugung der einfachen Modelle. Auch ein kleiner Priori-Parameter α scheint zu einer Präferenz der einfachen Modelle zu führen. Bei β scheinen extrem kleine oder große Werte dieses Phänomen hervorzurufen.

3. Gegeben sei eine Zufallsstichprobe X_1, \dots, X_n aus einer Normalverteilung mit Mittelwert μ und bekannter Varianz κ^{-1} , für die wir zwei Modelle vergleichen wollen. Im ersten Modell (M_1) sei der Parameter μ bei $\mu = \mu_0$ fixiert. Im zweiten Modell (M_2) sei der Parameter μ unbekannt mit Priori-Verteilung $\mu \sim N(\nu, \lambda^{-1})$, wobei ν und λ fixiert seien.

- a) Man berechne analytisch den Bayes-Faktor B_{12} von Modell M_1 im Vergleich mit Modell M_2 .

► Im Modell M_1 ist kein Parameter unbekannt, sodass die marginale Likelihood gleich der gewöhnlichen Likelihood

$$f(x | M_1) = \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\kappa}{2} \sum_{i=1}^n (x_i - \mu_0)^2\right)$$

ist. Die marginale Likelihood im Modell M_2 ist in (7.4) angegeben. Der Bayes-Faktor von M_1 gegenüber M_2 ergibt sich daraus zu

$$\begin{aligned}
 B_{12} &= \frac{f(x | M_1)}{f(x | M_2)} \\
 &= \left(\frac{n\kappa + \lambda}{\lambda}\right)^{\frac{1}{2}} \exp\left\{-\frac{\kappa}{2} \left(\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n\lambda}{n\kappa + \lambda} (\bar{x} - \nu)^2\right)\right\}.
 \end{aligned}
 \tag{7.2}$$

- b) Man berechne exemplarisch für die vier in der Fischstudie vorliegenden Datensätze den Bayes-Faktor für $\mu_0 = 0$, $\nu = 0$ und $\lambda = 1/100$.

- a) Man zeige dann, dass für den Bayes-Faktor B_{01} bei beliebiger Priori-Verteilung für μ gilt:

$$B_{01} \geq \exp \left\{ -\frac{1}{2} z^2 \right\},$$

wobei $z = x/\sigma$ unter Annahme von Modell M_0 standardnormalverteilt ist. Man nennt $\exp(-1/2z^2)$ den *Minimum-Bayes-Faktor*.

► Die Situation ist ähnlich wie in Aufgabe 3; κ^{-1} entspricht σ^2 , $n = 1$ und $\mu_0 = 0$. Die marginale Likelihood im Modell M_0 ist somit

$$f(x | M_0) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} z^2 \right\},$$

wobei $z = x/\sigma$ ist. Allerdings ist nun die Priori-Verteilung $f(\mu)$ von μ im Alternativ-Modell M_1 unbestimmt. Zur Bestimmung des Minimum-Bayes-Faktors von M_0 im Vergleich zu M_1 müssen wir $f(x | M_1)$ bzgl. der Wahl von $f(\mu)$ maximieren. Betrachtet man die Definition

$$f(x | M_1) = \int f(x | \mu) f(\mu) d\mu,$$

so wird deutlich, dass die Priori-Verteilung die Likelihood-Werte von μ mittelt. Intuitiv ist dann klar, dass die Ein-Punkt-Verteilung im ML-Schätzer $\hat{\mu}_{ML}$ die Priori-Verteilung ist, die $f(x | M_1)$ maximiert. Wir haben also hier wegen $\hat{\mu}_{ML} = x$

$$\begin{aligned} f(x | M_1) &\leq f(x | \hat{\mu}_{ML}) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \hat{\mu}_{ML}}{\sigma} \right)^2 \right\} \\ &= (2\pi\sigma^2)^{-\frac{1}{2}}, \end{aligned}$$

woraus

$$B_{01} = \frac{f(x | M_0)}{f(x | M_1)} \geq \exp \left\{ -\frac{1}{2} z^2 \right\}$$

