Computational Aspects of Quadrature Methods in Random Effects Models

Herwig Friedl
Institute of Statistics
University of Technology Graz, Austria
http://www.cis.tu-graz.ac.at/friedl

Outline

- Overdispersed GLMs
- EM Algorithm and Quadrature Approximation
- Non-Parametric Maximum Likelihood (NPML)
- Variance Components Models
- Random Parameter Models

Overdispersion

Example: Teenager Pregnancy Rates in NC Florida

Data: $n = 13$ counties, y_i teen preg's out of totals m_i.

Model: $y_i \overset{ind}{\sim} binomial(\mu_i, m_i)$,

$$g(\mu_i) = \log \frac{\mu_i}{m_i - \mu_i} = \beta_0$$

MLE: $\hat{\beta}_0 = -3.35$, but $dev = 89.9$ (df=12)

General problem of **overdispersion** in Poisson (counts) or binomial (frequencies) models:
Residual deviance (dev) from the model is much larger than the residual degree of freedom (df).

Interpret this as evidence that there are other factors varying which are not accounted for in the model, but which are associated with the response:
Omitted variables which have to be allowed for, e.g., $g(\mu_i) = \beta_0 + u_i^t \gamma$

Example: add fixed county effects
$\Rightarrow$ saturated model !!

A simple way of representing the extra variation is by adding an unobservable **random effect** z to the **linear predictor**

$$g(\mu_i) = x_i^t \beta + z_i$$

z_i are iid from some (unknown) distribution $F(z)$.

Example: add random county effects

$\Rightarrow \log \frac{\mu_i}{m_i - \mu_i} = \beta_0 + z_i$

$\Rightarrow$ MLE ??

MLE

Conditional model specification

$y_i | z_i \overset{ind}{\sim} F(y_i | z_i; \theta)$ (specified), density $f(y_i | z_i; \theta)$

$z_i \overset{iid}{\sim} N(0, 1)$ (assumed), density $f(z_i) = \phi(z_i)$

$$g(\mu_i) = x_i^t \beta + z_i \sigma_z = (x_i^t, z_i)\theta$$

with $\mu_i = E(y_i | z_i)$ and $\theta = (\beta^t, \sigma_z)^t$.

Marginal Maximum Likelihood Estimate (MLE)

$$\hat{\theta} = \underset{\theta}{\text{argmax}} (\log) f(y; \theta)$$

where

$$f(y; \theta) = \int f(y|z; \theta) f(z) dz.$$

EM Algorithm

Indirect maximization technique:

instead of maximizing $f(y;\theta)$ directly we iteratively maximize $Q(\theta|\theta^{(t)})$ wrt θ (keeping $\theta^{(t)}$ fix), giving $\hat{\theta}$ at convergence.

Generally:

$$Q(\theta|\theta^{(t)}) = E\Big(\log f(y,z;\theta)|y;\theta^{(t)}\Big)$$

Example (overdispersion): $f(y,z) = \prod_i f(y_i|z_i)f(z_i)$

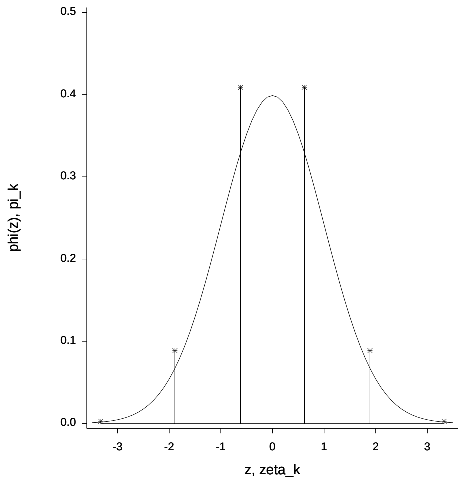
$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \int \log\Big(f(y_i|z_i;\theta)f(z_i)\Big) f(z_i|y_i;\theta^{(t)})dz_i$$

$$f(z_i|y_i;\theta^{(t)}) = \frac{f(y_i|z_i;\theta^{(t)})f^{(t)}(z_i)}{\int f(y_i|z_i;\theta^{(t)})f^{(t)}(z_i)dz_i}$$

If $f(z) = \phi(z)$, K-point **Gaussian-Quadrature** yields (discrete) approximation

$$\int h(z)\phi(z)dz \approx \sum_{k=1}^{K} h(\zeta_k)\pi_k$$

mass points ζ_k and masses π_k are **'known'** and available from tables.



