Efficient Data-augmented MCMC Methods for Binomial Logit Models

Agnes Fussl

In collaboration with
Sylvia Frühwirth-Schnatter and **Rudolf Frühwirth**

ROeS Conference
September 11^{th}, 2013

Outline

Binomial data

- binomial data arise e.g. when repeated measurements for each covariate pattern in the design matrix of a planned experiment or frequencies in the form of two-way/three-way contingency tables are observed
- each observation y_i is modeled as a realization from a binomial distribution with known repetition parameter N_i

Binomial logit regression model

$$y_i|\pi_i \sim \mathrm{BiNom}\left(N_i, \pi_i\right), \quad \log \frac{\pi_i}{1-\pi_i} = \log \lambda_i = \mathbf{x}_i \boldsymbol{\beta} \qquad (1)$$

- $y_i|\pi_i \sim \mathrm{BiNom}\left(N_i, \pi_i\right)$ is considered as the marginal distribution of an augmented model involving latent variables

Example

Caesarean birth data (Fahrmeir and Tutz, 2001)

yi	Ni	intercept	planned	riskfactors	antibiotics
1	18	1	0	1	1
11	98	1	1	1	1
0	2	1	0	0	1
28	58	1	0	1	0
23	26	1	1	1	0
8	40	1	0	0	0
0	9	1	1	0	0

y_i . . . number of women with infection
N_i . . . number of observed women in each group

$\sum N_i = 251$ observations $\Rightarrow$ reduced to $N = 7$ covariate patterns

Individual RUM representation

- consider each observation y_i as the aggregated number of successes of N_i independent binary experiments with outcomes $z_{1i}, \dots, z_{N_i,i}$ (Frühwirth-Schnatter and Frühwirth, 2007)
- z_{ni} follows a binary logit model with the same log odds ratio as in (1), i.e. $\Pr(z_{ni} = 1|\pi_i) = \pi_i$
- the binary outcomes $z_{1i}, \dots, z_{N_i,i}$ can be reconstructed easily from the binomial observation y_i
- to perform bayesian inference with data augmentation the binary logit model can be rewritten as ***random utility model*** (RUM) introduced by McFadden (1974) or as ***difference RUM*** (dRUM)
- for both representations the binary logit model results as marginal distribution of z_{ni}

Individual RUM representation

- for each binary observation z_{ni} we introduce the utilities $u_{0,ni}$ and $u_{1,ni}$ of choosing category 0 or 1 as latent variables:

Individual RUM version

$$u_{0,ni} = \epsilon_{0,ni}, \qquad \epsilon_{0,ni} \sim \mathcal{EV}, \tag{2}$$
$$u_{1,ni} = \log \lambda_i + \epsilon_{1,ni}, \qquad \epsilon_{1,ni} \sim \mathcal{EV} \tag{3}$$
$$z_{ni} = I\{u_{1,ni} > u_{0,ni}\},$$

- where $i = 1, \dots, N$ and $n = 1, \dots, N_i$ with independent extreme value distributed errors $\epsilon_{0,ni}, \epsilon_{1,ni}$
- ***disadvantage:*** very high-dimensional latent variable

Aggregated representations

- introduce a ***single aggregated latent*** $y_i^\star$ for each binomial observation y_i instead of the whole sequence $u_{1,1i}, \ldots, u_{1,N_i,i}$
- desirable properties of an aggregated representation:
  (1) latent equation should take the form of a regression-type model
  $$y_i^\star = \log \lambda_i + \varepsilon_i$$
  (2) error ε_i in the model has a distribution with a pdf that is known explicitly
  (3) it should be easy to simulate from the conditional distribution of $y_i^\star | y_i, \lambda_i$
- Frühwirth-Schnatter et al. (2009): **aggregated RUM representation** $\rightarrow$ the aggregation step is only applied to equation (3), modeling the utility of choosing 1