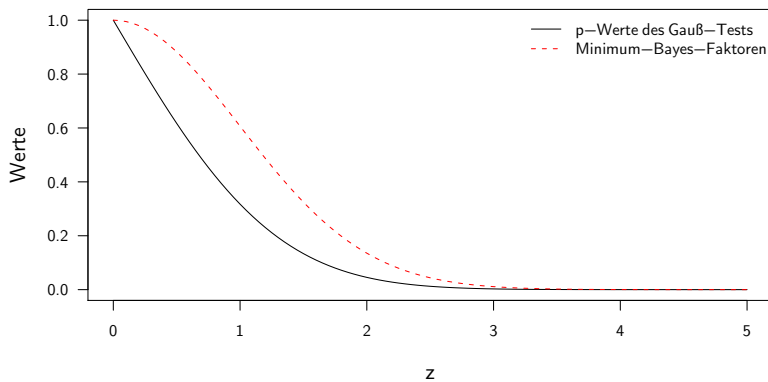
folgt.

- b) Man berechne für verschiedene Werte von $z \geq 0$ einerseits den zugehörigen p-Wert $2(1 - \Phi(z))$ des einfachen Gauß-Tests und andererseits den Minimum-Bayes-Faktor, und vergleiche die Ergebnisse.


```

► ## Min-Bayes-Faktor:
> mbf <- function(z)
+   exp(-1/2 * z^2)
> ## für diese Werte von z
> zgrid <- seq(0, 5, length = 101)
> ## P-Werte mit mbf vergleichen
> pvalues <- 2 * (1 - pnorm(zgrid))
> mbfvalues <- mbf(zgrid)
> ## und plotten
> matplot(zgrid, cbind(pvalues, mbfvalues), type = "l",
+   xlab = expression(z), ylab = "Werte")
> legend("topright", legend = c("p-Werte des Gauß-Tests", "Minimum-Bayes-Faktoren"),
+   col = 1:2, lty = 1:2, bty = "n")
> ## Vergleiche
> all(pvalues <= mbfvalues)
[1] TRUE
> zgrid[pvalues == mbfvalues]
[1] 0

```



Die p -Werte des Gauß-Tests sind also für alle Werte von z höchstens so groß wie die Minimum-Bayes-Faktoren. Lediglich für $z = 0$ ergibt sich die Gleichheit $p = \min B_{01} = 1$. Für $z > 3$ werden sie sich sehr ähnlich.

8 Prognose

1. Fünf Ärzte nehmen an einer Studie zur Untersuchung der Wirkung eines Medikaments gegen Migräne teil. Arzt $i = 1, \dots, 5$ behandelt n_i Patienten mit dem neuen Medikament, wobei sich zeigt, dass es bei y_i der Patienten einen positiven Effekt hat. Sei π die Wahrscheinlichkeit, dass ein Migräne-Erkrankter positiv auf das Medikament reagiert. Angenommen

$$n = (3, 2, 4, 4, 3) \quad \text{und} \quad y = (2, 1, 4, 3, 3).$$

- a) Geben Sie einen Ausdruck für die Likelihood $L(\pi)$ in der obigen Studie an.

► Nehmen wir an, dass sowohl die Patienten bei einem Arzt als auch die Studien-Ergebnisse der Ärzte unabhängig voneinander sind, können wir y_i als Realisation von

$$Y_i \stackrel{iid}{\sim} \text{Bin}(n_i, \pi), \quad i = 1, \dots, m = 5,$$

auffassen. Die Likelihood von π ist demnach

$$\begin{aligned} L(\pi) &= \prod_{i=1}^m \binom{n_i}{y_i} \pi^{y_i} (1 - \pi)^{n_i - y_i} \\ &\propto \pi^{m\bar{y}} (1 - \pi)^{m\bar{n} - m\bar{y}}, \end{aligned}$$

wobei die mittlere Zahl von behandelten Patienten pro Arzt mit \bar{n} bezeichnet ist.

- b) Spezifizieren Sie eine konjugierte Priori-Verteilung $f(\pi)$ für π und wählen Sie angemessene Werte für deren Parameter. Bestimmen Sie anschließend mit diesen Parametern die Posteriori-Verteilung $f(\pi | n, y)$.

