Anhang A

A.1 Bootstrap-Stichproben

Tabelle A.1: Bootstrap-Stichprobe nach den Vorgaben der Studie über koronarer Herzkrankheit, bei der eine Komponente einen besonders großen Schätzwert für das Relative Risiko im Profil-Likelihoodansatz hat.

Zentrum	Behandlungsarm		Kontrollarm	
i	x^T	n^T	x^C	n^C
1	47	380	53	350
2	47	1250	38	640
3	48	690	47	500
4	5	90	3	30
5	2	30	0	30
6	68	1240	71	1180
7	58	1930	53	890
8	26	340	15	350
9	70	1930	115	1920
10	88	1240	39	410
11	30	1140	46	1140
12	10	210	10	220
13	11	210	15	230
14	1	90	6	170
15	1833	38620	727	19420
16	225	1350	184	1330
17	19	890	32	860
18	47	1970	50	2060
19	1	150	8	150
20	35	2150	62	2100
21	33	1010	31	1120
22	3	100	0	50
23	13	340	1	340
24	152	4410	246	4390
25	69	3850	73	3740
26	1	190	1	190
27	4	1510	15	1560
28	88	13850	56	13800
29	28	10140	49	10040
30	5	5910	3	1500
31	191	27630	180	27590
32	2	100	2	100
33	2	20	3	30

Tabelle A.2: Bootstrap-Stichprobe nach den Vorgaben der Studie über koronarer Herzkrankheit, bei der eine Komponente einen besonders großen Schätzwert für das Relative Risiko im Normalverteilungsansatz hat.

Zentrum	Behandlungsarm		Kontrollarm	
i	x^T	n^T	x^C	n^C
1	53	380	50	350
2	98	1250	46	640
3	17	690	41	500
4	7	90	2	30
5	4	30	0	30
6	56	1240	84	1180
7	103	1930	52	890
8	22	340	16	350
9	122	1930	118	1920
10	106	1240	23	410
11	49	1140	54	1140
12	10	210	10	220
13	19	210	31	230
14	1	90	1	170
15	929	38620	786	19420
16	219	1350	183	1330
17	19	890	30	860
18	29	1970	49	2060
19	3	150	6	150
20	41	2150	56	2100
21	14	1010	33	1120
22	2	100	0	50
23	11	340	3	340
24	227	4410	262	4390
25	42	3850	58	3740
26	2	190	0	190
27	12	1510	13	1560
28	45	13850	68	13800
29	38	10140	40	10040
30	14	5910	0	1500
31	169	27630	175	27590
32	0	100	0	100
33	4	20	0	30

A.2 Ergebnisse einzelner Bootstrap-Stichproben

Das PL-Modell (Tabelle A.3) schätzt mit den Daten der ersten Bootstrap-Stichprobe (Tabelle A.1) in der drei- und vierkomponentigen Lösung eine Ausreißerkomponente mit einem Relativen Risiko von 12.36. Zu dieser Komponente gehört aber nur das 23. Zentrum. Nach dem NPMLE-Kriterium findet dieses Modell 4 Mischungskomponenten, wo außer der dritten Komponente alle zur vorgegebenen Verteilung $P = \begin{pmatrix} 0.6 & 0.9 & 1.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$ passen. Das ND-Modell (Tabelle A.4) ist bei dieser Studie robuster und schätzt mit drei Mischungskomponenten nach dem NPMLE-Kriterium fast die richtige a priori Verteilung. Bei den Daten der zweiten Bootstrap-Stichprobe (Tabelle A.2) schätzt das ND-Modell (Tabelle A.7) eine Ausreißerkomponente mit einem Relativen Risiko von 6.8. Zu dieser Mischungskomponente gehören die Zentren 5, 30 und 33. In diesen drei Zentren wurden keine Fälle im Kontrollarm beobachtet. Bei der dreikomponentige Lösung des PL-Modells (Tabelle A.6) liegen alle drei geschätzten Komponenten im Bereich der vorgegebenen Verteilung P. Das ML-Modell schätzt schon nach dem BIC-Kriterium deutlich mehr Mischungskomponenten. Die einzelnen Komponenten der dreikomponentige Lösung des ML-Modells unterscheiden aber sich deutlich von der vorgegebenen mischenden Verteilung P.

