

Kapitel 3

Schätzen der Parameter

In dieser Arbeit wird die Maximum-Likelihood-Schätzung angewendet. Mit der Einführung der unbeobachteten Indikatorvariablen Z kann die vollständige Likelihoodfunktion bestimmt werden. Z gibt an, zu welcher Unterpopulation θ_j gehört ($z_{ij} = 1$ i -te Studie gehört zur j -ten Unterpopulation, sonst ist $z_{ij} = 0$). Die Likelihoodfunktion besitzt folgende Gestalt

$$L(y, P, z) = \prod_{i=1}^k \prod_{j=1}^m [f(y_i, \theta_j) q_j]^{z_{ij}}, \quad (3.1)$$

Die entsprechende Log-Likelihoodfunktion ist

$$l(y, P, z) = \sum_{i=1}^k \sum_{j=1}^m [z_{ij} \log(f(y_i, \theta_j)) + z_{ij} \log(q_j)]. \quad (3.2)$$

Der Vorteil dieser Form liegt darin, dass die Gewichte und das Effektmaß getrennt voneinander maximiert werden können. Die geschätzten Gewichte sind

$$\hat{q}_j = \frac{1}{k} \sum_{i=1}^k z_{ij} \quad (3.3)$$

Dieser Schätzer für die Gewichte gilt in allen drei Modellen. Der Effektschätzer wird nun für jedes Modell aus der Log-Likelihoodfunktion abgeleitet.

3.1 Parameterschätzung im Normalverteilungsmodell

Die vollständige Log-Likelihoodfunktion im ND-Ansatz ist:

$$l_{ND}(y, \Psi) = \sum_{i=1}^k \sum_{j=1}^m z_{ij} \log\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_i}\right) - z_{ij} \frac{(y_i - \log \theta_j)^2}{2\hat{\sigma}^2} + z_{ij} \log(q_j) \quad (3.4)$$

Zur Maximierung von (3.4) wird die Ableitung nach θ_j gebildet, mit

$$\frac{\partial l_{ND}(y, \Psi)}{\partial \theta_j} = \sum_{i=1}^k z_{ij} \frac{(y_i - \log \theta_j)}{\hat{\sigma}^2 \theta_j} = \frac{1}{\theta_j} \sum_{i=1}^k \frac{z_{ij} y_i}{\hat{\sigma}^2} - \frac{\log \theta_j}{\theta_j} \sum_{i=1}^k \frac{z_{ij}}{\hat{\sigma}^2}. \quad (3.5)$$

Durch Nullsetzen der Ableitung ergibt sich für den Schätzer des Relativen Risikos folgende Formel:

$$\hat{\theta}_{NDj} = e^{\frac{\sum_{i=1}^k \frac{z_{ij} y_i}{\hat{\sigma}^2}}{\sum_{i=1}^k \frac{z_{ij}}{\hat{\sigma}^2}}}$$

Damit ist der Schätzer für das logarithmische Relative Risiko ein gewichteter Mittelwert. Die Gewichte sind als die Inverse der Varianz definiert.

3.2 Parameterschätzung im Profil-Likelihood-Ansatz

Die vollständigen Log-Likelihoodfunktion ist im PL-Ansatz

$$l_{Poi}(x, \Psi) = \sum_{i=1}^k \sum_{j=1}^m z_{ij} (x_i^T \log(\theta_j) - x_i \log(n_i^T \theta_j + n_i^C)) + z_{ij} \ln(q_j) \quad (3.6)$$

Die Ableitung von (3.6) nimmt folgende Form an:

$$\frac{\partial l_{Poi}(x, \Psi)}{\partial \theta_j} = \sum_{i=1}^k \frac{z_{ij} x_i^T}{\theta_j} - \sum_{i=1}^k \frac{z_{ij} x_i n_i^T}{n_i^T \theta_j + n_i^C}$$

Leider ist für den Schätzer $\hat{\theta}_j$ keine geschlossene Lösungsform zugänglich, da die nullgesetzte Ableitung die Form besitzt:

$$\hat{\theta}_j = \frac{\sum_{i=1}^k z_{ij} x_i^T}{\sum_{i=1}^k \frac{z_{ij} x_i n_i^T}{n_i^T \theta_j + n_i^C}} \quad (3.7)$$

Möglich ist aber eine Fixpunktiteration, mit

$$\theta_{PMLEj} = \frac{\sum_{i=1}^k z_{ij} x_i^T n_i^C (n_i^T \theta_{PMLEj} + n_i^C)^{-1}}{\sum_{i=1}^k z_{ij} x_i^C n_i^T (n_i^T \theta_{PMLEj} + n_i^C)^{-1}} := \Gamma(\theta_{PMLEj}) \quad (3.8)$$

Es kann gezeigt werden, dass die Lösung von $\Gamma(\theta_{PMLEj})$ eindeutig und immer ein Maximum ist. Das besondere an dieser Iteration ist die Ähnlichkeit mit dem wohlbekannten Mantel-Haenszel-Schätzer. Wird die Iteration im Homogenitätsfall mit eins gestartet, dann ist das Ergebnis des ersten Iterationsschritts (3.9) sogar der exakte Mantel-Haenszel-Schätzer.

$$\Gamma(1) = \frac{\sum_{i=1}^k \frac{x_i^T n_i^C}{n_i^T + n_i^C}}{\sum_{i=1}^k \frac{x_i^C n_i^T}{n_i^T + n_i^C}} = \theta_{MH} \quad (3.9)$$

Unter anderem rechtfertigt diese wichtige Eigenschaft die Untersuchung dieses PL-Ansatzes.

3.3 Parameterschätzung im Multi-Level-Modell

In diesem Modell werden zwei Varianten betrachtet. Einmal unter der Annahme, dass die Beobachtungen binomialverteilt sind und zum anderen, dass die Beobachtungen Poissonverteilt sind. Unter der ersten Annahme ergibt sich folgende vollständige Likelihoodfunktion:

$$L_{MLBin}(P) = \prod_{i=1}^k \prod_{j=1}^m (q_j f_{MLBin}(\alpha_j, \beta_j)_i)^{z_{ij}}. \quad (3.10)$$

Der parameterabhängige Log-Likelihood von (3.10) ist

$$l_{MLBin}(P) = \sum_{i=1}^k \sum_{j=1}^m z_{ij} \log(q_k) + z_{ij} (x_i^C \alpha_j - n_i^C \log(1 + e^{\alpha_j}) + x_i^T (\alpha_j + \beta_j) - n_i^T \log(1 + e^{\alpha_j + \beta_j})) \quad (3.11)$$

