Rao-Blackwellization of Sampling Schemes

George Casella
Biometrics Unit
Cornell University
Ithaca, New York 14853

Christian P. Robert
URA CNRS 1378
Université de Rouen and CREST
Insee, Paris France

BU-1252-M

July 1994

Rao-Blackwellization of Sampling Schemes

George Casella[1] and Christian P. Robert[2]

[1] *Cornell University, Ithaca, NY*

[2] *URA CNRS 1378, Université de Rouen and CREST, Insee, Paris*

Version 3.1 -- July 19,1994

Abstract

This paper compares estimators from two general simulation methods, the Accept-Reject and Metropolis algorithms, with their improved versions based on the Rao-Blackwell Theorem, that is, integration over the uniform random variables involved in the algorithms. We show how the Rao-Blackwellized versions of these algorithms can be implemented and illustrate the improvement brought by these new procedures through examples. We also compare the improved version of the Metropolis algorithm with ordinary and Rao-Blackwellized Importance Sampling procedures for independent and general Metropolis setups.

Keywords: Monte Carlo algorithm, Accept-Reject, Metropolis, Importance Sampling, convex losses, ancillary statistics, polynomial computing time.

This is technical report BU-1252-M in the Biometrics Unit, Cornell University, Ithaca, NY, 14853. Professor Casella's research was supported by National Science Foundation Grant No. DMS93-05547.

1. Introduction.

The Rao-Blackwell Theorem, a well-known result in mathematical statistics (see, e.g., Lehmann, 1983), shows how to improve upon any given estimator under every convex loss function. The improvement is obtained by calculating a conditional expected value, often involving integrating out an ancillary statistic. The appeal of this important theorem has been recently extended to simulation settings in the case of Markov Chain Monte Carlo methods by Gelfand and Smith (1990) and Liu, Wong and Kong (1992). They worked, for the most part, in the context of Gibbs sampling. The Rao-Blackwell Theorem was used to show that smoothed estimators, using the available conditional distributions, were an improvement over non-smoothed estimators. In fact, Liu, Wong and Kong were able to extend the applicability of the Rao-Blackwell Theorem to a non-independent case.

Many simulation methods rely on the side simulation of uniform $U_{[0,1]}$ random variables. When these uniform random variables can be integrated out, the estimators

resulting from these simulations are improved by virtue of the Rao-Blackwell theorem, since the uniform random variables are ancillary and hence do not bring information on the distribution of interest. We consider in this paper two general simulation algorithms, the Accept-Reject and the Metropolis algorithms. Estimators that are constructed from these procedures will typically depend on the ancillary uniform random variables, and we create improved procedures by integrating out these random variables. An important point is that the resulting procedures, called *Rao-Blackwellized procedures*, use all of the candidate random variables simulated while running the algorithm, and are independent of the ancillary uniform random variables. They improve the estimation of the quantity of interest by introducing weight factors. There is actually some similarity between these procedures and procedures resulting from an Importance Sampling algorithm, the weight factors being generally more elaborate in our case (although computable in a polynomial time). A theoretical comparison of these approaches (Rao-Blackwellization versus Importance Sampling) is yet to be undertaken, even though we give some elements of comparison in simulation studies.

The paper is composed as follows. In Section 2 we consider the Accept-Reject Algorithm and derive the corresponding weights for the Rao-Blackwellized version of the estimator. An example, illustrating the potential improvement, is also given. Section 3 similarly treats the Metropolis algorithm in the independent case using the same setup as Section 2, i.e. when a manageable distribution g is available to simulate from, as opposed to the distribution of interest, f. (In a sense, Accept-Reject and Metropolis are comparable in this setting since they make use of the same distribution couple (f, g) in order to generate samples from f. However, since Accept-Reject leads to the generation of a sample of random size, while Metropolis relies on a fixed sample, the comparison is not straightforward, or even relevant.) In Section 3 we also provide an example illustrating the magnitude of improvement possible in this case, and we see that the potential decrease in mean squared error can be quite impressive. Section 4 extends the Metropolis Rao-Blackwellization improvement to the general case, i.e. to the setup where the random variables actually simulated are not independent, and Section 5 discusses Importance Sampling based on the sample generated by the general Metropolis algorithm. In the general Metropolis case, we also give a Rao-Blackwellized version of the Importance Sampling estimator which turns out to have the same order of complexity as the other Rao-Blackwellized procedures. We give an example that shows that the Rao-Blackwellized Importance-Sampling estimator can dramatically improve upon the Rao-Blackwellized Metropolis estimator. However, the Importance Sampling approach does fall short of providing a true sample from the distribution of interest, contrary to Accept-Reject and Metropolis approaches. Lastly, Section 6 contains a discussion and some conclusions.

2. The Accept-Reject Algorithm.

The Accept-Reject algorithm is based on the following lemma.

Lemma 2.1 *If f and g are two densities, and there exists $M < \infty$ such that $f(x) \leq M g(x)$ for every x, the random variable X provided by the algorithm*

1. `Simulate` $Y \sim g(y)$;
2. `Simulate` $U \sim \mathcal{U}_{[0,1]}$ `and take` $X = Y$ `if` $U \leq f(Y)/Mg(Y)$; `otherwise, repeat step 1.`

is distributed according to f.

