DISCUSSION OF LOGISTIC MIXED MODEL PAPERS

Ralph E. Folsom, Research Triangle Institute
RTI, P. O. Box 12194, Research Triangle Park, NC 27709

I will begin my discussion with Dr. McCulloch's paper. He does an excellent job of outlining the strengths and weaknesses of alternative estimation methods. I will emphasize some of his points and relate them to the other two papers and my own experience.

The methods he contrasts include:

- MLE and REML
- Penalized Quasi-Likelihood (PQL)
- Generalized Estimating Equations (GEEs), and
- Bayes estimation.

I will not comment on the unadorned GEE approach since it does not estimate random effects for the small areas. The Singh and Wu paper overcomes this weakness of GEE's marginal or unconditional mean model by estimating additive random effects that can be appropriately bounded.

Maximum Likelihood's strengths include the consistency and asymptotic normality of the βs and $\sigma^2 s$. Likelihood ratio tests are also available for comparing nested $x\beta$ models. It's weaknesses include the observation that MLEs are difficult to calculate for all but the simplest two level nested random effect models.

Dr. McCulloch's (1997) JASA paper presents a Metropolis-Hastings type monte-carlo method for computing MLEs of the βs and $\sigma^2 s$ for more complex models. The computation burden for his method is similar to the Gibbs/Metropolis sampling of the Bayes solution.

Another weakness of the MLE solution is that mean squared errors for the small area predicted values that account for the sampling error in the estimated $\sigma^2 s$ (a'la. Prasad and Rao 1990) are not well developed. There are some recent unpublished results by Lahiri and Jiang for simple models. Also, bootstrapping may be required to produce interval estimates with good small sample properties.

Penalized Quasi-Likelihood's main strength is that the computations are relatively easy. It works best when the associated first or second order 'small σ' linearization behaves well; that is, when σ is small and $y_{ij} \sim Bin(n_{ij}, P_{ij})$ has $n_{ij} > 1$ for most-j and the expected number of hits per cluster-i

$$\left(\sum_j n_{ij} P_{ij} \right)$$

is greater than one. It's weaknesses include a tendency to substantially underestimate the $\sigma^2 s$ and to yield biased βs when σ is not small and the expected number of hits per cluster-i is substantially less than one for most-i. The evidence is fairly convincing that PQL cannot be recommended for small area estimation.

Bayes estimation's strength is its ability to estimate the exact (small sample) joint distribution of the βs, $\sigma^2 s$, and the random effects conditional on the data. It therefore, provides estimates of the tail percentiles, say the 2 .5 and 97.5 percentiles, for the exact distribution of the small area estimates.

The Bayes solution's weakness is that the joint posterior can fail to exist when improper uniform priors are used for the variance components as recommended by Zeger and Karim in their (1991) JASA paper. Use of proper priors for the variance components that are too diffuse or too close to improper can also cause failure of convergence or convergence to a posterior that depends on the prior. These numerical problems are likely to be extreme when the likelihood for a σ^2 parameter is relatively flat due to insufficient replication; for example, when not enough of the binomial $n_{ij}s$ are greater than one.

Turning to my discussion of the Ghosh, Natarajan, and Maiti paper, my comments concern the unified hierarchical Bayes models presented by the authors, particulary their binomial version where $y_{ij} \sim Bin(n_{ij}, P_{ij})$ with $logit(P_{ij}) = \theta_{ij}$ and

$$\theta_{ij} = x_{ij}^T \beta + u_i + \epsilon_{ij}$$

is linear in the fixed and random effects with the random small area effects

$$u_i \overset{ind}{\sim} N(O, \sigma_u^2)$$

and the random error terms or demographic domain effects

$$\epsilon_{ij} \overset{ind}{\sim} N(O, \sigma_\epsilon^2)$$

In the author's model, the conditional posteriors for β and the u_i random effects are all normal. The authors Gibbs sampling scheme is therefore much easier than Zeger and Karim's (1991) solution for the logistic mixed model. In the Zeger & Karim model which does not include the random error terms ϵ_{ij}, the β and u_i conditional posteriors are not known exactly and must be sampled via the Metropolis algorithm. For many area probability samples, I believe the author's model will be difficult to fit. For these samples useful $x's$ are often available from the Decennial Census for sample block groups-i with cluster sample sizes n_{i+} that are usually less than 10. With the j subscript typically indexing over 30 age by gender by race/ethnicity domains, the number of replicate observations per domain (n_{ij}) is seldom greater than 1.

