A Multinomial and a Uniform Unit-Level Model for Small Area Estimation*

Donald Malec[†]

Abstract

 The development and evaluation of a hierarchical multinomial model is presented. By limiting this type of model to include constraints at a higher geographic level, many of the design-based characteristics retained by a multinomial may still be retained while borrowing strength from a higher geographic-level model. An evaluation using data from the National Health Interview Survey (NHIS) to make estimates of access to care for the counties in States that border Mexico is presented. Here, the model uses sample segments as the unit-level and accounts for the differential selection of segments within PSU strata. Covariates available from the U.S. Census Bureau's "American Factfinder" are used in the county-level model. In addition, a brief evaluation of using a uniform model at the unit level, instead of a multinomial model, is made.

Key Words: Nonparametric Bayes

1. Introduction

The following provides a mulitnomial-type model that avoids the need to assume that the design-based small area estimates are asymptotically Normally distributed with estimated covariances assumed to be without error. These two assumptions both require a large sample size within a small area when the very nature of the problem is that the sample sizes are too small. Errors in these assumptions can affect the amount of borrowing from the larger model, as borrowing is both a function of the within small areas variances and the assumed within small area distribution. The constrained multinomial distribution used to model the within small area outcomes is based on an extension of the Bayesian bootstrap model (Rubin, 1981). Using only the Bayesian Bootstrap model will result in estimates and estimated variances of the small areas that are comparable to design-based methods. By modifying the underlying Bayesian bootstrap model by adding a constraint at a higher level level, many of the characteristics of the Bayesian bootstrap distribution are retained while still borrowing strength from a higher geographic-level model. Essentially, the asymptotic assumptions of Normality with fixed variances are replaced by a constrained Bayesian bootstrap model. In addition, the usefulness of specifying uniform models, with unknown boundaries as a substitute for a multinomial is briefly evaluated.

In related work, Malec (2005) proposed using a multinomial model with support based on the observed data but, instead of relying on the more nonparametric Bayesian Bootstrap model, a fully parametric hierarchical model was placed on each parameter of the multinomial. See Malec (2005), also, for further references to other earlier work in this area. The device of using models that assume the population support is the same as the sample support was introduced to survey sampling by Hartley and Rao (1968) who assumed the population is distributed as a multiple hypergeometric distribution with support equal to the unit level sample outcomes. Malec et al. (1999) use this approach to account for the effects of PPS sampling by assuming that the support of the sampled, sample selection probabili-

---

ties is equal to the unavailable support of the population selection probabilities.

Section 3 outlines the multinomial model with hierarchical constraints. Section 4 describes the estimation and algorithms used. Section 5 demonstrates estimation using an NHIS delayed medical care outcome for counties in four States and compares estimates with those using a Normal approximation within a small area. Section 6 contains a brief evaluation of using a uniform models within small areas, compared to using a multinomial model.

2. Background of National Health Interview Survey (NHIS)

The use of the Bayesian bootstrap model is best explained and illustrated through the application to outcomes from an actual survey. Although the method is general, the design-levels and instances in which the Bayesian Bootstrap is applied are design dependent. The 2006-2015 NHIS design and accompanying data will be used for illustration.

2.1 Salient features of the 2006-2015 NHIS

The following provides just enough information to enable the explanation of the model and the implementation of small area estimates. For a detailed description of the design, see Parsons et al. (2014).

The entire U.S. population of the civilian non-institutional population is split into a certainty part, where the first level of sampling begins within county, and a non-certainty part, where the first level of sampling begins at the county-level. The first two levels of sampling within the non-certainty strata are 1) sampling PSUs proportional to population size within strata, where PSUs are defined by a county or a group of contiguous counties and strata are defined by metropolitan status within State and 2) sampling of segments with equal probability within Minority Density Strata (MDS), where segments are defined by clusters of 8 - 16 nearby housing units and Minority Density Strata, are defined by grouping census blocks according to similar minority characteristics. Only the first level of sampling within the certainty part is needed to implement the model. In the certainty part, the Minority Density Strata are defined separately within each certainty county. Segments are defined and sampled in an analogous manner as in the noncertainty part.

2.2 Design and estimation simplifications for this initial work

In the noncertainty part, counties are used as if they were the PSUs, even when an actual NHIS PSU is composed of a group of counties. Segment estimates of totals are assumed fixed - i.e. weights at the segment level are used but the uncertainty due to screening, missing data and large-segment subsampling is ignored.

Notation:

Define:

i: State indicator

j: county indicator

k: minority density stratum indicator

ℓ: segment indicator

Within county sample terms:

Define:

m_{ij}: number of minority density strata in State i, county j.

n_{ijk}: number of sampled segments in state i, county j and minority density stratum k

N_{ijk}: total number of segments in (i, j, k) (will be estimated from NHIS by "up-weighting")

$y_{ijk\ell}$: count of number with characteristic, by age x race x sex (within segment missing value and screening adjustment applied and assumed fixed)

$x_{ijk\ell}$: count of total (also, after missing value and screening adjustment).

