Statistica Sinica **22** (2012), 1233-1251
doi:http://dx.doi.org/10.5705/ss.2010.221

α-STABLE LIMIT LAWS FOR HARMONIC MEAN ESTIMATORS OF MARGINAL LIKELIHOODS

Robert L. Wolpert and Scott C. Schmidler

Duke University

Abstract: The task of calculating marginal likelihoods arises in a wide array of statistical inference problems, including the evaluation of Bayes factors for model selection and hypothesis testing. Although Markov chain Monte Carlo methods have simplified many posterior calculations needed for practical Bayesian analysis, the evaluation of marginal likelihoods remains difficult. We consider the behavior of the well-known *harmonic mean estimator* (Newton and Raftery (1994)) of the marginal likelihood, which converges almost-surely but may have infinite variance and so may not obey a central limit theorem.

We illustrate the convergence in distribution of the harmonic mean estimator in typical applications to a one-sided stable law with characteristic exponent $1 < \alpha < 2$. While the harmonic mean estimator does converge almost surely, we show that it does so at rate $n^{-\epsilon}$ where $\epsilon = (\alpha - 1)/\alpha$ is often as small as 0.10 or 0.01. In such a case, the reduction of Monte Carlo sampling error by a factor of two requires increasing the Monte Carlo sample size by a factor of $2^{1/\epsilon}$, or in excess of $2.5 \cdot 10^{30}$ when $\epsilon = 0.01$, rendering the method entirely untenable. We explore the possibility of estimating the parameters of the limiting stable distribution to provide accelerated convergence.

Key words and phrases: Alpha stable, Bayes factors, bridge sampling, harmonic mean, marginal likelihood, model averaging.

1. Introduction

The task of calculating marginal likelihoods arises in a wide array of statistical inference problems. Models with missing or censored data, or latent variables, require computation of marginal likelihoods in the process of (often iterative, numerical) likelihood maximization. In the Bayesian paradigm, hypotheses are tested and models selected by evaluating their posterior probabilities or calculating *Bayes factors* (Kass and Raftery (1995)), and predictions made by averaging over models weighted by their associated posterior probabilities (Clyde and George (2004)). In each case the key quantity is the marginal probability density function $f_m(x)$ at the observed data vector x for each model m under consideration, and only in simple special cases can this be obtained analytically.

Although Markov chain Monte Carlo methods have broadened dramatically the class of problems that can be solved numerically using Bayesian methods, the problem of evaluating $f_m(x)$ remains difficult and has received considerable attention (Gelfand and Dey (1994); Newton and Raftery (1994); Chib (1995); Meng and Wong (1996); Raftery (1996); Gelman and Meng (1998); Chib and Jeliazkov (2001); Han and Carlin (2001); Meng and Schilling (2002); Sinharay and Stern (2005)). Newton and Raftery (1994) note that the *harmonic mean identity* $f_m^{-1} = \mathrm{E}_{\pi} f_m(x \mid \theta)^{-1}$ (where $\pi(\theta) = f_m(x \mid \theta)\pi_0(\theta)/f_m(x)$ denotes the posterior density for prior $\pi_0(\theta)$) implies that for (possibly dependent) posterior samples $\{\theta_j\} \sim \pi(\theta)\,d\theta$, the sample means of the inverse likelihood function $f_m(x \mid \theta_j)^{-1}$ converge almost surely to $f_m(x)^{-1}$, yielding the *harmonic mean estimator* (HME) for $f_m(x)$:

$$\hat{f}_m(x) \stackrel{\text{def}}{=} \Big(\frac{1}{n}\sum_{j=1}^{n} f_m(x \mid \theta_j)^{-1}\Big)^{-1}. \tag{1.1}$$

Newton and Raftery also note that $f_m(x\,|\,\theta_j)^{-1}$ can have infinite variance, in which case a central limit theorem does not apply to the partial sums S_n of the $f_m(x \mid \theta_j)^{-1}$ in (1.1). In fact, $f_m(x \mid \theta_j)^{-1}$ has finite variance only when $f_m(x \mid \theta)^{-1}$ is square-integrable with respect to $\pi(d\theta)$, i.e., when $\int \big\{\pi_0(\theta)/f_m(x \mid \theta)\big\}\,d\theta < \infty$; this can only happen when the prior $\pi_0(\theta)$ is *less diffuse* than the likelihood $f_m(x \mid \theta)$. Although this situation almost never arises in practice, nevertheless the HME remains popular due to its seductive simplicity, and it is used widely in such applications areas as Bayesian phylogenetics (Huelsenbeck and Ronquist (2001); Nylander et al. (2004); Drummond and Rambaut (2007)). Indeed, despite awareness that the HME may fail to obey a central limit theorem (Lartillot and Philippe (2006)), its use continues to be recommended in the literature. For example, Nylander et al. (2004) say "Although some workers have questioned the general stability of the harmonic mean estimator, it should be sufficiently accurate for comparison of models with distinctly different model likelihoods given that the sample from the posterior distribution is large", and support this by citing a popular introductory text on MCMC methods (Gamerman and Lopes (2006, Sec. 7.2.1)) that in turn suggests "The simplicity of [the HME] makes it a very appealing estimator and its use is recommended provided the sample is large enough." The HME also continues to be the subject of research (Raftery et al. (2007)). It is our goal in this paper to show that it is nearly always a practical impossibility to draw samples that are "large enough".

The behavior of the HME is also relevant to other methods aimed at related problems. For example, *bridge sampling* (Meng and Wong (1996)) estimates the ratio of normalizing constants for two unnormalized densities $p_1(z) = c_1^{-1}q_1(z)$ and $p_2(z) = c_2^{-1}q_2(z)$ by the identity

$$\lambda \stackrel{\text{def}}{=} \frac{c_1}{c_2} = \frac{\mathrm{E}_2(q_1(z)g(z))}{\mathrm{E}_1(q_2(z)g(z))} \approx \frac{n_1\sum_{j=1}^{n_2} q_1(z_{2j})g(z_{2j})}{n_2\sum_{j=1}^{n_1} q_2(z_{1j})g(z_{1j})} \stackrel{\text{def}}{=} \hat{\lambda}, \tag{1.2}$$

where $z_{ij} \sim p_i(z)\,dz$, for any $g(\cdot)$ such that $0 < \int_\Omega p_1(z)p_2(z)\,|g(z)|\,dz < \infty$. For $q_j(z)$ sharing identical compact support, the generalized HME is the special case in which $g(z) = (q_1(z)q_2(z))^{-1}$. This suggests that other choices of $g(\cdot)$ may also lead to similar problems. The numerator or denominator in (1.2) has infinite variance whenever

$$\int g(z)^2 p_1(z)^2 p_2(z)\,dz = \infty \qquad \text{or} \qquad \int g(z)^2 p_1(z) p_2(z)^2\,dz = \infty, \tag{1.3}$$

respectively (for example, if $g(z)^2 p_1(z)^2 p_2(z) \asymp k|z|^{-\alpha}$ as $z \to \infty$ for some $\alpha < 1$). Thus bridge sampling with a poor choice of $g(\cdot)$ may suffer problems similar to those of the HME. (Note however that Meng and Wong (1996) recommend iterative refinement of $\hat{\lambda}$, leading to finite variance on the second iteration.) Similarly, *path sampling* (Gelman and Meng (1998)) (or *thermodynamic integration*), with finite step sizes uses a sequence of distributions $p_1 = \tilde{p}_1$, $\tilde{p}_2, \ldots, \tilde{p}_k = p_2$, and may have $\mathsf{var}(\hat{\lambda}) = \infty$ if condition (1.3) occurs for *any* pair $(\tilde{p}_j, \tilde{p}_{j+1})$, $j \in \{1, \ldots, k\}$. Gelfand and Dey (1994) propose a generalization of the HME given by $\mathrm{E}_\pi\,\{\tau(\theta)/[f_m(x \mid \theta)\pi_0(\theta)]\}$ for an arbitrary proper density function $\tau(\cdot)$, which will be similarly problematic for any choice of $\tau(\cdot)$ for which $\int \tau(\theta)^2/\pi(\theta)^2\,d\theta = \infty$ — i.e., if τ has much heavier tails than the posterior (quite possible, with the t densities they recommend).

