Hierarchical Bayes estimation of mortality rates for disease mapping

Tapabrata Maiti

*Department of Mathematics and Statistics, University of Nebraska-Lincoln, Lincoln,
NE 68588-0323, USA*

Abstract

Mapping of incidence rates or mortality rates (relative risks) from diseases like cancer and leukemia is of primary importance in an epidemiological study. The usual procedure is to map the standardized mortality ratio (SMR) across different geographical regions. Direct use of SMR may not be worthwhile, particularly for small places, as it does not take into account the high variability for different population sizes over different regions and the spatial patterns of the regions under study. In this paper a hierarchical Bayes approach is presented in smoothing the relative risks and providing the measures of uncertainty associated with these estimates of relative risks. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Disease mapping; Empirical Bayes; Gibbs sampling; Hierarchical Bayes; Hyperparameter; Metropolis-Hastings; Relative risk; Small area

1. Introduction

One of the primary tools of an epidemiological study is mapping of incidence or mortality rates from diseases like cancer. Mapping of relative risks over the different geographical regions helps to have an idea of environmental determinants of a specific disease. As mentioned by Clayton and Kaldor (1987), the basic problem of mapping is the choice of an appropriate measure of relative risks. The inappropriateness of direct use of standardized mortality ratio (SMR) is discussed in Wallin (1984), Clayton and Kaldor (1987), Tsutakawa *et al.* (1985), Tsutakawa (1988), Cressie and Chan (1989) and Cressie (1992). The SMR suffers from high variability due to unequal bases from region to region. The crude estimates of relative risks for small areas are subject to large chance fluctuations due to lower level of aggregation of samples in the population. Tsutakawa (1988) and Manton *et al.* (1989) mentioned the adjustment of relative risks for small-area populations.

Most of the statistical techniques for smoothing the relative risks use empirical Bayes (EB) method. The basic idea from Efron and Morris (1975) is pooling the information across the regions through a suitable model using James–Stein (1961) estimators. In the EB approach, the posterior distributions of the parameters of interest given the data are first obtained assuming the model parameters are known. The model parameters are then estimated by suitable methods and inferences are made from the estimated posterior distributions (see Morris, 1983). Clayton and Kaldor (1987) and Manton *et al.* (1989) use Poisson likelihood and gamma prior for the parameters. Tsutakawa (1988) uses Poisson–gamma model with an additional random effect term in the mixed linear model. Clayton and Kaldor (1987) also propose non-parametric prior distribution which would be estimated by Laird's (1978) maximum likelihood approach. Cressie and Chan (1989) and Cressie (1992) use Gaussian model for both likelihood and prior to produce EB estimates of relative risks. Clayton and Kaldor (1987), Cressie and Read (1989) and Marshall (1991) consider the spatial patterns of regional data. One major problem of the EB approach is its failure to take into account a measure of uncertainty of the estimates. The estimated posterior variance does not take care of the extra variability due to estimation of the model parameters.

In the present article we consider hierarchical Bayes (HB) approach for estimating relative risks along with their measure of uncertainties. In the HB approach, together with the prior distributions of the parameters, suitable prior for the hyperparameters (or model parameters) are proposed and then inferences are made from the posterior distributions. In particular, a parameter is estimated by its posterior mean and its posterior variance is taken as the measure of error of the estimate. The HB method is easy to understand but often involves high-dimensional integration. We use Gibbs sampling in finding out the posterior densities.

The subsequent sections are as follows. In Section (2) we consider two hierarchical Bayes models and list the corresponding full conditional distributions required to carry out the Gibbs sampling. In Section (3) the HB methodology is illustrated by using the example of lip cancer data given in Clayton and Kaldor (1987). Concluding remarks are given in Section (4). HB analysis is often pursued with improper prior for the hyperparameters. This may lead to improper posterior distribution. In the appendix we establish the propriety of the posterior distribution corresponding to an improper hyperprior distribution.

2. Hierarchical Bayes modeling for relative risk estimation

In this section we consider two different modeling approaches as considered in Clayton and Kaldor (1987), Manton *et al.* (1989), Cressie and Chan (1989) and Cressie (1992). First, we consider Poisson likelihood, log-normal prior and then we consider spatial modeling of log-relative risks. For both cases we use proper but vague prior.

