Mischverteilungen in der Perspektive biometrischer Anwendungen

Dankmar Böhning
AG *Statistische Methoden* derAbteilung für Epidemiologie
Institut für Soziale Medizin des FB Humanmedizin
UKBF, Freie Universität Berlin

Präsentation in Verbindung mit der Besetzung einer Professur für „Statistik mit Schwerpunkten Biometrie und epidemiologische Methoden“ (C4) an der Fakultät für Mathematik der Universität Bremen

Inhalt

I. Einführung und Motivation

II. Theoretische Hintergründe
- II.1 *Charakterisierungen des NPMLE*
- II.2 *Algorithmen*

III. Anwendungen
- III.1 *Meta-Analyse*
- III.2 *Disease Mapping*
- III.3 *Binäre Regression*

I. Einführung und Motivation

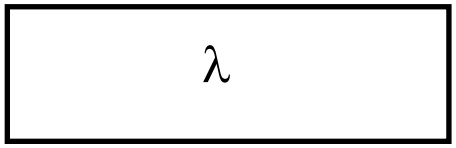
natürliche Genesis der *Mischverteilungen:*

unbeobachtete Heterogenität

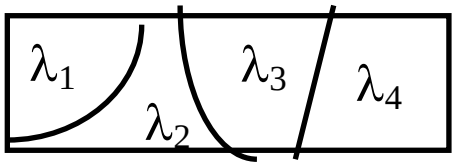
einparametrige Dichte $f(x,\lambda)$

λ Populationsparameter, x im Stichprobenraum

Homogeneität



Heterogenität



Dichte in Subpopulation j: $f(x,\lambda_j)$
latente Variable Z beschreibt Populationszugehörigk.

gemeinsame Dichte f(x,z)
mit $f(x,z) = f(x|z)f(z) = f(x,\lambda_z)p_z$

Marginal Dichte:

Mischungskern

$$f(x,P) = \sum_{z=1}^{k} f(x|z)f(z) = \sum_{j=1}^{k} f(x,\lambda_j)p_j = \int f(x,\lambda)\, P(d\lambda)$$

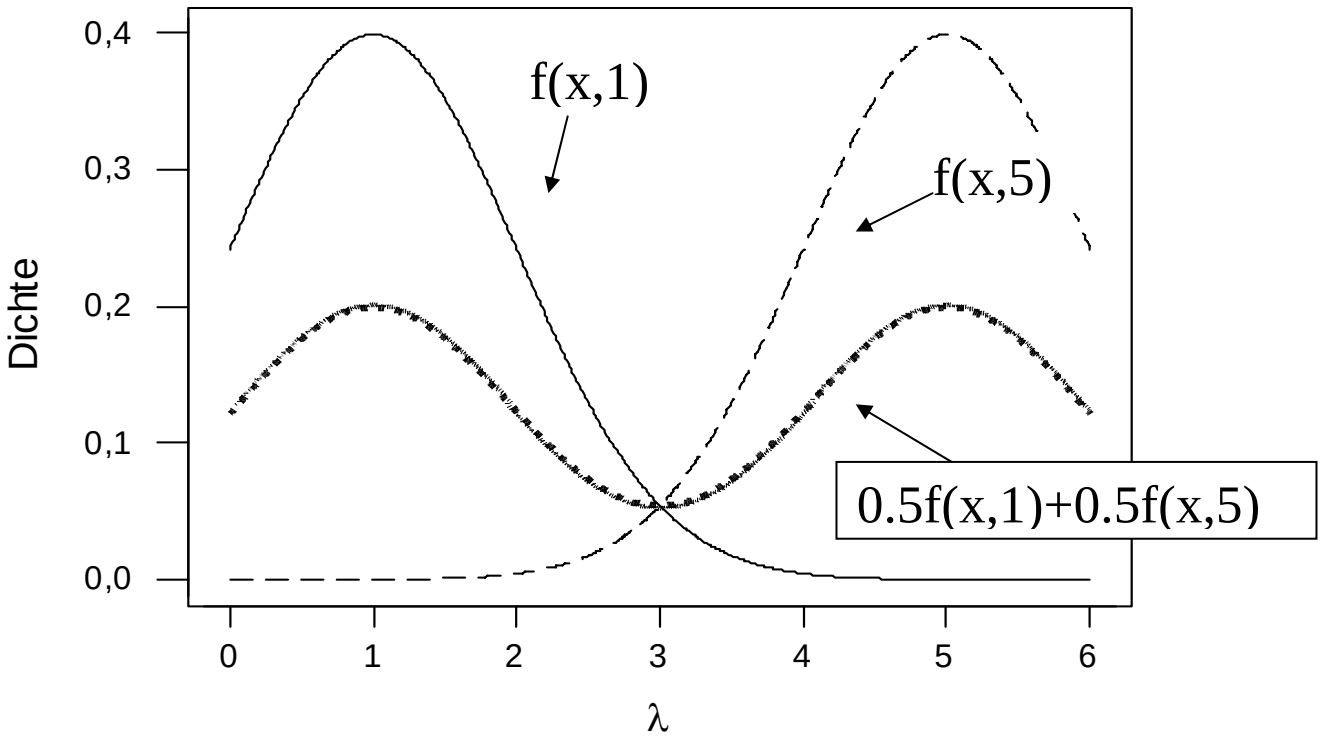
Mischende Verteilung

Mischverteilung: $\sum_{j=1}^{k} f(x,\lambda_j)p_j$

z.B. $X|\lambda$ normalverteilt,

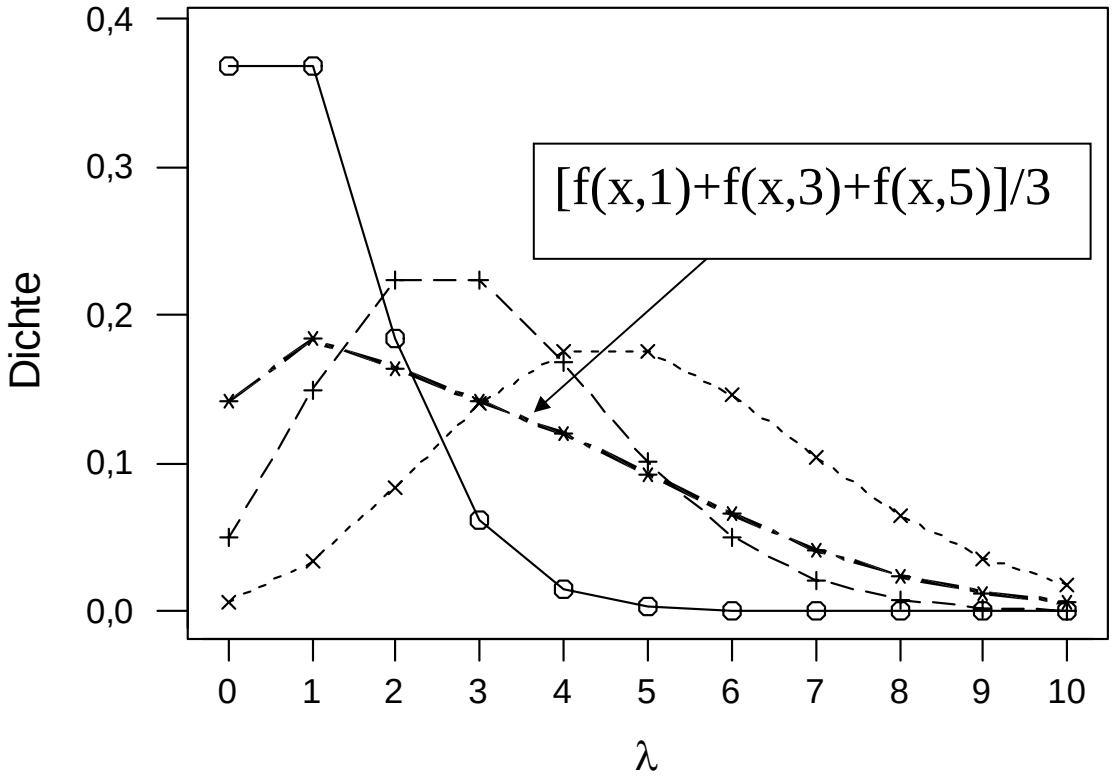
d.h. $f(x,\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{- ½ (x-\lambda)^2/\sigma^2\}$

und $k = 2$, $\sigma=1$, $\lambda_1=1$, $\lambda_2=5$



Mischverteilung: $\sum_{j=1}^{k} f(x,\lambda_j)p_j$

z.B. $X|\lambda$ Poisson verteilt,
d.h. $f(x,\lambda) = e^{-\lambda}\lambda^x/x!$
und $k = 3$, $\lambda_1=1$, $\lambda_2=3$, $\lambda_3=5$