Other methods where similar effects are likely to arise include methods for handling intractable normalizing constants in likelihood functions (Geyer (1991); Geyer and Thompson (1992); Geyer (1994)), including the recent algorithm of Møller et al. (2006) involving exact sampling and its generalizations (Andrieu et al. (2007)). All of these methods have in common the use of importance ratios which, if unbounded, can lead to infinite variance and inapplicability of a central limit theorem.

In this paper conditions are given and examples are shown to illustrate that $f_m(x \mid \theta_j)^{-1}$ may lie in the domain of attraction of a one-sided α-stable law of index (or characteristic exponent) $\alpha \in (1, 2]$. Only in problems with precise prior information and diffuse likelihoods is $\alpha = 2$, where the Central Limit Theorem applies and the S_n have a limiting Gaussian distribution with sample means converging at rate $n^{-1/2}$. In typical applications where the sample information exceeds the prior information, the limit law is stable of index α close to one, and convergence is very slow at rate $n^{-\epsilon}$ for $\epsilon = 1 - \alpha^{-1}$ close to zero.

In Section 2 we review the properties of α-stable laws we need. In Section 3 we introduce a sequence of increasingly realistic illustrative examples where the α-stable behavior of the HME can be studied analytically. In Section 4 we explore the possibility of improving on the HME's slow convergence rate by exploiting the approximately α-stable nature of its partial sums, but find that this approach remains uncompetitive with other existing alternatives. In Section 5 we discuss the implications of these results, including the emphatic recommendation that the harmonic mean estimator should not be used.

2. Stable Laws

Last century Paul Lévy (1925, Chap. VI) proved that the only possible limiting distributions for re-centered and rescaled partial sums $S_n = \sum_{j \le n} Y_j$ of independent identically-distributed random variables $\{Y_j\}$ are the *stable laws* $(S_n - a_n)/b_n \Rightarrow Z \sim \mathsf{St}_{\mathsf{A}}(\alpha, \beta, \gamma, \delta)$ with characteristic functions (in the (A) parametrization of Zolotarev (1986, *p.* 9))

$$\mathrm{E}\,[e^{i\omega Z}] = \begin{cases} \exp\left(i\delta\gamma\omega - \gamma|\omega|^{\alpha}\{1 - i\beta\,\mathrm{sgn}\,\omega\ \tan\frac{\pi\alpha}{2}\}\right) & \alpha \ne 1 \\ \exp\left(i\delta\gamma\omega - \gamma|\omega|\ \ \{1 + i\beta\,\mathrm{sgn}\,\omega\ \frac{2}{\pi}\log|\omega|\}\right) & \alpha = 1 \end{cases}, \tag{2.1}$$

for some index $\alpha \in (0, 2]$, skewness $-1 \le \beta \le 1$, rate $\gamma > 0$, and location $-\infty < \delta < \infty$. For non-zero β this family has a sharp discontinuity at $\alpha = 1$ (Cheng and Liu (1997)) because of the tangent term. The (M) parametrization $\mathsf{St}_{\mathsf{M}}(\alpha, \beta, \gamma, \delta) = \mathsf{St}_{\mathsf{A}}(\alpha, \beta, \gamma, \delta^*)$ of Zolotarev (1986, *p.* 11) overcomes this by shifting the location to $\delta^* = \delta - \beta\tan(\pi\alpha/2)$ leading to a continuous parametrization in all four parameters (we need this below). This location/scale family has a smooth unimodal density function $f(x)$, but only in a few special cases is it known in closed form. Still both density and distribution functions can be evaluated by numerical inversion of the Fourier transform; in Zolotarev's (M) parametrization, these are

$$f(z) = \frac{1}{\pi}\int_0^{\infty} e^{-\gamma\omega^{\alpha}} \cos\Big[\omega(z - \delta\gamma) + \beta\gamma\tan\tfrac{\pi\alpha}{2}(\omega - \omega^{\alpha})\Big]\, d\omega, \tag{2.2a}$$

$$F(z) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} e^{-\gamma\omega^{\alpha}} \sin\Big[\omega(z - \delta\gamma) + \beta\gamma\tan\tfrac{\pi\alpha}{2}(\omega - \omega^{\alpha})\Big]\, \omega^{-1}\, d\omega. \tag{2.2b}$$

Only for $\alpha > 1$ is $\mathrm{E}\,|Z|$ finite; in this case the mean is

$$\mathrm{E}\,[Z] = \delta^*\gamma = \delta\gamma - \beta\gamma\tan\tfrac{\pi\alpha}{2}.$$

For $\beta = 1$ and α just slightly above one, the situation we encounter, the mean $\mu \equiv \mathrm{E}\,[Z] \asymp \delta\gamma + 2\gamma/[\pi(\alpha - 1)]$ is large while the median M is close to zero, so the distribution is skewed far to the right (see Figure 1a), and estimating δ or $\mathrm{E}\,[Z]$ from sample averages is difficult.

If random variables $\{Y_j\}$ are iid with finite variance and tail probabilities that fall off sufficiently fast that $y^2\,\mathsf{P}[|Y_j| > y] \to 0$, the Central Limit Theorem applies and the only possible limiting distribution of $(S_n - a_n)/b_n$ for $S_n =$

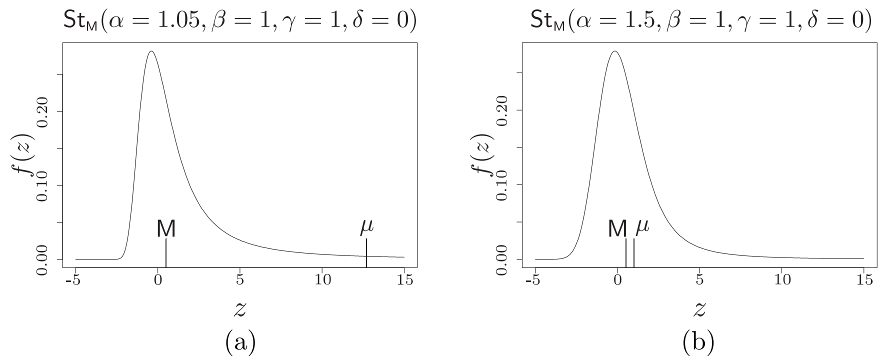


Figure 1. Stable pdf $f(z)$ for (a) $\alpha = 1.05$ and (b) $\alpha = 1.5$, with $\beta = \gamma = 1$ and $\delta = 0$, in Zolotarev's (M) parametrization. Mean and median are indicated by strokes labeled "μ" and "M" respectively.

$\sum_{j\le n} Y_j$ is normal, the special case $\alpha = 2$ of the stable. The limit is stable of index $\alpha \in (0, 2)$ if, instead, $\mathsf{P}[|Y_j| > y] = k(y)\, y^{-\alpha}$ as $y \to \infty$ for a slowly-varying function $k(\cdot) > 0$ (Gnedenko (1939), Feller (1971, IX.8), or Gnedenko and Kolmogorov (1968, Chap. 7, Sec. 35)); it is one-sided (or "fully skewed") stable if also $y^\alpha\, \mathsf{P}[Y_j < -y] \to 0$, whereupon $\beta = 1$. That is the case of interest to us below.