This gives for the marginal response density $f(y_i;\theta)$

$$\int f(y_i|z_i;\theta)\phi(z_i)dz_i \approx \sum_{k=1}^{K} f(y_i|\zeta_k;\theta)\pi_k =: f_K(y_i;\theta),$$

approximated by likelihood of a Finite Mixture Model. Moreover,

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= \sum_{i=1}^{n} \int \log\Big(f(y_i|z_i;\theta)f(z_i)\Big) f(z_i|y_i;\theta^{(t)})dz_i \\ &\approx \sum_{i=1}^{n}\sum_{k=1}^{K} \log\Big(f(y_i|\zeta_k;\theta)\pi_k\Big) w_{ik}^{(t)} \end{aligned}$$

with weights

$$w_{ik}^{(t)} = \frac{f(y_i|\zeta_k;\theta^{(t)})\pi_k}{\sum_{l=1}^{K} f(y_i|\zeta_l;\theta^{(t)})\pi_l} \approx P^{(t)}(z_i = \zeta_k|y_i).$$

In the M-step, $Q(\theta|\theta^{(t)})$ is maximized wrt θ, ie

$$\frac{\partial}{\partial\theta} Q(\theta|\theta^{(t)}) \approx \sum_{i=1}^{n}\sum_{k=1}^{K} \frac{\partial}{\partial\theta} \log\Big(f(y_i|\zeta_k;\theta)\Big) w_{ik}^{(t)}$$

How to interpret $f(y_i|\zeta_k;\theta^{(t)})$?

originally: $\mu_i = E(y_i|z_i)$ with

$$g(\mu_i) = x_i^t\beta + z_i\sigma_z$$

now: $\mu_{ik} = E(y_i|z_i = \zeta_k)$ with

$$g(\mu_{ik}^{(t)}) = x_i^t\beta^{(t)} + \zeta_k\sigma_z^{(t)}$$

Note: if $K = 1$, then $\pi = 1$, $\zeta = 0$ and $w_i = 1$,
(an ordinary GLM without random effects)

E-step: update weights w

M-step: expand data to length nK by replicating y and X each K times. Use weighted MLE

y	w		β		σ_z
y_1	w_{11}	x_{11}	$\dots$	x_{1p}	z_1
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_n	w_{n1}	x_{n1}	$\dots$	x_{np}	z_1
y_1	w_{12}	x_{11}	$\dots$	x_{1p}	z_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_n	w_{n2}	x_{n1}	$\dots$	x_{np}	z_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_1	w_{1K}	x_{11}	$\dots$	x_{1p}	z_K
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_n	w_{nK}	x_{n1}	$\dots$	x_{np}	z_K

Example: Teenager Pregnancy Rates

Model: $y_i|z_i \overset{ind}{\sim} binomial(\mu_i, m_i)$,

$$z_i \overset{iid}{\sim} N(0,1)$$

$$\log \frac{\mu_i}{m_i-\mu_i} = \beta_0 + z_i\sigma_z$$

Results:

K	$\hat\beta_0$	$\hat\sigma_z$	dev	iter
1	-3.35	(0.02)	89.9	$-$
2	-3.24	0.27	40.9	6
3	-3.41	0.25	42.4	8
4	-3.22	0.27	34.1	10
7	-3.43	0.26	30.5	21

Deviance: $dev = -2\Big(\log f_K(y;\hat\mu) - \log f_1(y;y)\Big)$,

$f_1(y;y)$ is the saturated binomial likelihood

$f_K(y;\hat\mu) = \sum_k f(y|\zeta_k;\hat\theta)\pi_k$

NPMLE

- If $f(z)=\phi(z)$: K-point **Gaussian-Quadrature**

$$\int f(y|z)\phi(z)dz \approx \sum_{k=1}^{K} f(y|\zeta_k)\pi_k$$

mass points ζ_k (locations) and masses π_k (mixture proportions) are **'known'** (in the likelihood of a finite mixture)! Add the variable ζ to the linear predictor.

- If $f(z)$ is unknown: we have to estimate all its prob's π_k on all its unknown support points ζ_k.

