Statistica Sinica **17**(2007), 895-907

POOL SIZE SELECTION FOR THE SAMPLING/IMPORTANCE RESAMPLING ALGORITHM

Kim-Hung Li

The Chinese University of Hong Kong

Abstract: The sampling/importance resampling algorithm is an approximate noniterative sampling method. The algorithm has been used on many occasions to select an approximate random sample of size m from a target distribution from M input random variates. The selection mechanism is an unequal probability sampling with weights being the importance weights. As the weights are random, sampling without replacement is not always possible and some input variates may have more than one copy in the final sample.Duplication of values in the final output is undesirable as it means dependence among the output variates. In this paper a general and simple determination rule for M is proposed. It keeps the duplication problem at a tolerably low level when a tight resampling method is used. We show that (a) $M = O(m)$ if and only if the importance weight is bounded above, (b) if the importance weight has a moment generating function, the suggested M is of order $O(m \ln(m))$, and (c) M may need to be as large as $O(m^{c/(c-1)})$ if the importance weight has finite c-th moment for a $c > 1$. A procedure is suggested to determine M numerically. The method is tested on the Pareto, Gamma and Beta distributions, and gives satisfactory results.

Key words and phrases: Importance weight, Monte Carlo sampling, resampling method, sample size, tight resampling algorithm.

1. Introduction

The increasing use of the Monte Carlo approach in statistics demands effective sampling algorithms. Difficult sampling problems are usually handled by approximate methods. Markov chain Monte Carlo methods (see for example Gilks, Richardson and Spiegelhalter (1996)), are a family of iterative sampling algorithms commonly used for this purpose.

Rubin (1987, 1988) proposed a noniterative sampling procedure and called it the sampling/importance resampling (SIR) algorithm. The SIR algorithm generates an approximately independent and identically distributed (i.i.d.) sample of size m from the target probability density function (pdf) $f(x)$. It starts by generating M (M is usually larger than or equal to m) random numbers from a pdf $h(x)$ as inputs to the algorithm. The output is a weighted sample of size m drawn from the M inputs, with weights being the importance weights $\omega(x)$. As expected, the output of the SIR algorithm is good if the inputs are good (i.e., $h(x)$ is close to $f(x)$) or if M is large compared to m. An advantage of the SIR algorithm is its simplicity. It can be easily understood and used as a general tool for full Bayesian analysis (Albert (1993)). Because of its simple structure, we know how $h(x)$ and M affect the quality of the outputs (Li (2004)). It is worth mentioning that Skare, Bølviken and Holden (2003) proposed an improved SIR algorithm. In this paper, we consider only the standard form of the SIR algorithm described above.

Not surprisingly, most of the applications of the SIR algorithm are in Bayesian computation. It has been successfully applied in many statistical problems, see for example Gelfand and Smith (1990), McAllister, Pikitch, Punt and Hilborn (1994), Newton and Raftery (1994), Raftery, Givens and Zeh (1995), Berzuini, Best, Gilks and Larizza (1997), Lopes, Moreira and Schmidt (1999), Davison and Louzada-Neto (2000), and Koop and Poirier (2001). Recently, Li (2004) considered the subject in detail.

A fundamental problem for the SIR algorithm is the determination of M. A good choice of M should depend on how close $h(x)$ is to $f(x)$. If $h(x) = f(x)$, we can set $M = m$. The poorer is $h(x)$ as an approximation of $f(x)$, the larger is M compared to m. Rubin (1987, 1988) showed that the SIR algorithm is exact when M/m approaches infinity. In practice, M is usually chosen subjectively: M/m might be one (Albert (1993), Gordon, Salmond and Smith (1993), Kitagawa (1996), Lancaster (1997)); Bunnin, Guo and Ren (2002) chose $M/m = 2$; Tan, Tian and Ng (2003) considered two examples with M close to m ($M/m \leq 1.25$); Rubin (1987) considered $M/m = 20$; Smith and Gelfand (1992) recommended $M/m \geq 10$ in their example.

A theoretical determination of M requires a concrete objective function. Lee (1997) suggested using M to control the mean squared error of the probability estimate. McAllister and Ianelli (1997) required that the maximum $\omega(X)$ be less than 0.04 of the total importance weight. As some inputs may be selected more than once in the unequal probability sampling step, we encounter a duplication problem. Many copies of a value in the output will lead to significant underestimation of standard error in the subsequent analysis when the outputs are used as if they are i.i.d. For this reason, Li (2004) proposed using M to keep the maximum number of duplicates in the output at a tolerably low level. We continue Li's (2004) work and give a general relation between M and m. In this paper, we show that M depends mainly on the right-tail behaviour of $\omega(X)$: if the variance of $\omega(X)$ is not finite, the required M is huge when compared with m; if $\omega(X)$ has a moment generating function, the suggested M is of order $O(m\ln(m))$, consistent with Li's (2004) finding when $\omega(X)$ follows a gamma distribution; the value M is $O(m)$ if and only if $\omega(X)$ is bounded above.

