Journal of Data Science **10**(2012), 175-193

Scheffé Style Simultaneous Credible Bands for Regression Surfaces with Application to Ache Honey Gathering

Timothy Hanson[1]* and Garnett P. McMillan[2]
[1] *University of South Carolina and* [2] *Portland VA Medical Center*

Abstract: We propose two simple, easy-to-implement methods for obtaining simultaneous credible bands in hierarchical models from standard Markov chain Monte Carlo output. The methods generalize Scheffé's (1953) approach to this problem, but in a Bayesian context. A small simulation study is followed by an application of the methods to a seasonal model for Ache honey gathering.

Key words: Generalized linear mixed model, random effects, simultaneous credible region.

1. Introduction and Motivation

Generalized linear models are a popular and widely used class of models encompassing standard normal-errors linear regression, Poisson regression, binomial regression, and many more. The addition of random effects broadens the scope and usefulness of these models allowing for dependent or grouped data. All of these models may be cast in the form of a general hierarchical model.

In longitudinal studies, it is often of interest to obtain a simultaneous credible band (SCB) for the regression surface for an average individual from the population or for a particular individual. These bands provide a region in which the entire regression surface lay with high probability and are useful as a diagnostic tool to check model adequacy, but also for making simultaneous statements about individual- or population-level mean response. To date there has been very little in the literature on computing these regions in the Bayesian or non-Bayesian setting in general hierarchical models. Recent contributions focusing on penalized splines include Crainiceanu *et al.* (2007), Brezger and Lang (2008), and Krivobokova, Kneib, and Claeskens (2010). Although cruder, the approaches we present here are more general, simpler to compute, and work well across several simulations in Section 3.

*Corresponding author.

In this paper we aim to provide some approximate methods for obtaining a SCB in a very general framework, with emphasis on introducing and empirically comparing the methods. The methods rely on the approach of Scheffé (1953) by first obtaining a crude but workable $1-\alpha$ highest posterior density region R^* for the parameters of interest, typically regression parameters and possibly a subject-specific random effect. Once the region R^* is computed, the SCB is from the minimum and maximum of a mean function over R^* for a set of covariates $\mathcal{X}$. The idea is simple, yet strangely absent from the literature so far.

In Section 2 we describe methods of obtaining regions of high posterior mass and computing the SCB, and recommend two methods for general use. Section 3 presents a modest simulation study in simple linear regression, penalized B-spline, and generalized nonlinear modeling situations. In Section 4 we provide an example of these approaches on a binomial regression model for Ache (ah-CHAY) honey gathering. Section 5 details our conclusions and future research.

2. Approximate Simultaneous Credible Bands

We propose a method based on using the functional form of the posterior density and three alternative methods for obtaining a representative subset R^* from regions of high posterior density. These subsets are used to obtain a SCB for an average or subject-specific mean function in a general hierarchical regression model.

We assume a general hierarchical model on data $y = (y_1, \cdots, y_n)'$ that includes generalized linear mixed models (GLMM):

$$\begin{array}{rcl} y_i|\beta,\gamma,\tau & \overset{ind.}{\sim} & f(y_i|\beta,\gamma), \\ \beta,\gamma|\tau & \sim & f(\beta)f(\gamma|\tau), \\ \tau & \sim & f(\tau). \end{array} \tag{1}$$

The components of γ are typically zero mean random effects that are specific to subjects or groups and we assume for each individual covariates x_i and z_i.

The posterior is proportional to the product

$$f(\beta,\gamma,\tau|y) \propto \left[\prod_{i=1}^{n} f(y_i|\beta,\gamma)\right] f(\beta)f(\gamma|\tau)f(\tau), \tag{2}$$

and inferences of interest are typically unavailable in closed-form. Inference can be obtained through any number of MCMC schemes (see, for example, Robert and Casella, 2004) resulting in a discrete approximation to the posterior distribution in the form of, say, M point masses $\{(\beta^j,\gamma^j,\tau^j)\}_{j=1}^{M}$ each with equal mass $1/M$.

For model (1), define h to be the mean function

$$E(y_i|\beta,\gamma) = h(x_i,z_i;\beta,\gamma).$$

In GLMM's we typically see additional structure of the form

$$E(y_i|\beta,\gamma) = h(x_i'\beta + z_i'\gamma),$$

where h is monotone.

For fixed z and x, we may be interested in posterior credible intervals for the mean $h(x,z;\beta,\gamma)$, easily obtained from the MCMC output. If several means are to be examined simultaneously over a set of covariate values $x \in \mathcal{X}$, then a simultaneous credible band $B(x)$ is useful. That is, we seek a relatively "small" region $B(x)$ such that $P(h(x,z;\beta,\gamma) \in B(x), \forall x \in \mathcal{X}|y) \geq 1-\alpha$. If some of the components are equal over all values in $\mathcal{X}$ and/or z contains zero elements, only a posterior subset of β,γ need be considered. Denote the parameters of interest to be θ and the remaining, "nuisance" parameters to be η. Redefine h then to be a function of the free parameters in $\mathcal{X}$ and those elements of γ corresponding to nonzero elements of z.

A standard approach to finding the band B is to first obtain a $1-\alpha$ posterior probability region R for θ that is in some sense "small". Then one obtains

$$L(x) = \inf_{\theta\in R}\{h(x;\theta)\} \quad \text{and} \quad U(x) = \sup_{\theta\in R}\{h(x;\theta)\}.$$

Define $B(x) = (L(x), U(x))$. Then $\theta \in R$ implies $h(x;\theta) \in B(x)$ for all $x \in \mathcal{X}$ and thus $P(h(x;\theta) \in B(x),\ \forall x \in \mathcal{X}|y) \geq 1-\alpha$.

