



Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships

Geof H. Givens; Adrian E. Raftery

Journal of the American Statistical Association, Vol. 91, No. 433 (Mar., 1996), 132-141.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28199603%2991%3A433%3C132%3ALAISFM%3E2.0.CO%3B2-%23>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of the American Statistical Association is published by American Statistical Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Journal of the American Statistical Association
©1996 American Statistical Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships

Geof H. GIVENS and Adrian E. RAFTERY

We consider adaptive importance sampling techniques that use kernel density estimates at each iteration as importance sampling functions. These can provide more nearly constant importance weights and more precise estimates of quantities of interest than the sampling importance resampling algorithm when the initial importance sampling function is diffuse relative to the target. We propose a new method that adapts to the varying local structure of the target. When the target has unusual structure, such as strong nonlinear relationships between variables, this method provides estimates with smaller mean squared error than alternative methods.

KEY WORDS: Bayesian statistics; Density estimation; Integral evaluation; Kernel method; Monte Carlo simulation; Nonparametric method.

1. INTRODUCTION

Consider the problem of finding $I_0 = \int g(\mathbf{x})t(\mathbf{x}) d\mathbf{x}$, where $t(\mathbf{x})$ is a probability density that is intractable. The direct Monte Carlo estimate of I_0 is $\sum_{i=1}^n g(\mathbf{x}_i)/n$, where $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is an iid sample from t . The integral can be rewritten as $I_0 = \int g(\mathbf{x})(t(\mathbf{x})/p(\mathbf{x}))p(\mathbf{x}) d\mathbf{x}$, where $p(\mathbf{x})$ is a tractable probability density that serves as an importance sampling distribution. This suggests the Monte Carlo estimate $\hat{I}_0 = \sum_{i=1}^n g(\mathbf{x}_i)(t(\mathbf{x}_i)/p(\mathbf{x}_i))/n$ as an estimator of I_0 , where $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is an iid sample from p . In many cases of interest, frequently when t is a posterior distribution from a Bayesian analysis, $t(\mathbf{x})$ is known only up to a constant of proportionality. In this case we want to find

$$I = \int g(\mathbf{x})t^*(\mathbf{x}) d\mathbf{x} \Big/ \int t^*(\mathbf{x}) d\mathbf{x}, \quad (1)$$

where $t^*(\mathbf{x}) = ct(\mathbf{x})$ for some positive constant c , and this ratio of integrals can be estimated using a ratio of estimates like \hat{I}_0 :

$$\begin{aligned} \hat{I} &= \frac{\sum_{i=1}^n g(\mathbf{x}_i)(t^*(\mathbf{x}_i)/p(\mathbf{x}_i))}{\sum_{i=1}^n t^*(\mathbf{x}_i)/p(\mathbf{x}_i)} \\ &= \sum_{i=1}^n w(\mathbf{x}_i)g(\mathbf{x}_i), \end{aligned} \quad (2)$$

where $w(\mathbf{x}_i) = (t^*(\mathbf{x}_i)/p(\mathbf{x}_i))/\sum_{i=1}^n (t^*(\mathbf{x}_i)/p(\mathbf{x}_i))$ and $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is an iid sample from p .

In a general sense, the quality of the estimator \hat{I} will often be good when p is close to t , and this condition provides results that are good when there are many choices of g to be investigated, although it may not be optimal for any one choice of g . This article is concerned with ways of obtaining

good densities p , and hence good estimators \hat{I} , when it is not feasible to sample from t itself.

Replacing the direct estimate with \hat{I} recalls the sampling importance resampling (SIR) algorithm of Rubin (1987, 1988) for obtaining an iid sample of size m from a distribution that approximates the target distribution $t(\mathbf{x})$ by sampling with replacement from an initial iid sample of size n from the importance sampling distribution, or *envelope*, $p(\mathbf{x})$ with weights $w(\mathbf{x}_i)$. Thus the integration problem may be recast as a sampling problem. For appropriate envelopes, \hat{I} is consistent as $n \rightarrow \infty$, and we show in Section 3.1 that for any m , the SIR method is exact as $n \rightarrow \infty$.

Adaptive importance sampling (Evans 1988; Kloek and Van Dijk 1978; Oh and Berger 1992; Smith, Skene, Shaw, and Naylor 1978; West 1992, 1993) is an effort to improve the final sample or estimate of I by iterating the importance sampling approach while updating the choice of envelope between stages based on current information about the target. Examples in this article demonstrate that the choice of updating strategy can have an important effect on the quality of the results.

This work was motivated by a problem that induced intractable posterior distributions that had support concentrated on a thin, curved manifold when starting from an extremely diffuse prior over a vast region. This problem was an application of the "Bayesian synthesis" framework (Givens 1993; Raftery, Givens, and Zeh 1995) developed for combining available evidence about the inputs and outputs of complex mechanistic models.

This article demonstrates that a general adaptive importance sampling algorithm may be specialized to better suit certain classes of target densities that have distinctive features. We explore one such specialization that offers some advantages when dealing with distributions with unusual structure, like those that may result from the Bayesian synthesis approach.

2. A GENERAL IMPORTANCE SAMPLING ALGORITHM

The adaptive importance sampling methods that we consider may be thought of as an alternative to simple SIR, where part of the sampling effort is used initially to refine

Geof H. Givens is Assistant Professor of Statistics, Department of Statistics, Colorado State University, Fort Collins, CO 80523. Adrian E. Raftery is Professor of Statistics and Sociology, Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195. This research was supported by the North Slope Borough, Alaska, the State of Alaska (through the Alaska Department of Fish and Game), and the National Oceanic and Atmospheric Administration (through the National Marine Mammal Laboratory and the Alaska Eskimo Whaling Commission). The authors are grateful to Judy Zeh for helpful discussions and to the associate editor and two anonymous referees for useful comments that have substantially improved this article.

the envelope and the remainder is then used to obtain a sample with the improved envelope. The methods proceed by first performing a simple importance sampling step to get an initial approximate sample from the target, then using kernel density estimation to estimate the target. The resulting mixture distribution then serves as an improved envelope for another stage of importance sampling. These steps can be repeated if the current envelope is still a poor one. The two adaptive importance sampling methods described in this article use different methods of kernel density estimation. A global method uses the same kernel for all points in the current sample. A local method uses, for each point, a different kernel chosen to mimic the local covariance structure of the target density in a neighborhood of that point. The local method often provides a more effective envelope for the next stage of importance sampling when the target distribution lies on a complicated, nearly lower-dimensional manifold compared to the initial envelope.

Let the target distribution be $t(\mathbf{x})$, where \mathbf{x} is d dimensional. A general v -stage importance sampling algorithm can be written as follows:

1. Let the initial envelope be $p^{(0)}(\mathbf{x})$ and set the stage index to $j = 0$.
2. Draw a sample of size n_j from $p^{(j)}(\mathbf{x})$. Call this sample $\mathbf{x}_1^{(j)}, \dots, \mathbf{x}_{n_j}^{(j)}$.
3. Form standardized importance weights

$$w_i^{(j)} = [t^*(\mathbf{x}_i^{(j)})/p^{(j)}(\mathbf{x}_i^{(j)})] / \sum_{i=1}^{n_j} (t^*(\mathbf{x}_i^{(j)})/p^{(j)}(\mathbf{x}_i^{(j)})) \quad (3)$$

for $i = 1, \dots, n_j$ if $t(\mathbf{x}) = ct^*(\mathbf{x})$ and c is unknown. If t is known exactly, then use (3) with $c = 1$.