This algorithm is widely used for simulation, often with some refinements (see Devroye, 1985) to increase the probability of acceptance at each step. Nonetheless, this method leads to the rejection of a part of the sample simulated from g, that is, although we simulate the values $Y_1, \ldots, Y_n$, the Y_i's for which $U_i > f(Y_i)/Mg(Y_i)$ are eliminated. We now propose an improvement upon the original Accept-Reject procedure which makes use of every simulated value.

First, note that we consider the distributions f and g to be given. For every couple (f, g) such that $f(x)/g(x)$ is bounded, the improved (or Rao-Blackwellized) procedure can be constructed and implemented. In fact, the comparison between algorithms based on different g's is not considered in this paper, and is even somehow less important. The choice of g should be based on a compromise between ease of simulation and closeness to the target; the more g is concentrated around the quantity of interest, $\mathbb{E}^f[h(X)]$, the faster the Accept-Reject algorithm converges.

Consider then a sequence $Y_1, Y_2, \ldots$ of i.i.d. random variables generated from g and a corresponding sequence $U_1, U_2, \ldots$ of uniform random variables. Given a function h, the Accept-Reject estimator of $\mathbb{E}^f[h(X)]$, based upon a sample $X_1, \ldots, X_t$ generated according to Lemma 2.1, is given by

$$\hat{\tau}_1 = \frac{1}{t} \sum_{i=1}^{t} h(X_i). \tag{2.1}$$

For a fixed sample size t, which is the number of accepted random variables X_j, the number of generated Y_i's is a random integer N satisfying

$$\sum_{i=1}^{N} \mathbb{I}_{U_i \leq w_i} = t \qquad \text{and} \qquad \sum_{i=1}^{N-1} \mathbb{I}_{U_i \leq w_i} = t - 1,$$

where we define $w_i = f(Y_i)/M\,g(Y_i)$. Since $\hat{\tau}_1$ can be written as

$$\hat{\tau}_1 = \frac{1}{t}\sum_{i=1}^{N} \mathbb{I}_{U_i \leq w_i} h(Y_i),$$

the conditional expectation

$$\hat{\tau}_2 = \frac{1}{t}\mathbb{E}\left[\sum_{i=1}^{N} \mathbb{I}_{U_i \leq w_i} h(Y_i) \middle| N, Y_1, \ldots, Y_N\right] \tag{2.2}$$

improves upon (2.1) by virtue of the Rao-Blackwell theorem. In fact, $\hat{\tau}_1$ and $\hat{\tau}_2$ are both unbiased but the expectation in (2.2) reduces the variance of $\hat{\tau}_2$.

Before deriving a manageable formula for the estimator $\hat{\tau}_2$, we digress for a moment and derive some of the necessary distributions. We skip much of the details of the derivations since they are, for the most part, straightforward applications of standard techniques. The conditional expectation calculations involve averaging over permutations of the realized sample, which is the main consideration in the derivation of the necessary distributions.

The joint distribution of $(N, Y_1, \ldots, Y_N, U_1, \ldots, U_N)$ is given by

$$\begin{aligned}
&P(N = n, Y_1 \leq y_1, \ldots, Y_n \leq y_n, U_1 \leq u_1, \ldots, U_n \leq u_n) = \\
&P(Y_1 \leq y_1, \ldots, Y_n \leq y_n, U_1 \leq u_1, \ldots, U_{n-1} \leq u_{n-1}, U_n \leq (u_n \wedge w_n), \sum_{i=1}^{n-1} \mathbb{I}_{U_i \leq w_i} = t-1) \\
&= \int_{-\infty}^{y_n} g(t_n)(u_n \wedge w_n)dt_n \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_{n-1}} g(t_1)\ldots g(t_{n-1}) \times \\
&\qquad \sum_{(i_1,\ldots,i_{t-1})} \prod_{j=1}^{t-1} (w_{i_j} \wedge u_{i_j}) \prod_{j=t}^{n-1} (u_{i_j} - wi_j)^+ dt_1 \ldots dt_{n-1}.
\end{aligned}$$

The last sum is over all $(t-1)$-tuples $(i_1, \ldots, i_{t-1})$ that make up the different partitions of $\{1, \ldots, n-1\}$ into $\{\{i_1, \ldots, i_{t-1}\}, \{i_t, \ldots, i_{n-1}\}\}$. This joint distribution then provides the distributions involved in $\hat{\tau}_2$. For instance, the conditional distribution of the U_i's is given by

$$\begin{aligned}
&f(u_1, \ldots, u_n | N = n, Y_1, \ldots, Y_n) \propto \\
&\left(\sum_{(i_1,\ldots,i_{t-1})} \prod_{j=1}^{t-1} \mathbb{I}_{u_{i_j} \leq w_{i_j}} \prod_{j=t}^{n-1} \mathbb{I}_{u_{i_j} > w_{i_j}} \prod_{j=1}^{t-1} w_{i_j} \prod_{j=t}^{n-1} (1 - w_{i_j})\right) \mathbb{I}_{u_n \leq w_n} \\
&\qquad \left(\sum_{(i_1,\ldots,i_{t-1})} \prod_{j=1}^{t-1} w_{i_j} \prod_{j=t}^{n-1} (1 - w_{i_j})\right)^{-1}.
\end{aligned}$$

