Bayesian Simultaneous Intervals for Small Areas: An Application to Variation in Maps

Erik B. Erhardt[1], Balgobin Nandram[2] & Jai Won Choi[3]

[1] Department of Mathematics and Statistics, 1 University of New Mexico, Albuquerque, New Mexico

[2] Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA, USA

[3] Statistical Consulting Service, 9504 Mary Knoll Drive, Rockville, MD, USA

Correspondence: Balgobin Nandram, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, USA. Tel: 1-508-831-5539. E-mail: balnan@wpi.edu

Abstract

Bayesian inference about small areas is of considerable current interest, and simultaneous intervals for the parameters for the areas are needed because these parameters are correlated. This is not usually pursued because with many areas the problem becomes difficult. We describe a method for finding simultaneous credible intervals for a relatively large number of parameters, each corresponding to a single area. Our method is model based, it uses a hierarchical Bayesian model, and it starts with either the $100(1-\alpha)\%$ (e.g., $\alpha = .05$ for 95%) credible interval or highest posterior density (HPD) interval for each area. As in the construction of the HPD interval, our method is the result of the solution of two simultaneous equations, an equation that accounts for the probability content, $100(1-\alpha)\%$ of all the intervals combined, and an equation that contains an optimality condition like the "equal ordinates" condition in the HPD interval. We compare our method with one based on a nonparametric method, which as expected under a parametric model, does not perform as well as ours, but is a good competitor. We illustrate our method and compare it with the nonparametric method using an example on disease mapping which utilizes a standard Poisson regression model.

Keywords: hierarchical Bayesian model, highest posterior density, Metropolis-Hastings algorithm, Nelder-Mead algorithm, Rao-Blackwellized estimator, simultaneous inference

1.Introduction

Simultaneous inference on parameters is an important problem; see Miller (1981). In Bayesian analysis the parameters are generally correlated, and so simultaneous interval estimation is of interest. For example, in Bayesian small area estimation in which a parameter indexes each area, the parameters are usually correlated, and simultaneous intervals should be reported. This is not usually done because the problem is generally difficult. We propose a method to construct $100(1-\alpha)\%$ (e.g., $\alpha = .05$ for 95%) simultaneous credible intervals for many small areas, and we apply our method to study variation in a problem on disease mapping.

This is not a new problem. By 1955 three principal investigators (Duncan, 1952; Scheffé, 1953; Tukey, 1953) brought the general principles of multiple comparisons into their current structure; see Miller (1981, p. 2). The basic technique of multiple comparisons can be separated into two groups: those which can provide confidence intervals (confidence regions), and those which provide tests of hypotheses. Our study contributes to the first of these groups. Simultaneous intervals based on standard normal theory form an optimal ellipsoid (Box & Tiao, 1973).

Why do we need simultaneous intervals? Consider two parameters μ_1 and μ_2. Let a 95% credible interval (CI) for μ_1 be (a_1, b_1) and a 95% CI for μ_2 be (a_2, b_2). Then the intersection $(a_1, b_1) \cap (a_2, b_2)$ does not form a set giving a 95% credible interval; in fact, it is at most 95%. Thus, we need to lengthen these individual intervals in an optimal manner. For n events A_i, $i, \ldots, n$, Bonferroni's inequality, $\Pr\left(\cap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \Pr(A_i) - (n-1)$, allows us to provide a lower bound to the probability of a simultaneous event (the intersection) in terms of the probabilities of the individual events (Miller, 1981, p. 8). For example, for just two 97.5% credible intervals, the simultaneous coverage has a lower bound of .95, and with 798 credible intervals, each with probability content 95%, the simultaneous

coverage of all intervals has a meaningless lower bound of −38.9! Thus, this method gives a meaningful result when the number of events is small and the probabilities of the individual events are sufficiently large. While the well known methods of Bonferroni, Tukey, Scheffé and others are reasonable for simultaneous intervals of a moderate number of parameters, they become overly conservative for a large number of parameters. Also, these methods are limited by their specific distributional assumptions.

Our method for constructing $100(1 − \alpha)\%$ simultaneous credible intervals starts with the individual $100(1 − \alpha)\%$ credible intervals or highest posterior density (HPD) intervals. Let $f(\theta|\underline{d})$ denote the posterior density of a parameter θ given data $\underline{d}$. An interval (a, b) is called a $100(1 − \alpha)\%$ credible interval if its posterior probability content is $1 − \alpha$; that is, $\int_a^b f(\theta|\underline{d})\, d\theta = 1 − \alpha$. Clearly, $100(1 − \alpha)\%$ credible intervals are not unique. There are two ways to construct credible intervals: numerical and sampling-based. Let $F(\theta|\underline{d}) = \int_{-\infty}^{\theta} f(t|\underline{d})\, dt$ be the cumulative distribution function (cdf), and let $F^{-1}(.|\underline{d})$ be the inverse cdf. Then $a = F^{-1}\left(\frac{\alpha}{2}|\underline{d}\right)$ and $b = F^{-1}\left(1 − \frac{\alpha}{2}|\underline{d}\right)$ give the $100(1 − \alpha)\%$ equal-tailed credible interval (a, b). An alternative and standard method is to draw a random sample of 1,000 values from $f(\theta|\underline{d})$. Place the values in ascending order, $\theta^{(1)} < \theta^{(2)} < \ldots < \theta^{(1000)}$. Then an estimate from these order statistics of the 95% credible interval is $(\theta^{(25)}, \theta^{(976)})$. This method works well for large samples (e.g., 1,000). A $100(1 − \alpha)\%$ credible interval (a, b) is a HPD interval if for any $\theta_1 \in (a, b)$ and $\theta_2 \notin (a, b)$, $f(\theta_1|\underline{d}) \geq f(\theta_2|\underline{d})$ (i.e., the height at any point of the density within the HPD interval is greater than the height at any point outside the interval). The $100(1 − \alpha)\%$ HPD interval is the shortest interval with $100(1 − \alpha)\%$ coverage.

For a unimodal posterior density the $100(1 − \alpha)\%$ HPD interval is obtained by solving two equations. Letting $F(a|\underline{d}) = \int_{-\infty}^{a} f(\theta|\underline{d})\, d\theta$, where $f(\theta|\underline{d})$ is a unimodal posterior density with its mode not on the boundary and both tails approaching zero monotonically, then the equations,

$$f(a|\underline{d}) \quad = \quad f(b|\underline{d}) \tag{1}$$

$$F(b|\underline{d}) − F(a|\underline{d}) \quad = \quad \int_a^b f(\theta|\underline{d})\, d\theta = 1 − \alpha \tag{2}$$

provide the HPD interval (a, b). Conditions (1) and (2) can be expressed using a single equation, for example, by solving for a,

$$f(a|\underline{d}) = f(b|\underline{d}) \quad = \quad f\left[F^{-1}\{F(a) + (1 − \alpha)\}\right].$$

Nandram (1993) describes a method for constructing simultaneous cuboid intervals for predicting k new observations. There a Bayesian normal-normal model is used to provide intervals based on the multivariate Student's t distribution. These intervals form a simple cuboid which engineers use, instead of the relatively more complex optimal ellipse (Box & Tiao, 1973). The bounds on each interval are obtained by solving a pair of simultaneous equations. The first equation satisfies the simultaneous probability content by forcing the difference of the values of the cumulative density function evaluated at the lesser interval bound from the greater interval bound to be $1 − \alpha$. The second equation satisfies optimality criterion of equal ordinates by forcing the difference of the values of the probability density function evaluated at the interval bounds to be zero. Thus the cuboids are optimized by constructing the smallest such k-dimensional cuboid by using HPD intervals in each dimension. However, this method considers only a small number of future observations. Also note that in this method the future observations have the same distribution, and they are correlated. Similarly, Nandram and Choi (2004) construct simultaneous concentration bands for quantile-quantile probability plots, accounting for the correlation of the order statistics and providing exact coverage probability. Comparisons of pointwise and Bonferroni concentration bands are given.

Besag et al. (1995) presents a method, henceforth BESAG, to calculate simultaneous credible regions based on order statistics (see Appendix A). The idea is to use samples drawn from the empirical distribution of each parameter of interest. The procedure is analogous to ordering each sample, counting from the minimum and maximum of each ordered sample a fixed number of ranks and use these order statistics to form the simultaneous interval. Because the method is nonparametric, it ignores the properties of the underlying distribution the sample was drawn from, but uses an assumption of symmetry. Therefore, the method conservatively makes the intervals wider than necessary. One would expect poor performance if there is a specific underlying parametric model.