Tabelle A.3: Ergebnisse der Bootstrap-Studie (Tabelle A.1) anhand des Profil-Likelihoodansatzes

Profil-Likelihood				
Komp.	1.			
θ	1.006647			
q	1.000000			
H	33			
Log-L.= -237.804746				
$\max GF = 49000.721604$				
BIC= -479.106000				
Komp.	1.	2.		
θ	.650141	1.225349		
q	.614341	.385659		
H	21	12		
Log-L.= -195.404056				
$\max GF = 10.936823$				
BIC= -401.297634				
Komp.	1.	2.	3.	
θ	.652048	1.220140	12.362823	
q	.624652	.344215	.031133	
H	24	8	1	
Log-L.= -193.643222				
$\max GF = 1.198333$				
BIC= -404.768983				
Komp.	1.	2.	3.	4.
θ	.629751	1.252899	12.361214	.987451
q	.547294	.229244	.031328	.192134
H	23	5	1	4
Log-L.= -192.968707				
$\max GF = 1.000000$				
BIC= -410.412967				

Tabelle A.4: Ergebnisse der Bootstrap-Studie (Tabelle A.1) anhand des Normalverteilungsansatzes

Normalverteilungsansatz			
Komp.	1.		
θ	1.004053		
q	1.000000		
H	33		
Log-L.= -64.274635			
$\max GF = 42886.601857$			
BIC= -132.045778			
Komp.	1.	2.	
θ	.656301	1.220598	
q	.610219	.389781	
H	21	12	
Log-L.= -24.467422			
$\max GF = 1.176821$			
BIC= -59.424367			
Komp.	1.	2.	3.
θ	.634090	1.253461	.989152
q	.530845	.261916	.207239
H	21	8	4
Log-L.= -23.814023			
$\max GF = 1.000001$			
BIC= -65.110584			

Tabelle A.5: Ergebnisse der Bootstrap-Studie (Tabelle A.1) anhand des Multi-Level Modells (Poissonlikelihood).

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	1.149666					
q	1.000000					
H	33					
Log-L.= -2660.480822						
$\max GF = 7.8734 \times 10^{214}$						
BIC= -5324.458151						
Komp.	1.	2.				
θ	.999998	.951760				
q	.773741	.226259				
H	26	7				
Log-L.= -723.441120						
$\max GF = 8.9043 \times 10^{86}$						
BIC= -1457.371762						
Komp.	1.	2.	3.			
θ	1.037340	.963059	1.114616			
q	.746514	.186288	.067199			
H	25	6	2			
Log-L.= -465.489269						
$\max GF = 7.1362 \times 10^{23}$						
BIC= -948.461076						
Komp.	1.	2.	3.	4.	5.	6.
θ	1.267216	1.052163	1.114600	.741079	.749692	.602014
q	.057214	.043968	.065983	.233596	.082769	.195226
H	1	1	2	10	3	6
Komp.	7.	8.	9.	10.	11.	12.
θ	.507871	12.037292	.422116	1.561378	1.016247	.912813
q	.068918	.035325	.031997	.038805	.099373	.046825
H	3	1	1	1	3	1
Log-L.= -245.530926						
$\max GF = 3.058382$						
BIC= -571.481526						

Tabelle A.6: Ergebnisse der Bootstrap-Studie (Tabelle A.2) anhand des Profil-Likelihoodansatzes

Profil-Likelihood				
Komp.	1.			
θ	.778942			
q	1.000000			
H	33			
Log-L.= -222.468445				
$\max GF = 169.329886$				
BIC= -448.433398				
Komp.	1.	2.		
θ	.598834	.994824		
q	.313784	.686216		
H	10	23		
Log-L.= -194.491887				
$\max GF = 2.253121$				
BIC= -399.473296				
Komp.	1.	2.	3.	
θ	.597249	.913066	1.137529	
q	.289145	.441153	.269703	
H	9	19	5	
Log-L.= -193.814083				
$\max GF = 2.368546$				
BIC= -405.110704				
Komp.	1.	2.	3.	4.
θ	.603533	.907664	1.138340	.313373
q	.244918	.458890	.269616	.026576
H	6	21	5	1
Log-L.= -193.366396				
$\max GF = .996415$				
BIC= -411.208344				

