Bayesian Generalized linear mixed models with data missing not at random

Overview:

- Two simple introductory examples of data missing not at random (MNAR)
- Missing mechanism and likelihood in the case of missing at random (MAR) as defined by Rubin (1976)
- Missing mechanism and Bayesian inference in the case of MAR as defined by Schafer(1997)
- Bayesian GLMMs with nonignorable nonresponse
- Selection model, with example
- Shared parameter model
- References

Random sample from a Bernoulli distribution with missing data

- Let $(y_1, \ldots, y_n)$ an iid sample from a $Bernoulli(p)$
- $p = \mathrm{E}(y_i) = P(y_i = 1)$, $0 < p < 1$
- $m < n$ observations are missing:

y_i	r_i
1	1
0	1
$\vdots$	$\vdots$
1	1
?	0
?	0
$\vdots$	$\vdots$
?	0

- We introduce indicator variables r_i:

$$r_i = \begin{cases} 1 & \text{if } y_i \text{ is observed (reported)} \\ 0 & \text{if } y_i \text{ is missing (not reported)} \end{cases}$$

- The indicator variables r_i are also random variables

- The missing process can be characterised through the conditional distributions of r_i given y_i:

$$
\begin{aligned}
P(r_i = 1|y_i = 1) &= \alpha_1 \qquad\qquad & P(r_i = 1|y_i = 0) &= \alpha_0 \\
P(r_i = 0|y_i = 1) &= 1 - \alpha_1 \qquad\qquad & P(r_i = 0|y_i = 0) &= 1 - \alpha_0
\end{aligned}
$$

with $0 < \alpha_0, \alpha_1 < 1$.

- Theorem of Bayes:

$$\mathrm{E}(y_i|r_i = 1) = P(y_i = 1|r_i = 1) = \frac{p\alpha_1}{p\alpha_1 + (1-p)\alpha_0} \tag{1}$$

and

$$\mathrm{E}(y_i|r_i = 0) = P(y_i = 1|r_i = 0) = \frac{p(1-\alpha_1)}{p(1-\alpha_1) + (1-p)(1-\alpha_0)} \tag{2}$$

The conditional expectations in (1) und (2) are equal iff $\alpha_0 = \alpha_1$.

- On the other hand

$$\mathrm{E}(y_i) = p = \mathrm{E}(y_i|r_i = 1)P(r_i = 1) + \mathrm{E}(y_i|r_i = 0)[1 - P(r_i = 1)] \tag{3}$$

- The interesting question from a statistical point of view is: *Can we estimate the probability or expectation p from the $n - m$ observed values?*

  Answer: Only if $\mathrm{E}(y_i|r_i = 1) = \mathrm{E}(y_i|r_i = 0)$ in (3), that is when $\alpha_0 = \alpha_1$ holds, since then: $p = \mathrm{E}(y_i|r_i = 1) \longrightarrow$ Missing (completely) at random M(C)AR

- What happens if $\alpha_0 \neq \alpha_1$? We can only estimate

  - $\mathrm{E}(y_i|r_i = 1)$
  - $P(r_i = 1)$

  by relative frequencies. But $\mathrm{E}(y_i|r_i = 0)$ is not identifiable from the observed data $\longrightarrow$ MNAR

  Example: $p = 0.4$, $\alpha_0 = 0.5$, $\alpha_1 = 0.9$. Then $\mathrm{E}(y_i|r_i = 1) = 0.55 > p \longrightarrow$ $(\sum_{i=1}^{n-m} y_i)/(n - m)$ (observed data) overestimates p

Motivation for three different approaches to the problem of MNAR data

1. We can make some vague assumption ($\longrightarrow$ Bayes) for α_0 und α_1 $\longrightarrow$ include missing data process in the estimation procedure for p

2. We assume $P(y_i|r_i = 0) = P(y_i|r_i = 1)$. Therefore we equate or constrain the unidentifiable parameter to an identifiable parameter. This is essentially the idea of pattern mixture models. Verbeke and Molenberghs (2000) gives an extensive and excellent overview about pattern mixture models in the context of linear mixed models and provides many references.

3. No such assumption is possible $\longrightarrow$ Compute bounds for p

   With (3) we have

$$p_{\min} = \underbrace{\mathrm{E}(y_i|r_i = 1)P(r_i = 1)}_{\text{if } \mathrm{E}(y_i|r_i = 0) = 0} < p < \underbrace{\mathrm{E}(y_i|r_i = 1)P(r_i = 1) + [1 - P(r_i = 1)]}_{\text{if } \mathrm{E}(y_i|r_i = 0) = 1} = p_{\max}$$

   Example continued: Using the concrete numbers and (3) we get

$$p_{\min} = 0.36 < p < 0.36 + 0.34 = 0.7 = p_{\max}$$

This results in two sources of uncertainty for estimating p:

- Uncertainty induced by the missing data through parameters which cannot be identified from the observed data
- Statistical uncertainty (variance) from the estimation procedure

This idea has been applied to more complex models (missing response and/or covariate data) e.g. by