Die Schätzer für α_j und β_j leiten sich aus den Nullgesetzten partiellen Ableitungen her, mit

$$\begin{aligned}\frac{\partial l_{MLBin}(P)}{\partial \beta_j} &= \sum_{i=1}^k z_{ij} \left(x_i^T - \frac{n_i^T e^{\alpha_j + \beta_j}}{1 + e^{\alpha_j + \beta_j}} \right) \stackrel{!}{=} 0 \\ e^{\alpha_j + \hat{\beta}_j} &= \frac{\sum_{i=1}^k z_{ij} x_i^T}{\sum_{i=1}^k z_{ij} (n_i^T - x_i^T)} \\ \hat{\beta}_j &= \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^T}{\sum_{i=1}^k z_{ij} (n_i^T - x_i^T)} \right) - \alpha_j\end{aligned}$$

$$\begin{aligned}\frac{\partial l_{MLBin}(P)}{\partial \alpha_j} &= \sum_{i=1}^k z_{ij} \left(x_i^C + x_i^T - \frac{n_i^C e^{\alpha_j}}{1 + e^{\alpha_j}} - \frac{n_i^T e^{\alpha_j + \beta_j}}{1 + e^{\alpha_j + \beta_j}} \right) \stackrel{!}{=} 0 \\ e^{\hat{\alpha}_j} &= \frac{\sum_{i=1}^k z_{ij} x_i^C}{\sum_{i=1}^k z_{ij} (n_i^C - x_i^C)} \\ \hat{\alpha}_j &= \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^C}{\sum_{i=1}^k z_{ij} (n_i^C - x_i^C)} \right)\end{aligned}\tag{3.12}$$

$$\hat{\beta}_j = \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^T \sum_{i=1}^k z_{ij} (n_i^C - x_i^C)}{\sum_{i=1}^k z_{ij} x_i^C \sum_{i=1}^k z_{ij} (n_i^T - x_i^T)} \right)\tag{3.13}$$

Nun kann anhand der Schätzer für α und β das Relative Risiko θ ermittelt werden, mit

$$\theta_j = \frac{\sum_{i=1}^k z_{ij} x_i^T \sum_{i=1}^k z_{ij} n_i^C}{\sum_{i=1}^k z_{ij} n_i^T \sum_{i=1}^k z_{ij} x_i^C}.$$

Unter Poissonverteilten Beobachtungen nimmt die vollständige Log-Likelihood folgende Form an:

$$l_{MLPoi}(P) = \prod_{i=1}^k \prod_{j=1}^m (q_j f_{MLPoi}(P)_i)^{z_{ij}}.$$

Analog ist die parameterabhängige Log-Likelihood

$$\begin{aligned}l_{MLPoi}(P) &= \sum_{i=1}^k \sum_{j=1}^m z_{ij} \log(q_k) \\ &\quad + z_{ij} \left(-e^{\alpha_j} n_i^C + x_i^C \alpha_j - e^{\alpha_j + \beta_j} n_i^T + x_i^T (\alpha_j + \beta_j) \right)\end{aligned}\tag{3.14}$$

Daraus lassen sich nun folgende Schätzer ableiten:

$$\alpha_j = \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^C}{\sum_{i=1}^k z_{ij} n_i^C} \right) \quad (3.15)$$

$$\beta_j = \log \left(\frac{\sum_{i=1}^k z_{ij} x_i^T \sum_{i=1}^k z_{ij} n_i^C}{\sum_{i=1}^k z_{ij} n_i^T \sum_{i=1}^k z_{ij} x_i^C} \right). \quad (3.16)$$

Eine Eigenschaft dieser zwei ML-Ansätze ist, dass sie unter Homogenität den gleichen Schätzer für das Relative Risiko liefern. Unter Heterogenität treten geringe Unterschiede auf. Dabei gilt die Aussage, je kleiner die Baselineparameter p_i^C und p_i^T in der multizentrischen Studie sind, desto ähnlicher sind sich die Schätzer für das Relative Risiko. Wäre die Zentrenzugehörigkeit bekannt und müsste nicht geschätzt werden, würden die zwei ML-Ansätze wieder die gleichen Ergebnisse liefern.

3.4 EM-Algorithmus bei einer Mischverteilung

Der EM (Expectation Maximization)-Algorithmus wurde in dem Artikel von Dempster [12] eingeführt und benannt. Neben dem Artikel beinhaltet das Buch von McLachlan and Krishnan [27] eine umfassende Darstellung der Entwicklung und Weiterentwicklung des EM-Algorithmus.

Die vollständige Log-Likelihoodfunktion (2.3) enthält die unbeobachtete und unbekannte Indikatorvariable Z . Im EM-Algorithmus besteht nun der E-Schritt darin, die Indikatorvariable durch den Erwartungswert $E(Z)$ zu ersetzen. Dieser ist $E(z_{ij}|y_i, q) = P_q(Z_{ij} = 1|Y_i = y_i)$. Die Anwendung des Bayesschen Theorems führt zu

$$\begin{aligned} P_q(Z_{ij} = 1|Y_i = y_i) &= \frac{P_q(Y_i = y_i|Z_{ij} = 1)P_q(Z_{ij} = 1)}{\sum_{j^*} P_q(Y_i = y_i|Z_{ij^*} = 1)P_q(Z_{ij^*} = 1)} \\ &= \frac{f(y_i, \theta_j)q_j}{\sum_{j^*} f(y_i, \theta_{j^*})q_{j^*}} \\ &= : e_{ij} \end{aligned} \quad (3.17)$$

Das Austauschen der Indikatorvariable Z durch e_{ij} führt zu der erwarteten vollständigen Log-Likelihoodfunktion

$$E(l(y, \Psi)) = \sum_{i=1}^k \sum_{j=1}^m [e_{ij} \log(f(y, \theta_j)) + e_{ij} \log(q_j)]. \quad (3.18)$$

Die Maximierung der Funktion (3.18) führt zum M-Schritt. Die Lösung für die Anteilswerte q_j ist in jedem der drei Modelle gleich, mit

$$q_j^{(new)} = \frac{1}{k} \sum_{i=1}^k e_{ij} = \frac{1}{k} \sum_{i=1}^k \frac{f(y_i, \theta_j) q_j}{\sum_{j^*} f(y_i, \theta_{j^*}) q_{j^*}} \quad (3.19)$$

Die Schätzung für das Relative Risiko ist je nach Modell folgend aufgeführt.