C.A.MAN bietet Mischverteilungsanalyse:

$P = \begin{pmatrix} \lambda_1 \ldots \lambda_k \\ p_1 \ldots p_k \end{pmatrix}$ *Schätzung der mischenden Verteilung* !

log-likelihood: $l(P) = \sum_{i=1}^{n} \log f(x_i,P)$

$$= \sum_{i=1}^{n} \log \left[\sum_{j=1}^{k} f(x_i,\lambda_j)\, p_j \right]$$

C.A.MAN stellt bereit: MLE für P: $\hat{P} = \begin{pmatrix} \hat{\lambda}_1 \ldots \hat{\lambda}_k \\ \hat{p}_1 \ldots \hat{p}_k \end{pmatrix}$

Viele Anwendungen (direkte) sind vom folgendem Typ:

Homogene Population	*Heterogene Population*
natürliche Dichte	*Mischverteilung*

C.A.MAN bietet *Mischverteilungsanalyse* für Mischungen von

- Normalverteilungen
- Expontentialverteilungen
- Poissonverteilungen
- Binomialverteilungen
- Geometrische Verteilungen
- u.a. mehr

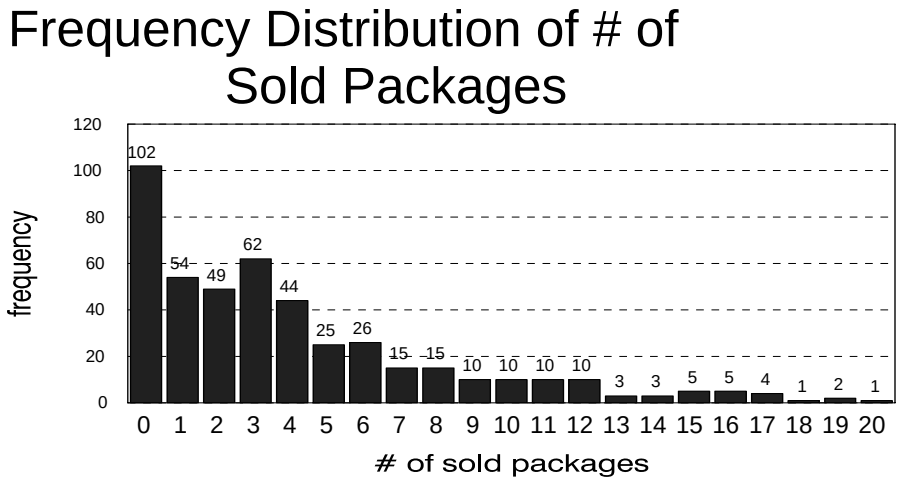
Ein einführendes Beispiel:
Advanced Methods of Marketing Research
by Richard P. Bagozzi (1995):

„data from a recent branded „hardcandy“ new product and concept test“
von Interesse: X = „Anzahl von verkauften Einheiten des neuen Produktes in den verschiedenen Verkaufsstellen innerhalb der letzten 7 Tage“

Tabelle 1 : Verteilunge der verkauften Einheiten

# von Einheiten	0	1	2	3	4	5	6	7	8	9
Häufigkeit	102	54	49	62	44	25	26	15	15	10

# von Einheiten	10	11	12	13	14	15	16	17	18	19	20
Häufigkeit	10	10	10	3	3	5	5	4	1	2	1



Typische Annahme für Anzahldaten $X|\lambda$:
Poisson-Verteilung $f(x,\lambda) = Po(x,\lambda) = e^{-\lambda}\lambda^{x}/x!$

Reference: Böhning, Schlattmann, Lindsay: „Recent Developments in C.A.MAN“, *Biometrics* (1992,1998)

Böhning, D.: Computer Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping, and Others. *Chapman & Hall* 1998.

Mischverteilungsmodelle bilden den natürlichen Rahmen der *Empirischen Bayes Verfahren*

Maritz and Lwin (1989)

Dichte $f(x,\lambda)$ mit *priori*- Verteilung $P = \begin{pmatrix} \lambda_1 \ldots \lambda_k \\ p_1 \ldots p_k \end{pmatrix}$ auf dem Parameterraum von λ

Bayessche Theorem:

posteriori-Dichte $p(\lambda_j/x) = f(x,\lambda_j)\, p_j \,/ \sum_{j=1}^{k} f(x,\lambda_j)p_j$

Problem in klassischen Bayesianischen Analysen:
P unbekannt!

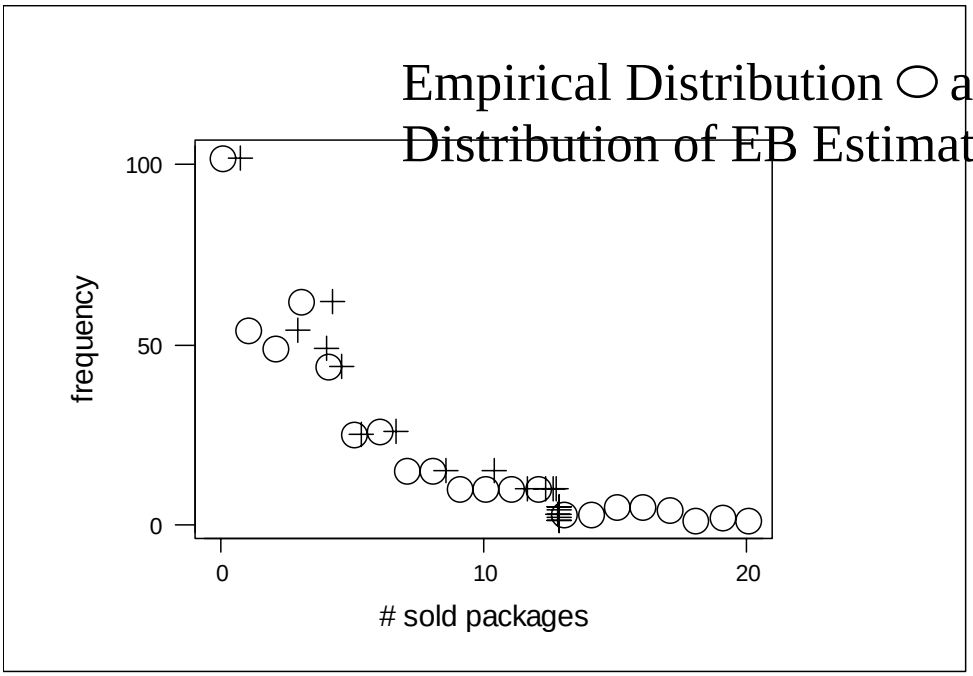
Empirische Bayes-Lösung:

Benutze Schätzung für P (z.B. $\hat{P} = \begin{pmatrix} \hat{\lambda}_1 \ldots \hat{\lambda}_k \\ \hat{p}_1 \ldots \hat{p}_k \end{pmatrix}$)
zur Berechung der posteriori-Verteilung

Dann: benutze EB-Schätzer $E(\lambda/x)$ an Stelle von x:

$$E(\lambda/x) = \sum_{j=1}^{k} \hat{p}\,(\lambda_j/x)\lambda_j$$

Vorteil: Varianzreduktion



Klassifikation basierend auf Mischverteilungen

Klassifikation in die Mischverteilungskomponenten:

klassifiziere x_i in diejenige Komponente j mit

$\hat{p}(\lambda_j/x_i) = \max_{1 \le l \le k} \hat{p}(\lambda_l/x_i)$

II. Theorie des NPMLE

II. Charakterisierung

log-likelihood *l* ist <u>konkav</u> auf der Menge *aller* diskreter Wahrscheinlichkeitsverteilungen Ω!