2.1. Stable laws for Markovian sequences

When the Y_j's are not independent but arise from a stationary Markov chain with invariant distribution π, one expects that similar results may hold (i.e., that $(S_n - a_n)/b_n$ can only have normal or α-stable limits in distribution) so long as any dependence for $Y_i,\ Y_j$ decays sufficiently fast in $|i-j|$. Ibragimov and Linnik (1971, Thm. 18.1.1) show that the partial sums of any *strongly mixing* sequence can converge only to a stable distribution with $0 < \alpha \le 2$. When the Y_j's have finite variance, CLTs hold under standard mixing conditions (Kipnis and Varadhan (1986); Chan and Geyer (1994); Roberts and Rosenthal (2004)). In particular if $\mathrm{E}_\pi(|Y_j|^{2+\epsilon}) < \infty$ for some $\epsilon > 0$, or $\epsilon \ge 0$ when the Markov chain is reversible, it suffices that the chain be *geometrically ergodic*, i.e., that for some number $\rho \in (0, 1)$ and function $M : \mathcal{X} \to [0, \infty)$, and all $n \in \mathbb{N}$, the total variation satisfies $\|K^n(x, \cdot) - \pi(\cdot)\| \le M(x)\rho^n$. When $\mathrm{E}_\pi(|Y_j|^2) = \infty$ one might similarly expect convergence to an α-stable law if dependence is not too strong.

Nagaev (1957) proved that when the Y_j form a stationary Markov chain satisfying *Döblin's condition* (Döblin (1938)) with marginal distribution belonging to a stable domain of attraction with $0 < \alpha < 2$, then S_n has a stable limit with index α. Davis (1983) shows this for any stationary sequence satisfying a condition weaker than strong mixing along with an extreme value dependence condition.

As Döblin's condition is equivalent to *uniform* ergodicity (Meyn and Tweedie (1993, Chap. 16)) with constant $M(x) \equiv M$, one suspects that weaker conditions (similar to the geometric ergodicity required for CLTs when $\mathrm{E}_{\pi}|Y_j|^2 < \infty$) may suffice. Recently Jara, Komorowski, and Olla (2009) proved convergence to α-stable for Markov chains under either of two distinct sets of conditions. The first requires geometric ergodicity plus an additional condition that the tails of the marginal transition density do not differ too much from the tails of the stationary distribution; the second does not require geometric ergodicity, but imposes a condition on the tails of coupling times.

At present it is unclear whether these assumptions can be weakened further. Since geometric ergodicity is a typical requirement for a central limit theorem in the finite variance case; however, it seems reasonable to require similar conditions for a stable limit in the infinite variance case.

3. Illustrative Examples

In this section we present a sequence of increasingly realistic illustrative examples where the α-stable behavior of the HME can be studied analytically. Although the HME is not *required* in these examples, since the marginal likelihoods can be obtained in closed form, this allows us to analyze precisely the convergence behavior of the HME. In 3.6, we argue that similar behavior can be expected in essentially all nontrivial cases, where exact calculations are generally unavailable.

3.1. Example 1: Gamma

Let $X_j \sim \mathsf{Ga}(a, \lambda)$ be independent draws from a Gamma distribution with shape parameter a and rate parameter λ, and set $Y_j \equiv \exp(X_j)$. Then

$$\begin{aligned}
\mathrm{E}\,[Y_j{}^p] &= (1 - \frac{p}{\lambda})^{-a} < \infty \qquad \text{if } p < \lambda; \\
\mathsf{P}[Y_j > y] &= \mathsf{P}[\lambda X_j > \lambda \log y] \\
&= \frac{\Gamma(a, \lambda \log y)}{\Gamma(a)} \\
&= (\lambda \log y)^{a-1} \exp(-\lambda \log y) \frac{[1 + O(1/\lambda \log y)]}{\Gamma(a)} \\
&= k(y) y^{-\lambda} \qquad \text{as } y \to \infty,
\end{aligned}$$

where $\Gamma(a, x)$ denotes the incomplete Gamma function (Abramowitz and Stegun (1964, §6.5.32)) and $k(\cdot)$ is slowly-varying. If $\lambda > 2$ then Y_j has finite variance and lies in the normal domain of attraction, while for $\lambda < 2$ the limit is one-sided stable of index $\alpha = \lambda$.

3.2. Example 2: normal

Now let Z_j be independent normal with mean $\mu \in \mathbb{R}$ and variance $V > 0$, and set $Y_j \equiv \exp(c\,Z_j{}^2)$. If $\mu = 0$ then $c\,Z_j{}^2 \sim c\,V\chi_1^2 = \mathsf{Ga}(a,\lambda)$ is Gamma distributed with shape $a = 1/2$ and rate $\lambda = 1/2cV$, so for $V > 1/4c$ the limiting distribution is again one-sided stable of index $\alpha = 1/2cV$, as in Example 1. Even for $\mu \neq 0$ the same limit follows from the calculation

$$\begin{aligned}
\mathsf{P}[Y_j > y] &= \mathsf{P}\big[|Z_j| > \sqrt{\frac{(\log y)}{c}}\big] \qquad \Big(\text{set } \eta \equiv \sqrt{\frac{(\log y)}{c}}\Big) \\
&= \Phi\Big(\frac{-\eta-\mu}{\sqrt{V}}\Big) + \Phi\Big(\frac{-\eta+\mu}{\sqrt{V}}\Big) \\
&\asymp \sqrt{\frac{V}{2\pi}}\left[\frac{\exp\big(-(\eta+\mu)^2/2V\big)}{\eta+\mu} + \frac{\exp\big(-(\eta-\mu)^2/2V\big)}{\eta-\mu}\right] \\
&= k(y)\,e^{-\eta^2/2V} = k(y)\,y^{-1/2cV} \qquad \text{as } y \to \infty,
\end{aligned} \tag{3.1}$$

for a slowly-varying $k(\cdot)$, where $\Phi(z)$ is the cumulative distribution function for the standard normal distribution (Abramowitz and Stegun (1964, §26.2.13)). Again Y_j lies in the domain of attraction of the one-sided stable distribution of index $\alpha = 1/2cV$.

Similarly if $Z_j \overset{\text{i.i.d.}}{\sim} \mathsf{No}_p(\mu,\Sigma)$ are p-variate normal with mean $\mu \in \mathbb{R}^p$ and positive-definite covariance matrix Σ, and $c > 0$, then $Y_j \equiv \exp\big\{c\|Z_j\|^2\big\}$ satisfies

$$\mathsf{P}[Y_j > y] = k(y)\,y^{-1/2cV}$$

for a slowly-varying $k(\cdot)$ with $V = \rho(\Sigma)$, the spectral radius (largest eigenvalue) of Σ. Again Y_j lies in the domain of attraction of the one-sided stable distribution of index $\alpha = 1/2cV$ (note that α does not depend on the mean μ or the dimension p).

3.3. Example 3: testing a normal hypothesis

Let $\{X_k\} \overset{\text{i.i.d.}}{\sim} \mathsf{No}(\theta,\sigma^2)$ be r normally-distributed iid replicates with known variance $\sigma^2 > 0$ but uncertain mean θ. Two models are entertained: M_0, under which $\theta \sim \mathsf{No}(\mu_0,\tau_0^2)$, and M_1, under which $\theta \sim \mathsf{No}(\mu_1,\tau_1^2)$ (the point null hypothesis with $\tau_0 = 0$ is included). The joint and marginal densities for the sufficient statistic $\bar{x}$ from the r replicates $\{X_k\}$ under model $m \in \{0,1\}$ are

$$\begin{aligned}
\pi_m(\theta,\bar{x}) &= (\frac{2\pi\sigma^2}{r})^{-1/2}(2\pi\tau_m^2)^{-1/2}e^{-r(\bar{x}-\theta)^2/2\sigma^2-(\theta-\mu_m)^2/2\tau_m^2}, \\
f_m(\bar{x}) &= [2\pi(\frac{\sigma^2}{r}+\tau_m^2)]^{-1/2}e^{-(\bar{x}-\mu_m)^2/2(\sigma^2/r+\tau_m^2)},
\end{aligned}$$

so under prior probabilities $\pi[M_0] = \pi_0$ and $\pi[M_1] = \pi_1$, the posterior probability of model M_0, and the posterior odds against M_0, are

$$\mathsf{P}[M_0 \mid \vec{x}] = \frac{\pi_0 f_0(\bar{x})}{\pi_0 f_0(\bar{x}) + \pi_1 f_1(\bar{x})} \qquad \text{and} \qquad \frac{\mathsf{P}[M_1 \mid \vec{x}\,]}{\mathsf{P}[M_0 \mid \vec{x}\,]} = \frac{\pi_1}{\pi_0}\frac{f_1(\bar{x})}{f_0(\bar{x})}.$$

