Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. http://www.jstatsoft.org/

Rgbp: Bayesian Hierarchical Modeling and Frequentist Method Check

Joseph Kelly
Harvard University

Carl Morris
Harvard University

Hyungsuk Tak
Harvard University

Abstract

Bayesian-frequentist reconciliation via Bayesian hierarchical modeling for Gaussian, Binomial, and Poisson data and frequentist method check for good coverage probability.

Keywords: hierarchical model, multi-level model, random effects mixed model, method check, coverage probability, normal, binomial, poisson, shrinkage, R.

1. Introduction

Rgbp uses Bayesian machinery to estimate a two-level model (a random-effects mixed model) and allows for a check of its frequentist properties via a repeated sampling procedure (which we call a "method check"). It is found that even in small samples our procedure yields good frequency properties. Also, this package will be useful for Bayesians who want to see a non-informative reference point before and after constructing their full-Bayesian hierarchical model. For frequentists, it will provide confidence intervals of a random-effect mixed model with good repeated sampling properties.

2. Three Feasible Types of Data

This package is intended to fit a multi-level model on the group-level (or unit-level) data in which each group-level (or unit-level) observation is believed to have the Normal, Poisson, or Binomial distribution. In this section, we will introduce three specific types of feasible datasets.

2.1. Normal: 8 School Data

Education Testing Service conducted randomized experiments in eight separate schools and obtained this dataset. It contains the coaching effects on SAT scores ($y_j, j = 1, \ldots, 8$) and standard errors ($se_j, j = 1, \ldots, 8$) of eight schools obtained after an analysis of covariance adjustment (Rubin, 1981).

```
R> y  <- c(28,  8, -3,  7, -1,  1, 18, 12)
R> se <- c(15, 10, 16, 11,  9, 11, 10, 18)
```

In the original paper, each school's coaching effect has approximately Normal sampling distribution with known sampling variance, *i.e.*, standard error of each school is assumed to be known or to be accurately estimated. So, it is reasonable to think that each (group-level) coaching effect is distributed as independent Normal distribution given the unknown mean μ_j and known standard error: $y_j|\mu_j \overset{ind}{\sim} \text{Normal}(\mu_j, se_j^2),\ j = 1, \ldots, 8$. **Rgbp** includes this dataset and can be called by typing '`R> data(schools)`' on R.

2.2. Poisson: 31 Hospital Data

This dataset is about the medical profiling evaluations for Coronary Artery Bypass Graft (CABG) surgeries of 31 New York hospitals conducted in 1992 (Morris and Lysy, 2012). It comprises of the number of deaths within a month of CABG surgeries in each hospital ($z_j,\ j = 1, \ldots, 31$) and total number of patients receiving CABG surgeries (case load) in each hospital ($n_j,\ j = 1, \ldots, 31$). The below code is an example of input based on the last ten hospital data.

```
R> z <- c( 14,   9,  15,  13,  35,  26,  25,  20,   35,   27)
R> n <- c(593, 602, 629, 636, 729, 849, 914, 940, 1193, 1340)
```

Considering the type of data, it makes sense to assume the number of deaths in each hospital has independent Poisson distribution given the unknown true rate parameter λ_j: $z_j|\lambda_j \overset{ind}{\sim} \text{Poisson}(n_j\lambda_j), j = 1, \ldots, 31$, where n_j can be interpreted as an exposure (not necessarily an interger). This dataset is also included in the package and can be called by '`R> data(hospital)`' on R.

2.3. Binomial: 18 Baseball Data

This dataset contains information about batting averages of 18 major league baseball players through their first 45 official at-bats of the 1970 season (Efron and Morris, 1975). Also, it has two covariates, League and Position, showing in which league and in which position each player was playing. In this paper, we will use Position for a tutorial purpose. For convenience, we transform this variable into a binary indicator, which is 1 if a player was a outfielder and 0 otherwise. The code below shows a way to make inputs. If we have more than one covariate, for example, `x1` and `x2`, then '`R> x <- cbind(x1, x2)`' will be the right input of the `gbp` function.

```
R> z <- c(18, 17, 16, 15, 14, 14, 13, 12, 11, 11, 10, 10, 10, 10, 10,  9,  8,  7)
R> n <- c(45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45)
R> x <- c( 1,  1,  1,  1,  1,  0,  0,  0,  0,  1,  0,  0,  0,  1,  1,  0,  0,  0)
```

The data indicate that independent Binomial distribution is appropriate for each player's number of hits among 45 at-bats conditioning on the unknown true batting average p_j: $z_j|p_j \overset{ind}{\sim} \text{Binomial}(n_j, p_j),\ j = 1, \ldots, 18$. This dataset is also a part of the package and can be called on R by '`R> data(baseball)`'.