- ND-Modell:

$$\theta_{NDj} = \frac{\sum_{i=1}^k \frac{e_{ij} x_i}{\hat{\sigma}^2}}{\sum_{i=1}^k \frac{e_{ij}}{\hat{\sigma}^2}} \quad (3.20)$$

- PL-Ansatz:

$$\theta_{PMLEj} = \frac{\sum_{i=1}^k e_{ij} x_i^T n_i^C (n_i^T \theta_{PMLEj} + n_i^C)^{-1}}{\sum_{i=1}^k e_{ij} x_i^C n_i^T (n_i^T \theta_{PMLEj} + n_i^C)^{-1}} \quad (3.21)$$

- ML-Modell (Binomial-Likelihood und Poisson-Likelihood):

$$\theta_{MLj} = \frac{\sum_{i=1}^k e_{ij} x_i^T \sum_{i=1}^k e_{ij} n_i^C}{\sum_{i=1}^k e_{ij} x_i^C \sum_{i=1}^k e_{ij} n_i^T} \quad (3.22)$$

Die Struktur des Algorithmus ist in Kurzfassung gegeben als:

Schritt 0: q und θ (α und β für das ML-Modell) sind die Startwertvektoren

Schritt 1: Berechne den E-Schritt nach (3.17)

Schritt 2: Berechne den M-Schritt nach (3.19), welcher zu q^{new} führt und je nach Modell (3.20, 3.21 und 3.22), welcher zu θ^{new} führt

Schritt 3: Setze $q = q^{new}$ und $\theta = \theta^{new}$. Gehe zum Schritt 1¹.

3.5 Eigenschaften nichtparametrischer Mischverteilungen

Bei der Maximierung treten zwei Probleme auf. Die eine Schwierigkeit besteht in dem Vorkommen mehrerer lokaler Extremwerte. Der EM-Algorithmus konvergiert immer zu einem lokalen Maximum, das aber nicht gleichzeitig das globale Maximum darstellt. Dies

¹Ein entsprechendes Abbruchkriterium ist im Schritt 3 notwendig.

hängt von der Wahl der Startwerte ab. Das zweite Problem besteht darin, dass die Anzahl der Mischungskomponenten unbekannt ist. Ein wichtiges Hilfsmittel zur Lösung dieser Probleme ist die Gradientenfunktion mit dem allgemeinen Maximum-Likelihood Satz über Mischverteilungen. Mit der Gradientenfunktion kann überprüft werden, ob die Anzahl der Mischungskomponenten ausreichend ist und ob die Lösung ein globales Maximum ist. Im nächsten Abschnitt werden die Gradientenfunktion und danach der ML-Satz eingeführt.

3.5.1 Gradientenfunktion

Ω ist die Menge aller diskreten Wahrscheinlichkeitsverteilungen. P und Q sind zwei diskrete Wahrscheinlichkeitsfunktionen aus Ω . Die Richtungsableitung der Log-Likelihoodfunktion $l(P)$ ² nach P in Richtung Q und kann geschrieben werden als

$$\begin{aligned}\Phi(P, Q) &= \lim_{\alpha \rightarrow 0} \frac{l((1-\alpha)P + \alpha Q) - l(P)}{\alpha} := \left. \frac{\partial}{\partial \alpha} l((1-\alpha)P + \alpha Q) \right|_{\alpha=0} \\ \frac{\partial}{\partial \alpha} l((1-\alpha)P + \alpha Q) \Big|_{\alpha=0} &= \sum_{i=1}^k \frac{f(y_i, Q) - f(y_i, P)}{f(y_i, P)} \\ &= \sum_{i=1}^k \frac{f(y_i, Q)}{f(y_i, P)} - k\end{aligned}\tag{3.23}$$

Gewöhnlich wird die Richtungsableitung durch k geteilt. Die Gradientenfunktion $d(\theta, P)$ leitet sich aus der Richtungsableitung her, mit

$$\begin{aligned}d(\theta, P) &: = \frac{1}{k} \sum_{i=1}^k \frac{f(y_i, Q_\theta)}{f(y_i, P)} \\ \text{mit } Q_\theta &= \begin{pmatrix} \theta \\ 1 \end{pmatrix}\end{aligned}$$

Je nach Modell hat die Gradientenfunktion folgende Form:

- ND-Modell:

$$d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{e^{-\frac{(y_i - \log \theta)^2}{2\sigma_i^2}}}{\sum_{j=1}^m e^{-\frac{(y_i - \log \theta_j)^2}{2\sigma_i^2}} q_j}\tag{3.24}$$

² $l(P) = \sum_i^n \log f(y_i, P)$

- PL-Ansatz:

$$d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{\frac{\theta^{x_i^T}}{(n_i^T \theta + n_i^C)^{x_i}}}{\sum_{j=1}^m \frac{\theta_j^{x_i^T}}{(n_i^T \theta_j + n_i^C)^{x_i}} q_j} \quad (3.25)$$

- ML-Modell (Binomial-Likelihood):

$$d(\vartheta^*, P^*) = \frac{1}{k} \sum_{i=1}^k \frac{\frac{e^{x_i^T \alpha + x_i^C (\alpha + \beta)}}{(1+e^\alpha)^{n_i^T} (1+e^{\alpha+\beta})^{n_i^C}}}{\sum_{j=1}^m \frac{e^{x_i^T \alpha_j + x_i^C (\alpha_j + \beta_j)}}{(1+e^{\alpha_j})^{n_i^T} (1+e^{\alpha_j + \beta_j})^{n_i^C}} q_j} \quad (3.26)$$

- ML-Modell (Poisson-Likelihood):

$$d(\vartheta^*, P^*) = \frac{1}{k} \sum_{i=1}^k \frac{e^{-e^\alpha n_i^C} (e^\alpha)^{x_i^C} e^{-e^{\alpha+\beta} n_i^T} (e^{\alpha+\beta})^{x_i^T}}{g(P^*)} \quad (3.27)$$

$$g(P^*) = \sum_{j=1}^m e^{-e^{\alpha_j} n_i^C} (e^{\alpha_j})^{x_i^C} e^{-e^{\alpha_j + \beta_j} n_i^T} (e^{\alpha_j + \beta_j})^{x_i^T} q_j$$

$$\text{mit } \vartheta^* = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \text{ und } P^* = \begin{pmatrix} \alpha_1 & \dots & \alpha_m \\ \beta_1 & \dots & \beta_m \\ q_1 & \dots & q_m \end{pmatrix}$$

3.5.2 Der allgemeine Maximum-Likelihood Satz von Mischverteilungen

In der Theorie der nichtparametrischen Verteilungsmodelle wird üblicherweise folgende wichtige Definition verwendet.

Definition 1 Jedes \hat{P} , welches die Log-Likelihoodfunktion $l(P)$ maximiert, wird Nicht-Parametrischer-Maximum-Likelihood-Schätzer (NPMLE) genannt.

Der folgende allgemeine Maximum-Likelihood Satz von Mischverteilungen (Böhning [8]) ist ein wichtiges Konstrukt in der nichtparametrischen Mischverteilungsstatistik.

Lemma 1 P und Q sind diskrete Verteilungsfunktionen. Es gilt

$$\Phi(P, Q) \geq l(Q) - l(P)$$

Beweis. $l(P)$ ist eine konkave Funktion (Böhning [8], Seite 36-37). ■

Satz 2 \hat{P} ist NPMLE dann und nur dann, wenn gilt

$$d(\theta, \hat{P}) = \frac{1}{k} \sum_{i=1}^k \frac{f(y_i, Q_\theta)}{f(y_i, P)} \leq 1$$

für alle θ im Parameterraum.