Tabelle A.7: Ergebnisse der Bootstrap-Studie (Tabelle A.2) anhand des Normalverteilungsansatzes

Normalverteilungsansatz					
Komp.	1.				
θ	.779153				
q	1.000000				
H	33				
Log-L.= −57.758167					
$\max GF$ = 181862.428541					
BIC= −119.012842					
Komp.	1.	2.			
θ	.996719	.599900			
q	.702030	.297970			
H	24	9			
Log-L.= −30.199229					
$\max GF$ = 4.769987					
BIC= −70.887980					
Komp.	1.	2.	3.		
θ	.982224	.599806	6.809290		
q	.597298	.300970	.101731		
H	20	10	3		
Log-L.= −27.335496					
$\max GF$ = 1.904453					
BIC= −72.153531					
Komp.	1.	2.	3.	4.	
θ	.979974	.604614	6.808151	.313071	
q	.606758	.270186	.101733	.021324	
H	21	8	3	1	
Log-L.= −27.082418					
$\max GF$ = 1.112698					
BIC= −78.640389					
Komp.	1.	2.	3.	4.	5.
θ	.906376	.603167	6.837582	.313229	1.115996
q	.429120	.241165	.100247	.022516	.206953
H	20	6	4	1	2
Log-L.= −26.560818					
$\max GF$ = 1.000001					
BIC= −84.590204					

Tabelle A.8: Ergebnisse der B.-Studie (Tabelle A.2) anhand des Multi-Level Modells (Poissonlikelihood).

Multi-Level-Modell (Poi)						
Komp.	1.					
θ	.854552					
q	1.000000					
H	33					
Log-L.= -2573.837735						
$\max GF = 4.5361 \times 10^{189}$						
BIC= -5151.171977						
Komp.	1.	2.				
θ	.843507	.719740				
q	.283402	.716598				
H	9	24				
Log-L.= -854.429437						
$\max GF = 6.4026 \times 10^{100}$						
BIC= -1719.348396						
Komp.	1.	2.	3.			
θ	.843034	.621540	.979244			
q	.281634	.312595	.405771			
H	9	10	14			
Log-L.= -458.612880						
$\max GF = 1.3084 \times 10^{36}$						
BIC= -934.708297						
Komp.	1.	2.	3.	4.	5.	6.
θ	.970072	.595940	.892374	1.090112	.679112	6.457682
q	.107649	.062824	.247168	.099095	.170527	.033352
H	5	2	9	3	6	1
Komp.	7.	8.	9.	10.		
θ	1.286447	.770454	.727698	.302771		
q	.109002	.090072	.048715	.031596		
H	3	2	1	1		
Log-L.= -246.968061						
$\max GF = 11.020904$						
BIC= -560.369765						

A.3 Ableitungen der Gradientenfunktion im ML-Modell

Beim Binomial-Likelihood sind die Ableitungen der Gradientenfunktion

$$\frac{\partial}{\partial\alpha}d(\alpha,\beta,P) = \frac{1}{k}\sum_{i=1}^{k}\frac{\frac{e^{x_i^C\alpha+x_i^T(\alpha+\beta)}(x_i^T+x_i^C+e^{\alpha+\beta}(x_i-n_i^T)+e^{\alpha}(x_i-n_i^C)+e^{2\alpha+\beta}(x_i-n_i))}{(1+e^{\alpha})^{n_i^C+1}\left(1+e^{\alpha+\beta}\right)^{n_i^T+1}}}{\sum_{j=1}^{m}\frac{e^{x_i^C\alpha_j+x_i^T(\alpha_j+\beta_j)}}{\left(1+e^{\alpha_j}\right)^{n_i^C}\left(1+e^{\alpha_j+\beta_j}\right)^{n_i^T}}q_j}$$