This paper is organized as follows. The SIR algorithm is outlined in Section 2. The relation between M and m is studied in Section 3. Then we consider the numerical determination of M in Section 4; and finally, a conclusion is made in Section 5. The proofs of the main results are given in the appendix.

2. SIR Algorithm

The SIR algorithm consists of two steps: a sampling step and an importance resampling step as given below.

Step 1. (Sampling Step) Generate $X_1, \ldots, X_M$ i.i.d. from $h(x)$, the support of which includes that of $f(x)$.

Step 2. (Importance Resampling Step) Draw m values $\{Y_1, \ldots, Y_m\}$ from $\{X_1, \ldots, X_M\}$ in such a way that

$$\mathrm{E}\,(q_i | X_1, \ldots, X_M) = \frac{m\omega(X_i)}{\sum_{j=1}^{M} \omega(X_j)} \qquad \text{for } i = 1, \ldots, M,$$

where q_i is the number of copies of X_i in $\{Y_1, \ldots, Y_m\}$, and $\omega(X_j) \propto f(X_j)/h(X_j)$ for all j.

We call $\{X_1, \ldots, X_M\}$ a pool of candidate values, $\{Y_1, \ldots, Y_m\}$ a resample, M the pool size, m the resample size, $h(x)$ the importance sampling pdf, and $\omega(X)$ the importance weight of X. Clearly $\omega(X)$ has finite expected value, which we denote by μ. The resample is used as an approximate random sample from $f(x)$.

A variety of resampling algorithms can be used in the importance resampling step to draw a weighted sample. Whether a resampling without replacement is possible or not depends on the random weights, $\omega(X)$. A necessary and sufficient condition for the existence of a without replacement resampling method is that

$$\frac{m\omega(X_i)}{\sum_{j=1}^{M} \omega(X_j)} \leq 1 \tag{1}$$

for all i.

A resampling algorithm is called tight if for $i = 1, \ldots, M$,

$$\left\lfloor \frac{m\omega(X_i)}{\sum_{j=1}^{M} \omega(X_j)} \right\rfloor \leq q_i \leq \left\lceil \frac{m\omega(X_i)}{\sum_{j=1}^{M} \omega(X_j)} \right\rceil,$$

where $\lceil v \rceil$ is the smallest integer larger than or equal to v, and $\lfloor v \rfloor$ is the largest integer less than or equal to v. A tight resampling algorithm always generates a sample without replacement whenever (1) holds (Li (2004)). Li (2004) considered how tight resampling algorithms might be constructed and recommended using them to reduce the variability introduced by the q_i's in the subsequent analysis. In this paper, we consider only tight resampling algorithms.

3. Relation between M and m

It was shown (Rubin (1987, 1988)) that the SIR algorithm generates i.i.d. sample from $f(x)$ when M/m goes to infinity. In other words, given any fixed m, the resample can be of as "high quality" as demanded when M is sufficiently large. Here high quality means two things. First, the distribution of any Y_i in the resample should be close to $f(x)$. Li (2004) discussed how the magnitude of M alone affects the closeness of the distribution of each Y_i to $f(x)$. Second, Y_i's should be "almost independent". Li (2004) used $\max_i q_i$ to measure the stochastic dependence. The smaller the $\max_i q_i$, the better the resample. Li (2004) proposed choosing M large enough that

$$\Pr(\max_i q_i \leq b) \geq 1-\gamma \tag{2}$$

for a given positive integer b ($b < m$), and a small positive value γ ($0 < \gamma < 1$). A necessary condition for (2) is that $M \geq m/b$. Therefore if m approaches infinity, M goes to infinity at least at the same order as m. When a tight resampling algorithm is used in the importance resampling step,

$$\Pr(\max_i q_i \leq b) \geq \Pr\left(\frac{m \max_i \omega(X_i)}{\sum_{j=1}^{M} \omega(X_j)} \leq b\right).$$

If $\Pr(c_1 \leq \omega(X) \leq c_2) = 1$ for $0 < c_1 \leq c_2 < \infty$, $\Pr(\max_i q_i \leq b)$ is one when $M \geq 1+(m-b)c_2/(bc_1)$. In particular, when $f(x) = h(x)$, the inequality becomes $M \geq m/b$, the best lower bound for M.