Thus the key for making inference and for computing the Bayes factor $B = f_1(\bar{x})/f_0(\bar{x})$ is the computation of each marginal density function

$$f(\bar{x}) = \Big[2\pi(\frac{\sigma^2}{r} + \tau^2)\Big]^{-1/2} e^{-(\bar{x}-\mu)^2/2(\sigma^2/r+\tau^2)} \tag{3.2}$$

at the observed data point $\bar{x}$. Set

$$Y_j \equiv \frac{1}{f(\bar{x} \mid \theta_j)} = \Big(\frac{2\pi\sigma^2}{r}\Big)^{1/2} e^{r(\bar{x}-\theta_j)^2/2\sigma^2} \propto \exp\Big(\frac{r}{2\sigma^2}\,(\theta_j - \bar{x})^2\Big). \tag{3.3}$$

Then Newton and Raftery's Harmonic Mean Estimator (1994)

$$\frac{n}{\sum_{j=1}^n 1/f(\bar{x} \mid \theta_j)} = \frac{1}{\bar{Y}_n} \to f(\bar{x}) \tag{3.4}$$

converges almost surely for any μ and τ. It is our goal to show that the convergence can be *very* slow.

For this conjugate model the Monte Carlo replicates $\{\theta_j\}$ have normal posterior distributions, so for iid draws the Y_j of (3.3) are the same as those of Example 2 above, with $c = r/2\sigma^2$ and $V = (r/\sigma^2 + 1/\tau^2)^{-1}$ (the conditional variance of θ_j given $\bar{x}$). The Central Limit Theorem applies and $\bar{Y}_n$ is asymptotically normal only if

$$\alpha = \frac{1}{2cV} = \frac{1}{2(r/2\sigma^2)(r/\sigma^2 + 1/\tau^2)^{-1}} = \Big(1 + \frac{\sigma^2}{r\tau^2}\Big)$$

exceeds 2, i.e., only if the prior variance τ^2 is *less* than the sampling variance σ^2/r. Otherwise, if $\sigma^2/r < \tau^2$, the limiting distribution of $\bar{Y}_n$ is one-sided α-stable with index $\alpha \in (1, 2)$. As the sample size r increases, so that the data contain substantially more information than the prior, then we are driven inexorably to the α-stable limit, with index $\alpha = (1 + \sigma^2/r\tau^2)$ just slightly above one.

Since $\mathrm{E}\,[Y_j] = 1/f(\bar{x})$, Lévy's limit theorem asserts that

$$\frac{\bar{Y}_n - 1/f(\bar{x})}{n^{(-1+\alpha^{-1})}} \Rightarrow Z \tag{3.5}$$

converges in distribution to a fully-skewed α-stable random variable $Z \sim \mathsf{St}_{\mathsf{A}}(\alpha, 1, \gamma, 0)$ of index $\alpha = (1 + \sigma^2/r\tau^2)$ for some rate $\gamma > 0$, whence

$$\bar{Y}_n \approx \frac{1}{f(\bar{x})} + Z\, n^{-1+\alpha^{-1}}.$$

Although $\bar{Y}_n \to 1/f(\bar{x})$ almost surely as $n \to \infty$, the convergence is only at rate $n^{-\epsilon}$ for $\epsilon = 1 - \alpha^{-1} = (1 + r\tau^2/\sigma^2)^{-1} < \sigma^2/r\tau^2$ (Kuske and Keller (2001)), and moreover the errors $Zn^{-1+\alpha^{-1}}$ have thick-tailed distributions with infinite moments of all orders $p \geq \alpha$ (Samorodnitsky and Taqqu (1994, Eqn. 1.2.8 and Prop. 1.2.15, 1.2.16)).

This has enormous consequences. For typical Monte Carlo applications in which a Central Limit Theorem applies, errors may be reduced by a factor of two (for example) by increasing the Monte Carlo sample size by a factor of just $2^2 = 4$. For the HME, this would instead require that the sample size be increased by a factor of $2^{1/\epsilon}$— a factor of 2,048 if the prior variance τ^2 is just ten times that of the likelihood σ^2/r, or of $2.53 \cdot 10^{30}$ in the common case of a prior distribution chosen to be be "vague" enough to have a variance exceeding that of the likelihood by a factor of one hundred.

Figure 2 illustrates the situation. On the left, Figure 2a shows a sequence of one million terms $Y_j = 1/f(\bar{x} \mid \theta_j)$ as faint spots (the five largest are circled for emphasis) along with a curve indicating their running average $\bar{Y}_n$, the inverse of the HME. The dashed horizontal line is the asymptote $1/f_m(\bar{x}) = 3.3005$. Figure 2b shows ten overlaid plots of a million steps each for the Harmonic Mean Estimators $1/\bar{Y}_n$ as solid curves, along with their asymptotic value $f_m(\bar{x}) = 0.3030$ as a thick dashed curve. Vertical scale is logarithmic; strikingly slow convergence is evident.

3.4. Example 4: linear models

Consider Bayesian analysis of the linear regression model $y = X\beta + \epsilon$ for design matrix $X_{r\times p}$ and iid measurement errors $\{\epsilon_i\} \overset{\text{i.i.d.}}{\sim} \mathsf{No}(0, \sigma^2)$. With σ^2 known and a standard conjugate $\mathsf{No}(\beta_0, \Sigma_0)$ prior distribution for β, the posterior density is normal $\pi(\beta \mid y, X) = \mathsf{No}_p(\beta_y, \Sigma_y)$ with mean and covariance

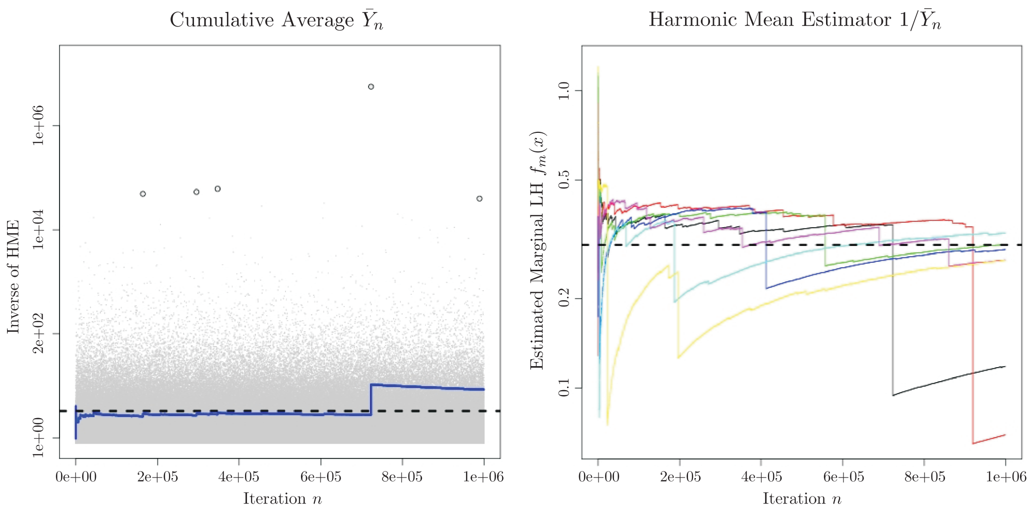
$$\begin{aligned}
\beta_y &= \beta_0 + \Sigma_0 X'[X\Sigma_0 X' + \sigma^2]^{-1}\,[y - X\beta_0],\\
\Sigma_y &= \Sigma_0 - \Sigma_0 X'[X\Sigma_0 X' + \sigma^2]^{-1} X\Sigma_0,
\end{aligned}$$

so $Z = y - X\beta$ has conditional (given y) distribution

$$y - X\beta \sim \mathsf{No}\big(y - X\beta_y,\ \sigma^2\, X\Sigma_0 X'\ [X\Sigma_0 X' + \sigma^2]^{-1}\big).$$

