Adding Weights to the Bernor Package

Yun Ju Sung Charles J. Geyer

September 30, 2005

This document details a trivial modification of the `bernor` package and is itself a trivial modification of the first part of the document "Monte Carlo Likelihood Approximation" supplied in the `doc` directory of that package.

1 Monte Carlo Likelihood Approximation

[Quote from "Monte Carlo Likelihood Approximation"] Let $f_\theta(x, y)$ be the complete data density for a missing data model, the missing data being x and the observed data being y. Suppose we have observed data $y_1, \ldots, y_n$ which are independent and identically distributed (IID) and simulations $x_1, \ldots, x_m$ which are IID from a known importance sampling distribution with density h.

The (observed data) log likelihood for this model is

$$l_n(\theta) = \sum_{j=1}^{n} \log f_\theta(y_j) \tag{1}$$

where

$$f_\theta(y) = \int f_\theta(x, y)\, dx$$

is the marginal for y. [End of Quote]

We modify this to allow for the possibility that y values are repeated many times. Suppose the value y_j is repeated w_j times. Then, purely for reasons of computational efficiency, we can rewrite (1) as

$$l_n(\theta) = \sum_{j=1}^{n} w_j \log f_\theta(y_j). \tag{2}$$

Note that the sample size is now $w_1 + \ldots + w_n$ (not n as before).

[Quote from "Monte Carlo Likelihood Approximation"] The Monte Carlo likelihood approximation for (1) is

$$l_{m,n}(\theta) = \sum_{j=1}^{n} \log f_{m,\theta}(y_j) \tag{3a}$$

where

$$f_{\theta,m}(y) = \frac{1}{m} \sum_{i=1}^{m} \frac{f_\theta(x_i, y)}{h(x_i)}. \tag{3b}$$

The maximizer $\hat{\theta}_{m,n}$ of (3a) is the Monte Carlo (approximation to the) MLE (the MCMLE). [End of Quote]

Of course, corresponding to our rewrite of (1) as (2), we now must rewrite (3a) as

$$l_{m,n}(\theta) = \sum_{j=1}^{n} w_j \log f_{m,\theta}(y_j) \tag{4a}$$

where (3b) remains the same (because it does not involve a sum over y_j).

Derivatives of (4a) are, of course,

$$\nabla^k l_{m,n}(\theta) = \sum_{j=1}^{n} w_j \nabla^k \log f_{m,\theta}(y_j)$$

where ∇ denotes differentiation with respect to θ, and derivatives of (3b) remain as they were given in "Monte Carlo Likelihood Approximation" since (3b) itself has not changed.

2 Asymptotic Variance

The asymptotic variance of $\hat{\theta}_{m,n}$, including both the sampling variation in y_1, …, y_n and the Monte Carlo variation in x_1, …, x_m is

$$J(\theta)^{-1} \left(\frac{V(\theta)}{n} + \frac{W(\theta)}{m} \right) J(\theta)^{-1} \tag{5}$$

where

$$V(\theta) = \operatorname{var}\{\nabla \log f_\theta(Y)\} \tag{6a}$$

$$J(\theta) = E\{-\nabla^2 \log f_\theta(Y)\} \tag{6b}$$

$$W(\theta) = \operatorname{var}\left\{ E\left[\frac{\nabla f_\theta(X \mid Y)}{h(X)} \;\middle|\; X \right] \right\} \tag{6c}$$

where X and Y here have the same distribution as x_i and y_j, respectively. This is the content of Theorem 3.3.1 in the first author's thesis.

The first two of these quantities have obvious "plug-in" estimators

$$\widehat{V}_{m,n}(\theta) = \frac{1}{w_1 + \cdots + w_n} \sum_{j=1}^{n} w_j \big(\nabla \log f_{\theta,m}(y_j)\big)\big(\nabla \log f_{\theta,m}(y_j)\big)^T \tag{7a}$$

$$\widehat{J}_{m,n}(\theta) = -\frac{1}{w_1 + \cdots + w_n} \sum_{j=1}^{n} w_j \nabla^2 \log f_{\theta,m}(y_j) \tag{7b}$$

The quantity (6c) has a natural plug-in estimator

$$\widehat{W}_{m,n}(\theta) = \frac{1}{m} \sum_{i=1}^{m} \widehat{S}_{m,n}(\theta, x_i) \widehat{S}_{m,n}(\theta, x_i)^T \tag{7c}$$

where

$$\widehat{S}_{m,n}(\theta, x) = \frac{1}{w_1 + \cdots + w_n} \sum_{j=1}^{n} w_j \bigl(\nabla \log f_\theta(x, y_j) - \nabla \log f_{\theta,m}(y_j) \bigr) \cdot \frac{f_\theta(x, y_j)}{f_{\theta,m}(y_j) h(x)} \tag{7d}$$

See equations (2.7) and (2.9) in the first author's thesis.

3 Method of Batch Means

The "Monte Carlo Likelihood Approximation" document goes on about a "method of batch means" estimator of (6c) In the present context this is almost unchanged. Equations (7a) and (7b) in that document remain the same. The only difference is that equation (6d) in that document is replaced by equation (7d) above.

Putting these together we obtain

$$\begin{aligned}
\widetilde{S}_{m,n,k}(\theta) &= \frac{1}{l} \sum_{i=(k-1)l+1}^{kl} \widehat{S}_{m,n}(\theta, x_i) \\
&= \frac{1}{l} \sum_{i=(k-1)l+1}^{kl} \frac{1}{w_1 + \cdots + w_n} \sum_{j=1}^{n} w_j \\
&\qquad \times \bigl(\nabla \log f_\theta(x_i, y_j) - \nabla \log f_{\theta,m}(y_j) \bigr) \cdot \frac{f_\theta(x_i, y_j)}{f_{\theta,m}(y_j) h(x_i)}
\end{aligned}$$

which the code, for efficiency reasons, reverses the order of summation

$$\begin{aligned}
&= \frac{1}{l(w_1 + \cdots + w_n)} \sum_{j=1}^{n} \sum_{i=(k-1)l+1}^{kl} \\
&\qquad \times w_j \bigl(\nabla \log f_\theta(x_i, y_j) - \nabla \log f_{\theta,m}(y_j) \bigr) \cdot \frac{f_\theta(x_i, y_j)}{f_{\theta,m}(y_j) h(x_i)}
\end{aligned}$$