3. Multi-level Structure

Our multi-level model, also called a conditionally independent hierarchical model (Kass and Steffey, 1989), is a very powerful tool for exploring the hierarchical sturucture in data. For example, we can think about a district-level hierarchy (bigger population) for 8 schools, the state-level hierarchy for 31 hospitals, and the position-level hierarchy for 18 baseball players. `gbp`, one of functions in **Rgbp**, fits such a hierarchical model whose first-level hierarchy has a distribution of observed data and second-level (bigger population hierarchy) has a conjugate prior distribution on the first-level parameter. Users can determine one of three types of multi-level models, such as Normal-Normal, Poisson-Gamma, and Binomial-Beta, based on their datasets.

3.1. Normal-Normal

`gbp` can construct a two-level Normal-Normal hierarchical model on the 8 school data. For reference, σ_j^2 below is assumed to be known or to be accurately estimated, and subscript j indicates j-th school in the dataset.

$$y_j|\mu_j \overset{ind}{\sim} \text{Normal}(\mu_j, \sigma_j^2), \tag{1}$$

$$\mu_j|\beta, A \overset{ind}{\sim} \text{Normal}(\mu_{0j}, A), \tag{2}$$

where $\mu_{0j} = x_j^T\beta,\ j = 1, \ldots, 8$, x_j is j-th school's covariate vector $(m \times 1)$, and m is the number of regression coefficients. Note that if there is no covariate then $x_{j,(1\times1)} = 1$ for an intercept term and so $\mu_{0j} = \mu_0 = \beta_0$ for all j, resulting in an exchangeable prior distribution. For reference, a paramter with a zero subscript, such as μ_{0j}, represents a mean parameter of the prior (second-level) distribution, *i.e.*, a prior mean. Also, Based on this conjugate prior distribution, it is easy to derive corresponding posterior distribution, which is the same as (1) $\times$ (2) up to a normalizing constant.

$$\mu_j|\mathbf{y}, \beta, A \overset{ind}{\sim} \text{Normal}(\ (1 - B_j)y_j + B_j\mu_{0j},\ (1 - B_j)\sigma_j^2), \tag{3}$$

where $B_j \equiv \frac{\sigma_j^2}{\sigma_j^2 + A},\ j = 1, \ldots, 8$, is called a shrinkages.

3.2. Poisson-Gamma

`gbp` is also able to build a Poisson-Gamma multi-level model on the 31 hospital data. Note that a constant, $1/r$, multiplied to the Gamma distribution below is a scale and a square bracket below indicates [mean, variance] of distribution. And for notational consistency, let's define $y_j \equiv z_j/n_j$ for all j.

$$z_j|\lambda_j \overset{ind}{\sim} \text{Poisson}(n_j\lambda_j), \tag{4}$$

$$\lambda_j|\beta, r \overset{ind}{\sim} \frac{1}{r}\text{Gamma}(\lambda_{0j}r) \sim \text{Gamma}[\lambda_{0j},\ \frac{\lambda_{0j}}{r}], \tag{5}$$

where $\log(\lambda_{0j}) = x_j'\beta$, and $j = 1, \ldots, 31$. Immediate posterior distribution of this Poisson-Gamma model is

$$\lambda_j|\mathbf{z}, \beta, r \overset{ind}{\sim} \frac{1}{r+n_j}\text{Gamma}(r\lambda_{0j} + n_j y_j) \sim \text{Gamma}[\lambda_j^*,\ \frac{\lambda_j^*}{r+n_j}], \tag{6}$$

where $\lambda_j^* \equiv (1-B_j)y_j + B_j\lambda_{0j}$, $B_j \equiv r/(r+n_j)$, and $y_j \equiv z_j/n_j$, $j = 1, \ldots, 31$.