2.1. Poisson log-normal model

Suppose there are D regions under consideration. The regions may be counties or districts. Let θ_i be the relative risk (incidence rate) of a disease for area i $(i = 1, \ldots, D)$. The problem is to find a suitable estimate for θ_i $(i = 1, \ldots, D)$ and the measure of error of the estimate. Let y_i denote the observed number of incidence in region i $(i = 1, \ldots, D)$ which is usually observed over a number of years. Let E_i be the number of persons-years at risk in area i $(i = 1, \ldots, D)$. A crude measure of relative risk like SMR is defined as $x_i = y_i / E_i$ for $i = 1, \ldots, D$. Though E_i's are estimated from age-specific death rates, in the present context we consider that E_i's are known. Let, conditional on θ_i, y_i's be independent Poisson random variables with parameter $(\theta_i E_i)$ and β_i's are independent and identical (iid) normal variables with mean μ and variance r^{-1} where $\beta_i = \log \theta_i$. Manton et al. (1989) considered gamma distributions for θ_i's $(i = 1, \ldots, D)$. To find out the empirical Bayes estimates Clayton and Kaldor (1987), Manton et al. (1989), Marshall (1991) and Cressie (1992) suggested various methods of estimation of the hyperparameters. Lahiri and Maiti (1996) propose the optimal estimating function approach. Marshall (1991) discussed the difficulties arising in iterative methods of estimation and proposed non-iterative ANOVA-type estimates for the prior mean and variance. Another shortcoming of EB approach is that the measure of uncertainty of the EB estimate based on the estimated posterior distribution underestimates the true measure as the naive EB approach fails to account for the uncertainty due to estimating the model parameters.

To avoid the difficulties that arise in EB approach, one may adopt the HB approach at the cost of some extra computation. In standard hierarchical Bayesian analysis, in the absence of any subjective prior, improper non-informative priors are used for the hyperparameters which may lead to improper posterior distributions. For choice of non-informative priors, see Berger (1985). For data analysis we consider proper, but vague, prior for the hyperparameters as considered by Arora and Lahiri (1996). We use the hierarchical models as follows.

Model MI:

(I) $y_i | \theta_i \sim$ Poisson $(\theta_i E_i)$; independently, $i = 1, \ldots, D$;

(II) $\log \theta_i | (\mu, r) \overset{\text{ind}}{\sim} N(\mu, r^{-1})$, $i = 1, \ldots, D$;

(III) Marginally μ and r are independently distributed with $\mu \sim$ Uniform $(-\infty, +\infty)$, $r \sim$ Gamma$(a/2, b/2)$, $a > 0$, $b > 0$,

where a random variable x is said to have Gamma(n, p) distribution if it has the p.d.f

$$f(x) = \frac{p^n}{\Gamma(n)} \exp(-px) x^{n-1}, \quad x > 0.$$

Instead of a proper gamma prior for r one may use a diffused prior (see Datta and Ghosh, 1991). We assume both a and b are known. Our objective is to use the posterior distribution of θ_i's given the observations y_i, $i = 1, \ldots, D$ to make inference about these parameters. If there is any covariate available in the data, one could easily incorporate it in modeling of θ_i's (see Ghosh et al., 1997).

The posterior distributions involve high-dimensional integration. To overcome this problem we use Gibbs sampling (see Geman and Geman, 1984; Gelfand and Smith, 1990). We estimate the posterior distributions using multiple-path iterative sampling, recommended by Gelman and Rubin (1992) as it provides a measure (potential scale reduction) to check the convergence of Gibbs sampling. We generate $m(\geqslant 2)$ independent parallel chains, each of length $2t$ with starting values from an overdispersed distribution. Then we remove the first t iterated observations from each of the m chains to elliminate the effects of initial choices. From the remaining t iterated observations we find the posterior means and variances and the criterion of monitoring the convergence of Gibbs sampling.

We now write down the full-conditional distributions to implement the Gibbs sampling.