There are a few attempts at simultaneous interval estimation in the context of small area estimation. Lui and Cumberland (1989) use simultaneous interval estimates in small domain estimation under the Bayesian paradigm. They use the Bonferroni method, the multivariate Student's t method and Scheffé's method and make comparisons.

Andrews and Birdsall (1988) compare three simultaneous confidence interval procedures: ordinary-χ^2, full-design, and Bayesian. They found that for their study, the Bayesian procedure has the best properties in terms of correct coverage with small average interval width having small variation over replications.

We use these $100(1 - \alpha)\%$ simultaneous intervals in an application on the variation in maps. Nandram, Sedransk and Pickle (2000) described a method to study variation in maps (see Section 4 of their paper) for the continental U.S. with 798 health service areas (HSAs). They use the 1,000 iterates of the 798 mortality rates from the output of a Markov chain Monte Carlo method to construct 1,000 maps. They study variation by finding the identity of the quantiles of a HSA in the mean map and over each of the 1,000 alternative maps. This method addresses directly the issue of how the apparent map patterns change, but it is difficult to present all the available information from 1,000 maps. Our idea is to construct three maps, the mean map (the one that is usually presented), and the end points of 95% simultaneous intervals (an upper map and a lower map) for all HSAs. We assess variation in the mean map estimator by the degree of similarity among the three maps.

The rest of the paper is as follows. In Section 2 we describe how to construct the $100(1-\alpha)\%$ simultaneous credible intervals. We describe two methods, one uses an equation on the probability content only, and the other has an additional optimality condition (i.e., an "equal ordinates" condition). In Section 3 we describe an application in disease mapping of 798 health service areas (a relatively large number of parameters for simultaneous inference). In Section 4 we describe a numerical example on chronic obstructive pulmonary disease of the 798 HSAs. We also compare our method with the nonparametric method (BESAG, see Besag et al., 1995); see Appendix A. Section 5 has concluding remarks, and we suggest future research.

2. Construction of Simultaneous Intervals

We propose to construct $100(1 - \alpha)\%$ simultaneous intervals by "stretching" individual HPD intervals until the desired content is obtained; this includes an optimality criterion. Unlike ellipsoidal regions, because the parameters are not identically distributed, the simultaneous intervals form a hyper-rectangle, not a cuboid.

Our methods require (a) the support of the individual posterior densities to be on the nonnegative part of the real line, (b) the posterior densities not to be flat in any interval, (c) the tails of the posterior densities to asymptote to zero monotonically, (d) the posterior density to be unimodal, and (e) the marginal posterior densities must be similar, especially in the tails. Although we do not consider a posterior density with its mode on the boundary of the parameter space, our method can be adjusted to meet this requirement. We consider a simple example. Let $\underline{x}$ be data, θ be the parameter of interest, and $\theta \mid \underline{x} \sim$ Gamma(δ, τ), where δ and τ are functions of the data $\underline{x}$. Then, this posterior density $f(\theta|\underline{x})$ is unimodal, provided $\delta > 1$, and the both tails satisfy (a)–(d), and when there are several posterior densities we expect them to be similar to this posterior density. In problems on mortality data analysis in a Bayesian model, the posterior densities of the mortalities typically satisfy our requirements. In our methods for simultaneous intervals, we assume that these conditions are met.

Generalizing the method for individual HPD intervals, we want to solve the system of equations,

$$\int_{a_\ell}^{b_\ell} \cdots \int_{a_1}^{b_1} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_\ell \quad = \quad 1 - \alpha, \tag{3}$$

$$f_1(a_1|\underline{d}) \quad = \quad f_1(b_1|\underline{d}),$$

$$\vdots$$

$$f_\ell(a_\ell|\underline{d}) \quad = \quad f_\ell(b_\ell|\underline{d}). \tag{4}$$

The combined probability content is ensured to be $100(1 - \alpha)\%$ by (3), and the equal ordinates conditions are ensured by (4). However, because we have nearly twice as many unknown quantities as the number of equations, a unique solution does not exist. Even if there were a unique solution, optimizing over such a large set of parameters (e.g., 798 in the disease mapping problem) is undoubtedly computationally demanding.

First, we relax the condition in (4) to obtain our first method, Method I. Here, our simplification is to provide one parameter to operate on the individual HPD intervals. Let $0 < \gamma \leq 1$ be a "stretching" factor on the HPD interval (a_i, b_i), giving the stretched interval higher content than the initial interval because

$$\int_{\gamma a_i}^{b_i/\gamma} f_i(\lambda_i|\underline{d}) \, \mathrm{d}\lambda_i \quad \geq \quad \int_{a_i}^{b_i} f_i(\lambda_i|\underline{d}) \, \mathrm{d}\lambda_i.$$

In this way we can begin with the $100(1 - \alpha)\%$ credible or HPD intervals (a_i, b_i) and stretch them until the desired

content is obtained. Thus, our objective is to solve the equation

$$\int_{\gamma a_\ell}^{b_\ell/\gamma} \cdots \int_{\gamma a_1}^{b_1/\gamma} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell = 1 - \alpha.$$

This can be obtained by determining the value for γ such that $\min_\gamma \Delta(\gamma)$ is zero, where

$$\Delta(\gamma) = \left| \int_{\gamma a_\ell}^{b_\ell/\gamma} \cdots \int_{\gamma a_1}^{b_1/\gamma} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell - (1 - \alpha) \right|, \tag{5}$$

and using the Nelder-Mead algorithm (Nelder & Mead, 1965) to force $\Delta(\gamma)$ in (5) down to zero. Clearly, the minimum of this function over γ is zero. We will call the method that starts with the individual credible intervals Method I-CRE, and the method that starts with the individual HPD intervals Method I-HPD.

The following argument shows that there is a unique solution to (5). Let $A_i' = \{\lambda_i : a_i < \lambda_i < b_i\}$ be the individual $100(1-\alpha)\%$ HPD credible interval defined on a set of densities with positive support, $\Pr(A_i'|\underline{d}) = 1-\alpha$, $i = 1, \ldots, \ell$. Let $A_i(\gamma) = \{\lambda_i : \gamma a_i < \lambda_i < b_i/\gamma\}$ be the stretched individual HPD interval with $\Pr(A_i(\gamma)|\underline{d}) \geq 1 - \alpha$, $i = 1, \ldots, \ell$, and $\Pr\left(\cap_{i=1}^\ell A_i(\gamma)|\underline{d}\right) = 1 - \alpha$ be the probability of the intersection. Observe that as γ decreases to 0, each $A_i(\gamma)$ is a sequence of increasing sets in γ so that $\cap_{i=1}^\ell A_i(\gamma)$ is a sequence of increasing sets in γ. So, letting $\underline{\lambda} = (\lambda_1, \ldots, \lambda_\ell)'$, by the continuity theorem,

$$\lim_{\gamma \to 0} \Pr\left(\cap_{i=1}^\ell A_i(\gamma)|\underline{d}\right) = \Pr\left(\lim_{\gamma \to 0} \cap_{i=1}^\ell A_i(\gamma)|\underline{d}\right) = \Pr\left(\underline{\lambda} \in \mathfrak{R}_+^\ell | \underline{d}\right) = 1.$$

That is, as γ goes to 0, the sequence of increasing sets $\cap_{i=1}^\ell A_i(\gamma)$ encompasses the entire space $\mathfrak{R}_+^\ell$, the positive ℓ-dimensional Euclidean space. Also, at $\gamma = 1$, $\Pr\left(\cap_{i=1}^\ell A_i(\gamma)|\underline{d}\right) < \Pr(A_i'|\underline{d}) = 1 - \alpha$. Thus, under our conditions, as $\gamma \to 0$, $\Pr\left(\cap_{i=1}^\ell A_i(\gamma)|\underline{d}\right)$ strictly increases from a value less than $1 - \alpha$ to 1. Therefore, there is a unique solution to $\Pr\left(\cap_{i=1}^\ell A_i(\gamma)|d_i\right) = 1 - \alpha$ for γ.