Further define:

$y_{ijk\ell}^*$ and $x_{ijk\ell}^*$ one of the observed, sampled pairs, $\ell = 1, \ldots, n_{ijk}$.

3. The Multinomial Model With Hierarchical Constraints

3.1 The Within County Model

Using the sample segment outcomes, $y_{ijk\ell}^*$ and $x_{ijk\ell}^*$, to define the support of the population model (as with the Bayesian Bootstrap) define:

$\theta_{ijk\ell}$: the probability that a segment takes the values of $\underline{y}_{ijk\ell}^*$ and $\underline{x}_{ijk\ell}^*$, $\ell = 1, \ldots, n_{ijk}$

That is, define:

$\theta_{jk\ell^*} = prob((y_{jk\ell}, x_{jk\ell}) = (y_{jk\ell^*}^*, x_{jk\ell^*}^*)))$, $\ell^* = 1, \ldots, n_{ijk}$ and $\ell = 1, \ldots, N_{ijk}$

so that $\sum_{\ell^*=1}^{n_{jk}} \theta_{jk\ell^*} = 1$

The population will be modeled as conditionally multinomial:

Define:

$N_{ijk\ell}$: the total number of segments that take values $\underline{y}_{ijk\ell}^*$ and $\underline{x}_{ijk\ell}^*$

So that, $P(\underline{N}_{ijk}|\underline{\theta}_{ijk}, N_{ijk.}) \propto \prod_{\ell=1}^{n_{ijk}} \theta_{ijk\ell}^{N_{ijk\ell}}$.

Substituting srs with replacement for the actual sequential sample, use:

$P(\underline{n}_{ijk}|\underline{\theta}_{ijk}, n_{ijk.}) \propto \prod_{\ell=1}^{n_{ijk}} \theta_{ijk\ell}^{n_{ijk\ell}}$.

3.2 The Between County Model

The distribution of the $\theta_{ijk\ell}$:

Define, at the county level:

$$
\begin{aligned}
Y_{ij} &= \sum_k N_{ijk.} \sum_{\ell=1}^{n_{ijk}} \theta_{ijk\ell} y_{ijk\ell}^* \\
X_{ij} &= \sum_k N_{ijk.} \sum_{\ell=1}^{n_{ijk}} \theta_{ijk\ell} x_{ijk\ell}^*, \text{ and} \\
R_{ij} &= Y_{ij}/X_{ij}
\end{aligned}
$$

Note that the values: $y_{ijk\ell}^*$'s and the $x_{ijk\ell}^*$'s are fixed. A prior for the $\theta_{ijk\ell}$'s is chosen to be informative for the R_{ij}, via a hierarchical model but similar in uninformativeness regarding the remaining free parameters of the $\theta_{ijk\ell}$'s as the Bayesian bootstrap prior is. Specifically, define $\mu_{ij} = \ln(R_{ij}/(1 - R_{ij}))$.

the between county model for μ_{ijd} is specified as a truncated Normal:

$$
\mu_{ij} \sim N(Z_{ij}\beta|\sigma^2), \ lb_{ij} \leq \mu_{ij} \leq ub_{ij},
$$

where $lb_{ij} = min \frac{\sum_k N_{ijk.} \sum_{\ell=1}^{n_{ijk}} \theta_{ijk\ell} y^*_{ijk\ell}}{\sum_k N_{ijk.} \sum_{\ell=1}^{n_{ijk}} \theta_{ijk\ell} x^*_{ijk\ell}}$ and $ub_{ij} = max \frac{\sum_k N_{ijk.} \sum_{\ell=1}^{n_{ijk}} \theta_{ijk\ell} y^*_{ijk\ell}}{\sum_k N_{ijk.} \sum_{\ell=1}^{n_{ijk}} \theta_{ijk\ell} x^*_{ijk\ell}}$,

where both maxima and minima are over the set of thetas: $\{\{\theta_{ijk\ell}\}_{\ell=1}^{n_{ijk}}\}_{k=1}^{m_{ij}}\}$.
(Note: these limits can be determined, numerically, using the R library "alabama" and routine: "constrOptim.nl").

No further hierarchical model is used for the remaining constrained parameters of the multinomial and the complete, joint prior for the multinomial parameters is specified as:

$$Pr(\underline{\theta}_{ij}) \propto \prod_k \prod_\ell \theta_{ijk\ell}^{-1} \times P(\underline{\mu}_{ij} | \beta ij, \Gamma_{ij}).$$

3.3 The Within county Model for counties not in Sample

The definition for the model within a county is explicitly based on the minority density strata that are defined according to the population demographics within the county in conjunction with the total number of segments in each stratum; neither of which are known for counties that haven't been sampled. In fact, minority density strata are not even constructed unless a county has been sampled. In order to model the within county distribution for an unsampled county, a Bayesian bootstrap approach is applied at the county level. That is, the possible configurations of MDS, their underlying multinomial parameters (apart from μ_{ij}) and the distribution of segments is assumed to consists of only the configurations that have been observed in the counties sampled in the non-certainty part.