Generally, it is hard to study (2) for finite m. In this paper, we consider only the case where m approaches infinity - that is, a selection rule for M to make

$$\liminf_{m\to\infty} \Pr(\max_i q_i \leq b) \geq 1-\gamma. \tag{3}$$

Denote the distribution function of $\omega(X)$ by $G(\omega)$. For any $0 \leq p \leq 1$, define $\xi_p = \min\{\omega : G(\omega) \geq p\}$. Clearly $G(\xi_p) \geq p$. Let $\alpha = -\ln(1-\gamma)$. Then

$$\begin{aligned}&\Pr(\max_i q_i \leq b)\\ &\geq \Pr\left(\omega(X_1),\ldots,\omega(X_M) \leq \xi_{1-\frac{\alpha}{M}}\right) \Pr(\max_i q_i \leq b \Big| \omega(X_1),\ldots,\omega(X_M) \leq \xi_{1-\frac{\alpha}{M}})\end{aligned}$$

$$\geq G^M(\xi_{1-\frac{\alpha}{M}})\Pr\left(\frac{m\max_i\omega(X_i)}{\sum_{j=1}^{M}\omega(X_j)}\leq b\middle|\omega(X_1),\ldots,\omega(X_M)\leq\xi_{1-\frac{\alpha}{M}}\right)$$

$$\geq (1-\frac{\alpha}{M})^M\Pr\left(\frac{m\xi_{1-\frac{\alpha}{M}}}{b}\leq\sum_{j=1}^{M}\omega(X_j)\middle|\omega(X_1),\ldots,\omega(X_M)\leq\xi_{1-\frac{\alpha}{M}}\right). \tag{4}$$

The last expression is a product of two terms. The first term converges to $(1-\gamma)$ as M approaches infinity. Inequality (3) holds if the second term goes to one. The second term involves the sum of $\omega(X_j)$ whose distribution depends on M because of the condition $\omega(X_j)\leq\xi_{1-\alpha/M}$. To study the limit of the second term, we need the following lemma, the proof of which is given in Appendix A.1.

Lemma 1. *Let $\varphi(\cdot)$ be a nonnegative nondecreasing real-valued function on $[0,\infty)$. If $\mathrm{E}(\varphi(\omega(X)))$ is finite, $\lim_{M\to\infty}\varphi(\xi_{1-a/M})/M = 0$ for any positive constant a.*

Using Lemma 1, we prove the following theorem in Appendix A.2.

Theorem 1. *Suppose that $\mathrm{E}(\omega(X)^c)$ is finite for a fixed c, $1\leq c\leq 2$. When a tight resampling algorithm is used,* (3) holds if M is chosen such that

$$\lim_{m\to\infty}\frac{M-\frac{m\xi_{1+\frac{\ln(1-\gamma)}{M}}}{b\mu}}{\sqrt{M\xi^{2-c}_{1+\frac{\ln(1-\gamma)}{M}}}}=\infty. \tag{5}$$

Since $E(\omega(X))$ exists, from Lemma 1, $\lim_{M\to\infty}\xi_{1+\ln(1-\gamma)/M}/M = 0$. This implies that

$$\lim_{M\to\infty}\frac{M}{\sqrt{M\xi^{2-c}_{1+\frac{\ln(1-\gamma)}{M}}}}=\lim_{M\to\infty}\sqrt{\frac{M^{c-1}}{\left(\frac{\xi_{1+\frac{\ln(1-\gamma)}{M}}}{M}\right)^{2-c}}}=\infty.$$

As a result, there always exists an M, a function of m, that satisfies (5).

Equation (5) provides valuable insight into the relation between M and m. From Theorem 1 with $c=1$, (5) becomes

$$\lim_{m\to\infty}\frac{M-\frac{m\xi_{1+\frac{\ln(1-\gamma)}{M}}}{b\mu}}{\sqrt{M\xi_{1+\frac{\ln(1-\gamma)}{M}}}}=\infty.$$

Therefore, (3) holds if $M\geq km\xi_{1+\ln(1-\gamma)/M}/(b\mu)$ for any fixed constant $k>1$. The following corollary considers the case when $c>1$.

Corollary 1. *If $E(\omega(X)^c)$ is finite for a $c > 1$, then $\lim_{m\to\infty} \Pr(\max_i q_i \le 1) = 1$ when a tight resampling algorithm is used and $M \ge \tau m^{c/(c-1)}$ for any fixed positive constant τ.*

In particular if $\omega(X)$ has finite variance, M may need to be of order $O(m^2)$. From Corollary 1, if $\mathrm{E}(\omega(X)^c)$ exists for any positive value c, a sufficient condition for $\lim_{m\to\infty} \Pr(\max_i q_i \le 1) = 1$ is that $M \ge \tau m^k$ for any fixed $k > 1$ and $\tau > 0$. In general, the condition cannot be weakened to $M > \tau m$ for any fixed $\tau > 0$. The reason is obvious from the following corollary, the proof of which is given in Appendix A.3.

Corollary 2. *Suppose $\mathrm{E}(\exp(t\omega(X)))$ is finite for a positive constant t. When a tight resampling algorithm is used, $\lim_{m\to\infty} \Pr(\max_i q_i \le b) = 1$ if $M \ge m\{\ln(m) + k\ln(\ln(m))\}/(bt\mu)$ for any fixed constant $k > 1$.*

Corollary 2 shows that when $\omega(X)$ has a moment generating function, the required M is of order $O(m\ln(m))$. This is consistent with the finding of Li (2004) who considered only the case that $\omega(X)$ follows a gamma distribution.