Hauptwerkzeug: <u>Richtungsableitung</u>

$$\Phi(P,Q) = \lim_{\alpha \to 0} \frac{l((1-\alpha)P + \alpha Q) - l(P)}{\alpha}$$

$$= \sum_{i=1}^{n} \frac{f(x_i,Q) - f(x_i,P)}{f(x_i,P)}$$

und insbesondere für die Einpunktverteilung an λ: Q_λ

<u>Gradientenfunktion</u>: $D_P(\lambda) = \Phi(P,Q_\lambda)$

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

führt auf: *mixture maximum likelihood theorem* (Lindsay 83a,b *Ann. Statist.*; Böhning 82 *Ann. Statist.*)

a) $\hat{P}$ is NPMLE $\Leftrightarrow$ $D_{\hat{P}}(\lambda) \leq 0$ for all λ

$$\Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i,\lambda)}{f(x_i,P)} \leq 1 \text{ for all } \lambda$$

b) $D_{\hat{P}}(\lambda) = 0$ for all support points λ of $\hat{P} = \begin{pmatrix} \hat{\lambda}_1 \dots \hat{\lambda}_k \\ \hat{p}_1 \dots \hat{p}_k \end{pmatrix}$

II.2 Algorithmen

zuverlässig konvergierende Algorithmen

VDM

vertex direction

$(1-\alpha)P + \alpha Q_\lambda$

step-length

$l((1-\alpha)P + \alpha Q_\lambda) - l(P)$ so groß wie möglich

$\approx \alpha D_P(\lambda) \rightarrow$ maximiere $D_P(\lambda)$ in λ!

VEM

bad vertex direction

$$P + \alpha P(\lambda^*) \{Q_\lambda - Q_{\lambda^*}\}$$

step-length

$l(P + \alpha P(\lambda^*) \{Q_\lambda - Q_{\lambda^*}\}) - l(P)$ so groß wie möglich!

$$\approx \alpha\, P(\lambda^*) \{D_P(\lambda) - D_P(\lambda^*)\}$$

$\rightarrow$maximiere $D_P(\lambda)$ in λ *und minimiere* $D_P(\lambda^*)$ *im Träger von P*!

VEM konvergiert wesentlich schneller als die VDM; Lesperance and Kalbfleisch 92 JASA: Bsp. mit #Iterationen: 2177 (VDM), 143 (VEM); Böhning 1986 *Metrika*, Böhning 1986 *JSPI*

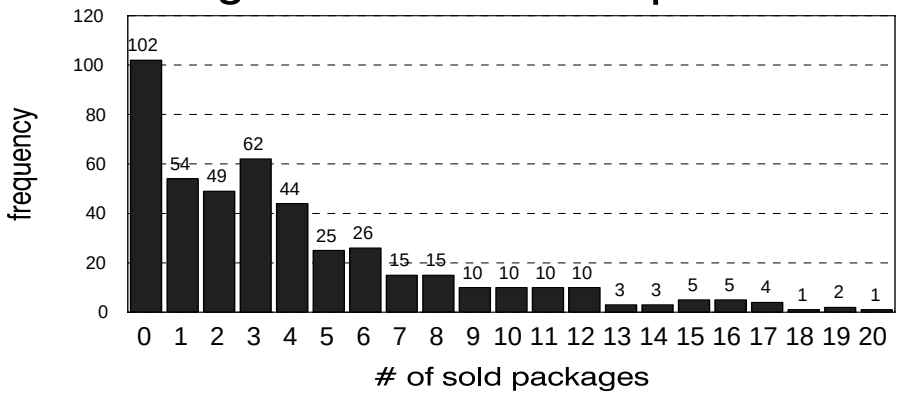
Praktische Realisierung

(Böhning, Schlattmann, Lindsay 1992, *Biometrics*)

Maximiere $l(P)$ im Simplex Ω aller W-verteilungen P auf dem Parameterraum von λ !

Phase I: Wähle ein approximierendes Gitter $\lambda_1,...,\lambda_L$ ($L \leq 50$)!
Maximiere $l(P)$ im Simplex Ω_{GRID} aller W-verteilungen P auf dem Gitter $\{\lambda_1,...,\lambda_L\}$ mit einem der o.g. Algorithmen!

Approximating Grid in Marketing Research Example



Hier: $\{\lambda_1,...,\lambda_L\} = \{0,1,2,...,20\}$ mit L=21
Mögliche Wahl der *Start* -Gewichte: $p_i = 1/21$

Veränderte algorithmische Strategie in der algorithmischen Maschine von C.A.MAN

Phase II: Benutze alle Parameterpunkte mit positivem Gewicht als Startwerte des EM Algorithmus: $\rightarrow$ NPMLE

Beispiel: (Marketing Research)

The program will use the following options
to compute NPMLE:
DATA-FILE:psu.dat
PARAMETER-GRID:DEFAULT
DISTRIBUTION:POISSON
ALGORITHM: VEM
STEP-LENGTH: FULL NR *(optimal step-length)*
ACCURACY: .000001
NUMBER OF ITERATIONS:12345

step 1560 max. dir. derivative 1.000001

gradient	1.0000 weight	**.1554** parameter	**.000000**
gradient	1.0000 weight	**.1272** parameter	**1.000000**
gradient	.9977 weight	.0000 parameter	2.000000
gradient	1.0000 weight	**.4026** parameter	**3.000000**
gradient	1.0000 weight	**.0725** parameter	**4.000000**
gradient	.9993 weight	.0000 parameter	5.000000
gradient	.9995 weight	.0000 parameter	6.000000
gradient	1.0000 weight	**.0540** parameter	7.000000
gradient	1.0000 weight	**.0914** parameter	**8.000000**
gradient	.9991 weight	.0000 parameter	9.000000
gradient	.9982 weight	.0000 parameter	10.000000
gradient	.9983 weight	.0000 parameter	11.000000
gradient	.9996 weight	.0000 parameter	12.000000
gradient	1.0000 weight	**.0968** parameter	**13.000000**
gradient	.9951 weight	.0000 parameter	14.000000
gradient	.9798 weight	.0000 parameter	15.000000
gradient	.9491 weight	.0000 parameter	16.000000
gradient	.9003 weight	.0000 parameter	17.000000
gradient	.8337 weight	.0000 parameter	18.000000
gradient	.7520 weight	.0000 parameter	19.000000
gradient	.6601 weight	.0000 parameter	20.000000

Log-Likelihood at iterate : -1130.61200

C.A.MAN did identify 7 grid points
with positive support

weight:	.1554	parameter:	.000000
weight:	.1272	parameter:	1.000000
weight:	.4026	parameter:	3.000000
weight:	.0725	parameter:	4.000000
weight:	.0540	parameter:	7.000000
weight:	.0914	parameter:	8.000000
weight:	.0968	parameter:	13.000000

Phase II:

step 3689 max. dir. derivative 1.000001
group: 1 weight: .0002 mean: .000000
group: 2 weight: .2440 mean: .204903
group: 3 weight: .4314 mean: 3.001951

group: 4 weight: .0713 mean: 3.001951
group: 5 weight: .0563 mean: 7.418164
group: 6 weight: .0951 mean: 7.418164
group: 7 weight: .1017 mean: 12.872540