Define: $\tilde{n}$ to be the number of non-certainty sampled counties. That is, $\tilde{n} = |\{(i, j) :$ state i and county j are in the non-certainty sample $\}|$. The probability that any particular county in the population of non-certainty counties will be the same as the county model observed for sampled county $(i, j)^*$ is defined as $\pi_{(i,j)^*}$. (i.e., the individual county-level models observed in the sampled counties are the only possible models that can occur across all counties.)

Since the population totals are known for all counties, this additional information can be used to adjust the unsampled county models to reflect a difference in population from the counties in sample. This is accounted for by assuming that the total number of segments in an unsampled county, N_{ij}, given that the county $(ij)^*$ model is it's selected model-type is $N_{ij} = N_{(ij)^*} pop_{ij}/pop_{(ij)^*}$, where pop are the corresponding total county population counts obtained from the census.

3.4 Prior Distributions

- prior distribution of $\underline{\theta}_{jk}$:
  As implied in section 3.2:
  $pr(\underline{\theta}_{jk}|\mu_{ij}) \propto \prod_k \prod_\ell \theta_{jk\ell}^{-1}$
  - an improper prior both because its parameters are out of scope to be a Dirichlet distribution and because it does not account for the conditional constraint.

- prior distribution for parameters of regression model (also section 3.2)
  $pr(\underline{\beta}) \propto constant$
  $\sigma \sim$ half-Cauchy

- prior distribution of non-certainty county-level models (section 3.3)
  $pr(\underline{\pi}) \propto \prod_k \pi_k^{-1}$

4. Posterior Inference and Estimation from the Multinomial Model

Posterior inference is made, numerically, from realizations of the joint posterior distribution following the Gibbs sampler steps:

1. $\underline{\theta}_{ij1}, \ldots, \underline{\theta}_{ijm_{ij}}, \mu_j | \beta, \sigma^2, data$

2. $\beta | \mu_{i,j}, \sigma^2$

3. $\sigma^2 | \mu_j, \beta$

4. $\underline{\pi} | data$

Step 3 in the Gibbs sampler chain is implemented using a Metropolis/Hastings step. Realizations of $\underline{\theta}_{ij1}, \ldots, \underline{\theta}_{ijm_{ij}}, \mu_{ij}$, in the Gibbs step 1 above, will be based on the conditional and marginal distributions listed in the following steps:

Additional MCMC steps:

1.1. $\theta_{i,j1}, \ldots, \theta_{i,jn_{ij}} | \mu_{i,j}, data$

1.2. $\mu_{i,j} | \beta, \sigma^2, data$

The following, briefly details how these steps were implemented:

4.1   Step 1.1: The Conditional Posterior of $\underline{\theta}_{ij}$ given μ_{ij} :

In most situations $n_{ijk\ell} = 1$, especially when $\underline{y}_{ijk\ell}$'s and $\underline{x}_{ijk\ell}$'s are complicated. For simplicity, it will be assumed that $n_{ijk\ell} = 1$ here. If not, a similar approach may be taken by coding the sampling based on Adaptive Rejection Sampling (Gilks and Wild, 1992) or similar univariate methods.

When $n_{ijk\ell} = 1$ use the "hit and run" algorithm in the R "hit and run" library as follows: The "hit and run" algorithm will draw $\theta_{i,j,1,1}, \ldots, \theta_{i,j,1,n_{ij1}}, \ldots, \theta_{i,j,m_{ij},1}, \ldots, \theta_{i,j,m_{ij},n_{ijm_{ji}}})/m_{ij}$ from a multivariate uniform distribution with the following side constraints:

$$\sum_{\ell=1}^{n_{ij1}} \theta_{ij,1,\ell} = 1$$

$$\vdots$$

$$\sum_{\ell=1}^{n_{ijm_{ij}}} \theta_{ij,m_{ij},\ell} = 1$$

$$\mu_{ij} = \frac{\sum_{k=1}^{m_{ij}} \sum_{\ell=1}^{n_{ijm_{ij}}} N_{ijk\ell} \theta_{ijk\ell} y_{ijk\ell}^*}{\sum_{k=1}^{m_{ij}} \sum_{\ell=1}^{n_{ijm_{ij}}} N_{ijk\ell} \theta_{ijk\ell} x_{ijk\ell}^*}$$

One draw will produce the appropriate set of $\underline{\theta}_{ij}$ from each of the county's minority density strata.

4.2 Step 1.2: The Marginal Posterior of μ_{ij}

Since the conditional posterior of $\underline{\theta}_{ij}$ is uniform (when all $n_{ijk\ell} = 1$):

$$f(\mu_{ij}|\beta, data) = \int_{\underline{\theta}_{ij} \in W(\mu_{ij})} f(\underline{\theta}_{ij}, \mu_{ij}|\beta, data) d\underline{\theta}_{ij} \tag{1}$$

$$\propto \int_{\underline{\theta}_{ij} \in W(\mu_{ij})} f(data|\underline{\theta}_{ij}, \mu_{ij}) f(\underline{\theta}_{ij}|\mu_{ij}) f(\mu_{ij}|\beta) d\underline{\theta}_{ij}$$

$$\propto \int_{\underline{\theta}_{ij} \in W(\mu_{ij})} f(\mu_{ij}|\beta) d\underline{\theta}_{ij} = |W(\mu_{ij})| \, f(\mu_{ij}|\beta),$$

where, $W(\mu_{ij})$ is the set of all $\underline{\theta}_{ij}$ that satisfy the side constraints given in section 4.1.

