A Case Study in Comparing Bayes Estimated Fixed Effects versus Frequentist Random Effects

Eric V. Slud, Univ. of Maryland, Math. Dept.
& U.S. Census Bureau,
Center for Stat. Research & Methodology

March 2019, UMD Stat Seminar

Joint work with Yves Thibaudeau

Disclaimer

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are the author's and not necessarily the Census Bureau's.

Overview

1. CPS Data on 1-month Labor Force Transitions

2. Generalized Logistic Model (3 outcome) with Random State Effects

3. Fixed effects, Random Effects, & Posterior Fixed Effects

4. Model Comparisons (AIC) & Predictive Metrics (ROC curves)

5. R Packages and Bayesian MCMC

6. Conclusions

Current Population Survey – Gross Labor Flows

Large survey (75,000 HU's per month) focused on Labor force (age 16+)

Longitudinal rotating panel design: in 4 months, out 8, in 4 (*issue of matching persons in HUs across months*)

`Labor` categories: 1=Employed, 2=Unemployed, 3=NILF

Other variables: geography (`ST`, 16 states NE, SW, South), `Educ` (College or no), `Age` (4 groups: 65+, 55-64, 35-54, 34-),

Data Structure and Problem

Y_6, Y_7 3-category labor status `Lab` (Emp, Unemp, NILF)
June and July 2017 (18K matched cases only)

$\texttt{Age}_6$, $\texttt{Educ}_6$ (= college indicator), `ST` (16 states)

Objective: estimate/predict # in small cross-classified cells, e.g. (NILF_6, EMP_7), 35-54, No-college, NewMexico

Large **X** cells obtained using survey weights: $\hat{N}_{\mathbf{x}} = \sum_i w_i I_{[\mathbf{X}_i = \mathbf{x}]}$

Model cross-classified conditional probabilities for Y_7 given predictors **X** (including Y_6)

Model – Generalized Logistic

Let y values $1 = \texttt{Emp}$, $2 = \texttt{Unemp}$, $3 = \texttt{NILF}$, and for $y = 1, 2$ define

$$P(Y_{7,i} = y \,|\, \mathbf{X}_i = \mathbf{x}) \;=\; \exp(\mathbf{x}'\beta^{(y)}) \,/\, \Big(1 + \sum_{z=1}^{2} \exp(\mathbf{x}'\beta^{(z)})\Big)$$

$\Rightarrow$ separate logistic regressions **Emp vs NILF**, **Unemp vs NILF**

Models to compare wrt covariate sets:

F0: $\texttt{Lab}_6\texttt{+Age}_6\texttt{+Educ}_6$, **F1: F0**+ST, **F2: F1**+$\texttt{Lab}_6\texttt{:Age}_6$

$2 \cdot (1 + 2 + 3 + 1)$ $\quad 14 + 2 \cdot 15$ $\quad 44 + 2 \cdot (3\dot{2})$ $\quad df$

with # coef's = 14 for **F0** , 44 for **F1**, 56 for **F2**

Fixed-Effect Model Comparisons

Comparisons for separate logistic models `Emp07 vs. NILF07` and `Unemp07 vs. NILF07` based on CPS June-July 2017 dataset.

		Emp07 vs NILF07				Unemp07 vs NILF07			
		df	Dev	ResDf	ResDev	df	Dev	ResDf	ResDev
	NULL			327	18857			333	1912
	F0	6	18365	321	492	6	1530	327	382
F1	ST	15	35	306	457	15	18	312	364
F2	Lab:Age	6	75	300	381	6	35	306	329
	Lab:ST	30	31	270	350	30	39	276	290

adequacy questionable OK

Random-Effect Models

Can instead assume $\beta^{(y)} = (\beta_{ST}^{(y)}, \gamma^{(y)})$ in models **(F1)**, **(F2)**, with **random** ST coefficients of State dummy indicator that label s_i of i'th subject is equal to j

$$\beta_{ST} = (u_j^{(y)},\ j = 2, \ldots, 16,\ y = 1, 2)$$

e.g., $\beta_{ST} \sim \mathcal{N}(0, V)$ for V of prescribed form, e.g. V diagonal with 1st and 2nd sets of 15 entries $\equiv\ \sigma^{(y)2},\ \ y = 1, 2$

Then estimate $\theta = (\gamma^{(y)},\ \sigma^{(y)2},\ \ y = 1, 2)$ by Max. likelihood.