Using this distribution we can calculate the probabilities of the events $\{U_i < w_i\}$ conditional on $N, Y_1, \ldots, Y_N$ and thus derive the weights of $h(Y_i)$ in the estimator $\hat{\tau}_2$. We have

$$\varrho_i = P(U_i \leq w_i | N = n, Y_1, \ldots, Y_n)$$
$$= w_i \sum_{(i_1,\ldots,i_{t-2})} \prod_{j=1}^{t-2} w_{i_j} \prod_{j=t-1}^{n-2} (1 - w_{i_j}) \Big/ \sum_{(i_1,\ldots,i_{t-1})} \prod_{j=1}^{t-1} w_{i_j} \prod_{j=t}^{n-1} (1 - w_{i_j}), \qquad (2.3)$$

while

$$\varrho_n = P(U_n \leq w_n | N = n, Y_1, \ldots, Y_n) = 1.$$

Here, the $(t-2)$-tuples $(i_1, \ldots, i_{t-2})$ indices of the sums mean that the sums are taken over all the different partitions of $\{1, \ldots, n-1\}/\{i\}$ into $\{\{i_1, \ldots, i_{t-2}\}, \{i_{t-1}, \ldots, i_{n-2}\}\}$.

The computation of $\hat{\tau}_2$ now follows quickly from equations (2.2) and (2.3), and is summarized in the following proposition. As can be seen, the complicating factors in the calculation are the sampling scheme and the random size of the sample. However, the resulting estimator can be thought of as an average over all of the possible permutations of the realized sample, with the permutations being weighted by their probabilities.

Proposition 2.2 *For $N = n$, The Rao-Blackwellized version of (2.1) is given by*

$$\hat{\tau}_2 = \frac{1}{t} \sum_{i=1}^{n} \varrho_i h(Y_i)$$

where

$$\varrho_i = P(U_i \leq w_i | N = n, Y_1, \ldots, Y_n)$$

is given by equation (2.3).

The computation of the ϱ_i's may appear formidable but these weights are easily derived from a recurrence relation which is of order n^2. For example, if we define

$$S_k(m) = \sum_{(i_1,\ldots,i_k)} \prod_{j=1}^{k} w_{i_j} \prod_{j=k+1}^{m} (1 - w_{i_j})$$

we can recursively calculate

$$S_k(m) = w_m S_{k-1}(m-1) + (1 - w_m) S_k(m-1)$$

and note that weight ϱ_i of (2.3) is given by

$$\varrho_i = w_i S_{t-2}(n-2) / S_{t-1}(n-1)$$

The Rao-Blackwellized procedure can thus be effectively computed to improve upon the original Accept-Reject procedure. Note also that a consequence of the above calculations is that N has marginally a negative binomial distribution,

$$N \sim \mathcal{N}eg(t, \frac{1}{M}),$$

therefore that the average size of the sample generated from g is tM.

The estimator $\hat{\tau}_2$ can be regarded as an Importance Sampling estimator, but one that uses weights that are based on every variable generated in the process. Compared with the classical Importance Sampling procedure,

$$\hat{\tau}_3 = \frac{1}{N} \sum_{i=1}^{N} \frac{f(Y_i)}{g(Y_i)} h(Y_i), \tag{2.4}$$

$\hat{\tau}_2$ is indeed more involved. However, $\hat{\tau}_3$ does not take into account the fact that N is random, and also lacks the original motivation of unbiasedness. A corrected (that is, unbiased) version of $\hat{\tau}_3$ would use modified weights which are of the degree of complexity of the ϱ_i's. To undertake the determination of the corrected weights is an extremely involved calculation, one that we suspect does not have a closed form solution.

We also note that

$$\begin{aligned} \mathrm{var}(\hat{\tau}_1) &= \mathrm{var}(\mathbb{E}[\hat{\tau}_1 | N, Y_1, \ldots, Y_N]) + \mathbb{E}[\mathrm{var}(\hat{\tau}_1 | N, Y_1, \ldots, Y_N)] \\ &= \mathrm{var}(\hat{\tau}_2) + \mathbb{E}[\mathrm{var}(\hat{\tau}_1 | N, Y_1, \ldots, Y_N)] \end{aligned}$$

so that the improvement brought by $\hat{\tau}_2$ over $\hat{\tau}_1$ is exactly $\mathbb{E}[\mathrm{var}(\hat{\tau}_1 | N, Y_1, \ldots, Y_N)]$. In some cases this quantity can be evaluated in closed form.

To illustrate the potential improvement from Rao-Blackwellization, we look at the following example.

Example 2.1 The target distribution is a Gamma distribution $\mathcal{G}a(\alpha, \beta)$. We set $\beta = 2\alpha$ so that the mean of the distribution is $1/2$. Although it is straightforward to simulate Gamma random variables when α is an integer—since they are then sums of exponential random variables—or when 2α is an integer—because of the relation with the χ^2_α distribution—, Gamma distributions associated with other noninteger α's are more delicate to deal with and an Accept-Reject algorithm may provide an easy simulation method.