Sampling from the marginal posterior of μ_{ij} is accomplished by, first, numerically approximating $|W(\mu_{ij})|$ as a discrete function of μ_{ij} over its range of values, matching the resulting discrete values of $|W(\mu_{ij})|$ to corresponding discrete values of the truncated normal distribution of μ_{ij} specified in section 3.2 and, finally, drawing one value from this discrete compound distribution. $|W(\mu_{ij})|$ is numerically approximated by drawing each $\underline{\theta}_{i,j,k}$ from a Dirichlet distribution with unit-defined parameters and then counting the relative frequency of $\mu_{ij} = \dfrac{\sum_{k=1}^{m_{ij}} \sum_{\ell=1}^{n_{ij}m_{ij}} N_{ijk\ell}\theta_{ijk\ell}y_{ijk\ell}^*}{\sum_{k=1}^{m_{ij}} \sum_{\ell=1}^{n_{ij}m_{ij}} N_{ijk\ell}\theta_{ijk\ell}x_{ijk\ell}^*}$

4.3 Step 3: the posterior distributions of $\beta|\mu_{i,j}, \sigma^2$ and $\sigma^2|\mu_{ij}, \beta$

In the example below, untruncated Normal distributions were used, instead of the correct truncated ones. Even though this was done for expediency it, possibly, offers a more direct comparison between the multinomial model and the model that uses a Normal approximation (see the next section). This is because the distribution for μ_{ij} in these two models are both the same (i.e., not truncated) leaving the comparison between the two unit-level likelihoods as the only two different parts of the model.

In this case, the conditional posterior of β is also Normal and can be sampled from directly. The conditional posterior of σ is sampled using a Metropolis/Hasting step where Jeffrey's prior is used to form the proposal distribution along with the distribution of μ_{ij}.

If implemented in a production setting, the truncated distribution would need to be used. A nearest neighbor MCMC may work in this case or, if not, the parameter spaces of β and of σ could also be discretized and a "Griddy-Gibbs" type algorithm (Ritter and Tanner, 1992) implemented.

4.4 Step 4: the posterior distributions of $\underline{\pi}|data$

This step is the same as the Bayesian Bootstrap and can be implemented by drawing $\underline{\pi}$ from a Dirichlet distribution with unit parameters.

5. An example Using the NHIS to Predict County-level Access to Care for the States Bordering Mexico

In this section the Constrained Bayesian Bootstrap model, described above, will be implemented and compared with an alternative model that assumes the county sample rate is Normally distributed with known variance. As an example, the county proportion of people who have delayed medical care is estimated in all the counties in the States bordering Mexico. Specific details about which years in the 2006-2015 NHIS design were

used and how 'delayed medical care''' is defined are not described here in order to avoid any misunderstanding that these estimates can be construed as anything but demonstration estimates. However, estimates such as these could be of use for targeted studies that do not want to "borrow" data from the entire U.S. For example, estimates of the health of the population along the border may be of adjunctive interest to the La Paz Agreement (see https://www.epa.gov/sites/production/files/2015-09/documents/lapazagreement.pdf). Estimates for other parts of the country focused on specific populations may also be of interest in future applications of small area methods (see https://www.hhs.gov/about/agencies/iea/regional-offices for on overview of the HHS health regions).

The basic unit used is the segment. Missing housing units within segments planned to be completely enumerated were imputed using the in segment-level weight: wtfa/(basewgt/10000) as were unsampled housing units, where within segment subsampling occurred. These segment totals (i.e. the total persons who delayed medical care and the total population) were treated as the complete data within a sampled segment with no within segment variation.

The segment totals, for sample segments, along with corresponding county-level information is the input into the model. As specified above, an MCMC approach making use of the "hit-and-run" algorithm is used for posterior inference.

5.1 Selection of County-level Covariates

Covariates for the linear regression of the μ_{ij} were selected using design-based estimates of μ_{ij}. The covariates under consideration were collected at the county level and obtained from the U.S. Census Bureau's "American Faultfinder" website. Decennial census, ACS and Economic Census values were all considered. Specifically the following covariates considered were:%white, %African American, %Asian, %American Indian Alaskan Native, %urban %poverty %Hispanic, latitude, longitude, metropolitan status, land area, shortest distance to Mexico, population density, the density of ambulatory care facilities per square miles and the density of ambulatory care facilities per persons.

As this is a demonstration and evaluation project, a detailed model was not considered. If small area estimates were actually needed, variable selection procedures such as "rpart" in R and the SAS procedure "Apaptivereg" could be useful. For this demonstration, an estimate of the county poverty rate obtained from the ACS was used as the single covariate. Poverty rate looked to be one of the important predictors based on a quick analysis.