Beweis.

1. „ \implies “

Wenn \hat{P} NPMLE ist, dann ist

$$\begin{aligned} \frac{l((1-\alpha)\hat{P} + \alpha Q) - l(\hat{P})}{\alpha} &\leq 0 \text{ für alle } \alpha \in [0, 1] \text{ und alle } Q \in \Omega \\ &\implies \Phi(\hat{P}, Q) \leq 0 \\ * &\implies \Phi(\hat{P}, Q_\theta) \leq 0 \end{aligned}$$

*) Q_θ ist in der Menge Q enthalten.

2. „ \impliedby “

Wenn $\Phi(P, Q_\theta) \leq 0$ für alle θ aus dem Parameterraum

$$\rightarrow 0 \geq \sup_{\theta} \Phi(P, Q_\theta)^{**} \geq \sup_Q \Phi(P, Q) \geq \sup_Q l(Q) - l(P) \quad (3.28)$$

da $\sup_Q l(Q) - l(P) \geq 0$, folgt aus der Ungleichung 3.28, dass P nur der NPMLE sein kann und

$$0 = \sup_{\theta} \Phi(P, Q_\theta) = \sup_Q \Phi(P, Q).$$

**)

$$\Phi(P, Q) = \sum_j q_j \Phi(P, Q_{\theta_j}) \leq \sum_j q_j \sup_{\theta} \Phi(P, Q_\theta) = \sup_{\theta} \Phi(P, Q_\theta)$$

diese Ungleichung gilt für jedes Q aus Ω , speziell auch für das Q welches $\Phi(P, Q)$ maximiert.

Damit endet der Beweis. ■

Mit Hilfe dieses Satzes kann nun ein Algorithmus entwickelt werden, der als Abbruchkriterium prüft, ob der NPMLE erreicht ist. Zwei Algorithmen mit globaler Konvergenz sind bei Böhning [8, 9] enthalten, zum einen der VDM (Vertex Direction Method)-

Algorithmus und zum anderen die VEM-Methode (Vertex Exchange Method). Der Vorteil dieser Algorithmen liegt in der globalen Konvergenz. Nachteilig ist die künstliche Aufblähung der Komponentenanzahl. Die Anzahl der Komponenten kann aber über eine Clusterbildung wieder reduziert werden. Ist der NPML erreicht, so gibt dieser auch gleichzeitig die Anzahl der Mischungskomponenten an. Damit sind theoretisch die zwei Probleme (Anzahl der Mischungskomponenten und globales Maximum) gelöst.

3.6 EM-Algorithmus mit Gradientenfunktions-Update

3.6.1 EM-Algorithmus mit Gradientenfunktions-Update bei bekannter Anzahl der Komponenten

In der Publikation von Böhning [9] wird ein Algorithmus mit Gradientenfunktions-Update (EMGFU) mit bekannter Anzahl der Mischungskomponenten vorgeschlagen. Die Idee besteht darin, den EM-Algorithmus mit der Gradientenfunktion zu kombinieren. Im Folgenden wird der EMGFU in Kurzfassung aufgeführt:

Schritt 0: Wähle die Anzahl der Komponenten m und den Startwertvektor

$$P = \begin{pmatrix} \theta_1 & \dots & \theta_m \\ q_1 & \dots & q_m \end{pmatrix}.$$

Schritt 1: Berechne mit dem EM-Algorithmus P_{EM}

Schritt 2: Bestimme θ_{\max} , indem $d(\theta, P_{EM})$ in θ maximiert wird.

Schritt 3: Bestimme θ_{\min} , so dass $l(P_{EM} + P_{EM}(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}])$ das größte Element in der Menge

$$\{l(P_{EM} + P_{EM}(\theta_j)[Q_{\theta_{\max}} - Q_{\theta_j}]) \mid j = 1, \dots, m\}$$

ist.

Schritt 4: $P = P_{EM} + P_{EM}(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$. Wenn $l(P) > l(P_{EM})$, dann gehe zu Schritt 1, sonst Stopp.

Bemerkung 1 Die inhaltliche Bedeutung von $P + P(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$ ist, dass in dem Vektor P das Element θ_{\min} durch θ_{\max} ersetzt wird.

Da die Anzahl der Komponenten im Allgemeinen unbekannt ist, kann der EMGFU nur im Spezialfall verwendet werden. Es gibt aber die Möglichkeit den EMGFU für den allgemeinen Fall geeignet zu modifizieren.

3.6.2 EM-Algorithmus mit Gradientenfunktions-Update bei unbekannter Anzahl der Komponenten (gEMGFU)

Die Idee ist, dass zuerst der homogene Fall betrachtet wird. Das heißt es gibt nur eine Komponente. Der Startwert für θ kann beliebig gewählt werden. Danach wird mit dem EM-Algorithmus P_{EM}^1 berechnet. Wie im Schritt 2 beim EMGFU wird θ_{\max} bestimmt. Ist $d(\theta_{\max}, P_{EM}^1) \leq 1$ kann der Algorithmus abgebrochen werden, da der NPMLE erreicht ist. Schritt 3 wird vom EMGFU unverändert übernommen. Der Schritt 4 wird ebenso beibehalten und erweitert. Zuerst wird der Austauschschritt durchgeführt, was zu einem P^{l*} führt. Ist $l(P_{EM}^{l*}) > l(P_{EM}^l)$, dann war der Austauschschritt erfolgreich und es wird zu Schritt 2 gesprungen. Wenn $l(P_{EM}^{l*}) \leq l(P_{EM}^l)$, dann konnte keine Verbesserung durch den Austausch erreicht werden. Im Schritt 5 wird nun eine Komponente hinzugefügt. Das erfolgt nach dem Prinzip des VEM-Algorithmus. Dazu folgende Illustration: $P^3 = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ q_1 & q_2 & q_3 \end{pmatrix}$, $\theta_{\max} = \theta_4$ und $\theta_{\min} = \theta_2$. θ_{\max} wird nun zu P^3 hinzugefügt. Somit entsteht $P^4 = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ q_1 & (1-\alpha)q_2 & q_3 & \alpha q_2 \end{pmatrix}$. Da sich die Gewichte zu eins aufsummieren müssen, wird das Gewicht von θ_{\min} in $(1-\alpha)q_2$ und in αq_2 geteilt. Die neu hinzugefügte Komponente erhält das Gewicht αq_2 und θ_{\min} erhält das Gewicht $(1-\alpha)q_2$. Für die Schätzung von α gibt es zwei Vorschläge. Eine Möglichkeit besteht darin, $l(P^l + P^l \alpha(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}])$ in α zu maximieren. Alternativ kann die Approximationsformel von Böhning [9] verwendet werden. Sie lautet

$$\alpha_{\max} = \frac{\sum_{i=1}^k g(y_i, \theta_{\max}, P)}{\sum_{i=1}^k g(y_i, \theta_{\max}, P)^2} \quad (3.29)$$

mit $g(y, \theta, P) = \frac{f(y, \theta) - f(y, P)}{f(y, P)}$.