3.3. Binomial-Beta

Binomial-Beta hierarchical model is the last model that `gbp` can fit. Again, a square bracket below indicates [mean, variance] of distribution.

$$z_j|p_j \overset{ind}{\sim} \text{Binomial}(n_j,\ p_j), \tag{7}$$

$$p_j|\beta, r \overset{ind}{\sim} \text{Beta}(rp_{0j},\ r(1-p_{0j})) \sim \text{Beta}[p_{0j},\ \frac{p_{0j}(1-p_{0j})}{r+1}], \tag{8}$$

where $\log(\frac{p_{0j}}{1-p_{0j}}) = x_j'\beta$ and $j = 1, \ldots, 18$. Then posterior distribution is

$$p_j|\mathbf{z}, \beta, r \overset{ind}{\sim} \text{Beta}(rp_{0j} + n_j y_j,\ r(1-p_{0j}) + n_j(1-y_j) \sim \text{Beta}\left[p_j^*,\ \frac{p_j^*(1-p_j^*)}{r+n_j+1}\right], \tag{9}$$

where $p_j^* \equiv (1-B_j)y_j + B_j p_{0j}$, $B_j \equiv \frac{r}{r+n_j}$, and $y_j \equiv \frac{z_j}{n_j}$, $j = 1, \ldots, 18$.

3.4. Hyper-prior Distribution

Hyper-prior distribution indicates a distribution of the second-level parameters, which plays an important role in deriving a full posterior distribution of all the parameters. `gbp` sets non-informative distributions on second-level parameters to let the data speak more about their estimation.

$$\beta \sim \text{Uniform on } \mathbf{R}^m, \quad A \text{ (or } \frac{1}{r}) \sim \text{Uniform}(0, \infty), \tag{10}$$

where m is the number of regression coefficients. For β, it is reasonable choice to take flat (non-informative) distribution because information about the location gets plentiful as the number of groups increases. Another flat prior on the second-level variance component A (or $1/r$) that we suggest here will guarantee a posterior propriety under the moderate conditions and will bring a good repeated sampling property to us.

4. Estimation

4.1. Shrinkage Estimation

Estimating shrinkage is a key part of our hierarchical modeling. As we can see in (3), (6), and (9), the posterior means are a linear function of shrinkage and the posterior variances are also

a linear (Gaussian), quadratic (Poisson), or cubic (Binomial) function of shrinkage. It implies that we cannot obtain $E(\mu_j|\mathbf{y})$ and $Var(\mu_j|\mathbf{y})$ without appropriate shrinkage estimation.

4.2. Adjustment for Density Maximization

When it comes to estimating a shrinkage, we can notice that it is a function of the second-level variance component, *i.e.*, $B_j \equiv \frac{\sigma^2}{\sigma^2+A} = B_j(A)$ for Gaussian and $B_j \equiv \frac{r}{r+n_j} = B_j(r)$ for Poisson and Binomial models.

In this case, the most common way is to obtain an MLE of the variance component with its asymptotic Normality and then to use Delta method for asymptotic Normal distribution of shrinkage, *i.e.*, $\hat{B}_{j,MLE} = B_j(\hat{A}_{MLE})$. But is the Normal approximation a good approximation for shrinkage that takes on a value between 0 and 1? Why do we fit a logistic regression, such as $\text{logit}(p) = x^T\beta$, instead of fitting a linear regression, $p = x^T\beta$?

Here the adjustment for density maximization (Morris and Tang, 2011), called ADM, comes. It assumes the Beta distribution for shrinkage because this distribution is also defined between 0 and 1. In the end, ADM estimates its posterior moments, *i.e.*, $E(B_j|\text{data})$ and $Var(B_j|\text{data})$, without any trouble that MLE can cause. Please refer to Morris and Tang (2011) for more apparent advantages for using ADM.

Once we estimate these two moments of shrinkage, we can also estimate the posterior moments given only data. Taking the Normal model as an example, $E(\mu_j|\mathbf{y}) = E(E(\mu_j|\mathbf{y}, r, A)|\mathbf{y})$ by Adam's law and $Var(\mu_j|\mathbf{y}) = E(Var(\mu_j|\mathbf{y}, r, A)|\mathbf{y}) + Var(E(\mu_j|\mathbf{y}, r, A)|\mathbf{y})$ by Eve's law. Note that both $E(\mu_j|\mathbf{y}, r, A)$ and $Var(\mu_j|\mathbf{y}, r, A)$ are a function of shrinkage in (3) and Adam's and Eve's laws are taking conditional expectation and variance over it given data.