Method 1

Following Wei and Tanner (1990) we define a region R to be the Highest Posterior Density (HPD) region of content $1-\alpha$ if $P(\theta \in R|y) = 1-\alpha$ and $f(\theta_1|y) \geq f(\theta_2|y)$ whenever $\theta_1 \in R$ and $\theta_2 \notin R$. We can numerically calculate $f(\theta|y)$ by noting that

$$f(\theta|y) \propto f(\theta)f(y|\theta) = f(\theta)\int f(y|\eta,\theta)f(\eta|\theta)d\eta.$$

For example, if $\theta = \beta$ and $\eta = (\gamma,\tau)$ we have

$$f(\beta|y) \propto f(\beta)\int f(y|\gamma,\beta)f(\gamma|\tau)f(\tau)d\gamma d\tau,$$

essentially (2) integrated with respect to γ and τ, which is easily approximated by either numerical or Monte Carlo integration. The latter is achieved by simulation assuming one can simulate from the priors $f(\tau)$ and $f(\gamma|\tau)$.

Another example, pertinent to computing a SCB for an individual, is when $\theta = (\beta,\gamma_k)$ and $\eta = (\gamma^{(k)},\tau)$ where $\gamma^{(k)}$ is the vector γ with the k^{th} component removed. Then we have

$$f(\beta,\gamma_k|y) \propto f(\beta)\left[\int f(\gamma_k|\tau)f(\tau)d\tau\right]\left[\int f(y|\gamma,\beta)f(\gamma^{(k)}|\tau)f(\tau)d\gamma^{(k)}d\tau\right].$$

Note that these integrals are similar to those required when computing Bayes factors, but are readily approximated.

In a similar context, Held (2004) suggests using the "Rao-Blackwellized" estimator

$$\hat{f}(\theta|y) \propto \frac{1}{M}\sum_{j=1}^{M}\frac{f(y|\theta,\eta^j)f(\theta,\eta^j)}{f(\eta^j|y)} \overset{\bullet}{\propto} \int f(\theta|\eta,y)P(d\eta|y) = f(\theta|y),$$

of Gelfand and Smith (1990). Typically, the function $f(\eta|y)$ is unknown and Held estimates $f(\eta|y)$ by $\hat{f}(\eta|y) = \sum_{j=1}^{M} f(\eta|\theta^j,y)/\mathrm{M}$, requiring on the order of M^2 operations to estimate $f(\theta|y)$ once. In two examples Held is able to find closed forms for $f(\eta|\theta,y)$, but for the general model (1) we consider this will typically not be the case.

On a side note, formula (2.5) in Held (2004) suggests the use of the median rather than the mean, but this does not provide a simulation consistent estimator of $f(\theta|y)$ in general. However, the mean (Held (2004)'s 2.4) is simulation consistent by the law of large numbers for ergodic chains and the law of total probability. If the sample size is large $\eta|y$ can be approximately normal and therefore the random variable $f(\theta|\eta,y)$ as a function of $\eta|y$ will also be approximately normal by the delta method. Then the median will approximately estimate the mean. Brezger and Lang (2008) use this approach to compute SCBs for penalized B-spline models and find that estimated contour probabilities based on the mean of the log-density "to be noticeably higher than the ones based on the median".

A finite, representative subset of R, call it R^*, is obtained by taking αM of the $\{\theta^j\}_{j=1}^{M}$ with the largest values of $f(\theta^j|y)$. Probability statements involving R are thus approximated using the set R^* and we obtain the approximate bands

$$L(x) = \min_{\theta\in R^*}\{h(x;\theta)\} \quad \text{and} \quad U(x) = \max_{\theta\in R^*}\{h(x;\theta)\}. \tag{3}$$

Alternatively, approximate regions of high posterior mass may be found using a smoothed empirical approximation to the posterior $f(\theta|y)$ based on the output $\{\theta^j\}_{j=1}^{M}$, for example using shifted histograms or frequency polygons as introduced by Scott (1985a, 1985b). Kernel-smoothed methods are impractical due to the typically high dimension of the vector β.

Method 2

Hanson and Johnson (2002) suggest simply removing $(1-\alpha)M$ of the points in $\{\theta^j\}_{j=1}^{M}$ with the largest Euclidean distance to the next nearest point. That is, for each MCMC iterate θ^j the distance $D_j = \min_{i\neq j}||\theta^j - \theta^i||$ is computed. An approximate, finite region R^* is obtained by removing from $\{\theta^j\}_{j=1}^{M}$ the $(1-\alpha)M$

MCMC realizations θ^j with the largest values of D_j. The SCB is then computed using (3). They suggest that this approach removes those discrete point masses that are literally "less dense" as measured by Euclidean distance. To compute all distances, $M(M-1)/2$ comparisons are required, e.g. half a million when $M = 1000$. A further downside of this approach is that the resulting approximate region will not be invariant to scale changes in the predictor variables; for consistency, one might consider standardizing covariates before using this method.

Method 3

An approach that is invariant to changes in the measurement scale of the predictors assumes that the posterior is approximately elliptical in shape. An approximate HPD 95% probability ellipsoid for $\theta|y$ is computed based on the Mahalanobis distance. Specifically, compute the posterior mean $\bar{\theta}$ and covariance $S_\theta = \frac{1}{M}\sum_{j=1}^{M}(\theta^j - \bar{\theta})(\theta^j - \bar{\theta})'$. Remove αM of the vectors from $\{\theta^j\}_{j=1}^{M}$ with the largest Mahalanobis distance from $\bar{\theta}$, $D_j = (\theta^j - \bar{\theta})' S_\theta^{-1} (\theta^j - \bar{\theta})'$. The resulting subset R^* has 95% of the posterior mass and is approximately HPD, assuming that $\theta|y$ is approximately multivariate normal. Again, the SCB is then computed using (3). This is analogous to the approach of Hauck (1983) for standard logistic regression.

As an alternative to the mean $\bar{\theta}$ we may also consider other measures of multivariate location that are more robust to skew such as the spatial median, the vector θ_{sm} that minimizes $\sum_{j=1}^{M} ||\theta_{sm} - \theta^j||$ or θ_{ch}, the convex hull median, the mean of data lying on the the innermost convex hull layer after successively removing data laying on outermost convex hulls. The Mahalanobis distance approach yields the smallest ellipsoid that contains $1-\alpha$ of the posterior mass centered at some measure of multivariate location; as an alternative to an ellipsoid we can consider a polytope quantile, essentially the smallest convex hull that encompasses $(1-\alpha)M$ of the $\{\theta^j\}_{j=1}^{M}$. The latter is less restricted in shape but is nonetheless convex. All of these statistics can be calculated, for example, in the "Multivariate Descriptive Statistics" package in Mathematica (Wolfram Research, Inc.).