$$\frac{\partial}{\partial\beta}d(\alpha,\beta,P) = \frac{1}{k}\sum_{i=1}^{k}\frac{\frac{e^{x_i^C\alpha+x_i^T(\alpha+\beta)}(x_i^T+e^{\alpha+\beta}(x_i^T-n_i^T))}{(1+e^{\alpha})^{n_i^C}\left(1+e^{\alpha+\beta}\right)^{n_i^T+1}}}{\sum_{j=1}^{m}\frac{e^{x_i^C\alpha_j+x_i^T(\alpha_j+\beta_j)}}{\left(1+e^{\alpha_j}\right)^{n_i^C}\left(1+e^{\alpha_j+\beta_j}\right)^{n_i^T}}q_j}$$

$$\frac{\partial}{\partial\alpha^2}d(\alpha,\beta,P) = \frac{1}{k}\sum_{i=1}^{k}\frac{t(t_1 * t_2 + t_4(t_3+t_5+t_6))}{\sum_{j=1}^{m}\frac{e^{x_i^C\alpha_j+x_i^T(\alpha_j+\beta_j)}}{\left(1+e^{\alpha_j}\right)^{n_i^C}\left(1+e^{\alpha_j+\beta_j}\right)^{n_i^T}}q_j}$$

$$\begin{aligned}
t &= e^{x_i^C\alpha+x_i^T(\alpha+\beta)}\,(1+e^{\alpha})^{-n_i^C-2}\left(1+e^{\alpha+\beta}\right)^{-n_i^T-2}\\
t_1 &= e^{\alpha}(1+e^{\alpha})(1+e^{\alpha+\beta})\\
t_2 &= -n_i^C + x_i + e^{\beta}(-n_i^T+x_i) + 2e^{\alpha+\beta}(-n_i+x_i)\\
t_3 &= e^{\alpha+\beta}(1+e^{\alpha})(-1-n_i^T)\\
t_4 &= x_i + e^{\alpha+\beta}(-n_i^T+x_i) + e^{\alpha}(-n_i^C+x_i) + e^{2\alpha+\beta}(-n_i+x_i)\\
t_5 &= e^{\alpha}(1+e^{\alpha+\beta})(-1-n_i^C)\\
t_6 &= (1+e^{\alpha})(1+e^{\alpha+\beta})x_i
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\beta^2}d(\alpha,\beta,P) &= \frac{1}{k}\sum_{i=1}^{k}\frac{t * t_1}{\sum_{j=1}^{m}\frac{e^{x_i^C\alpha_j+x_i^T(\alpha_j+\beta_j)}}{\left(1+e^{\alpha_j}\right)^{n_i^C}\left(1+e^{\alpha_j+\beta_j}\right)^{n_i^T}}q_j}\\
t &= e^{x_i^C\alpha+x_i^T(\alpha+\beta)}(1+e^{\alpha})^{-n_i^C}(1+e^{\alpha+\beta})^{-2-n_i^T}\\
t_1 &= e^{2(\alpha+\beta)}(n_i^T-x_i^T)^2 + \left(x_i^T\right)^2 - e^{\alpha+\beta}(n_i^T + 2n_i^T x_i^T - 2\left(x_i^T\right)^2)
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial\alpha\beta}d(\alpha,\beta,P) &= \frac{1}{k}\sum_{i=1}^{k}\frac{t(t_1+t_2+t_3)}{\sum_{j=1}^{m}\frac{e^{x_i^C\alpha_j+x_i^T(\alpha_j+\beta_j)}}{\left(1+e^{\alpha_j}\right)^{n_i^C}\left(1+e^{\alpha_j+\beta_j}\right)^{n_i^T}}q_j}\\
t &= e^{x_i^C\alpha+x_i^T(\alpha+\beta)}(1+e^{\alpha})^{-1-n_i^C}(1+e^{\alpha+\beta})^{-2-n_i^T}\\
t_1 &= x_i^T(x_i+e^{\alpha}(-n_i^C+x_i))\\
t_2 &= e^{2(\alpha+\beta)}(-n_i^T+x_i^T)(-n_i^T+x_i+e^{\alpha}(-n_i+x_i))\\
t_3 &= e^{\alpha+\beta}(e^{\alpha}n_i^T(-1+n_i^C-2x_i^T-x_i^C)-n_i^T(1+2x_i^T+x_i^C)\\
&\quad +2x_i^T(x_i+e^{\alpha}(-n_i^C+x_i)))\\
\frac{\partial}{\partial\beta\alpha}d(\alpha,\beta,P) &= \frac{\partial}{\partial\alpha\beta}d(\alpha,\beta,P)
\end{aligned}
$$