In Method II we simplify the problem in equations (3) and (4) to provide two parameters to operate on the individual HPD intervals. Let $0 < \gamma_1 \leq 1$ and $0 < \gamma_2 \leq 1$ be "stretching" factors on the HPD intervals (a_i, b_i). Then,

$$\int_{\gamma_1 a_i}^{b_i/\gamma_2} f_i(\lambda_i | \underline{d}) \, d\lambda_i \geq \int_{a_i}^{b_i} f_i(\lambda_i | \underline{d}) \, d\lambda_i.$$

In this way we can begin with the HPD intervals (or credible intervals) (a_i, b_i), and stretch them until the desired content is obtained together with an ordinate optimality criterion. Thus, our objective is to solve the set of two equations

$$\int_{\gamma_1 a_\ell}^{b_\ell/\gamma_2} \cdots \int_{\gamma_1 a_1}^{b_1/\gamma_2} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell = 1 - \alpha \tag{6}$$

$$f_i(\gamma_1 a_i | \underline{d}) = f_i(b_i/\gamma_2 | \underline{d}), \ i = 1, \ldots, \ell. \tag{7}$$

Note that we have two parameters and at least two equations, depending on how we wish to formulate our ordinate optimization criterion in equation (7).

The conditions in (7) are very tight because the distributions of $\lambda_i \mid \underline{d}$ can be very different, although we expect them not to be different. If given $\underline{d}$ the λ_i are exchangeable, the equations in (6) and (7) reduce to

$$\int_{\gamma_1 a_1}^{b_1/\gamma_2} \cdots \int_{\gamma_1 a_1}^{b_1/\gamma_2} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell = 1 - \alpha$$

$$f_1(\gamma_1 a_1 | \underline{d}) = f_1(b_1/\gamma_2 | \underline{d}).$$

This is similar to the situation for prediction in Nandram (1993), and it is easy to show that there is a unique solution (γ_1, γ_2) to this system of equations. However, in this case the assumption of exchangeability is unreasonable. Thus, there may be no (γ_1, γ_2) which satisfy (7). Therefore, although there are several values of (γ_1, γ_2) which satisfy (6), there may be no simultaneous solution to (6) and (7).

Thus, because we need the condition in (6), we will relax the condition in (7) to require

$$\int_{\gamma_1 a_\ell}^{b_\ell/\gamma_2} \cdots \int_{\gamma_1 a_1}^{b_1/\gamma_2} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell = 1 - \alpha \tag{8}$$

$$f_i(\gamma_1 a_i | \underline{d}) = f_i(b_i/\gamma_2 | \underline{d}) + e_i, \quad i = 1, \ldots, \ell, \tag{9}$$

where the e_i are small perturbations. However, we replace the conditions in (9) by

$$\ell^{-1} \sum_{i=1}^{\ell} f_i(\gamma_1 a_i | \underline{d}) = \ell^{-1} \sum_{i=1}^{\ell} f_i(b_i/\gamma_2 | \underline{d}) + \bar{e}, \text{ where } \bar{e} = \ell^{-1} \sum_{i=1}^{\ell} e_i.$$

Furthermore, for a reasonable small value of $\bar{e}$ we have

$$\ell^{-1} \sum_{i=1}^{\ell} f_i(\gamma_1 a_i | \underline{d}) \approx \ell^{-1} \sum_{i=1}^{\ell} f_i(b_i/\gamma_2 | \underline{d}).$$

In many applications we may have $\bar{e} \approx 0$, where given $\underline{d}$, the λ_i are nearly exchangeable. In fact, in our applications the λ_i are nearly exchangeable. Thus, our problem reduces to finding the solution (γ_1, γ_2) of

$$\int_{\gamma_1 a_\ell}^{b_\ell/\gamma_2} \cdots \int_{\gamma_1 a_1}^{b_1/\gamma_2} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell = 1 - \alpha \tag{10}$$

$$\ell^{-1} \sum_{i=1}^{\ell} f_i(\gamma_1 a_i | \underline{d}) \approx \ell^{-1} \sum_{i=1}^{\ell} f_i(b_i/\gamma_2 | \underline{d}). \tag{11}$$

Thus, we need to obtain the $\min_{\gamma_1, \gamma_2} \Delta(\gamma_1, \gamma_2)$, where

$$\Delta(\gamma_1, \gamma_2) = \left| \int_{\gamma_1 a_\ell}^{b_\ell/\gamma_2} \cdots \int_{\gamma_1 a_1}^{b_1/\gamma_2} f(\lambda_1, \ldots, \lambda_\ell | \underline{d}) \, d\lambda_1 \cdots d\lambda_\ell - (1 - \alpha) \right| + \left| \ell^{-1} \sum_{i=1}^{\ell} \left\{ f_i(\gamma_1 b_i | \underline{d}) - f_i(a_i/\gamma_2 | \underline{d}) \right\} \right|. \tag{12}$$

Under our general conditions, $\min_{\gamma_1, \gamma_2} \Delta(\gamma_1, \gamma_2) \approx 0$. Again, computation can be done by using the Nelder-Mead algorithm (Nelder & Mead, 1965) to force $\Delta(\gamma_1, \gamma_2)$ in (12) down to zero. Also, we will call the method that starts with the individual credible intervals Method II-CRE, and the method that starts with the individual HPD intervals Method II-HPD.

3. An Application on Disease Mapping

In this section we describe the disease mapping application in which there are 798 health service areas. We need the mean map, a lower bound map and an upper bound map so that the overall coverage is 95%. We use a hierarchical Bayesian model, a Poisson regression model, to study the mortalities over the 798 health service areas.

We briefly describe the model used by Nandram, Liu and Choi (2005), introduced by Christiansen and Morris (1997). Let n_i denote the population size, d_i denote the number of deaths and λ_i denote the mortality rate for the ith HSA, $i = 1, \ldots, \ell$, where $\ell = 798$. Then, the hierarchical Bayesian (Poisson regression) model is

$$d_i | \lambda_i \stackrel{ind}{\sim} \text{Poisson}(n_i \lambda_i), \tag{13}$$

$$\lambda_i | \alpha, \underline{\beta} \stackrel{ind}{\sim} \text{Gamma}\left(\alpha, \alpha e^{-\underline{x}_i' \underline{\beta}}\right), \qquad i = 1, \ldots, \ell, \tag{14}$$

where $\underline{x}_i$, a p-dimensional vector, are covariates with $x_{i1} \equiv 1$. Finally, the hyper-parameters α and $\underline{\beta}$ have proper prior densities

$$\pi(\alpha) = \frac{a_0}{(a_0 + \alpha)^2}, \qquad \alpha > 0 \text{ and } \underline{\beta} \sim \text{Normal}\left(\mu_{\underline{\beta}}, \Lambda_{\underline{\beta}}\right), \tag{15}$$

where a_0, $\mu_{\underline{\beta}}$ and $\Lambda_{\underline{\beta}}$ are to be specified. We have used the same noninformative specification of $\mu_{\underline{\beta}}$ and $\Lambda_{\underline{\beta}}$ as in Nandram, Liu, and Choi (2005) and we have taken $a_0 = 1$; see Albert (1988).

Using Bayes' theorem in (13), (14) and (15), the joint posterior density of all the parameters is

$$p(\underline{\lambda}, \alpha, \underline{\beta} | \underline{d}) \propto \frac{1}{(1 + \alpha)^2} \exp\left\{-\frac{1}{2}\left(\underline{\beta} - \mu_{\underline{\beta}}\right)' \Lambda_{\underline{\beta}}^{-1} \left(\underline{\beta} - \mu_{\underline{\beta}}\right)\right\} \times \prod_{i=1}^{\ell} \lambda_i^{d_i} e^{-n_i \lambda_i} \frac{\left(\alpha e^{-\underline{x}_i' \underline{\beta}}\right)^\alpha \lambda_i^{\alpha - 1} \exp\left\{-\left(\alpha e^{-\underline{x}_i' \underline{\beta}}\right) \lambda_i\right\}}{\Gamma(\alpha)}. \tag{16}$$

Nandram et al. (2005) showed how to use the Metropolis-Hastings algorithm (e.g., see Chib & Greenberg, 1995) to fit the model in (13), (14) and (15), and how to obtain samples from the posterior density in (16). At the suggestion of Gelman et al. (1996), they kept the jumping probabilities in all their applications in (.25, .50) by varying a tuning constant. They choose every tenth iterate to remove the autocorrelation to obtain $M = 1000$ iterates of $(\alpha, \underline{\beta})$ (i.e., $(\alpha^{(h)}, \underline{\beta}^{(h)})$, $h = 1, \ldots, M$).