For the **Non-Parametric Maximum Likelihood Estimation** (NPMLE) we treat ζ_k and π_k as **'unknown parameters'** which have to be estimated concurrently with β. This is now an exact mixture ML problem.

$$\int f(y|z)f(z)dz = \sum_{k=1}^{K} f(y|\zeta_k)\pi_k$$

With $\theta = (\beta^t, \zeta_1, \dots, \zeta_K)^t$ this results again in

$$Q(\theta,\pi|\theta^{(t)},\pi^{(t)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \log\Big(f(y_i|\zeta_k;\theta)\pi_k\Big) w_{ik}^{(t)}$$

E-step: update weigths

$$w_{ik}^{(t)} = \frac{f(y_i|\zeta_k^{(t)};\theta^{(t)})\pi_k^{(t)}}{\sum_{l=1}^{K} f(y_i|\zeta_l^{(t)};\theta^{(t)})\pi_l^{(t)}}$$

M-step: additional maximization of $Q(\theta,\pi|\theta^{(t)},\pi^{(t)})$ wrt π and subject to $\sum_k \pi_k = 1$ (Lagrange multiplier)

$$\pi_k^{(t)} = \frac{1}{n}\sum_{i=1}^{n} w_{ik}^{(t)}$$

To estimate the mass points we add the K-level factor $\zeta_1,\dots,\zeta_K$ to the linear predictor

$$g(\mu_{ik}^{(t)}) = x_i^t\beta^{(t)} + \zeta_k^{(t)}$$

y	w	β			ζ			
y_1	w_{11}	x_{11}	$\dots$	x_{1p}	1	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	w_{n1}	x_{n1}	$\dots$	x_{np}	1	0	$\dots$	0
y_1	w_{12}	x_{11}	$\dots$	x_{1p}	0	1	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	w_{n2}	x_{n1}	$\dots$	x_{np}	0	1	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_1	w_{1K}	x_{11}	$\dots$	x_{1p}	0	0	$\dots$	1
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	w_{nK}	x_{n1}	$\dots$	x_{np}	0	0	$\dots$	1

Example: Teenager Pregnancy Rates

Model: $y_i|z_i \overset{ind}{\sim} binomial(\mu_i, m_i)$,

$z_i \overset{iid}{\sim} F(z)$ (unknown)

$\log \frac{\mu_i}{m_i-\mu_i} = \beta_0 + z_i$

Results:

K	$\hat\beta_0$	$\hat\sigma_z$	dev	iter
1	-3.35	(0.02)	89.9	–
2	-3.24	0.27	40.9	8
3	-3.25	0.29	35.5	11
4	-3.23	0.34	**31.1**	12
7	-3.23	0.34	31.1	27

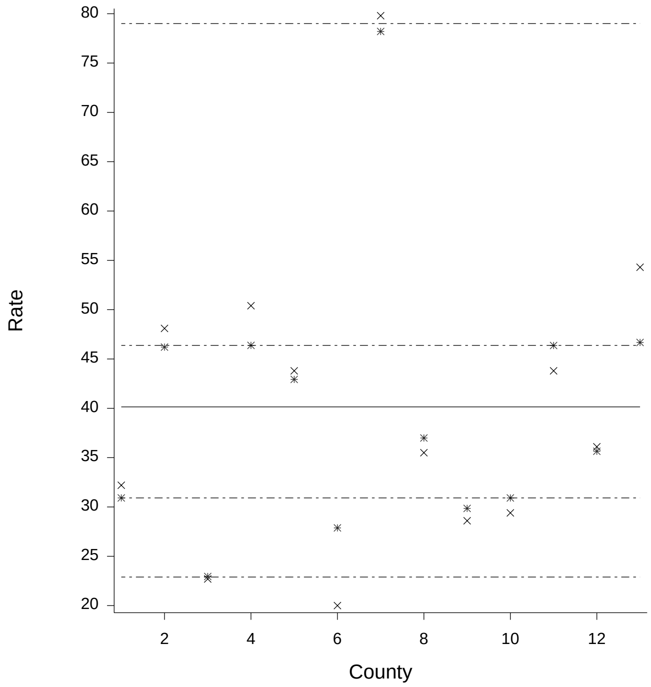
$\hat\sigma_z^2 = \sum_k (\hat\zeta_k - \overline{\hat\zeta})^2 \hat\pi_k$

Deviance: $dev = -2\Big(\log f_K(y;\hat\mu) - \log f_1(y;y)\Big)$,

$f_1(y;y)$ is the saturated binomial likelihood

$f_K(y;\hat\mu) = \sum_k f(y|\hat\zeta_k;\hat\theta)\hat\pi_k$

$K = 2 : \hat\zeta$	0.27	-0.26		
$\hat\pi$	0.50	0.50		
$3 : \hat\zeta$	0.28	-0.19	-0.50	
$\hat\pi$	0.49	0.38	0.13	
$\mathbf{4} : \hat\zeta$	0.78	0.21	-0.22	-0.52
$\hat\pi$	0.08	0.42	0.37	0.13
	weights w_{ik}			
Alachua			1.00	
Bradford		0.99	0.01	
Clay			0.01	0.99
Columbia		1.00		
Dixie		0.77	0.22	0.01
Gilchrist		0.03	0.54	0.43
Hamilton	0.98	0.02		
Lafayette		0.43	0.47	0.10
Levy		0.01	0.83	0.16
Marion			1.00	
Putman		1.00		
Suwanee		0.31	0.68	0.01
Union	0.03	0.92	0.05	