- Horowitz and Manski (2000)
- Horowitz and Manski (2001)
- Vansteelandt and Goetghebeur (2001)
- Manski (2003)
- Heumann(2003), Habilitation, Chapter 5

Random sample from a normal distribution with missing data

- $y \sim N(0,1)$

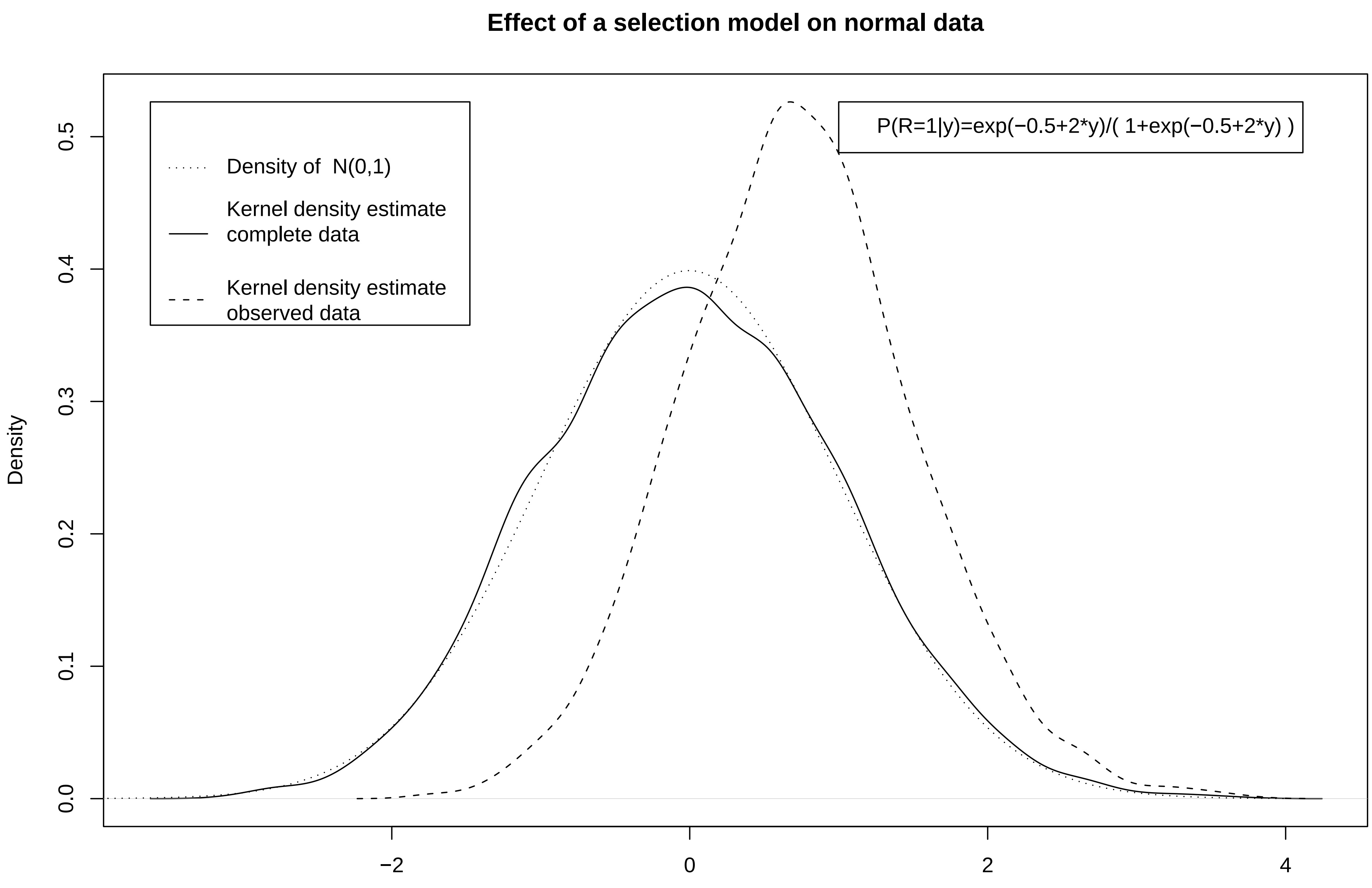
- Missing data process is parameterised with a logistic regression model:

$$\begin{aligned} \log\left(P(r_i = 1|y_i)/P(r_i = 0|y_i)\right) &= \beta_0 + \beta_1 y_i \ , \qquad \beta_0, \beta_1 \in \mathcal{R} \\ P(r_i = 1|y_i) &= \frac{\exp(\beta_0 + \beta_1 y_i)}{1 + \exp(\beta_0 + \beta_1 y_i)} \end{aligned}$$

- The situation is a variant of the sample selection model (Heckman, 1976), where we use the logit link instead of the probit link

- If the model is correctly specified (assumption of a normal distribution is correct and the missing data process is correctly specified by the logistic model) $\longrightarrow$ Maximum Likelihood estimation is possible

- Example: $\beta_0 = -0.5$, $\beta_1 = 2.0$


Effect of a selection model on normal data
Density of N(0,1)
Kernel density estimate complete data
Kernel density estimate observed data
P(R=1|y)=exp(−0.5+2*y)/(1+exp(−0.5+2*y))
Density
0.0
0.1
0.2
0.3
0.4
0.5
−2
0
2
4

Asymmetric treatment of missing data in regression models

- Missing response data or missing covariate data or both?