Method 4

Methods 1, 2, and 3 work for any $\mathcal{X}$ including the whole of the parameter space. Often $\mathcal{X}$ will be some compact subset (the same or lower dimension) of the parameter space. As a final method for such sets $\mathcal{X}$, we propose for a finite uniform grid $\mathcal{X}_0 \subset \mathcal{X}$ of size $\alpha M/2$, to sequentially take every $x \in \mathcal{X}_0$, compute

$$j_1 = \arg\max_j h(x;\theta^j) \quad \text{and} \quad j_2 = \arg\min_j h(x;\theta^j),$$

and remove the corresponding θ^{j_1} and θ^{j_2} from the remaining $\{\theta^j\}$; call the resultant set R^* and proceed with (3). We are selectively removing those offending θ^j's that make the SCB "large" over a representative subset $\mathcal{X}_0$ of $\mathcal{X}$.

Note that this approach can be easily adapted for the computation of one-sided credible regions, and hence credible bounds on a regression surface. A very simple, related method due to Besag *et al.* (1995) that sequentially removes θ^js − ignoring $h(x;\theta^j)$ − produces regions R^* that are hypercubes; see also Held (2004). We found this approach to be very conservative and do not pursue it further.

3. Simulations

3.1 Simple Linear Regression

We examine how the fastest approaches, Methods 3 and 4, perform in the simplest of situations: normal-errors linear regression with one predictor. Over the domain $[0, 10]$ we consider $m_1(x) = 0.2x - 1$, see Figure 1. The predictor values are $x_i = 0.1i$ for $i = 1, \cdots, 100$ and data generated according to

$$y_i = m_1(x_i) + e_i, \quad e_i \overset{iid}{\sim} N(0, \sigma^2),$$

for $i = 1, \cdots, 100$. Two values of σ were considered: 0.1 and 0.5. For each σ, 2000 data sets of size $n = 100$ were generated. For each data set a Gibbs sampler was run for 12000 iterations taking every third iterate, yielding $M = 4000$ iterates kept. Thus we removed $\alpha M = 200$ iterates to obtain R^* for both methods. Two different models for the mean were considered: the correct linear mean, $m(x) = \beta_1 + \beta_2 x_i$, and the richer but unnecessary quadratic, $m(x) = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$. We are interested in the actual coverage rates of 95% SCBs for both linear and quadratic models as well as their volumes.

Let X be the $n \times p$ design matrix for either model. Let $\beta = (\beta_1, \beta_2)'$ for linear and $\beta = (\beta_1, \beta_2, \beta_3)'$ for quadratic. For either, we assume $\beta \sim N_p(0, 1000 I_p)$ where $p = 2, 3$ independent of $\sigma^{-2} \sim \Gamma(a, b)$ where $a = b = 0.001$. The full conditional distributions are well known:

$$\beta|\sigma \sim N_p(\sigma^{-2} V^{-1} X'y, V^{-1}), \quad \text{where} \quad V^{-1} = [0.001 I_p + \sigma^{-2} X'X]^{-1},$$

and

$$\sigma^{-2}|\beta \sim \Gamma(a + 0.5n,\ b + 0.5||y - X\beta||^2).$$

Table 1 has coverage probabilities and volumes for Methods 3 and 4. The coverage probabilities have standard deviation about 0.005, so a 95% probability interval for coverage is the point estimate ±0.01. Overall, both methods perform

about the same, reaching the nominal coverage level of 0.95 within the margin of error of the simulation study. Figure 1, for $\sigma = 0.5$, shows that Methods 3 and 4 produce similar SCBs. Fitting the quadratic versus the linear function increases the volume by 20–30% but coverage stays the same. Increasing σ from 0.1 to 0.5 increases the volume by a factor of 5, but again, nominal levels are reached.

Table 1: Coverage and volume of 95% SCB from Methods 3 & 4: linear and quadratic mean functions

σ	Mean	Method	Coverage	Volume
0.1	linear	3	0.944	0.67 (0.58, 0.78)
		4	0.939	0.68 (0.59, 0.78)
	quadratic	3	0.955	0.91 (0.78, 1.04)
		4	0.942	0.89 (0.75, 1.01)
0.5	linear	3	0.953	3.4 (2.9, 3.8)
		4	0.952	3.4 (2.9, 3.9)
	quadratic	3	0.967	4.5 (3.9, 5.2)
		4	0.951	4.4 (3.8, 5.0)

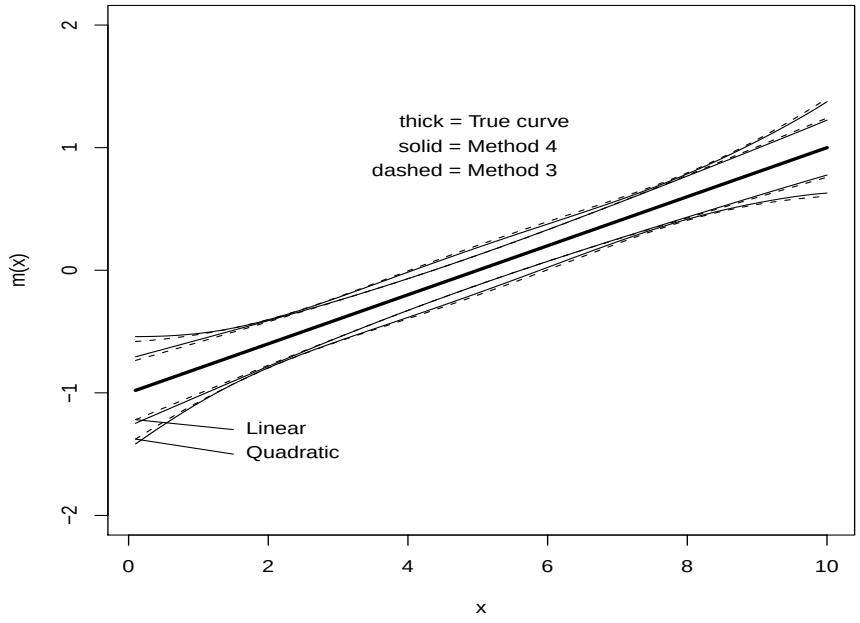


Figure 1: SCBs averaged over 2000 simulated data sets of size $n = 100$ where true mean is $m(x) = 0.2x - 1$ & $\sigma = 0.5$