4. If $j = v$, then sample $\mathbf{x}_1, \dots, \mathbf{x}_m$ with replacement from $\mathbf{x}_1^{(j)}, \dots, \mathbf{x}_{n_j}^{(j)}$ with weights $w_1^{(j)}, \dots, w_{n_j}^{(j)}$ and use this sample for inference. Otherwise, go to Step 5.
5. Update the envelope using the kernel density estimator

$$p^{(j+1)}(\mathbf{x}) = \sum_{i=1}^{n_j} w_i^{(j)} K(\mathbf{H}_i^{(j)-1}(\mathbf{x} - \mathbf{x}_i^{(j)})) / |\mathbf{H}_i^{(j)}|, \quad (4)$$

where K is the kernel, $|\cdot|$ is the absolute value of the determinant, and $\{\mathbf{H}_i^{(j)}\}$ ($i = 1, \dots, n_j$) are nonsingular $d \times d$ matrices that may depend on $\mathbf{x}_1^{(j)}, \dots, \mathbf{x}_{n_j}^{(j)}$.

6. Increment j and go to 2.

In this article we present one choice for the set $\{\mathbf{H}_i^{(j)}\}$ where these matrices do vary with i . This means that for each stage j , up to n_j different covariance matrices of size $d \times d$ must be estimated and (4) must be evaluated at n_{j+1} points. The algorithm requires $N = \sum_{j=0}^v n_j$ evaluations of the target. When these evaluations are extremely complicated, N may be the limiting factor in the procedure. Adaptive methods may be most useful when N is limited, because the nonadaptive SIR algorithm will eventually suffice when N is sufficiently large, regardless of how diffuse the envelope is. When N is limited, we may ask for a fixed N whether it is preferable to devote the entire sample to a single SIR run or to devote a portion of the sample to adaptive refinement of the envelope. Examples in this article demonstrate that adaptive envelope refinement is sometimes preferred.

2.1 The SIR Algorithm

The SIR algorithm corresponds to $v = 0$. The effectiveness of the SIR algorithm can be strongly influenced by the quality of the envelope (Geweke 1989; Hesterberg 1991; Oh and Berger 1991; Rubin 1987, 1988; Smith et al. 1987; Smith and Gelfand 1992; Stewart 1983).

Extreme dispersion of the envelope relative to the target is exacerbated by the dimensionality of the probability space when d is not small. In this case the SIR algorithm can be very inefficient, because most draws from the envelope will occur in regions where the target distribution has low probability. The resulting importance resample will then consist of only a few unique points repeated many times. Such a sample will often yield poor estimates of variability and probabilities and may not even give good estimates of location.

2.2 Global Adaptive Importance Sampling

For difficult target densities, it would be helpful to use a preliminary sample to identify the possible region of support of the target before making the final importance sample. Smith et al. (1987) and West (1992, 1993) have suggested using an iterative procedure where at each stage a new envelope is estimated using a nonparametric kernel density estimator with the sample from the previous stage.

Ideally, the sample at each stage would first be transformed in some way to improve the performance of the density estimator, typically by making the distribution of the transformed variables more nearly multivariate normal (Smith et al. 1987). The transformations are problem specific and rely on the investigator's knowledge of the target, but often they do not remove its troublesome features. The sample at each stage can be further transformed by covariance scaling (Smith et al. 1987; West 1992). The scaling is implicit when $\mathbf{H}_i^{(j)T} \mathbf{H}_i^{(j)} = h^{(j)} \hat{\Sigma}^{(j)}$ for $i = 1, \dots, n_j$, where $\hat{\Sigma}^{(j)}$ is the estimated covariance matrix of the j th sample and $h^{(j)}$ is some scale parameter.

Indeed, with $j > 1$ and $\mathbf{H}_i^{(j)} = \mathbf{H}^{(j)}$ ($i = 1, \dots, n_j$ for each j), the foregoing general algorithm includes methods similar to the adaptive importance sampling method of West (1992). West uses a multivariate t distribution for K and sets $\mathbf{H}^{(j)T} \mathbf{H}^{(j)} = h(n_j) \hat{\Sigma}^{(j)}$, where $\hat{\Sigma}^{(j)}$ is the Monte Carlo covariance estimate from the weighted sample and $h(n_j)$ is some scale parameter. West suggests that the scale parameter be a slowly decreasing function of sample size. To compensate for the natural overdispersion of a kernel estimate, West shrinks the kernel locations toward the sample mean.

We call methods where $\mathbf{H}_i^{(j)} = \mathbf{H}^{(j)}$, such as West's, *global adaptive importance sampling* (GAIS), because the same $\mathbf{H}^{(j)}$ is used for each kernel contribution at each sample point, and this $\mathbf{H}^{(j)}$ is based on the estimated global covariance structure of the target.

Typically, the performance of a kernel density estimator is relatively insensitive to the choice of the kernel's functional form (Silverman 1986). However, the choice of the matrices $\mathbf{H}_i^{(j)}$ can be important with the adaptive importance sampling method, because small differences in the

tails of the density estimate can have dramatic effects on its quality as an importance sampling envelope. In other words, for the creation of importance sampling envelopes, it is wise to avoid density estimators that can be light in the tails, even if such estimates are more precise overall.

If, after transformation, the target distribution has highly nonconvex contours or there are strong nonlinear relationships between variables, then the global kernel density estimate may be a poor one to use as an envelope. For example, consider the target distribution whose 95% minimum area probability region is shown in Figure 1. The contours of a kernel based on the global covariance estimate would resemble G and G^* . This kernel would not attribute mass ideally for sample points like z . A better envelope could be constructed with an estimator that more readily recognizes the obvious structure of this density and contributes mass with kernels that have contours like L and L^* .

2.3 Local Adaptive Importance Sampling

A density estimator that adapts to the local covariance structure of the sample may be better able to provide an envelope that covers the support region of an unusual target. Let $\mathbf{H}_i^{(j)T} \mathbf{H}_i^{(j)} = h_i^{(j)} \hat{\Sigma}_i^{(j)}(\lambda_i^{(j)})$, where $\hat{\Sigma}_i^{(j)}(\lambda_i^{(j)})$ is a covariance estimate for the $\lceil \lambda_i^{(j)} n_j \rceil$ nearest neighbors of the i th point in the j th sample ($2/n_j \leq \lambda_i^{(j)} \leq 1$) and $h_i^{(j)}$ are local scale factors. We measure neighborhoods by Mahalanobis (1936) distance. Each point is included among its own neighbors. We call this method *local adaptive importance sampling* (LAIS). GAIS is a special case of this method obtained when $\lambda_i^{(j)} = 1$ for all i and j and $h_i^{(j)}$ does not depend on i .

LAIS uses a *local kernel density estimator*, where K remains the same but $\mathbf{H}_i^{(j)}$ varies with each point (i) in the sample. Givens (1995) discussed the \mathcal{L}_1 consistency of the local kernel density estimator.

Consider the target density shown in Figure 1. The contour shape for two choices of kernel are shown centered at the point z near one tail of the target. G represents a GAIS kernel, which reflects the global covariance of the j th sample. L , representing a LAIS kernel, contributes mass near the point in a manner that more closely reflects the local structure of the target density. Indeed, if z were one of few points sampled in that region of the target, then the envelope based on the global kernel estimate would have a much lighter tail than the target in the direction where the target's tail has most mass. If point q was drawn at the next sampling stage, then GAIS would give q a very large importance weight. The single point q would play an exaggerated role

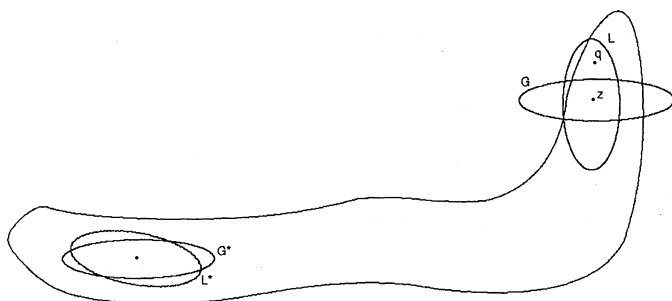


Figure 1. LAIS (L and L^*) and GAIS (G and G^*) Contribute Mass Differently.

in representing this region of the target. We would prefer a more diverse sample of points in this region. If points like q were not drawn by GAIS, then an important region of the target would not be discovered.

