The bootstrap and Markov chain Monte Carlo

Bradley Efron[*]
Stanford University

Abstract

This note concerns the use of parametric bootstrap sampling to carry out Bayesian inference calculations. This is only possible in a subset of those problems amenable to MCMC analysis, but when feasible the bootstrap approach offers both computational and theoretical advantages. The discussion here is in terms of a simple example, with no attempt at a general analysis.

Keywords: Bayes credible intervals, Gibbs sampling, MCMC, importance sampling

1 Introduction

Markov chain Monte Carlo (MCMC) is the great success story of modern-day Bayesian statistics. MCMC, and its sister method "Gibbs sampling," permit the numerical calculation of posterior distributions in situations far too complicated for analytic expression. Together they have produced a forceful surge of Bayesian applications in our journals, as seen for instance in Li et al. (2010), Chakraborty et al. (2010), and Ramirez-Cobo et al. (2010).

The point of this brief note is to say that in some situations the bootstrap, in particular the parametric bootstrap, offers an easier path toward the calculation of Bayes posterior distributions. An important but somewhat under-appreciated article by Newton and Raftery (1994) made the same point using nonparametric bootstrapping. By "going parametric," we can illustrate more explicitly the bootstrap/MCMC connection. The arguments here will be made mainly in terms of a simple example, with no attempt at the mathematical justifications seen in Newton and Raftery.

It is not really surprising that the bootstrap and MCMC share some overlapping territory. Both are general-purpose computer-based algorithmic methods for assessing statistical accuracy, and both enable the statistician to deal effectively with nuisance parameters, obtaining inferences for the interesting part of the problem. On the less salubrious side, both share the tendency of general-purpose algorithms toward overuse.

Of course the two methodologies operate in competing inferential realms: frequentist for the bootstrap, Bayesian for MCMC. Here, as in Newton and Raftery, we leap over that divide, bringing bootstrap calculations to bear on the Bayesian world. The working assumption is that we have a Bayesian prior in mind and the only question is how to compute its posterior distribution. Arguments about the merits of Bayesian versus frequentist analysis will not be taken up here, except for our main point that the two camps share some common ground.

Section 2 develops the connecting link between bootstrap and Bayesian calculations, mainly in terms of a simple component of variance model. The bootstrap approach to Bayesian inference for this model is carried out in Section 3, and then compared with Gibbs sampling. We end with some brief comments on the comparison.

2 Bootstrap and Bayes computations

We consider the following components of variance (or random effects) model:

$$\boldsymbol{a} \sim \mathcal{N}_n(\alpha_0 \boldsymbol{1}, \sigma_0^2 \boldsymbol{I}) \quad \text{and} \quad \boldsymbol{y}|\boldsymbol{a} \sim \mathcal{N}_n(\boldsymbol{a}, \boldsymbol{I}). \tag{2.1}$$

That is, $a_1, a_2, \ldots, a_n$ the unobserved random effects are an independent sample of size n from a normal distribution having mean α_0 and variance σ_0^2; we get to see the n independent normal variates y_i

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(a_i, 1) \qquad \text{for } i = 1, 2, \ldots, n. \tag{2.2}$$

Marginally, integrating out the unobserved random effects a_i,

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(\alpha_0, \sigma^2) \qquad i = 1, 2, \ldots, n, \tag{2.3}$$

where

$$\sigma^2 = \sigma_0^2 + 1 \tag{2.4}$$

is the *total variance*.

In this section we will concentrate on assessing the variability of the bivariate parameter

$$\beta = (\alpha_0, \sigma^2). \tag{2.5}$$

the maximum likelihood estimate (mle) of β is

$$\hat{\beta} = (\hat{\alpha}_0, \hat{\sigma}^2) = \left(\bar{y}, \sum_{i=1}^n (y_i - \bar{y})^2 \Big/ n \right), \tag{2.6}$$

$\bar{y} = \sum_1^n y_i/n$; $\hat{\beta}$ is sufficient for β, with distributions

$$\hat{\alpha}_0 \sim \mathcal{N}(\alpha_0, \sigma^2/n) \text{ independent of } \hat{\sigma}^2 \sim \sigma^2 \chi_{n-1}^2/n, \tag{2.7}$$

where χ_{n-1}^2 indicates a chi-squared variate having $n-1$ degrees of freedom.

A parametric bootstrap replication $\hat{\beta}^*$ is obtained by substituting $\hat{\beta} = (\hat{\alpha}_0, \hat{\sigma}^2)$ for the parameter value $\beta = (\alpha_0, \sigma^2)$ in (2.7),

$$\hat{\alpha}_0^* \sim \mathcal{N}\left(\hat{\alpha}_0, \hat{\sigma}^2/n\right) \text{ independent of } \hat{\sigma}^{2*} \sim \hat{\sigma}^2 \chi_{n-1}^2/n. \tag{2.8}$$

Here $\hat{\beta} = (\hat{\alpha}_0, \hat{\sigma}_0^2)$ is fixed at its observed value, while $\hat{\beta}^* = (\hat{\alpha}_0^*, \hat{\sigma}^{2*})$ varies according to (2.8). To harmonize notation with that for Bayesian inference we will denote $\hat{\beta}^*$ simply "β", that now indicates any point in the sample space $\mathcal{B}$ of possible β's, not necessarily a true value.

Suppose that $\theta = t(\beta)$ is any function of β, for example $\theta = \sigma^2/\alpha_0$, or

$$\theta = \begin{cases} 1 & \text{if } \sigma^2 \geq 2 \\ 0 & \text{if } \sigma^2 < 2. \end{cases} \tag{2.9}$$

Starting with a *prior density* $\pi(\beta)$ on β, Bayes theorem says that θ has posterior expectation, given $\hat{\beta}$,

$$E\{\theta|\hat{\beta}\} = \int_{\mathcal{B}} t(\beta)\pi(\beta)g_\beta(\hat{\beta})\,d\beta \Big/ \int_{\mathcal{B}} \pi(\beta)g_\beta(\hat{\beta})\,d\beta, \tag{2.10}$$

where $g_\beta(\hat{\beta})$ is the *likelihood function*: the density of $\hat{\beta} = (\hat{\alpha}_0, \hat{\sigma}^2)$ in (2.7) (now fixed at its observed value) as $\beta = (\alpha_0, \sigma^2)$ varies over $\mathcal{B}$. From (2.7) we calculate

$$g_\beta(\hat{\beta}) = c\exp\left\{-\frac{n}{2\sigma^2}\left[(\alpha_0 - \hat{\alpha}_0)^2 + \hat{\sigma}^2\right]\right\} \Big/ (\sigma^2)^{n/2} \tag{2.11}$$

for the components of variance problem, where c is a constant that does not depend upon β.

Define the *conversion ratio*

$$R(\beta) = g_\beta(\hat{\beta})/g_{\hat{\beta}}(\beta), \tag{2.12}$$

the denominator being the bootstrap density (2.8), with $\beta = (\alpha_0, \sigma^2)$ denoting $\hat{\beta}^* = (\hat{\alpha}_0^*, \hat{\sigma}^{2*})$. We can rewrite (2.10) as

$$E\{\theta|\hat{\beta}\} = \int_{\mathcal{B}} t(\beta)\pi(\beta)R(\beta)g_{\hat{\beta}}(\beta)\,d\beta \Big/ \int_{\mathcal{B}} \pi(\beta)R(\beta)g_{\hat{\beta}}(\beta)\,d\beta. \tag{2.13}$$

Now the integrals are over the bootstrap density $g_{\hat{\beta}}(\beta)$ rather than the likelihood function in (2.10). This allows $E\{\theta|\hat{\beta}\}$ to be directly estimated via bootstrap sampling.

Suppose we draw a parametric bootstrap sample by independently sampling B times according to (2.8),

$$g_{\hat{\beta}} \longrightarrow \beta_1, \beta_2, \ldots, \beta_B.$$

Let $R_i = R(\beta_i)$, $\pi_i = \pi(\beta_i)$, and $t_i = t(\beta_i)$. Both the numerator and denominator of (2.13) are expectations with respect to the density $g_{\hat{\beta}}(\beta)$. These can be estimated by bootstrap sample averages, leading to the estimate of $E\{\theta|\hat{\beta}\}$,

$$\hat{E}\{\theta|\hat{\beta}\} = \sum_{i=1}^{B} t_i \pi_i R_i \Big/ \sum_{i=1}^{B} \pi_i R_i. \tag{2.14}$$

In this way, any Bayesian posterior expectation can be evaluated from parametric bootstrap replications.

None of this is very novel, except for the focus on the parametric bootstrap: (2.14) is a standard importance sampling procedure, as described in Chapter 23 of Lange (2010). A connection between the nonparametric bootstrap and Bayesian inference was suggested under the name "Bayesian bootstrap" in Rubin (1981), and also in Section 10.6 of Efron (1982). Newton and Raftery (1994) make the connection more tangible, applying (2.14) with nonparametric bootstrap samples.

Parametric bootstrapping makes the connection more tangible still, since in favorable situations we can write down and examine explicit representations for the conversion factor $R(\beta)$ (2.12). For our components of variance problem, (2.11) and the related expression for $g_{\hat{\beta}}(\beta)$ yield

$$R(\beta) = \left(\frac{\sigma^2}{\hat{\sigma}^2}\right)^{3/2} \exp -\frac{n}{2}\left\{(\alpha_0 - \hat{\alpha}_0)^2\left(\frac{1}{\sigma^2} - \frac{1}{\hat{\sigma}^2}\right) + \left(\frac{\hat{\sigma}^2}{\sigma^2} - \frac{\sigma^2}{\hat{\sigma}^2}\right) + 2\log\frac{\sigma^2}{\hat{\sigma}^2}\right\}. \tag{2.15}$$

The factor R is interesting in its own right. It relates "sample space integration to parameter space integration," to quote Fraser et al. (2010), investigating a related problem.

A considerable advantage of bootstrap estimate (2.14), compared to its Gibbs sampling or MCMC analogues, is that an accuracy estimate for $\hat{E}\{\theta|\hat{\beta}\}$ is available from the same data used in its calculation. In (2.14) define

$$r_i = \pi_i R_i \quad \text{and} \quad s_i = t_i \pi_i R_i, \tag{2.16}$$

so

$$\hat{E}\{\theta|\hat{\beta}\} = \bar{s}/\bar{r}. \tag{2.17}$$

Let $\widehat{\text{Cov}}$ be the usual 2-by-2 sample covariance matrix calculated from the B vectors (s_i, r_i), say

$$\widehat{\text{Cov}} = \begin{pmatrix} \hat{c}_{ss} & \hat{c}_{sr} \\ \hat{c}_{rs} & \hat{c}_{rr} \end{pmatrix}. \tag{2.18}$$

(For example, $\hat{c}_{sr} = \sum(s_i - \bar{s})(r_i - \bar{r})/B$.) Then standard delta-method calculations yield

$$\text{var}\left\{\hat{E}\{\theta|\beta\}\right\} \doteq \frac{\hat{E}\{\theta|\beta\}^2}{B}\left(\frac{\hat{c}_{ss}}{\bar{s}^2} - 2\frac{\hat{c}_{sr}}{\bar{s}\bar{r}} + \frac{\hat{c}_{rr}}{\bar{r}^2}\right). \tag{2.19}$$

Formula (2.19) is possible because the bootstrap vectors (s_i, r_i) are drawn independently of each other, which is not the case in MCMC sampling.

3   An example

As an example of the bootstrap/Bayes computations of Section 2, a simulated data vector $\boldsymbol{y}$ was generated according to (2.1), with

$$n = 100, \quad \alpha_0 = 1, \quad \text{and} \quad \sigma_0^2 = 0.25. \tag{3.1}$$

Marginally, the independent obervations y_i had distribution

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(\alpha_0, \sigma^2) \qquad [\sigma^2 = 1.25] \tag{3.2}$$

as in (2.3). The maximum likelihood estimates (2.6) were

$$\hat{\beta} = \left(\hat{\alpha}_0, \hat{\sigma}^2\right) = (1.005, 1.295). \tag{3.3}$$

Here we will only concern ourselves with calculating the posterior distribution of σ^2.

$B = 25,000$ bootstrap replications (many times more than necessary, as we shall see), $\beta_1, \beta_2, \ldots, \beta_B$, were generated from the bootstrap distribution $g_{\hat{\beta}}(\beta)$ (2.8), and the conversion factors $R_i = R(\beta_i)$ (2.12) calculated. The prior density $\pi(\beta)$ was taken to be

$$\pi(\beta) = 1/\sigma^2. \tag{3.4}$$

(Or equivalently, $\pi(\beta) = c/\sigma^2$ for any positive constant c, since c factors out of Bayesian calculations, as seen in (2.10).) Formula (3.4) is Jeffreys' "uninformative" prior for inferences concerning the total variance σ^2. Finally, the adjusted factors $r_i = \pi_i R_i = \pi(\beta_i)R(\beta_i)$ were calculated. We now have $B = 25,000$ points β_i, each with associated weight r_i, and these can be used to estimate any posterior expectation $E\{\theta|\hat{\beta}\}$, as in (2.16)–(2.17).

Table 1 describes the posterior distribution of the total variance σ^2. The interval [0.75,2.15], which contains almost all of the bootstrap replications σ_i^2, has been divided into bins of width

4

	.8	.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1
Boot	107	497	1707	3744	5191	5268	4090	2484	1188	488	160	57	14	5
Bayes	14	171	1010	2969	4762	5275	4443	3029	1714	882	390	200	82	61

Table 1: *Top row*: bin counts for $B = 25,000$ parametric bootstrap replications of σ^2, as in (2.8); for example, 497 values σ_i^2 fell into second bin, $.85 \leq \sigma_i^2 < .95$. *Bottom row*: weighted bin counts (3.5), for posterior density of σ^2 given $\boldsymbol{y}$; using Jeffreys invariant prior density $1/\sigma^2$.

0.1 each. The top row of the table shows the bootstrap histogram counts, for example, 1707 of the σ_i^2 values fell into the third bin, $0.95 \leq \sigma_i^2 < 1.05$. The bottom row shows the Bayesian bin counts, weighted according to $r_i = \pi_i R_i$,

$$\text{Bayes}_k = B \cdot \sum_{k\text{th bin}} r_i \Big/ \sum_1^B r_i, \tag{3.5}$$

scaled to sum to B. The weights r_i were smaller than 1 in the third bin, resulting in fewer Bayes counts, 1010 compared to 1707. The opposite happens at the other end of the scale.
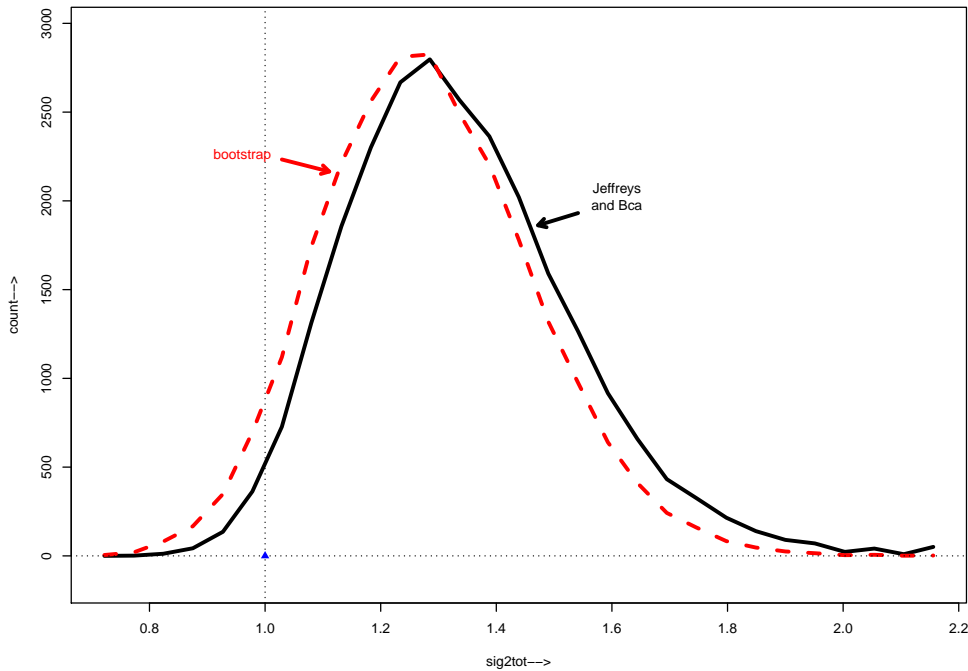

Figure 1: Distribution of total variance σ^2 given $\hat{\beta}$ (3.3); *dashed curve* counts from 25,000 parametric bootstrap replications (2.8), "Boot" row of Table 1; *solid curve* weighted counts (3.5), weights $r_i = \pi(\beta_i) R(\beta_i)$, with $\pi(\beta)$ the Jeffreys prior $1/\sigma^2$ and $R(\beta)$ the conversion factor (2.15). Corrected bootstrap distribution "Bca" agrees with Jeffreys.

Figure 1 graphs the Boot and Bayes counts of Table 1. We see that the Jeffreys Bayes distribution is shifted to the right of the Boot distribution. It can be shown that the Jeffreys prior $\pi(\beta) = 1/\sigma^2$ is almost exactly correct from a frequentist point of view: for example, the upper 90% point for the Jeffreys posterior distribution is almost exactly the upper endpoint of the usual 90% one-sided confidence interval for σ^2. (The Welch–Peers theory of 1963 shows this

5

kind of Bayes-frequentist agreement holding asymptotically for all priors of the form $\pi(\beta) = h(\alpha_0)/\sigma^2$, with $h(\cdot)$ any smooth positive function. The components of variance problem is unusual in allowing a simple expression for the Welch–Peers prior; see Section 6 of Efron (1993).)

α	se	cv	$\sigma^2[\alpha]$	se	cv
.025	.0007	.028	**1.01**	.0016	.0015
.050	.0011	.022	**1.05**	.0014	.0013
.100	.0016	.016	**1.10**	.0013	.0012
.160	.0021	.013	**1.15**	.0013	.0012
.500	.0034	.007	**1.32**	.0016	.0013
.840	.0033	.021	**1.52**	.0031	.0020
.900	.0031	.031	**1.59**	.0043	.0027
.950	.0027	.055	**1.69**	.0072	.0043
.975	.0024	.098	**1.78**	.0126	.0071

Table 2: Estimated posterior quantiles $\sigma^2[\alpha]$ for total variance σ^2, given data (3.3) and Jeffreys prior $\pi(\beta) = 1/\sigma^2$. Also shown are standard errors (se) and coefficients of variation (cv), obtained using (2.19), as explained in text.

Table 2 shows the estimated posterior quantiles $\sigma^2[\alpha]$ for σ^2, given data (3.3) and Jeffreys prior (3.4); for example, $\sigma^2[0.95] = 1.69$ is the empirical 95th percentile of the 25,000 bootstrap replications weighted according to $r_i = \pi_i R_i$ as in (3.5).

Table 2 also shows standard errors derived from (2.19), i.e., the variability due to working with $B = 25,000$ bootstrap replications (rather than $B = \infty$). As an example, let $\theta = t(\beta)$ in (2.9) equal 0 or 1 as σ_i^2 is greater or less than 1.69, the value of $\sigma^2[0.95]$. Following through (2.16)–(2.17) results in 0.0027 as the standard error of $\hat{E}\{\theta|\hat{\beta}\} = 0.95$, and this is the se value shown in the second column's next-to-bottom row, along with cv $= 0.055$ ($= 0.0027/(1-0.950)$). The standard error for $\hat{\sigma}^2[\alpha] = 1.69$ is then obtained from the usual delta-method approximation $0.0027/g[0.95] = 0.0072$, where $g[0.95]$ is the height of the Jeffreys prior density curve at 1.69.

All of the standard errors are small. If we had stopped at $B = 5000$ bootstrap replications, they would be multiplied by $\sqrt{5} = 2.2$, and would still indicate good accuracy. At $B = 1000$ the multiplying factor increases to 5, and we might begin to worry about simulation errors.

All of these calculations depend on the adjusted conversion factor $r_i = \pi_i R_i$, with $\pi_i = 1/\sigma_i^2$ and $R_i = R(\beta_i)$ as in (2.15). Figure 2 plots r_i versus σ_i^2 for the $B = 25,000$ bootstrap replications from (2.8). These are less or greater than 1 as σ_i^2 is less or greater than $\hat{\sigma}^2 = 1.30$, accounting for the rightward shift of the Bayes posterior density in Figure 1.

Much of the simulation standard error seen in Table 2 comes from the large values of r_i on the right side of Figure 2. Strategems for decreasing simulation error exist. In this case we might modify the bootstrap sampling recipe (2.8), say to

$$\hat{\sigma}^{2*} = 2\hat{\sigma}^2 \chi_{n-1}^2/n. \tag{3.6}$$

This would extend the dashed curve rightwards in Figure 1, reducing the large values of R_i and r_i. We will not worry further about estimation efficency here.

The Bca system of confidence intervals ("bias-corrected and adjusted," Efron, 1987) adjust the raw bootstrap distribution — represented by the dashed curve in Figure 1 — to achieve
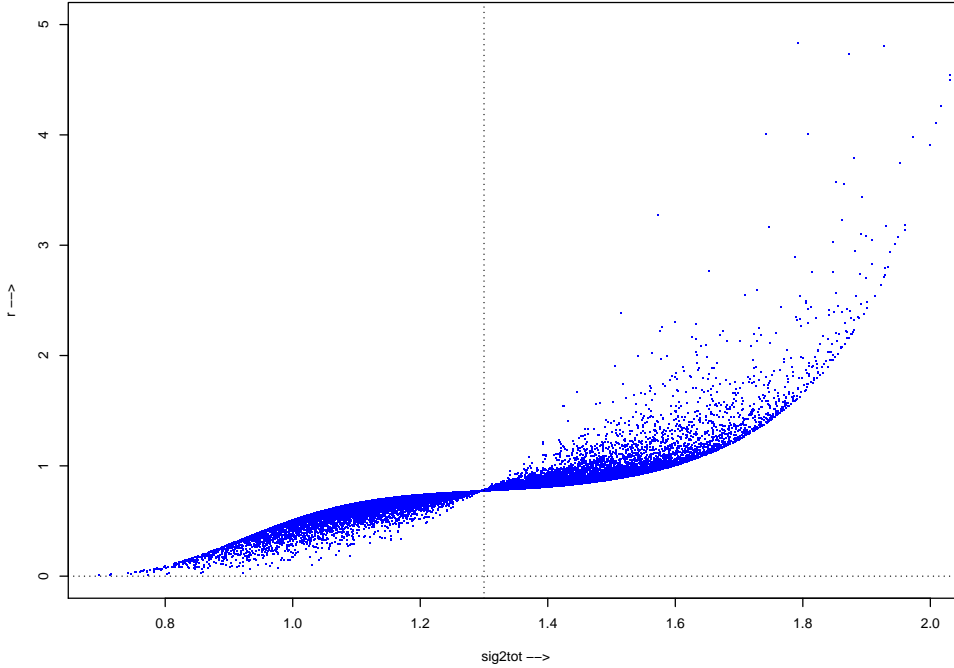
Figure 2: Adjusted conversion factor $r_i = \pi_i R_i$ for the $B = 25{,}000$ bootstrap replications, plotted versus the bootstrap replications σ_i^2; the factor is less or greater than 1 as σ_i^2 is less or greater than $\hat\sigma^2 = 1.30$, moving the Jeffreys posterior density rightward in Figure 1.

second-order accurate frequentist coverage: a level-α Bca interval has true average probability $\alpha + O(1/n)$, n the sample size, rather than $\alpha + O(1/\sqrt{n})$ for the usual first-order approximate intervals. The corresponding "confidence distribution" (e.g., assigning probability 0.01 to the region between the 0.90 and 0.91 Bca interval endpoints) requires a different conversion factor $R_{\mathrm{Bca}}(\beta)$, that does not involve any Bayesian input. In the example of Figure 1, the Bca confidence density almost perfectly matches the Jeffreys posterior density.

All of the calculations so far have concerned the total variance $\sigma^2 = \sigma_0^2 + 1$. We might be more interested in σ_0^2 itself, the variance of the random effects a_i in model (2.11). Simply shifting the posterior densities of Figure 1 a unit to the left has an obvious defect: some of the probability (0.051 for the raw bootstrap, 0.021 for Jeffreys/Bca) winds up below zero, defying the obvious constraint $\sigma_0^2 \geq 0$. We need a prior density $\pi(\beta)$ that has only positive support for σ_0^2.

The Jeffreys analogue $\pi(\beta) = 1/\sigma_0^2$ looks promising, but fails because of its infinite spike at zero. (Unlike σ^2, $\sigma_0^2 = \sigma^2 - 1$ has substantial likelihood near 0 given the observed data (3.3).) The toy example in Spiegelhalter et al. (1996, Sect. 2.3) suggests using prior density

$$\pi(\beta) : \alpha_0 \sim \mathcal{N}(0, V_0) \text{ independent of } \sigma_0^2 \sim \nu_0/G_{\nu_0}, \tag{3.7}$$

$V_0 = 10{,}000$ and $\nu_0 = 0.01$, in a situation similar to (2.11), with G_{ν_0} indicating a gamma distribution having shape parameter ν_0. Prior or hyperprior assumptions like (3.7) are familiar in the MCMC literature.

We can reweight our $B = 25{,}000$ bootstrap replications β_i to compute the posterior distribution of σ_0^2 for prior (3.7), as in (3.5); R_i is still provided by (2.15) (with $\sigma^2 = \sigma_0^2 + 1$); the Bayes prior is

$$\pi_i = c \exp\{-\nu_0/\sigma_{0i}^2\}/\sigma_{0i}^{2(\nu_0+1)}, \tag{3.8}$$

7

from the inverse gamma distribution in (3.7), the $\alpha_0 \sim \mathcal{N}(0, V_0)$ portion of the prior having almost no effect.
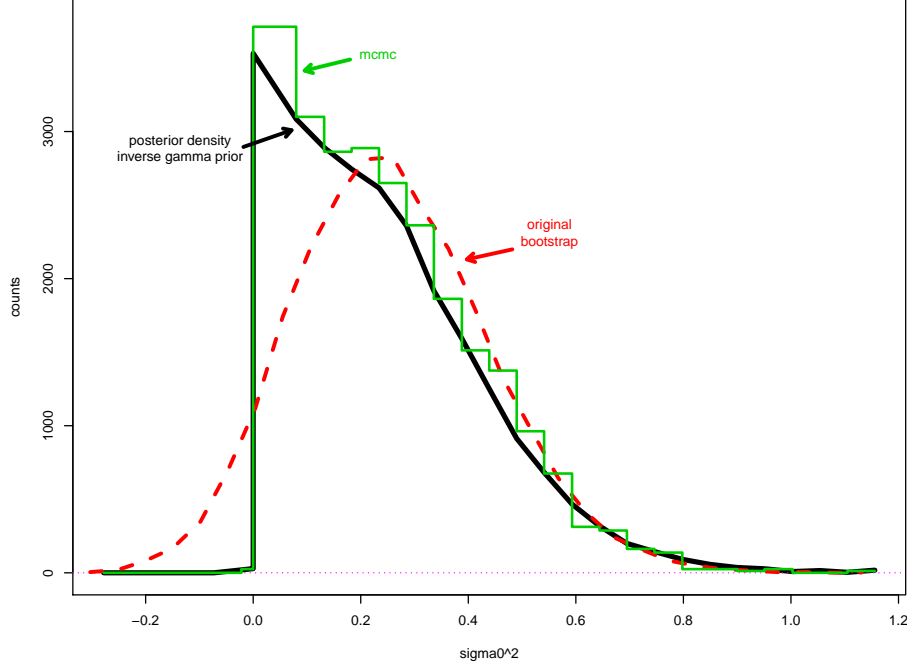

Figure 3: *Heavy pointed curve* indicates posterior density of σ_0^2 in model (2.1), data (3.3); inverse gamma prior (3.7) as calculated from weighted bootstrap replications. *Stepped curve* is same posterior density calculation by Gibbs sampling algorithm (3.9). *Dashed curve* shows unweighted bootstrap counts of Figure 2 shifted one unit left.

The heavy pointed curve in Figure 3 shows the posterior distribution of σ_0^2 under model (3.7), as calculated from the weighted bootstrap replications. It looks completely different than a simple leftward translation of the curves in Figure 1. Now the posterior density has its high point at the minimum allowable value $\sigma_0^2 = 0$, and decreases monotonically to the right.

A Gibbs sampling analysis was performed, as a check on the weighted bootstrap results. Given the data $\boldsymbol{y}$, or equivalently given the sufficient statistics $(\hat{\alpha}_0, \hat{\sigma}^2)$ (3.3), a Gibbs analysis cycles through the three unobserved quantities in (2.11), namely $\boldsymbol{a}$, α_0, and σ_0^2. The "complete conditional distribution" of each one given the other two is calculated from the specifications (2.1) and (3.7):

$$
\begin{aligned}
\boldsymbol{a}|\alpha_0, \sigma_0^2 &\sim \mathcal{N}_\nu\left(\left(\frac{\alpha_0}{\sigma_0^2}\mathbf{1} + \boldsymbol{y}\right)\Big/ h_1, \boldsymbol{I}/h_1\right) && \left[h_1 = \frac{1}{\sigma_0^2} + 1\right] \\
\alpha_0|\boldsymbol{a}, \sigma_0^2 &\sim \mathcal{N}\left(\frac{n\bar{a}}{\sigma_0^2}\Big/ h_2, 1/h_2\right) && \left[h_2 = \frac{1}{V_0} + \frac{n}{\sigma_0^2}\right] \\
\sigma_0^2|\alpha_0, \boldsymbol{a} &\sim \mathcal{N}(\nu_0 + S/2)/G_{\nu_0 + n/2} && \left[S = \sum_1^n (a_i - \bar{a})^2\right].
\end{aligned}
\tag{3.9}
$$

Here $n = 100$ is the sample size, $\bar{a} = \sum a_i/n$, $\mathbf{1}$ is a vector of n 1's, $\boldsymbol{I}$ is the $n \times n$ identity matrix, and G_ν indicates a gamma random variable with shape parameter ν. Starting from $\alpha_0 = 0$ and $\sigma_0^2 = 1$ as initial values, 25,000 cycles through (3.9) produced 25,000 values σ_{0i}^2.

8

Their step histogram in Figure 3 agrees reasonably well with the bootstrap results. (Formula (2.19) suggests that the bootstrap estimates are nearly exact.)

A comparison of Figure 3 with Figure 1 raises some important points:

- The inverse gamma prior distribution (3.7)–(3.8) is far from uninformative. The posterior distribution of $\sigma_0^2 = \sigma^2 - 1$ is strongly pulled toward zero, putting nearly 50% of its posterior probability on $\sigma_0^2 \leq 0.2$. This compares with 25% for $\sigma^2 \leq 1.2$ in the Jeffreys/Bca analysis of Figure 1.

- Changing $\nu_0 = 0.01$ to 0.001 in (3.7) substantially increases the posterior concentration of σ_0^2 near 0. Inverse gamma priors are especially convenient for Gibbs sampling implementation, as in (3.9), but should not be assigned casually.

- The mle $\hat{\sigma}^2 = \sum(y_i - \bar{y})^2/n$ in (2.6) ignores the constraint $\sigma^2 \geq 1$, the actual mle being $\max(1, \hat{\sigma}^2)$. This suggests that for the dashed curve in Figure 3 we could erase the portion below zero and replace it with a probability atom of size 0.051 at 0.

- More accurately, we could shift the Jeffreys/Bca density of Figure 1 one unit left, replacing the part below zero with an atom of size 0.021 at 0.

- In a genuine random effects application, we might well allocate some prior probability to $\sigma_0^2 = 0$ (as is *not* done in (3.7)). This makes for more difficult Bayesian analysis. The Bca approach of the previous remark is an attractive option here.

- Priors that are convenient for MCMC or Gibbs sampling applications play no special role in the bootstrap/Bayes approach of this paper. Convenience priors can be risky, in my opinion, and should be avoided in favor of frequentist methods in the absence of genuine Bayesian prior information.

References

Chakraborty, A., Gelfand, A. E., Wilson, A. M., Latimer, A. M. and John A. Silander, J. (2010). Modeling large scale species abundance with latent spatial processes. *Ann. Appl. Statist.* 4: 1403–1429, doi:10.1214/10-AOAS335.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics *38*. Philadelphia, Pa.: Society for Industrial and Applied Mathematics (SIAM).

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* 82: 171–200, with comments and a rejoinder by the author.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* 80: 3–26.

Fraser, D. A. S., Reid, N., Marras, E. and Yi, G. Y. (2010). Default priors for bayesian and frequentist inference. *J. Roy. Statist. Soc. Ser. B* 72: 631–654, doi:10.1111/j.1467-9868.2010.00750.x.

Lange, K. (2010). *Numerical Analysis for Statisticians*. Statistics and Computing. New York: Springer, 2nd ed.

Li, B., Nychka, D. W. and Ammann, C. M. (2010). The value of multiproxy reconstruction of past climate. *Journal of the American Statistical Association* 105: 883–895, doi:10.1198/jasa. 2010.ap09379.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B* 56: 3–48, with discussion and a reply by the authors.

Ramirez-Cobo, P., Lillo, R. E., Wilson, S. and Wiper, M. P. (2010). Bayesian inference for double Pareto lognormal queues. *Ann. Appl. Statist.* 4: 1533–1557, doi:10.1214/10-AOAS336.

Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* 9: 130–134.

Spiegelhalter, D. J., Best, N. G., Gilks, W. R. and Inskip, H. (1996). Hepatitis B: A case study im MCMC methods. In Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds), *Markov chain Monte Carlo in practice*, Interdisciplinary Statistics. London: Chapman & Hall, 21–43.

Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* 25: 318–329.