3.2 Poisson Regression with Nonlinear Mean

The second simulation investigates SCBs obtained from fitting a Poisson growth model with a nonlinear mean, specifically

$$y_i \overset{ind.}{\sim} \text{Poisson}(\mu_i), \quad \text{where} \quad \mu_i = \beta_1[1 - \exp(-\beta_2 x_i)],$$

for $i = 1, \cdots, 100$. The same predictor values $x_1, \cdots, x_{100}$ are used as in Section 3.1. Here, β_1 is a horizontal asymptote and β_2 affects curvature. The true values were set at $\beta = (\beta_1, \beta_2)' = (10, 0.5)'$, see Figure 2. The improper prior $p(\beta) \propto (\beta_1\beta_2)^{-1}$ was used yielding the posterior

$$\log p(\beta|y) = C - \beta_1 \sum_{i=1}^{n}[1-\exp(-\beta_2 x_i)] + y_i \sum_{i=1}^{n} \log\{\beta_1[1-\exp(-\beta_2 x_i)]\} - \log(\beta_1\beta_2),$$

where C is a constant that depends only on $(y_1, \cdots, y_n)'$. A standard Metropolis-within-Gibbs algorithm was implemented using random-walk proposals for β_1 and β_2 tuned to achieve roughly 40% acceptance for both. In total 2000 data sets of size 100 were generated. For each data set a Gibbs sampler was run for 12000 iterations taking every third iterate, yielding $M = 4000$ iterates kept. Thus we removed $\alpha M = 200$ iterates to obtain R^* for both methods.

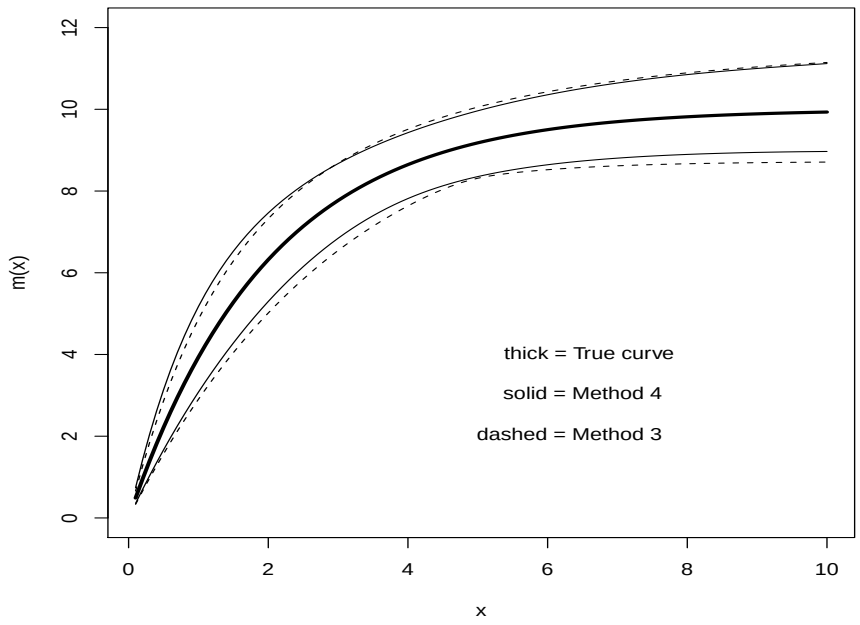


Figure 2: Poisson growth: SCBs averaged over 2000 simulated data sets of size $n = 100$

Figure 2 shows that Method 4 provides, on average, a slightly reduced volume SCB than 3. The median volume for Method 3 is 20.1 with 95% interval (18.6, 22.2); for Method 4 the median is 18.3 with 95% interval (17.3, 19.5). Coverage was 0.953 for Method 3 and 0.948 for Method 4. Both reach the nominal level but Method 4 provides a 10% smaller SCB.

3.3 Penalized B-splines

The Bayesian penalized B-spline model (Lang and Brezger, 2004) is considered in the final simulation for two regression functions. Over the domain $[0, 10]$ we consider $m_2(z) = \sin(0.2z\pi)$ and $m_1(z) = 0.2z - 1$, see Figures 3 and 4. The predictor values as before $x_i = 0.1i$ for $i = 1, \cdots, 100$ and data generated according to

$$y_i = m_j(x_i) + e_i, \quad e_i \overset{iid}{\sim} N(0, \sigma^2), \quad j = 1, 2,$$

for $i = 1, \cdots, 100$. Two values of σ were considered: 0.1 and 0.5. We are interested in the actual coverage rates of 95% SCBs for both functions $m_1(x)$ and $m_2(x)$ as well as their volumes from the B-spline approach. For each function and σ, 2000 data sets of size 100 were generated. For each data set a Gibbs sampler was run for 12000 iterations taking every third iterate, yielding $M = 4000$ iterates kept. Thus we removed $\alpha M = 200$ iterates to obtain R^* for both methods. We used 21 quadratic B-spline basis functions over 20 knots equispaced over $[0, 10]$. For completeness, the model is

$$y_i = m(x_i) + e_i = \sum_{j=1}^{21} \beta_j \phi_j(x_i) + e_i,$$

where $\phi_j(x) = \psi\{19x/10 + 3 - j\}$ and

$$2\psi(x) = x^2 I_{[0,1]}(x) + (-3 + 6x - 2x^2) I_{[1,2]}(x) + (3 - x)^2 I_{[2,3]}(x).$$

Note that constant, linear, and quadratic polynomials over $[0, 10]$ are special cases of the B-spline expansion, and so generalize these common trend functions. The design matrix X has ijth element $\phi_j(x_i)$ for $i = 1, \cdots, 100$ and $j = 1, \cdots, 21$.