To simulate the $\mathcal{G}a(\alpha, \beta)$ distribution, a reasonable candidate is the Gamma $\mathcal{G}a(a, b)$ distribution with $a = [\alpha]$ and $b = \beta a/\alpha$, if $\alpha > 1$. If $\alpha < 1$, the simulation of $\mathcal{G}a(\alpha, \beta)$ can be derived from the simulation of $\mathcal{G}a(\alpha+1, \beta)$ (see Devroye, 1985) and we can thus assume w.l.o.g. that $\alpha > 1$. (It is actually necessary to have $a < \alpha$ in order for M in Lemma 2.1 to be finite.) The choice $b = \beta a/\alpha$ improves the fit between the two distributions since both means match. We consider two cases which reflect different acceptance rates for the Accept-Reject algorithm:

- Case 1: $\alpha = 2.70$, $\beta = 5.40$, $a = 2$, $b = 1.78$ and $1/M = 0.9$;
- Case 2: $\alpha = 2.063$, $\beta = 4.126$, $a = 2$, $b = 3.13$ and $1/M = 0.3$.

For each case we estimated the mean (chosen to be 1/2) and a tail probability (chosen to be 5%) using both the simple Accept-Reject algorithm and its Rao-Blackwellized version. The results are presented in Tables 2.1 and 2.2, which shows the averages of both the Accept-Reject estimators and their Rao-Blackwellized counterparts.

Tables 2.1 and 2.2 about here

We also included mean squared errors estimates for the Accept-Reject estimator and the improvement brought by Rao-Blackwellizing. This improvement is measured by the percentage decrease in mean squared error. From both tables, it can be seen that the Rao-Blackwellization provides a substantial decrease in mean squared error, reaching 60% in the case where the acceptance rate of the algorithm is 0.3. The improvement is better at the lower Accept-Reject acceptance rate partially because the Rao-Blackwellized sample is about three times bigger, with approximately two thirds of the sample being discarded by the Accept-Reject algorithm. Another interesting observation is that the percent improvement in mean squared error remains constant as the Accept-Reject sample size increases, showing that the variance of the original Accept-Reject estimator does not approach the variance of the Rao-Blackwellized estimator when the sample size increases.

3. The Metropolis Algorithm in the Independent Case.

Similar to the Accept-Reject algorithm, the independent Metropolis algorithm constructs a sample $Z_1, \ldots, Z_n$ distributed according to a distribution f from a sample $Y_1, \ldots, Y_n$, generated according to a distribution g, by discarding some of the Y_i's. The main difference between Accept-Reject and Metropolis is that the random variables Z_i are not necessarily independent, but are marginally distributed according to f. But the Metropolis algorithm does not require the ratio f/g to be bounded. Introduced by Metropolis *et al.* (1953) and studied in Hastings (1970), Geyer (1992) and Tierney (1991, 1994), the Metropolis algorithm starts with a random variable Z_0 generated from f and generates a Markov chain (Z_n) as follows.

`Generate` $Z_{n+1}|Z_n$ `as`

$$Z_{n+1} = \begin{cases} Z_n & \texttt{with probability } 1 - \varrho_{n+1}, \\ Y_{n+1} \sim g(y) & \texttt{with probability } \varrho_{n+1}, \end{cases} \tag{3.1}$$

`where`

$$\varrho_{n+1} = \frac{f(Y_{n+1})/g(Y_{n+1})}{f(Z_n)/g(Z_n)} \wedge 1.$$

The assumption that Z_0 is generated from f is not very restrictive given that the Metropolis algorithm converges (in the ergodic sense) to the distribution f. Therefore, after a 'burn-in' period, the current simulation from the Metropolis algorithm can be considered to be approximately generated from the true distribution f. More formally, we can also consider that the Metropolis algorithm is initialized with the first acceptance by an Accept-Reject algorithm if f/g is bounded (see Mengersen and Tweedie, 1993, about the importance of this assumption for the geometric convergence of Metropolis algorithms).

A major difference between the Metropolis Algorithm and the Accept-Reject Algorithm of Section 2 is that the sample size n is now fixed. In Accept-Reject sampling we produce a sample of specified size t, but we generate a random number N of Y_i's. With the Metropolis algorithm we generate, and end up with, a fixed number n of variables Z_j's. To estimate a quantity $\mathbb{E}^f[h(Z)]$, the usual estimator (justified by the Ergodic Theorem) is

$$\hat{\tau}_4 = \frac{1}{n+1} \sum_{i=0}^{n} h(Z_i),$$

which only involves the Y_i's accepted by the Metropolis algorithm. Using the full sample of Y_i's, the estimator $\hat{\tau}_4$ can be written in the form

$$\hat{\tau}_4 = \frac{1}{n+1} \left(h(Z_0) + \sum_{i=1}^{n} \left\{ \mathbb{I}_{Z_i = Y_i} h(Y_i) + \mathbb{I}_{Z_i = Z_{i-1}} h(Z_{i-1}) \right\} \right)$$

$$= \frac{1}{n+1} \sum_{i=0}^{n} h(Y_i) \sum_{j=i}^{n} \mathbb{I}_{Z_j = Y_i}. \tag{3.2}$$

With the convention $Y_0 = Z_0$, this form incorporates all the Y_i's, and only depends on the Y_i's (and not the Z_i's). It also shows that $\hat{\tau}_4$ is a weighted average of the $h(Y_i)$'s with integer weights representing the number of times Z_j equals Y_i. Since the value of $\hat{\tau}_4$ is determined by the ancillary uniform variables linked with (3.1), we can apply Rao-Blackwell Theorem to integrate over the U_i's. If we define the quantities

$$w_i = f(Y_i)/g(Y_i), \qquad \rho_{ij} = (w_j/w_i) \wedge 1, \qquad \delta_i = P(Z_i = Y_i | Y_1, \ldots, Y_i)$$

and

$$\xi_{ii} = 1, \qquad \xi_{ij} = \prod_{t=i+1}^{j} (1 - \rho_{it}), \quad (i < j)$$

we get the following improvement upon $\hat{\tau}_4$.