- Makes a big difference! Why?

- A regression model only specifies $f(y|x;\theta)$ while the marginal distribution of the covariates is unspecified

- One possibility for MNAR response : provide a model for the missing data process $P(r_y|y,x;\psi)$ and use the selection model

$$f(y, r_y|x;\theta,\psi) = P(r_y|y,x;\xi)f(y|x;\theta)$$

  This has been used e.g. by Verbeke and Molenberghs (2000) for linear mixed models (LMMs)

- If covariates x are MNAR then estimating a regression model conditional on x is out-of-the-box possible if we use only the complete cases (CC analysis). One possible method is to model the joint distribution of y and x instead of the conditional distribution of y given x:

$$f(y, x, r_x|\theta, \psi, \xi) = P(r_x|y, x; \xi) f(y|x; \theta) f(x|\psi)$$

This has been used by Ibrahim, Lipsitz and Chen (1999) for Generalised Linear Models (GLMs)

- An interesting special case is if $P(r_x|y, x; \xi) = P(r_x|x; \xi)$. Then

$$f(y|x, r_x) = f(y|x) \tag{4}$$

  - A complete case analyses (CC) which indeed models $f(y|x, r_x = 1)$ gives a *consistent* estimate for θ.

Characterising the missing mechanism as introduced by Rubin (1976), Little and Rubin (1987) in the context of likelihood estimation

- Simplification: No distinction between response and covariates
- Split data y into the two parts $y = (y_{obs}, y_{mis})$
- Likelihood $f(y|\theta)$
- Missing mechanism
$$P(r|y;\xi) = P(r|y_{obs}, y_{mis};\xi)$$
- Assumption: $\theta \in \Theta$, $\xi \in \Xi \longrightarrow (\theta, \xi) \in \Theta \times \Xi$. θ and ξ are said to be distinct.
- The expression
$$f(r, y|\theta, \xi) = f(y_{obs}, y_{mis}|\theta)P(r|y_{obs}, y_{mis};\xi)$$
is called likelihood of the complete data (or *complete data likelihood*)

- The expression

$$f(r, y_{obs}|\theta, \xi) = \int f(y_{obs}, y_{mis}|\theta) P(r|y_{obs}, y_{mis}; \xi) dy_{mis}$$

is called likelihood of the observed data (or *observed data likelihood*)

- The missing mechanism is called missing at random (MAR), if

$$P(r|y_{obs}, y_{mis}; \xi) = P(r|y_{obs}; \xi)$$

does not depend on y_{mis}.

- Then:

$$\begin{aligned} f(r, y_{obs}|\theta, \xi) &= \int f(y_{obs}, y_{mis}|\theta) P(r|y_{obs}; \xi) dy_{mis} \\ &= f(y_{obs}|\theta) P(r|y_{obs}; \xi) \end{aligned}$$

If we are only interested in inference about the parameter θ and under the assumption that θ and ξ are distinct, inference can then be based on $f(y_{obs}|\theta)$ alone and the mechanism $P(r|y_{obs}; \xi)$ can be ignored. The mechanism is then called ignorable.

Extension to Bayesian inference as introduced by Schafer (1997)

- Assumption of independent priors on θ und ξ:

$$\pi(\theta, \xi) = \pi(\theta)\pi(\xi)$$

- Posterior distribution:

$$\pi(\theta, \xi | y_{obs}, r) \propto f(y_{obs}, r | \theta, \xi)\pi(\theta)\pi(\xi)$$

If MAR holds:

$$\pi(\theta, \xi | y_{obs}, r) \quad \propto \quad f(y_{obs} | \theta) P(r | y_{obs}; \xi)\pi(\theta)\pi(\xi)$$

It follows:

$$\pi(\theta | y_{obs}) \propto f(y_{obs} | \theta)\pi(\theta)$$

- Often $f(y_{obs}|\theta)$ is complicated compared to $f(y_{obs}, y_{mis}|\theta)$. Solution through Monte Carlo techniques, e.g. data augmentation (Tanner, 1991). For $s = 1, \ldots, S$:
  - Imputation step (I–step): draw from the conditional predictive distribution

$$y_{mis}^{(s)} \sim f(y_{mis}|y_{obs}, \theta^{(s)})$$

  - Probability step (P–step)

$$\theta^{(s+1)} \sim \pi(\theta|y_{obs}, y_{mis}^{(s)}) \propto f(y_{obs}, y_{mis}^{(s)}|\theta)\pi(\theta)$$

- If S is big enough, the sequences $\{\theta^{(s)}\}$ und $\{y_{mis}^{(s)}\}$ (after some burn-in) are draws from the distribution $\pi(\theta|y_{obs})$ and the unconditional predictive distribution $f(y_{mis}|y_{obs}) \longrightarrow$ proper imputations

Nonignorable nonresponse

- $P(r|y_{obs}, y_{mis}; \xi)$

- Then

$$f(r, y_{obs}|\theta, \xi) \quad = \quad \int f(y_{obs}, y_{mis}|\theta) P(r|y_{obs}, y_{mis}; \xi) dy_{mis}$$

can not be factored in one part which depends on θ and another part which depends on ξ. Inference about θ can not ignore the missing data mechanism.

