Journal of Statistical Software

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

Rgbp: An R Package for Gaussian, Poisson, and Binomial Hierarchical Modeling

Joseph Kelly
Harvard University

Hyungsuk Tak
Harvard University

Carl Morris
Harvard University

Abstract

Rgbp is an R package that utilizes approximate Bayesian machinery to provide a method of estimating two-level hierarchical models for Gaussian, Poisson, and Binomial data in a fast and computationally efficient manner. The main products of this package are point and interval estimates for each group's random effect, μ_j for Gaussian, λ_j for Poisson, or p_j for Binomial model, whose good frequency properties can be validated via its repeated sampling procedure called "frequency method checking." It is found that such Bayesian-frequentist reconciliation allows **Rgbp** to have attributes desirable from both perspectives, working well in small samples and yielding good coverage probabilities for its interval estimates.

Keywords: multilevel model, conjugate hierarchical generalized linear models, frequency method checking, coverage probability, shrinkage, R.

1. Introduction

Rgbp is an R package for estimating and validating a two-level model, also known as a conjugate hierarchical generalized linear model (Lee, Nelder, and Pawitan 2006; Rönnegård, Shen, and Alam 2010) or a conditionally independent hierarchical model (Kass and Steffey 1989), producing point and interval estimates of random effects. The estimation procedure utilizes approximate Bayesian machinery and the validation involves checking frequency properties of the procedure via repeated sampling (which we call "frequency method checking"). It is found that even in small samples our procedure yields good frequency properties in comparison to other methods such as Maximum Likelihood Estimation (MLE).

This package is useful for both frequentists and Bayesians. Bayesians are able to use the package to see results for a non-informative reference prior before and after constructing their own Bayesian hierarchical model and frequentists are provided with a procedure with good

repeated sampling properties.

2. Hierarchical structure

A two-level or multilevel model is a powerful tool for exploring the hierarchical structure in data. For example, we can imagine that there exists a district-level hierarchy (bigger population) for observed school-level data, or a state-level hierarchy for observed hospital-level data in a certain state.

`gbp`, one of the functions in **Rgbp**, fits such a hierarchical model whose first-level has distributions of observed data and whose second-level has conjugate distributions on the first-level parameters. The `gbp` function allows users to choose one of three types of hierarchical models, namely Normal-Normal, Poisson-Gamma, and Binomial-Beta.

2.1. Normal-Normal ("g"=Gaussian data)

The following is the general Normal-Normal hierarchical model assumed by `gbp`. For reference, V_j ($\equiv \sigma^2/n_j$) below is assumed to be known, and subscript j indicates the j-th group among k groups in the dataset.

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

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

where $\mu_{0j} = x_j^\top \beta_{(m\times 1)} = \beta_0 + \beta_1 x_{1,j} + \cdots + \beta_{m-1} x_{m-1,j}$, $j = 1, \ldots k$ and m is the number of regression coefficients including an intercept term. The notation $x_j^\top = (1, x_{1,j}, x_{2,j}, \ldots, x_{m-1,j})$ represents the j-th group's $m \times 1$ covariate vector whose first element is 1 for the intercept term of regression coefficients. It is assumed that the second-level variance A is unknown and that $\beta_{(m\times 1)}$ is also unknown unless otherwise specified. Note that if there are no covariates then $x_j = 1$, $(m = 1)$ and so $\mu_{0j} = \mu_0 = \beta_0$ for all j, resulting in an exchangeable prior distribution. For reference, a parameter with a zero subscript, such as μ_{0j}, represents a mean parameter of the prior (second-level) distribution, i.e., a prior mean. Based on this conjugate prior distribution it is easy to derive the corresponding posterior distribution

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

where $B_j \equiv V_j/(V_j + A)$, $j = 1, \ldots, k$, are called shrinkage factors. By integrating over μ_j for $j = 1 \ldots k$ a marginal distribution, useful for model checking (Section 6), can be derived

$$y_j|\beta, A \overset{ind}{\sim} \text{Normal}(\mu_{0j}, V_j + A). \tag{4}$$

2.2. Poisson-Gamma ("p"=Poisson data)

`gbp` is also capable of estimating a Poisson-Gamma hierarchical model, though its usage is limited to the case where the second-level mean is known ($m = 0$). For notational consistency, let's define $y_j \equiv z_j/n_j$ as the unbiased estimate of λ_j for the j-th group, where z_j is the number of successes (or failures) out of n_j trials. For this Poisson model n_j can be interpreted as the exposure of the j-th group and is not necessarily an integer. The notation r represents the

amount of prior information and n_j represents the amount of observed information. Note that the constant, $1/r$, multiplied to the Gamma distribution below is a scaling factor and the square brackets indicates the [mean, variance] of a distribution.

$$z_j|\lambda_j \overset{ind}{\sim} \text{Poisson}(n_j\lambda_j), \tag{5}$$

$$\lambda_j|r \overset{ind}{\sim} \frac{1}{r}\text{Gamma}(r\lambda_0) \sim \text{Gamma}\left[\lambda_0,\ \frac{\lambda_0}{r}\right], \tag{6}$$

where λ_0, the second-level mean, is known a priori and r is the only unknown hyper-parameter. The posterior distribution of this Poisson-Gamma model given the hyper-parameter r is

$$\lambda_j|\mathbf{y}, r \overset{ind}{\sim} \frac{1}{r+n_j}\text{Gamma}(r\lambda_0 + n_j y_j) \sim \text{Gamma}\left[\lambda_j^*,\ \frac{\lambda_j^*}{r+n_j}\right], \tag{7}$$

where $\lambda_j^* \equiv (1-B_j)y_j + B_j\lambda_0$ and $B_j \equiv r/(r+n_j)$ for $j = 1, \ldots, k$.

Note that the variance in (6) is linear in the mean, whereas a slightly different Poisson-Gamma specification has been used elsewhere (Christiansen and Morris 1997) that makes the variances in (6) and (7) quadratic functions of the mean respectively .

2.3. Binomial-Beta ("b"=Binomial data)

The Binomial-Beta hierarchical model is the last model that `gbp` can fit. Again, z_j is the number of successes (or failures) out of n_j trials of the j-th group and the square bracket [,] below indicates the [mean, variance] of a distribution. Unlike the Poisson-Gamma model, the second-level mean is either known ($m = 0$) or unknown ($m \geq 1$) a priori.

$$z_j|p_j \overset{ind}{\sim} \text{Binomial}(n_j,\ p_j), \tag{8}$$

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

where $\log(\frac{p_{0j}}{1-p_{0j}}) = x_j^\top\beta,\ j = 1, \ldots, k$, and $\beta_{(m\times 1)}$ and r are unknown hyper-parameters. Then the corresponding posterior distribution given $\beta_{(m\times 1)}$ and r is

$$p_j|\mathbf{y}, \beta, r \overset{ind}{\sim} \text{Beta}(rp_{0j} + n_j y_j,\ r(1-p_{0j}) + n_j(1-y_j)) \sim \text{Beta}\left[p_j^*,\ \frac{p_j^*(1-p_j^*)}{r+n_j+1}\right], \tag{10}$$

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

2.4. Hyper-prior Distribution

A hyper-prior distribution is a distribution assigned to the second-level parameters. With the goal of objectivity in mind our choice for hyper-priors assumes the following non-informative distributions:

$$\beta \sim \text{Uniform on } \mathbf{R}^m \ \text{ and } \ A \sim \text{Uniform}(0, \infty) \ \ (\text{or } \frac{1}{r} \sim \text{Uniform}(0, \infty)), \tag{11}$$

where m is the number of regression coefficients. As for β, the flat non-informative distribution above is a common choice to ensure good frequency properties. In the Gaussian case, the flat

prior distribution on the second-level variance A is chosen to produce good repeated sampling properties for point and interval estimates of random effects $\{\mu_j\}$, and to have its posterior propriety when $k \geq m + 3$ (Morris and Tang 2011).

In the other two cases, Poisson and Binomial, the flat prior distribution on $\frac{1}{r}$ induces the same improper prior distribution on shrinkages ($\pi(B_j) \propto B_j^{-2} dB_j$) as does A with Uniform$(0, \infty)$ for the Gaussian. As for the posterior propriety, we found that the resultant posterior distribution for the Binomial model is proper if there are at least two groups (among k groups), whose observed values (z_j) are neither 0 nor n_j, and the covariate matrix ($X_{(k\times m)}$), if any, has full rank ($rank(X) = m$).

The Poisson model with hyper-prior in (11) provides posterior propriety if there are two groups whose observed values exceed 0 and the second-level mean (λ_0) is known ($m = 0$). However, if λ_0 is unknown ($m \geq 1$) the posterior distribution for the Poisson model is improper with hyper-prior in (11) and this model should not be used. In cases where the Poisson model is used because it is an approximation to the Binomial, then we recommend staying with the Binomial model with the same hyper-prior distribution. The efficiency of these priors is explored in the examples in section 5.

3. Estimation

3.1. Shrinkage Estimation

Estimating the shrinkage factors $(B_1, \ldots, B_k)$ is the key estimation problem with the hierarchical models `gbp` assumes. As we can see in (3), (7), and (10), the posterior means are a linear function of the shrinkage factors (B_j) and the posterior variances are linear (Gaussian), quadratic (Poisson), and approximately cubic (Binomial) functions of B_j. A natural method then to estimate $E(\mu_j|\mathbf{y})$ and $Var(\mu_j|\mathbf{y})$ is to first estimate the shrinkage factors.

3.2. Approximation via adjustment for density maximization

It is noted that the shrinkage factors $(B_1, \ldots B_k)$ are a function of the second-level variance component, i.e., $B_j \equiv V_j/(V_j + A) = B_j(A)$ for Gaussian and $B_j \equiv r/(r + n_j) = B_j(r)$ for Poisson and Binomial models. A common approach to approximate the distribution of B_j is to find its MLE $B_j(\hat{A}_{MLE})$ (or $B_j(\hat{r}_{MLE})$) and to use its asymptotic Normality. This Normal approximation, however, is defined on $(-\infty, \infty)$ whereas $0 < B_j < 1$ and hence in small sample sizes this approximation can be quite flawed and can even result in point estimates lying on the boundary of the parameter space.

To continue with a maximization-based estimation procedure but to steer clear of aforementioned boundary issues we make use of adjustment for density maximization (ADM) (Morris 1988a; Christiansen and Morris 1997; Morris and Tang 2011). For our purposes we approximate the posterior distribution of a shrinkage factor with a Beta distribution, which allows us to finally obtain estimates of the posterior moments, i.e., of $E(B_j|\mathbf{y})$ and $Var(B_j|\mathbf{y})$, where the notation $\mathbf{y}$ represents observed data, without the boundary issues that MLE encounters. See Morris and Tang (2011).

3.3. Posterior Moment Estimation

Using the posterior mean and variance estimates of B_j, **Rgbp** estimates the posterior mean and variance of the random effects given only the data ($\mathbf{y}$). Taking the Normal-Normal model as an example, it estimates $E(\mu_j|\mathbf{y})$ and $Var(\mu_j|\mathbf{y})$ by the following identities

$$E(\mu_j|\mathbf{y}) = E(E(\mu_j|\mathbf{y},\beta,A)|\mathbf{y}), \tag{12}$$

$$Var(\mu_j|\mathbf{y}) = E(Var(\mu_j|\mathbf{y},\beta,A)|\mathbf{y}) + Var(E(\mu_j|\mathbf{y},\beta,A)|\mathbf{y}), \tag{13}$$

where both $E(\mu_j|\mathbf{y},\beta,A)$ and $Var(\mu_j|\mathbf{y},\beta,A)$ are linear functions of the shrinkage factors as specified in (3). We then estimate $E(\mu_j|\mathbf{y})$ and $Var(\mu_j|\mathbf{y})$ by plugging-in the approximately estimated two posterior moments of shrinkage factors.

3.4. Approximation to Posterior Distribution via Matching Moments

After estimating the two posterior moments, for example $E(\mu_j|\mathbf{y})$ and $Var(\mu_j|\mathbf{y})$, `gbp` approximates a posterior distribution of the mean effects given the data by assuming a reasonable distribution and matching moments. For the Normal-Normal model we actually estimate the first three moments and approximate $\pi(\mu_j|\mathbf{y})$ with a skewed-Normal distribution (Azzalini 2013). For the Poisson-Gamma model we approximate $\pi(\lambda_j|\mathbf{y})$ with a Gamma distribution and finally for the Binomial-Beta model we approximate the posterior distribution of p_j, $\pi(p_j|\mathbf{y})$, with a Beta distribution.

4. Frequency Method Checking

Like the two sides of the same coin, checking a statistical model comes hand in hand with fitting the model. If a fitted model cannot pass a validation or checking process, we usually go back and forth from estimation and checking steps iteratively. In this sense, checking a fitted model is an interactive procedure for the model justification.

There are two kinds of model justification processes; one is model checking and the other is method checking. The model checking is for the justification of the hierarchical model on a specific dataset. One possible question is, "Can this dataset benefit from such hierarchical modeling?" Christiansen and Morris (1996) answered this question by using the Negative-Binomial mixture model 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 two-level Poisson hierarchical model could successfully account for such additional variation.

Once we make sure that the hierarchical model is appropriate for our data, the following question will be about the validity of the interval estimates, "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?" **Rgbp** has a function called `coverage` to answer this question and it comprises of two parts, generating pseudo-datasets and estimating coverage probabilities.

From now on, the explanation will be based on the Normal-Normal model because the idea can be easily applied to the other two models.

4.1. Pseudo-data generation process

Figure 1 will be helpful to understand this process. As we can see in (2), the distribution

of each parameter ($\mu_j,\ j = 1, \ldots, k$) is completely determined by two hyper-parameters, A and $\beta_{(m\times 1)}$. So, once we fix these hyper-parameters at specific values, we can generate as many parameters as we want. Suppose we generated an N_{sim} (default: $N_{sim} = 100$) number of $\boldsymbol{\mu}_{(k\times 1)}$'s, (i.e., $\{\boldsymbol{\mu}^{(i)}_{(k\times 1)},\ i = 1, \ldots, N_{sim}\}$), from the prior distribution in (2), where values of A and β are specifically given and the superscript (i) indicates the i-th simulation. Then using the distribution of observed data in (1), we can also generate $\{\mathbf{y}^{(i)}_{(k\times 1)},\ i = 1, \ldots, N_{sim}\}$ given each $\boldsymbol{\mu}^{(i)}_{(k\times 1)}$. Note that we generate one dataset per one parameter set.

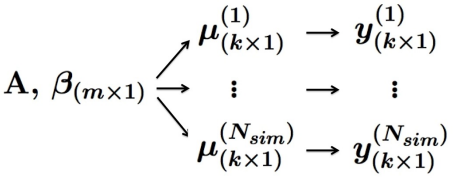

Figure 1: Pseudo-data generating process

4.2. Coverage estimation process

Next, `coverage` fits the Normal-Normal model N_{sim} number of times using the N_{sim} pseudo-datasets in order to obtain $N_{sim}\ \times k$ interval estimates, $\{(\hat{\mu}^{(i)}_{j,\ low},\ \hat{\mu}^{(i)}_{j,\ upp}),\ j = 1, \ldots, k,\ i = 1, \ldots, N_{sim}\}$. Here, we introduce both simple and Rao-Blackwellized unbiased coverage estimators but the function `coverage` in **Rgbp** displays only the Rao-Blackwellized one.

Simple unbiased coverage estimates

Let's define an indicator variable, $I^{(i)}_j$, which is 1 if the j-th group's interval estimate from the i-th pseudo-dataset includes $\mu^{(i)}_j$ and 0 otherwise. Then `coverage` estimates the coverage probability via the sample mean of these indicator variables for each group j, averaging over all the possible $\mu^{(i)}_j$ and $y^{(i)}_j$ for given A and β.

$$\bar{I}_j \equiv \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} I^{(i)}_j,\ j = 1, 2, \ldots, k. \tag{14}$$

This provides k simple unbiased coverage estimates. For example, $\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} I^{(i)}_1$ is the simple unbiased coverage estimate for the first group ($j = 1$), accounting for possible randomness in $\mu^{(i)}_1$ and $y^{(i)}_1$, $i = 1, \ldots, N_{sim}$, given specific values of A and β.

Each indicator has an independent and identically distributed (iid) Bernoulli($p_{cov,j}$) distribution, where $p_{cov,j}$ is the true coverage probability for the j-th group depending on specific values of A and β.

Then, its unbiased variance estimator for $Var(\bar{I}_j)$ is simply the sample variance of indicators over the number of simulations.

$$\widehat{Var}(\bar{I}_j) \equiv \frac{1}{N_{sim}(N_{sim} - 1)} \sum_{i=1}^{N_{sim}} (I^{(i)}_j - \bar{I}_j)^2,\ j = 1, 2, \ldots, k. \tag{15}$$

Rao-Blackwellized unbiased coverage estimates

Rao-Blackwellization improves accuracy of an unbiased estimator by taking more information (the sufficient statistic) into account. Based on this idea, we use $E(I_j^{(i)}|y_j^{(i)}, A, \beta)$, where A and β are given and $y_j^{(i)}$ is the sufficient statistic for the j-th group in the i-th pseudo-dataset.

$$\bar{I}_{RB,j} \equiv \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} E(I_j^{(i)}|y_j^{(i)}, A, \beta), \ j = 1, 2, \ldots, k. \tag{16}$$

The expectation part in (16) is $\Pr(\hat{\mu}_{j,low}^{(i)} \leq \mu_j^{(i)} \leq \hat{\mu}_{j,upp}^{(i)}|y_j^{(i)}, A, \beta)$, where $(\hat{\mu}_{j,low}^{(i)}, \hat{\mu}_{j,upp}^{(i)})$ is the j-th group's interval estimate on the i-th dataset. We can calculate this probability because we know the distribution of $(\mu_j^{(i)}|y_j^{(i)}, A, \beta)$ in (3). Note that conditioning on $y_j^{(i)}$ is equivalent to conditioning on $\mathbf{y}_{(k\times 1)}^{(i)}$ as long as A and β are known.

For example, we estimate the first group's true coverage probability $p_{cov,1}$ (which depends on specific values of A and β) by $\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} E(I_1^{(i)}|y_1^{(i)}, A, \beta)$, which is also unbiased but with smaller variance than the previous simple estimator, $\frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} I_1^{(i)}$.

Under the setting where one dataset $(\mathbf{y}_{(k\times 1)}^{(i)})$ is simulated per one true parameter set $(\boldsymbol{\mu}_{(k\times 1)}^{(i)})$, the unbiased variance estimator for $\mathrm{Var}(\bar{I}_{RB,j})$ follows.

$$\widehat{Var}(\bar{I}_{RB,j}) \equiv \frac{1}{N_{sim}(N_{sim}-1)} \sum_{i=1}^{N_{sim}} \left(E(I_j^{(i)}|y_j^{(i)}, A, \beta) - \bar{I}_{RB,j} \right)^2, \ j = 1, 2, \ldots, k. \tag{17}$$

5. Examples

5.1. Data of 31 hospitals with a known second-level mean

In this example we adopt the perspective of a person living in the state of New York (NY) who has been suffering from severe coronary heart disease. If this person must receive coronary artery bypass graft (CABG) surgery soon, he or she might want to find the most reliable hospital for such a procedure.

For this purpose, data were gathered from 31 hospitals in NY composed of the number of deaths $(\mathbf{z}_{(31\times 1)})$ for a specified period after CABG surgeries and the total number of patients $(\mathbf{n}_{(31\times 1)})$ receiving CABG surgeries in each hospital. For reference, caseloads (n_j) can be interpreted as exposures. These data can be loaded into R using the following code where the symbol '`R>`' represents a command prompt and is not to be typed into R.

```
R> z <- c( 3,   2,   5,  11,   9,  12,  12,   4,  10,  13,  14,   7,  12,
          11,  13,  22,  15,  11,  14,  11,  16,  14,   9,  15,  13,  35,
          26,  25,  20,  35,  27)
R> n <- c(67,  68, 210, 256, 269, 274, 278, 295, 347, 349, 358, 396, 431,
         441, 477, 484, 494, 501, 505, 540, 563, 593, 602, 629, 636, 729,
         849, 914, 940, 1193, 1340)
```

or

```
R> data(``hospital'')
R> z <- hospital$d
R> n <- hospital$n
```

In addition, suppose one knows that the state-level death rate per exposure of this surgery is 0.030 (λ_0). Using these data and the known second-level mean (λ_0), **Rgbp** provides point and interval estimates of the true death rate (λ_j) so that one can evaluate each hospital's reliability.

The independent Poisson distribution, i.e., $z_j|\lambda_j \overset{ind}{\sim} \text{Poisson}(n_j\lambda_j)$, $j = 1, \ldots, 31$, would be an ideal choice to describe these data based on the Poisson approximation when λ is small and n is relatively large.

Next, `gbp` fits the Poisson hierarchical model with the Gamma conjugate prior distribution as the NY state-level population distribution of the death rates whose mean is 0.030 ($\lambda_0 = 0.030$). For reference, the number of regression coefficients (m) is 0 because we do not need to estimate the prior mean (the second-level mean) via regression for this Poisson-Gamma model.

```
R> p <- gbp(z, n, mean.PriorDist = 0.03, model = "poisson")
R> p
```

```
Summary for each unit (sorted by n):
```

	obs.mean	n	prior.mean	shrinkage	low.intv	post.mean	upp.intv	post.sd
1	0.0448	67	0.03	0.911	0.0199	0.0313	0.0454	0.00653
2	0.0294	68	0.03	0.910	0.0189	0.0299	0.0435	0.00631
3	0.0238	210	0.03	0.765	0.0185	0.0285	0.0407	0.00566
4	0.0430	256	0.03	0.728	0.0225	0.0335	0.0467	0.00619
5	0.0335	269	0.03	0.718	0.0208	0.0310	0.0432	0.00573
6	0.0438	274	0.03	0.714	0.0229	0.0339	0.0472	0.00621
7	0.0432	278	0.03	0.711	0.0228	0.0338	0.0469	0.00617
8	0.0136	295	0.03	0.699	0.0157	0.0250	0.0366	0.00534
9	0.0288	347	0.03	0.663	0.0200	0.0296	0.0410	0.00536
10	0.0372	349	0.03	0.662	0.0222	0.0325	0.0446	0.00571
11	0.0391	358	0.03	0.656	0.0228	0.0331	0.0454	0.00579
12	0.0177	396	0.03	0.633	0.0165	0.0255	0.0363	0.00506
13	0.0278	431	0.03	0.613	0.0200	0.0292	0.0400	0.00511
14	0.0249	441	0.03	0.608	0.0191	0.0280	0.0387	0.00502
15	0.0273	477	0.03	0.589	0.0199	0.0289	0.0394	0.00499
16	0.0455	484	0.03	0.585	0.0256	0.0364	0.0491	0.00601
17	0.0304	494	0.03	0.580	0.0211	0.0302	0.0409	0.00506
18	0.0220	501	0.03	0.577	0.0180	0.0266	0.0369	0.00483
19	0.0277	505	0.03	0.575	0.0202	0.0290	0.0395	0.00494
20	0.0204	540	0.03	0.559	0.0173	0.0258	0.0358	0.00474
21	0.0284	563	0.03	0.548	0.0206	0.0293	0.0395	0.00485
22	0.0236	593	0.03	0.535	0.0187	0.0270	0.0369	0.00466
23	0.0150	602	0.03	0.532	0.0147	0.0230	0.0329	0.00466

```
24          0.0238  629       0.03      0.521    0.0188    0.0271    0.0368 0.00460
25          0.0204  636       0.03      0.518    0.0173    0.0254    0.0351 0.00455
26          0.0480  729       0.03      0.484    0.0286    0.0393    0.0516 0.00587
27          0.0306  849       0.03      0.446    0.0223    0.0303    0.0397 0.00445
28          0.0274  914       0.03      0.428    0.0208    0.0285    0.0374 0.00423
29          0.0213  940       0.03      0.421    0.0176    0.0249    0.0335 0.00407
30          0.0293 1193       0.03      0.364    0.0223    0.0296    0.0379 0.00397
31          0.0201 1340       0.03      0.338    0.0170    0.0235    0.0310 0.00360
colMeans             517      0.03      0.600    0.0201    0.0293    0.0403 0.00517
```

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 caseload (n_j), as shown above.

The output contains information about observed death rates (y_j), caseloads (n_j), known prior mean (λ_0), shrinkage estimates ($\hat{B}_j$), lower bounds of interval estimates ($\hat{\lambda}_{j,low}$, 2.5% percentiles of the approximate posterior distributions if 95% confidence level is given), approximate posterior means ($\hat{\lambda}_j = E(\lambda_j|\mathbf{y})$), upper bounds of interval estimates ($\hat{\lambda}_{j,upp}$, 97.5% percentiles of the approximate posterior distributions), and standard deviations of the approximate posterior distributions ($sd(\lambda_j|\mathbf{y})$).

As we can see in (7), the posterior mean, $(1 - B_j)y_j + B_j\lambda_0$, is a convex combination of the sample mean and prior mean ($\lambda_0 = 0.030$) with the shrinkage factor, $B_j \equiv r/(r + n_j)$, determining the weight. This makes intuitive sense because r and n_j can be interpreted as the degree (sample sizes) of prior and observed information respectively. If the second level has more information than the first level, i.e., ensemble sample size r exceeds individual sample size n_j, then the estimator will shrink towards the prior mean more than 50%. This is clear because, as caseload increases, shrinkage decreases, depending less on the NY state-level (second-level) information.

A function `summary` shows selective information on hospitals and more detailed estimation results, 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 posterior mode and standard deviation of $\alpha \equiv \log(1/r)$, follow.

```
R> summary(p)

Main summary:

                    obs.mean    n prior.mean shrinkage low.intv post.mean
Unit with min(n)      0.0448   67       0.03     0.911   0.0199    0.0313
Unit with median(n)   0.0455  484       0.03     0.585   0.0256    0.0364
Unit with max(n)      0.0201 1340       0.03     0.338   0.0170    0.0235
Overall Mean                  517       0.03     0.600   0.0201    0.0293

                    upp.intv  post.sd
                      0.0454  0.00653
                      0.0491  0.00601
                      0.0310  0.00360
```

```
                    0.0403   0.00517

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

  post.mode.alpha post.sd.alpha post.mode.r
1           -6.53         0.576         684
```

The output of `summary` also provides $\hat{r} = \exp(6.53) = 684$, which is an indicator of how valuable and informative the hypothetical second-level hierarchy is. It means that observed sample means of hospitals whose caseloads are less than 684 will shrink toward the prior mean (0.030) more than 50%. For example, the shrinkage estimate of the first hospital ($\hat{B}_1 = 0.911$) was calculated by 684 / (684 + 67), where 67 is its caseload (n_1), and its posterior mean is $(1-0.911)*0.0448+0.911*0.030 = 0.0313$. As for this hospital, using more information from the prior distribution is an appropriate choice because the observed amount of information (67) is far less than the amount of state-level information (684).

To obtain a graphical summary the function `plot` can be used, as seen in Figure 2.

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

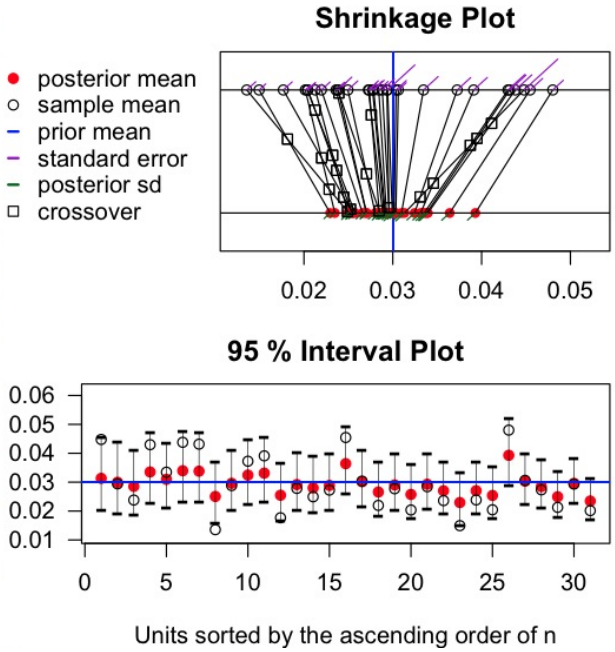

Figure 2: Shrinkage plot and 95% interval plot for 31 hospitals

In Figure 2 the regression towards the mean (RTTM) is obvious in the first graph; the observed sample means, empty dots on the upper horizontal line, are shrinking towards the known second-level mean (a blue vertical line at 0.030) to the different extents. Note that some hospitals' ranks have changed by shrinking much harder towards 0.030 than others.

For example, the empty square symbol at the crossing point of the two left-most lines (8th and 23rd hospitals on the list above) indicates that the seemingly safest hospital among 31 hospitals in terms of the observed death rate is probably not the safest in terms of the estimated posterior means.

Intuitively, switching ranks for these two hospitals makes sense. To be specific, their observed death rates (y_j, $j = 8, 23$) are 0.0136 and 0.0150 and caseloads (n_j, $j = 8, 23$) are 295 and 602 each. Considering solely the observed death rates may lead to an unfair comparison because the latter hospital handled twice the caseload. **Rgbp** accounts for this caseload difference, making the death rate estimate for the former hospital shrink toward the state-level mean (λ_0=0.030) much harder than that for the latter hospital.

Note that the point estimates are not enough to evaluate hospital reliability because one hospital may have a lower point estimate but bigger uncertainty (variance) than the other. In the second plot of Figure 2, the estimated 95% intervals are displayed. We see that each posterior mean (red dot) is between the sample mean (empty dot) and the second-level mean (a blue horizontal line). For reference, we could plot this 95% interval plot by the order of data input via `plot(p, sort = FALSE)`.

This 95% interval plot reveals that the 31st hospital has the lowest upper bound even though its point estimate ($\hat{\lambda}_{31} = 0.0235$) is slightly bigger than that of the 23rd hospital ($\hat{\lambda}_{23} = 0.0230$). The observed death rates for these two hospitals ($y_j, j = 23, 31$) are 0.0150 and 0.0201 and the caseloads ($n_j, j = 23, 31$) are 602 and 1340 each. The 31st hospital has twice the caseload, which leads to borrowing less information from the NY state-level hierarchy (or shrinking less toward the state-level mean, 0.030) with smaller variance. Based on the point and interval estimates, the 31st hospital seems the most reliable one among all candidates.

When fitting a model it is always a good idea to question how reliable the estimation procedure is. For example, does our procedure generate interval estimates that have good repeated sampling properties? To answer this question the `coverage` function generates pseudo-datasets assuming the estimated r (= 683.53) is a true value. For reference, we can designate any other value of r, for example $r = 600$, by adding another argument, `A.or.r = 600`, into the code below; `R> pcv <- coverage(p, A.or.r = 600, nsim = 1000)`.

In addition, `gbp` also provides interval estimates with different confidence levels, for example 90%. For this, we need to go back to the code for fitting the model, adding another argument, `Alpha = 0.9`; `R> p <- gbp(z, n, Alpha = 0.9, mean.PriorDist = 0.03, model = "poisson")`. Then, the code below will evaluate whether interval estimates achieve the 90% confidence level.

```
R> pcv <- coverage(p, nsim = 1000)
```

In Figure 3, the black horizontal line at 0.95 represents the nominal confidence level and the red circles indicate Rao-Blackwellized (RB) unbiased coverage estimates for 31 hospitals. The overall RB unbiased coverage estimate across all the hospitals is 0.953. And none of RB unbiased coverage estimates for 31 hospitals are less than 0.95 regardless of their caseloads (n_j). This result shows that the interval estimates for this particular dataset accurately achieves 95% confidence under repeated sampling. Note that these estimates depend on the given true value of r, the known prior mean, and the assumption that the model is true.

The following code provides 31 RB unbiased coverage estimates for each hospital.

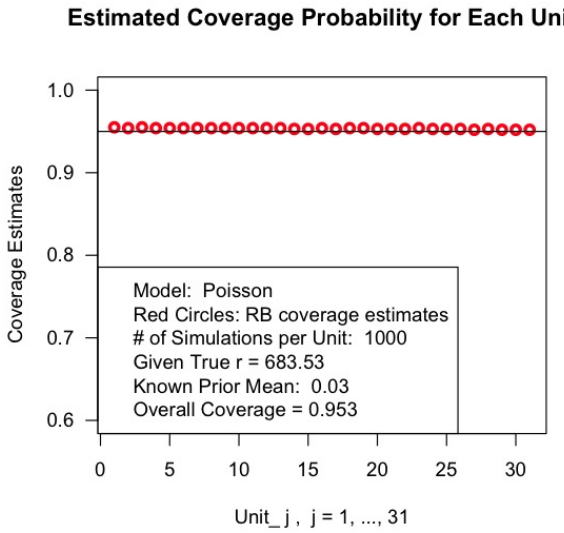

Figure 3: Coverage plot via frequency method checking for 31 hospitals

```
R> pcv$coverageRB

 [1] 0.955 0.954 0.955 0.954 0.954 0.954 0.954 0.954 0.954 0.954 0.954 0.954
[13] 0.954 0.953 0.953 0.954 0.953 0.954 0.954 0.953 0.953 0.953 0.954 0.953
[25] 0.953 0.953 0.952 0.953 0.952 0.952 0.952
```

And the code below shows 31 simple unbiased coverage estimates for each hospital.

```
R> pcv$coverageS

 [1] 0.949 0.960 0.958 0.958 0.950 0.945 0.960 0.953 0.960 0.956 0.955 0.946
[13] 0.955 0.954 0.960 0.965 0.955 0.952 0.960 0.956 0.959 0.955 0.964 0.960
[25] 0.945 0.942 0.947 0.960 0.940 0.946 0.956
```

The function `coverage` also calculates the standard errors for each hospital's RB unbiased coverage estimate defined in (17). The following code provides 31 standard errors for RB estimates.

```
R> pcv$se.coverageRB

 [1] 0.0016 0.0016 0.0014 0.0014 0.0013 0.0013 0.0013 0.0013 0.0012 0.0013 0.0012
[12] 0.0012 0.0011 0.0011 0.00110.0011 0.0011 0.0011 0.0011 0.0011 0.0010 0.0010
[23] 0.0010 0.0010 0.0010 0.0009 0.0009 0.0008 0.0008 0.0007 0.0007
```

Similarly, 31 standard errors for each simple unbiased coverage estimate defined in (15) are

```
R> pcv$se.coverageS
```

```
 [1] 0.0070 0.0062 0.0063 0.0063 0.0069 0.0072 0.0062 0.0067 0.0062 0.0065 0.0066
[12] 0.0072 0.0066 0.0066 0.0062 0.0058 0.0066 0.0068 0.0062 0.0065 0.0063 0.0066
[23] 0.0059 0.0062 0.0072 0.0074 0.0071 0.0062 0.0075 0.0072 0.0065
```

Taking the first hospital as an example, the variance estimate of RB unbiased coverage estimate is about 19 times smaller than that of simple one. It means that 1,000 RB unbiased coverage estimates are as precise as 19,000 simple unbiased coverage estimates in terms of estimating true coverage probability for the first hospital, $p_{cov,1}$.

For reference, two $31 \times 1,000$ matrices `raw.resultRB` and `raw.resultS`, each row of which is about each hospital, in `pcv` contain all the individual estimates, $I_j^{(i)}$ and $E(I_j^{(i)}|y_j^{(i)}, A, \beta)$.

Morris and Christiansen (1995) also investigated a similar ranking problem in hierarchical modeling, taking shrinkage into account.

5.2. Data of 8 schools with unknown second-level mean with no covariates

The Education Testing Service (ETS) conducted randomized experiments in eight separate schools (groups) to test whether students (units) SAT scores are effected by coaching. The dataset contains the estimated coaching effects on SAT scores ($y_j, j = 1, \ldots, 8$) and standard errors ($se_j, j = 1, \ldots, 8$) of the eight schools (Rubin 1981).

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

or

```
R> data(``schools'')
R> y  <- schools$y
R> se <- schools$se
```

Due to the nature of the test each school's coaching effect has an approximately Normal sampling distribution with known sampling variance, i.e., standard error of each school is assumed to be known. At the second hierarchy, the mean for each school is assumed to be drawn from a common Normal distribution and hence, we can use the Gaussian component of `gbp` to fit this Normal-Normal hierarchical model.

```
R> g <- gbp(y, se, model = "gaussian")
R> g

Summary for each unit (sorted by se):

         obs.mean   se prior.mean shrinkage low.intv post.mean upp.intv post.sd
5           -1.00  9.0      8.168     0.408  -13.297     2.737   16.692   7.634
2            8.00 10.0      8.168     0.459   -7.255     8.077   23.361   7.810
7           18.00 10.0      8.168     0.459   -1.289    13.484   30.821   8.176
4            7.00 11.0      8.168     0.507   -8.780     7.592   23.602   8.257
6            1.00 11.0      8.168     0.507  -13.027     4.633   20.131   8.441
1           28.00 15.0      8.168     0.657   -2.315    14.979   38.763  10.560
```

```
3           -3.00 16.0      8.168     0.685  -17.130     4.650   22.477  10.096
8           12.00 18.0      8.168     0.734  -10.208     9.189   29.939  10.227
colMeans          12.5      8.168     0.552   -9.163     8.168   25.723   8.900
```

This output from `gbp` summarizes the results. In this Normal-Normal hierarchical model the amount of shrinkage for each unit is governed by the shrinkage factor, $B_j = V_j/(V_j + A)$. As such, schools whose variation within the school (V_j) is less than the between school variation (A) will shrink greater than 50%. The results provided by `gpb` suggests that there is little evidence that the training provided much added benefit due to the fact that every school's 95% posterior interval contains 0. In the case where the number of groups is large **Rgbp** provides a summary feature:

```
R> summary(g)

Main summary:

                      obs.mean   se prior.mean shrinkage low.intv post.mean
Unit with min(se)        -1.00  9.0       8.17     0.408   -13.30      2.74
Unit with median(se)1     1.00 11.0       8.17     0.507   -13.03      4.63
Unit with median(se)2     7.00 11.0       8.17     0.507    -8.78      7.59
Unit with max(se)        12.00 18.0       8.17     0.734   -10.21      9.19
Overall Mean                   12.5       8.17     0.552    -9.16      8.17

                      upp.intv post.sd
                          16.7    7.63
                          20.1    8.44
                          23.6    8.26
                          29.9   10.23
                          25.7    8.90

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

  post.mode.alpha post.sd.alpha post.mode.A
1            4.77          1.14         118


Regression Summary:

      estimate   se z.val p.val
beta0    8.168 5.73 1.425 0.154
```

The summary provides results regarding the second level hierarchy parameters. It can be seen that the estimate of the second level mean, `beta0`, is not significantly different from 0 suggesting that there was no effect of the coaching program on SAT math scores.

Rgbp also provides functionality to plot the results of the analysis as seen in Figure 4. Plotting the results provides a visual aid to understanding but is only largely beneficial when the number of groups (k) is small.

```
R> plot(g)
```

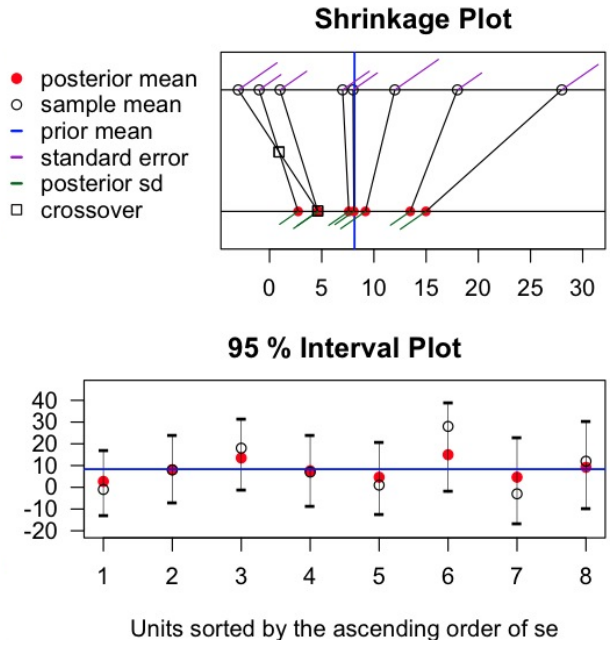

Figure 4: Shrinkage plot and 95% interval plot for 8 schools

The frequency method checking, assuming the model is correct, generates new pseudo-data from our assumed model. Unless otherwise specified, the procedure fixes the hyper-parameter values at their estimates ($\hat{A}$ and $\hat{\beta}_0$ in this example) and then simulates "true" θ_j for each group j. The model is then estimated and this is repeated an `nsim` number of times to estimate the coverage probabilities of the procedure.

```
R> gcv <- coverage(g, nsim = 1000)
```

As seen in Figure 5 the desired 95% confidence (black horizontal line at 0.95) is achieved (actually, exceeded) for each school in this example. Note that all the coverage estimates depend on the chosen true values of A and β_0, and the assumption that the model is valid.

In addition, Rao-Blackwellized (RB) unbiased coverage estimate and its standard error for each school can be gotten with the command below.

```
R> gcv$coverageRB

 [1] 0.966 0.959 0.967 0.960 0.959 0.962 0.960 0.966

R> gcv$se.coverageRB

 [1] 0.0013 0.0012 0.0013 0.0013 0.0011 0.0011 0.0010 0.0017
```

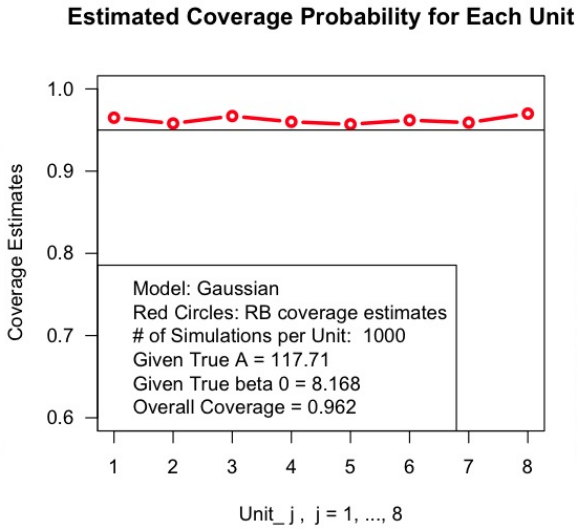

Figure 5: Coverage plot via frequency method checking for 8 schools

All the individual RB coverage estimates are saved in the $8 \times 1,000$ matrix, `gcv$raw.resultRB`, each row of which is about each school.

5.3. Data of 18 baseball players with unknown second-level mean and one covariate

The following dataset from the New York Times published on 26 April 1970 contains information on the batting averages and positions (outfielder=1,otherwise=0) of 18 major league baseball players through their first 45 official at-bats of the 1970 season (Efron and Morris 1975).

```
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)
```

or

```
R> data(``baseball'')
R> z <- baseball$Hits
R> n <- baseball$At.Bats
R> x <- ifelse(baseball$Position == "fielder", 1, 0)
```

Conditioning on the true batting average for each player we assume that the at-bats are independent and therefore, $z_j|p_j \overset{ind}{\sim} \text{Binomial}(45, p_j)$, $j = 1, \ldots, 18$. Our goal is to obtain point and interval estimates of the true batting average, p_j, for each player, whilst considering the additional information on whether the player is an outfielder or not. `gbp` provides a way to incorporate such covariate information seamlessly into the second-level hierarchy such that information is shared and regression towards the mean (RTTM) occurs within outfielders and non-outfielders separately.

```
R> b <- gbp(z, n, x, model = "binomial")
R> b

Summary for each unit (sorted by n):

          obs.mean  n   X1 prior.mean shrinkage low.intv post.mean upp.intv post.sd
1            0.400 45 1.00      0.310     0.715    0.248     0.335    0.429  0.0462
2            0.378 45 1.00      0.310     0.715    0.244     0.329    0.420  0.0448
3            0.356 45 1.00      0.310     0.715    0.240     0.323    0.411  0.0437
4            0.333 45 1.00      0.310     0.715    0.236     0.316    0.403  0.0429
5            0.311 45 1.00      0.310     0.715    0.230     0.310    0.396  0.0424
6            0.311 45 0.00      0.233     0.715    0.179     0.256    0.341  0.0415
7            0.289 45 0.00      0.233     0.715    0.175     0.249    0.331  0.0400
8            0.267 45 0.00      0.233     0.715    0.171     0.243    0.323  0.0388
9            0.244 45 0.00      0.233     0.715    0.166     0.237    0.315  0.0380
10           0.244 45 1.00      0.310     0.715    0.210     0.291    0.379  0.0432
11           0.222 45 0.00      0.233     0.715    0.161     0.230    0.308  0.0377
12           0.222 45 0.00      0.233     0.715    0.161     0.230    0.308  0.0377
13           0.222 45 0.00      0.233     0.715    0.161     0.230    0.308  0.0377
14           0.222 45 1.00      0.310     0.715    0.202     0.285    0.375  0.0441
15           0.222 45 1.00      0.310     0.715    0.202     0.285    0.375  0.0441
16           0.200 45 0.00      0.233     0.715    0.155     0.224    0.302  0.0377
17           0.178 45 0.00      0.233     0.715    0.148     0.218    0.297  0.0381
18           0.156 45 0.00      0.233     0.715    0.140     0.211    0.292  0.0389
colMeans          45 0.44       0.267     0.715    0.191     0.267    0.351  0.0410
```

Note that the shrinkage estimates are the same for all players due to the fact that they are determined solely by the relative amount of information between the first-level and the second-level hierarchies, ($\hat{B}_j \equiv \hat{r}/(\hat{r}+45) = 113/(113+45) = 0.715$).

```
R> summary(b)

Main summary:

                              obs.mean  n    X1 prior.mean shrinkage low.intv
Unit with min(obs.mean)          0.156 45 0.000      0.233     0.715    0.140
Unit with median(obs.mean)1      0.244 45 0.000      0.233     0.715    0.166
Unit with median(obs.mean)2      0.244 45 1.000      0.310     0.715    0.210
Unit with max(obs.mean)          0.400 45 1.000      0.310     0.715    0.248
Overall Mean                           45 0.444      0.267     0.715    0.191


                              post.mean upp.intv post.sd
                                  0.211    0.292  0.0389
                                  0.237    0.315  0.0380
                                  0.291    0.379  0.0432
                                  0.335    0.429  0.0462
```

```
                                    0.267    0.351  0.0410

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

  post.mode.alpha post.sd.alpha post.mode.r
1           -4.73         0.957         113


Regression Summary:

      estimate    se  z.val p.val
beta0   -1.194 0.131 -9.129 0.000
beta1    0.389 0.187  2.074 0.038
```

From the `Regression Summary`, one of the outputs of `summary`, we see that the two prior means distinguishing outfielders from other positions are significantly different (p-value for $\hat{\beta}_1$= 0.038). Also, the positive sign of $\hat{\beta}_1$ indicates that the population mean batting average for all outfielders tends to be higher than that for those in the other positions (estimated odds ratio = exp(0.389)=1.48).

```
R> plot(b)
```

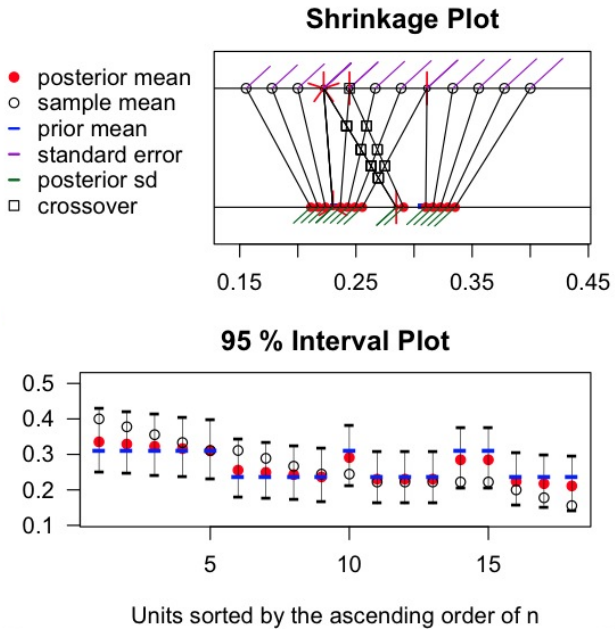

Figure 6: Shrinkage plot and 95% interval plot for 18 baseball players

It is evident in the shrinkage plot in Figure 6 that shrinkage occurs from the sample means (empty dots) on the upper horizontal line towards the two prior means, 0.233 and 0.310. For

reference, the short red line symbols on dots are for when two or more points have the same mean and are plotted over each other. For example, five players (from the 11th player to the 15th) have the same sample mean (0.222) and at this point on the upper horizontal line, there are short red lines toward five directions.

The 95% interval plot shows the range of true batting average for each player, which clarifies the regression toward the mean (RTTM) within two groups. The 10th, 14th, and 15th players, for example, are outfielders but their observed batting averages are far lower than the first five outfielders. This can likely be attributed to their bad luck because their observed batting averages are close to the lower bounds of their interval estimates. RTTM suggests that their batting averages will shrink towards the expected prior mean of outfielders (0.310) in the long run.

As in the previous examples in Section 5.1 and 5.2, in order to check the level of trust in these interval estimates, we can proceed to frequency method checking by assuming the estimates, 112.95 for $\hat{r}$ and (-1.194, 0.389) for $\hat{\beta}$, are given values.

```
R> bcv <- coverage(b, nsim = 1000)
```

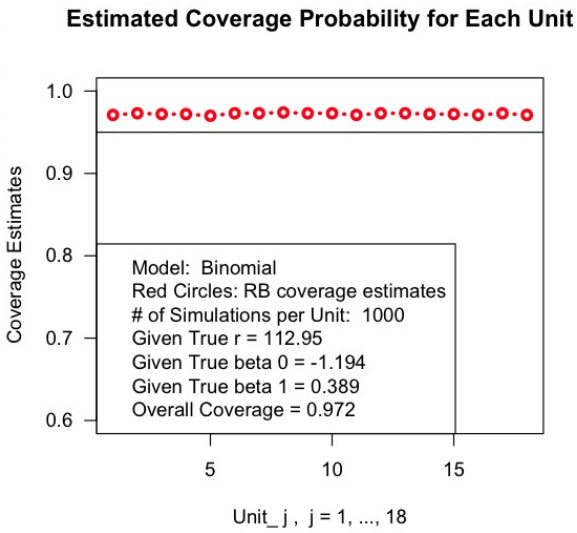

Figure 7: Coverage plot via frequency method checking for 18 players

For reference, to do the frequency method checking at different true values, we need to specify additional arguments in the coverage function. For example, if we want to try different true values, either 100 for r or (-1, 0.2) for (β_0, β_1), the additional arguments are `A.or.r = 100` and `reg.coef = c(-1, 0.2)`; `coverage(b, A.or.r = 100, reg.coef = c(-1, 0.2), nsim = 1000)`.

Finally, in Figure 7, we see that the overall Rao-Blackwellized unbiased coverage estimate is 0.972 (across all the players), conservatively satisfying the definition of the 95% confidence interval. Note that each coverage estimate depends on given true values of r and $\beta_{(2\times1)}$, and the assumption that the model is valid.

The Rao-Blackwellized unbiased coverage estimates and their standard errors for each player follow.

```
R> bcv$coverageRB

 [1] 0.971 0.973 0.972 0.972 0.970 0.973 0.973 0.974 0.973 0.973 0.971 0.973
[13] 0.973 0.972 0.972 0.971 0.973 0.971

R> bcv$se.coverageRB

 [1] 0.0015 0.0012 0.0013 0.0014 0.0016 0.0010 0.0012 0.0010 0.0010 0.0013
[11] 0.0015 0.0013 0.0019 0.0013 0.0014 0.0015 0.0011 0.0014
```

All the simulation results are saved in the $18 \times 1,000$ matrix, `bcv$raw.resultRB`, each row of which is for each player.

6. Discussion and summary

Rgbp is an R package for estimating and validating two-level Gaussian, Binomial and Poisson hierarchical models. The package aims to provide a procedure that is computationally efficient with good frequency properties and includes "frequency method checking" functionality to examine repeated sampling properties and to test that the method is valid at specified hyper-parameter values.

As an alternative to other maximization based estimation methods such as MLE and REML, **Rgbp** provides point and interval estimates of parameters via ADM. Using the ADM approach, with our specified choice of priors, protects from cases of overshrinkage and undercoverage from which the aforementioned methods suffer from (Morris 1988b).

A benefit of **Rgbp** is that it produces non-random output and so results are easily reproduced and compared across studies. In addition to being a standalone analysis tool the package can be used as an aid in a broader estimation procedure. For example, by checking the similarity of output of **Rgbp** and that of another estimation procedure (such as MCMC) the package can be used as a confirmatory tool to check whether the alternative procedure has been programmed correctly. In addition, the parameter estimates obtained via **Rgbp** can be used to initialize a MCMC thus decreasing time to convergence.

Due to its speed and ease of use, **Rgbp** can be used as a method of preliminary data analysis. Such results may tell statisticians and practitioners alike whether a more intensive method in terms of implementation and computational time, such as MCMC, is needed.

In addition to the built in "frequency method checking" procedure the package can be used to undergo "model checking". For example, in the Gaussian hierarchical model, the assumed marginal distribution of the data is given in (4). By substituting the point estimates of A and β from the package into this marginal distribution a test can be constructed to see whether the data follow the marginal distribution suggested by the hierarchical model.

7. Acknowledgments

The authors thank Professor Cindy Christiansen, Professor Phil Everson and the 2012 class of Harvard's Stat 324r: Parametric Statistical Inference and Modeling for their valuable inputs.

References

Azzalini A (2013). *R package* **sn***: The skew-normal and skew-t distributions (version 0.4-18)*. Università di Padova, Italia. URL http://azzalini.stat.unipd.it/SN.

Christiansen C, Morris C (1996). "Fitting and Checking a Two-Level Poisson Model: Modeling Patient Mortality Rates in Heart Transplant Patients." In D Berry, D Stangl (eds.), *Bayesian Biostatistics*, pp. 467–501. CRC press.

Christiansen C, Morris C (1997). "Hierarchical Poisson Regression Modeling." *Journal of the American Statistical Association*, **92**(438), pp. 618–632. ISSN 01621459. URL http://www.jstor.org/stable/2965709.

Efron B, Morris C (1975). "Data Analysis Using Stein's Estimator and its Generalizations." *Journal of the American Statistical Association*, **70**(350), pp. 311–319. ISSN 01621459. URL http://www.jstor.org/stable/2285814.

Kass RE, Steffey D (1989). "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)." *Journal of the American Statistical Association*, **84**(407), pp. 717–726. ISSN 01621459. URL http://www.jstor.org/stable/2289653.

Lee Y, Nelder JA, Pawitan Y (2006). *Generalized linear models with random effects: a unified analysis via h-likelihood*. Chapman & Hall/ CRC, New York.

Morris C (1988a). "Approximating Posterior Distributions and Posterior Moments." In J Bernardo, MH DeGroot, DV Lindley, AFM Smith (eds.), *Bayesian Statistics 3*, pp. 327–344. Oxford University Press.

Morris C (1988b). "Determining the Accuracy of Bayesian Empirical Bayes Estimates in the Familiar Exponential Families." In S Gupta, J Berger (eds.), *Statistical Decision Theory and Related Topics IV*, pp. 251–263. Springer-Verlag.

Morris C, Christiansen C (1995). "Hierarchical Models for Ranking and for Identifying Extremes, With Application." In J Bernardo, J Berger, A Dawid, A Smith (eds.), *Bayesian Statistics 5*, pp. 227–296. New York: Oxford University Press.

Morris C, Tang R (2011). "Estimating Random Effects via Adjustment for Density Maximization." *Statistical Science*, **26**(2), pp. 271–287. ISSN 08834237. URL http://www.jstor.org/stable/23059992.

Rönnegård L, Shen X, Alam M (2010). "hglm: A Package for Fitting Hierarchical Generalized Linear Models." *The R Journal*, **2**(2), 20–28. ISSN 20734859.

Rubin DB (1981). "Estimation in Parallel Randomized Experiments." *Journal of Educational Statistics*, **6**(4), pp. 377–401. ISSN 03629791. URL http://www.jstor.org/stable/1164617.

Affiliation:

Joseph Kelly
Department of Statistics
Harvard University
1 Oxford Street, Cambridge, MA
E-mail: kelly2@fas.harvard.edu

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

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

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