4.3. Approximation to Posterior Distribution via Matching Moments

After estimating two posterior moments, $E(p_j|\mathbf{z})$ and $Var(p_j|\mathbf{z})$, `gbp` reasonably approximates a posterior distribution given data, *i.e.*, $p_j|\mathbf{z}$ (or $\mu_j|\mathbf{y}$, $\lambda_j|\mathbf{z}$), by matching these two moments with its parameters. To be specific, we assumed $p_j|\mathbf{z}$ had another Beta(a_1, a_0) distribution and matched two estimated moments, $E(p_j|\mathbf{z})$ and $Var(p_j|\mathbf{z})$, with two parameters, a_1 and a_0, of this Beta distribution.

5. Method Check

Like the two sides of the same coin, checking a statistical model always comes with fitting a model. If a fitted model cannot pass a checking process, we usually go back to the fitting process and come back to the checking process iteratively. In this sense, checking a fitted model is an interactive procedure for the model justification.

There are two kinds of model justification process; one is a model check and the other is a method check. The model check is for the justification of a hierarchical modeling on a specific dataset. One possible question is, "Can this dataset benefit from such a multi-level modeling?" Christiansen and Morris (1996) answered this question by using a mixture model,

$z_j|\beta, r \sim$ Negative-Binomial, on Poisson data to justify the second-level hierarchy. They found that their data had more variation than expected of the first-level Poisson distribution and Poisson hierarchical model could successfully account for such additional variation.

Once we are sure that the hierarchical modeling can be appropriate for our data, the following quetsion will be about the validity of interval estimates, the final product of this multi-level modeling. "Does the 95% (can be specified differently) confidence interval obtained via this Bayesian model-fitting process achieve 95% confidence level for any true parameter values?" Our answer is "yes" and **Rgbp** has a function to assure this point. From now on, all the explanations will be based on the Binomial model.

5.1. Pseudo-data Generation Process

Figure 1 will be helpful to understand this process. As we can see in (8), the distribution of each true batting average ($p_j,\ j = 1, \ldots, 18$) depends on two hyper-parameters, r and $\beta_{(m\times 1)}$, where m is the number of regression coefficients. So, once we fix these hyper-parameters at specific values, we can generate true batting averages. Suppose we sampled 500 $\mathbf{p}_{(18\times 1)}$'s, *i.e.*, $\{\mathbf{p}^{(i)}_{(18\times 1)},\ i = 1, \ldots, 500\}$ from the prior distribution in (8), where r and β are given. Then, we can also generate $\{\mathbf{z}^{(i)}_{(18\times 1)},\ i = 1, \ldots, 500\}$ given each p_{ij}, where i indicates i-th pseudo-dataset and j does j-th player. Next, `coverage` fits the Binomial hierarchical model 500 times on $\{(\mathbf{z}^{(i)}_{(18\times 1)}, \mathbf{n}_{(18\times 1)}),\ i = 1, \ldots, 500\}$ to obtain 500×18 interval estimates. After this generating process, as for the first player, we will have 500 pairs of ($z_{i1},\ p_{i1}$) and 500 interval estimates, $(\hat{p}_{i1,low}, \hat{p}_{i1,upp})$, $i = 1, \ldots, 500$.

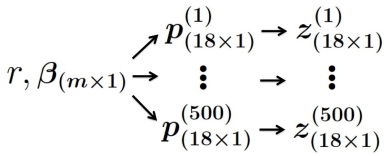


Figure 1: Pseudo-data generating process

5.2. Unbiased Coverage Probability

Based on the generated pseudo-datasets, let's define an indicator variable, I_{ij}, which is 1 if j-th player's interval estimate from i-th pseudo-dataset includes p_{ij} and 0 otherwise. One way to estimate the coverage probability is taking average over these indicator variables. We call it an unbiased coverage probability estimate. For example, $\bar{I}_1 = \sum_{i=1}^{500} I_{i1}/500$ is the estimated coverage probability for the first player.