Packages to do this ML estimation discussed in later slide

Random-Effect Models, continued

Suppose $\mathbf{X}_i$ includes ST dummy components; for states j, denote $\beta^{(y)}_{ST,j} = u^{(y)}_j$ for $y = 1, 2$, and $\underline{u} \equiv \beta_{ST}$, $\underline{u}_j = \{u^{(y)}_j\}^2_{y=1}$.

For $\mathbf{X}_i = \mathbf{x} = (\xi, s)$ predictor values within state s, have random

$$P(Y_i = y \,|\, \mathbf{X}_i = \mathbf{x}, \underline{u}) = \pi(y, \mathbf{x}, \underline{u}, \gamma) \equiv e^{\xi'\gamma^{(y)} + u^{(y)}_s} \Big/ \Big(1 + \sum_{z=1}^{2} e^{\xi'\gamma^{(z)} + u^{(z)}_s}\Big)$$

and Best Linear Unbiased Predictor (for fixed parameter values)

of unconditional $P(Y = y \,|\, \mathbf{X} = \mathbf{x})$ is $= E_\theta\Big(\pi(y, \mathbf{x}, \underline{u}, \gamma) \,\Big|\, \{Y_i, \mathbf{X}_i\}_i\Big)$

This becomes the Empirical Bayes predictor when $\widehat{\theta}$ replaces θ.

Bayesian Point of View

Within fixed-effects model **(F1)** *or* **(F2)** *adopt prior either flat or with fictitious observations (*n_y *in* `Lab`$_7$ *status* $y = 1, 2, 3$ *with covariates* **X** *replaced by average covariate).*

Posterior distributiions for all coefficients jointly obtained by MCMC: posterior distributions of state effects are like those of other parameters, but enter predictions by being averaged within $\pi(y, \mathbf{x}, \underline{u}, \gamma)$

Using State-Effect Coefficients for Prediction

Fixed-effect case: $\widehat{N}_{(\mathbf{x},y)} = \sum_i w_i \, I_{[\mathbf{X}_i=\mathbf{x}]} \, \widehat{P}(Y_i = y \,|\, \mathbf{X}_i = \mathbf{x},\, \beta) \Big|_{\beta=\widehat{\beta}}$

Random-effect case: with $\mathbf{x} = (\xi, s)$,

$$\widehat{N}_{(\mathbf{x},y)} = \sum_i w_i \, I_{[\mathbf{X}_i=\mathbf{x}]} \, E\Big(\pi(y, \mathbf{x}, \underline{u}, \gamma) \,\Big|\, \{Y_i, X_i\}_i\Big) \Big|_{\theta=\widehat{\theta}}$$

Bayes prediction: like fixed-effect case but $\widehat{P}(Y_i = y \,|\, \mathbf{X}_i = \mathbf{x},\, \widehat{\beta})$ replaced by posterior-averaged $\pi(y, \mathbf{x}, \underline{u},\, \gamma)$

Alternate Interpretations of State Effects

(1) Fixed Effects – unknown constants estimated directlly

$$\widehat{\beta} \;=\; (\widehat{\gamma}^{(y)}, \, \underline{\widehat{u}}^{(y)}, \; y = 1, 2) \sim \mathcal{N}(\, \beta, \, (I(\beta))^{-1})$$

Fixed *state effects* $u_j^{(y)}$ (coefficients for $I_{[s_i=j]}, \;\; j = 1, \ldots, 16$)

(2) Random state effects $u_j^{(y)} \sim \mathcal{N}(0, \, \sigma^{(y)2})$

model random variation around mean state effect $\mu^{(y)}, \; y = 1, 2$

(3) Posterior-distributed $\underline{u}$ effects based on simple prior or simple variants)