5.2 County Estimates of the Proportion who Delayed Medical Care

Following the model and procedures outlined above, the county proportions who delayed medical care were sampled from their posterior distribution. Figure 1 provides a map of the posterior mean of the proportion of persons who delayed medical care. This map is a demonstration that the method can work. However, the algorithm is slow, taking about 18 hours of computation time to run for 1500 iterations. In addition, the sampled counties with more than forty segments were based on a Normal approximation to avoid an even lengthier computation time.
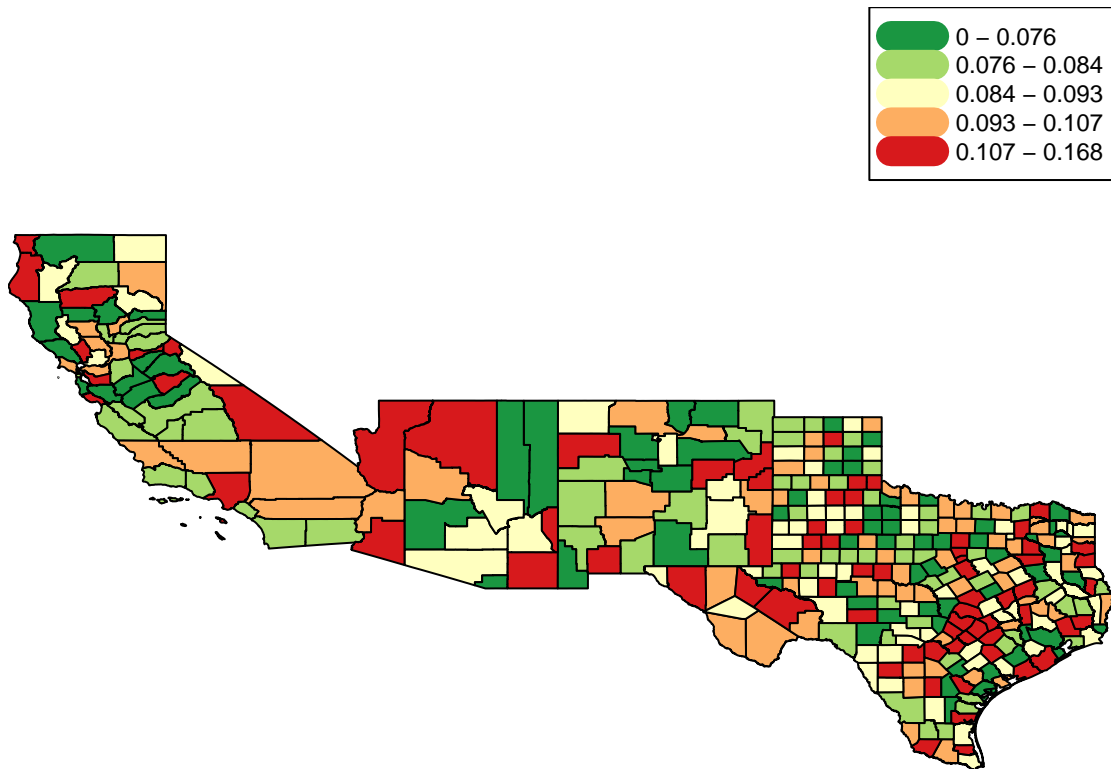
Figure 1: Posterior Means of the Proportion Who Delayed Medical Care: The Constrained Multinomial Model

5.2.1 A Posterior Predictive Check of the Model

The unit-level model allows one to create a new, posterior predicative sample that can be used to assess how well the model can replicate selected, observed sample quantities. In the following, posterior predicted samples of segments were created and county-level, design-based estimates and their corresponding design-based estimate of variance were calculated. Using these repeated predicted samples, 95% credible intervals can be made for both the design-based estimates and their variances or standard errors.

Figure 2 displays these intervals for the counties in sample, excluding those large counties in which a Normal approximation was used instead of using the constrained multinomial. The county estimates are ordered by the number of sampled segments in each county and presented on an ordinal scale to be able to view each county clearly. As can be seen, most observed values lie within the intervals predicted by the model.
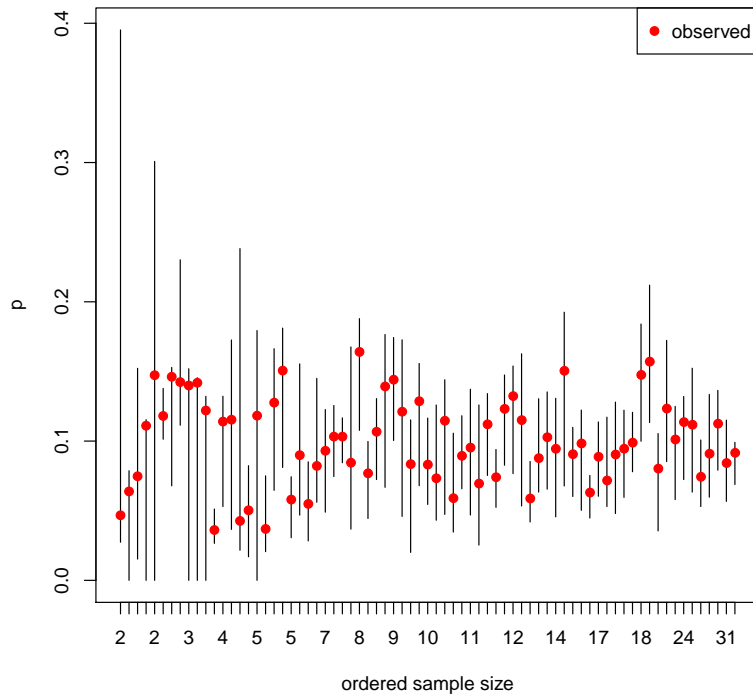
Figure 2: 95% Bayesian posterior predicted intervals of the design-based sample county proportion who delayed medical care by county sample size