Let D be a bidiagonal matrix of dimension 20×21 with non-zero elements $d_{i,i} = 1$ and $d_{i,i+1} = -1$ for $i = 1, \cdots, 20$, i.e. a first-difference matrix. The prior on $\beta = (\beta_1, \cdots, \beta_{21})'$ is given by

$$p(\beta|\lambda) \propto \lambda^{21/2} \exp(-0.5\lambda ||D\beta||^2),$$

which encourages an overall level of smoothing based on the precision λ. The model is completed by assuming $\lambda \sim \Gamma(c, d)$ independent of $\sigma^{-2} \sim \Gamma(a, b)$; we take $a = b = c = d = 0.001$. The full conditional distributions are

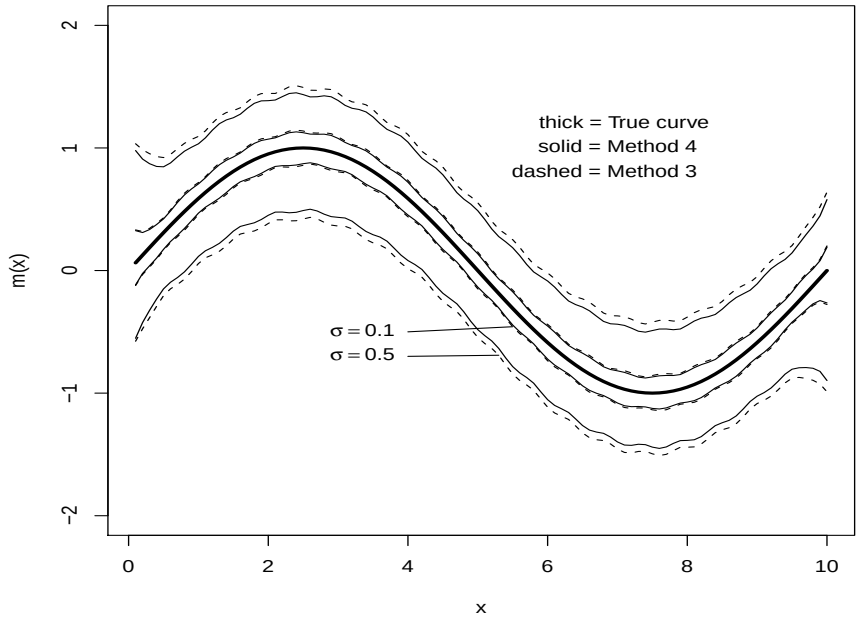


Figure 3: Credible bands averaged over 2000 simulated data sets where true mean is $\sin(0.2x\pi)$

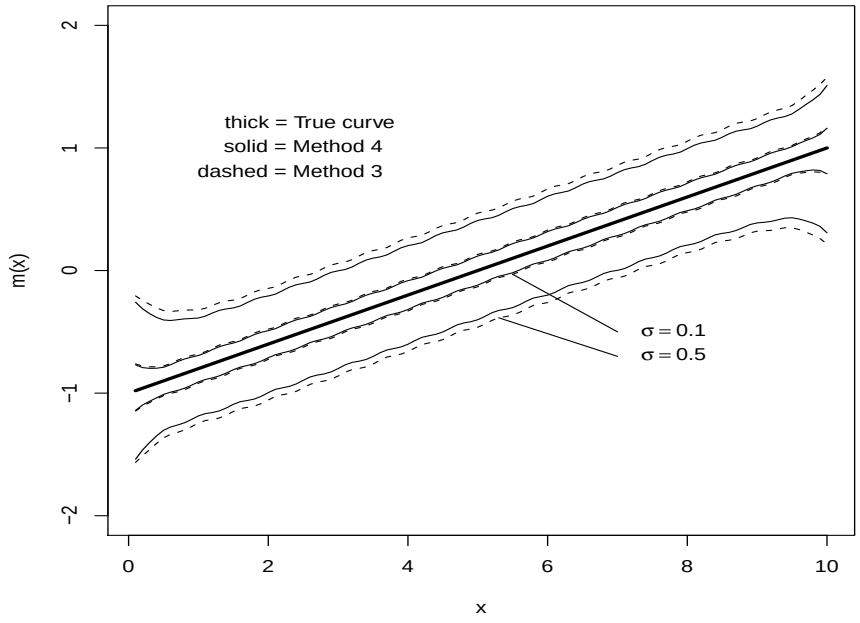


Figure 4: Credible bands averaged over 2000 simulated data sets where true mean is $0.2z - 1$

$$\beta|\sigma,\lambda \sim N_{21}(\sigma^{-2}V^{-1}X'y, V^{-1}), \text{ where } V^{-1} = [\lambda D'D + \sigma^{-2}X'X]^{-1},$$

$$\sigma^{-2}|\beta \sim \Gamma(a + 0.5n,\ b + 0.5||y - X\beta||^2),$$

and

$$\lambda|\beta \sim \Gamma(c + 0.5(21),\ d + 0.5||D\beta||^2).$$

Figures 3 and 4 show the posterior 95% SCBs averaged over the 2000 simulated data sets. Table 2 suggests Method 4 outperforms 3 in terms of volume. Both methods are conservative in coverage, but Method 4 comes closer to the nominal level of 0.95.