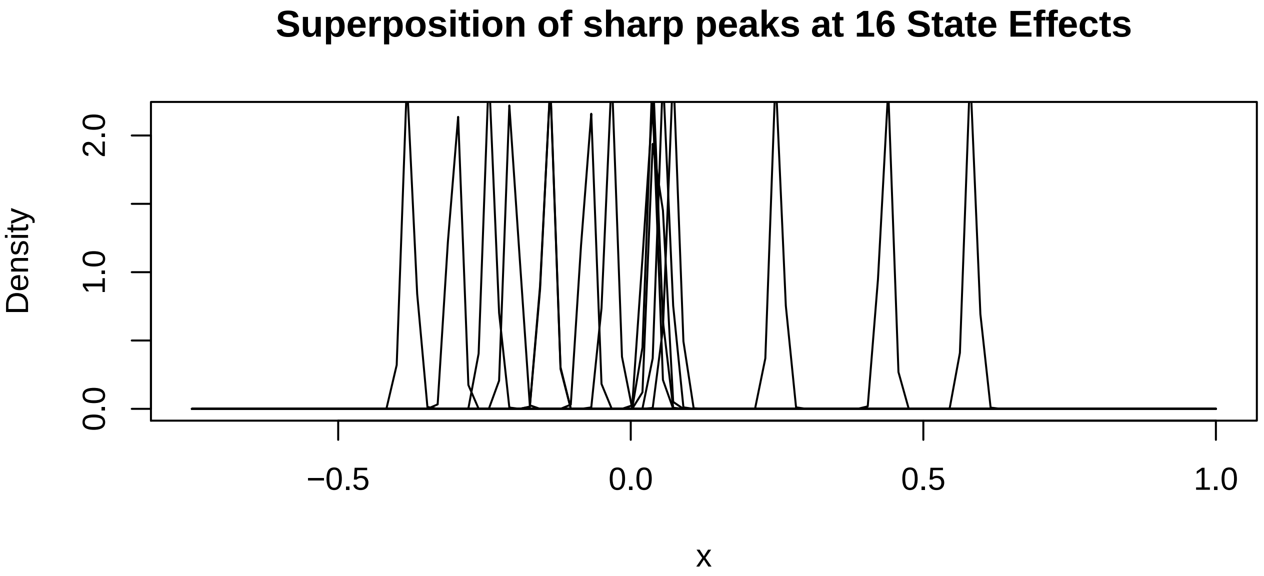
Summary of Results Related to State Effects

(a) Correlations across different ML state effects small (almost all < 0.05), but between same-state effects for Emp_7 and Unemp_7 outcomes around 1/3. Same holds for Bayes posteriors.

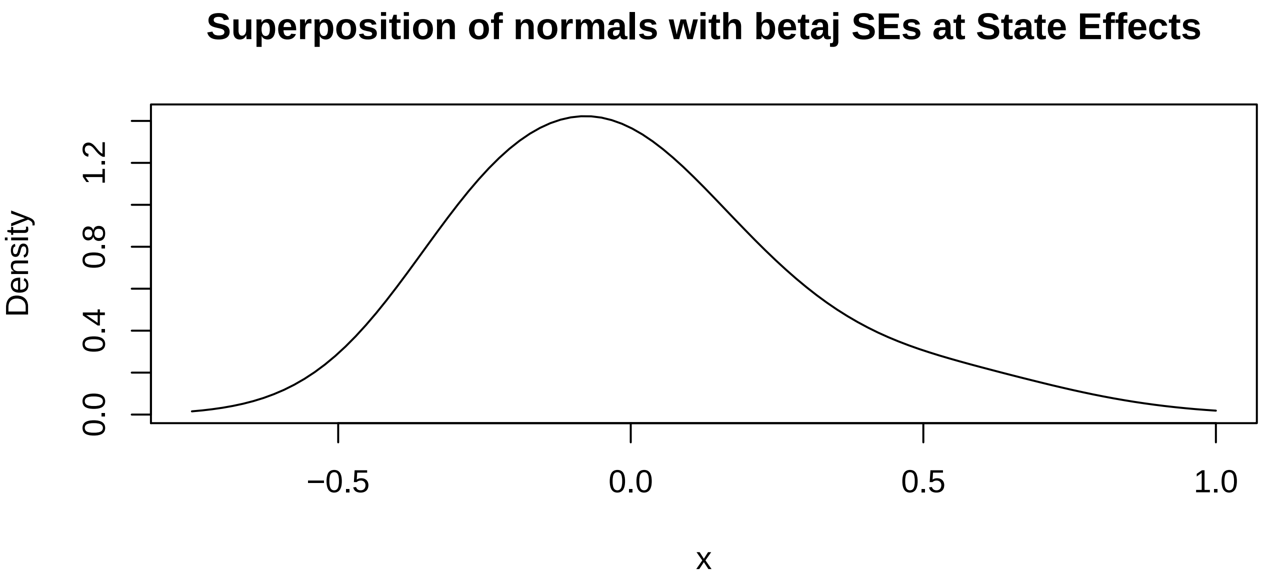
(b) Models with all indep. state effects did not fit materially worse than the fixed-effect models (several criteria ... prediction criteria can be applied to Bayes fits as well).