In (16), conditional on $\alpha, \underline{\beta}$, and $\underline{d}$, the λ_i are independent gamma random variables with

$$g_i(\lambda_i | \alpha, \underline{\beta}, \underline{d}) = \frac{\left(n_i + \alpha e^{-x_i'\beta}\right)^{d_i+\alpha} \lambda_i^{d_i+\alpha-1} \exp\left\{-\left(n_i + \alpha e^{-x_i'\beta}\right)\lambda_i\right\}}{\Gamma(d_i + \alpha)}, \tag{17}$$

$\lambda_i \geq 0$, and the corresponding posterior distribution function is $G_i(a) = \int_0^a g_i(\lambda_i | \alpha, \underline{\beta}, \underline{d}) \, d\lambda_i$, $a \geq 0$. The conditional posterior density in (17) permits us to obtain Rao-Blackwellized estimators of the λ_i, thereby facilitating the construction of the maps. Also note that $g_i(\lambda_i | \alpha, \underline{\beta}, \underline{d})$ depend on i (i.e., the λ_i are not conditionally identically distributed as in Nandram (1993)).

Thus, we construct the posterior mean map using Rao-Blackwellized estimators for the λ_i. Letting $r_i = d_i/n_i$, $i = 1, \ldots, \ell$, denote the observed mortality rate and letting $\Lambda_i = n_i\left(n_i + \alpha e^{-x_i'\beta}\right)^{-1}$, the conditional posterior mean of λ_i is $E(\lambda_i | \alpha, \underline{\beta}, d_i) = \Lambda_i r_i + (1-\Lambda_i)e^{x_i'\beta}$. This is a weighted average of the observed mortality rate r_i and the prior mortality rate $e^{x_i'\beta}$. It follows that the posterior mean (unconditional) of λ_i is $E(\lambda_i | \underline{d}) = E_{\Omega | \underline{d}_i}\left\{\Lambda_i r_i + (1 - \Lambda_i)e^{x_i'\beta}\right\}$ and the Rao-Blackwellized estimator of λ_i is

$$\widehat{E(\lambda_i | \underline{d})} = M^{-1} \sum_{h=1}^{M} \left\{\Lambda_i^{(h)} r_i + \left(1 - \Lambda_i^{(h)}\right) e^{x_i'\beta^{(h)}}\right\},$$

where $\Lambda_i^{(h)} = n_i\left(n_i + \alpha^{(h)} e^{-x_i'\beta^{(h)}}\right)^{-1}$ and $\Omega^{(h)} = \left(\alpha^{(h)}, \underline{\beta}^{(h)}\right)$, $h = 1, \ldots, M$, are the M iterates obtained from the Metropolis-Hastings sampler. A map of these $\widehat{E(\lambda_i | \underline{d})}$ is an estimator of the mean map.

The individual credible intervals or HPD intervals are easy to obtain. To construct the HPD intervals (a_i, b_i), for the ith HSA, we minimize the function $\Delta(a_i, b_i)$ over (a_i, b_i), where

$$\Delta(a_i, b_i) = \left| M^{-1} \sum_{h=1}^{M} \left\{G_i\left(b_i | d_i, \Omega^{(h)}\right) - G_i\left(a_i | d_i, \Omega^{(h)}\right)\right\} - (1 - \alpha) \right|$$
$$+ \left| M^{-1} \sum_{h=1}^{M} g_i\left(a_i | d_i, \Omega^{(h)}\right) - M^{-1} \sum_{h=1}^{M} g_i\left(b_i | d_i, \Omega^{(h)}\right) \right|, \tag{18}$$

and $g_i(\cdot)$ and $G_i(\cdot)$ are given by (17). These individual intervals are to be used to construct the simultaneous intervals. To construct the credible intervals one does not need any optimization. For each $(\alpha^{(h)}, \underline{\beta}^{(h)})$, we generate one iterate $\lambda_i^{(h)}$, $i = 1, \ldots, \ell$, $h = 1, \ldots, M = 1000$ from (17). Then, the 95% credible interval for the ith area is $(\lambda_i^{(25)}, \lambda_i^{(976)})$, $i = 1, \ldots, \ell$. Thus, if the two methods lead to the same 95% simultaneous intervals, one can start with the credible intervals, thereby avoiding some computational effort.

We obtain our simple Method I for simultaneous intervals by finding the value of γ minimizing $\Delta(\gamma)$ over parameter γ, where

$$\Delta(\gamma) = \left| M^{-1} \sum_{h=1}^{M} \left[\prod_{i=1}^{\ell} \left\{G_i\left(b_i/\gamma | d_i, \Omega^{(h)}\right) - G_i\left(\gamma a_i | d_i, \Omega^{(h)}\right)\right\} \right] - (1 - \alpha) \right|. \tag{19}$$

We obtain the Method II with the optimality condition by minimizing the function $\Delta(\gamma_1, \gamma_2)$ over (γ_1, γ_2), where

$$\Delta(\gamma_1, \gamma_2) = \left| M^{-1} \sum_{h=1}^{M} \left[\prod_{i=1}^{\ell} \left\{G_i\left(b_i/\gamma_2 | d_i, \Omega^{(h)}\right) - G_i\left(\gamma_1 a_i | d_i, \Omega^{(h)}\right)\right\} \right] - (1 - \alpha) \right|$$
$$+ \left| \ell^{-1} \sum_{i=1}^{\ell} M^{-1} \sum_{h=1}^{M} g_i\left(\gamma_1 a_i | d_i, \Omega^{(h)}\right) - M^{-1} \sum_{h=1}^{M} g_i\left(b_i/\gamma_2 | d_i, \Omega^{(h)}\right) \right|. \tag{20}$$

Instead of (20) we have used

$$
\begin{aligned}
\tilde{\Delta}(\gamma_1, \gamma_2) \;=\; & \left| M^{-1} \sum_{h=1}^{M} \left[\prod_{i=1}^{\ell} \left\{ G_i\left(b_i/\gamma_2 \,|\, d_i, \Omega^{(h)}\right) - G_i\left(\gamma_1 a_i \,|\, d_i, \Omega^{(h)}\right) \right\} \right] - (1 - \alpha) \right| \\
& + \ell^{-1} \sum_{i=1}^{\ell} \left| M^{-1} \sum_{h=1}^{M} g_i\left(\gamma_1 a_i \,|\, d_i, \Omega^{(h)}\right) - M^{-1} \sum_{h=1}^{M} g_i\left(b_i/\gamma_2 \,|\, d_i, \Omega^{(h)}\right) \right|.
\end{aligned}
\tag{21}
$$

Note that by the triangle inequality (21) is a stronger condition. But, our results show very minor differences.

The minimizations in (18), (19) and (20) or (21) are obtained using the Nelder-Mead algorithm (Nelder & Mead, 1965). In Section 4 we apply these methods to study variation in maps using data on chronic obstructive pulmonary diseases.

4. A Numerical Example

We use data in Nandram, Liu and Choi (2005) for white males with chronic obstructive pulmonary diseases (COPD) in age classes 8, 9, and 10 (i.e., white males older than 65 years). There are 798 HSAs over the continental U.S., and we have the data on the population sizes and the number of deaths in these age classes combined. So our 95% simultaneous credible intervals consist of 798 intervals for the mortality rates, which form a hyper-rectangle. Four potential risk factors were selected in Nandram et al. (2000) for COPD, and used by Nandram, Liu and Choi (2005). These risk factors, used as covariates, help to explain the spatial patterns of COPD and its constituent diseases (asthma, chronic bronchitis, and emphysema). Here $x_0 \equiv 1$ is used for an intercept and the covariates are: x_1 is the white male lung cancer rate per 1,000 population; x_2 is the square root of (population density/10^4); x_3 is the square root of (elevation/10^4); and x_4 is (annual rainfall/100).

Using the model in Section 3, Nandram, Liu and Choi (2005) showed that the 95% credible intervals for each of the regression coefficients do not contain zero (i.e., all the covariates are significant). Nandram, Liu and Choi (2005) showed that inference about the mortality rates is not much sensitive to the specifications of a_0, $\mu_{\underline{\beta}}$ and $\Lambda_{\underline{\beta}}$. Using a cross-validation and a deviance measure, they also showed that the Poisson regression model fits reasonably well. Similar diagnostics are discussed in Nandram, Sedransk and Pickle (1999, 2000). Also, the 95% credible interval for α when all 798 HSAs are used is (19.5, 36.8); showing that the conditional posterior distributions of the λ_i are all unimodal.

Again, we fit the model described in Section 3 using the Metropolis-Hastings algorithm exactly as described by Nandram, Liu and Choi (2005) to obtain M iterates of $(\alpha, \underline{\beta})$ from the posterior density $p(\alpha, \underline{\beta}|\underline{d})$. In most situations we have taken $M = 1000$ to get the sample $\Omega^{(h)} = (\alpha^{(h)}, \underline{\beta}^{(h)})$, $h = 1, \ldots, M$. From these $\Omega^{(h)}$ we have obtained the posterior mean map and the 95% simultaneous credible intervals for the mortality rates λ_i, $i = 1, \ldots, 798$, as described in Section 3. All maps are colored with the same five-color monochromatic color scheme with light representing a low rate and dark representing a high rate. All intervals have 95% coverage. In the three maps made up by the simultaneous intervals, the mean map is always presented between the lower and upper maps as a basis of reference, and to study variation.

4.1 Comparisons

To compare our methods, we have studied the simultaneous probability contents of all methods, the values of γ in Method I and the values of (γ_1, γ_2) in Method II, and the maps based on the 95% simultaneous credible intervals. We note that the probability content is calculated under the Poisson regression model in Section 3 as follows. Let (a_i, b_i), $i = 1, \ldots, \ell$, be a set of intervals for the mortalities of the $\ell = 798$ HSAs. Then, the combined probability content of the hyper-rectangle formed by these intervals is

$$
C \;=\; \int_{\underline{\beta}} \int_{\alpha} \left[\prod_{i=1}^{\ell} \left\{ G_i(b_i|d_i, \Omega^{(h)}) - G_i(a_i|d_i, \Omega^{(h)}) \right\} \right] p(\alpha, \underline{\beta}|\underline{d}) \, d\alpha \, d\underline{\beta},
$$

where $G_i(\cdot|d_i, \Omega^{(h)})$ is given near (17) and $p(\alpha, \underline{\beta}|\underline{d})$ is the posterior density of $(\alpha, \underline{\beta})$ under the Poisson regression model. Thus, a Monte Carlo consistent estimator of C is

$$
\hat{C} \;=\; M^{-1} \sum_{h=1}^{M} \left[\prod_{i=1}^{\ell} \left\{ G_i\left(b_i/\gamma \,|\, d_i, \Omega^{(h)}\right) - G_i\left(\gamma a_i \,|\, d_i, \Omega^{(h)}\right) \right\} \right].
$$

For all our methods $M = 1000$ suffices.

The simultaneous probability content by region for the BESAG method (Besag et al., 1995) using 10,000, 15,000, 20,000, and 25,000 iterates is given in Table 1. To highlight the behavior of BESAG method, we have separated out the U.S. into its 12 regions (Note that the 95% credible intervals for α by region are all to the right of 15.3). Under the Poisson regression model we observe that the contents of the intervals do not converge to the nominal value of 95% as the number of iterates increases from 10,000 to 25,000. For some of the regions BESAG method converges to the wrong value, and for others perhaps 25,000 iterates are too small. For example, for W. N. Central-North and Mountain-North the probability contents are 93.3% and 93.4% respectively, lower than the nominal value of 95%. It appears that BESAG method is not appropriate for very high dimensions. Of course, we are operating in a parametric situation, not favorable to BESAG method. Also, convergence for extreme order statistics or quantiles is generally slow; see Serfling (1980, Ch. 2) for a discussion of the asymptotic behavior of order statistics and quantiles.

To assess BESAG method further we have performed a small simulation study. We have drawn samples from the independent conditional posterior densities

$$\lambda_i \mid d_i, \alpha, \underline{\beta} \quad \sim \quad \text{Gamma}\left(d_i + \alpha, n_i + \alpha e^{-\underline{x}_i'\underline{\beta}}\right), \quad i = 1, \dots, \ell. \tag{22}$$

We set α and $\underline{\beta}$ at their posterior means, obtained when the full Poisson regression model in Section 3 is fit to all the COPD data. We use the (n_i, d_i), $i = 1, \dots, \ell = 798$, as in the COPD data. The importance of this exercise is that we know precisely that the data come from (22), so that this simple exercise shows the defect of BESAG method. Letting $\lambda_i^{(h)}$, $h = 1, \dots, M$, $i = 1, \dots, \ell$, denote the samples obtained, we use BESAG method to compute the 95% simultaneous credible intervals for $\lambda_1, \dots, \lambda_\ell$, denoted by (a_i, b_i), $i = 1, \dots, \ell$. Then, the simultaneous probability content of these intervals is

$$\prod_{i=1}^{\ell} \left\{ G_i\left(b_i \mid d_i, \alpha, \underline{\beta}\right) - G_i\left(a_i \mid d_i, \alpha, \underline{\beta}\right) \right\},$$

where $G_i(\cdot \mid d_i, \alpha, \underline{\beta})$ are the incomplete gamma functions obtained from (22). We apply this simulation to the 798 HSAs and each region separately. For each design point we repeat the experiment 100 times, and Figure 1, in which the means and standard deviations (reasonably small) are reported, shows very interesting results, much in concordance with Table 1. We obtain the simultaneous coverage for $M = 1000, 5000, 10000, 15000, 20000$, and 25000 for all 798 HSAs and by region. Generally, BESAG method has very poor simultaneous coverage for small sample sizes (e.g., $M = 1000$). There is convergence for small regions with small values of HSAs by about $M = 10000$, but for larger regions convergence has not been realized by $M = 25000$ (e.g., regions 4, 5, 6, 8). Also, for all HSAs the combined coverage at $M = 1000$ is .217, and it increases to .941 at $M = 25000$.

Table 1. An assessment of the nonparametric method (Besag et al., 1995) using the simultaneous probability content by region based on the Poisson regression model using 10,000, 15,000, 20,000, and 25,000 iterates

Region		ℓ	BESAG10	BESAG15	BESAG20	BESAG25
All	All	798	.878	.950	.976	.987
New England	1	23	.945	.946	.947	.949
Middle Atlantic	2	49	.936	.946	.943	.942
S. Atlantic-North	3	38	.950	.944	.949	.947
S. Atlantic-South	4	88	.951	.941	.954	.956
E. S. Central	5	88	.933	.942	.941	.950
E. N. Central	6	121	.927	.941	.948	.954
W. N. Central-North	7	45	.930	.931	.934	.933
W. N. Central-South	8	105	.936	.946	.955	.952
W. S. Central	9	115	.931	.941	.951	.959
Mountain-South	10	40	.945	.948	.948	.949
Mountain-North	11	38	.933	.935	.934	.934
Pacific	12	48	.943	.946	.949	.948

NOTE: The probability contents for various methods are computed under the Poisson regression model. To assess

convergence, in a Monte Carlo integration procedure we have used sample sizes of 10,000, 15,000, 20,000 and 25,000 from the importance function.

The simultaneous probability content of five types of intervals is given in Table 2. The intervals are the individual equal-tailed credible interval (CI), individual highest posterior density interval (HPD), and Method I-HPD simultaneous interval each using 10,000 iterates, and the nonparametric interval described by Besag et al. (1995) from Appendix A using 10,000 iterates (BESAG10) and 25,000 iterates (BESAG25). Obvious details to note are the approximately zero simultaneous contents for the CI and HPD intervals for all regions, and the less than 0.15 simultaneous content for each region. Method I-HPD is the only method that consistently obtains the 95% simultaneous coverage for all cases. Our other three methods also have the nominal coverage value of 95%. BESAG10 underestimates and BESAG25 overestimates for all regions showing the importance of a large number of iterates; 10,000 was too few for the correct content and 25,000 shows the conservative nature of the method. For most individual regions, BESAG25 is comparable to Method I-HPD.

The values of γ that optimize (5) for Method I and the (γ_1, γ_2) pair that optimize (12) for Method II are given in Table 3. As expected, the values for γ and (γ_1, γ_2) are smaller when more HSAs are considered simultaneously. To compare the sensitivity of the values of γ and (γ_1, γ_2) on the condition of starting with individual HPD intervals, simultaneous intervals based on the credible intervals were obtained. There are differences between values of γ and (γ_1, γ_2) when starting with credible intervals versus using HPD intervals. The difference is not large for Method I where no ordinate optimality criterion is specified. However, there is a small difference for Method II which compensates for asymmetric densities to obtain equal ordinates when starting with the credible intervals. This small difference is attributed to the rough symmetry of our individual distributions. If the individual distributions are highly skewed, this difference will be even greater.
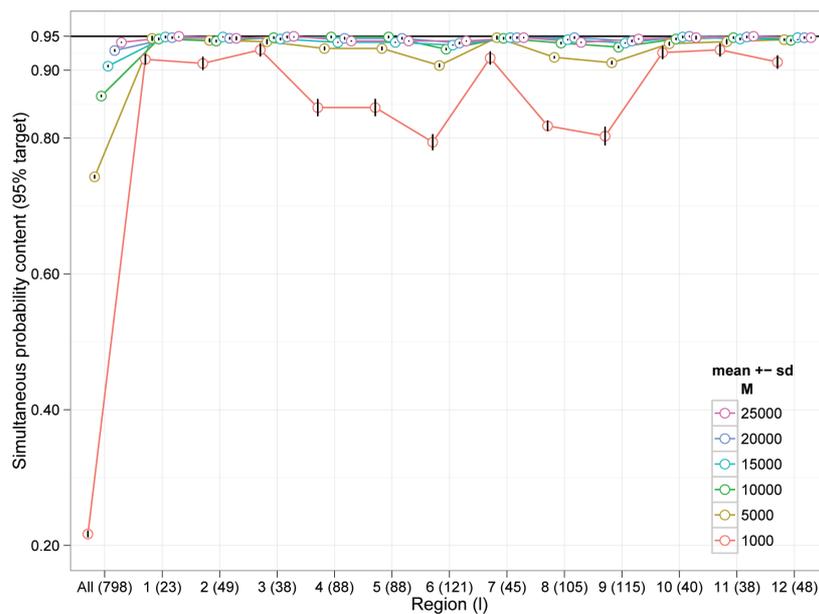
Figure 1. A simulation study to assess the simultaneous probability content of BESAG method by region. Mean and (sd) of the 100 simultaneous probability contents are reported

NOTE: The $(n_i, d_i), i = 1, \dots, \ell = 798$, with α and $\tilde{\beta}$ specified in the COPD data are used to generate 100 datasets of the 798 mortalities from their conditional posterior gamma densities. BESAG method then computes 95% simultaneous intervals for all HSAs and by regions for samples of sizes 1,000, 5,000, 10,000, 15,000, 20,000, and 25,000. Finally, the simultaneous probability content is computed using the conditional posterior gamma densities. This is done for each of the 100 datasets, and in the plot we report the mean±sd of the 100 simultaneous probability contents.

Table 2. A comparison of simultaneous probability content of different methods by region

Region	ℓ	CI	HPD	Method I-HPD	BESAG10	BESAG25
All	798	.000	.000	.950	.878	.987
1	23	.301	.307	.950	.945	.949
2	49	.081	.081	.950	.936	.942
3	38	.140	.142	.950	.950	.947
4	88	.011	.011	.950	.951	.956
5	88	.011	.011	.950	.933	.950
6	121	.002	.002	.950	.927	.954
7	45	.099	.099	.950	.930	.933
8	105	.005	.005	.950	.936	.952
9	115	.003	.003	.950	.931	.959
10	40	.127	.129	.950	.945	.949
11	38	.145	.142	.950	.933	.934
12	48	.086	.085	.950	.943	.948

NOTE: CI indicates the individual 95% credible intervals and the table shows their simultaneous content, HPD indicates the individual 95% HPD intervals and the table shows their simultaneous content, Method I-HPD simultaneous intervals contain 95% probability, BESAG10 and BESAG25 indicate the BESAG method using 10,000 and 25,000 iterates and the table shows their simultaneous content based under the Poisson regression model.

Table 3. A comparison of γ that optimizes (19) for Method I and (γ_1, γ_2) that optimize (20) for Method II starting from HPD and credible intervals

Reg	ℓ	Method I-HPD γ	Method II-HPD γ_1	γ_2	Method I-CRE γ	Method II-CRE γ_1	γ_2
All	798	.771	.771	.771	.762	.759	.775
1	23	.936	.936	.936	.937	.938	.937
2	49	.919	.919	.919	.919	.918	.921
3	38	.911	.911	.911	.910	.908	.912
4	88	.866	.866	.866	.864	.862	.867
5	88	.896	.896	.896	.896	.895	.896
6	121	.885	.885	.885	.884	.883	.886
7	45	.895	.895	.895	.900	.902	.897
8	105	.840	.839	.840	.837	.836	.840
9	115	.868	.868	.868	.867	.865	.869
10	40	.866	.866	.866	.863	.858	.868
11	38	.899	.899	.899	.902	.903	.900
12	48	.909	.909	.909	.909	.908	.910

4.2 Variation

In practice, the mean map (more appropriately posterior mean map) is presented without any measure of variability. In fact, the mean map is one of several maps that can be presented; the Markov chain Monte Carlo procedure gives $M = 1000$ maps because there are M values of the vectors $\lambda_1, \ldots, \lambda_M$. A natural summary is the posterior mean map estimator, a realization we call the mean map. How much confidence can we put on the mean map as an estimator of the true map? One way to answer this question is to use three maps; 1000 maps are too many. There are two possibilities. First, one can obtain the 95% HPD intervals for each HSA, and use the end points of these intervals to construct two maps, the lower map and an upper map. Then the degree of similarity among the lower map, mean map and upper map is a measure of the variation of the mean map estimator. This method has the problem that the overall coverage is much smaller than 95% (close to 0%). Thus, the second possibility is to use the 95% simultaneous credible intervals as a measure of the variation.

We have drawn the maps for 95% simultaneous credible intervals corresponding to the four methods: Method I-CRE, Method I-HPD, Method II-CRE, and Method II-HPD. As these maps are very similar we present the maps corresponding to Method I-HPD in Figure 3. For comparison we also study the map based on the individual HPD intervals in Figure 2.

Individual 95% HPD Interval Map

Deaths/1,000
- 4.409 - 9.205
- 4.034 - 4.409
- 3.691 - 4.034
- 3.279 - 3.691
- 1.787 - 3.279
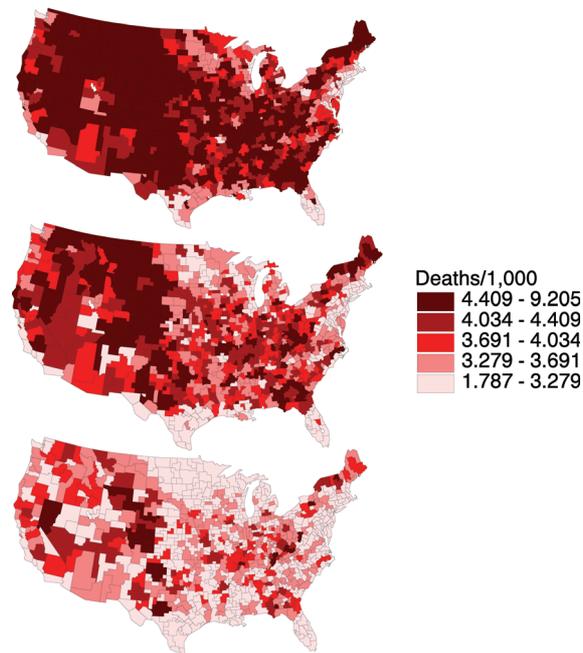
Figure 2. Choropleth maps corresponding to the 95% individual HPD intervals for COPD mortalities of white males at least 65 years old; the legend is based on the quintiles of the mean map (middle), and the lower (bottom) and upper (top) maps are the end points of the 798 credible intervals (hyper-rectangle)

95% Simultaneous Interval Map

Deaths/1,000
- 4.409 - 11.936
- 4.034 - 4.409
- 3.691 - 4.034
- 3.279 - 3.691
- 1.378 - 3.279
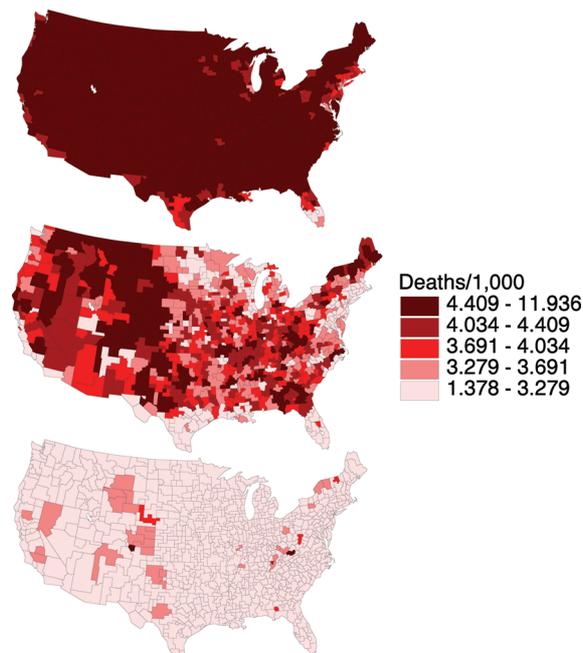
Figure 3. Choropleth maps based on Method I-HPD corresponding to the 95% simultaneous credible intervals for COPD mortalities of white males at least 65 years old; the legend is based on the quintiles of the mean map (middle), and the lower (bottom) and upper (top) maps are the end points of the 798 credible intervals (hyper-rectangle) which have combined probability content of 95%

The mean map, the map of the parameter estimates, is the middle map of the three. In practice, the mean map is

typically presented alone. The legend is based on the quintiles of the mean map. To illustrate how one might use the interval maps presented, select a HSA from the mean map. Find the corresponding HSA on both the lower and upper maps. The variation of the particular HSA selected can be interpreted from the colors in the lower to the upper map. A HSA that does not change color (or possibly changes just one color) can be said to have low variation, because inference is unchanged from map to map. A HSA that changes from light in the lower map to dark in the upper map can be said to have high variation.

First, consider the map based on the individual HPD intervals in Figure 2. Many HSAs with the least variation in terms of color difference between the lower and upper map can be easily identified by observing areas with the lightest color in the upper map or darkest color in the lower map. Some of the areas with little variation (one color difference) are HSA 608 (Douglas, MN – Stevens, MN), HSA 33 (Henrico (Richmond), VA – Prince Edward, VA), HSA 417 (Oklahoma (Oklahoma City), OK – Pottawatomie, OK), and HSA 548 (Jackson (Kansas City), MO – Clay, MO). Some of the areas with high variation (three color difference) are HSA 160 (Richland (Columbia), SC – Lexington, SC), HSA 302 (Dubuque (Dubuque), IA – Stephenson, IL), HSA 18 (Fayette (Lexington – Fayette), KY – Perry, KY), and HSA 802 (Santa Cruz (Santa Cruz), CA).

Second, consider the maps corresponding to Method I-HPD in Figure 3. Some of the areas with little variation (one color difference) are HSA 200 (Dade (Miami), FL – Broward, FL) and HSA 263 (Scott, TN). Some of the areas with moderate variation (two color difference) are HSA 538 (Webb (Laredo), TX – Jim Hogg, TX) and HSA 103 (Orleans, VT – Caledonia, VT). Some of the areas with high variation (three color difference) are HSA 516 (Val Verde, TX – Maverick, TX) and HSA 807 (Kern (Bakersfield), CA).

While the HPD map underestimates the variation as compared with the simultaneous maps, we have found some common features. Some HSAs exhibiting the least variation in both the individual HPD and Method I maps are (lightest in all maps) HSA 221 (Palm Beach (West Palm Beach), FL – St. Lucie, FL), HSA 94 (New York, NY – Kings (New York), NY), (darkest in all maps) HSA 39 (Mercer, WV – Tazewell, VA), and HSA 812 (Fremont, CO – Custer, CO). A selection of HSAs exhibiting the most variation in both maps (with increasing mean colors 2 through 5) are HSA 124 (Kent, MD), HSA 440 (Beckham, OK – Roger Mills, OK), HSA 261 (McDowell, NC), and HSA 643 (Greeley, KS – Wichita, KS).

We have also illustrated the variation in a tabular form. Table 4 shows how many HSAs are in each combination of quintiles between each pair of three maps for both individual HPD and Method I-HPD interval maps. For example, consider two cells in Table 4(a), Individual HPD: Lower to Upper. For cell $(1, 1)$, 48 HSAs are in quintile 1 on the lower map and in quintile 1 on the upper map, indicating low variation since the areas stay within their quintile. For cell $(1, 5)$, 155 HSAs are in quintile 1 on the lower map and in quintile 5 on the upper map, indicating high variation since the areas are in a low quintile in the lower map and in a high quintile in the upper map. Note that Tables 4(a) and (d) compare the intervals for individual interval maps and simultaneous maps. Because our areas are highly variable, we observe only the extremes in the Method I-HPD Tables (d–f).

Table 4. Number of HSAs in the quintiles of each pair of three maps, Lower to Upper, Mean to Upper, and Lower to Mean, for the 95% individual HPD interval (a–c) and the Method I-HPD corresponding to the 95% simultaneous credible intervals (d–f)

<table>
<tr><th colspan="6">(a) Individual HPD: Lower to Upper</th></tr>
<tr><th>Lower</th><th colspan="5">Upper Quintile</th></tr>
<tr><th>Quintile</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>48</td><td>88</td><td>74</td><td>77</td><td>155</td></tr>
<tr><td>2</td><td>0</td><td>0</td><td>10</td><td>41</td><td>148</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>0</td><td>5</td><td>82</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>45</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>25</td></tr>
</table>

<table>
<tr><th colspan="6">(d) Method I-HPD: Lower to Upper</th></tr>
<tr><th>Lower</th><th colspan="5">Upper Quintile</th></tr>
<tr><th>Quintile</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>5</td><td>6</td><td>18</td><td>49</td><td>683</td></tr>
<tr><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>30</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>4</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td></tr>
</table>

<table>
<tr><th colspan="6">(b) Individual HPD: Mean to Upper</th></tr>
<tr><th>Mean</th><th colspan="5">Upper Quintile</th></tr>
<tr><th>Quintile</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>48</td><td>75</td><td>25</td><td>11</td><td>0</td></tr>
<tr><td>2</td><td>0</td><td>13</td><td>57</td><td>53</td><td>36</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>2</td><td>57</td><td>101</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>0</td><td>2</td><td>157</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>161</td></tr>
</table>

<table>
<tr><th colspan="6">(e) Method I-HPD: Mean to Upper</th></tr>
<tr><th>Mean</th><th colspan="5">Upper Quintile</th></tr>
<tr><th>Quintile</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>5</td><td>6</td><td>18</td><td>49</td><td>81</td></tr>
<tr><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>159</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>160</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>159</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>161</td></tr>
</table>

<table>
<tr><th colspan="6">(c) Individual HPD: Lower to Mean</th></tr>
<tr><th>Lower</th><th colspan="5">Mean Quintile</th></tr>
<tr><th>Quintile</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>159</td><td>151</td><td>90</td><td>38</td><td>4</td></tr>
<tr><td>2</td><td>0</td><td>8</td><td>67</td><td>90</td><td>34</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>3</td><td>28</td><td>56</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>0</td><td>3</td><td>42</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>25</td></tr>
</table>

<table>
<tr><th colspan="6">(f) Method I-HPD: Lower to Mean</th></tr>
<tr><th>Lower</th><th colspan="5">Mean Quintile</th></tr>
<tr><th>Quintile</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th></tr>
<tr><td>1</td><td>159</td><td>159</td><td>160</td><td>159</td><td>124</td></tr>
<tr><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>30</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>4</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>0</td><td>2</td></tr>
</table>

NOTE: This is a numerical study of the variation in the estimator of the mean map.

5. Concluding Remarks

We have described four methods to construct the $100(1 − \alpha)$% simultaneous credible intervals for a relatively large number of small areas. One type of method arises from a single equation that holds the coverage probability, and the other type of method is the solution of a pair of simultaneous equations, one with the coverage probability and the other an optimality condition. Each of these types either starts with the individual $100(1 − \alpha)$% credible intervals or HPD intervals, and uses the Nelder-Mead algorithm to obtain the $100(1 − \alpha)$% simultaneous intervals. We prefer the method based on credible intervals with two equations because (a) it uses less computational effort and (b) it incorporates a sort of optimality condition. Both (a) and (b) are desirable properties for $100(1 − \alpha)$% simultaneous credible intervals.

Our method performs better than BESAG, the nonparametric method of Besag et al. (1995). We have demonstrated this fact using a small simulation study and an application on disease mapping with a standard Poisson regression model. When simultaneous intervals are required for a large number of areas, the nonparametric method needs large samples to converge. It is not within our interest to investigate the behavior of the nonparametric method for a small number of areas; but we believe that it will perform well. We must admit that the nonparametric method is simple, and in principle needs moderate amount of computation, but our method is worth the extra effort. The application is a nonstandard one because it needs 95% credible intervals for 798 parameters. In this application, our method assesses variation in the posterior mean map by constructing a lower bound map and an upper bound map so that the probability that the mean is between these two is 95%. We believe that this construction is novel.

Our work can find further application in bioinformatics in which simultaneous interval estimation will be useful. Problems in bioinformatics are typically of high dimensions, much higher dimensions than the one we have considered. Thus, standard methods like Bonferroni method are not appropriate, and further research is needed.

We mention two problems similar to the one in this paper. First, one can use spatial models to construct $100(1−\alpha)$%

simultaneous credible intervals as in our proposed method. For example, as in Besag et al. (1991), a pair-wise difference prior (PDP) can be assumed on the λ_i. The additional difficulty is that the λ_i are not independent given the other parameters and the data. Second, a problem of much interest in small area estimation is to construct the $100(1 - \alpha)\%$ credible intervals for small area proportions in a three stage hierarchical model, as in the Poisson regression model; see Nandram (1998) and Nandram and Erhardt (2005). Both of these problems are worth future investigations.

