



## **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 Schools' Data

Education Testing Service conducted randomized experiments in eight separate schools and obtained this dataset. It contains the coaching effects on SAT scores ( $y_j, j = 1, \dots, 8$ ) and standard errors ( $se_j, j = 1, \dots, 8$ ) 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 with unknown mean and known standard error:  $y_j | \mu_j \stackrel{ind}{\sim} \text{Normal}(\mu_j, se_j^2)$ ,  $j = 1, \dots, 8$ . **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 ( $z_j, j = 1, \dots, 31$ ) and total number of patients receiving CABG surgeries (case load) in each hospital ( $n_j, j = 1, \dots, 31$ ). 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 will make sense to assume the number of deaths in each hospital has independent Poisson distribution with unknown parameter:  $z_j | \lambda_j \stackrel{ind}{\sim} \text{Poisson}(n_j \lambda_j)$ ,  $j = 1, \dots, 31$ , where  $n_j$  can also 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 players through their first 45 official at-bats of the 1970 season (Efron and Morris, 1975). Also, it has one binary covariate indicating whether a player was in the National league (1) or the American league (0). The code below shows a way to make inputs using the first 10 players.

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

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

### 3. Multi-level Structure

Hierarchical model is a very powerful tool to (justification of hierarchical model for hierarchical data??)

**gbp**, one of functions in **Rgbp**, fits such a hierarchical model whose first-level hierarchy is a distribution of observed data and second-level is 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,  $\sigma_j^2$  below is assumed to be known or to be accurately estimated, and subscript  $j$  indicates  $j$ -th school in the dataset.

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

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

where  $\mu_{0j} = x'_j \beta$ ,  $j = 1, \dots, 8$ , and a parameter with a zero sub-index represents it is a mean parameter of the prior (second-level) distribution. Based on this conjugate prior distribution, it is easy to derive corresponding posterior distribution.

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

where  $B_j \equiv \frac{\sigma_j^2}{\sigma_j^2 + A}$ ,  $j = 1, \dots, 8$  is called a shrinkages and we will explain it in the following section.

#### 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 multiplied to the notation representing Gamma distribution is a scale and a square bracket below indicates [mean, variance] of distribution.

$$z_j | \lambda_j \stackrel{ind}{\sim} \text{Poisson}(n_j \lambda_j),$$

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

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

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

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

### 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.

$$z_j | p_{0j} \stackrel{ind}{\sim} \text{Binomial}(n_j, p_j),$$

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

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

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

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

### 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 a non-informative distribution on second-level parameters to let the data speak more about their estimation.

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

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

## 4. Estimation

### 4.1. Shrinkage Estimation

Estimating shrinkage is a key part of our work because as we can see posterior mean is a linear function of shrinkage and posterior variance is also a linear (Gaussian), quadratic (Poisson), or cubic (Binomial) function of shrinkage. Once we estimate it, we can approximate the posterior moments given only data, for example,  $E(\mu_j | \mathbf{y})$  for Gaussian model.

### 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.*,  $B_j \equiv \frac{\sigma_j^2}{\sigma_j^2 + A} = B_j(A)$  for Gaussian and  $B_j \equiv \frac{r}{r + n_j} = B_j(r)$  for Poisson and Binomial models.

In this case, one way is to obtain an MLE of the variance component with its asymptotic Normal distribution and to use Delta method for asymptotic Normal distribution of shrinkage. But is the Normal approximation a good approximation for shrinkage that takes on values between 0 and 1?

Here Adjustment for Density Maximization (Morris and Tang, 2011), called ADM, comes. It assumes the Beta distribution for shrinkage, which the authors believe is a much better approximation for it, and estimates its posterior moments, *i.e.*,  $E(B_j|\text{data})$  and  $Var(B_j|\text{data})$ .

### 4.3. Approximation to Posterior Distribution via Matching Moments

After estimating two posterior moments, `gbp` reasonably approximates posterior distribution given only data, *i.e.*,  $(\mu_j|\text{data})$ ,  $(\lambda_j|\text{data})$ , or  $(p_j|\text{data})$ , by matching two moments with its parameters. The reason we said ‘reasonably approximates’ is that it is hard to find a closed form of posterior distribution given only data unless it is a Gaussian model. So, we assumed  $(p_j|\text{data})$  has also  $Beta(a_1, a_0)$  distribution and matched two estimated moments,  $E(p_j|\text{data})$  and  $Var(p_j|\text{data})$ , with two parameters,  $a_1$  and  $a_0$ , of this Beta distribution.

## 5. Method Check

### 5.1. Pseudo-data Generation Process

### 5.2. Unbiased Coverage Probability

### 5.3. Rao-Blackwellized Coverage Probability

## 6. Example

### 6.1. Known Prior Mean

Using 10 hospital data with PR

### 6.2. Unknown Prior Mean and No Covariate

Using 8 school data with GR

### 6.3. Unknown Prior Mean and One Covariate

Using 18 baseball data with BR

## 7. Discussion

## 8. Acknowledgments

## 9. Reference

- Efron, B. and Morris, C. (1975). Data Analysis Using Stein's Estimator and its Generalizations. *Journal of the American Statistical Association*. **70**. 311-319.
- Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*. **6**. 377-401.
- Morris, C. and Tang, R. (2011). Estimating Random Effects via Adjustment for Density Maximization. *Statistical Science*. **26**. 271-287.
- Morris, C. and Lysy, M. (2012). Shrinkage Estimation in Multilevel Normal Models. *Statistical Science*. **27**. 115-134.

### Affiliation:

Joseph Kelly  
 Department of Statistics  
 Harvard University  
 1 Oxford Street, Cambridge, MA  
 E-mail: [kelly2@fas.harvard.edu](mailto: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](mailto:morris@fas.harvard.edu)

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