Zusammenfassend ist nachstehender Algorithmus entstanden:

Schritt 0: Wähle den Startwertvektor $P^1 = \begin{pmatrix} \theta \\ 1 \end{pmatrix}$

Schritt 1: Berechne mit dem EM-Algorithmus P_{EM}^l

Schritt 2: Bestimme θ_{\max} , indem $d(\theta, P_{EM}^l)$ in θ maximiert wird.

Schritt 3: Bestimme θ_{\min} , so dass $l(P_{EM}^l + P_{EM}^l(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}])$ das größte Element in der Menge

$$\{l(P_{EM}^l + P_{EM}^l(\theta_j)[Q_{\theta_{\max}} - Q_{\theta_j}]) \mid j = 1, \dots, m\}$$

ist.

Schritt 4: Bestimme $P^{l*} = P_{EM}^l + P_{EM}^l(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$. Berechne mit EM-Algorithmus P_{EM}^{l*} . Ist $l(P_{EM}^{l*}) > l(P_{EM}^l)$, dann $P_{EM}^l = P_{EM}^{l*}$ und gehe zu Schritt 2

Schritt 5: Bestimme α nach (3.29), und bestimme $P^{l+1} = P_{EM}^l + \alpha P_{EM}^l(\theta_{\min})[Q_{\theta_{\max}} - Q_{\theta_{\min}}]$. Gehe zu Schritt 1

Für diesen Maximierungsalgorithmus kann kein Konvergenzbeweis angegeben werden. Da sich der gEMGFU-Algorithmus aus dem EM-Algorithmus und der global-konvergenten *Vertex Exchange* Methode zusammensetzt, liegt die Vermutung einer globalen Konvergenz des gEMGFU sehr nahe. In den später folgenden Simulationsexperimenten und bei den hier vorgestellten multizentrischen Studien konvergierte der Algorithmus in jedem Fall zum nichtparametrischen Maximum-Likelihood-Schätzer. Ein wichtiger Vorteil des gEMGFU gegenüber dem EM-Algorithmus ist, dass keine expliziten Startwerte notwendig sind. Einen weiteren Vorteil stellen die Zwischenergebnisse dar. Es zeigte sich, dass in den hier dargestellten Zwischenergebnissen der multizentrischen Studien die Likelihoodfunktion in derselben Dimension nicht weiter verbessert werden konnte, auch wenn der nichtparametrische Maximum-Likelihood-Schätzer noch nicht erreicht ist. Das heißt, die Likelihoodfunktion kann nur verbessert werden, wenn eine weitere Komponente hinzugefügt wird. Ein Nachteil besteht in der langsamen Konvergenzgeschwindigkeit des EM-Algorithmus, der im gEMGFU in jedem Durchlauf neu angewendet wird.

3.7 gEMGFU in FORTRAN

Als Programmiersprache zur Schätzung der Parameter wird FORTRAN benutzt. FORTRAN benötigt für numerische Rechenoperationen sehr wenig Zeit. Für besonders aufwendige Berechnungen, wie in unserem Fall ist diese Programmiersprache besonders gut geeignet. Doch die veraltete MS-DOS Umgebung des Programms wirkt sich nachteilig aus, da PC-Nutzer heutzutage mit Maus und Menüsteuerung arbeiten. In der FORTRAN-Entwicklungsumgebung *FORTRAN PowerStation 4.0* sind Programmroutinen integriert, wie zum Beispiel *QuickWin*, mit deren Hilfe leicht ein menügesteuertes Programm erstellt

werden kann. Dennoch sind die Möglichkeiten für die Entwicklung eines benutzerfreundlichen Programms sehr begrenzt. Für komplexere Programme wird der Programmierer schnell an die Grenzen von FORTRAN stoßen und auf die Programmiersprache C oder C++ zurückgreifen müssen.

In diesem Abschnitt wird die Umsetzung von bestimmten Teilen des Algorithmus in FORTRAN beschrieben. Besonderes hervorgehoben werden zum einen die numerische Stabilität der einzelnen Berechnungsschritte und zum anderen die Maximierung der Gradientenfunktion.

3.7.1 Numerische Stabilität

An vielen Stellen kommt es im gEMGFU vor, dass die Dichte oder der Log-Likelihood der Mischverteilung berechnet werden muss. Als Beispiel dient der E-Schritt (3.17) im EM-Algorithmus. Beim PL-Ansatz hat der E-Schritt folgende Form

$$e_{ij} = \frac{\frac{\theta_j^{x_i^T}}{(n_i^C + n_i^T \theta_j)^{x_i}} q_j}{\sum_{j^*=1}^m \frac{\theta_{j^*}^{x_i^T}}{(n_i^C + n_i^T \theta_{j^*})^{x_i}} q_{j^*}}. \quad (3.30)$$

Es kann vorkommen, dass Nenner und Zähler von (3.30) sehr klein sind, wie zum Beispiel beim 4. Zentrum in der Beta-Blocker-Studie (Tabelle 1.1 mit $x_4^T = 102$; $x_4^C = 127$; $n_4^T = 1533$; $n_4^C = 1520$). Mit zwei Mischungskomponenten ($\theta_1 = 0.7685$; $\theta_2 = 1.128979$; $q_1 = 0.905401$; $q_2 = 0.094591$) ergeben sich für den Zähler und den Nenner bei e_{41} folgende Werte:

$$\begin{aligned} \frac{\theta_1^{x_4^T}}{n_4^C + n_4^T \theta_1} q_1 &= 3.8 \times 10^{-798} \\ \frac{\theta_1^{x_4^T}}{n_4^C + n_4^T \theta_1} q_1 + \frac{\theta_2^{x_4^T}}{n_4^C + n_4^T \theta_2} q_2 &= 3.813278 \times 10^{-798} \end{aligned}$$

FORTRAN rundet diese kleinen Zahlen ab, so dass am Ende das Programm versucht 0/0 zu berechnen. Das führt zu einem Fehler. Durch folgende Umformungen der Formel (3.30)

kann dieses Problem vermieden werden.

$$\begin{aligned}
 e_{ij}^{PL} &= \frac{\frac{\theta_j^{x_i^T}}{(n_i^C + n_i^T \theta_j)^{x_i}} q_j}{\sum_{j^*=1}^m \frac{\theta_{j^*}^{x_i^T}}{(n_i^C + n_i^T \theta_{j^*})^{x_i}} q_{j^*}} \\
 &= \left(\sum_{j^*=1}^m \exp \left(x_i \log \left(\frac{n_i^C + n_i^T \theta_j}{n_i^C + n_i^T \theta_{j^*}} \right) + x_i^T \log \frac{\theta_{j^*}}{\theta_j} + \log \frac{q_{j^*}}{q_j} \right) \right)^{-1}.
 \end{aligned}$$