Figure 3 presents the corresponding results for the sample design-based standard errors of the estimates and their corresponding observed values. Perhaps not surpassingly, the predicted sample standard errors do not do cover the observed values very well when there are only around two segments sampled per county.
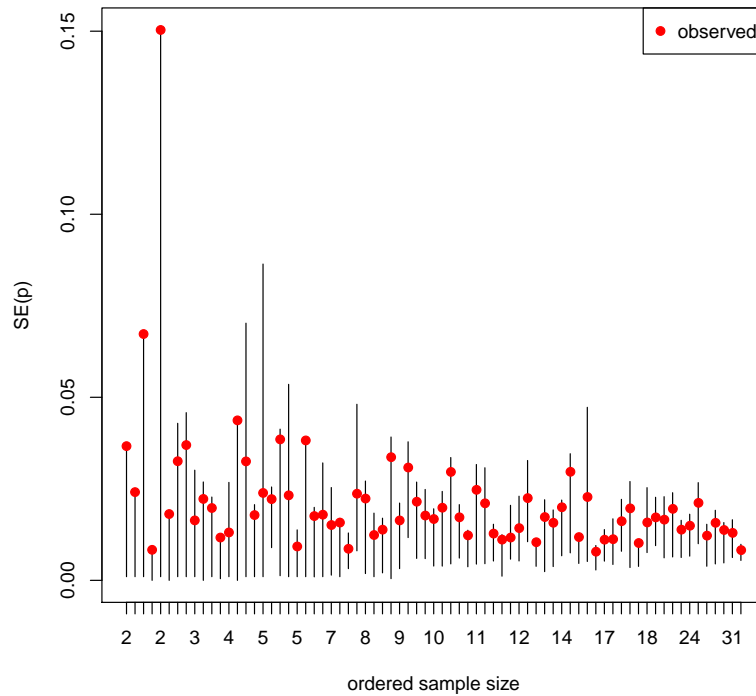
Figure 3: 95% Bayesian posterior predicted intervals of the design-based standard errors estimates of the sample county proportion who delayed medical care by county sample size

5.3 Comparison of Estimates from the Constrained Multinomial Model with a Normal Approximation Model

As mentioned in the introduction, usually design-based estimates constructed at the small area-level are assumed to be Normally distributed with known variance and then used as the input into a small area model. Using the exact same model at the county level, the two approaches are compared. Figure 4 plots the posterior mean of μ_{ij} based on the two methods for all counties in the four states, both in and out of sample. Although the estimates fall close to the line of equality, they do deviate from model-to-model and the deviations follows a pattern. This example demonstrates that there is an effect of using the Normal approximation as opposed to using a multinomial. Further work would need to be done in order to evaluate whether one approach is better than another.
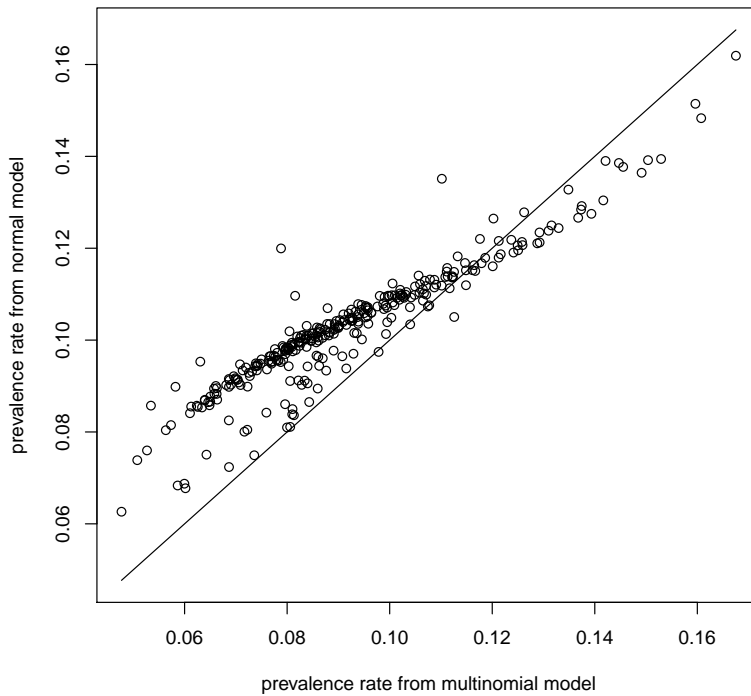
Figure 4: posterior means of μ_{ij}