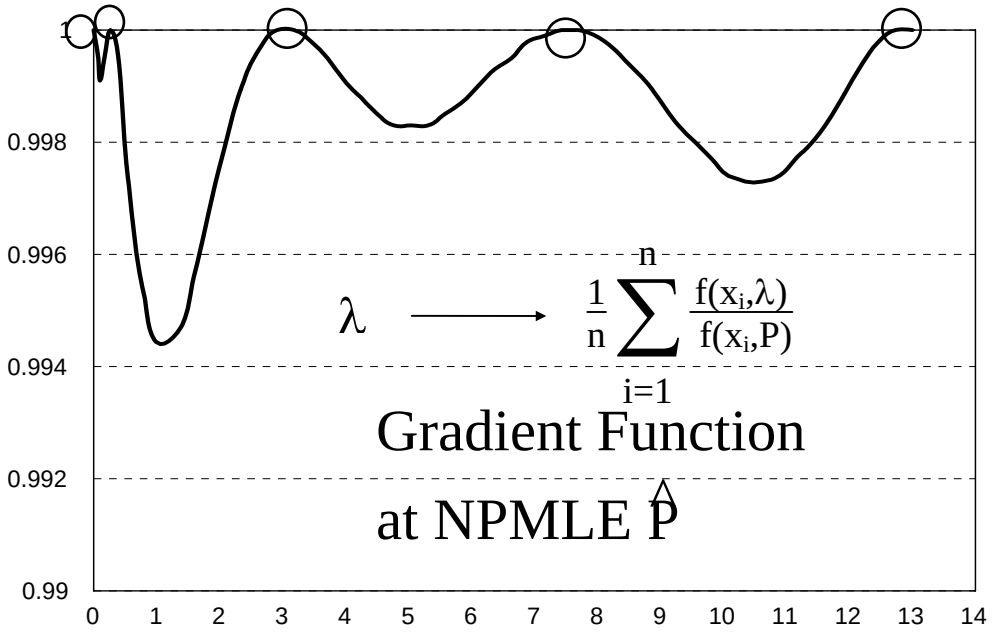
The NPMLE consists of 5 support points

Result after combining equal estimates:

weight: .0002 parameter: .000000
weight: .2440 parameter: .204903
weight: .5027 parameter: 3.001951
weight: .1514 parameter: 7.418164
weight: .1017 parameter: 12.872540

Log-Likelihood at iterate : -1130.07100

Log-Likelihood at iterate : -1130.61200 (based on appr. grid: $l(\hat{P}_{Grid})$)



III. Biometrische Anwendungen

III.1 Meta-Analyse

Ziel der Meta-Analyse:

Statistische Analyse einer Vielzahl von Einzelergebnissen (von unterschiedlichen Studien) mit dem Ziel einer integrativen Darstellung

Maßzahl von Interesse: (oft in der Epidemiologie)

Odds-Ratio ψ , log (*Odds-Ratio*) $\lambda=\log(\psi)$

n unabhängige Studien (Kohorten-, Fall-Kontroll) mit Schätzern:

$$\hat{\lambda}_1,\ldots,\hat{\lambda}_n$$

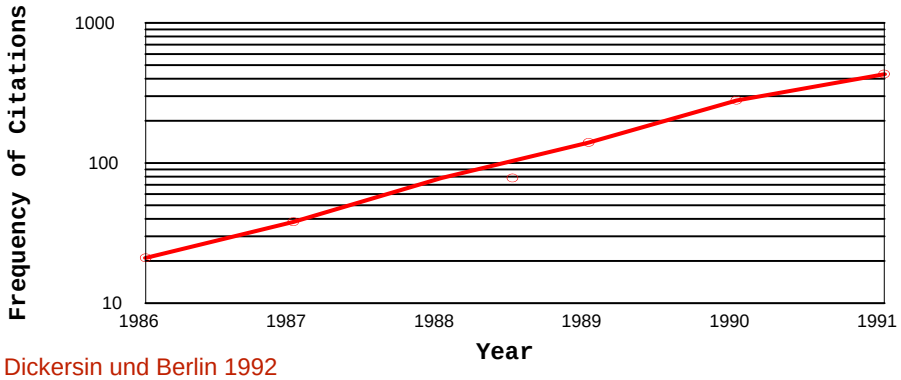
gewichteten Schätzer $\hat{\lambda}_{pool} = w_1\hat{\lambda}_1+\ldots +w_n\hat{\lambda}_n$

häufig: w_j *proportional zu* $1/\mathrm{var}(\hat{\lambda}_j)$

viele pros und cons! Review von (*Dickersin und Berlin 1992*, Epidemiologic Reviews)

Results of a MEDLINE-Search

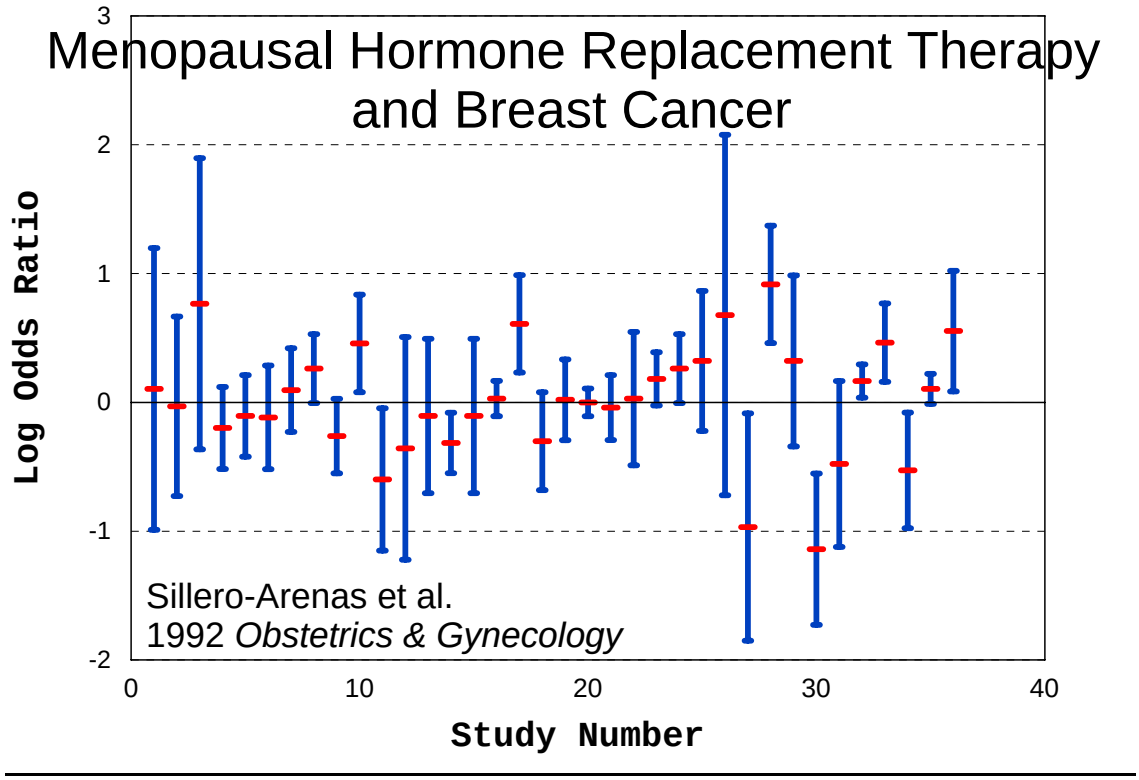
Search Term "Meta-Analysis"



after Dickersin und Berlin 1992

Beispiel

1999



<u>Homogeneität oder Heterogenität</u>

gepoolter Schätzer <u>angebracht</u> falls Homogenität vorliegt

$(\lambda_1=\lambda_2=...=\lambda_n)$

aber was tun, im Falle der Heterogenität?