All existing ad hoc rules for M are on the choice of the resampling ratio, M/m. The following theorem studies the situation when M/m is close to a constant. We use $\Phi(\cdot)$ to denote the standard normal distribution function, and z_p to denote the lower $100p\%$ point for the standard normal distribution (i.e., $\Phi(z_p) = p$).

Theorem 2. *When a tight resampling algorithm is used, there is an M which is of order $O(m)$ to make* (3) *hold, if and only if $\omega(X)$ is bounded above. If $\omega(X)$ is bounded above,* (3) *holds when*

$$M \ge \left\{ \sqrt{\frac{\xi_1 m}{b\mu} + \Big(\frac{\sigma z_{1-\gamma}}{2\mu}\Big)^2} + \frac{\sigma z_{1-\gamma}}{2\mu} \right\}^2, \tag{6}$$

where σ^2 is the variance of $\omega(X)$.

4. Numerical Determination of M

As a practical matter, we need to determine M numerically. Equation (5) does not lead to a unique value of M because it is not clear how to quantify the concept of divergence to infinity. An important step in the proof of Theorem 1 in Appendix A.2 has that

$$\liminf_{m\to\infty} \Pr(\max_i q_i \le b)$$
$$\ge \liminf_{m\to\infty} \Big(1 + \frac{\ln(1-\gamma)}{M}\Big)^M \Phi\left(\frac{\sqrt{M}\Big\{\mu - \frac{m\xi_{1+\frac{\ln(1-\gamma)}{M}}}{bM}\Big\}}{\sqrt{\xi^{2-c}_{1+\frac{\ln(1-\gamma)}{M}} E(\omega(X)^c) - \mu^2}} \right). \tag{7}$$

The first factor in (7) is $(1-\gamma-O(1/M))$. The expression in (7) is $(1-\gamma-O(1/M))$ if the second factor is $(1-O(1/M))$. This suggests a choice of M to be the smallest positive integer, not less than $\lceil m/b \rceil$, such that

$$\frac{\sqrt{M}\Big\{\mu - \frac{m\xi_{1+\frac{\ln(1-\gamma)}{M}}}{bM}\Big\}}{\sqrt{\xi^{2-c}_{1+\frac{\ln(1-\gamma)}{M}} E(\omega(X)^c) - \mu^2}} \geq z_{1-\frac{\epsilon}{M}}$$

for a positive constant ϵ. This is equivalent to finding M^*, the smallest positive integer satisfying the inequality

$$M \geq \psi(M) \equiv \frac{m\xi_{1+\frac{\ln(1-\gamma)}{M}}}{b\mu} + z_{1-\frac{\epsilon}{M}} \sqrt{M\Big\{\xi^{2-c}_{1+\frac{\ln(1-\gamma)}{M}} \frac{E(\omega(X)^c)}{\mu^2} - 1\Big\}}. \tag{8}$$

To determine M^* numerically, choose $M_0 \leq M^*$, say $M_0 = \lceil m/b \rceil$. Define $M_i = \lceil \psi(M_{i-1}) \rceil$ for $i = 1, \ldots$. It can be shown that (i) the function $\psi(M)$ is nondecreasing on M, (ii) $M_0 \leq M_1 \leq M_2 \leq \ldots$, (iii) $M_i \leq M^*$ for all i, and (iv) if $M_i = M_{i-1}$, then $M^* = M_i$. As $\{M_i\}$ is a nondecreasing sequence of integers and is bounded above by M^*, from (iv), M_i must be equal to M^* after a finite number of iterations.

When $\omega(X)$ is bounded above, we can use either (6) or (8) to find M. Usually different values of M are suggested. In the ideal case, $h(x) = f(x)$, both (6) and (8) suggest correctly that $M \geq m/b$.

Li (2004) considered the case in which $\omega(X)$ follows a Gamma(θ, β) distribution. He showed that when a tight resampling algorithm is used, (2) is fulfilled if

$$M \Pr(V > b/m \mid V \sim \text{Beta}(\theta, (M-1)\theta)) \leq \gamma. \tag{9}$$

The value M can then be determined as the smallest M value satisfying (9).

We study the performance of the M's suggested by (6), (8), and (9) using three families of distributions, namely the Pareto(s, β), the Gamma(θ, β), and the Beta$(1, \theta)$ distributions. Their density functions are $(s/\beta)(1+x/\beta)^{-s-1}$ for $x \geq 0$, $x^{\theta-1}\exp(-x/\beta)/\{\Gamma(\theta)\beta^\theta\}$ for $x \geq 0$, and $\theta(1-x)^{\theta-1}$ for $0 \leq x \leq 1$ respectively. The Pareto distribution has finite v-th moment only when $v < s$. The Gamma distribution has a moment generating function. The Beta distribution takes values in [0,1].

In the study, we take $\gamma = 0.05$ and $m = 1,000$. To exclude cases in which M is massive, we choose b to be the smallest positive integer such that the corresponding M^* is less than $1,000,000$ when $\epsilon = 1$. Table 1 lists the suggested value of M for different distributions when (6), (8), or (9) is used. For each M, we give $\hat{p}$ which is the relative frequency of $(\{m \max_i \omega(X_i)/\sum_{j=1}^M \omega(X_j)\} \leq b)$ in 1,000 simulation runs. It estimates a lower bound for $\Pr(\max_i q_i \leq b)$.

Table 1. The suggested value of M and its corresponding $\hat{p}$, the estimate of $\Pr(\max_i q_i \le b)$ ($\gamma = 0.05$, $m = 1,000$, $\epsilon = 1$)

Distribution	Formula used		b	Suggested M		$\hat{p}$	
Pareto$(1.5, \beta)_*$	(8) with $c = 1.4$		49	963050		0.980	
Pareto$(2, \beta)_*$	(8) with $c = 1.9$		5	832226		0.949	
Pareto$(2.5, \beta)$	(8) with $c = 2$		2	458764		0.954	
Pareto$(5, \beta)$	(8) with $c = 2$		1	63520		0.957	
Pareto$(10, \beta)$	(8) with $c = 2$		1	25050		0.956	
Gamma$(0.1, \beta)$	(8) with $c = 2$	(9)	1	105627	101009	0.979	0.955
Gamma$(0.5, \beta)$	(8) with $c = 2$	(9)	1	23254	22373	0.969	0.952
Gamma$(1, \beta)$	(8) with $c = 2$	(9)	1	12862	12418	0.961	0.948
Gamma$(2, \beta)$	(8) with $c = 2$	(9)	1	7549	7320	0.968	0.955
Gamma$(10, \beta)$	(8) with $c = 2$	(9)	1	2931	2873	0.973	0.949
Beta$(1, 1)$	(8) with $c = 2$	(6)	1	2088	2043	1.000	0.952
Beta$(1, 2)$	(8) with $c = 2$	(6)	1	3123	3065	1.000	0.997
Beta$(1, 5)$	(8) with $c = 2$	(6)	1	5638	6109	0.993	1.000
Beta$(1, 10)$	(8) with $c = 2$	(6)	1	7970	11159	0.980	1.000
Beta$(1, 20)$	(8) with $c = 2$	(6)	1	9928	21229	0.977	1.000

* Variance is not finite.

From Table 1, the selection rules for M work quite well. Most of the $\hat{p}$'s are larger than $(1 - \gamma) = 0.95$. When the variance of $\omega(X)$ does not exist, the M's are huge. For the Pareto$(1.5, \beta)$ distribution, even if $M/m = 963.05$, there is about 2% chance that there are values with 50 or more copies in the resample. This huge M^* value is in line with Corollary 1, which implies that for the Pareto distribution the required M may need to be of order $O(m^\eta)$ for any $\eta > 3$. The SIR algorithm should not be used when the variance of $\omega(X)$ does not exist. For the Gamma distribution with θ less than or equal to 0.5, M/m is larger than 20, the threshold considered by Rubin (1987). For the Gamma distributions, the performance of (9) looks good and is superior to (8). For the Beta$(1, \theta)$ distributions, (8) is better than (6) when $\theta \ge 5$.

To use (8), we need to choose two constants, c and ϵ. If the variance of $\omega(X)$ is finite, an obvious choice of c is 2. Fortunately ϵ, another user-specified constant, appears only in the second term of $\psi(M)$. From Lemma 1, the second term of $\psi(M)$ is of order less than M. As M^* is the smallest integer satisfying (8), M^* is of the same order as the first term of $\psi(M)$. Thus, the second term is small compared with the first term, lessening the effect of different choices of ϵ on M^*. To numerically investigate the effect of ϵ, we note that M^* decreases as ϵ increases. Therefore, to study the effect of different $\epsilon \in [u_1, u_2]$, we need only compare the M^* for $\epsilon = u_1$ with that for $\epsilon = u_2$. If the ratio is close to one, the effect is minor. We consider only the distributions which have finite variance in

Table 1. Take $u_1 = 0.1$ and $u_2 = 20$. The ratio of the M^* for $\epsilon = 0.1$ to that for $\epsilon = 20$ is computed. The values of the ratio for different distributions are displayed in Figure 1. All the ratios are close to one.

1.00*	P(2.5)
·	G(010), P(005)
1.01*	B(020), G(0.1), G(0.5), G(001), G(002), P(010)
·	B(002), B(005), B(010)
1.02*	B(001)

Figure 1. The ratio of M^* for $\epsilon = 0.1$ to that for $\epsilon = 20$ when $m = 1,000$, $b = 1$, $\gamma = 0.05$ and $c = 2$ (B(θ): Beta(1, θ); G(θ): Gamma(θ, β); P(s): Pareto(s, β))

We may want to determine M directly from Corollaries 1 or 2. However, applying the inequalities in the corollaries numerically requires specifying constants that have significant effect on the suggested M. For Corollary 1, the constant is τ; for Corollary 2, the constant is t (usually t can be chosen from an interval). Caution must be taken in using the inequalities. To demonstrate the problem, let us consider Corollary 2. It suggests selecting a constant $k > 1$, and a value of t, and setting

$$M = \left\lceil \frac{m\{\ln(m) + k\ln(\ln(m))\}}{bt\mu} \right\rceil. \tag{10}$$

As an example, suppose $m = 1,000$, $b = 1$ and $\omega(X)$ follows the Gamma$(2, 1)$ distribution. The moment generating function exists for $t < 1$. Take $k = 1.1$ and $t = 0.95$. Equation (10) gives $M = 4,755$. This choice of M is poor because based on 1,000 simulated random vectors of $\{\omega(X_1), \ldots, \omega(X_{4755})\}$, the estimate of $\Pr(\max_i q_i \leq b)$ is only 0.026. The reason for this poor performance is that we use an asymptotic inequality, $\xi_{1+\ln(1-\gamma)/M} \leq \ln(M)/t$ (see Appendix A.3) in the proof of the corollary. This inequality is equivalent to $\exp\{-(1 - G(\ln(M)/t))M\} \geq 1 - \gamma$. We should check whether $\exp\{-(1 - G(\ln(M)/t))M\}$ is sufficiently close to one before using the value M in (10). For $M = 4,755$ and $t = 0.95$, $\exp\{-(1 - G(\ln(M)/t))M\} = 0.0172$, showing that the suggested M is not large enough for the use of (10). If we choose a smaller t, say $t = 0.6$, (10) gives $M = 7,529$. The corresponding value of $\exp\{-(1-G(\ln(M)/t))M\} = 0.9595$, reasonably close to one. From Table 1, this choice of M is close to that suggested by (8).

5. Conclusion

The SIR algorithm is an easy-to-use approximate sampling method. It is an attractive tool for the Bayesian computation when we can sample from a decent approximate of $f(x)$. We select M to make the values in the resample "less dependent", or formally, to fulfil requirement (3). This paper suggests a simple selection rule for M when the resampling algorithm is tight. A simulation study shows that it gives reliable result.

The SIR algorithm is a reasonable choice of approximate sampling method when the importance weight has a moment generating function. The pool size M is of manageable size because M/m diverges slowly at order $O(\ln(m))$. The divergence rate of M/m gets large when $\omega(X)$ does not have a moment generating function. It is not recommended to use the SIR algorithm when the variance of $\omega(X)$ is not finite since M/m may diverge at the order $O(m)$. Table 1 demonstrates that the required M is huge when variance does not exist.

In practice, we do not know the distribution of the importance weight. A method then is to fit a parametric distribution to $\omega(X)$, and find the required M for the fitted distribution. As an example, consider a univariate case with $f(x)$ a posterior pdf and $h(x)$ the pdf of a normal distribution. Under mild conditions, $f(x)$ is close to a normal pdf. A plausible model is that $\omega(X)$ has the same distribution as $\exp(\theta_1 + \theta_2 Z + \theta_3 Z^2)$ for a standard normal random variable Z. We can estimate the parameters θ_1, θ_2, and θ_3 from the observed $\omega(X)$'s. An overdispersed $h(x)$ is preferred, as it likely leads to a negative θ_3 and thus a bounded $\omega(X)$.

Acknowledgement

The author is grateful to an associate editor and two referees for their helpful comments.

Appendix. Proofs

A.1. Proof of Lemma 1

Suppose that Lemma 1 is false, then there is a subsequence $a < M_1 < M_2 < \ldots$ such that $\varphi(\xi_{1-a/M_j})/M_j$ converges to a positive value v (v can be infinity). Furthermore, we can assume that $M_j \geq 2M_{j-1}$ for all j. If U is $U(0,1)$, ξ_U has the same distribution as $\omega(X)$. Therefore, $\mathrm{E}\left(\varphi(\omega(X))\right) = \int_0^1 \varphi(\xi_u)du \geq \sum_{j=1}^{\infty} \varphi(\xi_{1-a/M_j})\{(1 - a/M_{j+1}) - (1 - a/M_j)\} = a\sum_{j=1}^{\infty}\{\varphi(\xi_{1-a/M_j})/M_j\}(1 - M_j/M_{j+1}) \geq (a/2)\sum_{j=1}^{\infty}\{\varphi(\xi_{1-a/M_j})/M_j\} = \infty$, contradicting the fact that $\mathrm{E}\left(\varphi(\omega(X))\right)$ is finite.