Bayesian inference in generalised linear models with random effects (GLMM) and nonignorable nonresponse

- A non Bayesian approach using Monte–Carlo EM has been introduced by Ibrahim and Lipsitz (2001), but in detail only for the normal model

- Application in general for dependent outcomes:
  - Longitudinal data (Panel data)
  - Multilevel models: childs in a class, classes in a school, schools in school district, reading competition
  - Spatial and space-time models, additive models, e.g. Fahrmeir, Kneib and Lang (2003), Kamman and Wand (2003)

- Definition of a GLMM, Stiratelli, Laird and Ware (1984), Breslow and Clayton (1993), Fahrmeir and Tutz (2001)
  - $i = 1, \ldots, N$ individuals or units
  - At each individual i we observe n_i measurements: response y_{ij} and a vector of covariates, which is transformed to design vectors x_{ij} $(1 \times p)$ and z_{ij} $(1 \times q)$.

– Distributional assumption: The distribution of y_{ij} comes from an exponential family

$$f(y_{ij}|\theta_{ij},\phi) = \exp\{[y_{ij}\theta_{ij} - b(\theta_{ij})]/a(\phi) + c(y_{ij},\phi)\}$$

– Structural assumption:

$$\mu_{ij} = \mathrm{E}(Y_{ij}|\theta_{ij},\phi) = b'(\theta_{ij}) = h(\eta_{ij}) \ ,$$

or

$$g(\mu_{ij}) = \eta_{ij}$$

where

$$\eta_{ij} = x'_{ij}\beta + z'_{ij}b_i$$

A priori

$$b_i \overset{\text{iid}}{\sim} N(0,D)$$

We call the $(p \times 1)$ vector β fixed effects and the $(q \times 1)$ vector b_i the individual specific random effects
Canonical link:

$$\theta_{ij} = \eta_{ij} = x'_{ij}\beta + z'_{ij}b_i$$

Example: logit link for binary data

$$\theta_{ij} = \eta_{ij} = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$$

- Therefore

$$f(y_{ij}|x_{ij}, z_{ij}; b_i, \beta, \phi)$$

models the conditional distribution of y_{ij} given the random effects b_i

- Likelihood under the assumption of conditional independence: y_{ij} and y_{ik}, $j \neq k$ are conditionally independent given the random effects b_i (additionally, independence between individuals i is assumed)

$$L(\beta, b_1, \ldots, b_N, \phi|y) = \prod_{i=1}^{N}\left\{\prod_{j=1}^{n_i} f(y_{ij}|x_{ij}, z_{ij}; b_i, \beta, \phi)\right\}$$

- Bayesian inference, posterior distrubution

$$p(\beta, b_1, \ldots, b_N, \phi, D|y) \propto \underbrace{L(\beta, b_1, \ldots, b_N, \phi|y)}_{\text{Likelihood}} \underbrace{p(\beta)p(\phi)\left\{\prod_{i=1}^{N} p(b_i|D)\right\} p(D)}_{\text{Prior}}$$

- In the following: $\phi = 1$, dependence on x and z is suppressed, flat prior for β:

$$p(\beta, b_1, \ldots, b_N, D|y) \propto L(\beta, b_1, \ldots, b_N, |y) \left\{ \prod_{i=1}^{N} p(b_i|D) \right\} p(D)$$

- Choices for the prior $p(D)$:
  - Wishart distribution
  - a priori independent random effect components: product of q Gamma distributions
  - Log-normal distribution

$$\begin{aligned} b_{il}|\alpha_l &\sim N(0, \exp(\alpha_l)) \qquad l = 1, \ldots, q \\ \alpha_l &\sim N(0, a_l) \qquad l = 1, \ldots, q, \qquad a_l \text{ fixed constant }. \end{aligned}$$

- Posterior distribution with log-normal prior

$$p(\beta, b_1, \ldots, b_N, D|Y) \propto L(\beta, b_1, \ldots, b_N|Y) \left\{ \prod_{i=1}^{N} p(b_i|\alpha) \right\} p(\alpha|a) \tag{5}$$

with $\alpha = (\alpha_1, \ldots, \alpha_q)$, $a = (a_1, \ldots, a_q)$.