► Man weiß bzw. sieht an der Form der Likelihood, dass die Beta-Verteilung $\text{Be}(\alpha, \beta)$ mit Kern

$$f(\pi) \propto \pi^{\alpha-1} (1 - \pi)^{\beta-1}$$

konjugiert ist, vgl. auch die Bemerkung nach Beispiel 5.4. Mittels $\alpha = \beta = 1/2$ erhält man die nichtinformative Jeffreys' Priori, die zur Posteriori-Verteilung

$$\pi | n, y \sim \text{Be}(m\bar{y} + 1/2, m\bar{n} - m\bar{y} + 1/2)$$

führt.

- c) Ein sechster Arzt möchte mit $n_6 = 5$ Patienten an der Studie teilnehmen. Bestimmen Sie die posteriori-prädiktive Verteilung für y_6 (die Anzahl der fünf Patienten, auf die das Medikament einen positiven Effekt haben wird).

- Zunächst müssen wir die Dichte der posteriori-prädiktiven Verteilung,

$$f(y_6 | n_6, y, n) = \int_0^1 f(\pi | y, n) f(y_6 | \pi, n_6) d\pi,$$

bestimmen. Dies ist die Definition der Dichte der Beta-Binomial-Verteilung

$$y_6 | n_6, y, n \sim \text{BeB}(n_6, m\bar{y} + 1/2, m\bar{n} - m\bar{y} + 1/2),$$

deren analytischer Form in Tabelle A.1 man sich zur Übung vergewissern sollte. Für die vorliegenden Daten können wir somit ein Prognoseintervall sowie den Erwartungswert der Prognoseverteilung als Punktprognose für y_6 berechnen:

```
> ## vorliegende Beobachtungen
> n <- c(3, 2, 4, 4, 3)
> y <- c(2, 1, 4, 3, 3)
> ## Posteriori-Parameter sind (unter Jeffreys' Priori)
> alphaStar <- sum(y) + 0.5
> betaStar <- sum(n - y) + 0.5
> ## Posteriori-prädiktive Vtlg. ist BeBin mit diesen P. und
> nNew <- 5
> ## als neue Anzahl behandelter Patienten.
> (expectation <- nNew * alphaStar / (alphaStar + betaStar))
[1] 3.970588
> ## ist Punktprognose, Intervall über Vtlgs-Funktion:
> library(VGAM, warn.conflicts = FALSE)
> rbind(0:5,
+       pbetabin.ab(0:5, size = nNew, alphaStar, betaStar))
      [,1]      [,2]      [,3]      [,4]      [,5] [,6]
[1,] 0.00000000 1.00000000 2.00000000 3.00000000 4.00000000 5
[2,] 0.001729392 0.01729392 0.08673568 0.2824352 0.6412176 1
```

Das 2.5%-Quantil ist somit 2 und das 97.5%-Quantil 5, sodass das 95%-Prognoseintervall gleich $[2, 5]$ ist. Natürlich kann dieses wegen der diskreten Verteilung von Y_6 das nominale Niveau nicht einhalten; die prognostizierte Wahrscheinlichkeit, dass Y_6 in diesem Intervall liegt, ist vielmehr

$$1 - P(Y_6 \leq 1) = 0.9827.$$

- d) Berechnen Sie auch die prädiktive Likelihood-Prognose.

- Die erweiterte Likelihood-Funktion ist

$$L(\pi, y_6) = \binom{n_6}{y_6} \pi^{m\bar{y} + y_6} (1 - \pi)^{m\bar{n} + n_6 - m\bar{y} - y_6}.$$