(i) $[\theta_i|(\underset{\sim}{y},r,\mu)] \propto \theta_i^{y_i-1}\exp[-\theta_i E_i - \frac{r}{2}(\log\theta_i - \mu)^2], \quad i = 1,\ldots,D,$

(ii) $\mu|(\underset{\sim}{\theta},\underset{\sim}{y},r) \sim N(\frac{1}{D}\sum_{i=1}^{D}\log\theta_i,(Dr)^{-1}),$

(iii) $r|(\underset{\sim}{\theta},\underset{\sim}{y},\mu) \sim \text{Gamma}[\frac{1}{2}(D+a),\frac{1}{2}\{\sum_{i=1}^{D}(\log\theta_i - \mu)^2 + b\}].$

The conditional distribution (i) is known only upto a multiplicative constant. Though the generation of samples are easy for the distributions (ii) and (iii), generation of samples from (i) needs special effort. One cannot use the adaptive rejection sampling of Gilks and Wild (1992) as the distribution (i) is not log-concave. This situation is handled by Metropolis–Hastings algorithm recommended by Chib and Greenberg (1995). Metropolis–Hastings was originally developed by Metropolis et al. (1953) and was generalized by Hastings (1970). We assume that at least one y_i is strictly positive so that the posterior distribution corresponding to improper prior for μ is proper (proof is in the appendix).

2.2. Log-normal model with spatial effect

In this section we consider the conditional autoregression (CAR) model for the log-relative risks (see Clayton and Kaldor, 1987). Cressie and Chan (1989) use Gaussian modeling with random Markov field property. Let S_i denote the set of neighbourhood areas of area i $(i = 1,\ldots,D)$. Recall $\beta_i = \log\theta_i$ $(i = 1,\ldots,D)$. The CAR model is defined as

$$E(\beta_i|\beta_j, j \in S_i) = \mu_i + \sum_{j \in S_i} C_{ij}h_i(\beta_j), \quad i = 1,\ldots,D.$$

Here μ_i's are large-scale variation and C_{ij}'s $j \in S_i$ are small-scale variation of spatial dependence model (see Cressie and Chan, 1989). We take $\mu_i = \mu$ $(i = 1,\ldots,D)$ and $C = ((C_{ij})) = \rho W = \rho((W_{ij}))$. Here W is the adjacency matrix of the map and W_{ij}'s depend on the location of ith region $(i = 1,\ldots,D)$. Consequently, $C_{ij} = C_{ji}, C_{ij} = 0$ for $j \notin S_i$ in our model. We also define $C_{ii} = 0$. Cressie and Chan (1989) have made some general assumptions about C. Let $h_i(\beta_j) = \beta_j - \mu$ and $\text{Var}(\beta_i|\beta_j, j \in S_i) = v^{-1}$.

Let $(\lambda_{max})^{-1}$ denote the maximum value of ρ in the CAR model. Then we assume the following hierarchical model.

Model MII:

(I) $y_i|\theta_i \sim \text{Poisson}(\theta_i E_i)$, independently, $i = 1, \ldots, D$;

(II) $\beta_i|\beta_j, j \in S_i, \rho, v \sim N[\mu + \rho \sum_{j \in S_i}(\beta_j - \mu), v^{-1}]$, $i = 1, \ldots, D$.

(III) ρ, μ and v are mutually independently distributed, respectively, with $\rho \sim U(0, 1/\lambda_{max})$, $\mu \sim \text{Uniform}(-\infty, +\infty)$ and $v \sim \text{Gamma}(g, h), g > 0, h > 0$.

We assume both g and h are known. Here also we have used non-informative prior. As in Section (2.1) we use Gibbs sampling to obtain the posterior distribution of θ_i, $i = 1, \ldots, D$. The full conditional distributions are as follows:

(i) $[\theta_i|(\underset{\sim}{y}, v, \mu, \rho, \theta_j, j \in S_i)] \propto \theta_i^{y_i-1} \exp[-\theta_i E_i - \frac{r}{2}(\log \theta_i - A_i \mu - B_i)^2]$, $i = 1, \ldots, D$,

(ii) $\mu|(\underset{\sim}{\theta}, \underset{\sim}{y}, v, \rho) \sim N\left(\dfrac{\sum_{i=1}^D A_i(\log \theta_i - B_i)}{\sum_{i=1}^D A_i^2}, (v \sum_{i=1}^D A_i^2)^{-1}\right)$,

(iii) $v|(\underset{\sim}{\theta}, \underset{\sim}{y}, \mu, \rho) \sim \text{Gamma}[\frac{1}{2}(D + a), \frac{1}{2}\{\sum_{i=1}^D (\log \theta_i - A_i \mu - B_i)^2 + b\}]$,