- Example: model with random intercept: $z_{it} = 1$, $q = 1$, α, a. Prior:

$$
\begin{aligned}
&\left\{\prod_{i=1}^{N} p(b_i|\alpha)\right\} p(\alpha|a) = \\
&= \left\{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\exp(\alpha)}} \exp\left[-\frac{1}{2}\frac{b_i^2}{\exp(\alpha)}\right]\right\} \frac{1}{\sqrt{2\pi a}} \exp\left(-\frac{1}{2}\frac{\alpha^2}{a}\right) \\
&= (2\pi\exp(\alpha))^{-\frac{N}{2}} \left\{\prod_{i=1}^{N} \exp\left[-\frac{1}{2}\frac{b_i^2}{\exp(\alpha)}\right]\right\} (2\pi a)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{\alpha^2}{a}\right) \qquad (6)
\end{aligned}
$$

Nonignorable missing: selection model

- Focus on one subject i
- Density given b_i in the case of complete data

$$f(y_i|\beta, b_i, D) = \prod_{j=1}^{n_i} f(y_{ij}|\beta, b_i) \ ,$$

- Selection model:

$$f(r_i|y_i, \gamma)$$

$r_i = (r_{i1}, \ldots, r_{i,n_i})$

- Density of y_i, r_i and b_i:

$$f(y_i, r_i, b_i|\beta, \gamma, D) = f(y_i|\beta, b_i)p(b_i|D)f(r_i|y_i, \gamma) \ .$$

Partition y_i in an observed and a missing part

$$y_i = (y_{i,o}, y_{i,m}) \ ,$$

where $y_{i,o(bs)}$ and $y_{i,m(is)}$ have dimensions n_i^o und n_i^m with $n_i^o + n_i^m = n_i$

- Concrete pattern is not mentioned in the notation. Example: Let $n_i = 3$. No distinction between the patterns $r = (0,1,0)$ and $r = (1,0,0)$. In this case $n_i^o = 1$ and $n_i^m = 2$.

- With partitioned y_i:

$$
\begin{aligned}
& f(y_{i,o}, y_{i,m}, r_i, b_i | \beta, \gamma, D) \\
= & \underbrace{\left\{ \prod_{j_o=1}^{n_i^o} f(y_{ij_o} | \beta, b_i) \right\}}_{\text{lik. contr. of obs. data}} \underbrace{\left\{ \prod_{j_m=1}^{n_i^m} f(y_{ij_m} | \beta, b_i) \right\}}_{\text{lik. contr. of missing data}} \times \\
& \underbrace{p(b_i|D)}_{\text{random effect}} \quad \underbrace{f(r_i | y_{i,o}, y_{i,m}, \gamma)}_{\text{missing model}} \ .
\end{aligned}
$$

- Conditional predictive distribution of the missing data, y_{im}, given the observed data and the parameters is proportional to the joint density:

$$
f(y_{i,m} | y_{i,o}, r_i, b_i; \beta, \gamma, D)
$$

$$
\begin{aligned}
&\propto \left\{ \prod_{j_o=1}^{n_i^o} f(y_{ij_o}|\beta, b_i) \right\} \left\{ \prod_{j_m=1}^{n_i^m} f(y_{ij_m}|\beta, b_i) \right\} p(b_i|D) f(r_i|y_{i,o}, y_{i,m}, \gamma) \\
&\propto \left\{ \prod_{j_m=1}^{n_i^m} f(y_{ij_m}|\beta, b_i) \right\} f(r_i|y_{i,o}, y_{i,m}, \gamma)
\end{aligned}
$$

- Imputation step: draw from

$$
f(y_{i,m}|y_{i,o}, r_i, b_i; \beta, \gamma, D) \propto \underbrace{\left\{ \prod_{j_m=1}^{n_i^m} f(y_{ij_m}|\beta, b_i) \right\}}_{\text{lik. contrib. of missing data}} \underbrace{f(r_i|y_{i,o}, y_{i,m}, \gamma)}_{\text{missing model}}
$$

Algorithm

Repeat for $s = 1, \ldots, S$:

- Imputation step (I-step): replace the missing values by drawing from the conditional predictive distribution for all $i = 1, \ldots, N$

$$y_{im}^{(s)} \sim f(y_{i,m} | y_{i,o}, r_i, b_i^{(s)}; \beta^{(s)}, \gamma^{(s)}, D^{(s)})$$

- Probability step (P-step): Given the filled in and now complete data draw new parameters from the posterior distribution

$$(\beta, b_1, \ldots, b_N, D, \gamma)^{(s+1)} \sim p(\beta, b_1, \ldots, b_N, D, \gamma | y_o, y_m^{(s)})$$

with

$$\begin{aligned} & p(\beta, b, D, \gamma | y_o, y_m, r) \\ \propto \quad & L(\beta, \gamma, b(D) | y_o, y_m, r) \left\{ \prod_{i=1}^{N} p(b_i | D) \right\} p(D) p(\beta) p(\gamma) \ , \end{aligned}$$

where

$$L(\beta, \gamma, b(D)|y_o, y_m, r) = \prod_{i=1}^{N} f(y_{i,o}, y_{i,m}|\beta, b_i) f(r_i|y_{i,o}, y_{i,m}, \gamma)$$

is the likelihood of the completed data.