Wenn man y_6 beobachtet hätte, wäre der ML-Schätzer von π gleich dem arithmetischen Mittel

$$\hat{\pi}(y_6) = \frac{m\bar{y} + y_6}{m\bar{n} + n_6},$$

sodass die prädiktive Likelihood gleich

$$L_p(y_6) = L(\hat{\pi}(y_6), y_6) \\ = \binom{n_6}{y_6} \left(\frac{m\bar{y} + y_6}{m\bar{n} + n_6} \right)^{m\bar{y} + y_6} \left(\frac{m\bar{n} + n_6 - m\bar{y} - y_6}{m\bar{n} + n_6} \right)^{m\bar{n} + n_6 - m\bar{y} - y_6}$$

ist. Die Likelihood-Prognoseverteilung

$$f_p(y_6) = \frac{L_p(y_6)}{\sum_{y=0}^{n_6} L_p(y)}$$

kann daraus numerisch berechnet werden:

```
> ## prädiktive Likelihood
> predLik <- function(yNew, nNew)
+ {
+   sumY <- sum(y) + yNew
+   sumN <- sum(n) + nNew
+   pi <- sumY / sumN
+
+   logRet <- lchoose(nNew, yNew) + sumY * log(pi) + (sumN - sumY) * log(1 - pi)
+   return(exp(logRet))
+ }
> ## Werte der diskreten Prognosedichte sind also
> prognoseWkeiten <- predLik(0:5, 5)
> (prognoseWkeiten <- prognoseWkeiten / sum(prognoseWkeiten))
[1] 0.001720862 0.015495708 0.069189316 0.195223635 0.358595044 0.359775434
> ## Verteilungsfunktion ist
> cumsum(prognoseWkeiten)
[1] 0.001720862 0.017216571 0.086405887 0.281629522 0.640224566 1.000000000
Die Werte der diskreten Dichte- bzw. Verteilungsfunktion sind sehr ähnlich zu
denen aus der Bayes-Prognose. Wiederum ist das 95%-Prognoseintervall gleich
[2, 5]; als Punktprognose erhalten wir hier
> sum((0:5) * prognoseWkeiten)
[1] 3.972803
```

und damit nur einen leicht anderen Wert als 3.9706 aus der Bayes-Prognose.

2. Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer $N(\mu, \sigma^2)$ -Verteilung, aus der eine weitere Beobachtung $Y = X_{n+1}$ vorhergesagt werden soll. Sowohl der Erwartungswert μ als auch die Varianz σ^2 seien unbekannt.

a) Bestimmen Sie zunächst die Plug-in-Prognoseverteilung.

► Im Gegensatz zu Beispiel 8.2 ist hier auch die Varianz σ^2 unbekannt. Nach Beispiel 4.3 lauten die ML-Schätzer

$$\hat{\mu}_{ML} = \bar{x} \quad \text{und} \quad \hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (8.1)$$

Als Plug-in-Prognoseverteilung erhalten wir somit

$$Y \sim N\left(\bar{x}, \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right).$$

- b) Berechnen Sie nun die prädiktive Likelihood- und die Bootstrap-Prognoseverteilung.

► Die erweiterte Likelihoodfunktion ergibt sich zu

$$L(\mu, \sigma^2, y) = f(y | \mu, \sigma^2) \cdot L(\mu, \sigma^2) \\ \propto (\sigma^2)^{-\frac{n+1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left((y - \mu^2) + n(\bar{x} - \mu)^2 + \sum_{i=1}^n (x_i - \bar{x})^2 \right) \right\}.$$

Die ML-Schätzungen aus den vervollständigten Daten sind

$$\hat{\mu}(y) = \frac{n\bar{x} + y}{n+1} \quad \text{und} \quad \hat{\sigma}^2(y) = \frac{1}{n+1} \left(\sum_{i=1}^n (x_i - \hat{\mu}(y))^2 + (y - \hat{\mu}(y))^2 \right)$$

und führen zur prädiktiven Likelihood

$$L_p(y) = L(\hat{\mu}(y), \hat{\sigma}^2(y), y),$$

die allerdings nur in der konkreten Datensituation numerisch zur Prognosedichte $f(y) = L_p(y) / \int L_p(u) du$ normiert werden kann.