Table 2: Coverage and volume of 95% SCB from Methods 3 & 4

Mean function	σ	Method	Coverage	Volume
sine wave	0.1	3	0.975	2.9 (2.5, 3.2)
		4	0.955	2.6 (2.3, 3.0)
	0.5	3	0.993	11.0 (9.6, 12.3)
		4	0.975	9.7 (8.5, 10.9)
line	0.1	3	0.987	2.5 (2.2, 2.9)
		4	0.965	2.3 (2.0, 2.6)
	0.5	3	0.994	9.5 (8.3, 10.7)
		4	0.980	8.2 (7.2, 9.3)

4. Example: Ache Honey Gathering

The Ache are a group of indigenous people currently living in a small number of reservation settlements in Eastern Paraguay. Traditionally, the Ache were full time nomadic hunter-gatherers, subsisting on wild plants and animals that could be collected in the semi-tropical rain forest. The Ache had only hostile and sporadic encounters with outsiders until peaceful contact and resettlement occurred in the mid-1970s. Intensive lumber harvesting has destroyed most of the original Ache forest habitat. In 1991 the Paraguayan government, with help from the Nature Conservancy and various development agencies, created the Mbaracayu wildlife reserve as the largest tract of undisturbed forest in Eastern Paraguay. The law creating the Mbaracayu reserve as a legal entity states that the Ache have permission for unlimited use within its boundaries as long as traditional hunting and gathering practices are maintained and their impact on the forest ecology, including plant and game animal depletion, are closely monitored (Hill *et al.*, 1997; Hill and Padwe, 1998). The current analysis examines the effects of Ache honey collection on the Mbaracayu forest.

Small bands of Ache will periodically trek into the Mbaracayu reserve and stay anywhere from two to 24 days. While on trek, dependence on wild foods, including armadillo, Capuchin monkey, paca, peccary, palm tree products, fruit, and insects, is complete. The Ache also consume about 14 different varieties of wild honey. Honey collection is a labor-intensive process, requiring that one cut down the tree, typically a large, ancient hardwood, within which a bees' nest is built. While honey collection is within the scope of traditional Ache land use, the rate at which these trees are cut down poses a concern to conservation biologists.

A careful balance between wilderness conservation, particularly in the beleaguered rainforests, and the interests of indigenous people requires that policy decisions be based on carefully scrutinized evidence. Policy makers might consider curbing Ache access to the Mbaracayu reserve to limit hardwood tree felling for honey collection, but will only do if rates exceed acceptable limits and if there is a great deal of certainty about daily rates of tree felling and honey collection. To this end it is of interest to determine 1) the daily rate of honey collection for an average Ache adult male, how this rate varies over the year, and the accuracy of this model; and 2) determine the highest daily rate of honey collection over the year that we are 95% certain an average Ache man will not exceed.

The second author collected forest trek data for all residents of one Ache colony between September 12, 1997 and September 12, 1998 (McMillan, 2001). Only men, older than 17 years, who had spent more than 10 days on forest trek during the year, are included in this analysis. The final sample consists of 26 Ache men and 912 person-days on trek, with each man contributing between 10 and 91 total person-days on trek. The event that one (or more) honey trees were cut down and the date was documented for each man on each day of forest trek. The data are available from either author upon request.

We model the daily probability that a man cuts down a honey tree using logistic regression with a random effect for each man. The seasonal component is modeled as a cosine function with one 12 month period (Stolwijk, Straatman, and Zielhuis, 1999),

$$f(x) = \alpha \cos\left(\frac{2\pi x}{12} - \kappa\right),$$

where α is the amplitude, κ the horizontal shift, and x is the month of the year (1 = January, 2 = February, etc.). $f(x)$ is transformed to the linear function $f(x) = \beta_1 \sin(2\pi x/12) + \beta_2 \cos(2\pi x/12)$ and is included within the linear predictor of the logistic regression model. Details of this modeling framework are found in Stolwijk *et al.* (1999). Define y_{ij} as Bernoulli indicating the event that male i cut down one or more trees on day j while on forest trek and x_{ij} as the month

of this date. The data model is then

$$\begin{aligned} y_{ij}|p_{ij} &\overset{ind.}{\sim} Bernoulli(p_{ij}), \\ \text{logit}(p_{ij}) &= \beta_0 + \beta_1 \sin(2\pi x_{ij}/12) + \beta_2 \cos(2\pi x_{ij}/12) + \gamma_i, \\ \gamma_i|\tau &\overset{ind.}{\sim} N(0, \tau^2), \\ \tau &\sim U(0, 200). \end{aligned}$$

Informative, independent priors are available for some of the parameters of the logistic regression model. Namely, the mean function is expected to peak in late spring, during the highest flowering season, when bee activity and honey production are highest and is lowest in late fall when bee hives are largely dispersed or dormant. Priors on the regression effects are

$$\beta_0 \sim N(0, 1000), \quad \beta_1 \sim N(0, 100) \quad \text{and} \quad \beta_2 \sim N(0.28, 10).$$

A binary age indicator, with age = 50 years set as the demarcation, was originally included in the analysis but results showed a wide credible interval that included zero. The age indicator was subsequently omitted from the analysis.

We are also interested in the point during the year at which the probability of cutting down one or more honey trees is greatest, $\kappa^* = \arg\max_{x\in[0,12)} h(x;\theta)$, and the amplitude $\alpha^* = \{\max_{x\in[0,12)} h(x;\theta) - \min_{x\in[0,12)} h(x;\theta)\}/2$ of the mean function. The parameters κ^* and α^* are analogous to κ and α but on the month and probability scales respectively.