A.2. Proof of Theorem 1

As μ always exists, from Lemma 1, $\lim_{M\to\infty} \xi_{1-\alpha/M}/M = 0$. Therefore, for any fixed M_1, there is $M_2 > M_1$ such that whenever $M > M_2$, we have

$\xi_{1-\alpha/M_1}/M_1 > \xi_{1-\alpha/M}/M$. Write $\overline{\omega(X)}$ for the sample mean of $\omega(X_1),\ldots,\omega(X_M)$. For $M > M_2$, the second factor in (4) is

$$\Pr\Big(\frac{m\xi_{1-\frac{\alpha}{M}}}{bM} \le \overline{\omega(X)} \mid \omega(X_1),\ldots,\omega(X_M) \le \xi_{1-\frac{\alpha}{M}}\Big)$$
$$\ge \Pr\Big(\frac{m\xi_{1-\frac{\alpha}{M_1}}}{bM_1} \le \overline{\omega(X)} \mid \omega(X_1),\ldots,\omega(X_M) \le \xi_{1-\frac{\alpha}{M_1}}\Big).$$

By the Central Limit Theorem, the probability in the last expression converges to

$$\Phi\left(\frac{\sqrt{M}\Big\{\mathrm{E}\left(\omega(X) \mid \omega(X) \le \xi_{1-\frac{\alpha}{M_1}}\right) - \frac{m\xi_{1-\frac{\alpha}{M_1}}}{bM_1}\Big\}}{\sqrt{\mathrm{Var}\left(\omega(X) \mid \omega(X) \le \xi_{1-\frac{\alpha}{M_1}}\right)}}\right).$$

From (5), for sufficiently large M_1, $\mathrm{E}\left(\omega(X) | \, \omega(X) \le \xi_{1-\alpha/M_1}\right) > m\xi_{1-\alpha/M_1}/(bM_1)$. As

$$\mathrm{E}\left(\omega(X)^2 | \, \omega(X) \le \xi_{1-\frac{\alpha}{M_1}}\right) \le \xi^{2-c}_{1-\frac{\alpha}{M_1}} \mathrm{E}\left(\omega(X)^c | \, \omega(X) \le \xi_{1-\frac{\alpha}{M_1}}\right) \le \xi^{2-c}_{1-\frac{\alpha}{M_1}} \mathrm{E}\left(\omega(X)^c\right),$$

we have

$$\liminf_{m\to\infty} \Pr\Big(\frac{m\xi_{1-\frac{\alpha}{M}}}{bM} \le \overline{\omega(X)} \mid \omega(X_1),\ldots,\omega(X_M) \le \xi_{1-\frac{\alpha}{M}}\Big)$$
$$\ge \liminf_{m\to\infty} \Phi\left(\frac{\sqrt{M_1}\Big\{\mathrm{E}\left(\omega(X) \mid \omega(X) \le \xi_{1-\frac{\alpha}{M_1}}\right) - \frac{m\xi_{1-\frac{\alpha}{M_1}}}{bM_1}\Big\}}{\sqrt{\mathrm{Var}\left(\omega(X) \mid \omega(X) \le \xi_{1-\frac{\alpha}{M_1}}\right)}}\right)$$
$$\ge \liminf_{m\to\infty} \Phi\left(\frac{\sqrt{M_1}\Big\{\mu - \frac{m\xi_{1-\frac{\alpha}{M_1}}}{bM_1}\Big\}}{\sqrt{\xi^{2-c}_{1-\frac{\alpha}{M_1}} \mathrm{E}\left(\omega(X)^c\right) - \mu^2}}\right).$$

Condition (5) implies that the probability above converges to one. This completes the proof.

A.3. Proof of Corollary 2

Let γ be any value lying between 0 and 1. From Lemma 1, $\lim_{M\to\infty} \exp(t\xi_{1-\alpha/M})/M = 0$. Therefore, for sufficiently large M, $\xi_{1-\alpha/M} \le \ln(M)/t$. The function $\{M - m\ln(M)/(bt\mu)\}/\sqrt{M}$ is increasing with respect to M when $M \ge 8$. Thus if M satisfies the inequality in the Corollary, for sufficiently large m,

$$\frac{\Big\{M - \frac{m\xi_{1-\frac{\alpha}{M}}}{b\mu}\Big\}}{\sqrt{M}} \ge \frac{\Big\{M - \frac{m\ln(M)}{bt\mu}\Big\}}{\sqrt{M}}$$

$$\geq \frac{\left\{\frac{m\{\ln(m)+k\ln(\ln(m))\}}{bt\mu} - \frac{m\{\ln(m)+\ln\{\ln(m)+k\ln(\ln(m))\}-\ln(bt\mu)\}}{bt\mu}\right\}}{\sqrt{\frac{m\{\ln(m)+k\ln(\ln(m))\}}{bt\mu}}}$$

$$= \frac{\sqrt{\frac{m}{bt\mu}}\{k\ln(\ln(m)) - \ln\{\ln(m) + k\ln(\ln(m))\} + \ln(bt\mu)\}}{\sqrt{\ln(m) + k\ln(\ln(m))}},$$

which approaches infinity as m goes to infinity. This choice of M applies to all γ. The Corollary follows from Theorem 1 with $c = 2$.