Zur Bestimmung der Bootstrap-Prognoseverteilung benötigen wir die Verteilung der ML-Schätzer in (8.1). Wir benutzen dazu das Ergebnis aus der Theorie der linearen Modelle, dass

$$\hat{\mu}_{ML} = \bar{X} \sim N(\mu, \sigma^2/n) \quad \text{unabhängig von} \quad \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

ist (vgl. Aufgabe 3 in Kapitel 3). Daraus folgt nämlich mit $\chi^2(d) = G(d/2, 1/2)$

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \sim G\left(\frac{n-1}{2}, \frac{n}{2\sigma^2}\right).$$

Anschließend müssen wir die Bootstrap-Dichte von y gegeben $\theta = (\mu, \sigma^2)^T$,

$$g(y; \theta) = \int \int f(y | \hat{\mu}_{ML}, \hat{\sigma}_{ML}^2) f(\hat{\mu}_{ML} | \mu, \sigma^2) f(\hat{\sigma}_{ML}^2 | \mu, \sigma^2) d\hat{\mu}_{ML} d\hat{\sigma}_{ML}^2 \\ = \int_0^\infty \int_{-\infty}^\infty f(y | \hat{\mu}_{ML}, \hat{\sigma}_{ML}^2) f(\hat{\mu}_{ML} | \mu, \sigma^2) d\hat{\mu}_{ML} f(\hat{\sigma}_{ML}^2 | \mu, \sigma^2) d\hat{\sigma}_{ML}^2,$$

bestimmen. Das innere Integral können wir als marginale Likelihood im Normal-Normal-Modell identifizieren und mit (7.4) ergibt sich

$$\int_{-\infty}^\infty f(y | \hat{\mu}_{ML}, \hat{\sigma}_{ML}^2) f(\hat{\mu}_{ML} | \mu, \sigma^2) d\hat{\mu}_{ML} \\ = (2\pi\hat{\sigma}_{ML}^2)^{-\frac{1}{2}} \left(\frac{1/\sigma^2}{1/\hat{\sigma}_{ML}^2 + 1/\sigma^2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\hat{\sigma}_{ML}^2} \left(\frac{1/\sigma^2}{1/\hat{\sigma}_{ML}^2 + 1/\sigma^2} (y - \mu)^2 \right) \right\}. \quad (8.2)$$

Den zweiten Integrationsschritt kann man aber nicht mehr einfach analytisch vornehmen. Um letztendlich auf $g(y; \hat{\theta}_{ML})$ zu kommen, bietet sich deshalb die Monte-Carlo-Integration an: Wir ziehen viele Zufallszahlen $(\hat{\sigma}_{ML}^2)^{(i)}$ aus der $G((n-1)/2, n/(2\hat{\sigma}_{ML}^2))$ -Verteilung, setzen diese in (8.2) ein und berechnen den Mittelwert an den gewünschten Werten von y . Dies funktioniert natürlich wieder nur bei Vorliegen eines konkreten Datensatzes x_1, \dots, x_n .

- c) Bestimmen Sie schließlich unter Annahme der Referenz-Priori $f(\mu, \sigma^2) \propto \sigma^{-2}$ die Bayes-Prognoseverteilung.

► Wie in Beispiel 5.19 ist es praktisch, zur Präzision $\kappa = (\sigma^2)^{-1}$ und der zugehörigen Referenz-Priori $f(\mu, \kappa) \propto \kappa^{-1}$ überzugehen, die formal einer Normal-Gamma-Verteilung $NG(0, 0, -1/2, 0)$ entspricht. Die resultierende Posteriori-Verteilung können wir somit aus (5.19) ablesen – diese ist wieder eine Normal-Gamma-Verteilung, deren Parameter zu ν^* , λ^* , α^* und β^* aufdatiert wurden. In Aufgabe 2a) in Kapitel 7 hatten wir bereits die priori-prädiktive Verteilung der Daten in diesem Modell hergeleitet. Ein Vorteil der Benutzung von konjugierten Priori-Verteilungen ist aber gerade, dass aus der priori-prädiktiven Verteilung auch die posteriori-prädiktive Verteilung hergeleitet werden kann, indem man zum einen die Priori-Parameter durch die Posteriori-Parameter ersetzt und zum anderen die Likelihood entsprechend anpasst. Dies wollen wir zunächst tun. Dazu setzen wir in (7.1) $n = 1$, $\hat{\sigma}_{ML}^2 = 0$ sowie $\bar{x} = y$, da wir ja nur eine Beobachtung y prognostizieren wollen. Anschließend ersetzen wir die Priori-Parameter durch die Posteriori-Parameter und erhalten