3. EVALUATION OF IMPORTANCE SAMPLING STRATEGIES

3.1 Estimation and Simulation Convergence

Consider the case when interest is limited to the expected value of a single function g . The estimator \hat{I} converges almost surely as $n \rightarrow \infty$ to $I = \mathbb{E}g(\mathbf{x})$ if this expectation is finite and the support of $p(\mathbf{x})$ includes that of $t(\mathbf{x})$. Hesterberg (1991) showed that \hat{I} is also asymptotically normal in this case, and minimization of the asymptotic variance leads to identification of $p'(\mathbf{x}) \propto |g(\mathbf{x}) - \mathbb{E}g(\mathbf{x})|t(\mathbf{x})$ as an optimal envelope. However, $p'(\mathbf{x})$ cannot be used in practice because $t(\mathbf{x})$ and $\mathbb{E}g(\mathbf{x})$ are unknown, and often investigation of many choices of g is of interest. Zhang (1993) has examined simulation convergence of one nonparametric adaptive importance sampling method.

The adaptive methods are not meant for lengthy iteration (large v) but instead are intended to obtain quick, drastic envelope refinement in one or two iterations. The adequacy of the final sample is assured through the final importance sampling step. In this sense, then, adaptive importance sampling is strategically different from alternative Monte Carlo integration procedures such as Markov chain Monte Carlo methods (Hastings 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953). In fact, adaptive importance sampling can be an attractive alternative to Markov chain Monte Carlo methods, because these latter methods do not provide an independent sample and may experience convergence difficulties when the number of iterations is not sufficiently large.

To compare SIR, GAIS, and LAIS, it is also interesting to consider simulation consistency, by which we mean whether these methods produce final samples from distributions that converge to t as $v \rightarrow \infty$ and possibly $n_j \rightarrow \infty$.

Consider first the SIR algorithm, where we sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ iid from p , then calculate the importance weights $w_i = t(\mathbf{Y}_i)/(p(\mathbf{Y}_i) \sum_i t(\mathbf{Y}_i)p(\mathbf{Y}_i)^{-1})$, and finally resample $\mathbf{X}_1, \dots, \mathbf{X}_m$ with replacement from $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ with probabilities w_1, \dots, w_n . Let t and p be multivariate probability densities on \mathbb{R}^k , and let F be the distribution function on \mathbb{R}^k corresponding to the target t . If $\{\mathbf{x}: p(\mathbf{x}) \neq 0\} \supseteq \{\mathbf{x}: t(\mathbf{x}) \neq 0\}$, then \mathbf{X}_i converges in distribution to F as $n \rightarrow \infty$. Using intuitive arguments, Rubin (1988) and Gelman and Rubin (1992) have noted this convergence; a direct proof is straightforward. Fix m . Suppose that the first step of the SIR algorithm produces a weighted sample, $\mathbf{Y} = \{\mathbf{Y}_i\} = \{(Y_{i1}, \dots, Y_{ik})\}$ ($i = 1, \dots, n$), with weights w_1, \dots, w_n . Let $F_n(\mathbf{x}|\mathbf{y})$ be the empirical distribution function on \mathbb{R}^k corresponding to the weighted sample \mathbf{Y} , namely

$$F_n(\mathbf{x}|\mathbf{y}) = \sum_{i=1}^n w_i 1(x_1 \leq y_{i1}, \dots, x_k \leq y_{ik}).$$

Conditional on \mathbf{Y} , each \mathbf{X}_i is a random variable with discrete distribution $F_n(\mathbf{x}|\mathbf{y})$. Let h be any bounded,

continuous function on \mathcal{R}^k . Then $\mathbb{E}_F h(\mathbf{x})$ exists, and $\sum_{i=1}^n w_i h(\mathbf{Y}_i) \rightarrow \mathbb{E}_F h(\mathbf{x})$ according to the strong law of large numbers, as noted by Geweke (1989). Because $\int h(\mathbf{x}) dF_n(\mathbf{x}|\mathbf{y}) \rightarrow \int h(\mathbf{x}) dF(\mathbf{x})$ for any bounded, continuous h , the conditional distribution of any $\mathbf{X}_i|\mathbf{Y}, F_n(\mathbf{x}|\mathbf{y})$ converges to that of a random variable \mathbf{X} with target distribution F (Billingsley 1986). Because this limiting distribution does not depend on \mathbf{Y} , the unconditional limiting distribution of \mathbf{X}_i is also F .

Next we consider simulation convergence of the adaptive methods. Although one might compare $p^{(j)}$ to p' or t as the number of iterations increased, both GAIS and LAIS end with a final importance sampling step. Thus one may always obtain simulation consistency with these methods by letting $n_v \rightarrow \infty$, as long as the support of $p^{(v)}(\mathbf{x})$ includes that of $t(\mathbf{x})$, regardless of the convergence properties of $p^{(j)}(\mathbf{x})$. Because $p^{(v)}(\mathbf{x})$ is a kernel density estimate, simulation consistency can be achieved by using unbounded kernels.

3.2 Monte Carlo Evaluation

Here we consider how to examine the results of Monte Carlo comparisons between methods. For adaptive methods, evaluation of how well a density $p^{(j)}(\mathbf{x})$ serves as an importance sampling envelope for a target $t(\mathbf{x})$ should occur at each stage of the procedure. In particular, envelope monitoring is necessary to identify when the current envelope is not sufficiently diffuse. However, poor interstage monitoring strategies will often manifest themselves in the final results because of feedback between the sampling and envelope estimation steps. A poor envelope produces a poor sample, which in turn may produce another poor envelope. Thus here we focus on evaluating the results given at the final stage of a method.

We assess the quality of an importance sampling strategy in two ways: by the bias and variance of estimates produced with it and by the diversity of the sample that the final envelope provides. The first assessment is made by comparing the mean squared errors (MSE's) of Monte Carlo estimates of selected quantities of interest. Sample diversity is related to the variability of the weights and is discussed in the remainder of this section.

Usually there are many quantities of interest in an analysis, including means, modes, standard deviations, and probability intervals. Although each quantity of interest would specify a different optimal sampling envelope, adequate results could be obtained for all with a single envelope. The resulting sample should allow efficient estimation of all quantities of interest and also allow the researcher to later examine new quantities that may be relevant. Some quantities of interest may be marginal quantiles, which cannot be expressed directly as the expected value of some function $g(\mathbf{x})$. A quantile estimate would follow easily from a sample from the target density.

Although envelopes that differ from the target may actually be better for estimation of a single quantity of interest, the importance sampling envelope that results in a resample exactly from the target is the target itself. Consider the unstandardized weight function $\tilde{w}(\mathbf{x}) = t(\mathbf{x})/p(\mathbf{x})$ with cdf $F_p(\tilde{w})$, and assume for the remainder of this section that the

support of $p(\mathbf{x})$ includes that of $t(\mathbf{x})$. Using t itself as the envelope produces constant weights, so $F_t(\tilde{w}) = 1_{\{\tilde{w} \geq 1\}}$, where $1_{\{a \geq b\}}$ is 1 if $a \geq b$ and zero otherwise. Thus $p(\mathbf{x})$ can be evaluated by assessing how much F_p differs from F_t . Define

$$D = \int_0^1 (F_p^{-1}(u) - 1)^2 du \quad (5)$$

as a measure of the distance from F_p to F_t . Note that $D = \text{var}(\tilde{w})$, which was recommended as a performance measure by Hammersley and Handscomb (1964), and $D + 1 = \mathbb{E}_t \tilde{w}$, which was suggested by Geweke (1989). Note also that D has the desirable property that it is very sensitive to the case where a few small regions of the envelope produce extremely large weights. D is not equivalent to integrated squared error (ISE), mean ISE (MISE), or the L_1 norm, common measures of distance between densities.