Variance Components Models

Two-level nested data structure: e.g., patient j is nested in clinic i. Data y_{ij} and $y_{ij'}$ are correlated because they stem from the same clinic i. To account for that, they share the same random effect

$$g(\mu_{ij}) = x_{ij}^t \beta + z_i.$$

Let $y_i = (y_{i1}, \ldots, y_{in_i})$ with joint density $f(y_i;\theta)$. Because the y_is are independent

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \log\Big(f(y_i|\zeta_k;\theta)\pi_k\Big) w_{ik}^{(t)}$$

Assume that observations in a cluster are conditionally independent

$$f(y_i|z_i;\theta) = \prod_{j=1}^{n_i} f(y_{ij}|z_i;\theta)$$

NPMLE gives

$$Q(\theta,\pi|\theta^{(t)},\pi^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{k=1}^{K} \log\Big(f(y_{ij}|\zeta_k;\theta)\pi_k\Big) w_{ik}^{(t)}$$

with weights computed at the cluster level (E)

$$\begin{aligned} w_{ik}^{(t)} &= \frac{f(y_i|\zeta_k^{(t)};\theta^{(t)})\pi_k^{(t)}}{\sum_{l=1}^K f(y_i|\zeta_l^{(t)};\theta^{(t)})\pi_l^{(t)}} \\ &= \frac{\pi_k^{(t)} \prod_{j=1}^{n_i} f(y_{ij}|\zeta_k^{(t)};\theta^{(t)})}{\sum_{l=1}^K \pi_l^{(t)} \prod_{j=1}^{n_i} f(y_{ij}|\zeta_l^{(t)};\theta^{(t)})} \end{aligned}$$

and applied at the individual level (M).

Maximizing Q wrt π subject to $\sum_k \pi = 1$ gives

$$\pi_k^{(t)} = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^{n} n_i w_{ik}^{(t)}$$

Example: Bacteria Counts in Graz, Austria

Question: cfu increase near compost arrangements C?

Known: Concentration depends on weather conditions (temperature T, humidity H)

Observations from 2 living areas (with/without C)

Over 1 year the activity of micro-organisms was monitored every 2 weeks

Data: $y_i = (y_{i4}, y_{i5}, y_{i6})$, $i = 1, \ldots, n = 50$

	$C=0$			$C=1$		
date	H	T	y	H	T	y
08.03	29	50	2,1,1	23	50	1,1,0
22.03	37	45	1,3,2	38	45	4,2,0
05.04	31	54	5,1,1	32	54	0,2,0
19.04	33	70	1,2,1	38	68	2,2,7
etc						
08.08	53	82	2,0,0	48	82	2,3,1
22.08	43	84	0,2,0	45	84	2,1,2

Gauge: 6-stage air sampler, 6 petri dishes each with same agar medium, filters with increasing fineness between stages:

1-3: particles which are usually filtered out in the nose, pharynx and trachea

4-6: small micro-organisms which could be inhalled into the lung and reaches bronchi and alveoli

Model: $y_{ij}|z_i \overset{ind}{\sim} Poiss(\mu_{ij})$, $i = 1, \ldots, 50$,

$$j = 4, 5, 6$$

$$z_i \overset{iid}{\sim} F(z) \text{ (unknown)}$$

$$\log \mu_{ij} = \beta_t t_i + \beta_{t^2} t_i^2 + stage_j * site_i + z_i$$

Results:

K	1	$\hat{\beta}_t$	$\hat{\beta}_{t^2}$	dev
1	$-1.21(.59)$	$.076(.021)$	$-.0007(.0002)$	258.8
2	$-1.88(.44)$	$.093(.034)$	$-.0008(.0003)$	**224.1**
3	-1.88	$.093$	$-.0008$	224.1

	$Poisson$		$K=2$	
	$refer$	$comp$	$refer$	$comp$
$stg4$	$-.18(.21)$	$-.12(.20)$	$-.09(.23)$	$-.12(.20)$
$stg5$	$.22(.19)$	$-.17(.20)$	$.31(.21)$	$-.17(.20)$
$stg6$	$-.48(.22)$	$-(-)$	$-.39(.25)$	$-(-)$

$\hat{\zeta}$	-1.02	-2.11
$\hat{\pi}$	0.21	0.79

Random Coefficient Models

Observation specific random intercept and slope(s).

