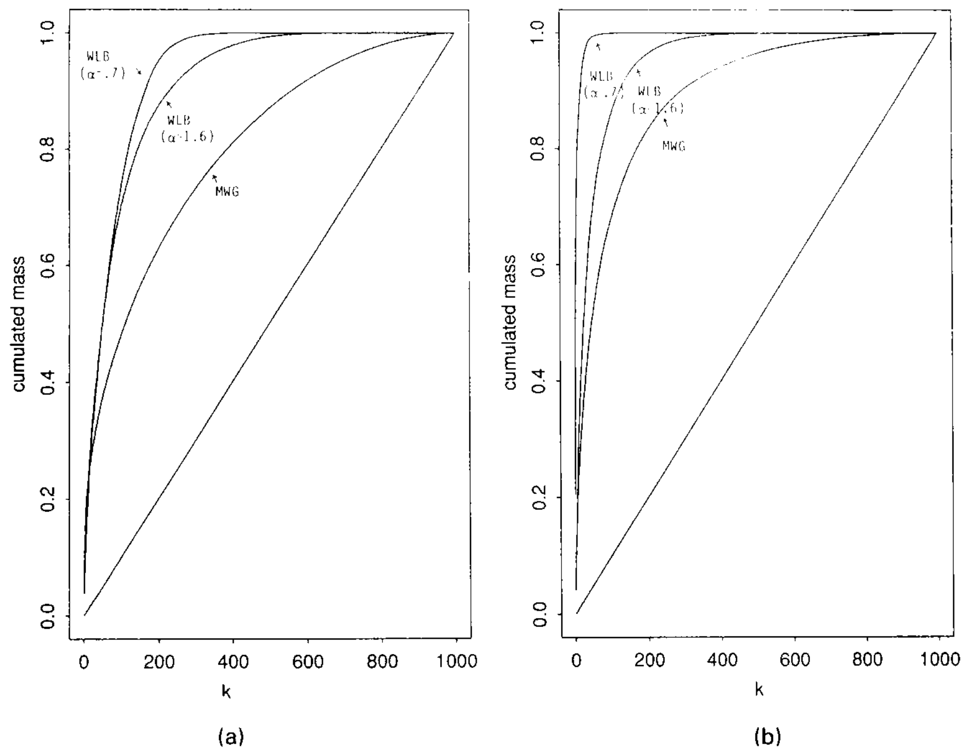

Fig. 15. Total mass assigned to the k largest weights *versus* k for (a) a flat prior for β and (b) Jeffreys's prior for β

evaluation of which requires $4n$ IB evaluations; maximization can be done via grid search or possibly a derivative-based method since $\partial L/\partial \beta_i$, $\partial L/\partial w$ and $\partial L/\partial \nu$ can be directly computed. The Bayesian model is completed by taking independent uniform priors for w and ν and, given w, a flat or Jeffreys prior for β.

We applied this model to a data set described in Aitkin *et al.* (1989) ($n = 33$). Allowing for interaction, β becomes four dimensional. Aitkin *et al.* (1989) fit, among other models, an exponential generalized linear model which corresponds to our base-line model. We fit the six-dimensional Bayesian model in two ways. We ran a weighted likelihood bootstrap (WLB) using 1000 weights drawn from Dir($\alpha\mathbf{1}$) with $\alpha = 0.7$ and $\alpha = 1.6$ thus obtaining 1000 samples approximately from the posterior under a flat prior. We ran an adaptive Metropolis-within-Gibbs (MWG) algorithm for both a flat and a Jeffreys prior following Müller's (1991) suggestions, stopping each of five strings after 500 iterations using the last 200 iterations from each again to obtain 1000 samples approximately from the posterior. Run times for the WLB on an IBM 3094 computer were 220 min for each α: for the MWG 40 min. Using either the WLB or the MWG samples we obtained the importance weights u_j as defined by the authors.

In Fig. 15 we plot the total mass assigned to the k largest weights against k for the two WLBs and the MWG. The MWG weights are dramatically better than the WLB weights even under a flat prior. The latter has much more mass attached to far fewer points.

Andrew Gelman (University of California, Berkeley): As the authors point out, the weighted likelihood bootstrap (WLB) can be used in place of the Gibbs sampler or Metropolis algorithm in a wide class of problems for which it is a fairly close fit to the target posterior distribution (so that, after importance resampling, the distribution will be almost exactly correct). One tricky point in application seems to be determining whether the importance ratios are sufficiently close for the method to be accurate. As with any approximate method that is based on overdispersion, the WLB has the potential for an even wider range of practical applicability: for problems in which the WLB simulation distribution is not a close fit, it may still be useful to use the sampling–importance resampling (SIR) samples as starting points for a Markov chain simulation. For example, Section 4.2 of Gelman and Rubin (1992) illustrates SIR samples (although not from the WLB) used successfully as a starting distribution for parallel runs of a Gibbs sampler.

Variation of importance ratios also seems like a potential difficulty in the estimates of the marginal likelihood presented in Section 7. Meng and Wong (1993) present a similar method for estimating marginal likelihoods and normalizing constants that uses samples from two distributions to achieve a lower variance than the harmonic mean estimator.

Let $p_i(\theta)$, $\theta \in \Theta_i$, $i=1$, 2, be two densities, each of which is known up to a normalizing constant: $p_i(\theta)=q_i(\theta)/z_i$, $i=1$, 2. The following identity is fundamental to our approach:

$$\frac{z_1}{z_2}=\frac{E_2\{q_1(\theta)\,\alpha(\theta)\}}{E_1\{q_2(\theta)\,\alpha(\theta)\}}, \tag{20}$$

where E_i denotes the expectation with respect to the p_i ($i=1$, 2) and $\alpha(\theta)$ is an arbitrary function defined on the common support, $\Theta_1 \cap \Theta_2$, such that the two expectations are finite and non-zero. Given draws from both densities and a choice of α, the numerator and denominator on the right-hand side of equation (20) can be simulated easily.

The 'harmonic mean' (13) corresponds to choosing $q_1(\theta)=p(x\,|\,\theta)\,p(\theta)$, $q_2(\theta)=p(\theta)$ and $\alpha(\theta)=\{p(x\,|\,\theta)\,p(\theta)\}^{-1}$. Similarly, equation (14) corresponds to the same choice of q_i, $i=1$, 2, with $\alpha(\theta)=p(\theta)^{-1}$. For these choices of α, as the authors noted, the resulting simulation may be unstable because the corresponding integrands are not necessarily square integrable. Furthermore, they provide legitimate estimates only when the prior $p(\theta)$ is proper. By suitable choices of α in equation (20), however, all these problems can be avoided. For example, choosing $\alpha=1/\surd(q_1 q_2)$ leads to

$$\frac{z_1}{z_2}=\frac{E_2\,[\surd\{q_1(\theta)/q_2(\theta)\}]}{E_1\,[\surd\{q_2(\theta)/q_1(\theta)\}]}, \tag{21}$$

where both integrands are always square integrable with respect to the corresponding densities. Implementing equation (20) with the optimal choice of α is also straightforward, as is detailed in Meng and Wong (1993). Continuous extensions of these methods are presented in Gelman and Meng (1993).

A. P. Grieve (ZENECA Pharmaceuticals, Macclesfield): On the basis of their experiences the authors suggest that the overdispersion parameter α should be chosen 'rather close to 1'. How are we to interpret this in the light of their use of values for α of 0.7 and 1.6? Is this effectively a recommendation to use purely uniform Dirichlet weights?

To understand to what extent a blanket adoption of uniform Dirichlet weights is reasonable I have applied the weighted likelihood bootstrap (WLB) to the family of distributions

$$\exp\left\{\frac{X\theta-b(\theta)}{a(\phi)}+c(X,\,\phi)\right\}$$

and have assumed

(a) that ϕ is known, which gives the exponential family with canonical parameter θ, and
(b) that interest centres on making inferences about the mean value parameter $\mu=E(X)=b'(\theta)$.

For data x_i ($i=1, \ldots, n$) and a given set of random uniform Dirichlet weights y_i ($i=1, \ldots, n$), the WLB gives, as a random estimate of μ,

$$\tilde{\mu}=\sum_{i=1}^{n} y_i x_i. \tag{22}$$

Using standard properties of the uniform distribution on an n-dimensional simplex it is possible to show that

$$E(\tilde{\mu})=\bar{x}, \qquad \operatorname{var}(\tilde{\mu})=\frac{n-1}{n+1}\,\frac{s^2}{n}.$$

There are two features of this result which are potentially disturbing, both of which arise because the WLB does not use all available information. Firstly, irrespective of which member of the exponential family I am interested in, the WLB simulates from the same data-based distribution, although the posteriors are very different in shape. Secondly in the concrete example of a normal distribution with known variance σ^2 we would presumably be wanting to simulate from a distribution with mean $\bar{x}$ and