The covariance matrix for $Z \equiv y - X\beta$ has spectral radius $V = \sigma^2 R/(R + \sigma^2)$ where $R = \rho(X\Sigma_0 X')$ is the largest eigenvalue of $X\Sigma_0 X'$. The inverse likelihood function of β for the observed y is

$$f(y \mid \beta, X)^{-1} = (2\pi\sigma^2)^{\frac{n}{2}} \exp\left\{\frac{1}{2\sigma^2}\|y - X\beta\|^2\right\} \propto \exp(c\|Z\|^2)$$



(a) (b)

Figure 2. HME illustrations for samples of size r from Normal $\{X_i\} \overset{\text{i.i.d.}}{\sim} \mathsf{No}(\theta, \sigma^2)$ distribution with conjugate $\theta \sim \mathsf{No}(0, \tau^2)$ prior, with $\tau^2 = 10(\sigma^2/r)$. Dots in (a) indicate individual values Y_j (five largest are circled), solid curve shows cumulative average. Curves in (b) show ten independent replicates of HME $1/\bar{Y}_n$. Dashed horizontal lines are asymptotes $1/f_m(\bar{x}) \approx 3.3005$ and $f_m(\bar{x}) \approx 0.3030$, respectively. Vertical axis is displayed in log scale.

for $c = 1/2\sigma^2$, so for iid draws $\beta_j \sim \mathsf{No}(\beta_y, \Sigma_y)$ from the posterior distribution the quantity $Y_j = f(y \mid \beta_j, X)^{-1}$ is in the domain of attraction of the one-sided α-stable distribution with index $\alpha = 1/2cV = 1 + \sigma^2/\rho(X\Sigma_0 X')$ whenever $\alpha < 2$. The Central Limit Theorem applies and $\bar{Y}_n$ converges at rate $n^{-1/2}$ only when $\sigma^2 \geq \rho(X\Sigma_0 X')$, i.e., when the measurement error variance exceeds the prior predictive variance. In the more typical case where $\sigma^2 \ll R = \rho(X\Sigma_0 X')$, each reduction of Monte Carlo sampling error by a factor or two would require that the number of MC samples be increased by a factor of $2^{1+R/\sigma^2}$.

3.5. Example 5: natural exponential families

The distribution of iid random variables $\{X_j\}$ taking values in some measurable space $\mathcal{X}$ is a *natural exponential family* if it has density functions of the form

$$f(x \mid \theta) = e^{\theta \cdot T(x) - A(\theta)}\, h(x)$$

for some p-dimensional statistic $T : \mathcal{X} \to \mathbb{R}^p$, real-valued $A : \Theta \to \mathbb{R}$, and nonnegative $h : \mathcal{X} \to \mathbb{R}_+$ (Brown (1986, Chap. 1)). It is well-known that $T_+ \equiv \sum T(x_i)$ is a sufficient statistic for an iid random sample $\vec{x} = (x_1, \cdots, x_r)$, that (under regularity conditions) $\nabla A(\hat{\theta}) = \bar{T} \equiv T_+/r$ at the maximum likelihood estimator (MLE) $\hat{\theta}(\vec{x})$, and that the single-observation Fisher information matrix coincides with the observed information, $I_\theta = \nabla^2 A(\theta)$. If A is smooth it follows that

$$A(\hat{\theta} + \varepsilon) = A(\hat{\theta}) + \varepsilon \cdot \bar{T} + \tfrac{1}{2}\varepsilon' I_{\hat{\theta}}\, \varepsilon + o(|\varepsilon|^2) \tag{3.6}$$

for $\hat{\theta},\ \hat{\theta} + \varepsilon \in \Theta$. Such a family admits a conjugate family of prior distributions with density function

$$\pi_0(\theta) = e^{\theta \cdot \tau - \beta A(\theta) - c(\tau, \beta)}$$

indexed by those *shape parameters* $\tau \in \mathbb{R}^p$ and *prior sample sizes* $\beta \in \mathbb{R}$ for which $c(\tau, \beta) \equiv \log \int_\Theta e^{\theta \cdot \tau - \beta A(\theta)}\, d\theta$ is well-defined and finite. The posterior density for θ in this conjugate family is

$$\pi_r(\theta) = e^{\theta \cdot (\tau + T_+) - (\beta + r) A(\theta) - c(\tau + T_+,\, \beta + r)},$$

which attains its maximum at the maximum a posteriori (MAP) estimator $\tilde{\theta}$ where $(\beta + r)\nabla A(\tilde{\theta}) = (\tau + r\bar{T})$. By Taylor's theorem it follows that the posterior distribution of $\zeta \equiv \sqrt{r}(\theta - \tilde{\theta})$ has density

$$\begin{aligned} \zeta \mid \vec{x} &\sim r^{-p/2}\, e^{(\tilde{\theta} + \zeta/\sqrt{r}) \cdot (\tau + r\bar{T}) - (\beta + r) A(\tilde{\theta} + \zeta/\sqrt{r}) - c(\tau + r\bar{T},\, \beta + r)} \\ &\propto e^{-\frac{1}{2}(1 + \beta/r)\zeta' I_{\tilde{\theta}} \zeta + o(1/r)} \qquad \text{as } r \to \infty, \end{aligned}$$

so θ has approximately a normal posterior distribution with mean $\tilde{\theta}$ and covariance $\{(r + \beta) I_{\tilde{\theta}}\}^{-1}$, while the inverse likelihood at $\theta_j \sim \pi_r(\theta)$ is

$$f(\vec{x} \mid \theta_j)^{-1} \propto e^{r[A(\theta_j) - \theta_j \cdot \bar{T}]} \propto e^{\frac{r}{2}(\theta_j - \tilde{\theta})' I_{\tilde{\theta}} (\theta_j - \tilde{\theta}) + o(1)} \qquad \text{as } r \to \infty.$$

By Section 3.2 the inverse HME $\bar{Y}_n$ converges in distribution to the α-stable $\mathsf{St_A}(\alpha, 1, \gamma^{1-\alpha^{-1}}, f(\vec{x})^{-1})$ distribution with mean $f(\vec{x})^{-1}$ and index $\alpha = 1 + \beta/r$ whenever $r > \beta$, so $(\bar{Y}_n - f(\vec{x})^{-1})$ has approximately the same distribution as $n^{-\epsilon} Z$ for some $Z \sim \mathsf{St_A}(\alpha, 1, \gamma, 0)$. Again $\bar{Y}_n$ (resp., the HME $1/\bar{Y}_n$) converges to $f(\vec{x})^{-1}$ (resp., $f(\vec{x})$) at rate $n^{-\epsilon}$.

For example, if $\{X_j\} \overset{\text{i.i.d.}}{\sim} \mathsf{Bi}(k, p)$ have binomial distributions and p the conjugate prior $p \sim \mathsf{Be}(a, b)$, then $\bar{Y}_n$ has an approximate α-stable distribution with index $\alpha = 1 + (a + b)/(kr)$, while if $\{X_j\} \overset{\text{i.i.d.}}{\sim} \mathsf{Po}(\lambda)$ are Poisson distributed with $\lambda \sim \mathsf{Ga}(a, b)$, then $\alpha = 1 + b/r$ and convergence is at rate $n^{-\epsilon}$ for $\epsilon = (1 + r/b)^{-1} < b/r$.