5.3. Rao-Blackwellized Coverage Probability

Based on the definition of Rao-Blackwellization, we define $E(I_{ij}|z_{ij}, r, \beta)$, where r and β were given at first (see 5.1) and z_{ij} is a sufficient statistic. This expectation is the same as $\Pr(\hat{p}_{ij,low} \leq p_{ij} \leq \hat{p}_{ij,upp}|z_{ij}, r, \beta)$, where ($\hat{p}_{ij,low},\ \hat{p}_{ij,upp}$) is j-th player's interval estimates on the i-th dataset. We can calculate this probability because we know the distribution

of $p_{ij}|z_{ij}, r, \beta$ in (9). Note that conditioning on z_{ij} is equivalent to conditioning on $\mathbf{z}$ because we know r and β. Then, we can estimate the first player's coverage probability by $\sum_{i=1}^{500} E(I_{i1}|z_{i1}, r, \beta)/500$, which is way more accurate than the previous one.

6. Example

6.1. 31 Hospitals: Known Second-level Mean

Suppose you live in the New York state (NY) and have been suffering from severe coronary heart disease (hopefully not). If you are supposed to receive the coronary artery bypass graft (CABG) surgery soon, you might want to find the most famous hospital for dealing with such a surgery. On top of that, if you can figure out each hospital's abililty to handle this surgery, it will be useful for your decision to choose a hospital.

For this purpose, you gathered data of 31 hospitals in NY composed of the number of deaths within a month of CABG surgeries and total number of patients receiving CABG surgeries in each hospital. In addition, while you were looking for such information, suppose you could know that the state-level death rate per exposure of this surgery in the past ten years was 0.02.

The multi-level modeling that assumes a bigger population-level hierarchy will be insightful in this problem. Here, we presume a state-level (NY) hierarchy governing the true death rates of CABG surgery of all the hospitals in NY. This perspective is to view the true death rates of those 31 hospitals as sampled from the state-level population distribution whose mean is 0.02. For reference, a model check can be useful for evaluating the validity of such a view point (Christiansen and Morris, 1996).

Assuming an additional hierarchy is reasonable, a model-fitting process begins. Since the true death rate per exposure after CABG surgery might be small and the caseloads (n_j) look relatively much bigger than the number of deaths (z_j), the Poisson distribution would be our first choice to describe the uncertainty in our data. Next, `gbp` will help us fit the Poisson multi-level model with the Gamma conjugate prior distribution on the true death rate λ_j whose mean is 0.02 ($\lambda_0 = 0.02$) as described in 3.2 . For reference, the number of regression coefficients (m) is 0 because we do not need to estimate the prior mean via any regression.

```
R> p <- gbp(z, n, mean.PriorDist = 0.02, model = "pr")
R> p

Summary for whole observations:

          obs.mean    n prior.mean shrinkage low.intv post.mean upp.intv post.sd
1            0.045   67       0.02     0.834    0.011     0.024    0.042   0.008
2            0.029   68       0.02     0.832    0.010     0.022    0.038   0.007
3            0.024  210       0.02     0.616    0.011     0.021    0.035   0.006
4            0.043  256       0.02     0.568    0.017     0.030    0.046   0.008
5            0.033  269       0.02     0.556    0.015     0.026    0.041   0.007
```

6	0.044	274	0.02	0.551	0.018	0.031	0.047	0.008
7	0.043	278	0.02	0.548	0.018	0.030	0.047	0.007
8	0.014	295	0.02	0.533	0.008	0.017	0.029	0.005
9	0.029	347	0.02	0.492	0.014	0.024	0.038	0.006
10	0.037	349	0.02	0.491	0.017	0.029	0.043	0.007
11	0.039	358	0.02	0.484	0.018	0.030	0.045	0.007
12	0.018	396	0.02	0.459	0.010	0.019	0.030	0.005
13	0.028	431	0.02	0.438	0.015	0.024	0.037	0.006
14	0.025	441	0.02	0.433	0.013	0.023	0.035	0.005
15	0.027	477	0.02	0.414	0.015	0.024	0.036	0.006
16	0.045	484	0.02	0.410	0.023	0.035	0.050	0.007
17	0.030	494	0.02	0.405	0.016	0.026	0.039	0.006
18	0.022	501	0.02	0.402	0.012	0.021	0.032	0.005
19	0.028	505	0.02	0.400	0.015	0.025	0.036	0.005
20	0.020	540	0.02	0.384	0.012	0.020	0.031	0.005
21	0.028	563	0.02	0.374	0.016	0.025	0.037	0.005
22	0.024	593	0.02	0.362	0.014	0.022	0.033	0.005
23	0.015	602	0.02	0.358	0.010	0.017	0.026	0.004
24	0.024	629	0.02	0.348	0.014	0.023	0.033	0.005
25	0.020	636	0.02	0.346	0.012	0.020	0.030	0.005
26	0.048	729	0.02	0.316	0.027	0.039	0.053	0.007
27	0.031	849	0.02	0.284	0.019	0.028	0.038	0.005
28	0.027	914	0.02	0.269	0.017	0.025	0.035	0.005
29	0.021	940	0.02	0.264	0.014	0.021	0.030	0.004
30	0.029	1193	0.02	0.220	0.020	0.027	0.036	0.004
31	0.020	1340	0.02	0.201	0.014	0.020	0.027	0.003
colMeans	0.029	517	0.02	0.438	0.015	0.025	0.037	0.006