Drawing from the posterior distribution

- Duane, Kennedy, Pendleton and Roweth (1987), Neal (1993): Hybrid Monte Carlo (HMC) algorithm
- Metropolis algorithm
- Uses the gradient of the log-posterior distribution
- Simultaneous update of all parameters, including the random effects (contrary to Gibbs sampling or single site Metropolis)
- One additional auxiliary variable for each parameter
- Advantage: suppresses random walk behaviour of usual Metropolis algorithms and is therefore more efficient
- Performance in simulation studies was good in general, but problems occur if the covariates are scaled extremely different (standardisation can help)

Application (not really): Longitudinal study, Ohio children data

- Analysed e.g. by Zeger, Liang and Albert (1988) by GEE, and by Fahrmeir and Tutz (2001) as GLMM with random intercept.

- $N = 537$ childs were examined at the ages of 7, 8, 9, and 10 years ($n_i = \text{const} = 4$) whether the suffer from a respiratory infection ($y_{ij} = 1$) or not ($y_{ij} = 0$), $j = 1, 2, 3, 4$.

- Primary interest was in the effect of the covariate $x_{ij}^{smoking}$: "smoking behaviour of the mother" ($1 =$ Mother smokes, $-1 =$ Mother doesn't smoke), which is not time varying: $x_{ij}^{smoking} = x_i^{smoke}$

- Generate missing data according to the model

$$\text{logit} P(r_{ij} = 1 | y_{ij}, x_i^{smoking}) = \gamma_0 + \gamma_1 y_{ij} + \gamma_2 x_i^{smoking}$$

- I.e.: $\gamma_0 = 1$, $\gamma_1 = -1$, $\gamma_2 = -1$, that is the probability of observing a response is highest if the mother doesn't smoke and the child has no respiratory infection.

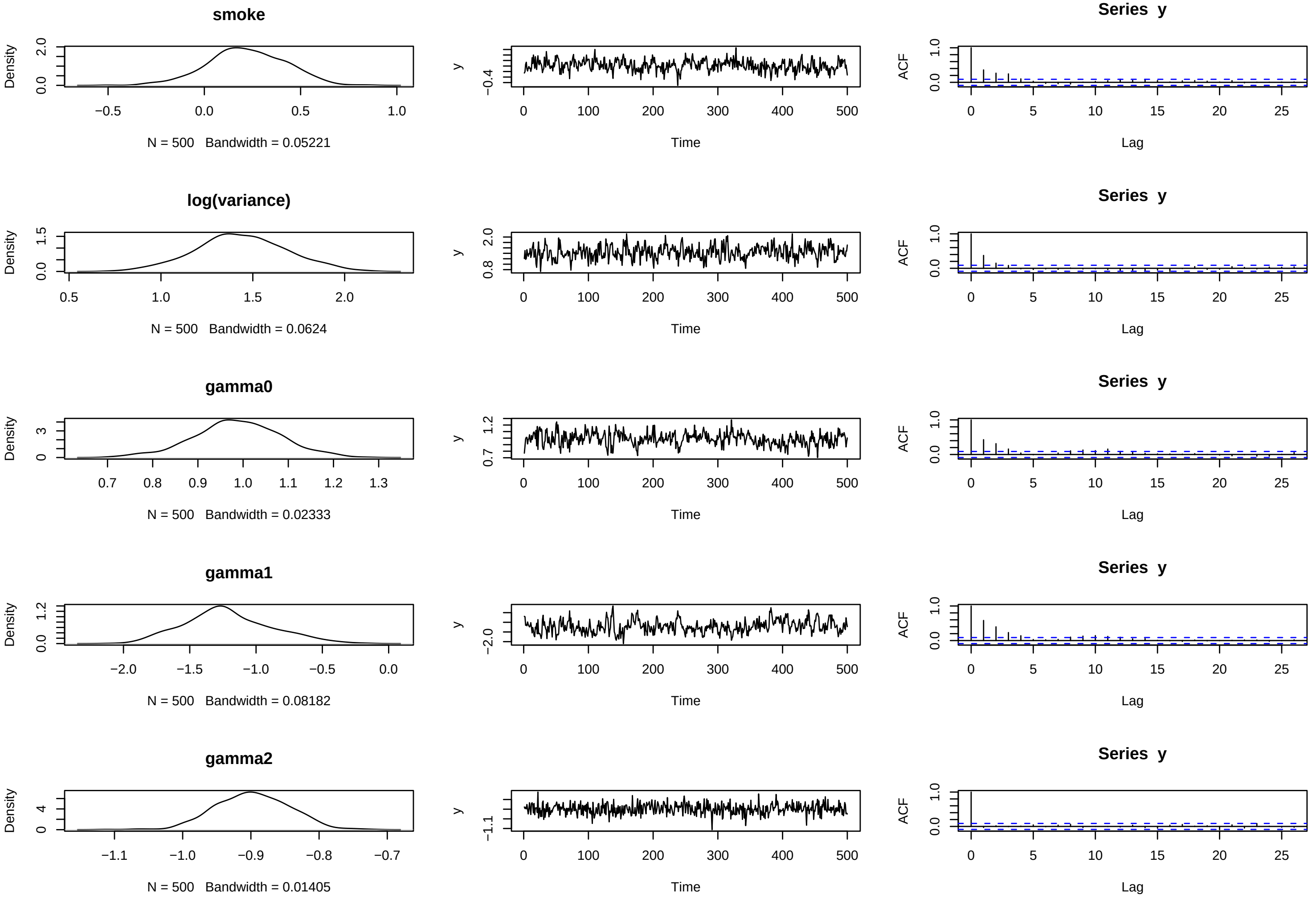
- Results for one run: From 2148 observations, 610 (28%) are missing