(c) MLEs of state fixed intercept effects (minus average effect) from -0.38 to 0.58 for Emp_7 outcome, and from -0.72 to 0.43 for Unemp_7.

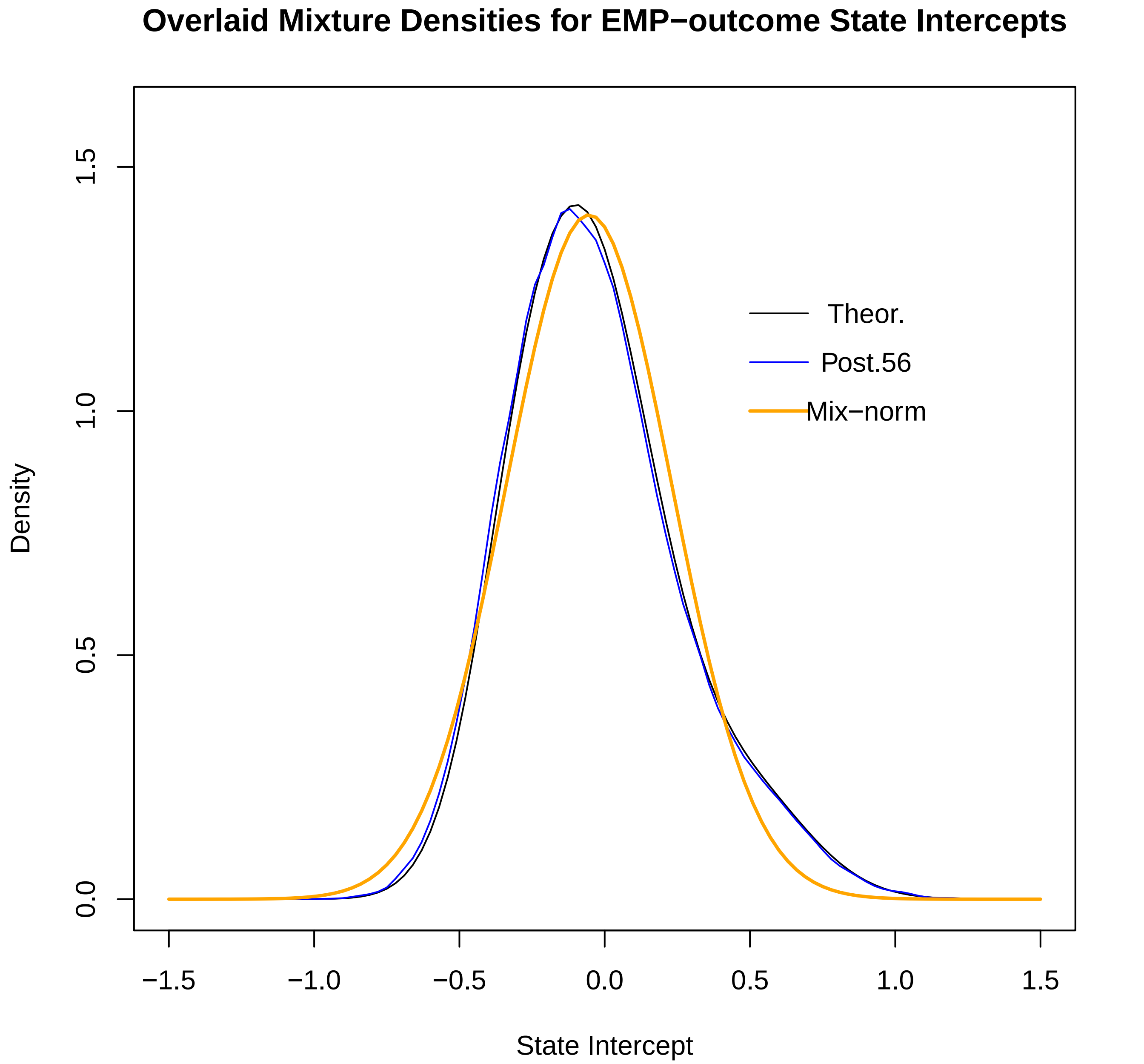
(d) In separate random-intercept logistic-regression model fits for $\texttt{Emp}_7$ vs. $\texttt{NILF}_7$, found $\hat{\sigma} = \mathbf{0.164}$ with $SE = \mathbf{0.060}$, and for $\texttt{Unemp}_7$ vs. $\texttt{NILF}_7$ found $\hat{\sigma} = .033$ with $SE = 0.28$.



Superposition of sharp peaks at 16 State Effects
Density
0.0
1.0
2.0
−0.5
0.0
0.5
1.0
x



Superposition of normals with betaj SEs at State Effects
Density
0.0
0.4
0.8
1.2
−0.5
0.0
0.5
1.0
x



Overlaid Mixture Densities for EMP−outcome State Intercepts
Density
State Intercept
0.0
0.5
1.0
1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Theor.
Post.56
Mix−norm

Bernstein-von Mises Theorem & Implications

B-vM Theorem says under large-sample regularity conditions for MLE theory, for general continuous broadly supported priors, the posterior density given data of the parameter θ within a well-specified parametric model is approximately distributed as

$$\sqrt{n}\,(\theta - \widehat{\theta}) \sim \mathcal{N}\Big(0,\, (\mathrm{I}(\widehat{\theta}))^{-1}\Big)$$

Two implications:

(1) can check applicability of asymptotic ML theory by checking whether posterior distributions for parameters look normal, and

(2) Metropolis-Hastings algorithm computations become easy and accurate.

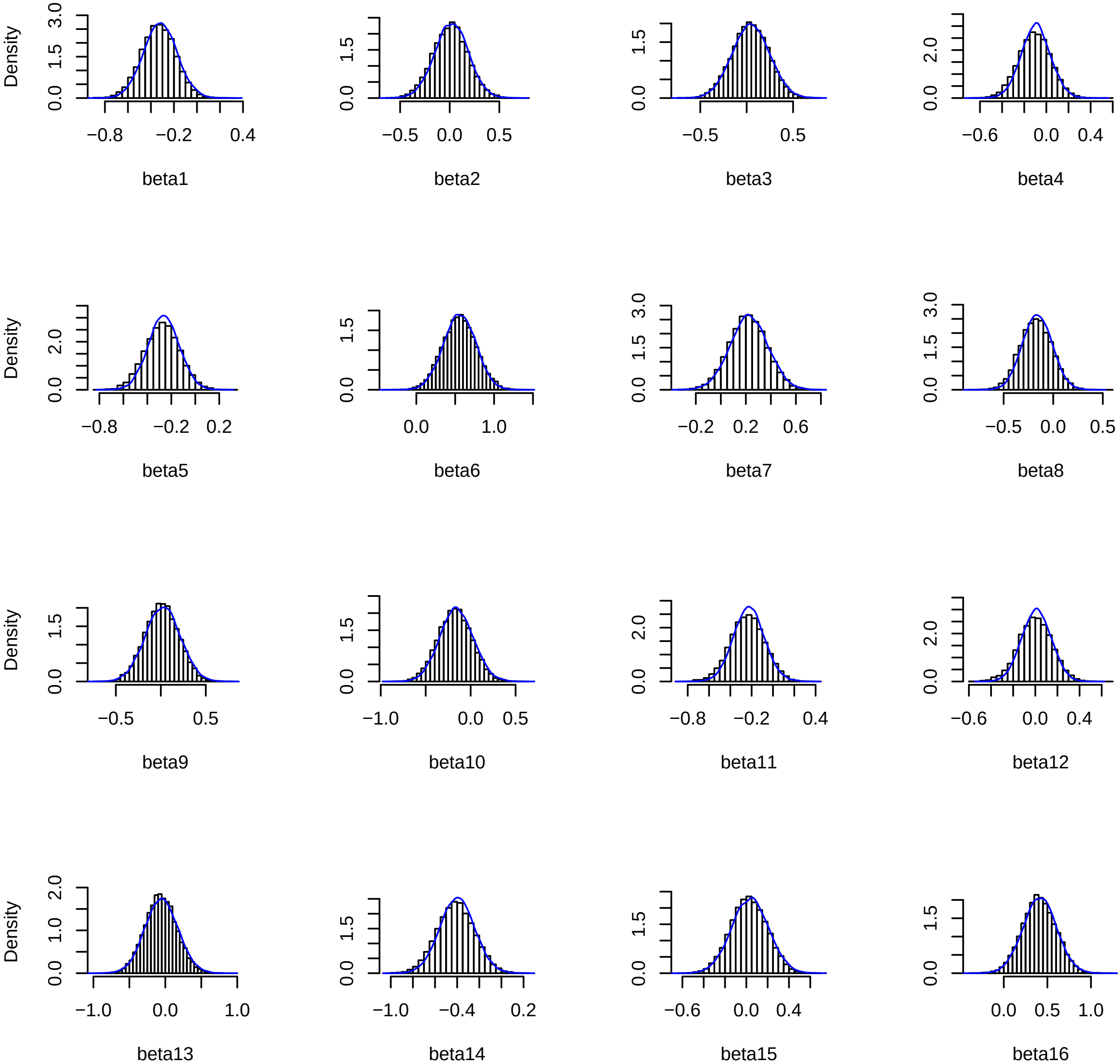
Technique for MCMC in Large Samples

Problem: to simulate Markov Chain $\theta^{(t)}$ with density $g(\theta) = \pi(\theta|\underline{\mathbf{y}}) \propto \pi(\theta)\ L(\underline{\mathbf{y}}, \theta)$ as equilibrium distribution

Algorithm: at each step, simulate independent $\tau_t \sim \mathcal{N}(\widehat{\theta},\ (\hat{\mathrm{I}}(\widehat{\theta}))^{-1})$ and $U_t \sim \text{Unif}(0,1)$, and define for $t \geq 1$:

$$\theta^{(t)} = \theta^{(t-1)} + (\tau_t - \theta^{(t-1)})\, I\Big[U_t \leq \frac{g(\tau_t)\, q(\theta^{(t-1)})}{g(\theta^{(t-1)})\, q(\tau_t)}\Big]$$

That is, a very good proposal density $q \approx g$ is the B-vM asymptotic normal ! *Acceptance ratio* $= Pr(bracket)$ high even for high-dimensional θ.


Density
beta1
beta2
beta3
beta4
beta5
beta6
beta7
beta8
beta9
beta10
beta11
beta12
beta13
beta14
beta15
beta16

Software Packages for Analysis

- `SAS PROC GLIMMIX`, `NLINMIX` logistic, generalized logistic models & random effects (Stroup 2013), adaptive Gaussian quadrature
- generalized logistic fixed effects in `R` packages `nnet`, `mlogit`
- `glmmML`, `GLMMadaptive`, `lme4` (function `glmer`): accurate logLik & MLEs approximation, logistic regression with 1 random effect
- other `R` packages do more general models with approx. logLik
- Bayes packages `BRugs`, `rjags` in `R`, or `Stan` outside

Some Conclusions

Very different analyses of somewhat different models (fixed vs mixed) give different interpretations but similar conclusions

Can use Bayes to check status of large-sample theory predictions and random-effect model specifications

'Hybrid' survey and model-based conclusions still of great research interest at the statistical agencies

References

Hand, D. and Till, R. (2001), *Machine Learning* journal
(multi-outcome ROCs)

Pfeffermann, Skinner, & Humphreys (1998), *JRSS A*

Rao and Molina (2015), *Small Area Estimation*, 2nd ed.

Stroup, W. (2013), *Generalized Linear Mixed Models*, CRC

Thibaudeau, Slud & Gottschalck (2017), *Ann. Appl. Stat.*

Thank you !

evs@math.umd.edu
Eric.V.Slud@census.gov