Für den ND-Ansatz ergibt sich die Formel

$$\begin{aligned}
 e_{ij}^{ND} &= \frac{\frac{1}{\sqrt{2\pi\hat{\sigma}_i}} e^{-\frac{(y_i - \log \theta_j)^2}{2\hat{\sigma}_i^2}} q_j}{\sum_{j^*=1}^m \frac{1}{\sqrt{2\pi\hat{\sigma}_i}} e^{-\frac{(y_i - \log \theta_{j^*})^2}{2\hat{\sigma}_i^2}} q_{j^*}} \\
 &= \left(\sum_{j^*=1}^m \exp \frac{(y_i - \log \theta_j)^2 - (y_i - \log \theta_{j^*})^2}{2\hat{\sigma}_i^2} + \log \frac{q_{j^*}}{q_j} \right)^{-1},
 \end{aligned}$$

und für das ML-Modell ist

$$\begin{aligned}
 e_{ij}^{MLB} &= \frac{\frac{e^{x_i^C \alpha_j}}{(1+e^{\alpha_j})^{n_i^C}} \times \frac{e^{x_i^T (\alpha_j + \beta_j)}}{(1+e^{\alpha_j + \beta_j})^{n_i^T}} q_j}{\sum_{j^*=1}^m \frac{e^{x_i^C \alpha_{j^*}}}{(1+e^{\alpha_{j^*}})^{n_i^C}} \times \frac{e^{x_i^T (\alpha_{j^*} + \beta_{j^*})}}{(1+e^{\alpha_{j^*} + \beta_{j^*}})^{n_i^T}} q_{j^*}} \\
 &= \left(\sum_{j^*=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1}, \text{ mit} \\
 t_1 &= x_i^C (\alpha_{j^*} - \alpha_j) + x_i^T (\alpha_{j^*} + \beta_{j^*} - \alpha_j - \beta_j) \\
 t_2 &= n_i^C \log \left(\frac{1 + e^{\alpha_j}}{1 + e^{\alpha_{j^*}}} \right) + n_i^T \log \left(\frac{1 + e^{\alpha_j + \beta_j}}{1 + e^{\alpha_{j^*} + \beta_{j^*}}} \right) \\
 t_3 &= \log \left(\frac{q_{j^*}}{q_j} \right)
 \end{aligned}$$

$$\begin{aligned}
e_{ij}^{MLPoi} &= \frac{e^{-e^{\alpha_j} n_i^C} (e^{\alpha_j})^{x_i^C} \times e^{-e^{\alpha_j + \beta_j} n_i^T} (e^{\alpha_j + \beta_j})^{x_i^T} \times q_j}{\sum_{j^*=1}^m e^{-e^{\alpha_{j^*}} n_i^C} (e^{\alpha_{j^*}})^{x_i^C} \times e^{-e^{\alpha_{j^*} + \beta_{j^*}} n_i^T} (e^{\alpha_{j^*} + \beta_{j^*}})^{x_i^T} \times q_{j^*}} \\
&= \left(\sum_{j^*=1}^m \exp(t_1 + t_2 + t_3) \right)^{-1} \\
t_1 &= n_i^C (e^{\alpha_j} - e^{\alpha_{j^*}}) + n_i^T (e^{\alpha_j + \beta_j} - e^{\alpha_{j^*} + \beta_{j^*}}) \\
t_2 &= x_i^C (\alpha_{j^*} - \alpha_j) + x_i^T (\alpha_{j^*} + \beta_{j^*} - \alpha_j - \beta_j) \\
t_3 &= \log \left(\frac{q_{j^*}}{q_j} \right)
\end{aligned}$$

Wie im ND-Ansatz könnte es auch im PL- und ML-Modell passieren, dass bei bestimmten Konstellationen durch Null geteilt wird. Beim PL-Modell tritt dieser Fall ein, wenn im Homogenitätsfall im Kontrollarm in jedem Zentrum keine Fälle beobachtet worden sind. Dieser Fall ist zwar rein hypothetisch, kann aber in einer Simulationsstudie durchaus vorkommen. Im Heterogenitätsfall kann, selbst wenn nur in einem Zentrum im Kontrollarm kein einziger Fall aufgetreten ist, ebenfalls durch Null geteilt werden. Dies tritt ein (siehe 3.8), wenn alle $e_{ij} = 0$ sind und das $e_{nj} = 1$ (n ist das Zentrum wo $x_n^C = 0$). Das heißt, das n -te Zentrum wird als einziges zu einer Mischungskomponente zugewiesen. Auch wenn im Behandlungsarm in einigen Zentren keine Fälle auftraten, könnten diese einer Mischungskomponente zugewiesen werden, welche ein Relatives Risiko von Null schätzt. Ein Relatives Risiko von Null ist aber nicht interpretierbar, so dass auch dieser Fall ausgeschlossen werden muss. Um diese Fehlerquellen zu beseitigen wird in jedem Modell zu jeder Zelle $\frac{1}{2}$ addiert, wenn $x_i = 0$. Zusätzlich zu den beschriebenen Fehlerquellen, kann es im ML-Modell vorkommen, dass bei $n_i = x_i$ durch Null geteilt wird (siehe 3.12 und 3.13). Daher wird im ML-Modell, wenn $x_i = 0$ oder $n_i = x_i$ zu jeder Zelle $\frac{1}{2}$ addiert. Im ND-Ansatz (nur bei binomialverteilten Beobachtungen) tritt bei $x_i^T = n_i^T$ und $x_i^C = n_i^C$ der Fall ein, dass die geschätzte Varianz im i -ten Zentrum gleich Null ist. Da die Gewichte des Schätzers für das Relative Risiko aus dem Inversen der Varianz gebildet werden, darf diese Konstellation nicht auftreten. Somit wird in dem Fall wieder $\frac{1}{2}$ zu jeder Zelle addiert. Diese Verfahrensweise birgt in sich den Nachteil, dass besonders wenn sehr häufig die Konstante addiert wird, ein Bias hinzugefügt wird. Als Beispiel dazu dienen die beiden Abbildungen 3.1 und 3.2. Die Bilder stammen aus der Simulation 5 unter Homogenität. An verschiedenen Punkten im Intervall von 0.01 bis 0.99 ist der Bias vom Relativen Risiko des ML- und PL-Ansatzes berechnet worden. Als Grundlage dieser Bootstrapsimulation wird die CALGB-Studie verwendet. Die Funktionsweise der Simulation wird in einem späteren Kapitel noch genau beschrieben. Deutlich ist im Bild 3.2 zu