Par.	Gauss–Hermite* $(m = 10)$	HMC	HMC with missing Response
$\beta^{rauchen}$	0.19 (0.11)	0.19 (0.14)	0.20 (0.21)
σ_b	2.14 (0.20)	2.19 (0.18)	2.11 (0.25)
γ_0	—	—	0.97 (0.10)
γ_1	—	—	-0.93 (0.35)
γ_2	—	—	-0.92 (0.05)

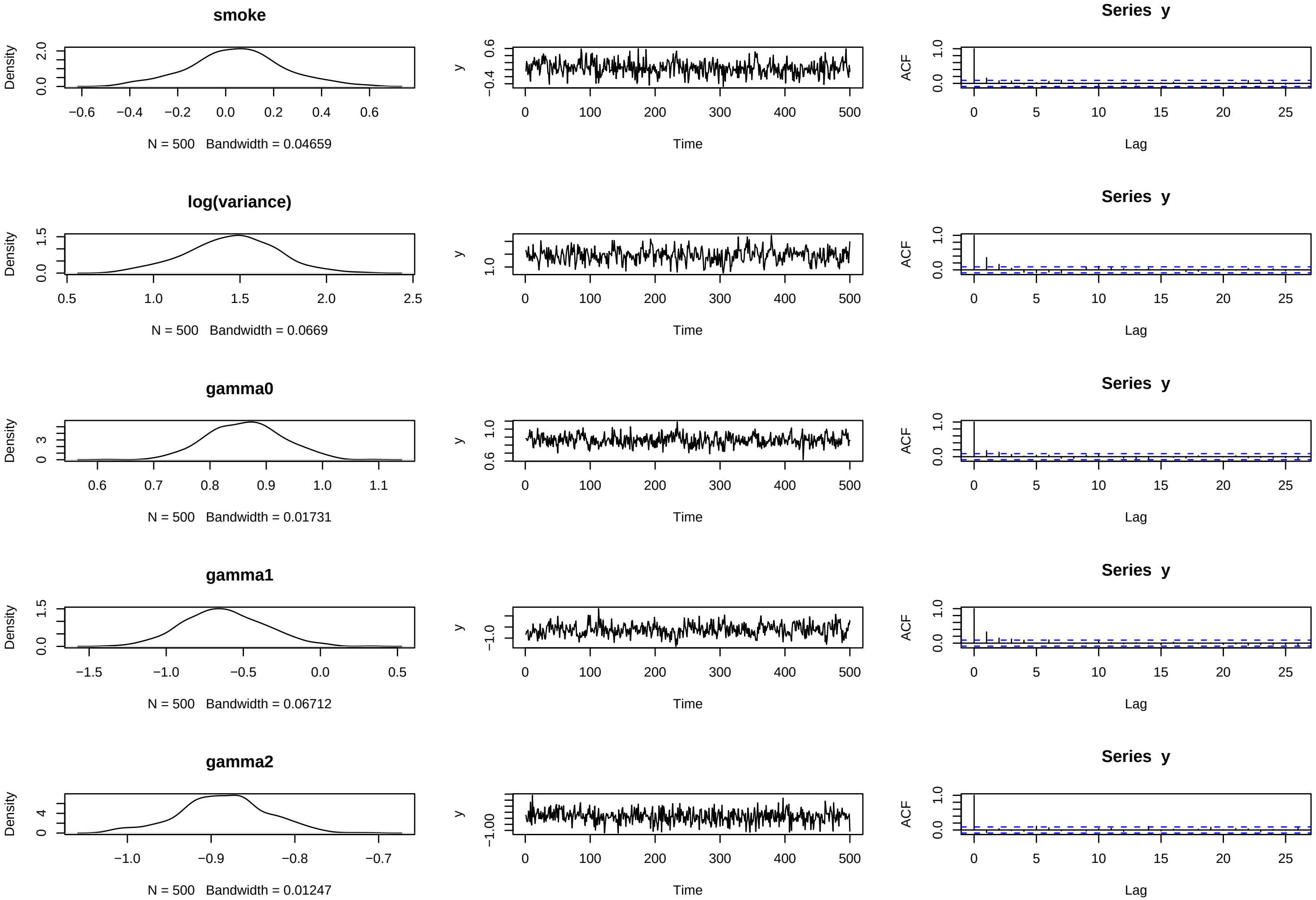
* Source: Fahrmeir and Tutz (2001), Chapter 7

Two other runs with informative priors on γ_1 and γ_2

- Simulation 1: $p(\gamma_1) = N(0,1)$, $p(\gamma_2) = N(0,1)$



- Simulation 2: $p(\gamma_1) = N(0, 0.1)$, $p(\gamma_2) = N(0, 5)$



Parameter estimates in the two runs

Simulation 1:

```
Smoke:            0,2193 (0,2009)
log(variance):    1,4383 (0,2457)
gamma0:           0,9871 (0,0955)
gamma1:          -1,2129 (0,3277)
gamma2:          -0,8976 (0,0541)
```