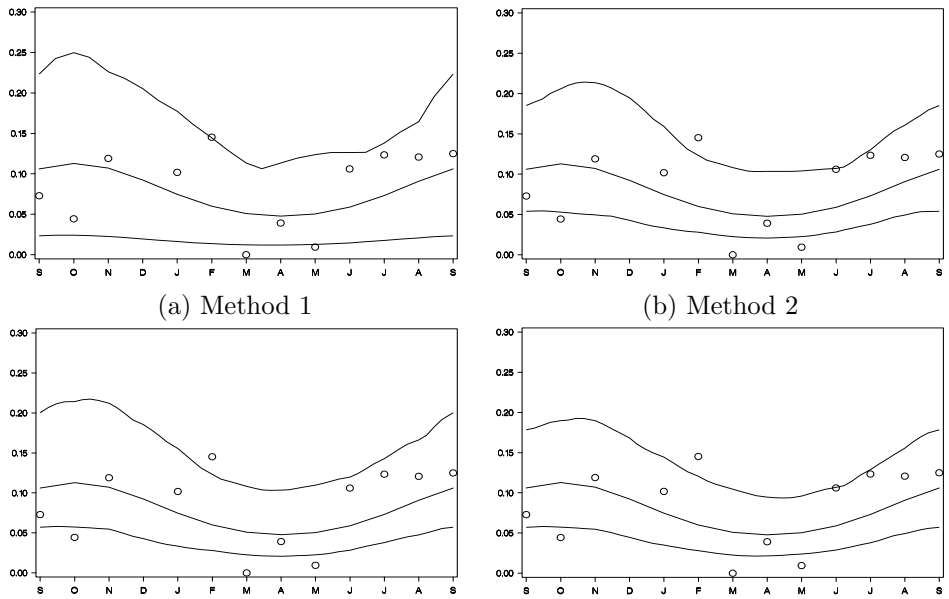
The logistic regression model was fit in WinBUGS. Two starting values were used and the model was run for 100,000 iterations. A modified Gelman-Rubin convergence diagnostic (Brooks and Gelman, 1998) was examined and showed good convergence of the individual chains. Posterior means, standard deviations, medians, 2.5%, and 97.5% quantiles from the last 50,000 iterations of the second chain are shown in Table 3. Posterior distributions for the amplitude α^* and shift κ^* were computed from the posterior distribution of β.

Table 3: Posterior results for Ache data

Par.	Mean	S.D.	2.5%	Median	97.5%
β_0	−2.53	0.26	−3.05	−2.52	−2.06
β_1	−0.40	0.18	−0.77	−0.39	−0.04
β_2	0.24	0.18	−0.08	0.23	0.59
τ	0.83	0.27	0.43	0.80	1.45
α^*	0.04	0.02	0.01	0.04	0.07
κ^*	9.8	0.95	8.0	10.0	11.0

Figure 5 show the estimated mean function for a randomly selected individual from the population along with the estimated 95% SCB using each method described; here, $\mathcal{X} = [0, 12)$. Overlayed on each plot are the average of the 26 Ache sample proportions of days in each month in which one or more trees were chopped down over the observed sample. These averages are thus unweighted with respect to the total days spent on trek per man or over all men in the sample. Figure 6 shows the same data and mean function with all SCBs together; the dashed line is the SCB of Method 4.

For these data Method 1, based on an approximation to the HPD region for $\theta|y$ as defined by Wei and Tanner (1990), provides the largest SCB. Held (2004) points out that simultaneous credible regions R calculated in the manner of Method 1 may not accurately reflect the true support of the posterior distribution and "often include areas which are not supported by the posterior at all". This may help explain why perhaps the most natural approach yields the least useful SCB of the methods presented for these data. In contrast Method 4, which depends on $\mathcal{X}$, provides the tightest SCB. We note that Method 1 takes on the order of hours to compute whereas the Methods 3 and 4 take a few seconds at most, and Method 2 a bit longer depending on M.



(a) Method 1 (b) Method 2

(c) Method 3 (d) Method 4

Figure 5: SCB for probability of cutting down one or more honey trees, Methods 1–4, September–September

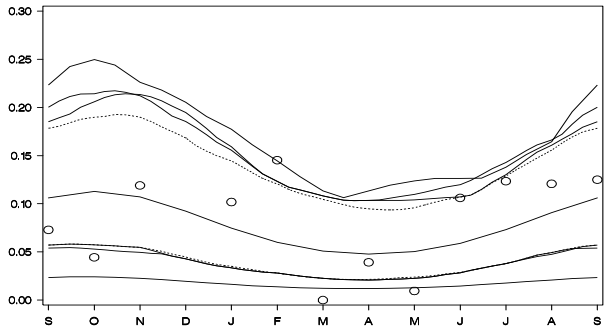


Figure 6: Per-capita daily probability of collecting honey, Methods 1–4 superimposed

Figure 5 show some degree of deficiency in the basic one-cosine seasonal model. The Method 1 SCB does not include the observed rates for March and May, and none of the other SCBs include the October and February data. The one-cosine seasonal function appears to poorly describe the observed seasonal honey collection process. Since the Ache recognize at least 14 kinds of honey, each of which is produced by a unique insect species that occupy distinct ecological and seasonal niches, a mixture of two or more cosine functions might better fit the data. A two cosine seasonal model 95% SCB from Method 4, including periods of both 6 and 12 months is shown in Figure 7. This is a decidedly better fit to the data, although the March and June data points remain outside the 95% SCB. If this model is accepted, then we are 95% certain that the per-capita daily probability of collecting honey never exceeds about 0.24 in mid summer (January-February) and about 0.21 in early spring (August-September). We are 95% certain that the per-capita probabilities never exceed about 0.06 in late fall / early winter (May-June).

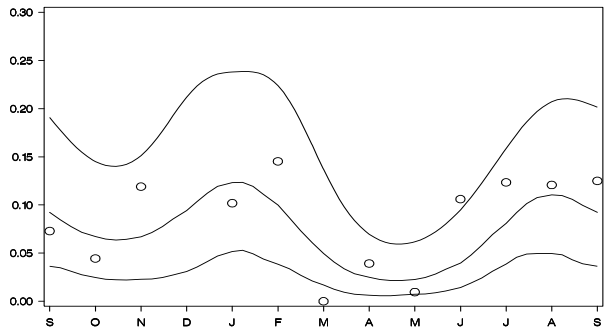


Figure 7: Ache data: two cosine model, Method 4

5. Conclusions

We have introduced several methods for obtaining an approximate SCB in hierarchical models from standard MCMC output. All four methods are easy to code and implement, although Methods 3–4 are an order of magnitude faster than Method 2, which is an order of magnitude faster than Method 1. Our simulation study only varied some of the model parameters, for example we only considered $n = 100$, but is large enough in scope to give a good feel for overall performance and validation. Overall, Method 4 provides the smallest volume while reaching nominal coverage; Methods 3 and 4 are also extremely fast to compute. Method 3 is particularly easy to implement; we have included sample SAS 9.2 code using the new PROC MCMC that implements Method 3 for a nonlinear Poisson regression example in the Appendix.