Figure 5 presents a similar comparison using the posterior standard error of μ_{ij}. Again, there is a pattern suggesting that the choice of within small area model can have an effect. Note that the right side of the figure, where the constrained multinomial model exhibits larger standard errors than the approximate normal model, consists of the counties that were not in sample.
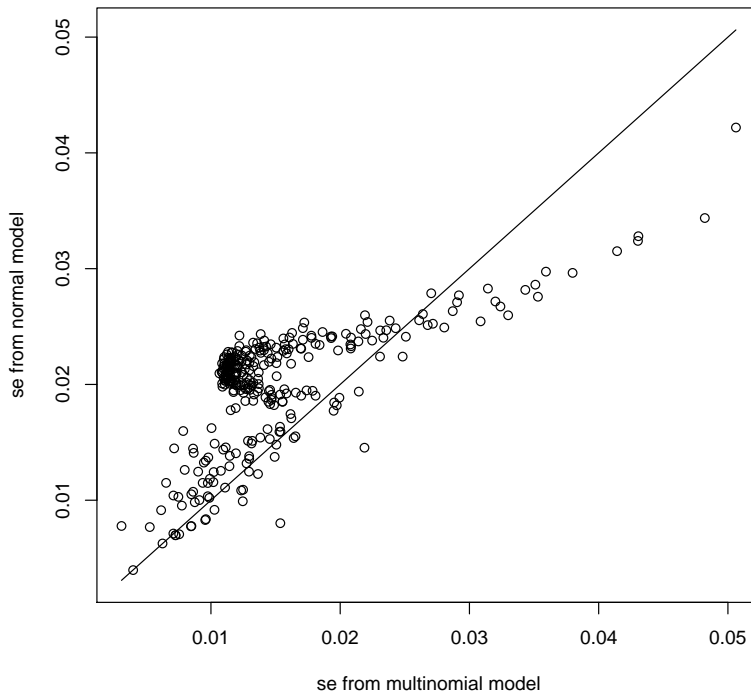
Figure 5: posterior standard errors of μ_{ij}

6. An Evaluation Between Using a Uniform Model within small areas, compared to using a multinomial

The following evaluates the possibility of using a uniform model, with unknown upper and lower bounds as a possible way to model the sample distribution within MDS strata. The uniform distribution has the advantage over the Bayesian-bootstrap-type multinomial model in that its distributional range can be extended beyond the sample range. As a quick check, models only at the county level were used for comparison. A multinomial model without county-level constraints is specified within each MDS within county and used to compare to a uniform model, also without county constraints. The multinomial model can be shown to replicate the design-based sample means and variances since the sample of segments is considered SRS within minority density stratum.

- Simple Multinomial model used:
  - For each ℓ within MDS i, j, k, $prob((y_{ijk\ell}, x_{ijk\ell}) = (y_{ijk\ell*}, x_{ijk\ell*})) = \theta_{ijk\ell*}$
  - $pr(\underline{\theta_{ijk}}) \propto \prod_\ell \theta_{ijk\ell}^{-1}$

- Uniform model used:
  - $(y_{ijk\ell}/x_{ijk\ell}) \sim U(ar_{ijk}, br_{ijk}), \qquad 0 < ar_{ijk} < br_{ijk} < 1$
  - $(x_{ijk\ell}) \sim U(ax_{ijk}, bx_{ijk}), \qquad 0 < ax_{ijk} < bx_{ijk}$
  - all independent
  - $p(ar_{ijk}, br_{ijk}, ax_{ijk}, bx_{ijk}) \propto constant$

Figures 6 and 7 show the respective 95% credible intervals of the predicted sample prevalences of delayed care for the simple multinomial model and for the uniform model,

along with the observed sample prevalences. It is clear that the uniform model (fig. 7) is a poor choice as evidenced in the non-coverage of the observed sample prevalences for counties with a large sample. This is because the posterior predicted mean from the uniform model is based on the midpoints of its ranges. So, even though the population range is extended beyond the sample range, the effect of a sample modes are not accounted for.
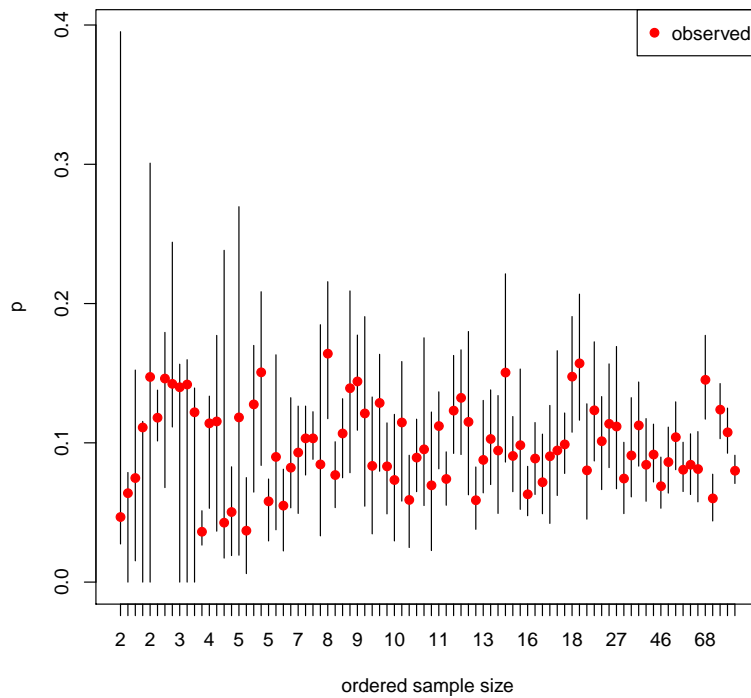