3.5.1. Non-conjugate priors

Of course the HME is unnecessary for conjugate distributions in exponential families, where the marginal likelihood $f(\vec{x}) = e^{c(\tau+T_+,\beta+r)-c(\tau,\beta)}$ is available explicitly, but the prior distributions commonly recommended for such problems (e.g., Ramsay and Novick (1980)) have *heavier* tails than conjugate priors, and so may be expected to lead to even slower convergence.

Consider replacing the conjugate prior $\theta \sim \mathsf{No}(\mu, \sigma^2)$ in the normal mean example $\{X_i\} \overset{\text{i.i.d.}}{\sim} \mathsf{No}(\theta, \sigma^2)$ of Section 3.3, for example, with a noncentral Student t prior distribution $\theta \sim t_\nu(\mu, \tau^2)$ with $\nu > 0$ degrees of freedom. We can write this prior in hierarchical form as $\theta \mid \zeta \sim \mathsf{No}(\mu, \tau^2/\zeta)$ with $\zeta \sim \mathsf{Ga}(\nu/2, \nu/2)$ and use (3.1) with $c = r/2\sigma^2$ and $V = (r/\sigma^2 + \zeta/\tau^2)^{-1}$ to write the posterior tail probabilities for $Y \equiv 1/f(\bar{x} \mid \theta)$ as

$$\begin{aligned}
\mathsf{P}[Y > y \mid \vec{x}, \zeta] &= k(y)\, y^{-1/2cV} = k(y)\, y^{-1-\zeta\sigma^2/r\tau^2}, \\
\mathsf{P}[Y > y \mid \vec{x}] &= k(y)\, y^{-1} \int_0^\infty \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \zeta^{(\nu/2)-1} e^{-\zeta\nu/2-\zeta(\sigma^2/r\tau^2)\log y}\, d\zeta \\
&= k(y)\, y^{-1} \big[1 + (2\frac{\sigma^2}{r}\nu\tau^2) \log y\big]^{-\nu/2} \\
&= \tilde{k}(y)\, y^{-1}
\end{aligned}$$

as $y \to \infty$ for slowly-varying $k(\cdot)$ and $\tilde{k}(\cdot)$. By Lévy's limit theorem (Gnedenko (1939); Döblin (1940)) the sample means $\bar{Y}_n$ converge in distribution to the fully-skewed $\mathsf{St_A}(1, 1, \gamma, 0)$ distribution with $\alpha = \beta = 1$ for some $\gamma > 0$ at a rate slower than any power of n.

3.6. Example 6: Bernstein-von Mises

Under suitable regularity conditions *every* posterior distribution is asymptotically normally distributed (Le Cam (1956, Sec. 6), van der Vaart (1998, Thm. 10.1)), and every likelihood function is asymptotically normal, so stable limiting behavior can be expected to arise in nearly all efforts to apply the Harmonic Mean Estimator to compute Bayes factors for large sample sizes and relatively vague prior information with, for any $\epsilon > 0$, convergence slower than rate $n^{-\epsilon}$ if r is sufficiently large.

4. Exploiting α-Stability

In this section we explore whether knowledge of this α-stable limiting behavior can be exploited to improve convergence of the HME by estimating parameters of the limiting α-stable distribution. Instead of estimating $1/f(\bar{x}) \approx \bar{Y}_n$ directly from ergodic averages, we may try to estimate the uncertain parameters α, γ, δ for the fully-skewed stable $\bar{Y}_n \sim \mathsf{St_M}(\alpha, 1, \gamma n^{1-\alpha}, \delta)$, whereupon we could base an estimate of $1/f(\bar{x}) = \mathrm{E}\,[\bar{Y}_n] = \delta\gamma - \gamma n^{1-\alpha} \tan \pi\alpha/2$ on estimates $\hat{\alpha}$, $\hat{\gamma}$, and $\hat{\delta}$.

4.1. Estimation of stable parameters

A wide variety of methods have been proposed for estimating the parameters α, β, γ, and δ of the α-stable distribution from random samples $\{Z_i\}$. Reviewed by Borak, Härdle and Weron (2005), these include the quantile approach of McCulloch (1986), the maximum likelihood approach of Nolan (2001) (based on numerical evaluation of the pdf using the inverse Fourier transform of the ch.f.), regression methods based on the empirical ch.f. estimate (or "ECE") (Kogon and Williams (1998)), wavelet approaches (Antoniadis, Feuerverger and Gonçalves (2006)), and tail-behavior methods based on extremes (Hill (1975)). Many of these consider only the symmetric ($\beta = 0$, $\delta = 0$) case and most break down when α is close to one.

A promising approach pioneered by Press (1972), improved by Kogon and Williams (1998), described by Borak, Härdle and Weron (2005, Sec. 1.5.3), and further developed by Besbeas and Morgan (2008), begins with the observation that the modulus $|\chi(\omega)| = \exp(-\gamma|\omega|^\alpha)$ of the α-stable ch.f. $\chi(\omega) \stackrel{\text{def}}{=} \mathrm{E}\exp(i\omega Z)$ depends on only two of the parameters, suggesting that one might base estimates $\hat\alpha$ and $\hat\gamma$ on the modulus of the empirical characteristic function

$$\hat\chi(\omega) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{j\le n} e^{i\omega Y_j}$$

at specified points $\omega = \omega_i$ by linear regression of

$$\log\big(-\log|\chi(\omega_i)|\big) = \log\gamma + \alpha\log|\omega_i| \tag{4.1a}$$

on $\log|\omega_i|$. Similarly, $\arg\chi(\omega)$ for the fully-skewed α-stable (in the (M) parametrization) is

$$\begin{aligned}\arctan\left(\frac{\Im(\chi(\omega))}{\Re(\chi(\omega))}\right) &= \delta\gamma\omega - \gamma\tan\tfrac{\pi\alpha}{2}\big(\omega - |\omega|^\alpha\operatorname{sgn}\omega\big)\\ &= \delta\gamma\omega - \gamma\frac{2}{\pi}\omega\log|\omega| - (\alpha-1)\,\frac{w}{\pi}\,(\log|\omega|)^2 + o(\alpha-1)\\ &\approx \delta\gamma\omega - \gamma\frac{2}{\pi}\omega\log|\omega| \quad \text{ for } \alpha\approx 1,\end{aligned} \tag{4.1b}$$

suggesting that a subsequent estimate $\hat\delta$ might be achieved by regressing $\big\{\arg\hat\chi(\omega_i)/\hat\gamma + (2/\pi)\omega_i\log|\omega_i|\big\}$ on ω_i. Note that "$\arg\hat\chi(\omega_i)$" must be evaluated using the principal value of the arctangent, with a branch cut along the negative real axis, due to the periodicity of the tangent function.

To illustrate the method we return to Example 3 from Section 3.3, evaluating the marginal likelihood $f(\bar x)$ for the model $X_i \stackrel{\text{i.i.d.}}{\sim} \mathsf{No}(\theta, \sigma^2 = 1)$ with prior distribution $\theta \sim \mathsf{No}(\mu = 0, \tau^2 = 1)$, for a sample $\{X_i\}$ of size $r = 10$ with sample mean $\bar{x}$. Since $\bar{x}$ is sufficient with likelihood

$$f(\bar{x} \mid \theta) = (\frac{2\pi}{r})^{-1/2} \exp\big\{ - \frac{(\bar{x}-\theta)^2}{2/r}\big\},$$

and $\bar{x}$ has marginal distribution $\bar{x} \sim \mathsf{No}(0, 1+r^{-1})$, the exact marginal likelihood for $r = 10$ is available analytically, $f(\bar{x}) = (2.2\pi)^{-1/2} e^{-\bar{x}^2/2.2}$. The HME will estimate this by $1/\bar{Y}_n$, where $\bar{Y}_n = (1/n)\sum_{j\le n} Y_j$ with $Y_j = 1/f(\bar{x} \mid \theta_j) = \sqrt{\pi/5}\ e^{5(\theta_j - \bar{x})^2}$ for iid draws θ_j from the posterior distribution; here $\theta \mid \bar{x} \sim \mathsf{No}\big(10\ \bar{x}/11, 1/11\big)$. In Section 3.3 we found that Y_j is in the domain of attraction of the α-stable $\mathsf{St_A}(\alpha, 1, \gamma, f(\bar{x})^{-1})$ distribution with $\alpha = 1 + \sigma^2/r\tau^2 = 1.10$, so we should expect the HME to behave badly. The marginal likelihood $f(\bar{x})$ may be expressed in the form

$$f(\bar{x}) = \frac{1}{\mathrm{E}\,[Y_j]} = \frac{1}{\gamma(\delta - \tan \pi\alpha/2)} \approx \frac{1}{\gamma(\delta + 20/\pi)}. \tag{4.2}$$