(iv) $\rho|(\underset{\sim}{\theta}, \underset{\sim}{y}, \mu, v) \sim \text{TN}\left(\dfrac{\sum_{i=1}^D k_i d_i}{\sum_{i=1}^D k_i^2}, (r \sum_{i=1}^D k_i^2)^{-1}\right)$,

where TN denotes the truncated normal distribution. The posterior distribution (iv) is truncated outside the range $(0, 1/\lambda_{max})$. $A_i = 1 - \rho \sum_{j \in S_i} W_{ij}$, $B_i = \rho \sum_{j \in S_i} W_{ij} \log \theta_j$, $d_i = \log \theta_i - \mu$, $k_i = \sum_{j \in S_i} W_{ij} d_j$. In this case also the posterior distribution (i) is known only upto a multiplicative constant. The conditional distribution (i) is not log-concave. So, we use the Metropolis–Hastings algorithm to generate samples from (i). The posterior distributions are proper, assuming at least one y_i is strictly positive.

3. An illustration with lip cancer data

In this section a HB analysis is carried out in the example of Clayton and Kaldor (1987). The observed (y) and the expected (E) number of cases of lip cancer registered during the six years period (1975–1980) for each of 56 counties of Scotland are given. We consider the first 54 counties as we obtain the 'expected' number of incidence E_i for ith county using the relation $E_i = y_i/x_i, i = 1, \ldots, D$, where x_i denotes the SMR for county i, $i = 1, \ldots, D$. Here x_i's are multiplied by 100. The y_i's are zero for the last two counties. We exclude those counties as E_i's become zero. The adjacency matrix W is obtained from the last column of Table 1 of Clayton and Kaldor (1987). Our data analysis is presented in Table 1 below. The Table 1, column 4 corresponds to the estimates from Model MI and column 5 corresponds to the estimates from Model MII. The standard errors (s.e.) of the estimates are given in columns 6 and 7, respectively, for Model MI and Model MII.

To implement the Gibbs sampler of Gelman and Rubin (1992), we consider $m = 10$ independent sequences each of length $2t = 200$. The convergence of Gibbs sampler is assessed by using the potential scale reduction of Gelman and Rubin (1992) which

Table 1
Hierarchical Bayes estimates of lip cancer incidence in Scotland by county