For reference, we need to type '`R> print(p, sort = FALSE)`' instead of '`R> p`' in order to list hospitals by the order of data input in the above output. '`R> p`' automatically sorts the output by the increasing order of n_j.

The output contains information about sample mean (`obs.mean`), caseload (`n`), known prior mean (λ_0), shrinkage (B_j), lower interval, posterior mean ($E(\lambda_j|\mathbf{z})$), upper interval, and standard deviation of posterior distribution ($sd(\lambda_j|\mathbf{z})$).

As we can see in (6), the posterior mean ($(1 - B_j)y_j + B_j\lambda_0$) is a linear function of shrinkage ($\equiv r/(r + n_j)$), locating between the sample mean and prior mean ($\lambda_0 = 0.02$). It makes sense because r can be interpreted as the amount of prior information and n_j as the amount of observed information. If the second level has more information than the first level, then the sample mean shrinks towards the prior mean more than 50%. This point is clear in the above output; as caseload increases, shrinkage decreases, depending less on the prior information.

A function "`summary`" shows selective information on hospitals and more detailed estimation result as below. To be specific, it displays some hospitals (not all as above) with minimum, median, and maximum caseloads (n_j). On top of that, more specific estimation results, such

as the estimation result of $\alpha \equiv \log(1/r)$, follow. Note that when we do not know the prior mean in advance unlike this hospital problem, `gbp` fits a regression model and the summary of regression fit will appear.

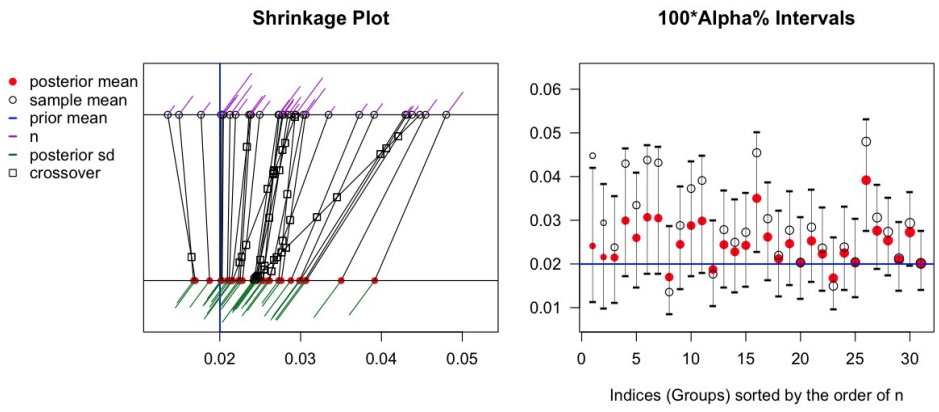
```
R> summary(p)
```

```
Main summary:

                  obs.mean    n prior.mean shrinkage intv.low post.mean intv.upp post.sd
Unit w/ min(n)       0.045   67       0.02     0.834    0.011     0.024    0.042   0.008
Unit w/ median(n)    0.045  484       0.02     0.410    0.023     0.035    0.050   0.007
Unit w/ max(n)       0.020 1340       0.02     0.201    0.014     0.020    0.027   0.003
Overall Mean         0.029  517       0.02     0.438    0.015     0.025    0.037   0.006


Second-level Variance Component Estimation Summary:
alpha = log(A) for Gaussian and log(1/r) for Binomial and Poisson data:

  post.mode.alpha post.sd.alpha
1          -5.818         0.411
```

Since estimated α is -5.818, we can easily calculate $r = \exp(5.818) = 336$, which will be helpful for understanding how each shrinkage of hospital was determined by comparing this value to the amount of observed information (n_j).