Simulation 2:

```
Smoke:            0,0455 (0,1958)
log(variance):    1,4531 (0,2574)
gamma0:           0,8629 (0,0666)
gamma1:          -0,623  (0,2626)
gamma2:          -0,8853 (0,0522)
```

The estimate for $\beta(Smoke)$ is shrunken to 0 using a prior for γ_1 which is more concentrated around zero (supports the MAR assumption)

Some remarks

- In the model

$$\text{logit}P(r_{ij} = 1 | y_{ij}, x_i^{smoking}) = \gamma_0 + \gamma_1 y_{ij} + \gamma_2 x_i^{smoking}$$

  is implicitly assumed, that missing does not depend on neither whether missing has occured (or not) at other time points nor on the response at other time points.

- This type of models is called outcome dependent missing models

- In general the full joint distribution of the missing indicators has to be modeled, e.g. by a sequence of univariate conditional distributions in the context of longitudinal data

- Special attention has to be given to different data situations:

  - Intermittent missing or only drop out
  - Equidistant time points or unequally spaced time points
  - Clustered data

- Models of the type

$$\mathrm{logit} P(r_{ij} = 1 | y_{ij}, x_i^{smoking}) = \gamma_0 + \gamma_1 \mathrm{E}(y_{ij}) + \gamma_2 x_i^{smoking}$$

would also be possible. The assumption of distinctness is violated.

Shared parameter models

- Shared parameter models are another example for models where the assumption of distinctness (independence of the priors of the data model and the missing model) is violated

- Shared parameter models have been proposed e.g. by Have, Kunselman, Pulkstenis and J.R. (1998), but not in a Bayesian version

- Example:

  Data model:

  $$Y_{ij}|b_{0i}, b_{1i} \sim N(\beta_0 + b_{i0} + (\beta_1 + b_{i1})t_{ij}, \sigma^2)$$

  where t_{ij} are the times of measurement.

  Missing model:

  $$\operatorname{logit} P(r_{ij} = 1 | x_{ij}, z_{ij}, \gamma, b_{i0}, b_{i1}) = \gamma_0 + \gamma_1^{\frac{1}{2}} b_{i1} + x'_{ij}\gamma$$

- Interpretation in the example: probability that the response is observed is (ceterus paribus) higher for individuals with a high individual random slope if $\gamma_1 > 0$.

The missing data problem in a wider context

- Causal inference with (and without?) counterfactual outcomes, potential outcomes
- Heterogeneous treatment effects, e.g. to control the efficiency of employments incentives.
- Randomised clinical studies: drop-out plus non-compliance

People working on such topics in the econometric community include e.g.: Angrist, Heckman, Imbens, Vytlacil

Literatur

Breslow, N. E. and Clayton, D. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**: 9–25.

Duane, S., Kennedy, A., Pendleton, B. J. and Roweth, D. (1987). Hybrid monte carlo, *Physics Letters B* **195**(2): 216–222.

Fahrmeir, L., Kneib, T. and Lang, S. (2003). Penalized additive regression for space-time data: a bayesian perspective, *SFB386 – Discussion Paper, ftp://ftp.stat.uni-muenchen.de/pub/sfb386/paper305.ps.Z 305*.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2 edn, Springer–Verlag, New York.

Have, T., Kunselman, A., Pulkstenis, E. and J.R., L. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out, *Biometrics* **54**: 367–383.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of Economic and Social Measurement* **5**: 475–492.

Horowitz, J. L. and Manski, C. F. (2000). Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data, *Journal of the American Statistical Association* **95**(449): 77–88.

Horowitz, J. and Manski, C. (2001). Imprecise identifcation from incomplete data, *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*.

Ibrahim, J. G., Lipsitz, S. R. and Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable, *Journal of the Royal Statistical Society, Series B* **61**(1): 173–190.

Ibrahim, Joseph G. abd Chen, M.-H. and Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable, *Biometrika* **88**(2): 551–564.

Kamman, E. E. and Wand, M. P. (2003). Geoadditive models, *Journal of the Royal Statistical Society C (Appl. Stat.)* **to appear**.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*, Springer, New York.

Neal, R. M. (1993). Probabilistic inference using markov chain monte carlo methods, *Technical Report CRG-TR-93-1*, Department of Computer Science, University of Toronto.

Rubin, D. B. (1976). Inference and missing data, *Biometrika* **63**: 581–592.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, Chapman and Hall, London.

Stiratelli, R., Laird, N. M. and Ware, J. H. (1984). Random effects models for serial observations with binary responses, *Biometrics* **40**: 961–971.

Tanner, M. A. (1991). *Tools for Statistical Inference*, Springer–Verlag, New York.

Vansteelandt, S. and Goetghebeur, E. (2001). Analyzing the sensitivity of generalized linear models to incomplete outcomes via the ide algorithm, *Journal of Computational and Graphical Statistics* **10**: 656–672.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*, Springer, New York.

Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating approach, *Biometrics* **44**: 1049–1060.