For illustration we took a typical value of $\bar{x}$, the marginal median value $\bar{x} = 0.7074$ for $|\bar{x}|$, and drew $n = 10^6$ Monte Carlo replicates θ_i from the posterior distribution of θ given $\bar{x}$; we evaluated the HME $1/\bar{Y}_n$ and the ECE estimate $\{\hat{\gamma}[\hat{\delta} - \tan(\pi\hat{\alpha}/2)]\}^{-1}$ from (4.2). Summary statistics from 100 replications of this experiment are given in Table 1. Histograms for MCMC samples of size 10^5, 10^6, 10^7, 10^8 are shown in Figure 3, and empirical convergence rates for their inter-quartile ranges (IQRs) are illustrated in Figure 4. The ECE converges at the CLT rate of $1/\sqrt{n}$ (as expected, since the CLT applies to the estimators $\hat{\alpha}$, $\hat{\gamma}$, and $\hat{\delta}$), substantially faster than the HME.

Although the ECE converges at rate $1/\sqrt{n}$, it does not seem competitive with other methods. Figure 3 also shows the IQR for a simple version of bridge sampling (with q_1 the unnormalized posterior, q_2 the prior, and bridge function $g = 1/q_2$ in (1.2)) equivalent to importance sampling using the prior. This converges at the same $1/\sqrt{n}$ rate as ECE, since the CLT applies, but it requires a sample-size more than 2000 times smaller for the same precision. The relative efficiency would be larger still for optimal bridge sampling (Meng and Wong (1996, Thm. 1, Sec. 3)). Thus it appears that using the knowledge of the HME's α-stable behavior to try and improve on it still does not yield an approach competitive with other methods. For that reason we do not pursue the ECE further.

5. Discussion

It is well-known that the Harmonic Mean Estimator may converge slowly, because its inverse $\bar{Y}_n$ is the sample mean of terms which may fail to be square-integrable and hence the Central Limit Theorem may not apply.

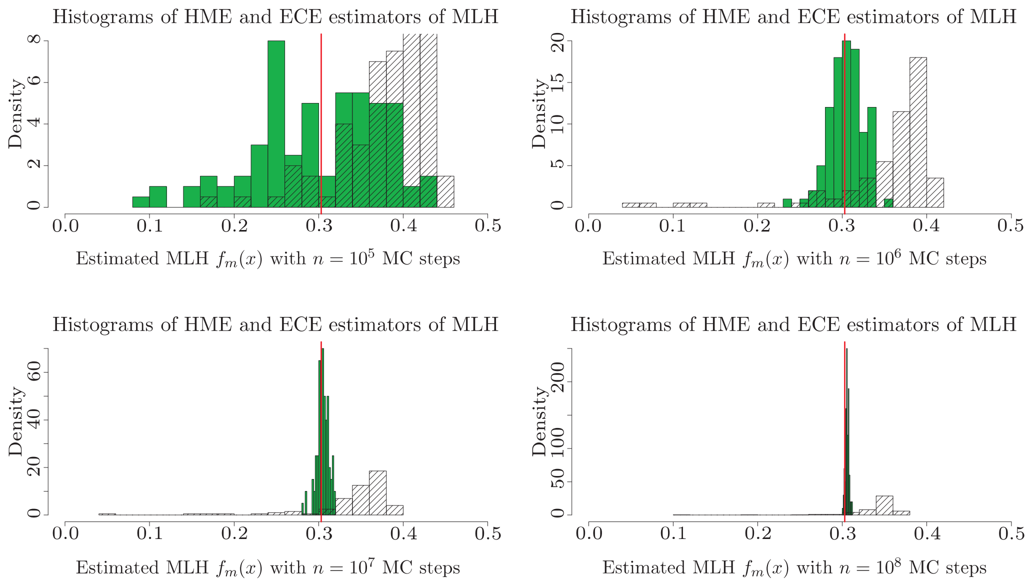


Figure 3. Histogram comparison of 100 replicated estimates of marginal likelihood $f(\bar{x})$ for Harmonic Mean Estimator (diagonal shaded) and Empirical Characteristic function Estimator (solid shaded), each based on a Monte Carlo sample of indicated size. True value is shown as vertical line.

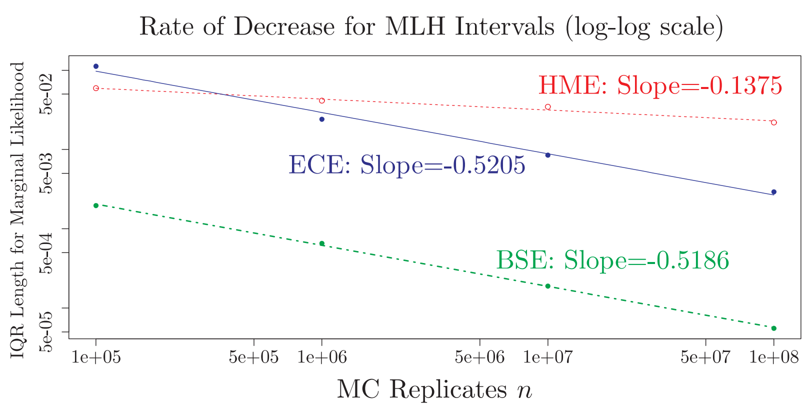


Figure 4. Empirical rates at which posterior Inter-quartile Range widths for the marginal likelihood $f(\bar{x})$ shrinks with Monte Carlo sampling size n, for Harmonic Mean Estimator (HME, dotted OLS regression line suggesting interval length decay rate is about $\propto n^{-0.14}$), Empirical Characteristic function Estimator (ECE, solid line) and simple form of a Bridge Sampling Estimator (BSE, dash-dot line). Both ECE and BSE slopes are within MC sampling error of $-1/2$, showing convergence at rate $1/\sqrt{n}$.

Table 1. Median and inter-quartile range for 100 replicates of estimating Marginal Likelihood $f_0(\bar{x})$ from $n = 10^6$ Monte Carlo replicates for Harmonic Mean Estimator and Empirical Characteristic function Estimator.

	q_{50}	q_{25}	q_{75}
Truth:	0.3030		
HME:	0.3742	[0.3465,	0.3877]
ECE:	0.3055	[0.2933,	0.3174]

We have shown a much stronger cause for alarm about the method— that in a wide variety of applications the inverse $\bar{Y}_n$ of the Harmonic Mean Estimator based on n Monte Carlo replicates (either iid or following a geometrically ergodic Markov chain) converges in distribution to a fully-skewed α-stable probability distribution for some α just slightly above one. Although $\bar{Y}_n$ does converge almost-surely to the marginal likelihood $1/f(\vec{x})$, it does so at rate $n^{-\epsilon}$ for $\epsilon = 1-\alpha^{-1}$ close to zero. For conjugate priors and exponential families, for example, we can evaluate $\epsilon = (1 + r/\beta)^{-1} < \beta/r$ explicitly, where r is the number of iid observations $\{X_j\}$ contributing to the likelihood and where β is the "prior sample size" for the conjugate prior distribution (the shape parameters are immaterial).