While the joint posterior may exist if only one cluster-i by domain cell-j has n_{ij} greater than 1, this case will undoubtedly lead to a posterior for σ^2 that is heavily influenced by the prior for σ^2. When the error terms ϵ_{ij} are removed from the model due to insufficient replication so that $\theta_{ij} = x_{ij}^T \beta + u_i$ has just the small area u_i random effect, then the β and u_i posteriors are no longer normal and one must use the time consuming Metropolis routine.

Turning to Ghosh, Natarajan, and Maiti's model averaging example, I found the idea of incorporating model uncertainty into the posterior variance of small area estimates (SAEs) very appealing. On the other hand, I was uncertain about how to set the prior probabilities for two models. I would be inclined to set $\pi = (1-\pi) = 1/2$. Also, I wondered how one would develop the strong prior belief that the interaction parameter vector $\beta_2^{(M_1)} = 0$ before looking at the current data. This was reflected by making $\beta_2^{(M_1)}$'s prior mean vector zero and its prior covariance matrix η_2^2 small. I think I would need experience with a previous incarnation of the data to form such a strong prior belief.

Turning to the Singh and Wu paper, I thought the authors' formulation of additive random effects for logistic mixed models was quite ingenious. Their work seems to restore the unconditional or marginal mean models of the GEE method to a viable competitor for small area estimation. Singh and Wu show that additive random effects can be rigorously defined as differences between hierarchical conditional means. In this fashion they legitimize an additive error structure motivated by the first and second order 'small σ' partitioning of random effects used in the marginal mean version of PQL, namely MQL. The authors carefully parameterize the random effect variances so that they add to the correct unconditional variance, namely $\eta_{ij}(1-\eta_{ij})$ where η_{ij} is the unconditional mean of y_{ij}.

To motivate the existence of additive random effects whose ranges are properly constrained, Singh and Wu show that beta random effect variables v_{1i}, v_{2j}, and v_{12ij} can be defined for the salamander data so that their ranges are suitably constrained and the proper mean and variance relationships are preserved. The authors also ensure that by construction the associated covariance matrix for the additive errors $Z(\lambda)\underset{\sim}{v} + \underset{\sim}{e}$, namely,

$$\Sigma = Z(\lambda_1)\Sigma_v Z(\lambda_1)^T + \Sigma_e(\lambda_2)$$

is positive definite. They can then use normal theory MLE or REML solution equations to get consistent estimators of the unknown parameters λ_1 and λ_2 in Σ. The familiar GEE type optimum estimation equations then yield consistent βs.

My main quarrel with the authors regards their mse estimator for the small area predictors. I think they should have acknowledged that their Empirical Bayes (EB) type mean-squared-error estimator may be a serious underestimate since it ignores the sampling error in the λ variance parameters. While there is hope that a Prashad and Rao type correction could be developed for their model, it is interesting to note that the standard mixed model formulation with Z known and Σ_v a function of λ parameters does not apply for the Singh and Wu covariance matrix where Z is a function of unknown parameters. I would also like to see the authors test the small sample properties of their solution on simulated binomial data with additive beta random effects. Finally, given a good mse estimate for the Singh & Wu small area predictor vector, $\hat{\mu} = \hat{\eta} + Z(\lambda_1)\hat{v}$, one will still probably need to bootstrap to be assured of good small sample interval estimates for the elements of $\hat{\mu}$. This may largely offset the relative ease of the author's estimation methods compared to the hierarchical Bayes solution.

In summary, I believe that good interval estimates are critical to small area predictions since SAEs are highly subject to misuse. I have come to believe that given careful attention to the use of proper and not too diffuse priors, the full Bayes solution is currently the method of choice in this regard.

References

McCulloch, C.E. (1997). "Maximum Likelihood Algorithms for Generalized Linear Models." *Journal of the American Statistical Association*, 93, pp. 162-170.

Prasad, N.G.N. and J.N.K. Rao (1990). "The Estimation of Mean Squared Errors of Small-area Estimators." *Journal of the American Statistical Association*, 85, pp. 163-171.

Zeger, S.L. and M.R. Karim (1991). "Generalized Linear Models with Random Effects: A Gibbs Sampling Approach." *Journal of the American Statistical Association*, 86, pp. 79-86.