We also need a graphical summary that can give us valuable insight buried in pile of numbers and a function '`plot`' is exactly for this purpose.

```
R> plot(p)
```



The regression towards the mean (RTTM) is obvious in the left-side graph; the observed sample means are shrinking towards the known second-level mean (a blue vertical line at 0.02) to the different extents. Note that some hospitals' ranks have changed by shrinking much

harder towards 0.02 than others. For example, the empty square at the crossing point of the two left-most lines (8th and 23rd hospitals on the list above) indicates that seemingly safest hospital among 31 hospitals in terms of the observed sample mean was not actually safer than the second safest hospital. Without this hierarchical modeling, we might have made a wrong decision in choosing a hospital.

Intuitively, the result of multi-level modeling makes more sense than that of naive sample mean. For example, suppose there are two hospitals, whose sample means (z_j/n_j) are 0 and 0.01 and caseloads (n_j) are 1 and 100 each. Do you believe that the former hospital is better than the latter and are you going to choose the former hospital? Borrowing information from state-level hierarchy seems reasonable for the former hospital because it is hard to judge their true death rate per exposure with just one caseload. Though somewhat extreme, this is what happens for the two left-most hostpitals on the first plot and this is why hierarchical modeling is a reasonable choice for this dataset.

The estimated 95% intervals are displayed on the right-side plot. We can clearly see that all the posterior means (red dots) are between sample mean (empty dots) and second-level mean (a blue horizontal line). Overall, as the caseload (size of dots) increases towards the right and as the posterior mean gets closer to 0, the length of interval gets shorter, which the formula of posterior variance in (6) implies. But the most important information from this plot would be the visualized range of each hospital's true death rate per exposure.

This plot can add one more valuable point, which we could not have noticed, to the previous plot based mostly on point estimates. Let's look at the 8th and the 31st hospitals on the graph. The point estimate of the true death rate per exposure of the 31st hospital is higher than the one of the 8th. But, the upper bound of interval estimate of this 31st hospital is lower than that of the 8th. This interval plot makes the 31st hospital emerge as one of your candidates. (Could you regard this 31st hospital as a possible candidate before you observe this plot?)

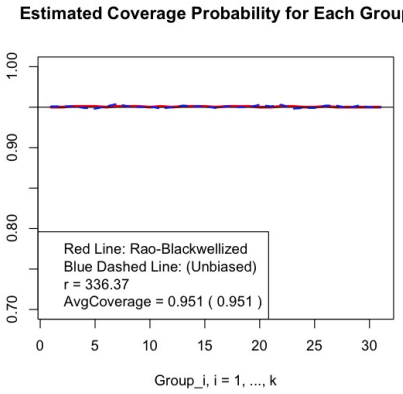
Also it reveals that the 23rd hospital whose estimated true rate was the smallest has also the smallest upper bound. If you are a risk-avoider, this hospital will attract you most strongly. And if you already chose this 23rd hospital compared to the 8th from the shrinkage plot, your decision might become stronger at this point, excluding the 8th hospital with more certainty.

Then, how reliable are these intervals? Does our procedure to generate interval estimates have good repeated sampling property? The method check below will answer this question.

```
R> pcv <- coverage(p, nsim = 10000)
```

This `coverage` function generated 10,000 pseudo-datasets assuming the estimated r (= 336.37) as a given value. For reference, we can try any other value of r, for example $r = 200$, by replacing above code with '`R> pcv <- coverage(p, A.or.r = 200, mean.PriorDist = 0.02, nsim = 10000)`.

It estimated coverage probabilities by unbiased estimates (a blue dashed line) and Rao-



Blackwellized estimates (a red line). Both lines are indistingishable from the horizontal line at 0.95 and the estimated overall average coverage rate is 0.951 based on both estimation methods. Finally, this method check showed that the interval estimates from the suggested multi-level modeling on this particular dataset has a very good repeated sampling property, achieving pre-specified 95% confidence level.