County	y	x	HB estimate		Standard error		Adjacent counties
			MI	MII	MI	MII	
1	9	652.2	458.3	413.3	163.0	137.9	5,9,11,19
2	39	450.3	424.6	404.7	70.1	66.2	7,10
3	11	361.8	303.0	292.0	92.6	81.9	6,12
4	9	355.7	285.9	256.1	96.8	75.6	18,20,28
5	15	352.1	304.6	314.5	80.9	83.7	1,11,12,13,19
6	8	333.3	270.4	269.2	96.1	81.2	3,8
7	26	320.6	302.5	301.5	56.7	58.0	2,10,13,16,17
8	7	304.3	249.3	238.2	92.0	72.9	6
9	6	303.0	239.1	254.1	92.1	86.9	1,11,17,19,23,29
10	20	301.7	275.6	277.8	61.8	59.3	2,7,16,22
11	13	295.5	259.4	278.8	70.3	70.3	1,5,9,12
12	5	279.3	225.4	259.4	93.3	86.7	3,5,11
13	3	277.8	201.3	245.0	105.9	91.2	5,7,17,19
14	8	241.7	216.3	189.2	75.1	54.9	31,32,35
15	17	216.8	202.6	189.9	45.9	44.2	25,29,50
16	9	197.8	180.5	194.5	58.8	50.0	7,10,17,21,22,29
17	2	186.9	161.6	209.7	84.5	76.6	7,9,13,16,19,29
18	7	167.5	156.5	164.5	55.0	47.5	4,20,28,33,55,56
19	9	162.7	152.8	190.7	47.1	48.0	1,5,9,13,17
20	7	157.7	149.2	166.6	52.0	45.0	4,18,55
21	16	153.0	147.0	148.9	34.6	32.4	16,29,50
22	31	136.7	136.0	142.3	22.8	22.7	10,16
23	11	125.4	125.7	118.2	34.8	31.9	9,29,34,36,37,39
24	7	124.6	126.9	100.2	42.8	31.2	27,30,31,44,47,48,55,56
25	19	122.8	122.7	127.7	28.2	24.4	15,26,29
26	15	120.1	119.7	118.5	29.4	26.6	25,29,42,43
27	7	115.9	118.6	123.8	39.5	35.1	24,31,32,55
28	10	111.6	115.4	121.4	33.9	31.9	4,18,33,45
29	16	111.3	111.5	109.0	27.3	24.9	9,15,16,17,21,23,25,26,34,43,50
30	11	107.8	110.2	95.1	29.8	25.1	24,38,42,44,45,56
31	5	105.3	111.4	96.8	43.1	33.6	14,24,27,32,35,46,47
32	3	104.2	110.7	123.7	49.6	43.9	14,27,31,35
33	7	99.6	101.7	112.2	31.9	31.2	18,28,45,56
34	8	93.8	99.3	71.2	31.8	22.4	23,29,39,40,42,43,51,52,54
35	11	89.3	92.8	94.8	25.3	22.7	14,31,32,37,46
36	9	89.1	94.3	93.4	27.8	24.3	23,37,39,41
37	11	86.8	89.4	89.0	23.5	22.3	23,35,36,41,46
38	8	85.6	91.1	80.6	28.9	23.1	30,42,44,49,51,54
39	6	83.3	91.3	86.2	30.4	26.7	23,34,36,40,41
40	4	75.9	87.0	73.6	32.8	25.1	34,39,41,49,52
41	10	53.3	59.6	55.0	15.1	14.2	36,37,39,40,46,49,53
42	8	50.7	59.7	60.9	17.0	15.9	26,30,34,38,43,51
43	2	46.3	74.7	84.2	34.6	30.0	26,29,34,42
44	6	41.0	51.3	54.6	16.9	14.1	24,30,38,48,49
45	19	37.5	42.3	47.9	9.1	9.5	28,30,33,56
46	3	36.6	52.1	55.8	19.6	18.0	31,35,37,41,47,53
47	2	35.8	62.2	52.0	27.0	18.6	24,31,46,48,49,53
48	3	32.1	50.1	52.8	19.0	17.0	24,44,47,49
49	28	31.6	34.2	32.1	6.4	5.8	38,40,41,44,47,48,52,53,54

Table 1. (Continued)

County	y	x	HB estimate		Standard error		Adjacent counties
			MI	MII	MI	MII	
50	6	30.6	42.5	58.2	13.8	15.7	15,21,29
51	1	29.1	65.7	67.6	37.6	24.0	34,38,42,54
52	1	27.6	64.5	60.9	37.1	25.5	34,40,49,54
53	1	17.4	51.5	53.9	28.3	20.5	41,46,47,49
54	1	14.2	50.8	48.3	22.1	17.9	34,38,49,51,52

Note: The reported standard errors are scale transformed.

is near unity for all the scalar estimands. The procedure is not sensitive to the initial choices of the parameters.

The estimates of θ_i's $(i = 1, \ldots, D)$ are very similar for both the models. The dispersions among the estimates are slightly larger than those for the EB estimates. We use $(\lambda_{max})^{-1} = 0.175$ which is the maximum value of ρ in a CAR process (see Clayton and Kaldor, 1987). In our case the estimate of ρ comes out as 0.164. The spatial effects are reflected from the estimates of some of the counties e.g., the counties 24, 30, 31 are affected by several low-risk areas while the counties 17, 32 and 33 are affected by several high-risk areas.

The standard errors of the estimates for Model II are smaller than those for Model I. This is because Model II exploits the spatial structure of the data. For the counties 1, 13, 17, 32, 43, 46, 47, 48, 51, 52, 53, and 54 the standard errors are high due to insufficient representation of obsereved number of incidences. However, the counties 2, 7, 22, 45, and 49 have small standard errors due to availability of more information.

The values of the potential scale reduction of Gelman and Rubin (1992) are close to unity for all the scalar estimands.

4. Concluding remarks

This paper presents the HB estimation procedure for an ensemble of parameters related to mapping of relative risks across several non-overlapping regions. We have used non-informative priors for hyperparameters but one could also use informative priors to incorporate prior information of the hyperparameters, if available. In the EB method if one uses the estimated posterior variance as a measure of uncertainty for the estimate, it leads to severe underestimation of the true posterior variance since it does not take into account the uncertainty involved in the estimation of mean and variance of prior parameters (see Berger 1985; Prasad and Rao, 1990; Lahiri and Rao, 1995). Unlike the EB method, the HB method accounts for this uncertainty by assigning the distributions of prior parameters. Moreover, the HB approach provides standard errors along with the point estimates (see Ghosh et al., 1998).