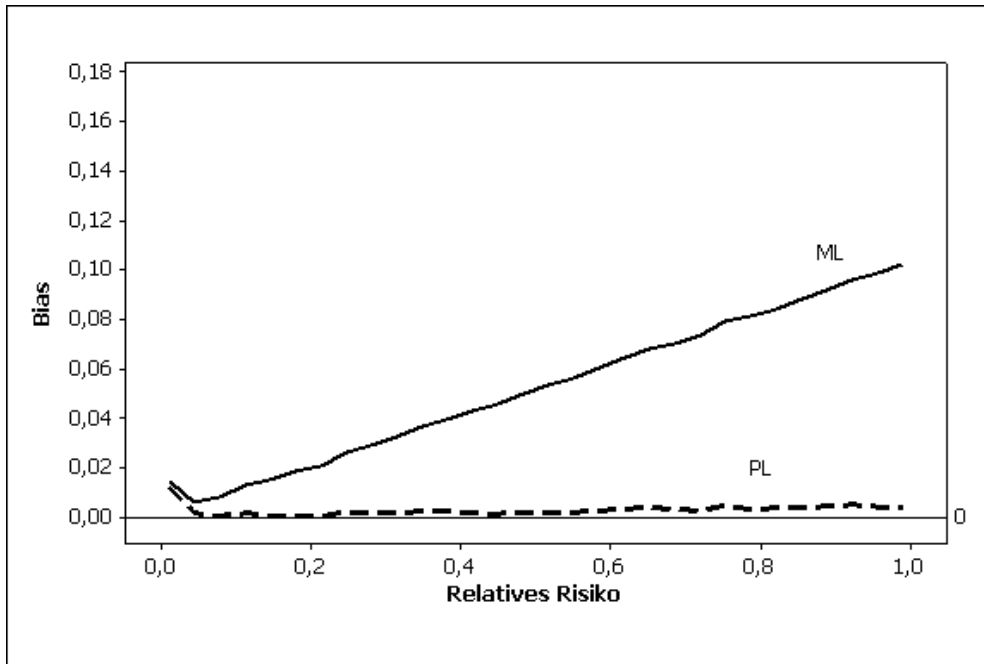


Abbildung 3.1: CALGB-Studie (Bias vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;0.99]$ berechnet, die generierten x^T , x^C -Werte sind binomialverteilt, additive Konstante ist 0.00000001)

erkennen, wie sich der Bias hauptsächlich bei kleineren Relativen Risiko-Werten erhöht hat. Es stellt sich hier die Frage, warum wird die Konstante 0.5 und nicht 0.00000001 gewählt. Die Antwort liegt in der Modellierung der unbeobachteten Heterogenität. Bei einer Konstanten mit einem Wert von 0.00000001, bekämen die Zentren mit $x_i = 0$ einen sehr starken Einfluss und würden einer separaten Komponente zugeordnet. Diese Komponente hätte dann extreme Werte. Entweder sie strebt gegen Null oder gegen unendlich. Besonders würde es beim PL-Modell in der Komponentenanalyse 5.2 auffallen, weil dort dann die erste Komponente stark unterschätzt und die letzte Komponente stark überschätzt werden würde. Um dies zu umgehen wird die Konstante mit dem Wert 0.5 gewählt.

Wird im ND-Modell zu einem Zentrum eine Konstante addiert, dann ändert sich auch die Formel der Varianzschätzung. Die vollständige Varianzformel ist binomialverteilten Beobachtungen bei $x_i^T = 0$ und $x_i^C > 0$ (entsprechend bei $x_i^T > 0$ und $x_i^C = 0$ oder bei $x_i^T = 0$ und $x_i^C = 0$)

$$\text{Var}(y_i) = \frac{0.5n_i^T}{(0.5 + n_i^T)} + \frac{1}{x_i^C} - \frac{1}{n_i^C} \quad (3.31)$$

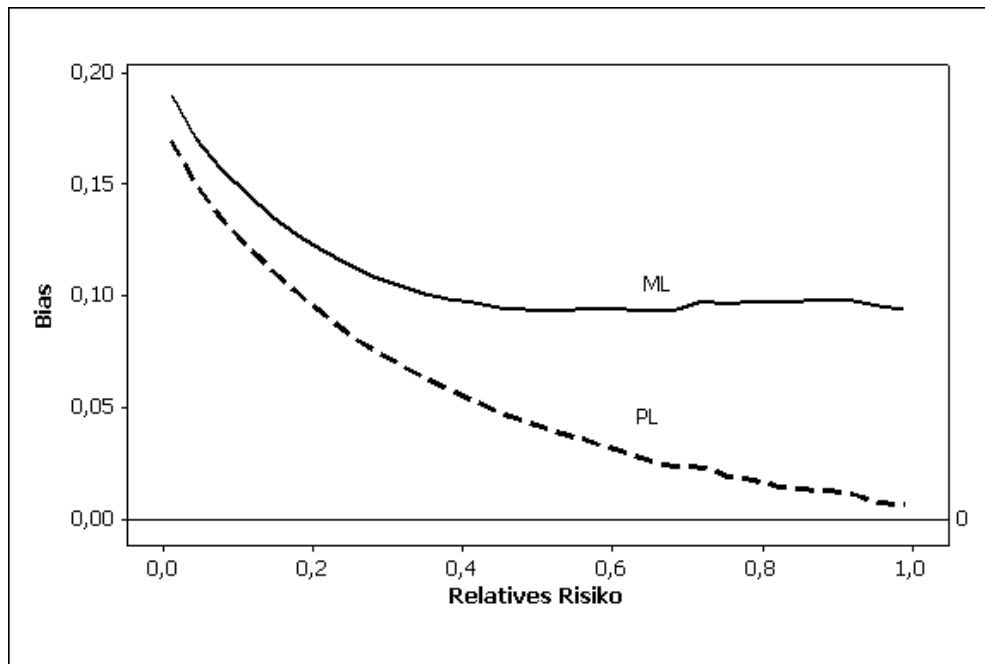


Abbildung 3.2: CALGB-Studie (Bias vom Relativen Risiko-Schätzer wurde an 30 Stellen gleichverteilt im Intervall $[0.01;0.99]$ berechnet, die generierten x^T , x^C - Werte sind binomialverteilt, additive Konstante ist 0.5)

und bei $x_i^T = n_i^T$ und $x_i^C = n_i^C$

$$\text{Var}(y_i) = \frac{0.5n_i^T(0.5 + n_i^T)}{(0.5 + n_i^T(1 + n_i^T))^2} + \frac{0.5n_i^C(0.5 + n_i^C)}{(0.5 + n_i^C(1 + n_i^C))^2} \quad (3.32)$$

Im Falle von Poissonverteilten Beobachtungen gilt bei $x_i^T = 0$ und $x_i^C > 0$ (entsprechend bei $x_i^T > 0$ und $x_i^C = 0$ oder bei $x_i^T = 0$ und $x_i^C = 0$) die veränderte Varianzformel

$$\text{Var}(z_i^*) = \frac{0.5n_i^T(1 + n_i^T)}{(0.5 + n_i^T)^2} + \frac{1}{x_i^C}. \quad (3.33)$$