Mischverteilungsmodellierung

das übliche Modell im homogenen Fall

$$\hat{\lambda}_i \sim \varphi(\,(x - \lambda)/\sigma_i),\ \varphi \text{ stand normal},\ \sigma_i = \text{var}(\hat{\lambda}_i)$$

dann: gepoolte Schätzer ist MLE, wobei $\hat{\lambda}_i$ die geschätzte Effektmaßzahl in der i-ten Studienpopulation ist

Analog bedeutet das Zulassen von Heterogenität in den $\hat{\lambda}_i$

$$\hat{\lambda}_i \sim \sum_{j=1}^{k} f(x,\lambda_j)p_j = \sum_{j=1}^{k} \varphi(\,(x - \lambda_j)/\sigma_i)\ p_j$$

Beispiel 1

The program will use the following options
to compute NPMLE:
DATA-FILE:oesall.dat
PARAMETER-GRID:COMPUTED
DISTRIBUTION:*NORMAL-DIFFERENT VARIANCE*
ALGORITHM: VEM
STEP-LENGTH: FULL NR
ACCURACY: .000010
NUMBER OF ITERATIONS: 1500

Approximation based on grid with 25 mean parameters

step 25 max. dir. derivative 1.000000

gradient	.4735 weight	.0000 parameter	-1.200000
gradient	.5115 weight	.0000 parameter	-1.108333
gradient	.5471 weight	.0000 parameter	-1.016667
gradient	.5813 weight	.0000 parameter	-.925000
gradient	.6155 weight	.0000 parameter	-.833333

gradient	.6511 weight	.0000 parameter	-.741667
gradient	.6893 weight	.0000 parameter	-.650000
gradient	.7310 weight	.0000 parameter	-.558333
gradient	.7769 weight	.0000 parameter	-.466666
gradient	.8271 weight	.0000 parameter	-.375000
gradient	.8802 weight	.0000 parameter	-.283333
gradient	.9319 weight	.0000 parameter	-.191666
gradient	.9749 weight	.0000 parameter	-.100000
gradient	**1.0000 weight**	**.2893 parameter**	**-.008333**
gradient	**1.0000 weight**	**.7107 parameter**	**.083334**
gradient	.9727 weight	.0000 parameter	.175000
gradient	.9222 weight	.0000 parameter	.266667
gradient	.8566 weight	.0000 parameter	.358334
gradient	.7846 weight	.0000 parameter	.450000
gradient	.7123 weight	.0000 parameter	.541667
gradient	.6427 weight	.0000 parameter	.633334
gradient	.5763 weight	.0000 parameter	.725000
gradient	.5128 weight	.0000 parameter	.816667
gradient	.4517 weight	.0000 parameter	.908334
gradient	.3931 weight	.0000 parameter	1.000000

Log-Likelihood at iterate : -16.92961

C.A.MAN did identify 2 grid points
with positive support

weight: .2893 parameter: -.008333
weight: .7107 parameter: .083334

weight: .4854 parameter: .046702
weight: .5146 parameter: .046702

Voll iterierter NPMLE:

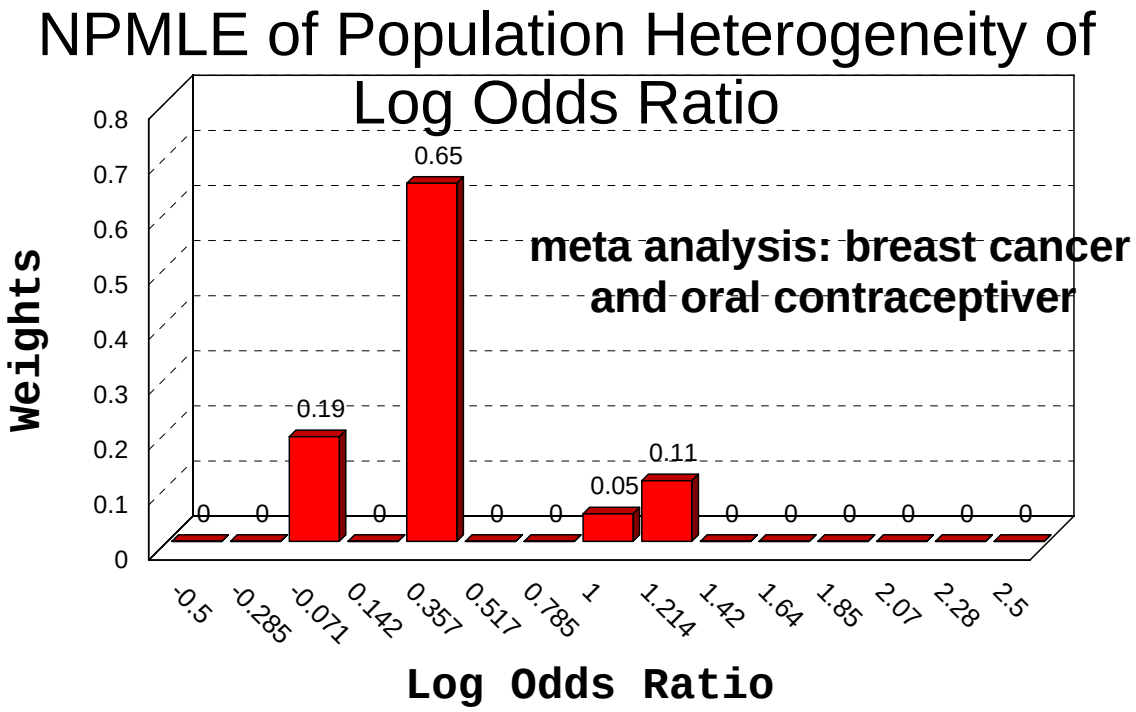
The NPMLE consists of 1 support points
Result after combining equal estimates:

weight: 1.0000 parameter: .046702

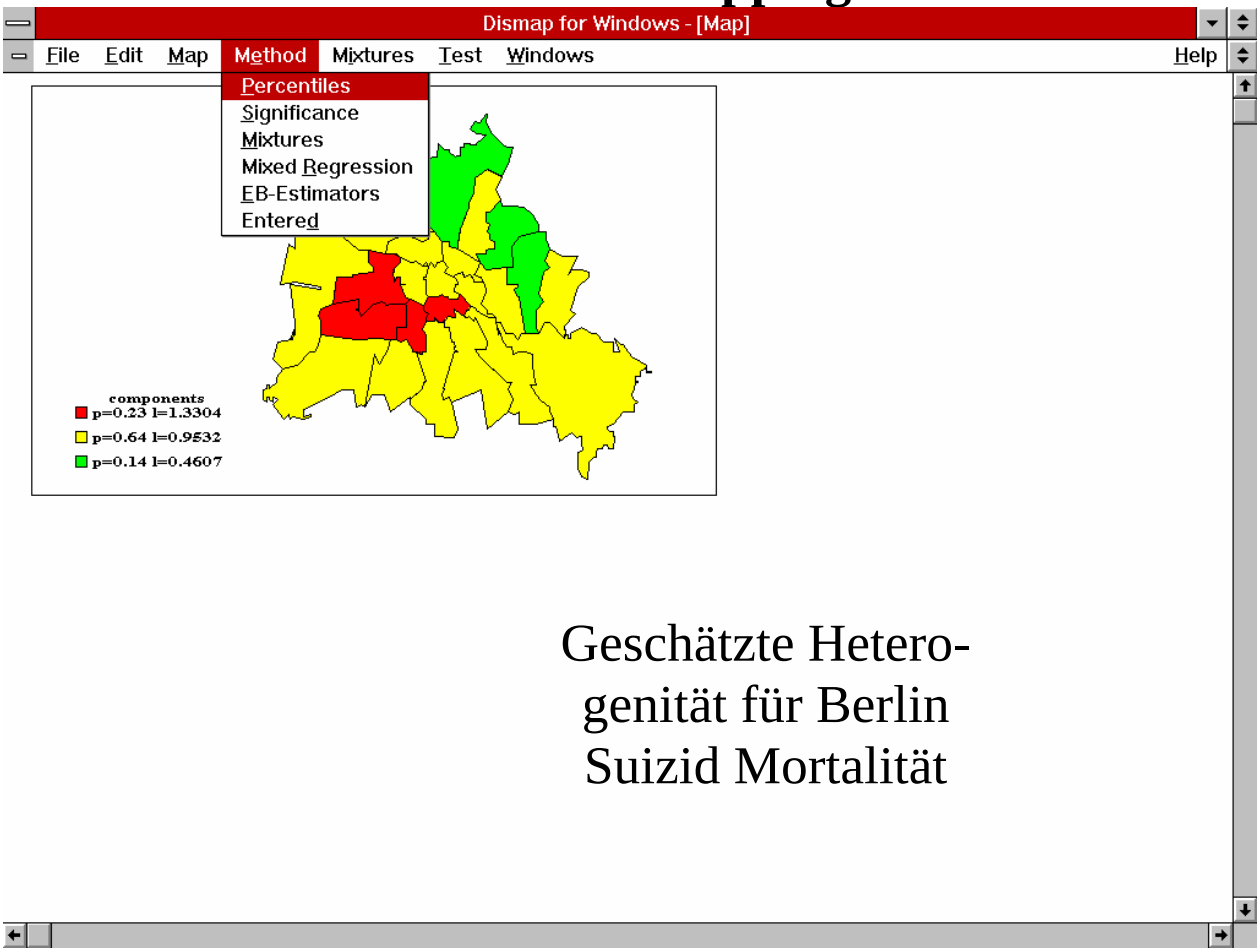
Log-Likelihood at iterate : -16.80735

Festhalten: keine Stützung von *Heterogeneität*!

Beispiel 2



.2 Disease Mapping



Geschätzte Hetero-
genität für Berlin
Suizid Mortalität

konventionelle *biometrische Methode:*
(basierend auf der SMR = O/E)

für die Region i: O_i ist Poisson mit Parameter λE_i

$$f(x_i,\lambda) = Po(o_i,\lambda E_i) = \exp(\lambda E_i)(\lambda E_i)^{o_i} / o_i!$$

mit $x_i = o_i/E_i$ *(beob. SMR),* λ *(theoret. SMR)*

Klassification basiert auf dem P-Wert unter einer *homogenen* Poissonverteilung

$$P(O_i \geq o_i) = Po(o_i,\lambda E_i) + Po(o_i+1,\lambda E_i) + \ldots$$

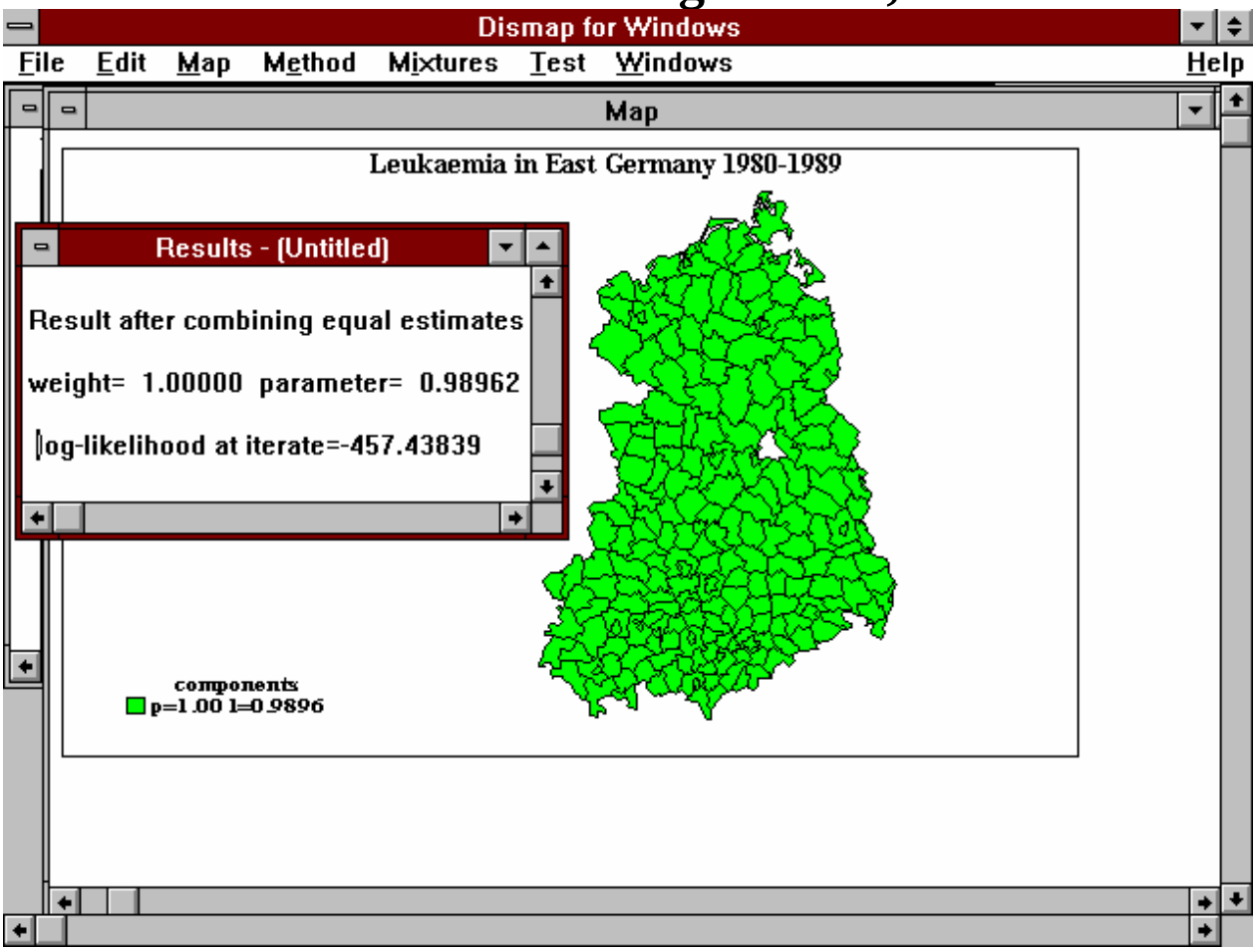
Nachteil: erzwingt *immer* eine Risikostruktur auf der Karte (,obwohl dies auch ein Zufallseffekt sein kann)

Lösung: Zulassen von Heterogeneität, d.h. Modellierung einer Mischung für $x_i = o_i/E_i$:

$$O_i \sim \sum_{j=1}^{k} f(x_i,\lambda_j)p_j = \sum_{j=1}^{k} Po(o_i,\lambda_j E_i)\, p_j$$

here $\hat{k} = 3$, $\hat{\lambda}_1=1.33$,
$\hat{p}_1=0.22$, $\hat{\lambda}_2=0.95$, $\hat{p}_2=0.64$, $\hat{\lambda}_3=0.46$, $\hat{p}_3=0.14$

Beispiel einer geschätzten homogenen Struktur
Leukämie in der ehemaligen DDR, 1980-1989



Model für Raten:

für Region i: o_i Fälle beobachtet, N_i unter Risiko

O_i Poisson mit Parameter λN_i

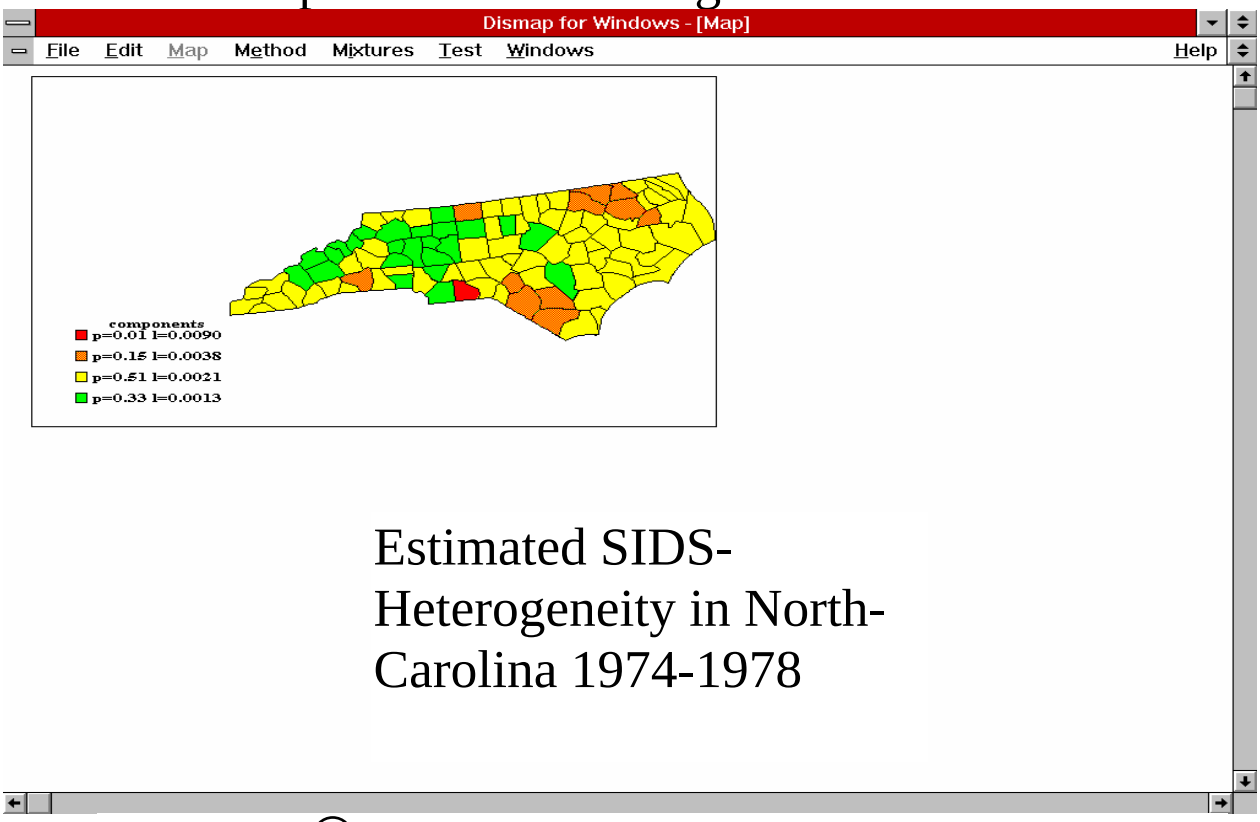
$$f(x_i,\lambda) = Po(o_i,\lambda N_i) = \exp(\lambda N_i)(\lambda N_i)^{o_i} / o_i!$$

with $x_i = o_i/N_i$ *(beob. Rate),* λ *(Populationsrate)*

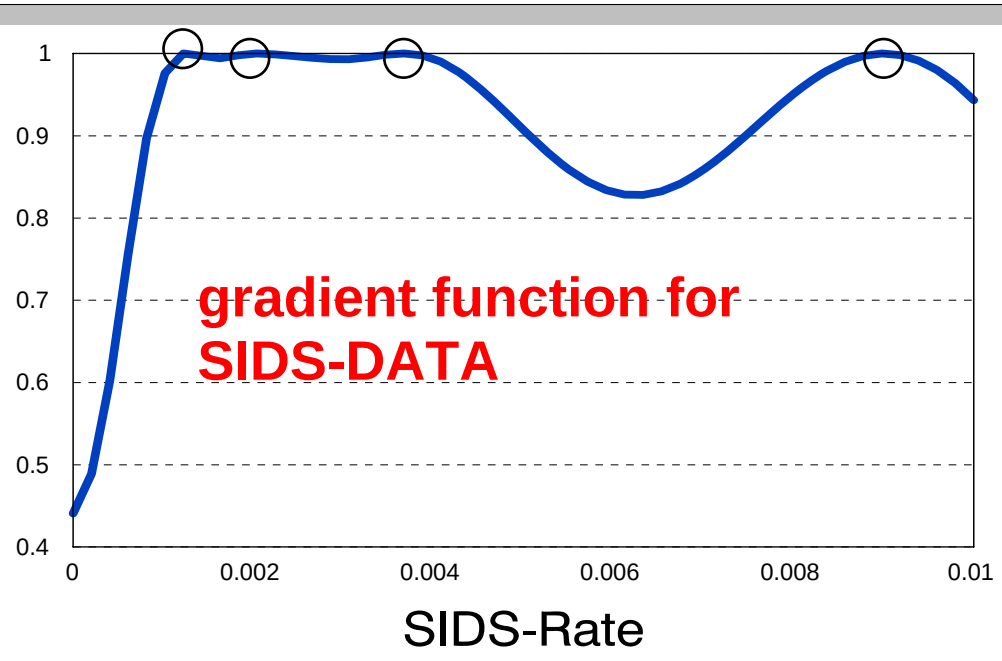
Modellierung der Heterogeneität wie gehabt!

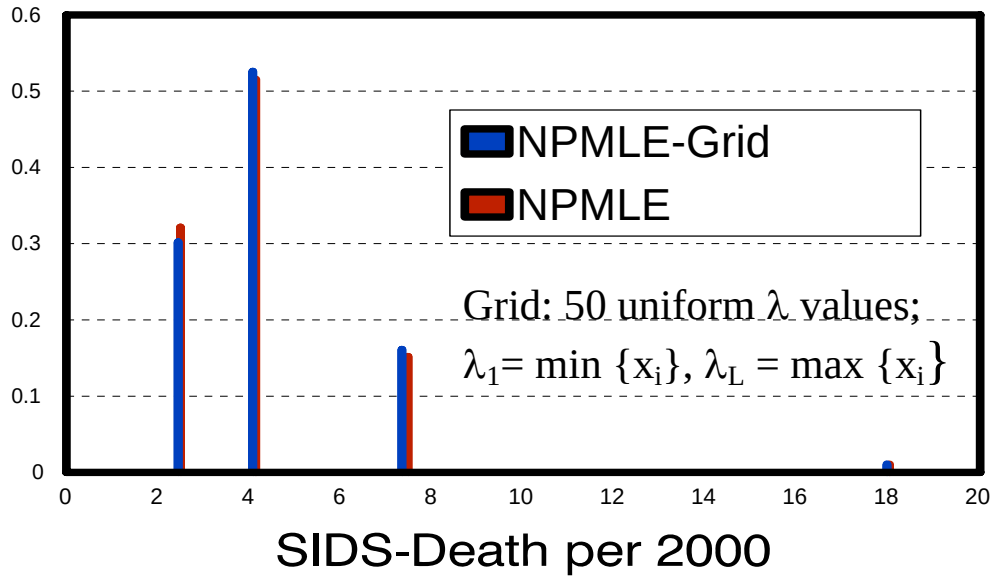
Beispiel: (Symons et al. 1982 *Biometrics*)

SIDS Rate in North Carolina: Vorschlag eines 2-Komponenten Mischungsmodells



Estimated SIDS-Heterogeneity in North-Carolina 1974-1978





Key Referenz:
Schlattmann & Böhning (1992) *Stat. in Med.*
Schlattmann, Dietz & Böhning (1996) *Stat. in Med.*

III.3 Binäre Regression

y binär, d.h. y=1 (Fall) oder y=0 (Kontrolle), **x** Kovariat

Ziel: Art der Abhängigkeit $p_{\mathbf{x}}$ vom Kovariat **x** darstellen, wobei

$p_x = P(Y=1/\mathbf{x})$

Bayes Theorem liefert

$$p_x = \frac{f_1(\mathbf{x})\, P(Y=1)}{f_1(\mathbf{x})\, P(Y=1) + f_0(\mathbf{x})\,(1-P(Y=1))} = \frac{f_1(\mathbf{x})\, p}{f_1(\mathbf{x})\, p + f_0(\mathbf{x})\,(1-p)}$$

mit $f_1(\mathbf{x}) = f(\mathbf{x}|Y=1)$ und $f_0(\mathbf{x}) = f(\mathbf{x}|Y=0)$

offenbar

Schätzbarkeit von Prävalenz p hängt vom Studientyp ab!