$$\begin{aligned} f(y|x) &= \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \left(\frac{\lambda^*}{\lambda^* + 1}\right)^{\frac{1}{2}} \frac{\Gamma(\alpha^* + 1/2)(\beta^*)^{\alpha^*}}{\Gamma(\alpha^*)} \left\{ \beta^* + \frac{(\lambda^* + 1)^{-1} \lambda^* (\nu^* - y)^2}{2} \right\}^{-(\alpha^* + \frac{1}{2})} \\ &\propto \left(\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{(n+1)^{-1} n (\bar{x} - y)^2}{2} \right)^{-\left(\frac{n-1}{2} + \frac{1}{2}\right)} \\ &= \left(\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{-\frac{n}{2}} \left(1 + \frac{2n(y - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2 (n+1)2} \right)^{-\frac{n}{2}} \\ &\propto \left(1 + \frac{(y - \bar{x})^2}{(n-1) \left(1 + \frac{1}{n}\right) \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^{-\frac{(n-1)+1}{2}}, \end{aligned}$$

was dem Kern der t -Verteilung aus dem Nachtrag zu Beispiel 8.6 entspricht. Damit haben wir die Form der Bayes-Prognoseverteilung $f(y|x)$ hergeleitet.

3. Beweisen Sie die Murphy Zerlegung (8.8) des Brier-Scores.

► Zunächst halten wir eine nützliche Identität fest. Da die Realisationen y_i binär sind, ist das Mittel aus den quadrierten Beobachtungen (\bar{y}^2) gleich dem Mittel aus den ursprünglichen Beobachtungen (\bar{y}). Mit dem Verschiebungssatz folgt daraus

$$\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \bar{y}^2 - \bar{y}^2 = \bar{y}(1 - \bar{y}).$$

Nun betrachten wir die in J Gruppen eingeteilten Beobachtungen y_{ji} , $i = 1, \dots, n_j$, $j = 1, \dots, J$, mit Gruppenmittelwerten

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ji}$$

und prognostizierten Wahrscheinlichkeiten π_j . Insgesamt sind es $N = \sum_{j=1}^J n_j$ Realisationen mit Gesamtmittelwert (oder relativer Häufigkeit) \bar{y} . Wir setzen die Murphy Zerlegung an und benutzen das eben gezeigte Resultat:

$$\begin{aligned} (8.8) &= \bar{y}(1 - \bar{y}) + S - M \\ &= \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ji} - \bar{y})^2 + \frac{1}{N} \sum_{j=1}^J n_j (\bar{y}_j - \pi_j)^2 - \frac{1}{N} \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 \\ &= \frac{1}{N} \sum_{j=1}^J \left(\sum_{i=1}^{n_j} (y_{ji} - \bar{y})^2 \right) + n_j (\bar{y}_j - \pi_j)^2 - n_j (\bar{y}_j - \bar{y})^2. \end{aligned}$$

Das Ziel ist der mittlere Brier-Score

$$BS = \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ji} - \pi_j)^2,$$

sodass wir diesen in den ersten Term einbauen:

$$\sum_{i=1}^{n_j} (y_{ji} - \bar{y})^2 = \sum_{i=1}^{n_j} (y_{ji} - \pi_j)^2 + 2(\pi_j - \bar{y}) \sum_{i=1}^{n_j} (y_{ji} - \pi_j) + n_j (\pi_j - \bar{y})^2$$

Somit haben wir

$$(8.8) = BS + \frac{1}{N} \sum_{j=1}^J 2(\pi_j - \bar{y})(n_j \bar{y}_j - n_j \pi_j) + n_j (\pi_j - \bar{y})^2 + n_j (\bar{y}_j - \pi_j)^2 - n_j (\bar{y}_j - \bar{y})^2$$

und durch Ausmultiplizieren der quadratischen Terme überzeugt man sich leicht davon, dass die einzelnen Summanden Null ergeben, sodass die Murphy Zerlegung des Brier-Scores bewiesen ist.