A new aggregated representation - Step 1

- the aggregation step is also applied to equation (2), modeling the utility of choosing 0
- we **aggregate the individual utilities** for each category:
$$\exp(-y_{0i}^\star) = \sum_{n=1}^{N_i} \exp(-u_{0,ni}), \quad \exp(-y_{1i}^\star) = \sum_{n=1}^{N_i} \exp(-u_{1,ni})$$
- as they are both sums of independent exponential random variables, the aggregated utilities are independent apriori, each following a Gamma distribution:
$$\exp(-y_{0i}^\star) \sim \mathcal{G}(N_i, 1), \quad \exp(-y_{1i}^\star)|\lambda_i \sim \mathcal{G}(N_i, \lambda_i) \qquad (4)$$

A new aggregated representation - Step 2

- **taking the negative logarithm** for both expressions in (4), we obtain latent equations in form of a regression-type model:

$$y_{0i}^\star = \epsilon_{0i}, \tag{5}$$
$$y_{1i}^\star = \log \lambda_i + \epsilon_{1i}, \tag{6}$$

where $\epsilon_{ki} = -\log \xi_{ki}$ with $\xi_{ki} \sim \mathcal{G}(N_i, 1)$ follows the negative log-Gamma distribution with shape parameter N_i for $k = 0, 1$

- the **single aggregated latent variable** is then defined as

$$y_i^\star = y_{1i}^\star - y_{0i}^\star$$

Aggregated dRUM representation

Aggregated dRUM version

$$y_i^\star = \log \lambda_i + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{LG}(N_i), \tag{7}$$

- where $\varepsilon_i = \epsilon_{1i} - \epsilon_{0i}$ and $\mathcal{LG}(\alpha)$ is the Type III generalized logistic distribution with parameter α (see Balakrishnan, 1992)
- the first two central moments are given by:

$$\mathrm{E}(\varepsilon_i|N_i) = 0, \qquad \mathrm{V}(\varepsilon_i|N_i) = 2\psi'(N_i)$$

Data-augmented independence MH

- the error ε_i in (7) is approximated by the normal distribution $\mathcal{N}(0, 2\psi'(N_i))$
- the resulting posterior of β is used as proposal
- the proposal density is independent of the previous draw β^{old}, but depends on the latent variable $\mathbf{z} = (y_1^\star, \ldots, y_N^\star)$
- high acceptance rate as the distribution of ε_i is approximately normal $\mathcal{N}(0, 2N_i)$ for large N_i

Auxiliary mixture sampler

- error ε_i in (7) is approximated by a scale mixture of normal distributions, all component means equal to 0:

$$f_{\mathcal{LG}}(\alpha) = \frac{\Gamma(2\alpha)e^{-\alpha\varepsilon}}{\Gamma(\alpha)^2(1+e^{-\varepsilon})^{2\alpha}} \approx q_\alpha = \sum_{r=1}^{R(\alpha)} w_r(\alpha)\,\varphi(0, s_r^2(\alpha)),$$

  where $\varphi(0, s^2)$ denotes a normal density with mean 0 and variance s^2

- the number of components $R(\alpha)$, the weights $w_r(\alpha)$ and the variances $s_r^2(\alpha)$ depend on $\alpha = N_i$

Hybrid auxiliary mixture (HAM) sampling

- a new sampler combining both data-augmented MH and auxiliary mixture sampling
- in cases where the ratio y_i/N_i is neither close to 0 nor close to 1, the normal approximation in the MH algorithm will give a contribution α_i to the acceptance rate $\alpha = \prod_{i=1}^{t} \alpha_i$ close to 1
- for extreme ratios $y_i/N_i \leq c_{low}$ and $y_i/N_i \geq c_{up}$ (e.g. $c_{low} = 0.05$, $c_{up} = 0.95$) α_i will be considerably smaller
- **idea:** use the mixture approximation only for extreme ratios of y_i/N_i and the normal approximation of the MH sampler otherwise

Application

- application of the different MCMC sampler to the Caesarean birth data and a simulated data set to compare the different approaches
- ***MCMC details:***
  - independent standard normal prior for each regression coefficient
  - 10000 draws from the posterior distribution after a burn-in of 2000 iterations
- ***'quality criteria':*** runtime T, effective sampling size ESS, effective sampling rate ESR, acceptance rate α

Caesarean Birth Data $N = 7$, $\min N_i = 2$, $\max N_i = 98$, $\sum N_i = 251$, $d = 4$				
Sampler	α (%)	**T** (sec)	**med ESS** (total draws)	**med ESR** (draws/sec)
Agg. dRUM-MH	97.0	3.5	1390.1	**393.8**
Agg. dRUM-AM		4.1	1465.0	356.5
Agg. dRUM-HAM	99.3	4.8	1453.7	302.2
Agg. RUM-MH	86.1	3.0	598.5	202.2
Indiv. dRUM-AM		7.9	1761.5	223.0
Indiv. dRUM-MH	68.1	5.1	1388.0	271.6

- HAM sampler: $c_{low} = 0.01$ and $c_{up} = 0.99$

Simulated Data $N = 490$, $\min N_i = 1$, $\max N_i = 126$, $\sum N_i = 25803$, $d = 10$				
Sampler	α (%)	**T** (sec)	**med ESS** (total draws)	**med ESR** (draws/sec)
Agg. dRUM-MH	97.2	11.4	947.3	**83.2**
Agg. dRUM-AM		13.0	974.4	75.1
Agg. dRUM-HAM	98.6	15.0	888.2	59.2
Agg. RUM-MH	82.2	9.5	409.8	43.2
Indiv. dRUM-AM		504.2	1141.8	2.3
Indiv. dRUM-MH	50.3	264.8	704.4	2.7

- HAM sampler: $c_{low} = 0.15$ and $c_{up} = 0.85$

Conclusions

- the aggregation step yields a considerable reduction of computing time compared to the individual dRUM
- the modifications in the aggregated dRUM lead to a remarkable gain in efficiency
- **data-augmented MH sampler is first choice**, but we propose to use the AM or HAM sampler for more complex models, where the acceptance rates might be smaller

R package `binomlogit` (Fussl, 2013)

- includes the data-augmented **independence MH sampler**, the **auxiliary mixture sampler** and the **HAM sampler** based on the dRUM representation of the **binomial logit model**
- plus a routine for the MH-sampler **for binary data** is enclosed
- all four functions simulate the posterior distribution of the regression coefficients of the logit model and return the MCMC draws
- for **technical details** concerning the algorithms see the paper by **Fussl, Frühwirth-Schnatter and Frühwirth (2013)**

References

Balakrishnan, N. (ed.) (1992). *Handbook of the Logistic Distribution*, New York: Marcel Dekker.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*, 2nd Ed. Springer Series in Statistics, Berlin: Springer.

Frühwirth-Schnatter, S. and R. Frühwirth (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis 51*, 3509-3528.

Frühwirth-Schnatter, S. and R. Frühwirth (2010). Data augmentation and MCMC for binary and multinomial logit models. In Kneib, T. and Tutz, G. (Eds.): *Statistical Modelling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir*, pp. 111-132, Heidelberg: Physica-Verlag.

Frühwirth-Schnatter, S. , R. Frühwirth, L. Held and H. Rue (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing 19*, 479-492.

Fussl, A. (2013). *binomlogit: Efficient MCMC for Binomial Logit Models*. R package version 1.1, http://CRAN.R-project.org/package=binomlogit

Fussl, A., S. Frühwirth-Schnatter and R. Frühwirth (2013). Efficient MCMC for Binomial Logit Models. *ACM Transactions on Modeling and Computer Simulation 23*, Number 1, Article 3, 21 Pages.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka, P. (Ed.): *Frontiers of Econometrics*, pp. 105-142, New York: Academic.

Scott, S.L. (2011). Data augmentation, frequentist estimation, and the Bayesian analysis of multinomial logit models. *Statistical Papers 52*, 87-109.