Figure 6: prediction of the design-based sample county proportion who delayed medical care by county sample size using simple multinomial model
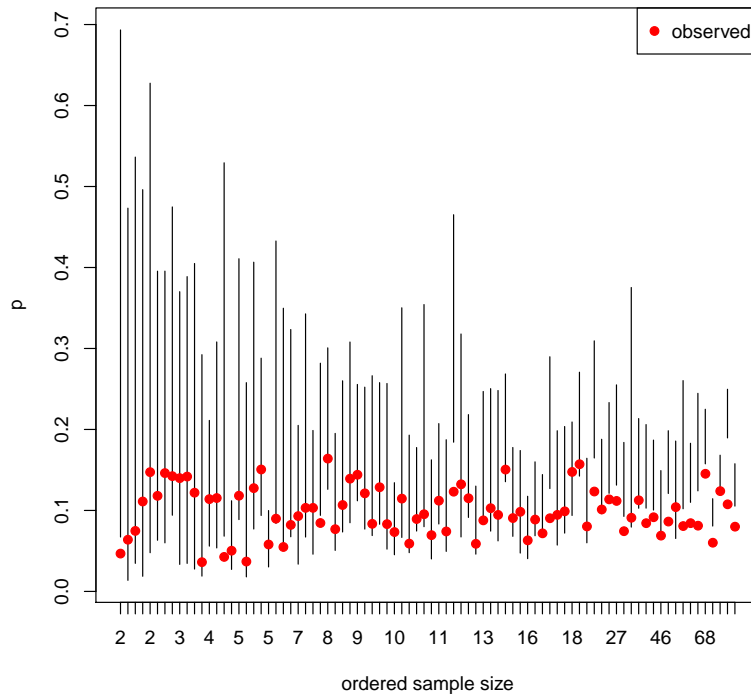
Figure 7: prediction of the design-based sample county proportion who delayed medical care by county sample size using uniform model

7. Summary

It was shown that a constrained multinomial model that combines the nonparametric benefits of the Bayesian Bootstrap with the a hierarchical model for small areas can be successfully implemented. Based on the preliminary model fitting experience with implementing this approach in the estimation of delayed medical care from the National Health Interview Survey, it is shown that the model adequately fits the observed sample, in terms of being able to replicate the county sample prevalence and estimated sample variances successfully, when the sample sizes are not too small. A possible remedy to correct problems in adequately modeling the variability in small areas with very small sample sizes could be to extend the hierarchical model of the constrained multinomial distribution across small areas via a relatively non-parametric distribution.

The "hit and run" algorithm can be computationally very slow with decreasing speed as the sample size increases. Substituting a large sample Normal approximation could be adequate for the counties with large sample but doing so, precludes the ability to make unit level predictions and unit level model checks for those counties that are approximated.

Using a uniform distribution in place of a multinomial distribution, as a way to extend the population range, did not work. Using the data example as a check, the ranges were extended but the predictive model checks were poor. A possible improvement may be to replace the multinomial model with a continuous discrete distribution, with uncertain boundaries between observations but with constant probabilities inside any given boundary.

REFERENCES

Hartley, H.O. and Rao, J.N.S. (1968). A New Estimation Theory for Sample Surveys. Biometrika, 55, 547557.

Malec, Donald, Davis, William W. and Cao, Xin. Model-based small area estimates of overweight prevalence using sample selection adjustment. Statistics in Medicine. Volume 18, Issue 23, 31893200.

Malec, Donald (2005). Small Area Estimation from the American Community Survey Using a Hierarchical Logistic Model of Persons and Housing Units. Journal of Official Statistics, 21, 3, 411-432.

Parsons VL, Moriarity C, Jonas K, et al.(2014) Design and estimation for the National Health Interview Survey, 2006-2015. National Center for Health Statistics. Vital Health Stat 2(165).

Ritter, Christian; Tanner, Martin A. (1992). "Facilitating the Gibbs Sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler". Journal of the American Statistical Association. 87 (419): 861868

Rubin, D.B. (1981). The Bayesian Bootstrap. Annals of Statistics, 9, 130134.