Beim Poisson-Likelihood sind die Ableitungen der Gradientenfunktion

$$
\begin{aligned}
\frac{\partial d(\theta,P)}{\partial\alpha} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{m}\exp\left(t_1+t_2+t_3\right)\right)^{-1}\\
t_1 &= n_i^C\left(e^{\alpha}-e^{\alpha_j}\right)+n_i^T\left(e^{\alpha+\beta}-e^{\alpha_j+\beta_j}\right)\\
t_2 &= x_i^C\left(\alpha_j-\alpha\right)+x_i^T\left(\alpha_j+\beta_j-\alpha-\beta\right)+\log q_j\\
t_3 &= -\log(-e^{\alpha}n_i^C-e^{\alpha+\beta}n_i^T+x_i^C+x_i^T)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial d(\theta,P)}{\partial\beta} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{m}\exp\left(t_1+t_2+t_3\right)\right)^{-1}\\
t_1 &= n_i^C\left(e^{\alpha}-e^{\alpha_j}\right)+n_i^T\left(e^{\alpha+\beta}-e^{\alpha_j+\beta_j}\right)\\
t_2 &= x_i^C\left(\alpha_j-\alpha\right)+x_i^T\left(\alpha_j+\beta_j-\alpha-\beta\right)+\log q_j\\
t_3 &= -\log(-e^{\alpha+\beta}n_i^T+x_i^T)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial d(\theta,P)}{\partial\alpha^2} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{m}\exp\left(t_1+t_2+t_3\right)\right)^{-1}\\
t_1 &= n_i^C\left(e^{\alpha}-e^{\alpha_j}\right)+n_i^T\left(e^{\alpha+\beta}-e^{\alpha_j+\beta_j}\right)\\
t_2 &= x_i^C\left(\alpha_j-\alpha\right)+x_i^T\left(\alpha_j+\beta_j-\alpha-\beta\right)+\log q_j\\
t_3 &= -\log(e^{2\alpha}\left(n_i^C\right)^2+2e^{2\alpha+\beta}n_i^Cn_i^T+e^{2(\alpha+\beta)}\left(n_i^T\right)^2\\
&\quad +\left(x_i^C+x_i^T\right)^2-e^{\alpha}n_i^C(1+2x_i^C+2x_i^T)-e^{\alpha+\beta}n_i^T(1+2x_i^C+2x_i^T))
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial d(\theta,P)}{\partial\beta^2} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{m}\exp\left(t_1+t_2+t_3\right)\right)^{-1}\\
t_1 &= n_i^C\left(e^{\alpha}-e^{\alpha_j}\right)+n_i^T\left(e^{\alpha+\beta}-e^{\alpha_j+\beta_j}\right)\\
t_2 &= x_i^C\left(\alpha_j-\alpha\right)+x_i^T\left(\alpha_j+\beta_j-\alpha-\beta\right)+\log q_j\\
t_3 &= -\log(e^{2(\alpha+\beta)}\left(n_i^T\right)^2+\left(x_i^T\right)^2-e^{\alpha+\beta}n_i^T(1+2x_i^T))
\end{aligned}$$

$$\begin{aligned}
\frac{\partial d(\theta,P)}{\partial\alpha\beta} &= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{m}\exp\left(t_1+t_2+t_3\right)\right)^{-1}\\
t_1 &= n_i^C\left(e^{\alpha}-e^{\alpha_j}\right)+n_i^T\left(e^{\alpha+\beta}-e^{\alpha_j+\beta_j}\right)\\
t_2 &= x_i^C\left(\alpha_j-\alpha\right)+x_i^T\left(\alpha_j+\beta_j-\alpha-\beta\right)+\log q_j\\
t_3 &= -\log(e^{2\alpha+\beta}n_i^C n_i^T+e^{2(\alpha+\beta)}\left(n_i^T\right)^2+\left(x_i^T\right)\left(x_i^C+x_i^T\right)\\
&\quad -e^{\alpha}n_i^C x_i^T-e^{\alpha+\beta}n_i^T(1+x_i^C+2x_i^T))
\end{aligned}$$