4. Untersuchen Sie, ob die Bewertungsregel

$$S(f(y), y_b) = -f(y_b)$$

bei einer binären Beobachtung Y korrekt ist.

► Sei wie in Beispiel 8.16 die tatsächliche Verteilung der Beobachtung $Y_b \sim B(\pi_2)$ und f die (diskrete) Dichte der Prognoseverteilung $Y \sim B(\pi_1)$. Der unter der wahren Verteilung erwartete Score $S(f(y), y_b) = -f(y_b)$ ist dann

$$\begin{aligned} \mathbb{E}[S(f(y), Y_b)] &= -\mathbb{E}[f(Y_b)] \\ &= -f(0) \cdot (1 - \pi_2) - f(1) \cdot \pi_2 \\ &= \pi_1(1 - 2\pi_2) + \pi_2 - 1 \end{aligned}$$

und somit als Funktion von π_1 eine Gerade mit Steigung $1 - 2\pi_2$. Falls diese positiv bzw. negativ ist, ist das Minimum also bei $\pi_1 = 0$ bzw. $\pi_1 = 1$ (vgl. Beispiel 8.15 zum absoluten Score). Dies ist nicht (immer) gleich π_2 , sodass die Bewertungsregel nicht korrekt ist.

5. Zeigen Sie, dass sich der CRPS bei einer normalverteilten Prognose zu (8.9) umformen lässt. Verwenden Sie hierzu die Formeln für den Erwartungswert der Halb-Normal-Verteilung aus Anhang A.3.

► Die Prognoseverteilung ist die Normalverteilung $N(\mu, \sigma^2)$, aus der Y_1 und Y_2 unabhängig voneinander entstammen. Daraus folgt

$$Y_1 - y_b \sim N(\mu - y_b, \sigma^2) \quad \text{und} \quad Y_1 - Y_2 \sim N(0, 2\sigma^2),$$

und die Beträge der obigen Zufallsvariablen sind halb-normalverteilt mit den gleichen Parametern. Der CRPS ist somit

$$\begin{aligned} CRPS(f(y), y_b) &= \mathbb{E} |Y_1 - y_b| - \frac{1}{2} \mathbb{E} |Y_1 - Y_2| \\ &= 2\sigma\varphi\left(\frac{\mu - y_b}{\sigma}\right) + (\mu - y_b) \left\{ 2\Phi\left(\frac{\mu - y_b}{\sigma}\right) - 1 \right\} - \frac{1}{2}(2\sqrt{2}\sigma\varphi(0) + 0) \\ &= 2\sigma\varphi\left(\frac{y_b - \mu}{\sigma}\right) + (\mu - y_b) \left\{ 2 \left[1 - \Phi\left(\frac{y_b - \mu}{\sigma}\right) \right] - 1 \right\} - \frac{\sqrt{2}\sigma}{\sqrt{2\pi}} \\ &= 2\sigma\varphi(\tilde{y}_b) + \left(\frac{\mu - y_b}{\sigma}\right) \sigma(1 - 2\Phi(\tilde{y}_b)) - \frac{\sigma}{\sqrt{\pi}} \\ &= \sigma \left(\tilde{y}_b(2\Phi(\tilde{y}_b) - 1) + 2\varphi(\tilde{y}_b) - \frac{1}{\sqrt{\pi}} \right). \end{aligned}$$

Literaturverzeichnis

Davison A. C. (2003) Statistical Models. Cambridge University Press, Cambridge.