6.2. 8 Schools: Unknown Second-level Mean and No Covariate

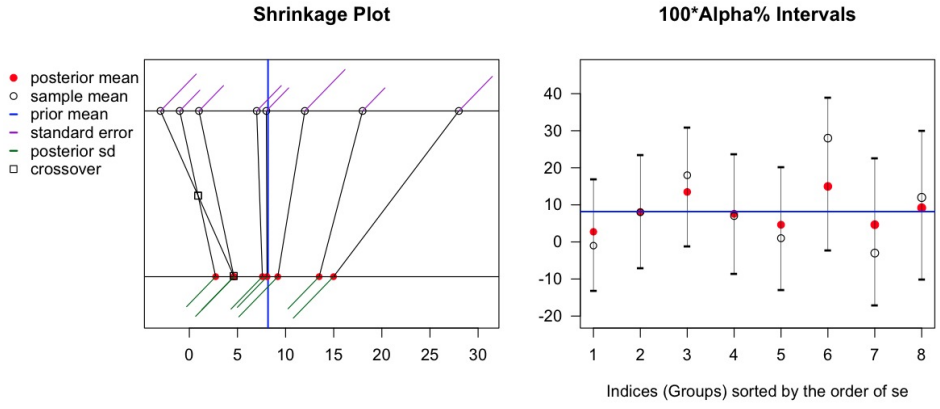


Figure 2: Shrinkage and interval plots of 8 schools

6.3. 18 Baseball Players: Unknown Prior Mean and One Covariate

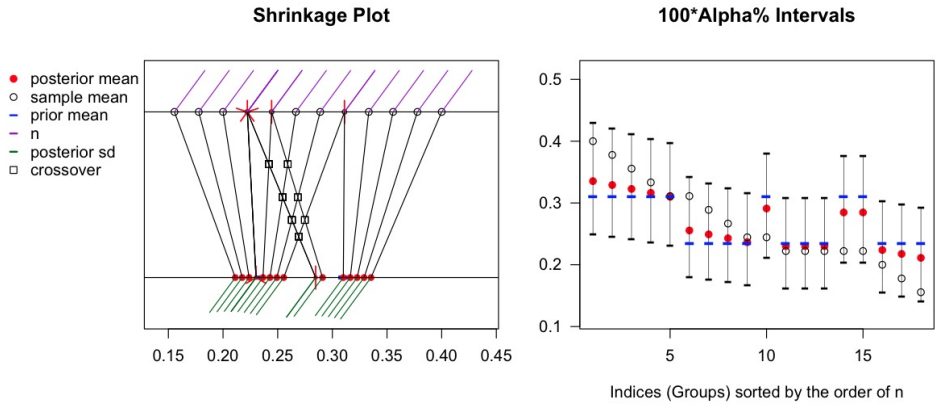


Figure 3: Shrinkage and interval plots of 18 baseball players

7. Discussion

8. Acknowledgments

9. Reference

1. Christiansen, C. and Morris, C. (1996). "Fitting and Checking a Two-Level Poisson Model: Modeling Patient Mortality Rates in Heart Transplant Patients," in *Bayesian Biostatistics*, eds. D. Berry and D. Stangl, New York: Marcel Dekker, pp. 467-561.

2. Efron, B. and Morris, C. (1975). "Data Analysis Using Stein's Estimator and its Generalizations." *Journal of the American Statistical Association*. **70**. 311-319.

3. Kass, R. and Steffey, D. (1989). "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)." *Journal of the American Statistical Association*. **84**. 717-726.

4. Morris, C. and Tang, R. (2011). "Estimating Random Effects via Adjustment for Density Maximization." *Statistical Science*. **26**. 271-287.

5. Morris, C. and Lysy, M. (2012). "Shrinkage Estimation in Multilevel Normal Models." *Statistical Science*. **27**. 115-134.

6. Rubin, D. B. (1981). "Estimation in parallel randomized experiments." *Journal of Educational Statistics*. **6**. 377-401.

Affiliation:

Joseph Kelly
Department of Statistics
Harvard University
1 Oxford Street, Cambridge, MA
E-mail: kelly2@fas.harvard.edu
URL: http://www.people.fas.harvard.edu/~kelly2/

Carl Morris
Department of Statistics
Harvard University
1 Oxford Street, Cambridge, MA
E-mail: morris@fas.harvard.edu

Hyungsuk Tak
Department of Statistics
Harvard University
1 Oxford Street, Cambridge, MA
E-mail: hyungsuk.tak@gmail.com

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd