Communications of the Korean Statistical Society
2010, Vol. 17, No. 6, 845–852
DOI: 10.5351/CKSS.2010.17.6.845

Accuracy Measures of Empirical Bayes Estimator for Mean Rates

Kwang Mo Jeong[1,a]

[a]Department of Statistics, Pusan National University

Abstract

The outcomes of counts commonly occur in the area of disease mapping for mortality rates or disease rates. A Poisson distribution is usually assumed as a model of disease rates in conjunction with a gamma prior. The small area typically refers to a small geographical area or demographic group for which very little information is available from the sample surveys. Under this situation the model-based estimation is very popular, in which the auxiliary variables from various administrative sources are used. The empirical Bayes estimator under Poisson-gamma model has been considered with its accuracy measures. An accuracy measure using a bootstrap samples adjust the underestimation incurred by the posterior variance as an estimator of true mean squared error. We explain the suggested method through a practical dataset of hitters in baseball games. We also perform a Monte Carlo study to compare the accuracy measures of mean squared error.

Keywords: Disease rate, Poisson-gamma model, inverse dispersion parameter, negative binomial, empirical Bayes, small area estimation, mean squared error, bootstrap sample.

1. Introduction

A study on mortality or disease rates to display the geographical variability of disease is very popular in epidemiological research. The outcomes of counts commonly occur in the area of disease mapping for mortality rates or disease rates. The small area typically refers to a small geographical area or demographic group for which very little information is available from sample surveys.

Model based estimation under small samples has received a considerable importance in recent years. Small area(or domain) generally refers to a subgroup of a population from which samples are drawn. The usual direct estimators for a small area are based on data from only the sample units in the area and are likely to yield unacceptably large standard errors due to small sizes. This makes it necessary to borrow strength from related areas to find more accurate estimates for a given area. For example, the synthetic estimator formally described by Gonzalez (1973) is traditionally used for small area estimation because of its simplicity and applicability to general sampling designs with increased accuracy in estimation. We may refer to Ghosh and Rao (1994) for general discussions about small area estimation.

In this paper we focus our attention to the empirical Bayes(EB) method to model count data. Efron and Morris (1975) were the first to apply EB method based on the idea of pooling information across areas to reduce the total mean squared error(MSE). A Poisson distribution is usually assumed to model disease rates in conjunction with gamma prior. Clayton and Kaldor (1987) proposed EB estimation procedures using a Poisson likelihood and gamma prior framework in testing for geographical excess

This work was supported for two years by a Pusan National University Research Grant.
[1] Professor, Department of Statistics, Pusan National University, Jangjeon-Dong, Kumjung-Gu, Pusan 609-735, Korea. E-mail: kmjung@pusan.ac.kr

risk. Similar work on disease rates has been done by many researchers such as Tsutakawa *et al.* (1985) and Marshall (1991). Although EB estimators are widely used in various settings, limited research has been done regarding the accuracy measures of the EB estimator. A naive EB approach measures its uncertainty by the estimated posterior variance. As discussed by Ghosh and Rao (1994) this measure can lead to severe underestimation of the true posterior variance under a prior distribution of related parameters. As remedies to this problem there has been significant research such as the jackknife method by Jiang *et al.* (2002), and also the Taylor expansion of MSE by Lahiri and Maiti (2002). We may refer to Rao (2003) for the general discussion on the small area estimation.

We suggest an alternative method using bootstrap samples to adjust the underestimation incurred by the posterior variance. Other bootstrapping techniques have been proposed by Laird and Louis (1987), Butar and Lahiri (2003) under different settings to account for the estimation of integrated Bayes risk.

2. Poisson-Gamma Model for Counts Data

We consider a population which are partitioned into N areas indexed by i, $i = 1, 2, \ldots, N$. We assume that the outcomes of our interest denote the number of events such as diseases recorded over a period of years or smaller subregions. Let n_i be the period (or number) of these years or subregions, which is the so called person-years in terms of disease mapping. In the literature of disease mapping the disease counts are modeled as Poisson variates with mean rate θ_i. If we let y_i be the cumulated counts over n_i units, then y_i follows a Poisson distribution with mean $n_i\theta_i$. In a Bayesian framework the mean rate θ_i itself is assumed to have a certain distribution. If we assume a gamma prior which is conjugate of Poisson distribution, this kind of framework is usually called a Poisson-gamma model. We simply denote the Poisson distribution with mean $n_i\theta_i$ by $y_i \,|\, \theta_i \sim \text{Poi}(n_i\theta_i)$.

A Poisson-gamma model can be formulated as

$$\theta_i = \mu_i\gamma_i, \tag{2.1}$$

where the error term γ_i follows a gamma distribution with shape parameter ϕ and scale $1/\phi$ and μ_i is the mean of θ_i which can be explained in terms of auxiliary variables. We note that the γ_i has mean 1 and variance $1/\phi$. The parameter ϕ is sometimes called an inverse dispersion parameter. Jeong and Yang (2009) discussed the problem of low means affecting the estimation of ϕ when sample sizes are small.

Let $x_1, \ldots, x_p$ be the auxiliary variables(covariates), and we assume that the mean μ_i is explained by a linear predictor

$$\eta(\boldsymbol{\beta}; \boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \tag{2.2}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$. Without loss of generality we assume the log link relationship given by

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}. \tag{2.3}$$

The joint density of y_i and θ_i is written as

$$\begin{aligned} f(y_i, \theta_i) &= \frac{e^{-n_i\theta_i}(n_i\theta_i)^{y_i}}{y_i!} \frac{\theta_i^{\phi-1} e^{-\theta_i\left(\frac{\phi}{\mu_i}\right)}}{\Gamma(\phi)} \left(\frac{\phi}{\mu_i}\right)^{\phi} \\ &= \frac{n_i^{y_i} (\phi/\mu_i)^{\phi}}{y_i!\Gamma(\phi)} \theta^{y_i+\phi-1} e^{-\theta_i\left(n_i+\frac{\phi}{\mu_i}\right)}, \end{aligned} \tag{2.4}$$

where $\Gamma(\cdot)$ is the Euler gamma function defined by $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$. By integrating out the joint density of (2.4) with respect to θ_i we obtain the marginal density of y_i. We remark that the joint density $f(y_i, \theta_i)$ and $f(y_i)$ depend on ϕ, μ_i and hence on $\boldsymbol{\beta}$ but we omit their dependence to simplify the notation. In a routine way we find the marginal of y_i by integrating out with respect to θ_i

$$f(y_i) = \frac{\Gamma(y_i + \phi)}{y_i!\Gamma(\phi)} \left(\frac{\phi}{n_i\mu_i + \phi}\right)^{\phi} \left(\frac{n_i\mu_i}{n_i\mu_i + \phi}\right)^{y_i}. \tag{2.5}$$

The marginal density in (2.5) denotes a negative binomial distribution. It is an easy task to find the conditional density of θ_i given y_i having the form

$$f(\theta_i \,|\, y_i) \propto \theta_i^{y_i+\phi-1} e^{-\theta_i\left(n_i + \frac{\phi}{\mu_i}\right)}. \tag{2.6}$$

The conditional density of (2.6) denotes the gamma distribution with shape $y_i + \phi$ and scale $\{n_i + (\phi/\mu_i)\}^{-1}$. The posterior mean and variance are respectively given by

$$E(\theta_i \,|\, y_i, \boldsymbol{x}_i) = \frac{y_i + \phi}{n_i + (\phi/\mu_i)} \tag{2.7}$$

and

$$\text{Var}(\theta_i \,|\, y_i) = \frac{y_i + \phi}{\{n_i + (\phi/\mu_i)\}^2}. \tag{2.8}$$

In a Bayesian approach we take the posterior mean of (2.7) as an estimator of θ_i when the parameters are assumed to be known.

3. Empirical Bayes Estimator

3.1. Estimation of parameters

To estimate the unknown parameters we discuss the maximum likelihood(ML) method applied to the marginal density of y_i in (2.5) using the Newton-Raphson algorithm. The marginal log-likelihood function of $\boldsymbol{y} = (y_1, \ldots, y_N)'$ denoted by $l(\phi, \boldsymbol{\beta} \,|\, \boldsymbol{y})$ can be rewritten as

$$l(\phi, \beta \,|\, \boldsymbol{y}) \propto \sum_{i=1}^{N} \left\{ \log\left(\frac{\Gamma(y_i + \phi)}{\Gamma(\phi)}\right) + \phi \log\left(\frac{\phi}{n_i\mu_i + \phi}\right) + y_i \log\left(\frac{n_i\mu_i}{n_i\mu_i + \phi}\right) \right\}. \tag{3.1}$$

We note that the first term on the right side of (3.1) can be simplified by using the relation

$$\log\left(\frac{\Gamma(y_i + \phi)}{\Gamma(\phi)}\right) = \sum_{h=0}^{y_i-1} \left\{ \log\left(1 + \frac{h}{\phi}\right) + \log(\phi) \right\}. \tag{3.2}$$

If we apply the Newton-Raphson algorithm based on the gradient elements of log-likelihood function the ML estimates of ϕ and β can be obtained in an iterative way. The subroutine function such as NLPTR in SAS can be implemented by PROC IML.

By substituting the estimates of parameters into (2.7) we propose an EB estimator of θ_i given by

$$\hat{\theta}_i^{EB} = \frac{y_i + \hat{\phi}}{n_i + \left(\hat{\phi}/\hat{\mu}_i\right)}, \tag{3.3}$$

where $\hat{\mu}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip})$ from the relationship in (2.3). On the other hand, the naive direct ML estimator of θ_i, based only on y_i, is given by $\tilde{\theta}_i^{ML} = y_i/n_i$. The EB estimator $\hat{\theta}_i^{EB}$ in (3.3) can be written in terms of both the mean $\hat{\mu}_i = \eta(\hat{\boldsymbol{\beta}}; \boldsymbol{x}_i)$ and the ML estimator $\tilde{\theta}_i^{ML}$ as

$$\begin{aligned}\hat{\theta}_i^{EB} &= \left(\frac{\hat{\phi}}{n_i\hat{\mu}_i + \hat{\phi}}\right)\hat{\mu}_i + \left(\frac{n_i\hat{\mu}_i}{n_i\hat{\mu}_i + \hat{\phi}}\right)\tilde{\theta}_i^{ML} \\ &= \delta_i\hat{\mu}_i + (1-\delta_i)\tilde{\theta}_i^{ML},\end{aligned} \tag{3.4}$$

where $\delta_i = \hat{\phi}/(n_i\hat{\mu}_i + \hat{\phi})$ with $0 < \delta_i < 1$. As we see in (3.4) when $\hat{\phi}$ is relatively small compared to n_i or $\hat{\mu}_i$ the $\hat{\theta}_i^{EB}$ is more shrunken to $\tilde{\theta}_i^{ML}$.

3.2. Accuracy measures

Steffey and Kass (1991) conjectured that the MSE of EB estimator is approximately equal to the posterior variance. By substituting the unknown parameters $\hat{\phi}$ and $\hat{\boldsymbol{\beta}}$ into (2.8) we find the estimated posterior variance, hereafter denoted by V_i^{EB}, is of the form

$$V_i^{EB} = \frac{n_i\tilde{\theta}_i^{ML} + \hat{\phi}}{\left(n_i + \hat{\phi}/\hat{\mu}_i\right)^2}. \tag{3.5}$$

As discussed by Ghosh and Rao (1994) this posterior variance fails to take account of the uncertainty about the parameters $\boldsymbol{\beta}$ and ϕ in which the form of their prior distribution is not specified in the EB approach unlike in the HB approach. We note that

$$\mathrm{Var}(\theta_i\,|\,\boldsymbol{y}) = E_{\boldsymbol{\beta},\phi}[\mathrm{Var}(\theta_i\,|\,y_i,\boldsymbol{\beta},\phi)] + \mathrm{Var}_{\boldsymbol{\beta},\phi}[E(\theta_i\,|\,y_i\,,\boldsymbol{\beta},\phi)], \tag{3.6}$$

where $E_{\boldsymbol{\beta},\phi}$ and $\mathrm{Var}_{\boldsymbol{\beta},\phi}$ respectively denote the expectation and variance with respect to the posterior distribution of $\boldsymbol{\beta}$ and ϕ given the data $\boldsymbol{y}$. The accuracy measure V_i^{EB} in (3.5) is a good approximation only to the first term of the right side of (3.6), but the second variance term is ignored in the naive EB approach.

We suggest an alternative accuracy measure of $\hat{\theta}_i^{EB}$ by the bootstrap method. Let $(\boldsymbol{x}^{*(b)}, \boldsymbol{y}_i^{*(b)})$ be the b^{th} bootstrap sample generated by the following parametric procedure. Based on the given dataset $(\boldsymbol{x}, \boldsymbol{y})$ the bootstrapped outcome variate $y_i^{*(b)}$ is generated from a Poisson distribution having mean $n_i\hat{\theta}_i^{EB}$ and the covariate $\boldsymbol{x}^{*(b)}$ is taken to equal the given $\boldsymbol{x}$. Based on this bootstrap sample we find the b^{th} EB estimator, denoted by $\hat{\theta}_i^{(b)EB}$, having the same form as (3.3), where $b = 1, 2, \ldots, B$ and B denote the number of bootstrap replications. Define the average of bootstrap estimators over all B replications as

$$\hat{\theta}_i(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_i^{(b)EB}. \tag{3.7}$$

We propose a bootstrap based accuracy measure of EB estimator by the relationship

$$V_i^{Boot} = \frac{1}{B}\sum_{b=1}^{B}\mathrm{Var}\left(\theta_i\,|\,y_i^{*(b)}\right) + \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_i^{(b)EB} - \hat{\theta}_i(\cdot)\right)^2. \tag{3.8}$$

The second term on the right side of (3.8) has been added to adjust the underestimation by the posterior variance of (3.5) as an accuracy measure. The bootstrap estimator is promising even if further studies

Table 1: Data of hitters against pitcher Mike Mussina

Hitter	At-bats	Hits	Estimators			Accuracy Measures		
			$\tilde{\theta}_i^{ML}$	$\hat{\theta}_i^{EB}$	$\hat{\theta}^{HB}$	V_i^{ML}	V_i^{EB}	V_i^{Boot}
R. Hidalgo	11	7	0.636	0.662	0.278	0.05785	0.05596	0.11195
A. Cintron	5	3	0.600	0.657	0.263	0.12000	0.11272	0.21364
B. Roberts	26	14	0.535	0.553	0.296	0.02071	0.02060	0.04193
R. Ibanez	21	10	0.476	0.496	0.279	0.02268	0.02273	0.04584
F. Catalanotto	56	26	0.464	0.472	0.312	0.00829	0.00831	0.01663
R. White	11	5	0.455	0.493	0.265	0.04132	0.04166	0.07917
M. Huff	5	2	0.400	0.485	0.256	0.08000	0.08326	0.17691
F. Thomas	78	30	0.385	0.391	0.299	0.00493	0.00496	0.01053
P. Burrell	10	3	0.300	0.353	0.253	0.03000	0.03265	0.06569
J. Canseco	61	18	0.295	0.305	0.264	0.00484	0.00493	0.00987
B.J. Surhoff	40	10	0.250	0.265	0.250	0.00625	0.00650	0.01338
A. Soriano	8	2	0.250	0.320	0.250	0.03125	0.03628	0.07884
H. Baines	35	7	0.200	0.218	0.240	0.00571	0.00610	0.01282
T. Hafner	10	2	0.200	0.261	0.247	0.02000	0.02411	0.05176
C. Fielder	42	7	0.167	0.183	0.231	0.00397	0.00427	0.00891
S. Posednick	6	1	0.167	0.268	0.247	0.02778	0.03919	0.08835
B. Mueller	23	0	0.000	0.035	0.214	0.00000	0.00146	0.00455
J. Kent	6	0	0.000	0.121	0.240	0.00000	0.01773	0.05075

on its performance in frequentist sense are needed. As a comparison we finally define the accuracy measure of ML estimator $\tilde{\theta}_i^{ML}$ based on Poisson distribution by

$$V_i^{ML} = \frac{\tilde{\theta}_i^{ML}}{n_i}. \tag{3.9}$$

3.3. An illustrative example

The first three columns of Table 1 shows the record of 18 selected hitters against Mike Mussina who has been pitching for 16 years. He faced 576 batters 5 or more times through July 23, 2006. The dataset comes from Stern and Sugano (2007) in which a hierarchical beta-binomial approach has been used to model the results of batter-pitcher matchups. The number of observed trials is quite small for any given batter-pitcher combination. Baseball fans and even baseball professionals have a tendency to draw stroong conclusions based on these small samples without considering the variability in the ability of a pitcher across different batters. There is considerable variability in the outcomes for different hitters ranging from Muller who has 0 hits in 23 attempts to Hidalgo who has 7 hits in 11 attempts.

From column 4 to column 9 we list the estimators of hitting rates for each player with their accuracy measures. As a comparison the hierarchical Bayes estimator $\hat{\theta}^{HB}$ under beta-binomial model, which was obtained by Stern and Sugano (2007), has been added in column 6. The estimated abilities of hits by the empirical Bayes method varies according to the observed averages and sample sizes, *i.e.*, at-bats. We note that Mueller's lack of success in 23 attempts is reflected by the EB estimate of success probability 0.035 than the estimate 0.121 for Kent who has had no hits but only 6 attempts thus far. The estimated accuracy measure by the bootstrap method is about 2 times larger than the posterior variance. We also note that accuracy of direct ML estimator is very similar to the posterior variance by the EB estimator.

Table 2: The MSE and their estimates when sample sizes are 5 or 10.

ϕ	Area no.	Sample size	MSE		Estimates of MSE		
			MSE($\tilde{\theta}_i^{ML}$)	MSE($\hat{\theta}_i^{EB}$)	mse($\tilde{\theta}_i^{ML}$)	mse($\hat{\theta}_i^{EB}$)	mse^{Boot}
3.0	4	10	0.11186	0.10363	0.1229	0.09896	0.16545
	8	10	0.11772	0.10317	0.1143	0.09123	0.15238
	12	10	0.11084	0.09064	0.1256	0.09957	0.16769
	16	10	0.25388	0.19072	0.2258	0.15021	0.21919
	20	5	0.27667	0.18210	0.2498	0.17047	0.25498
	24	5	0.27013	0.20615	0.2369	0.16007	0.23762
	28	5	0.21844	0.15280	0.2369	0.15582	0.22636
5.0	4	10	0.11468	0.09323	0.1169	0.07956	0.11606
	8	10	0.12300	0.09703	0.1216	0.08545	0.12290
	12	10	0.14794	0.11882	0.1107	0.07762	0.11417
	16	10	0.25172	0.17163	0.2526	0.13598	0.16706
	20	5	0.23008	0.14537	0.2458	0.14006	0.17976
	24	5	0.28415	0.20063	0.2166	0.11923	0.15220
	28	5	0.24918	0.14825	0.2486	0.13599	0.17123
7.0	4	10	0.12676	0.08268	0.1201	0.07533	0.10234
	8	10	0.11767	0.08267	0.1179	0.07275	0.09743
	12	10	0.11816	0.08891	0.1226	0.07798	0.10707
	16	10	0.22672	0.13656	0.2255	0.10710	0.12128
	20	5	0.18766	0.10571	0.2227	0.10386	0.11667
	24	5	0.24471	0.13090	0.2461	0.11957	0.13534
	28	5	0.24768	0.13437	0.2282	0.10994	0.12538

4. A Monte Carlo Study

We consider a single covariate and a simple linear predictor of the form $\eta(\boldsymbol{\beta}; x) = \beta_0 + \beta_1 x$. The covariate values are generated from uniform distribution over the interval $(-1, 1)$. We take the values of coefficients as $\beta_0 = 0$, $\beta_1 = 1$. The random error γ_i in the relationship (2.1) is taken from a gamma distribution with shape ϕ and scale $1/\phi$. We consider several ϕ values of $\phi = 3.0, 5.0, 7.0$. Finally, the outcome variate y_i has been generated from Poi$(n_i\theta_i)$, where $\theta_i = \mu_i\gamma_i$ and $\log(\mu_i) = \beta_0 + \beta_1 x$. Sample sizes of the first design are 5 or 10 and those of second design are 3, 5 and 7 according to area numbers. Both the number of Monte Carlo iterations and bootstrap replications are taken as $R = 500$ and $B = 500$, respectively. The number of areas is $N = 30$.

We define the approximate true MSE of an estimator of Poisson mean over iterations of simulation. For an EB estimator $\hat{\theta}_i^{EB}$ the true MSE is defined by

$$\mathrm{MSE}\left(\hat{\theta}_i^{EB}\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_{i(r)}^{EB} - \theta_i\right)^2.$$

Similarly, we can define the true MSE of $\tilde{\theta}_i^{ML}$, which is denoted by MSE($\tilde{\theta}_i^{ML}$). To compare the accuracy measures of MSE over iterations we define the estimated mean squared error. Firstly, the naive MSE estimate of $\hat{\theta}_i^{EB}$ is defined by

$$\mathrm{mse}\left(\hat{\theta}_i^{EB}\right) = \frac{1}{R}\sum_{r=1}^{R} V_{i(r)}^{EB},$$

where $V_{i(r)}^{EB}$ is an accuracy measure given by (3.5) at the r^{th} iteration of simulation. Similarly, the MSE estimate for the $\tilde{\theta}_i^{ML}$ is denoted by mse($\tilde{\theta}_i^{ML}$), but we simply use a notation mse^{Boot} as an estimate of MSE using bootstrap samples.

Table 3: The MSE and their estimates when sample sizes are 3, 5 or 7.

ϕ	Area no.	Sample size	MSE		Estimates of MSE		
			MSE($\tilde{\theta}_i^{ML}$)	MSE($\hat{\theta}_i^{EB}$)	mse($\tilde{\theta}_i^{ML}$)	mse($\hat{\theta}_i^{EB}$)	mse^{Boot}
3.0	4	7	0.16803	0.13900	0.17629	0.12589	0.19078
	8	7	0.17359	0.14171	0.16200	0.11524	0.17236
	12	5	0.23700	0.19411	0.23104	0.14844	0.20761
	16	5	0.23069	0.15800	0.24496	0.16380	0.24169
	20	5	0.26224	0.19548	0.23608	0.15710	0.23045
	24	3	0.35609	0.23632	0.39978	0.22258	0.29130
	28	3	0.35049	0.20310	0.38600	0.20529	0.26029
5.0	4	7	0.16939	0.10932	0.16661	0.09929	0.12634
	8	7	0.11571	0.09905	0.15510	0.09157	0.11643
	12	5	0.29023	0.18707	0.23424	0.12209	0.14821
	16	5	0.21735	0.14232	0.22312	0.12350	0.15076
	20	5	0.28201	0.17692	0.23736	0.12786	0.15989
	24	3	0.39753	0.21148	0.40822	0.17254	0.18718
	28	3	0.32953	0.17749	0.36644	0.14948	0.15892
7.0	4	7	0.14034	0.09705	0.16384	0.08230	0.10038
	8	7	0.16321	0.10604	0.16498	0.08315	0.09897
	12	5	0.23447	0.12937	0.22408	0.09618	0.10584
	16	5	0.23121	0.13240	0.23616	0.10457	0.11711
	20	5	0.22813	0.13648	0.23536	0.10297	0.11198
	24	3	0.38204	0.16442	0.37911	0.11866	0.11685
	28	3	0.40735	0.19232	0.40178	0.13544	0.13418

We listed the MSEs of $\tilde{\theta}_i^{ML}$ and $\hat{\theta}_i^{EB}$ with their corresponding estimates according to sample sizes and values of ϕ in Table 2 and Table 3 The values of MSE($\tilde{\theta}_i^{ML}$) are larger than those of MSE($\hat{\theta}_i^{EB}$) regardless of ϕ and sample sizes. The MSEs and their estimates have a tendency to decrease as ϕ and sample sizes increase. Large value of ϕ means that the variability between areas are smaller and hence we expect more accurate estimation. The estimates mse^{Boot} based on bootstrap samples are about 50% larger than the mse($\hat{\theta}_i^{EB}$) which in general underestimates the true MSE($\hat{\theta}_i^{EB}$).

5. Conclusion

We considered an EB estimator of mean rate for count data under Poisson-gamma model. The unknown parameters can be estimated by the ML method which may be implemented through a subroutine used in statistical packages such as SAS. As a method to adjust the underestimation incurred by the posterior variance we also suggested an alternative accuracy measure based on bootstrap samples.

We explained the suggested method through a practical example of hitting data in baseball games. We designed an experiment to do a small scale Monte Carlo study according to various values of inverse dispersion parameter and sample sizes. The bootstrap estimator approximates the true MSE more closely in contrast to the underestimation by the posterior variance. In this paper we have not studied the performance of a bootstrap method compared to other approaches such as the jackknife or Tayor expansion of MSE.

Other kinds of models such as a beta-binomial could be compared with the Poisson-gamma model in explaining the mean rates of events. An empirical best linear unbiased estimator or hierarchical Bayes estimators under various models will be good alternatives to the proposed method.

References

Butar, F. B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimator, *Journal of Statistical Planning and Inference*, **112**, 63–76.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics*, **43**, 671–681.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations, *Journal of the American Statistical Association*, **70**, 311–319.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal, *Statistical Science*, **9**, 55–93.

Gonzalez, M. E. (1973). Use and evaluation of synthetic estimators, *In Proceedings of the Social Statistics Section*, 33–36.

Jeong, K. M. and Yang, H. R. (2009). Dispersion parameter of Poisson-Gamma model in the small area estimation, *Journal of the Korean Data Analysis Society*, **11**, 23–32.

Jiang, J., Lahiri, P. and Wan, S. M. (2002). A unified jackknife theory for empirical best prediction with M-estimation, *The Annals of Statistics*, **30**, 1782–1810.

Lahiri, P. and Maiti, T. (2002). Empirical Bayes estimation of relative risks in disease mapping, *Calcutta Statistical Association Bulletin*, **53**, 213–223.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples, *Journal of the American Statistical Association*, **82**, 739–750.

Marshall, R. J. (1991). Mapping disease and mortality rates using empirical Bayes estimators, *Applied Statistics*, **40**, 283–294.

Rao, J. N. K. (2003). *Small Area Estimation*, Wiley.

Steffey, D. and Kass, R. E. (1991). Comment on Robinson, G. K., "That BLUP is a Good Thing – The estimation of random effects," *Statistical Science*, **6**, 45–47.

Stern, H. S. and Sugano, A. (2007). Inference about batter-pitcher matchups in baseball from small samples. In: Albert, J. and Koning, R.H., *Statistical Thinking in Sports*, Chapman & Hall, 153–165.

Tsutakawa, R. K., Shoop, G. L. and Marienfeld, C. J. (1985). Empirical Bayes estimation of cancer disease rates, *Statistics in Medicine*, **4**, 201–212.

Received August 2010; Accepted October 2010