A.4. Proof of Theorem 2

[Necessity] Suppose that $M \leq am$ for a positive constant a. Then

$$\Pr(\max_i q_i \leq b) = \Pr\left(\frac{m \max_i \omega(X_i)}{\overline{\omega(X)}M} \leq b\right) \leq \Pr\left(\frac{\max_i \omega(X_i)}{\overline{\omega(X)}} \leq ab\right).$$

By the Strong Law of Large Numbers, $\overline{\omega(X)}$ converges to μ. The last expression has finite limit inferior not less than $(1-\gamma)$ only if $\omega(X)$ is bounded above. [Sufficiency] This can be proved with Theorem 1. To prove that (6) is sufficient for (3), we find

$$\Pr(\max_i q_i \leq b) = \Pr\left(\frac{m \max_i \omega(X_i)}{\overline{\omega(X)}M} \leq b\right) \geq \Pr\left(\frac{m\xi_1}{\overline{\omega(X)}M} \leq b\right).$$

By the Central Limit Theorem, the last expression converges to $\Phi(\sqrt{M}\{\mu - m\xi_1/(Mb)\}/\sigma)$, which is greater than or equal to $(1-\gamma)$ if $\sqrt{M}\{\mu - m\xi_1/(Mb)\}/\sigma \geq z_{1-\gamma}$. Equation (6) follows from simple manipulation of the last inequality.

References

Albert, J. H. (1993). Teaching Bayesian statistics using sampling methods and MINITAB. *Amer. Statist.* **47**, 182-191.

Berzuini, C., Best, N. G., Gilks, W. R. and Larizza, C. (1997). Dynamic conditional independence models and Markov chain Monte Carlo methods. *J. Amer. Statist. Assoc.* **92**, 1403-1412.

Bunnin, F. O., Guo, Y. and Ren, Y. (2002). Option pricing under model and parameter uncertainty using predictive densities. *Statist. Comput.* **12**, 37-44.

Davison, A. C. and Louzada-Neto, F. (2000). Inference for the poly-Weibull model. *The Statistician* **49**, 189-196.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds.) (1996). *Markov Chain Monte Carlo in Practice.* Chapman & Hall, London.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F.* **140**, 107-113.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5**, 1-25.

Koop, G. and Poirier, D. J. (2001). Testing for optimality in job search models. *Econometrics J.* **4**, 257-272.

Lancaster, T. (1997). Exact structural inference in optimal job-search models. *J. Bus. Econom. Statist.* **15**, 165-179.

Lee, D. (1997). Selecting sample sizes for the sampling/importance resampling filter. *Proceedings of the Section on Bayesian Statistical Science*, ASA, 72-77.

Li, K.-H. (2004). The sampling/importance resampling algorithm. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.* (Edited by A. Gelman and X.-L. Meng), 265-276. Wiley, London.

Lopes, H. F., Moreira, A. R. B. and Schmidt, A. M. (1999). Hyperparameter estimation in forecast models. *Comput. Statist. Data Anal.* **29**, 387-410.

McAllister, M. K. and Ianelli, J. N. (1997). Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. *Canad. J. Fisheries and Aquatic Sci.* **54**, 284-300.

McAllister, M. K., Pikitch, E. K., Punt, A. E. and Hilborn, R. (1994). A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Canad. J. Fisheries and Aquatic Sci.* **51**, 2673-2687.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *J. Roy. Statist. Soc. Ser. B* **56**, 3-48.

Raftery, A. E., Givens, G. H. and Zeh, J. E. (1995). Inference from a deterministic population dynamics model for bowhead whales. *J. Amer. Statist. Assoc.* **90**, 402-416.

Rubin, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. Discussion of Tanner and Wong (1987). *J. Amer. Statist. Assoc.* **82**, 543-546.

Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. In *Bayesian Statistics* **3** (Edited by J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith), 395-402. Oxford University Press, Oxford.

Skare, Ø., Bølviken, E. and Holden, L. (2003). Improved sampling-importance resampling and reduced bias importance sampling. *Scand. J. Statist.* **30**, 719-737.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling-resampling perspective. *Amer. Statist.* **46**, 84-88.

Tan, M., Tian, G.-L. and Ng, K. W. (2003). A noniterative sampling method for computing posteriors in the structure of EM-type algorithms. *Statist. Sinica* **13**, 625-639.

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong.

E-mail: khli@cuhk.edu.hk

(Received May 2005; accepted April 2006)