Index

A

AB0-System 38
AIC 78
Anpassungstest 41, 44
arithmetisches Mittel 12
autoregressiver Prozess 4

B

Bayes-Faktor 83, 84
 Minimum- 85
Bayes-Schätzer 59
Bayesianisches hierarchisches Modell 71
Beta-Binomial-Verteilung 88
Beta-Verteilung 54, 65, 87
Bewertungsregel 92
BIC 78
Binomialverteilung 35, 87
Blutgruppen 38
Bootstrap-Prognoseverteilung 90
Brier-Score 91
Burn-In 74

C

Cauchy-Verteilung 2
 χ^2 -Statistik 44
Cramér-Rao-Schranke 19
CRPS 93

D

Delta-Regel 26, 36, 40
Dichtetransformationssatz 42, 63, 67

E

erwartungstreu 16, 19, 24
Exponentialfamilie 21
Exponentialverteilung 9, 18, 20, 22, 57, 78

F

Faktorisierungssatz 8, 9, 22
Faltungssatz 68
Fehler 1. Art 18
Fischstudie 79, 83
Fisher's z-Transformation 15
Fisher-Information 5, 30, 34

G

Gamma-Verteilung 19, 59, 61, 77
Gauß-Test 85
Geometrische Verteilung 7, 25
Gibbs Sampler 72

Gleichverteilung 7, 15, 51
Gradientenverfahren 36

H

Halb-Normalverteilung 93
HPD-Intervall 50, 66, 70, 74

I

Indikatorfunktion 8
Invers-Gamma-Verteilung 72

J

Jeffreys' Priori 58, 87

K

kanonischer Parameter 21, 22
Kohortenstudie 29
Konfidenzintervall 15
Konfidenzregion 40
Konturplot 31
Konvergenz 44
Korrelation 15, 35

L

Laplace-Approximation 61, 63
Lebensdauern 57, 77
Leibnizregel 60
Likelihood-Funktion
 erweiterte 88, 90
Likelihood-Konfidenzintervall 26
Likelihood-Quotienten-Statistik 44
Likelihoodfunktionen 1, 58
Lindley's Paradox 84

M

marginale Likelihood 49, 79, 83, 85, 90
Markov-Ketten Monte-Carlo 71
Matrixinversion 35
Meta-Analyse 72
Minimalsuffizienz 8
Mischverteilung 54
Modellwahl 83, 84
 Bayesianische 79
Monte-Carlo-Schätzung 66, 70, 74, 91
Multinomialverteilung 39
Murphy Zerlegung 91

N

Normal-Gamma-Verteilung 79, 91

Normal-Normal-Modell 52, 71, 83, 90
Normalverteilung 13, 22, 52, 83, 84, 89, 93
 bivariate 33
numerische Integration 63, 68, 70
numerische Optimierung 40

P

p-Wert 86
Pareto-Verteilung 51
Plug-in-Prognoseverteilung 89
Poisson-Verteilung 16, 21, 29, 61
posteriori-prädiktive Verteilung 87, 91
prädiktive Likelihood-Prognose 88, 90
priori-prädiktive Verteilung *siehe* marginale
 Likelihood
Priori-Verteilung
 informative 52, 54
 konjugierte 52, 54, 61, 66, 79, 87, 91
 nichtinformative 57, 87
Profil-Likelihood 30, 32, 36
Profil-Likelihood-Intervall 38, 67
Prognoseintervall 88, 89
Punktprognose 88, 89

R

Rückfangmethode 6, 11
Referenz-Priori 91
relative Chance 65
 logarithmierte 71
Reparametrisierung 30, 39
Risiko
 logarithmiertes relatives 35
 relatives 35

S

Satz von Bayes 79
Score-Gleichungen 31, 34
Score-Konfidenzintervall 16
Score-Test 20
Sensitivitätsanalyse 82
Shrinkage 76
Signifikanztest 16, 17, 21, 42
Simulation 57, 65, 68
Stichprobenvarianz 13, 90
Suffizienz 8, 23, 51

T

t-Verteilung 91
Taxis 23, 49
Taylor-Approximation 45

U

uneigentlich 58

V

Verlustfunktion 59
Verschiebungssatz 92
Vertrauensintervall 55, 57, 66, 74
vollständig bedingte Verteilung 71

W

Wald-Konfidenzintervall 19, 25, 32, 36, 67
Weibull-Verteilung 77

Z

zensierte Zufallsvariable 58, 77

Methoden der statistischen Inferenz

Likelihood und Bayes

Held, L.

2008, XII, 304 S. 60 Abb., Softcover

ISBN: 978-3-8274-1939-2