Though a number of models have been suggested in the literature, we have considered only two of them to demonstrate the HB method in estimating relative risks and the measures of uncertainty of the estimates.

Acknowledgements

Appendix

We will show that the joint posterior distribution of the parameters is proper. We have to show that $[(\underline{\theta}, \mu, r)/\underline{y}]$ is integrable.

Case I: $y_i > 0$ for all i. Let $u_i = \log \theta_i$, $i = 1, \ldots, D$.

Then we have

$$[(\underline{u}, \mu, r)/\underline{y}] \propto \exp \left\{ - \sum_{i=1}^{D} (E_i e_i^{u_i} - y_i u_i) \right\}$$

$$\times \exp \left\{ -\frac{r}{2} \left(\sum_{i=1}^{D} (u_i - \mu)^2 + b \right) \right\} r^{((D+a)/2)-1}$$

$$\int \cdots \int \exp \left\{ - \sum_{i=1}^{D} (E_i e_i^{u_i} - y_i u_i) \right\}$$

$$\times \exp \left\{ -\frac{r}{2} \left(\sum_{i=1}^{D} (u_i - \mu)^2 + b \right) \right\} r^{((D+a)/2)-1} \, dr \, d\mu \prod du_i$$

$$= \text{const.} \int \cdots \int \exp \left\{ - \sum_{i=1}^{D} (E_i e_i^{u_i} - y_i u_i) \right\}$$

$$\times \exp \left\{ -\frac{r}{2} \left(\sum_{i=1}^{D} (u_i - \bar{u})^2 + b \right) \right\} r^{(D+a-1)/2} \, dr \prod du_i$$

$$= \text{const.} \int \cdots \int \exp \left\{ - \sum_{i=1}^{D} (E_i e^{u_i} - y_i u_i) \right\}$$

$$\times \left\{ b + \sum_{i=1}^{D} (u_i - \bar{u})^2 \right\}^{-(D+a+1)/2} \prod du_i$$

assuming $a + D > 0, b > 0$.

Now, for $y_i > 0$,

$$\exp(y_i u_i - E_i e_i^{u_i}) < \exp(-c u_i^2) \quad \text{for large } u_i > 0, c > 0$$

and

$$\exp(y_i u_i - E_i e_i^{u_i}) < \exp(y_i u_i) \quad \text{for small } u_i < 0.$$

Then it follows that the posterior is proper.

Case II: At least one y_i is strictly positive.

Assume p of the y_i's are strictly positive and without loss of generality let $y_1 > 0, \ldots,$ $y_p > 0$. Then we have

$$\int \cdots \int [(\underline{\theta}, \mu, r)/\underline{y}]\, dr\, d\mu \prod d\theta_i$$

$$\propto \int \cdots \int \prod_{i=1}^{p} \{\exp(-\theta_i E_i)\theta_i^{y_i-1}\exp\{-\tfrac{r}{2}(\log\theta_i - \mu)^2\}r^{1/2}\} \prod_{i=1}^{p} d\theta_i$$

$$\times \int \cdots \int \prod_{i=p+1}^{D} \{\exp(-\theta_i E_i)\theta_i^{y_i-1}\exp\{-\tfrac{r}{2}(\log\theta_i - \mu)^2\}r^{r/2}\} \prod_{i=p+1}^{D} d\theta_i$$

$$\times \exp(-\tfrac{br}{2})r^{a/2-1}\, d\mu\, dr.$$

Now,

$$\int \cdots \int \prod_{i=p+1}^{D} \{\exp(-\theta_i E_i)\theta_i^{y_i-1}\exp\{-\tfrac{r}{2}(\log\theta_i - \mu)^2\}r^{r/2}\} \prod_{i=p+1}^{D} d\theta_i$$

$$\leqslant \int \cdots \int \prod_{i=p+1}^{D} \{\theta_i^{-1}\exp\{-\tfrac{r}{2}(\log\theta_i - \mu)^2\}r^{1/2}\} \prod_{i=p+1}^{D} d\theta_i$$