The estimator $\hat{D} = n^{-1} \sum_{i=1}^n (\tilde{w}(\mathbf{x}_i) - 1)^2$ is consistent for D when $\mathbf{x}_1, \dots, \mathbf{x}_n$ form an iid sample from $p(\mathbf{x})$ and c is known. Similarly, a consistent estimator for D when c is unknown is $\hat{D} = n \sum_{i=1}^n (w(\mathbf{x}_i) - n^{-1})^2$. Comparison of competing envelopes can be made via the ratio of their corresponding values of \hat{D} .

The constancy of the importance weights can also be gauged by the number of unique points in the final sample. We standardize this measure by using $U = Q/(N(1 - \exp(-m/N)))$, where Q is the number of unique points in the final sample and the denominator is an approximation to $\mathbb{E}Q$ in the desirable situation when the weights are constant and m is the size of the final sample from the complete collection of $N = \sum_j n_j$ candidates ever considered (Lewontin and Prout 1956). Standardization by $\mathbb{E}Q$ penalizes the adaptive methods for sampling from only the final n_v candidates but does not penalize SIR, which samples from the full collection of N candidates.

Samples that arise from weights with small \hat{D} and that result in large U are desirable.

4. MONTE CARLO EXAMPLES

We compare Monte Carlo results from simulated examples in two and three dimensions. We also briefly examine results from a very difficult ten-dimensional application motivated by the whale modeling work of Raftery et al. (1995).

4.1 Simulated Examples

4.1.1 Algorithm Specifications. Multivariate Gaussian kernels were used for all density estimation. The total number of target evaluations was $N = \sum_{j=0}^v n_j$. Scale parameters were chosen relative to Terrell's (1990) maximal smoothing span, denoted by $T(n_j, d)$ for a d -dimensional sample of size n_j , to help insure that constructed envelopes erred toward diffusion. For GAIS, we set $h^{(j)} = s^2 T(n_j, d)$, where $s > 0$ is a parameter to be chosen. For LAIS, we set $\lambda_i^{(j)} = \lambda$ for all i and j , and $h_i^{(j)} = s^2 T(n_j, d)/\lambda$ for all i . This choice for $h_i^{(j)}$ balances increasing neighborhood size, λ , against the increased dispersion of a larger neighborhood, as measured by $\hat{\Sigma}_i^{(j)}$. Sample covariance estimates provided $\hat{\Sigma}^{(j)}$ and $\hat{\Sigma}_i^{(j)}$. If, for a given point, the prespecified λ resulted in a noninvertible neighborhood covariance matrix,

then λ was gradually increased to alleviate the problem in that instance and then reset for the next point. We did not use West's (1992) method of shrinking resampled points toward the mean. When the target distribution has nonconvex contours, such shrinkage can cause severe misrepresentation of the target.

4.1.2 Monte Carlo Study Design and Display of Results.

For the examples that we examined, we present results for two levels of N and three methods: SIR, GAIS, and LAIS with $\lambda = .5$. Additional runs indicated that $\lambda = .5$ typified LAIS performance except for very small values ($\lambda \leq .2$), for which LAIS performance was degraded due to under-smoothing. As λ increased toward 1, the LAIS and GAIS strategies grew more similar. The choices of v, n_j, s, m , and N differ between examples. The targets that we examined are each parameterized by a dispersion parameter, a , for which we give results for values corresponding to strong and weak target structure.

The performance measures in each example are U, \hat{D} , and the MSE for estimation of potential quantities of interest.

We ran ten replications with each (N, λ) pairing and ten corresponding replications of a SIR run with each choice of N . We standardized the MSE and \hat{D} results by dividing by the mean result for the SIR runs with the same N . In the tables describing the MSE for estimating quantities of interest, values less than 1 are favorable because they indicate that a method estimated the quantity of interest with less MSE than the corresponding SIR runs. Smaller values are also favorable for \hat{D} . The results for U are not standardized relative to SIR, so larger values of U (near to 1) are favored, because they indicate more unique points in the final sample.

4.2 Example 1: Two-Dimensional Structure

4.2.1 Description. The first test case was the target density $f_a(x, y)$ such that X had a marginal uniform distribution on $[-1, 4]$ and $(Y|X) \sim N(|X|, .09a^2)$. Figure 2 shows samples drawn from this density when $a = .75$ and $a = 3.5$. For small a , f_a is exactly the type of target for which LAIS is intended, because it shows a strong non-linear relationship between x and y (see Fig. 2). In higher

dimensions, this type of density is seen in the whale modeling work of Raftery et al. (1995) examined in Section 4.4. Larger values of a produce densities with less distinct structure. The initial envelope was uniform over $[-4, 7] \times [-4, 8]$.

We fixed $m = 500$ and tabled results for $N = 1,250$ and $N = 5,000$. The number and size of stages were (400, 850) for $N = 1,250$ and (1,500, 1,000, 1,000, 1,500) for $N = 5,000$. Further investigation of the choice of the number of iterations can be found in Section 4.2.3. We used $s = .7$.

In addition to U and \hat{D} , we also examine MSE for estimates of $\mathbb{E}Y = \mu_y = 1.7$ and $p_1 = \Pr[X \leq 0] = .2$. Successful runs should produce precise unbiased estimates of p_1 and μ_y . When the envelopes resemble f_a , \hat{D} will be small and U will be near 1.

4.2.2 Results. Table 1 shows that the adaptive methods were more successful than SIR at providing constant importance weights and a diverse sample. For the less dispersed case ($a = .75$), LAIS avoided extreme importance weights and provided up to three times more unique points in the final resample than did SIR. As the target became

more dispersed, the differences between methods became less extreme.

Table 2 shows that the best method for estimation of μ_y and p_1 depended on the degree of structure in the target. When the target was diffuse ($a = 3.5$), the additional variability of the adaptive methods made SIR appear to be the best method. However, with strong target structure ($a = .75$), LAIS appeared to be an attractive alternative that generally provided estimates with lower MSE than did SIR or GAIS. GAIS performed poorly in this case, because it sometimes generated extreme importance weights.

4.2.3 Number of Stages. To investigate the importance of number of stages, v , and the total number of target evaluations, N , we ran additional trials with the target in this example with $a = .75$.

We let v vary from 0 to 5 with $n_j = 1,000$ for each j and $m = 1,000$. Thus N varied from 1,000 to 6,000. We ran

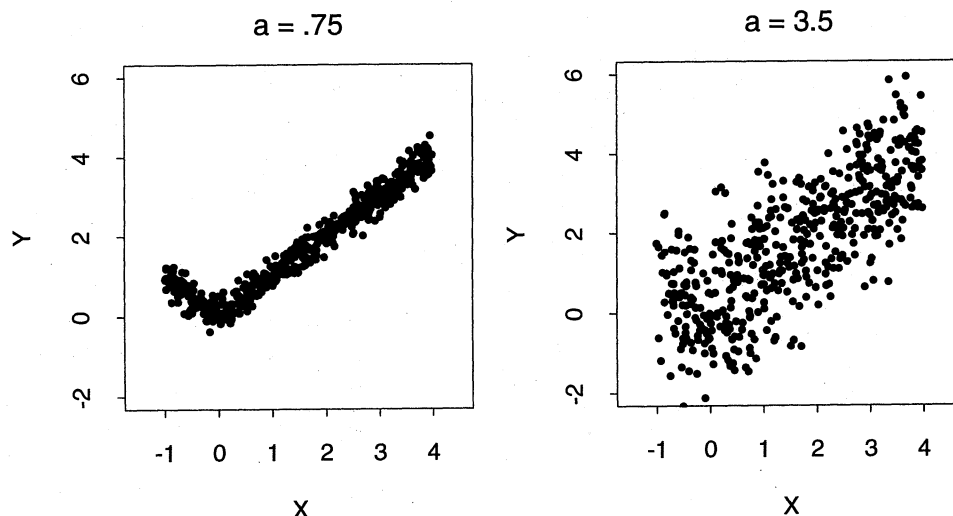


Figure 2. Samples of Size 500 From f_a in Example 1.

Table 1. Sample Diversity Performance Measures for Example 1

Method	N	Standardized \hat{D}		U	
		a = .75	a = 3.5	a = .75	a = 3.5
SIR	1,250	1.00	1.00	.24	.46
GAIS	1,250	5.94	.12	.56	.85
LAIS	1,250	.22	.16	.77	.85
SIR	5,000	1.00	1.00	.54	.76
GAIS	5,000	.09	.20	.87	.88
LAIS	5,000	.02	.04	.87	.87

30 replications of SIR, GAIS, and LAIS. The resulting numbers of unique points and $MSE(\hat{\mu}_y)$ are shown in Figure 3. LAIS adapts quickly; only one stage was necessary to provide any achievable improvement. LAIS performance was consistently better than SIR, providing smaller MSE's and more constant importance weights with smaller N . SIR performance will improve with increasing N . A SIR run with $N = \sum_j n_j$ draws will eventually outperform the adaptive methods if their stage sampling sizes (n_j) remain fixed while the number of stages increases. About $N = 27,000$ points would be necessary for SIR to provide the same number of unique points achieved by LAIS with $N = 2,000$.

On average, GAIS also provided more unique points than SIR but occasionally generated extreme importance weights and hence small numbers of unique points. In these instances estimation performance was degraded, as is evident in the panel of Figure 3 for $MSE(\hat{\mu}_y)$, where GAIS performance fluctuates drastically. This example was deliberately chosen with strong local target structure and diffuse initial envelope to emphasize the differences between methods. These results indicate that N is probably less closely related to performance than is the single one or two largest values of n_j .

One attractive feature of SIR is that it is noniterative. In this example the adaptive methods also do not require much iteration to derive a suitably refined envelope.

4.2.4 Computation Time. For this example, with $N = 1,250$, the computation times for SIR and LAIS were .11 and 1.97 seconds on a Sun SPARC 20 m61. GAIS speeds are between these, depending on coding.

SIR may be favored when target evaluations are computationally inexpensive, as they are in this simple example. However, the MSE of a SIR estimate is reduced by increasing the number of target evaluations, and when these evaluations are computationally expensive, switching to an adaptive method such as LAIS may provide a net decrease

Table 2. Standardized MSE Performance Measures for Example 1

Method	N	Std. $MSE(\hat{\mu}_y)$		Std. $MSE(\hat{p}_1)$	
		a = .75	a = 3.5	a = .75	a = 3.5
SIR	1,250	1.00	1.00	1.00	1.00
GAIS	1,250	50.16	.29	3.36	.90
LAIS	1,250	1.34	1.01	.52	1.71
SIR	5,000	1.00	1.00	1.00	1.00
GAIS	5,000	.17	1.50	1.04	2.81
LAIS	5,000	.29	.77	.38	2.36

in computation time because the increased computation required by the importance sampling method can be offset by a large reduction in the total number of sample points needed to attain a desired MSE.

4.3 Example 2: Trivariate Normal Mixture

4.3.1 Description. The second example was a mixture of two trivariate normal distributions with distribution function $G_a = \frac{1}{3}F_1 + \frac{2}{3}F_2$. F_1 was a trivariate normal distribution function with mean vector $(a, 1, a)$ and covariance matrix with diagonal $(a^2, 1/4, 1/4)$ and zeros elsewhere. F_2 had mean vector $(0, 0, 0)$ and covariance matrix with diagonal $(1/4, 1, a^2)$ and zeros elsewhere. Let g_a be the density corresponding to G_a . Large values of a give highly structured densities; small values give diffuse densities. We table results for $a = 1$ and $a = 6$. A pairwise scatterplot of a sample of 500 from g_a is shown in Figure 4 for $a = 6$. The strong structure in these data is reminiscent of the Stanford Linear Accelerator data described by Friedman and Tukey (1974) which was used to demonstrate the Prim-9 graphics system (Stanford Linear Accelerator Center 1973). The distribution shown in Figure 4 challenges the simple SIR method and GAIS, because these methods are unable to adapt to the local structure of the density. When $a = 1$, a scatterplot of a sample from g_a does not show much structure at all. The initial envelope was a trivariate normal distribution with the same mean vector and covariance matrix as G_a .

We tabled results for $N = 2,000$ and $N = 9,000$. The interim sample sizes were split evenly between stages, m was fixed at $m = 800$, and s was 1. We used two stages for the low value of N and three stages for the high value of N .

We consider estimation of three quantities of interest, in addition to the measures \hat{D} and U . These quantities are $\mathbb{E}Z = \mu_z = a/3$, $p_2 = \Pr[X > a \text{ and } Y > 1] = .0857$, and $\omega = \text{cor}(X, Z) = .265$, and $.315$ for $a = 1$ and $a = 6$.

4.3.2 Results. Table 3 shows the results for the \hat{D} and U performance measures. For the simple target ($a = 1$), the SIR method produced more unique points than any adaptive method, suggesting that the envelope resembled the target sufficiently for reliable estimation. As the target structure, a , increased (and hence the envelope became a poorer representation of the target), the SIR method did not provide as many unique points as the adaptive methods in the final sample, though its value of \hat{D} improved relative to the other methods. Generally, LAIS performed better than GAIS. All three methods produced more nearly uniform weights as N increased.

Table 4 shows the MSE results for this example. For the simple target ($a = 1$), no method was superior for estimation of all quantities of interest. In many cases, SIR worked about as well as the adaptive methods. However, for the very structured target ($a = 6$), the LAIS method generally had lower MSE's than SIR, and the GAIS method had greater MSE's than SIR. The high MSE's for the GAIS method were largely caused by excess variability due to extreme importance weights generated when GAIS failed to adequately approximate the structure of the tails of the target. For example, for estimation of $\mu_z(6) = 2$, when

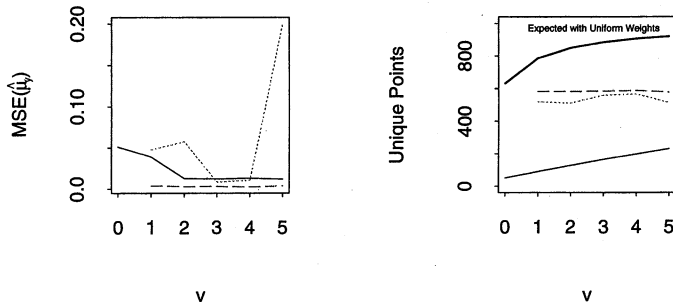


Figure 3. $MSE(\hat{\mu}_y)$ and Numbers of Unique Points for Each Method (Solid Line, SIR; Dotted Line, GAIS; Dashed Line, LAIS) With Increasing Number of Stages, v , for Example 1 with $a = .75$. The total number of target evaluations is $N = 1,000(v + 1)$.

$N = 9,000$, the sample means (and standard deviations) of the 10 replicates were 2.01 (.28) for SIR, 1.92 (.44) for GAIS, and 1.94 (.18) for LAIS.

Generally, the results in Table 4 confirm that GAIS performance is degraded when there is a strong nonlinear relationship between the variables of the target density. LAIS can adapt to this structure, providing estimates with lower MSE than those provided by SIR or GAIS.

We consider estimation of p_2 in further detail. Figure 5 summarizes the $MSE(\hat{p}_2)$ results for this example. In this figure we include additional results for $s = .7, \lambda = .2, N = 4,500$, and $a = 3$. The area of each circle is proportional

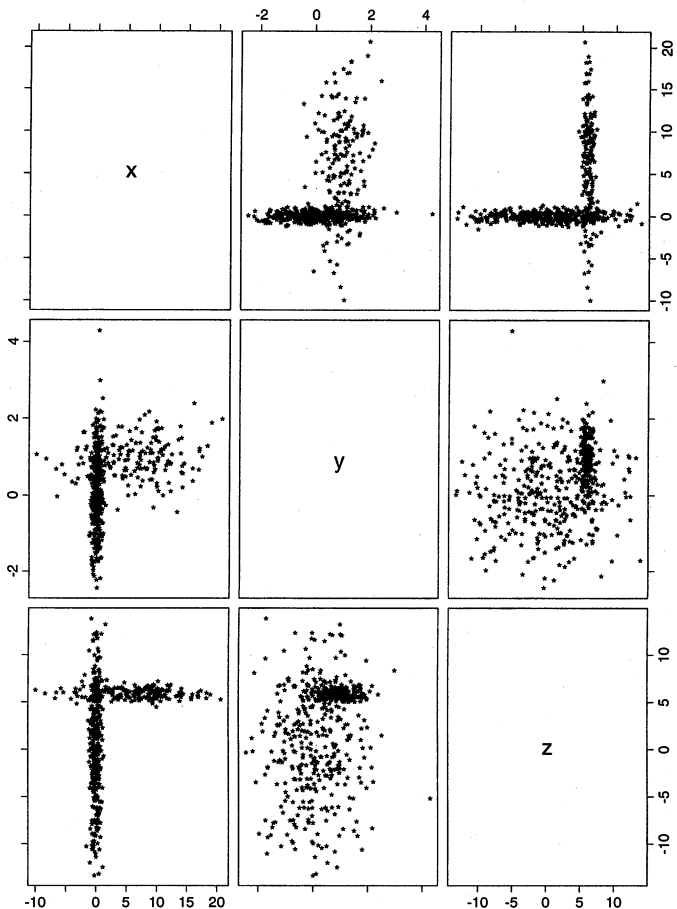


Figure 4. Sample of Size 500 From g_a When $a = 6$ for Example 2.

Table 3. Sample Diversity Performance Measures for Example 2

Method	N	Standardized \hat{D}		U	
		a = 1	a = 6	a = 1	a = 6
SIR	2,000	1.00	1.00	.90	.49
GAIS	2,000	.88	1.04	.77	.44
LAIS	2,000	.27	1.39	.79	.63
SIR	9,000	1.00	1.00	.96	.76
GAIS	9,000	.13	1.02	.91	.68
LAIS	9,000	.12	.96	.90	.83

to the ratio of MSE for estimating p_2 for each method relative to the corresponding SIR runs. When the ratio is less than 1, the circle is shaded, indicating that the method yielded smaller MSE than did SIR. The dominant feature of this figure is the poor performance of GAIS relative to SIR and LAIS when target structure is high ($a = 6$). This occurs because GAIS constructs envelopes which do not adequately cover the tails of highly structured targets. We also see that for low target structure and large samples, the gain achieved by adaptive envelope refinement is more than offset by the added variability relative to SIR. For two trials with the smallest sample size, N , and the smallest span, s , LAIS exhibits variability due to undersmoothing targets of moderate or high structure.

4.4 Example 3: Bayesian Analysis of Whale Population Model

Examples 1 and 2 are simple in the sense that they are easy to implement and the targets are low dimensional and easy to visualize. However because of their structure, they are challenging tests of method performance.

In this section we consider a complex, ten-dimensional target distribution which is the result of the Bayesian analysis of a population dynamics model for bowhead whales (Givens 1993; Raftery et al. 1995). The model projects an age- and sex-stratified whale population through time. The target distribution is the joint posterior distribution for the model's ten parameters. Like the earlier examples, the target is highly structured in some dimensions, and nearly multivariate Gaussian in others. Figure 6 shows two of the pairwise relationships between model parameters.

The larger number of dimensions makes this example a difficult one, especially because the envelope (prior distribution) is diffuse due to scientists' limited knowledge of

Table 4. Standardized MSE Performance Measures for Example 2

Method	N	Std. $MSE(\hat{\mu}_z)$		Std. $MSE(\hat{p}_2)$		Std. $MSE(\hat{\omega})$	
		a = 1	a = 6	a = 1	a = 6	a = 1	a = 6
SIR	2,000	1.00	1.00	1.00	1.00	1.00	1.00
GAIS	2,000	.62	2.29	1.86	.49	.15	1.83
LAIS	2,000	.63	.85	.93	.17	1.26	.41
SIR	9,000	1.00	1.00	1.00	1.00	1.00	1.00
GAIS	9,000	.73	2.63	1.62	.28	.77	1.91
LAIS	9,000	.85	.47	1.10	.14	.37	.28

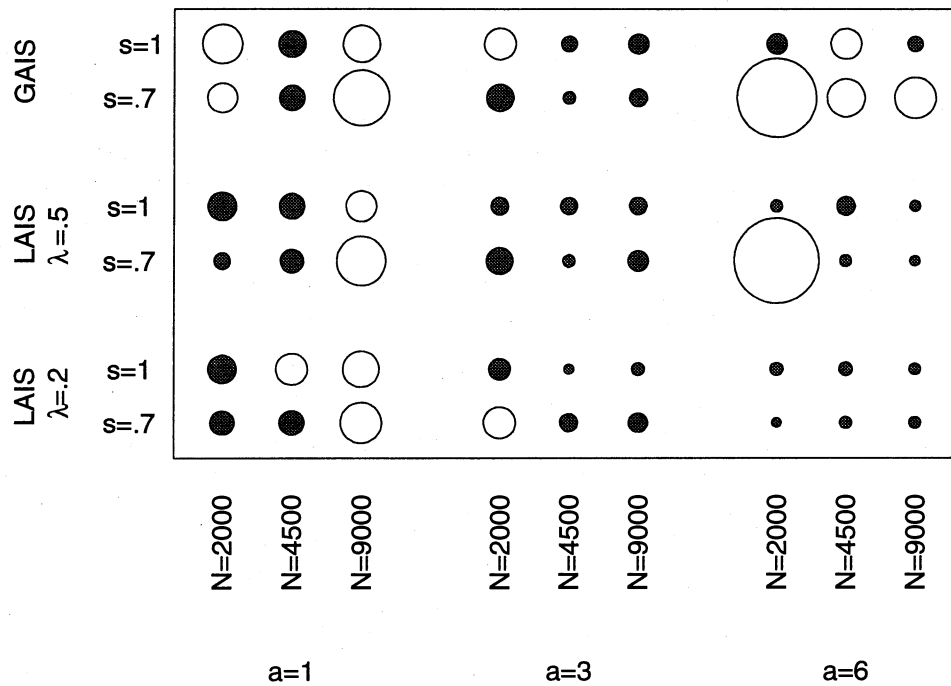


Figure 5. Graphical Summary of $MSE(\hat{p}_2)$ Results for Example 2. Standardized $MSE(\hat{p}_2)$ is proportional to the area of the circle. Shaded circles correspond to better performance than SIR.

the bowhead species. In the first example, there is, roughly speaking, a loss of one dimension from the envelope to the target; but in the bowhead problem, the target lies near a manifold that is many dimensions simpler. This means that it is extremely difficult to hit high probability regions of the target with a few draws from the envelope, and hence it is difficult for the adaptive methods to obtain good estimates of covariance structure.

We ran two trials of the LAIS, GAIS, and SIR methods for this example. With SIR, the total number of model simulations was $N = 20,000$, from which we resampled $m = 1,500$ for the final data. With LAIS and GAIS, $N = 20,000$ total model simulations were divided by taking an initial sample of 12,000 and an intermediate sample of 8,000. The final resamples were also $m = 1,500$ points. Neighborhoods consisted of 3,500 points for LAIS. We im-

plemented a subsampling strategy for LAIS to reduce the number of components in the mixture distribution $p^{(j)}(\mathbf{x})$ by 60%. Because of the large sample sizes used, the additional Monte Carlo variability introduced in LAIS estimation by this approximation should be small.

SIR provided the most unique points in each run (203 and 220), compared to GAIS (148 and 171) and LAIS (168 and 176). However, the LAIS method was able to provide approximate samples from the target that produced marginal density estimates that were generally as good as those of SIR and better than those of GAIS.

Figure 7 shows the marginal cumulative posterior distribution functions for bowhead whale replacement yield for the best run of each method. Replacement yield is the greatest number of whales that can be removed from the population in a year without causing negative population

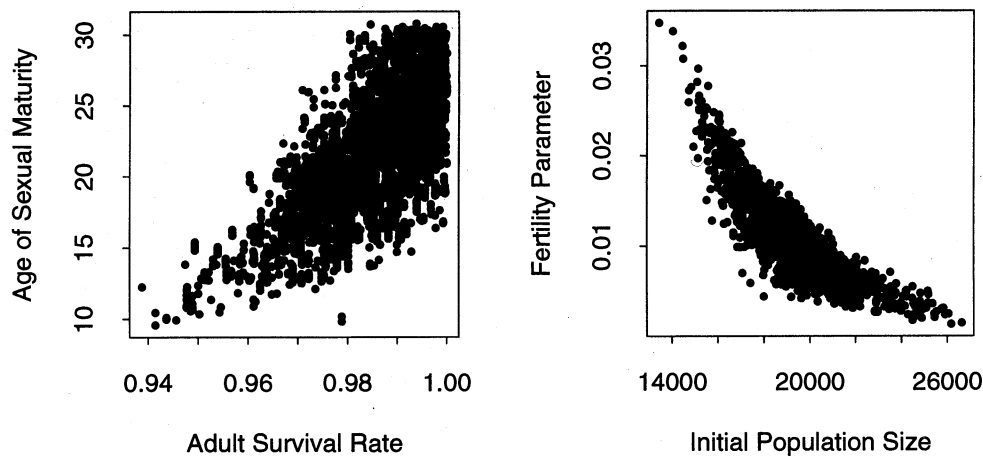


Figure 6. Two Examples of Structure in the Posterior Distribution for Parameters of the Bowhead Whale Population Model for Example 3.

growth; it is an important quantity for setting aboriginal subsistence whaling quotas on the bowhead species. The heavy solid line represents the result from a SIR run that used an N ten times greater than that used for these trials. The true marginal distribution is unknown, but we can take the heavy line as a "gold standard" of performance. In this figure the SIR and LAIS results are qualitatively similar to each other and near the gold standard, but the GAIS distribution is skewed considerably toward higher values of replacement yield. In practice, this GAIS error could lead to overestimation of replacement yield, which in turn could lead to overoptimistic quotas, with resulting endangerment of the species.

Although LAIS apparently avoids the difficulties GAIS experiences in this example, the results were not ideal. LAIS was less successful at providing a good importance sampling envelope along dimensions in the parameter space where there was not a high degree of structure. Further work may identify what features of the population model application cause the most difficulty for the adaptive methods and how the methods can be improved to meet the demands of such a challenging problem.

5. DISCUSSION

5.1 Conclusions

The LAIS method introduced in this article is intended for a general inference problem where a sample from a

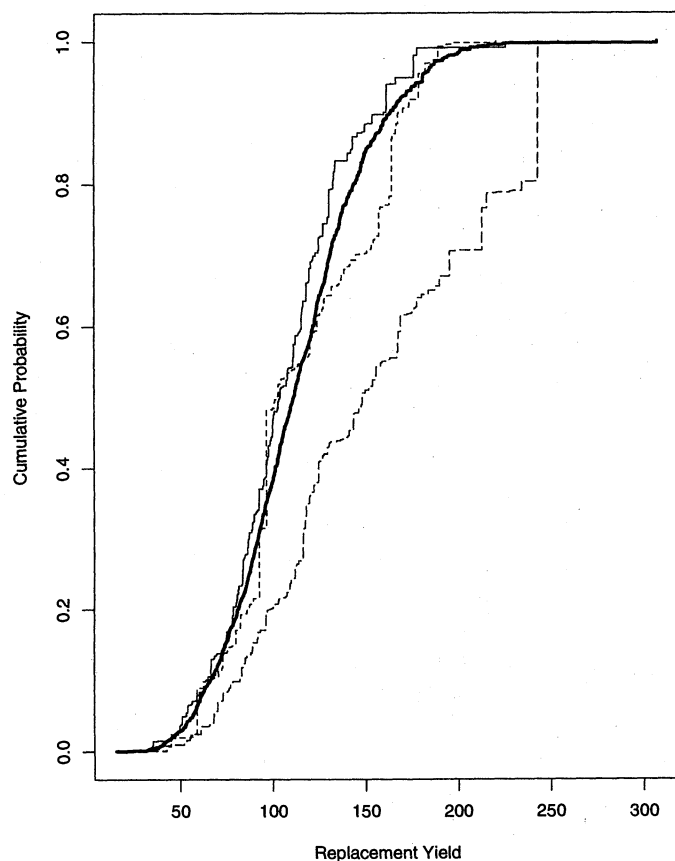


Figure 7. Estimated Marginal Cumulative Distribution of Replacement Yield for Each Method (Solid Line, SIR; Dotted Line, LAIS; Dashed Line, GAIS), and the Gold Standard (Heavy Line).

target distribution is desired for exploring and answering questions about numerous quantities of interest, where the target distribution exhibits a strong nonlinear relationship between variables and where the total number of target evaluations is limited. In this sense LAIS is a specialized tool, but its performance indicates that the general approach of adaptive importance sampling can successfully be tailored to certain classes of problems. Modifying the envelope selection strategy appears to be one promising approach to specialization. For nonparametric envelope selection, such specialization is essentially a density estimation problem in which the types of errors a density estimator is prone to make may be more important than the overall accuracy of the estimator. Other density estimators, such as those surveyed by Izenman (1991) and Silverman (1986) may provide effective envelope strategies for other classes of problems.

For the types of targets examined here, we can draw several conclusions. When target evaluations are quick and the available envelope is sufficiently informative or the target is sufficiently diffuse, the ease and simplicity of SIR are difficult to match. We ran a full set of simulations with a standard bivariate normal target and found that the adaptive methods generally outperformed SIR. However the differences between SIR and the adaptive methods were not as great as with the examples in this article.

If the envelope is vast or the number of draws is limited, and if the target exhibits a strong nonlinear relationship between variables, then an adaptive method such as LAIS may be the best choice. In these cases it appears that devoting a portion of the Monte Carlo sampling effort to envelope refinement reduces MSE more than devoting the entire sampling effort to SIR.