In the frequentist realm, obtaining thinner bands than through Scheffé's method was elegantly accomplished by Casella and Strawderman (1980) for the standard linear normal-errors model and for logistic regression by Piegorsch and Casella (1988); the latter method as well as the method of Hauck (1983) appeal to the asymptotic normality of maximum likelihood estimators in these models. It is possible that these approaches could be modified to deal with MCMC output as well.

Appendix: SAS Code to Implement Method 3

```
title1 'Nonlinear Poisson Regression'; title2
"Example 52.4 from PROC MCMC documentation";
   data calls;
      input weeks calls @@;
      datalines;
   1   0   1   2   2   2   2   1   3   1   3   3
   4   5   4   8   5   5   5   9   6  17   6   9
   7  24   7  16   8  23   8  27
   ;
run;

proc mcmc data=calls outpost=callout thin=10 seed=53197 ntu=1000
nmc=20000;
      parms alpha -4 beta 1 gamma 2;
      prior alpha ~ normal(-5, sd=0.25);
      prior beta  ~ normal(0.75, sd=0.5);
      prior gamma ~ gamma(3.5, scale=12);
      lambda = gamma*logistic(alpha+beta*weeks);
      model calls ~ poisson(lambda);
run;


*** Use principal components analysis to quickly compute
**** Mahalanobis distance from each theta_j to the mean vector;
proc princomp data=CallOut std out=out noprint;
     var alpha beta gamma;
run; data mahalanobis_to_mean;
     set out;
    D = sqrt(uss(of prin:));
***** Mahalanobis distance = sqrt(uncorrect sum of squares of principal components ****;
```

```
run;
************ Compute percentiles for the set of mahalanobis distances and pull out the
highest 95%;
********* Dataset RStar is composed theta_js in R*;
proc rank data = mahalanobis_to_mean
                out = RStar(where=(RankD <=94)) /*percentiles are 0 to 99. */
                groups = 100;
    var D;
    ranks RankD;
run;

***** compute mean for each theta_j in R* over observed range of weeks;
Data PlotInit1;
    set RStar;
        do weeks = 1  to 8;
            h =  gamma*logistic(alpha+beta*weeks);
            output;
        end;
run;

****** extract max and min h by week;
**** These are L(x) and U(x) from (3) in the paper;
proc sql;
    create table SBC
        as select weeks, min(h) as LWeeks, max(h) as UWeeks
            from PlotInit1
            group by weeks;
quit;

********** Combine SBC with observed data for plotting;
Data PlotData;
    set SBC calls;
run; ods listing; goptions reset = all ftext = swiss; symbol1 value
= circle i = none c = black; symbol2 value = none i = join c = red;
symbol3 value = none i = join c = red; proc gplot data = PlotData;
    plot (calls LWeeks UWeeks)*weeks / overlay legend;
run; quit;
```

Acknowledgements

Garnett McMillan's research support provided by National Science Foundation Dissertation Improvement Grant #SBR-9727656.

References

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3-66.

Brezger, A. and Lang, S. (2008). Simultaneous probability statements for Bayesian P-splines. *Statistical Modelling* **8**, 141-168.

Brooks S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434-455.

Casella, G. and Strawderman, W. E. (1980). Confidence bands for linear regression with restricted predictor variables. *Journal of the American Statistical Association* **75**, 862-868.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A. and Goodner, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265-288.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.

Hanson, T. and Johnson, W. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* **97**, 1020-1033.

Hauck, W. W. (1983). A note on confidence bands for the logistic response curve. *American Statistician* **37**, 158-160.

Held, L. (2004). Simultaneous posterior probability statements from Monte Carlo output. *Journal of Computational and Graphical Statistics* **13**, 20-35.

Hill, K. and Padwe, J. (1998). Sustainability of Ache Hunting in the Mbaracayu Reserve, Paraguay. In *Hunting for Sustainability in Tropical Forests* (Edited by J. G. Robinson and E. L. Bennett), 75-105. Columbia University Press, New York.

Hill, K., Padwe, J., Bejyvagi, C., Bepurangi, A., Jakugi, F., Tykuarangi, R. and Tykuarangi, T. (1997). Impact on hunting on large vertebrates in the Mbaracayu reserve, Paraguay. *Conservation Biology* **11**, 1339-1353.

Krivobokova, T., Kneib, T. and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association* **105**, 852-863.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**, 183-212.

McMillan, G. P. (2001). *Ache Residential Grouping and Social Foraging.* Ph.D. Dissertation, Department of Anthropology, University of New Mexico, Albuquerque, New Mexico.

Piegorsch, W. W. and Casella, G. (1988). Confidence bands for logistic regression with restricted predictor variables. *Biometrics* **44**, 739-750.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edition. Springer, New York.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40**, 87-104.

Scott, D. W. (1985a). Average shifted histograms: effective nonparametric density estimators in several dimensions. *Annals of Statistics* **13**, 1024–1040.

Scott, D. W. (1985b). Frequency polygons: theory and application. *Journal of the American Statistical Association* **80**, 348-354.

Stolwijk, A. M., Straatman, H. and Zielhuis, G. A. (1999). Studying seasonality by using sine and cosine functions in regression analysis. *Journal of Epidemiology and Community Health* **53**, 235-238.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* **22**, 1701-1728.

Wei, G. C. G. and Tanner, M. A. (1990). Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika* **77**, 649-652.

Received April 28, 2011; accepted August 25,2011.

Timothy Hanson
Department of Statistics
University of South Carolina
Columbia, SC 29208
hansont@stat.sc.edu

Garnett P. McMillan
National Center for Rehabilitative Auditory Research
Portland VA Medical Center
Portland, OR 97239
Garnett.Mcmillan@va.gov