$$\log (p_{\mathbf{x}}/(1-p_{\mathbf{x}})) = \log (f_1(\mathbf{x})/f_0(\mathbf{x})) + \log ((p/(1-p))$$

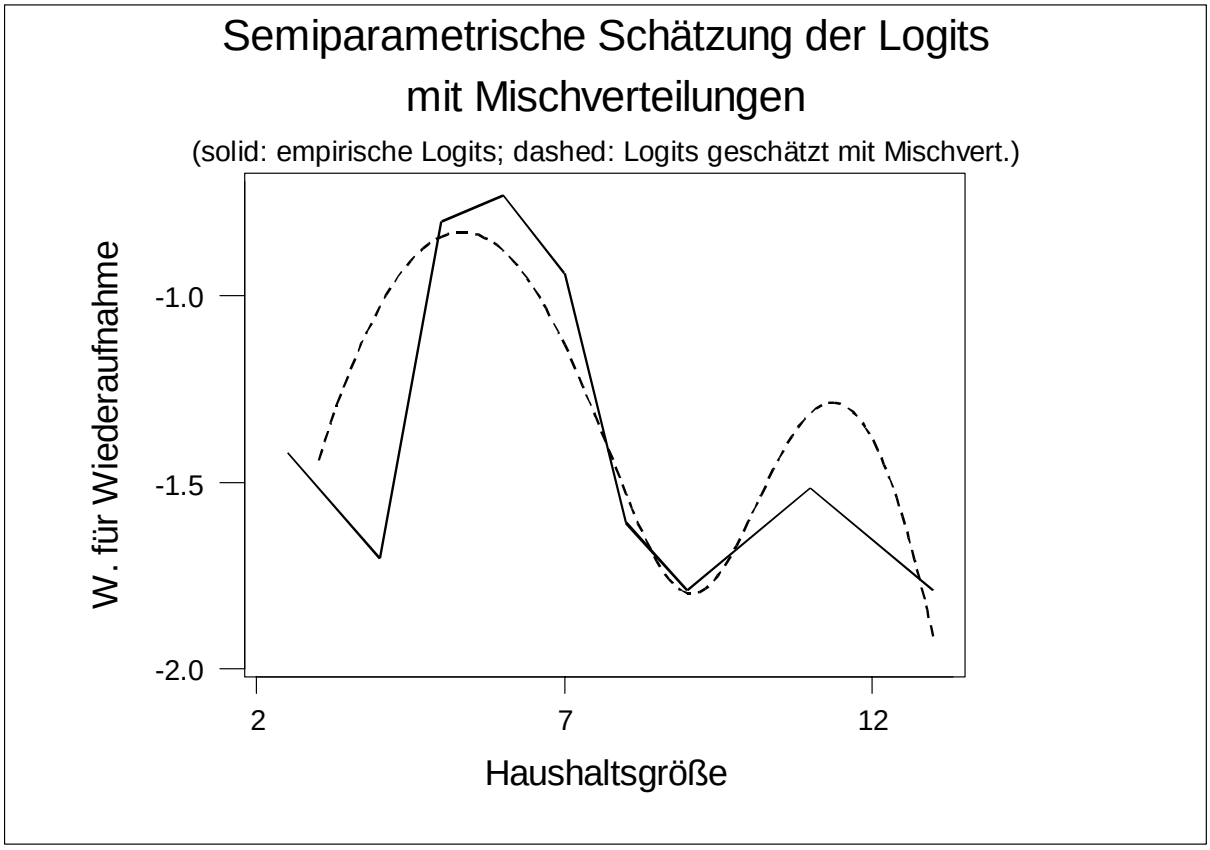
Idee:

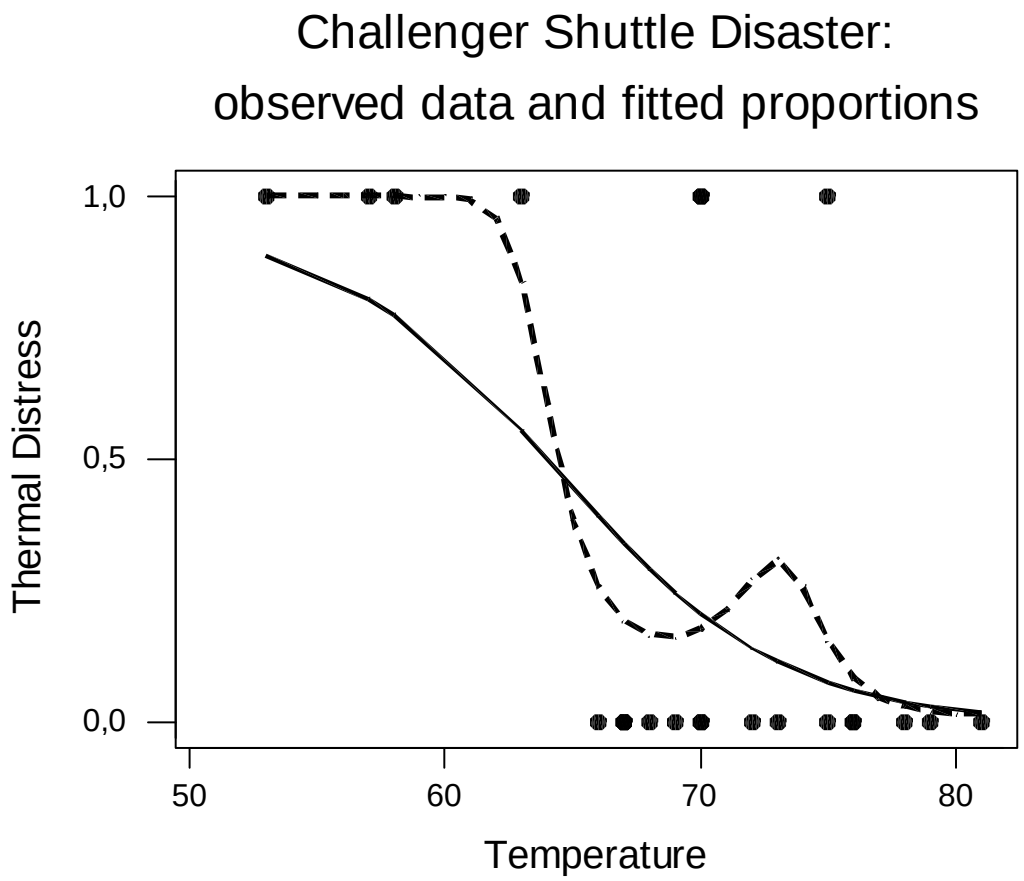
schätze $f_i(\mathbf{x})$ durch $f_i(\mathbf{x}, \hat{P})$ und erhalte $\log (\hat{p}_{\mathbf{x}}/(1-\hat{p}_{\mathbf{x}}))$!

Beispiel (Paquing 1995)

Follow-up Studie zur Auswirkung verschiedener Faktoren auf die Wiederaufnahmewahrscheinlichkeit in die Klinik *innerhalb von sechs Monaten* nach Entlassung bei an Schizophrenie erkrankten Patienten aus dem Raum Metro Manila

x Größe des Haushaltes (# Personen)
n = 251,
Besonderheit: für jedes x gibt es Replikationen!




Challenger Shuttle Disaster:
observed data and fitted proportions
Thermal Distress
1,0
0,5
0,0
50
60
70
80
Temperature

Perspektiven

Viele Aspekte nicht erwähnt, die im Zusammenhang mit C.A.MAN *verfügbar sind:*

Modellierung von Fruchtbarkeitsstudien
Medizinische Diagnostik

Überdispersionstest
Likelihood-Ratio Tests

Viele Aspekte nicht erwähnt, welche nicht verfügbar sind in C.A.MAN

Weitergehende Berücksichtigung von Kovariaten
Bootstrapping der Likelihood Ratio Statistik
zur weiteren Untersuchung der Anzahl der Komponenten
Mischungen von Normalverteilungen mit verschiedenen, unbekannten Varianzen