The adaptive methods described in this article estimate envelopes and quantities of interest from only the most recent sample. Oh and Berger (1992) and Zhang (1993) described similar parametric and nonparametric methods that use the cumulative collection of samples at every stage. Implementing this cumulative strategy with LAIS could further enhance its performance.

The reliance on kernel density estimates for envelopes introduces several potential weaknesses into the methodology. The resulting mixture distributions may have thousands of components, though far fewer contribute significant probability. This complexity slows computation. The methods of West (1992) to "collapse" complicated mixtures to simpler mixtures are an important computational aid. Higher-dimensional targets also present a problem. Kernel density estimation is not entirely effective in higher dimensions (Silverman 1986). Finally, we have found very similar results with normal and Epanechnikov (1969) kernels, but other families such as t and split-normal distributions may be useful as well.

5.2 Related Research

Kloek and Van Dijk (1978), Evans (1988), and Oh and Berger (1992) have explored parametric adaptive importance sampling methods. The envelope is assumed to be f_{γ} ,

where $\gamma \in \Gamma$ and the best choice of γ is estimated at each step of iteration. This process is iterated until the estimates of γ meet some convergence criterion. At this point, some or all of the samples from previous stages may be used to estimate quantities of interest, or additional importance sampling can be done using the estimated optimal value of γ .

There has been much more extensive study of nonadaptive importance sampling methods for the integral problem. Discussion of importance sampling, variance reduction techniques, efficiency, and applications have been provided by Berger (1985), Davis and Rabinowitz (1984), Geweke (1988, 1989), Shao (1988), Stewart (1979; 1983), and Van Dijk and Kloek (1980, 1983).

The griddy Gibbs sampler (Ritter and Tanner 1991) and the adaptive rejection sampling method of Gilks and Wild (1992) are two other Monte Carlo approaches to the problem. These are not based on the SIR algorithm, but they do attempt similar intermediate estimation of the target.

[Received October 1993. Revised April 1995.]

REFERENCES

- Abdous, B., and Theodorescu, R. (1989), "On the L^1 Strong Consistency of a New Class of Kernel Probability Density Estimators," *Statistics and Decisions*, 7, 313–331.
- Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), New York: Springer-Verlag.
- Billingsley, P. (1986), *Probability and Measure* (2nd ed.), New York: John Wiley.
- Davis, P. J., and Rabinowitz, P. (1984), *Methods of Numerical Integration*, Orlando, FL: Academic Press.
- Devroye, L. (1987), *A Course in Density Estimation*, Boston: Birkhauser.
- Devroye, L., and Györfi, L. (1985), *Nonparametric Density Estimation: The L_1 View*, New York: John Wiley.
- Epanechnikov, V. A. (1969), "Nonparametric Estimators of a Multivariate Probability Density," *Theory of Probability and Its Applications*, 14, 153–158.
- Evans, M. (1988), "Monte Carlo Computation of Marginal Posterior Quantiles," technical report, University of Toronto, Dept. of Statistics.
- Friedman, J. H., and Tukey, J. W. (1974), "A Projection Pursuit Algorithm for Exploring Data Analysis," *IEEE Transactions on Computers*, C-23, 881–889.
- Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–511.
- Geweke, J. (1988), "Antithetic Acceleration of Monte Carlo Integration Methods in Bayesian Inference," *Journal of Econometrics*, 38, 73–90.
- (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.
- Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337–348.
- Givens, G. H. (1993), "A Bayesian Framework and Importance Sampling Methods for Synthesizing Multiple Sources of Evidence and Uncertainty Linked by a Complex Mechanistic Model," unpublished Ph.D. dissertation, University of Washington, Dept. of Statistics.
- (1995), "Consistency of the Local Kernel Density Estimator," *Statistics and Probability Letters*, 25, 55–61.
- Hammersley, J. M., and Handscomb, D. C. (1964), *Monte Carlo Methods*, London: Methuen.
- Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.
- Hesterberg, T. (1991), "Importance Sampling for Bayesian Estimation," in *Computing and Graphics in Statistics, IMA Volumes in Mathematics and Its Applications*, eds. A. Buja and P. Tukey, New York: Springer-Verlag, pp. 63–75.
- Izenman, A. J. (1991), "Recent Developments in Nonparametric Density Estimation," *Journal of the American Statistical Association*, 86, 205–224.
- Kloek, T., and Van Dijk, H. K. (1978), "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46, 1–20.
- Lewontin, R. C., and Prout, T. (1956), "Estimation of the Number of Different Classes in a Population," *Biometrics*, 12, 211–223.
- Mahalanobis, P. C. (1936), "On the Generalized Distance in Statistics," *Proceedings of the National Institute of Science, India*, 12, 49–55.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1092.
- Oh, M. S., and Berger, J. O. (1992), "Adaptive Importance Sampling in Monte Carlo Integration," *Journal of Statistical Computation and Simulation*, 41, 143–168.
- Raftery, A. E., Givens, G. H., and Zeh, J. E. (1995), "Inference from a Deterministic Population Dynamics Model for Bowhead Whales" (with discussion), *Journal of the American Statistical Association*, 90, 402–430.
- Ritter, C., and Tanner, M. (1991), "The Griddy Gibbs Sampler," Technical Report 878, University of Rochester, Division of Biostatistics.
- Rubin, D. B. (1987), Comment on "The Calculation of Posterior Distributions by Data Augmentation," by M. A. Tanner and W. H. Wong, *Journal of the American Statistical Association*, 82, 543–546.
- (1988), "Using the SIR Algorithm to Simulate Posterior Distributions," in *Bayesian Statistics 3*, eds. J. M. Bernardo et al., Oxford, U.K.: Clarendon Press, pp. 395–402.
- Shao, J. (1988), "Monte Carlo Approximations in Bayesian Decision Theory," *Journal of the American Statistical Association*, 84, 733–741.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.
- Smith, A. F. M., and Gelfand, A. E. (1992), "Bayesian Statistics Without Tears: A Sampling–Resampling Perspective," *The American Statistician*, 46, 84–88.
- Smith, A. F. M., Skene, A. M., Shaw, J. E. H., and Naylor, J. C. (1987), "Progress With Numerical and Graphical Methods for Practical Bayesian Statistics," *The Statistician*, 36, 75–82.
- Stanford Linear Accelerator Center (1973), "Prim-9" (film), Bin 88 Productions.
- Stewart, L. (1979), "Multiparameter Univariate Bayesian Analysis," *Journal of the American Statistical Association*, 74, 684–693.
- (1983), "Bayesian Analysis Using Monte Carlo Integration—A Powerful Methodology for Handling Some Difficult Problems," *The Statistician*, 32, 195–200.
- Terrell, G. R. (1990), "The Maximal Smoothing Principle in Density Estimation," *Journal of the American Statistical Association*, 85, 470–477.
- Van Dijk, H. J., and Kloek, T. (1980), "Further Experience in Bayesian Analysis Using Monte Carlo Integration," *Journal of Econometrics*, 14, 307–328.
- (1983), "Experiments With Some Alternatives for Simple Importance Sampling in Monte Carlo Integration," in *Bayesian Statistics 2*, eds. J. M. Bernardo et al., Amsterdam: North-Holland, pp. 495–510.
- West, M. (1992), "Modelling with Mixtures," in *Bayesian Statistics 4*, eds. J. M. Bernardo et al., Oxford, U.K.: Oxford University Press, pp. 503–524.
- (1993), "Approximating Posterior Distributions by Mixtures," *Journal of the Royal Statistical Society, Ser. B*, 55, 409–422.
- Zhang, P. (1993), "Nonparametric Importance Sampling," technical report, University of Pennsylvania, Dept. of Statistics.