$$= (2\pi)^{(D-p)/2}$$

So,

$$\int \cdots \int [(\underline{\theta}, \mu, r)/\underline{y}]\, dr\, d\mu \prod d\theta_i$$

$$\leqslant (2\pi)^{(D-p)/2}$$

$$\times \int \cdots \int \prod_{i=1}^{p} \{\exp(-\theta_i E_i)\theta_i^{y_i-1}\exp\{-\tfrac{r}{2}(\log\theta_i - \mu)^2\}r^{1/2}\} \prod_{i=1}^{p} d\theta_i$$

$$\times \exp(-\tfrac{br}{2})r^{a/2-1}\, d\mu\, dr$$

$$< \infty \quad \text{(Follows from case I).}$$

References

Arora, V., Lahiri, P., 1997. On the superiority of the Bayesian method over the BLUP in small area estimation problems. Statistica Sinica 7, 1053–1063.

Berger, O.J., 1985. Statistical decision theory and Bayesian analysis. Springer, New York.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings Algorithm. Amer. Statist. 49, 327–335.

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

Cressie, N., 1992. Smoothing regional maps using empirical Bayes predictors. Geograph. Anal. 24, 75–95.

Cressie, N., Chan, N.H., 1989. Spatial modeling of regional variables. J. Amer. Statist. Assoc. 84, 393–401.

Cressie, N., Read, T.R.C., 1989. Spatial data analysis of regional counts. Biometrical J. 31, 699–719.

Datta, G.S., Ghosh, M., 1991. Bayesian prediction in linear models: application to small area estimation. Ann. Statist. 19, 1746–1770.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Gelman, A.E., Rubin, D., 1992. Inference from iterative simulation (with discussion). Statist. Sci. 7, 457–511.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6, 721–741.

Ghosh, M., Natarajan, K., Stroud, T.W.F., Carlin, B.P., 1996. Generalized linear models for small area estimation. J. Amer. Statist. Assoc., to appear.

Ghosh, M., Natarajan, K., Kim, D., Waller, L.A., 1997. Hierarchical bayes GLM's for the analysis of spatial data: an application to disease mapping. Research Rep., University of Florida.

Hastings, W.K., 1970. Monte Carlo sampling methods using Marcov chains and their applications. Biometrika 57, 1317–1340.

James, W., Stein, C., 1961. Estimation with quadratic loss. In: Proc. 4th Berkeley Symp., vol. 1, University of California Press, Berkeley, CA, pp. 361–379.

Lahiri, P., Maiti, T., 1996. Empirical Bayes estimation of mortality from diseases for small area. Invited paper in Statistical Research in the 21st Century, Montreal, Canada.

Lahiri, P., Rao, J.N.K., 1995. Robust estimation of mean squared error of small area estimators. J. Amer. Statist. Assoc. 90, 758–766.

Laird, N.M., 1978. Nonparametric maximum likelihood estimation of a mixing distribution. J. Amer. Statist. Assoc. 73, 805–811.

Manton, K.G., Woodbury, M.A., Stallard, E., Riggan, W.B., Creason, J.P., Pellom, A.C., 1989. Empirical Bayes procedures for stabilizing maps of US cancer mortality rates. J. Amer. Statist. Assoc. 84, 637–50.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Morris, C., 1983. Natural exponential families with quadratic variance functions: Statistical theory. Ann. Statist. 11, 515–529.

Marshall, R.J., 1991. Mapping disease and mortality rates using empirical Bayes estimators. Appl. Statist. 40, 283–294.

Prasad, N.G.N., Rao, J.N.K., 1990. On the estimation of mean square error of small area predictors. J. Amer. Statist. Assoc. 85, 163–171.

Tsutakawa, R.K., 1988. Mixed model for analysing geographic variability in mortality rates. J. Amer. Statist. Assoc. 83, 37–42.

Tsutakawa, R.K., Shoop, G.L., Marienfeld, C.J., 1985. Empirical Bayes estimation of cancer mortality rates. Statist. Med. 4, 201–212.

Wallin, E., 1984. Isarithmic maps and geographical disaggregation. In: Proc. Internat. Symp. on Spatial Data Handling. Zurich, Switzerland, 1984, pp. 209–217.

Gilks, Wild, 1992. Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41, 337–348.

Effron, B., Morris, C., 1975. Data analysis using Stein's estimator and its generalizations. J. Amer. Statist. Assoc. 70, 311–319.