Proposition 3.1 *The Rao-Blackwellized version of $\hat{\tau}_4$ is given by*

$$\hat{\tau}_5 = \frac{1}{n+1} \sum_{i=0}^{n} \varphi_i h(Y_i),$$

where φ_i is the expected number of times Y_i occurs in the sample, and is given by

$$\varphi_i = \delta_i \sum_{j=i}^{n} \xi_{ij},$$

with

$$\delta_i = \sum_{j=0}^{i-1} \delta_j \xi_{j(i-1)} \rho_{ji}. \qquad (i > 0)$$

Proof. First, the probability that Y_i occurs at least once in the chain is δ_i, given by

$$\begin{aligned}
\delta_i &= P(Z_i = Y_i) \\
&= \sum_{j=0}^{i-1} P(Z_i = Y_i | Z_{i-1} = Y_j) P(Z_{i-1} = Y_j) \\
&= \sum_{j=0}^{i-1} \rho_{ji} P(Z_{i-1} = Y_j) = \sum_{j=0}^{i-1} \rho_{ji} \delta_j \xi_{j(i-1)},
\end{aligned}$$

since the probability $P(Z_i = Y_j)$ $(i > j)$ is given by

$$
\begin{aligned}
P(Z_i = Y_j) &= P(Z_i = Y_j | Z_{i-1} = Y_j) P(Z_{i-1} = Y_j) \\
&= P(Z_i = Y_j | Z_{i-1} = Y_j) \ldots P(Z_j = Y_j) \\
&= (1 - \rho_{ji}) \ldots \delta_j
\end{aligned}
$$

Moreover, once $Y_i = y_i$ is accepted, y_i remains in the sequence as the value of $z_i, \ldots, z_t$ until a new y_{t+1} is accepted. Therefore, using (3.2), the expected number of times y_j occurs in the sample $z_0, \ldots, z_n$ is indeed

$$\delta_j \left(1 + \xi_{j(j+1)} + \xi_{j(j+2)} + \ldots + \xi_{jn}\right). \qquad \blacksquare\blacksquare$$

Again, note that despite its intricate form, the improved estimator $\hat{\tau}_5$ only requires the computation of $n(n-1)/2$ ξ_{ij}'s and n probabilities δ_i's. Therefore, it is highly manageable. An open question about $\hat{\tau}_5$ is the improvement brought not only over $\hat{\tau}_4$ but also over the Importance Sampling estimator (2.4), since $(n/n+1)\hat{\tau}_3 + (h(z_0)/n+1)$ could also be used in this setup. To partially answer this question, and also to evaluate the potential improvement of Rao-Blackwellizing in this case, we again look at an example.

Example 3.1 The target distribution here is a Student's t distribution with 3 degrees of freedom, from which we estimate the mean and a 5% tail probability. The estimation of these quantities is based on an independent Metropolis algorithm, with candidate distribution a Cauchy distribution. (The Cauchy distribution is both easy to simulate from and results in a finite supremum of the ratio $f(y)/g(y)$.)

We compare the usual Metropolis estimate $\hat{\tau}_4$ with its Rao-Blackwellized improvement $\hat{\tau}_5$ and with an Importance Sampling estimate. We do not use the Importance Sampling estimator $\hat{\tau}_3$ of (2.4) however, but rather the version given by

$$\hat{\tau}_6 = \frac{\sum_{i=1}^n \frac{f(Y_i)}{g(Y_i)} h(Y_i)}{\sum_{i=1}^n \frac{f(Y_i)}{g(Y_i)}},$$

which seems to perform better than $\hat{\tau}_3$. Note that $\hat{\tau}_6$ does not depend on any ancillary random variables, so Rao-Blackwellization will not improve it.

Tables 3.1 and 3.2 about here

The results, presented in Tables 3.1 and 3.2, are similar to the previous example. For the sample sizes examined, the Rao-Blackwellized estimator yields a 40-50 % decrease in mean squared error over the ordinary Metropolis mean. What is most surprising is that the Importance Sampling estimator yields an improvement that is comparable to the Rao-Blackwellized Metropolis estimator. This, perhaps, indicates that as an estimation technique, the Metropolis mean may not be very desirable. Of course, the Metropolis algorithm has other uses, such as providing a sample from the target distribution. We also note that the unbiased Importance Sampling estimator (2.4) did not provide an improvement that was comparable to Rao-Blackwellized Metropolis. Thus, if it is desired to retain the property of unbiasedness, the Rao-Blackwellized estimator is the choice.

4. The Metropolis Algorithm in the General Case

Section 3 presented the Metropolis algorithm in the particular case when the new random variable Y_i is generated according to $g(y)$, independently of Z_{i-1}. This perspective is justified when f and g are considered as given parameters. However, the Metropolis algorithm is often used in settings when Y_i is generated according to a conditional distribution $g(y|Z_{i-1})$. We consider in this section the extension of the previous results to this framework and show that a Rao-Blackwellization is also feasible, although the weights are more complex than for the procedure obtained in Proposition 3.1.

The extension from the independent Metropolis algorithm to the general Metropolis algorithm uses a conditional distribution $g(y|z)$ such that the transition from Z_n to Z_{n+1} is

$$Z_{n+1} = \begin{cases} Z_n & \text{with probability } 1 - \varrho_{n+1}, \\ Y_{n+1} \sim g(y_{n+1}|Z_n) & \text{with probability } \varrho_{n+1}, \end{cases}$$

where

$$\varrho_{n+1} = \frac{f(Y_{n+1})/g(Y_{n+1}|Z_n)}{f(Z_n)/g(Z_n|Y_{n+1})} \wedge 1.$$

As long as the support of $g(.|z)$ contains the support of f for all z in the support of f, convergence (ergodicity) is guaranteed.

The random variable generation in this algorithm creates dependencies between the Y_i's which do not even form a Markov chain, the distribution of Y_i depending on $Y_1, \ldots,$ Y_{i-1}. However, this more complex structure does not prevent us from representing the Metropolis algorithm as the construction of a sample $Z_0, Z_1, \ldots, Z_n$ distributed according to f if $Z_0 \sim f$. The sample is derived from the generation of two samples $Y_0 = Z_0, Y_1, \ldots, Y_n$ and $U_1, \ldots, U_n$, the second sample being i.i.d. uniform and thus ancillary for the estimation of $\mathbb{E}^f[h(Z)]$.

The joint distribution of the two samples $(Y_0, \ldots, Y_n)$ and $(U_1, \ldots, U_n)$ is rather involved. In order to give an idea of the complexity of this joint distribution, consider the special case when $n = 4$,

$$\begin{aligned} & f(u_1, \ldots, u_4, y_0, y_1, \ldots, y_4) \propto \\ & \quad \mathbb{I}_{[0,1]^4}(u_1, \ldots, u_4) f(y_0) g(y_1|y_0) \left\{ \mathbb{I}_{u_1 < \rho_{01}} g(y_2|y_1) \left(\mathbb{I}_{u_2 < \rho_{12}} g(y_3|y_2) \left[\mathbb{I}_{u_3 < \rho_{23}} g(y_4|y_3) \right. \right. \right. \\ & \qquad \left. + \mathbb{I}_{u_3 > \rho_{23}} g(y_4|y_2) \right] + \mathbb{I}_{u_2 > \rho_{12}} g(y_3|y_1) \left[\mathbb{I}_{u_3 < \rho_{13}} g(y_4|y_3) + \mathbb{I}_{u_3 > \rho_{13}} g(y_4|y_1) \right] \Big) \\ & \qquad + \mathbb{I}_{u_1 > \rho_{01}} g(y_2|y_0) \left(\mathbb{I}_{u_2 < \rho_{02}} g(y_3|y_2) \left[\mathbb{I}_{u_3 < \rho_{23}} g(y_4|y_3) + \mathbb{I}_{u_3 > \rho_{23}} g(y_4|y_2) \right] \right. \\ & \qquad \quad \left. \left. + \mathbb{I}_{u_2 > \rho_{02}} g(y_3|y_0) \left[\mathbb{I}_{u_3 < \rho_{03}} g(y_4|y_3) + \mathbb{I}_{u_3 > \rho_{03}} g(y_4|y_0) \right] \right) \right\}, \end{aligned}$$

where $(i \leq j)$

$$\rho_{ij} = P(Z_j = Y_j | Z_{j-1} = Y_i) = \frac{f(Y_j)/g(Y_j|Y_i)}{f(Y_i)/g(Y_i|Y_j)} \wedge 1.$$

Therefore, the dependence between the Y_i's creates a dependence between the U_i's, conditionally on the Y_i's, which complicates the derivation of the weights $P(Z_i = Y_j)$. For instance, in the above example,

$$P(Z_3 = Y_2 | Y_0, \ldots, Y_4) \propto \\ \{\rho_{01} g(Y_2|Y_1) \rho_{12} g(Y_3|Y_1) + (1-\rho_{01}) g(Y_2|Y_0) \rho_{02} g(Y_3|Y_1)\} (1-\rho_{23}) g(Y_4|Y_2).$$

However, we are still able to derive a Rao-Blackwellized version of the Metropolis estimator,

$$\hat{\tau}_7 = \frac{1}{n+1} \sum_{i=0}^{n} h(Z_i),$$

as shown by the following result. Consider first the quantities

$$\overline{\rho}_{ij} = \rho_{ij} g(Y_{j+1}|Y_j), \quad \underline{\rho}_{ij} = (1-\rho_{ij}) g(Y_{j+1}|Y_i), \qquad (i < j < n)$$

$$\xi_{jj} = 1, \quad \xi_{jt} = \prod_{\ell=j+1}^{t} \underline{\rho}_{j\ell}, \qquad (j < t < n)$$

$$\delta_0 = 1, \quad \delta_j = \sum_{t=0}^{j-1} \delta_t \xi_{t(j-1)} \overline{\rho}_{tj}, \quad \delta_n = \sum_{t=0}^{n-1} \delta_t \xi_{t(n-1)} \rho_{tn}, \qquad (j < n)$$

$$\omega_n^j = 1, \quad \omega_i^j = \overline{\rho}_{ji} \omega_{i+1}^i + \underline{\rho}_{ji} \omega_{i+1}^j. \qquad (0 \leq j < i < n)$$

Proposition 4.1 *The Rao-Blackwellized version of the general Metropolis algorithm estimator is*

$$\hat{\tau}_8 = \frac{\sum_{i=0}^{n} \varphi_i h(Y_i)}{\sum_{i=0}^{n-1} \delta_i \xi_{i(n-1)}},$$

where $(i < n)$

$$\varphi_i = \delta_i \left\{ \sum_{j=i}^{n-1} \xi_{ij} \omega_{j+1}^i + \xi_{i(n-1)} (1 - \rho_{in}) \right\}$$

and $\varphi_n = \delta_n$.

Proof. As in the independent case, $\hat{\tau}_7$ can be written

$$\frac{1}{n+1} \sum_{i=0}^{n} h(Y_i) \sum_{j=i}^{n} \mathbb{I}_{Z_j = Y_i}.$$

The conditional expectation of the above indicator variables can be written as

$$\begin{aligned} P(Z_j = Y_i | Y_0, Y_1, \ldots, Y_n) &= \mathbb{E}[\mathbb{I}_{Z_i=Y_i} \mathbb{I}_{U_{i+1}>\rho_{i(i+1)}} \cdots \mathbb{I}_{U_j>\rho_{ij}} | Y_0, Y_1, \ldots, Y_n] \\ &= \mathbb{E}\left[\left\{ \mathbb{I}_{Z_{i-1}=Y_{i-1}} \mathbb{I}_{U_i>\rho_{(i-1)i}} + \mathbb{I}_{Z_{i-2}=Y_{i-2}} \mathbb{I}_{U_i>\rho_{(i-2)i}} \mathbb{I}_{U_{(i-1)}>\rho_{(i-2)(i-1)}} + \ldots \right\} \right. \\ &\qquad \left. \mathbb{I}_{U_{i+1}>\rho_{i(i+1)}} \cdots \mathbb{I}_{U_j>\rho_{ij}} | Y_0, Y_1, \ldots, Y_n \right]. \end{aligned}$$

Therefore, conditionally on $Y_0, Y_1, \ldots, Y_n$, the event $\{Z_j = Y_i\}$ appears as the set of all the possible sequences of $(U_1, \ldots, U_i)$ leading to the acceptance of Y_i, of the sequences $(U_{i+1}, \ldots, U_j)$ corresponding to the rejection of $Y_{i+1}, \ldots, Y_j$ and of the sequences $(U_{j+1}, \ldots, U_n)$ constrained by $Z_j = Y_i$. That is,

$$\{Z_j = Y_i\} = \bigcup_{k=0}^{i-1} B_k^{i-1}(U_1, \ldots, U_{i-1}) \cup \left\{ U_i < \rho_{ki}, U_{i+1} > \rho_{i(i+1)}, \ldots, U_j > \rho_{ij} \right\},$$

with $(0 \le k \le t)$

$$B_k^t(U_1, \ldots, U_t) = \bigcup_{m=0}^{k-1} B_m^{k-1}(U_1, \ldots, U_{k-1}) \cup \{U_k < \rho_{mk}, U_{k+1} > \rho_{k(k+1)}, \ldots, U_t > \rho_{kt}\}$$

and

$$B_0^1 = \{U_1 > \rho_{01}\}, \quad B_1^1 = \{U_1 < \rho_{01}\}.$$

If we integrate $\{Z_j = Y_i\}$ with respect to the U_t's, we then get a quantity proportional to

$$\delta_i \prod_{t=i+1}^{j} \underline{\rho}_{it} \omega_{j+1}^i$$

since the probability that $U_t < \rho_{jt}$ is proportional to $\overline{\rho}_{jt}$, while the probability that $U_t > \rho_{jt}$ is proportional to $\underline{\rho}_{jt}$. The weighting factor in $\hat{\tau}_8$ is derived from the following expression

$$\begin{aligned} 1 = \sum_{i=0}^{n} P(Z_n = Y_i) &\propto \left[\sum_{i=0}^{n-1} \delta_i \xi_{i(n-1)} (1 - \rho_{in}) + \delta_n \right] \\ &= \sum_{i=0}^{n-1} \delta_i \xi_{i(n-1)}. \end{aligned}$$

■■

Therefore, despite the correlation between the Y_i's, the conditional expectation of the Metropolis procedure approximately has the same formal structure than in the independent case and requires an amount of computation of the order of n^2.

5. Importance Sampling Estimators for General Metropolis Samples

This section considers the Rao-Blackwellization of an Importance Sampling estimator based on the Metropolis algorithm. While it is possible to implement an Importance Sampling formula in this setup, as the true marginal distribution of the Y_i's can be derived explicitly, the expression of this Importance Sampling estimate is quite intricate. A simpler approach is to consider that the Y_i's $(i > 1)$ are actually generated from the conditional distributions $g(y|Z_{i-1})$ and then use the weights

$$\omega_i = f(Y_i)/g(Y_i|Z_{i-1}).$$

While formally correct, this solution is particularly unsatisfactory since the resulting estimator

$$\hat{\tau}_9 = \frac{1}{n+1}\sum_{j=0}^{n} \omega_j h(Y_j)$$

still depends on the ancillary uniform random variables through the Z_i's.

