ensures good convergence for the expectation of any bounded function. Can the authors demonstrate a similar result for their procedure? We believe that a requirement of asymptotic normality can be used to clarify large subsections of the literature on Bayesian simulations. It can, for example, suggest situations where the Gibbs sampler would converge within a practical time limit.

We congratulate the authors on some outstandingly novel ideas.

Jun S. Liu and Donald B. Rubin (Harvard University, Cambridge): Newton and Raftery's examples appear to provide striking evidence for the potential utility of the weighted likelihood bootstrap (WLB) for simulating non-normal likelihoods or posterior distributions, but their arguments do not provide explanations for this subasymptotic effect. Their implicit claim is that the WLB tracks the posited likelihood of θ, whereas it approximately tracks the posterior distribution of θ under a discrete approximation to the model that generated the data, irrespective of the posited model!

Consider two likelihoods, normal(θ, 1) and exponential(θ), where θ is the population mean, with fixed data. For both posited likelihoods, the value of θ being simulated is the weighted mean of the sample values, where the weights are independent of the posited likelihood; hence, the WLB distribution of θ is the same when the likelihood is normal as when it is exponential. Because the implied WLB specification for the data is a discrete approximation to the model that generated the data, the WLB distribution of θ tends to follow the shape of the posited likelihood for θ when the empirical distribution of the data approximates the model underlying this likelihood. Thus, regardless of which likelihood is posited, if the data look like a normal sample (or exponential sample), the WLB distribution of θ will tend to look like the posterior distribution of θ under a normal likelihood (or exponential likelihood) with a diffuse prior on θ. For large n, both WLB distributions look normal but will only have the correct scales if $\alpha \rightarrow 1$, thus implying that the Bayesian bootstrap (BB) specification is the only asymptotically acceptable weight distribution. Because this large sample restriction on the WLB holds in general, in the following general argument we assume a BB weight distribution.

When doing the WLB, a distribution function P with point masses on the observed data points is generated by the BB, denoted $P \sim [P|X, \text{BB}]$. Let $M_0 = \{f_\theta : \theta \in \Theta\}$ denote the posited model and $\mathscr{F}$ the space of all distributions. The maximization step in the WLB is equivalent to finding a θ such that the distance from f_θ to P is minimized. The model assumption M_0 affects the WLB only by inducing a particular mapping from $\mathscr{F}$ to the parameter space Θ, i.e. $\tilde{\theta} = \theta(P)$. Conditional on a fixed data set, and given a fixed function $\tilde{\theta}$, the WLB distribution of $\tilde{\theta}$ is the same for all posited models that induce the same function $\tilde{\theta}$. If the space $\mathscr{F}$ is partitioned into different classes of models indexed by M, then

$$[\theta(P)|X, \text{BB}] = \int [\theta(P)|X, \text{BB}, M]\,[M|X, \text{BB}]\,\mathrm{d}M = \int [\theta(P)|X, M]\,[M|X, \text{BB}]\,\mathrm{d}M,$$

because $[\theta(P)|X, \text{BB}, M] = [\theta(P)|X, M]$ for all M with positive support for the observed X. Thus, the WLB draws $\tilde{\theta}$ from a posterior distribution that mixes over all possible models under a diffuse prior. If models that are relatively well supported by the data under the BB specification, i.e. models with relatively large values of $[M|X, \text{BB}]$, yield posterior distributions $[\theta(P)|X, M]$ similar to $[\theta(P)|X, M_0]$, then the WLB distribution of $\tilde{\theta}$ will be close to $[\theta(P)|X, M_0]$, which is the targeted posterior distribution of θ.

Albert Y. Lo (State University of New York, Buffalo): The paper demonstrates advantages of the weighted likelihood bootstrap (WLB) for posterior inference in smooth parametric models. The choice of non-Dirichlet weights, i.e. non-exponential Y_i, affects the quality of the WLB approximations and is particularly interesting. It turns out that the accuracy of the WLB depends on the weights only through the coefficient of skewness of Y_1. (This is also found when using non-exponential weights in Rubin's (1981) Bayesian bootstrap; see Lo (1991, 1993).) In the WLB setting, the maximization of the weighted likelihood amounts to finding the roots θ^* of $\Sigma\, Y_i l_i'(\theta) = 0$, where $l_i'(\theta) = (\partial/\partial\theta) \log f(X_i|\theta)$ and $Y_1, \ldots, Y_n$ are independent and identically distributed non-negative random variables. A Taylor argument gives

$$\sigma(Y_1)^{-1} n^{1/2} I_n(\hat{\theta})(\theta^* - \hat{\theta}) = n^{-1/2}\, \Sigma\{Y_i/\sigma(Y_1)\}\, l_i'(\hat{\theta}) + R_n. \tag{23}$$

Conditional on the data, the distribution of $n^{-1/2}\,\Sigma\{Y_i/\sigma(Y_1)\}\, l_i'(\hat{\theta})\{n^{-1}\, \Sigma\, l_i'(\hat{\theta})^2\}^{-1/2}$, and hence of $\sigma(Y_1)^{-1} n^{1/2} I_n(\hat{\theta})^{1/2}(\theta^* - \hat{\theta})$, has an expansion