Only simple structures considered; no crossed effects!

$f(y; z) = \prod_i f(y_i|z_i) f(z_i)$ always required.

Model:

$$g(\mu_{ij}) = x_{ij}^t \beta + z_{0i} + \tilde{x}_{ij} z_{1i}$$

$z_i = (z_{0i}, z_{1i}) \overset{iid}{\sim} F_2(z)$, an unknown bivariate cdf.

For Gaussian Quadrature this would require multiple integration!

NPML again results in K mass points $\zeta_k = (\zeta_{0k}, \zeta_{1k})$ with masses π_k.

The same considerations as before gives now

$$g(\mu_{ijk}) = x_{ij}^t \beta + \zeta_{0k} + \tilde{x}_{ij} \zeta_{1k}$$

mass points enter the model as a 'factor' and a 'factor:variable' interaction.

For the simple model $g(\mu_i) = z_{0i} + x_i z_{1i}$, the NPMLE is based on

$$g(\mu_{ik}) = \zeta_{0k} + x_i \zeta_{1k}$$

In the M-step use weighted MLE on the expanded data:

y	w	ζ_0				ζ_1			
y_1	w_{11}	1	0	$\ldots$	0	x_1	0	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	w_{n1}	1	0	$\ldots$	0	x_n	0	$\ldots$	0
y_1	w_{12}	0	1	$\ldots$	0	0	x_1	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	w_{n2}	0	1	$\ldots$	0	0	x_n	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_1	w_{1K}	0	0	$\ldots$	1	0	0	$\ldots$	x_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	w_{nK}	0	0	$\ldots$	1	0	0	$\ldots$	x_n

Example: Epilepsy Data

59 epileptics assigned to 2 treatment (t) groups

$y_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})$ number of seizures of ith patient

b = log(baseline/4) and 4 more observations monitored at visit v $= (j - 2.5)/5$, $j = 1, 2, 3, 4$.

Model: $y_{ij}|z_i \overset{ind}{\sim} NegBin(\mu_{ij}, \alpha)$,

$$z_i = (z_{0i}, z_{1i}) \overset{iid}{\sim} F_2(z) \text{ (unknown)}$$

$$\log \mu_{ij} = \beta_1 \mathsf{b}_i + \beta_2 \mathsf{t}_i + \beta_3 \mathsf{b}_i{:}\mathsf{t}_i + \beta_4 \log(\mathsf{age}_i) + z_{0i} + \mathsf{v}_{ij} z_{1i}$$

Independence models: $K = 1$, $n = 236$ observations

• Poisson: $\alpha \to \infty$
saturated: $\ell = -383.0$
fixed z_0, z_1: $\ell = -817.9$; $dev = 869.9$ (df=230)
fixed z_{0i}, z_{1i}: $\ell = -523.4$; $dev = 280.7$ (df=118)

• Negative Binomial: α estimated
saturated: ℓ=? (-514.1^1, -407.2^2)
^{1}fixed z_0, z_1: $\ell = -647.5$; $dev^* = 266.8$ (df=229)
^{2}fixed z_{0i}, z_{1i}: $\ell = -513.9$; $dev^* = 213.5$ (df=117)

Dependence Model: $K > 1$, $n = 59$ epileptics

Conditional Negative Binomial, given random intercept and random slope:

K	$-\ell$	$\hat{\alpha}$	$\log(\mathsf{b}/4)$	t	$\log(\mathsf{b}/4)\!:\!\mathsf{t}$	$\log(\mathsf{age})$
1	647.5	2.8	0.90	−0.90	0.35	0.56
2	627.0	5.7	0.95	−0.85	0.21	0.63
3	**620.8**	7.4	0.92	−0.52	0.15	0.58
6	617.7	9.2	0.92	−0.48	0.12	0.65
8	617.5	9.6	0.92	−0.49	0.14	0.56

$\ell = \log f_K(y; \hat{\mu}, \hat{\alpha})$, estimated marginal log-likelihood.

If the deviance is twice the difference between ℓ and the 'maximum' saturated Negative Binomial log-likelihood (-383.0 for the Poisson), then $dev = 475.8$ (df=223) for $K = 3$, and
$dev^1 = 529.0$ (df=229), $dev^2 = 261.8$ (df=117).

K	$Intercept$	$v - slope$	$corr$
2	$-1.86(.39)$	$-.30(.08)$	1.00
3	$-1.78(.47)$	$-.25(.07)$	$-.92$
6	$-2.02(.52)$	$-.26(.89)$	$-.19$
8	$-1.75(.52)$	$-.25(.90)$	$-.22$