Computer Applications for Small Area Estimation, Part 2

Jerzy Wieczorek
Small Area Estimation Research Group, CSRM
1/31/2013

Question from last time

- Is normality required for (non-Bayesian) Fay-Herriot model? **No**. This would be enough:
- $y_i = Y_i + e_i$, $Y_i = X_i^T \beta + u_i$,
- $E(e_i) = E(u_i) = 0$, $V(e_i) = \sigma_i^2$, $V(u_i) = \sigma_M^2$
- **Without** normality, we can estimate BLUP (best linear unbiased predictor) and get moment-based variance estimates. **With** normality, BLUP is also BP, and we can use MLE or REML.

Outline

- Bayesian statistics refresher
- Intro to PROC MCMC and WinBUGS
- Bayesian area-level (Fay-Herriot) model
- Examples in SAS and in R + WinBUGS
- Further resources

Bayes refresher

- "All classical inference statements … are probability statements about x given θ, phrased so as to appear to be probability statements about θ."
  —Anthony O'Hagan
- Example: given $x_1, ..., x_n$ random sample with unknown mean μ, we use $\bar{x}|\mu \sim N(\mu, \sigma^2/n)$ to infer plausible values of μ

Bayes refresher

- Bayesians talk directly about posterior distribution of $\mu|\bar{x}$. (At Census, we usually want to report the mean, standard error, and CI endpoints of $\mu|\bar{x}$.)
- This requires specifying a prior distribution for μ. For many problems, can find a noninformative prior that gives similar inferences as classical/frequentist approach.

Bayes: exact vs. Monte Carlo

- (a) Is posterior $p(\theta|x)$ a standard distribution (Normal, Beta, etc.) we can get exactly? Then we know mean, stderr, and CI endpoints exactly.
- (b) Is posterior NOT a standard distribution? We draw a big sample from posterior, then summarize this sample: mean of posterior distribution is approx. the sample mean of the draws, the CI is 5% & 95% quantiles, etc.

Bayes: types of Monte Carlo

- (b1) Posterior is a non-standard distribution but we can draw samples directly? Then we do so.
- (b2) Posterior is a non-standard distribution and we can't draw samples directly, but we can evaluate the posterior distribution function? Use Markov Chain Monte Carlo (MCMC) methods to draw samples indirectly.

Bayes: MCMC overview (simpler!!!)

- You have an initial parameter value $\theta^{(t)}$. Propose a new parameter value $\tilde{\theta}^{(t+1)}$.
- Sometimes accept the proposed new value as the next draw in your MCMC sample: $\theta^{(t+1)} = \tilde{\theta}^{(t+1)}$ …and sometimes reject: $\theta^{(t+1)} = \theta^{(t)}$
- There's lots of math behind what's meant by "propose" and "sometimes", but SAS or WinBUGS will take care of this for you.

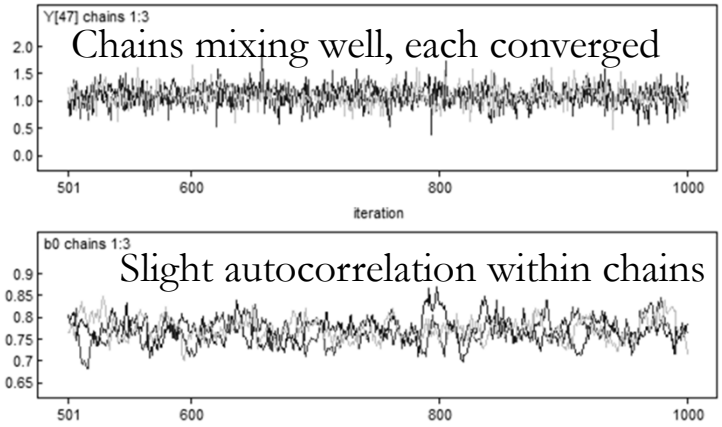
Bayes: MCMC overview: what is "sometimes"?

- Pick an initial parameter value $\theta^{(t)}$ and evaluate the posterior $p^{(t)}$. Propose a new parameter value $\tilde{\theta}^{(t+1)}$ and evaluate posterior there too: $p^{(t+1)}$
- If $p^{(t+1)} > p^{(t)}$, accept new parameters as a draw in your sample: $\theta^{(t+1)} = \tilde{\theta}^{(t+1)}$
- If $p^{(t+1)} < p^{(t)}$, make a randomized decision:
  - with prob. $1 - (p^{(t+1)}/p^{(t)})$, 'reject': $\theta^{(t+1)} = \theta^{(t)}$
  - with prob. $p^{(t+1)}/p^{(t)}$, 'accept': $\theta^{(t+1)} = \tilde{\theta}^{(t+1)}$

Bayes: MCMC overview

- End with a chain of values: $\theta^{(t)}, \theta^{(t+1)}, \theta^{(t+2)}, \ldots$
- Mathematically, can prove that this chain is a sample from the posterior distribution
- In practice, need to check "convergence": does the chain behave like a simple random sample from the high-probability parts of the posterior? Or is it stuck in a low-probability area, or autocorrelated, etc.?

Bayes: MCMC overview



Checking MCMC "convergence"

- Run multiple chains from different starts; see if they mix well and if each looks like white noise
- Use Gelman's $\hat{R}$ criterion: convergence isn't good unless it's close to 1.0 (i.e. below 1.1)
- Check if MC error is small relative to post.stdev
- If convergence is slow, try reparameterizing your model — may be able to reduce correlation among parameters

Model checking

- Same as for non-Bayes models: check the raking factors, compare to a "truth deck," etc.
- Sensitivity check on priors: re-run with different priors and see if your estimates change much
- Posterior predictive checks: if you draw new data using the posterior distribution, does it "look like" your original data?

~~Complications~~

- Your data are not normal as given, but are approximately normal on a transformed scale?
  $\log(y_i) \sim N(X_i^T \beta, \sigma_i^2 + \sigma_M^2)$
  No problem with Bayes! Just back-transform your MCMC samples to original scale before taking means, quantiles, etc.

PROC MCMC and WinBUGS

- Both tools let you just specify the form of the model, and they work out the MCMC details behind the scenes.
- WinBUGS seems faster, more robust, easier to tweak, and better at handling missing values. Also, in SAS you have to worry about whether each variable is in a dataset vs. in an array.
- But SAS is more familiar for many of us…

PROC MCMC

- PARMS statements initialize the parameters; use a separate statement for each block of likely-to-be-correlated parameters
- PRIOR statements define priors on parameters
- MODEL statement shows how your observed data relates to the parameters

R + WinBUGS

- R is an open-source statistical software package.
- WinBUGS is a standalone program for Bayesian statistical modeling.
- You can use WinBUGS on its own, but I prefer to call WinBUGS from within R:
  re-running a script is more reproducible than remembering clicks on a screen.
- On Linux, JAGS replaces WinBUGS.

Area-level (Fay-Herriot) model

- Sampling model: $y_i \sim N(Y_i, \sigma_i^2)$
- Linking model: $Y_i \sim N(X_i^T \beta, \sigma_M^2)$
- We want posterior distribution of each Y_i
- Just add uninformative priors. Ideally:
- $p(\beta) \propto 1$ and $p(\sigma_M) \propto 1$
- In practice:
- $\beta \sim N(0, 1000000)$ and $\sigma_M \sim \text{Unif}(0, 1000)$

Area-level (Fay-Herriot) model

- Alternate parameterization:
- $y_i \sim N(Y_i, \sigma_i^2)$, $Y_i = X_i^T \beta + u_i$, $u_i \sim N(0, \sigma_M^2)$
- Mathematically, this is the same model.
  But practically, the choice of how you specify it in code can affect the MCMC convergence.
- See code examples and Gelman (2006) paper for yet another way to reparameterize.

Further resources

- Gentle, thorough Bayes intro: Kruschke (2011)
- More detailed Bayes intro: Gelman et al. (2003)
- Hierarchical Bayes chapter of Rao (2003)
- PROC MCMC examples: Mukhopadhyay & McDowell (2011), SAS 9.2 user's guide
- Priors for hierarchical variances: Gelman (2006)
- Software: SAS 9.2, R, WinBUGS, JAGS

Bibliography

- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). Bayesian Data Analysis, 2nd ed. New York: Chapman & Hall/CRC.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, vol. 1, no. 3, p. 515-533.
- Kruschke, J. (2011). Doing Bayesian Data Analysis. Academic Press/Elsevier.
- Mukhopadhyay, P.K., and McDowell, A. (2011). Small Area Estimation for Survey Data Analysis Using SAS Software. Proceedings of the SAS Global Forum 2011 Conference. http://support.sas.com/resources/papers/proceedings11/336-2011.pdf
- Rao, J.N.K. (2003). Small Area Estimation. New York: Wiley.
- SAS 9.2 user's guide: MCMC random effects & mixed effects
  http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_mcmc_sect048.htm
  http://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_mcmc_sect006.htm