Appendix

A. Nonparametric Simultaneous Credible Intervals

For convenience, we reproduce the nonparametric method of Besag et al. (1995).

Denoting the stored sample by $\{x_i^{(t)} : i = 1, \dots, \ell; \ t = 1, \dots, M\}$, order $\{x_i^{(t)} : t = 1, \dots, M\}$ separately for each component i, to obtain order statistics $x_i^{[t]}$ and ranks $r_i^{(t)}, t = 1, \dots, M$. For fixed $k \in \{1, \dots, M\}$, let t^* be the smallest integer such that $x_i^{[M+1-t^*]} \le x_i^{(t)} \le x_i^{[t^*]}$, for all i, for at least k values of t; t^* is equal to the kth order statistic from the set $a^{(t)} = \max\left\{\max_i r_i^{(t)}, M + 1 - \min_i r_i^{(t)}\right\}, t = 1, \dots, M$, that is, $t^* = a^{[k]}$.

Then $\left\{\left[x_i^{[M+1-t^*]}, x_i^{[t^*]}\right] : i = 1, \dots, \ell\right\}$ are a set of simultaneous credible regions containing at least $100k/M\%$ of the empirical distribution.

B. Posterior Correlation Among $\lambda_1, \dots, \lambda_\ell$

We have $E(\lambda_i | \alpha, \underline{\beta}, \underline{d}) = \Lambda_i r_i + (1 - \Lambda_i)e^{x_i'\beta} = g_i, i = 1, \dots, \ell$. Also, by construction, $\text{Cov}(\lambda_i, \lambda_{i'} | \alpha, \underline{\beta}, \underline{d}) = 0, i \ne i'$, and $\text{Var}(\lambda_i | \alpha, \underline{\beta}, \underline{d}) = n_i^{-1}\Lambda_i g_i, i = 1, \dots, \ell$.

Then, using the conditional variance formula, $\text{Var}(\lambda | \underline{d}) = E''(n_i^{-1}\Lambda_i g_i) + \text{Var}''(g_i)$, and using the conditional covariance formula, $\text{Cov}(\lambda_i, \lambda_{i'} | \underline{d}) = E''(g_i g_{i'}) - E''(g_i)E''(g_{i'})$, where $E''(\cdot)$ and $\text{Var}''(\cdot)$ are expectation and variance over the posterior density of $\alpha, \underline{\beta} | \underline{d}$. Thus, letting $\rho_{ii'} = \text{Corr}(\lambda_i, \lambda_{i'} | \underline{d})$, we have

$$\rho_{ii'} = \frac{E''(g_i g_{i'}) - E''(g_i)E''(g_{i'})}{[E''(n_i^{-1}\Lambda_i g_i) + \text{Var}''(g_i)]^{\frac{1}{2}}[E''(n_{i'}^{-1}\Lambda_{i'} g_{i'}) + \text{Var}''(g_{i'})]^{\frac{1}{2}}},$$

$i \ne i'$ each $1, \dots, \ell$.

Finally, using the M iterates $\{(\alpha^{(h)}, \underline{\beta}^{(h)}), \ h = 1, \dots, M\}$ from the joint posterior density $\alpha, \underline{\beta} | \underline{d}$, we can construct a simulation consistent estimator of $\rho_{ii'}$. Now, $\bar{g}_i = M^{-1}\sum_{h=1}^M g_i^{(h)}$ and $s_i^2 = (M-1)^{-1}\sum_{h=1}^M (g_i^{(h)} - \bar{g}_i)^2$ are simulation consistent estimators of $E''(\bar{g}_i)$ and $\text{Var}''(\bar{g}_i)$, also $M^{-1}\sum_{h=1}^M g_i^{(h)} g_{i'}^{(h)}$ and $M^{-1}\sum_{h=1}^M n_i^{-1}\Lambda_i^{(h)} g_i^{(h)}$ are simulation consistent estimators of $E''(g_i g_{i'})$ and $E''(n_i^{-1}\Lambda_i g_i)$, respectively. Thus, a simulation consistent estimator of $\rho_{ii'}$ is

$$\hat{\rho}_{ii'} = \frac{M^{-1}\sum_{h=1}^M g_i^{(h)} g_{i'}^{(h)} - \bar{g}_i \bar{g}_{i'}}{[M^{-1}\sum_{h=1}^M n_i^{-1}\Lambda_i^{(h)} g_i^{(h)} + s_i^2]^{\frac{1}{2}}[M^{-1}\sum_{h=1}^M n_{i'}^{-1}\Lambda_{i'}^{(h)} g_{i'}^{(h)} + s_{i'}^2]^{\frac{1}{2}}},$$

$i \ne i'$ each $1, \dots, \ell$.

Acknowledgements

References

Albert, J. H. (1988). Bayesian estimation methods for Poisson means using hierarchical log linear model. In Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M., editors, *Bayesian Statistics, 3*. Proceedings of the Third Valencia International Meeting on Bayesian Statistics, 519-531.

Andrews, R. W., & Birdsall, W. C. (1988). Simultaneous confidence intervals: A comparison under complex sampling. In *Design and Analysis of Repeated Surveys Section*. Proceedings of the Survey Research Methods Section, American Statistical Association, 240-244. Retrieved from https://www.amstat.org/sections/srms/Proceedings/papers/1988_042.pdf

Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science, 10*, 3-66. http://dx.doi.org/10.1214/ss/1177010123

Besag, J., York, & Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*, 1-59. http://dx.doi.org/10.1007/BF00116466

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, MA: Addison-Wesley.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician, 49*(4), 327-335. http://dx.doi.org/10.1080/00031305.1995.10476177

Christiansen, C. L., & Morris, C. N. (1997). Hierarchical Poisson regression modeling. *Journal of the American Statistical Association, 92*, 618-632. http://dx.doi.org/10.1080/01621459.1997.10474013

Duncan, D. B. (1952). On the properties of the multiple comparisons test. *Virginia Journal of Science, 3*, 49-67.

Gelman, A., Roberts, G. O., & Gilks, W. R. (1996). *Bayesian Statistics, 5*, 599-607. New York: Oxford University Press.

Lui, K. J., & Cumberland, W. G. (1989). A Bayesian approach to small domain estimation. *Journal of Official Statistics, 5*(2), 143-156. Retrieved from http://www.jos.nu/Articles/abstract.asp?article=52143

Miller, R. G. (1981). *Simultaneous Statistical Inference*. New York: Springer-Verlag.

Nandram, B. (1993). Bayesian cuboid prediction intervals: An application to tensile-strength prediction. *Journal of statistical planning and inference, 44*, 167-180. http://dx.doi.org/10.1016/0378-3758(95)95021-N

Nandram, B. (1998). A Bayesian Analysis of the Three-Stage Hierarchical Multinomial Model. *Journal of Statistical Computation and Simulation, 61*, 97-126. http://dx.doi.org/10.1080/00949659808811904

Nandram, B., & Choi, J. W. (2004). Simultaneous concentration bands for continuous random samples. *Statistica Sinica, 14*, 1209-1219. Retrieved from http://www3.stat.sinica.edu.tw/statistica/J14N4/J14N413/J14N413.html

Nandram, B., & Erhardt, E. B. (2005). Fitting Bayesian Two-Stage Generalized Linear Models Using Random Samples via the SIR Algorithm. *Sankhya, 66*, 733-755. Retrieved from http://www.jstor.org/stable/25053398

Nandram, B., Liu, J., & Choi, J. W. (2005). A comparison of the posterior choropleth maps for disease mapping. *Journal of Data Science, 3*, 47-68. Retrieved from http://www.jds-online.com/file_download/64/JDS-193.pdf

Nandram, B., Sedransk, J., & Pickle, L. W. (2000). Bayesian analysis and mapping of mortality rates for chronic obstructive pulmonary disease. *Journal of the American Statistical Association, 95*, 1110-1118. http://dx.doi.org/10.1080/01621459.2000.10474307

Nandram, B., Sedransk J., & Pickle, L. W. (1999). Bayesian Analysis of Mortality Rates for U. S. Health Service Areas. *Sankhya, 61*, 145-165. Retrieved from http://www.jstor.org/stable/25053073

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*, 308-313. http://dx.doi.org/10.1093/comjnl/7.4.308

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika, 40*, 87-104. http://dx.doi.org/10.1093/biomet/40.1-2.87

Serfling, R. J. (1980). *Approximation Theorems in Mathematical Statistics*. New York: Wiley.

Tukey, J. W. (1953). The problem of multiple comparisons. *Unpublished Manuscript.*