3.7.2 Die Maximierung der Gradientenfunktion

Ein wichtiger Punkt im gEMGFU ist die Berechnung des globalen Maximums der Gradientenfunktion. Da die Gradientenfunktion mehrere lokale Maxima besitzt, ist es nicht leicht, den globalen Extremwert zu finden. Durch die Daten der Studie ist das Intervall gegeben, in der die Lösung enthalten sein muss. Im ND- und im PL-Ansatz ist das Intervall $[a, b]$ definiert durch die untere Grenze $a = \min_i(\frac{x_i^T n_i^C}{n_i^T x_i^C})$ und durch die obere Grenze $b = \max_i(\frac{x_i^T n_i^C}{n_i^T x_i^C})$ (bei $x_i^C > 0$ für $i = 1, \dots, k$). Die Idee besteht nun darin, das Intervall zu

splitten. Danach wird in jedem Teilintervall $[a_i, b_i]$ geprüft, ob das Vorzeichen der Ableitung der Gradientenfunktion $\frac{\partial}{\partial \theta} d(\theta, P)$ wechselt. Die Ableitung der Gradientenfunktion ist:

- Für den ND-Ansatz

$$\frac{\partial}{\partial \theta} d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{e^{-\frac{(y_i - \log \theta)^2}{2\sigma_i^2}} \frac{(y_i - \log \theta)}{\sigma_i^2}}{\sum_{j=1}^m e^{-\frac{(y_i - \log \theta_j)^2}{2\sigma_i^2}} q_j}$$

- Für den PL-Ansatz

$$\frac{\partial}{\partial \theta} d(\theta, P) = \frac{1}{k} \sum_{i=1}^k \frac{(-n_i^T \theta x_i^C + n_i^C x_i^T) \theta^{-1+x_i^T} (n_i^T \theta + n_i^C)^{-1-x_i}}{\sum_{j=1}^m \frac{\theta_j^{x_i^T}}{(n_i^T \theta_j + n_i^C)^{x_i}} q_j}$$

Ein lokales Maximum ist dann in dem Intervallstück enthalten, wenn gilt, dass

$$\frac{\partial}{\partial \theta} d(a_i, P) \geq 0 \text{ und } \frac{\partial}{\partial \theta} d(b_i, P) < 0.$$

Nachdem alle interessanten Intervallstücke gefunden worden sind, wird in diesen das jeweilige lokale Maximum mit Hilfe der Intervallhalbierungsmethode berechnet. Damit sind alle lokalen Maxima bestimmt. Das globale Maximum θ_{\max} ist dann der größte Wert von $d(\theta_{LMi}, P)$, mit θ_{LM} als Vektor aller lokalen Maxima. Beim ML-Modell kann dieses Verfahren nicht angewendet werden, da die Gradientenfunktion in dem Fall von 2 Variablen abhängt. Das Intervall, in dem die Lösung enthalten sein muss, ist gegeben durch $[\alpha_{\min}, \alpha_{\max}] \times [\beta_{\min}, \beta_{\max}] \in \mathbb{R}^2$, mit

$$\begin{aligned} \alpha_{\min} &= \min_i \left(\frac{x_i^C}{n_i^C} \right), \quad \alpha_{\max} = \max_i \left(\frac{x_i^C}{n_i^C} \right) \\ \beta_{\min} &= \min_i \left(\frac{x_i^T n_i^C}{x_i^C n_i^T} \right), \quad \beta_{\max} = \max_i \left(\frac{x_i^T n_i^C}{x_i^C n_i^T} \right) \\ \text{mit } i &= 1, \dots, k. \end{aligned}$$

Über ein Rastergitter auf dem Intervall aus 100×100 Punkten wird der größte Wert $\theta_{\max}^{ML} = \begin{pmatrix} \alpha_{\max} \\ \beta_{\max} \end{pmatrix}$ der Gradientenfunktion gesucht. θ_{\max}^{ML} stellt damit eine grobe Näherung des globalen Maximums dar. Um den globalen Extremwert möglichst gut zu schätzen, wird θ_{\max}^{ML} mit Hilfe des Newtonverfahrens verbessert. Das Newtonverfahren ist definiert

als

$$\begin{aligned}
 x^{n+1} &= x^n - f'(x^n)^{-1} f(x^n) \\
 \text{mit } f(x) &= \begin{pmatrix} u(x_1, x_2) \\ v(x_1, x_2) \end{pmatrix} \text{ und} \\
 f'(x) &= \begin{pmatrix} u_{x_1}(x_1, x_2) & u_{x_2}(x_1, x_2) \\ v_{x_1}(x_1, x_2) & v_{x_2}(x_1, x_2) \end{pmatrix}
 \end{aligned}$$

Mit dem Newtonverfahren wird die Nullstelle von der Ableitung der Gradientenfunktion $\frac{\theta}{\partial\theta^{ML}}d(\theta^{ML}, P) = \begin{pmatrix} \frac{\theta}{\partial\alpha}d(\alpha, \beta, P) \\ \frac{\theta}{\partial\beta}d(\alpha, \beta, P) \end{pmatrix}$ berechnet. Damit ist $f(\theta^{ML}) = \frac{\theta}{\partial\theta^{ML}}d(\theta^{ML}, P)$ und $f'(\theta^{ML}) = \begin{pmatrix} \frac{\theta}{\partial\alpha\alpha}d(\alpha, \beta, P) & \frac{\theta}{\partial\alpha\beta}d(\alpha, \beta, P) \\ \frac{\theta}{\partial\beta\alpha}d(\alpha, \beta, P) & \frac{\theta}{\partial\beta\beta}d(\alpha, \beta, P) \end{pmatrix}$. Das Inverse einer 2×2 Matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ kann mittels folgender Formel bestimmt werden

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Daraus folgt, dass

$$\begin{aligned}
 f'(\theta^{ML})^{-1} &= \frac{1}{\frac{\partial}{\partial\alpha\alpha}d(\alpha, \beta, P)\frac{\partial}{\partial\beta\beta}d(\alpha, \beta, P) - \frac{\partial}{\partial\alpha\beta}d(\alpha, \beta, P)^2} \\
 &\quad \begin{pmatrix} \frac{\partial}{\partial\beta\beta}d(\alpha, \beta, P) & -\frac{\partial}{\partial\alpha\beta}d(\alpha, \beta, P) \\ -\frac{\partial}{\partial\beta\alpha}d(\alpha, \beta, P) & \frac{\partial}{\partial\alpha\alpha}d(\alpha, \beta, P) \end{pmatrix} \quad (3.34)
 \end{aligned}$$

Zur vollständigen Bestimmung von $f(\theta^{ML})$ und $f'(\theta^{ML})$ sind alle einzelnen Ableitungen im Anhang angegeben.