Figures 3, 4 suggest that the slow convergence rate might be overcome by exploiting the α-stable nature of the HME partial sums and basing estimates of the marginal likelihood on parameter estimates for the nearly α-stable distribution of summands. While this approach does appear to converge at rate $1/\sqrt{n}$ in the Monte Carlo sample size n, it does not appear to be competitive with other existing approaches such as bridge sampling (Meng and Wong (1996)), for example, since even our toy examples require Monte Carlo sample sizes of tens or hundreds of millions for adequate performance.

Our argument breaks down, and indeed the HME works well, only for problems in which the prior information exceeds that of the data— for example, in conjugate exponential family problems with prior sample size β exceeding the data sample size r, or normal distribution problems in which the prior precision exceeds that of the combined data. In typical applications where the data are tens or hundreds of times more informative than the prior distribution, the convergence of the HME is at rate $n^{-\epsilon}$ for ϵ as small as 0.10 or 0.01. Under these conditions, reducing the Monte Carlo sampling error by a factor of two would require increasing the Monte Carlo sample size by a factor of $2^{1/\epsilon}$ that may easily exceed 10^{30}. We feel this makes the method entirely indefensible. Optimistic suggestions that the HME may work for "sufficiently large Monte Carlo samples" are correct, of course— but securing "sufficiently large" samples may require geological time.

Acknowledgements

The authors would like to thank Xiao-Li Meng, Nicholas Polson, Adrian Raftery, and Christian Robert for helpful conversations and encouragement, and an anonymous referee for helpful suggestions. This work was supported in part by the National Science Foundation under grants DMS-0635449, DMS-0757549, and PHY-0941373, National Institutes of Health grant NIH-1R01GM090201-01, and National Aeronautics and Space Administration grant NNX09AK60G. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, NIH, or NASA.

References

Abramowitz, M. and Stegun, I. A., eds. (1964). *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards, Washington, D.C. Reprinted in paperback by Dover (1974); on-line at `http://www.nr.com/aands/`.

Andrieu, C., Berthelsen, K. K., Doucet, A. and Roberts, G. O. (2007). Posterior sampling in the presence of unknown normalising constants: An adaptive pseudo-marginal approach. Unpublished.

Antoniadis, A., Feuerverger, A. and Gonçalves, P. (2006). Wavelet-based estimation for univariate stable laws. *Ann. Inst. Statist. Math.* **58**, 779-807.

Besbeas, P. and Morgan, B. J. T. (2008). Improved estimation of the stable laws. *Statist. Comput.* **18**, 219-231.

Borak, S., Härdle, W. and Weron, R. (2005). Stable distributions. In *Statistical Tools for Finance and Insurance*, (edited by P. Cízek, W. Härdle, and R. Weron), Chap. 1, 21-44. Springer-Verlag, New York, NY. On-line at `http://fedc.wiwi.hu-berlin.de/xplore/tutorials/`.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families.* Inst. Math. Statist., Hayward, CA.

Chan, K. S. and Geyer, C. J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1747-1758.

Cheng, R. C. H. and Liu, W. B. (1997). A continuous representation of the family of stable law distributions. *J. Roy. Statist. Soc. Ser. B* **59**, 137-145.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc.* **90**, 1313-1322.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *J. Amer. Statist. Assoc.* **96**, 270-281.

Clyde, M. A. and George, E. I. (2004). Model uncertainty. *Statist. Sci.* **19**, 81-94.

Davis, R. A. (1983). Stable limits for partial sums of dependent random variables. *Ann. Probab.* **11**, 262-269.

Döblin, W. (1938). Sur deux problèmes de M. Kolmogoroff concernant les chaînes dénombrables. *Bulletin de la Société Mathématique de France* **66**, 210-220.

Döblin, W. (1940). Sur l'ensemble de puissances d'une loi de probabilité. *Studia Mathematica* **9**, 71-96.

Drummond, A. J. and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Feller, W. (1971). *An Introduction to Probability Theory and its Application*, Vol. 2. Wiley, New York, NY.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, New York, NY.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *J. Roy. Statist. Soc. Ser. B* **56**, 501-514.

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13**, 163-185.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computing Science and Statistics: Proc.* 23$^{\mathrm{rd}}$ *Symp. Interface*, 156-163.

Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *J. Roy. Statist. Soc. Ser. B* **56**, 261-274.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. Roy. Statist. Soc. Ser. B* **54**, 657-699.

Gnedenko, B. V. (1939). On the theory of domains of attraction of stable laws. *Uchenye Zapiski Moskov. Gos. Univ. Matematika* **30**, 61-81.

Gnedenko, B. V. and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA.

Han, C. and Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: A comparative review. *J. Amer. Statist. Assoc.* **96**, 1122-32.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.

Huelsenbeck, J. P. and Ronquist, F. R. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationay Sequences of Random Variables*. Wolters-Noordhoff, Gronigen, DE.

Jara, M., Komorowski, T. and Olla, S. (2009). Limit theorems for additive functionals of a Markov chain. *Ann. Appl. Probab.* **19**, 2270-2300.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773-796.

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1-19.

Kogon, S. M. and Williams, D. B. (1998). Characteristic function based estimation of stable distribution parameters. In *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, (edited by R. J. Adler, R. E. Feldman, and M. S. Taqqu), 311-338. Birkhäuser, Boston, MA.

Kuske, R. A. and Keller, J. B. (2001). Rate of convergence to a stable law. *SIAM J. Appl. Math.* **61**, 1308-1323.

Lartillot, N. and Philippe, H. (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology* **55**, 195-207.

Le Cam, L. M. (1956). On the asymptotic theory of estimation and testing of hypotheses. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, (edited by J. Neyman), vol. 1, 129-156. University of California Press, Berkeley, CA.

Lévy, P. (1925). *Calcul des Probabilités.* Gauthier-Villars, Paris, FR.

McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Commun. Stat. B–Simul.* **15**, 1109-1136.

Meng, X.-L. and Schilling, S. (2002). Warp bridge sampling. *J. Comput. Graph. Statist.* **11**, 552-586.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statist. Sinica* **6**, 831-860.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer-Verlag, New York, NY.

Møller, J., Pettitt, A. N., Berthelsen, K. K. and Reeves, R. W. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93**, 451-458.

Nagaev, S. V. (1957). Some limit theorems for stationary Markov chains. *Theory Probab. Appl.* **2**, 378-406.

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.

Nolan, J. P. (2001). Maximum likelihood estimation of stable parameters. In *Lévy Processes: Theory and Applications* (Edited by O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick), 379-400. Birkhäuser.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P. and Nieves-Aldrey, J. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology* **53**, 47-67.

Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *J. Amer. Statist. Assoc.* **67**, 842-846.

Raftery, A. E. (1996). Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice*, (edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter), 163-188. Chapman and Hall, New York, NY.

Raftery, A. E., Newton, M. A., Satagopan, J. M. and Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In *Bayesian Statistics 8* (Edited by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West), 371-416. Oxford Univ. Press, Oxford, UK.

Ramsay, J. O. and Novick, M. R. (1980). PLU robust Bayesian decision theory: Point estimation. *J. Amer. Statist. Assoc.* **75**, 901-907.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surveys* **1**, 20-71.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York, NY.

Sinharay, S. and Stern, H. S. (2005). An empirical comparison of methods for computing Bayes factors in generalized linear mixed models. *J. Comput. Graph. Statist.* **14**, 415-435.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press, Cambridge, UK.

Zolotarev, V. M. (1986). *One-dimensional Stable Distributions.* American Mathematical Society, Providence, RI.

Department of Statistical Science, Duke University, Durham NC 27708-0251, USA.

E-mail: wolpert@stat.duke.edu

Department of Statistical Science, Duke University, Durham NC 27708-0251, USA.

E-mail: schmidler@stat.duke.edu

(Received October 2010; accepted July 2011)