The developments of the previous section can be exploited to build up an improved version of $\hat{\tau}_9$, by integrating out the U_i's in $\hat{\tau}_9$. In fact, the conditional expectation of $g(Y_i|Z_{i-1})^{-1}$ can be derived from the proof of Proposition 4.1, as

$$\mathbb{E}[g(Y_i|Z_{i-1})^{-1}|Y_0, Y_1, \ldots, Y_n] = \sum_{j=0}^{i-1} \mathbb{E}[g(Y_i|Y_j)^{-1}\mathbb{I}_{Z_{i-1}=Y_j}|Y_0, Y_1, \ldots, Y_n]$$

and

$$\mathbb{E}[\mathbb{I}_{Z_{i-1}=Y_j}|Y_0, Y_1, \ldots, Y_n] \propto \delta_j \xi_{j(i-1)} \omega_i^j. \qquad (0 < i \le n, 0 \le j < n)$$

Therefore, the Rao-Blackwellized version of $\hat{\tau}_9$ is

$$\hat{\tau}_{10} = \frac{1}{n+1}\left\{h(Z_0) + \frac{f(Y_1)}{g(Y_1|Y_0)}h(Y_1) + \frac{\sum_{i=2}^{n}\sum_{j=0}^{i-1} f(Y_i)\delta_j\xi_{j(i-2)}(1-\rho_{j(i-1)})\omega_i^j h(Y_i)}{\sum_{i=0}^{n-1}\delta_i\xi_{i(n-1)}}\right\}.$$

The comparison between this Importance Sampling estimator and the expected Metropolis procedure seems too formidable to be undertaken analytically, so we again illustrate the relative performances of these different procedures through a Student's t – Cauchy example. At this level of generality, we still point out the strong similarity between $\hat{\tau}_8$ and $\hat{\tau}_{10}$. Both estimators are actually taking advantage of all the random variables which have been simulated, although in slightly different ways, the expected Metropolis version being maybe more "complete" in its incorporation of the dependencies of each y_i on the past and on the future. We also again note that these estimators are free of dependence on the ancillary random variables, a desirable property that cannot be enjoyed by the "ordinary" Importance Sampling Estimator $\hat{\tau}_9$.

Example 5.1 The target distribution is again a Student's t distribution with 3 degrees of freedom from which we estimate the mean and the 5% tail probability. The estimation of these quantities is based on a Metropolis algorithm, where now the candidate distribution is a Cauchy distribution centered at the previous random variable Z_{n-1},with scale parameter σ^2. This is a rather academic and inefficient sampling scheme example, but the results are quite interesting.

Because of the somewhat involved nature of both the estimators and the comparisons we have provided an extended simulation study. We ran 50,000 simulations with two different acceptance rates, which are obtaind by choosing different values of σ. When $\sigma = 0.4$, the average acceptance rate is 0.327 and, for $\sigma = 3$, it is 0.749.

Tables 5.1 and 5.2 about here

We compare the usual Metropolis estimate $\hat{\tau}_7$ and its Rao-Blackwellized improvement with the Rao-Blackwellized Importance Sampling estimate $\hat{\tau}_{10}$. (Recall that the ordinary Importance Sampling estimate $\hat{\tau}_9$ depends on the values of the ancillary random variables and is dominated by $\hat{\tau}_{10}$, although the improvement seems small.) In any case, the results presented in Tables 5.1 and 5.2 are rather surprising. First, for a high acceptance rate like 0.75, the mean squared error improvement brought by Rao-Blackwellization upon $\hat{\tau}_7$ is quite minimal, being approximately 0.4% for the mean estimation and 7% for the tail probability estimation. This improvement gets more substantial for the lower acceptance rate since the relative decrease in mean squared error reaches 25% in the best case. The second and maybe even more interesting outcome of these simulation results is the major improvement brought by the use of the Importance Sampling estimate over the corresponding Metropolis and Rao-Blackwellized Metropolis estimates. In fact, the accompanying decrease in mean squared error improves as the sample size increases, and is about 95% for $n = 100$. Therefore, in this case, Importance Sampling appears as a significant improvement upon its Metropolis counterpart.

6. Conclusion.

We have seen that the outputs of simulation schemes such as Accept-Reject and Metropolis algorithms can be improved by use of the entire set of simulated random variables, thanks to the Rao-Blackwell Theorem. This improvement indeed relies on the recycling of the "wasted" simulated random variables and it represents a (statistically) better management of resources. Although the computational implementation may seem involved, the Rao-Blackwellized versions can be easily programmed via recursion relations with computing times that are quadratic (in the sample size).

The Rao-Blackwellizations presented in this paper are essentially non-parametric, in the sense that they neither depend on the form of the density or of the estimated function. In such a non-parametric setting, the Rao-Blackwellized estimator can be perceived as an UMVUE, being symmetric in the order statistics (Lehmann, 1983, Section 2.4). This property should be contrasted with a more familiar and parametric Rao-Blackwellization, often seen in Gibbs sampling (see Gelfand and Smith, 1990). In fact, although the Gibbs sampler is indeed a special case of the Metropolis algorithm, it enjoys the property that the acceptance probability ϱ_n is always equal to 1. Therefore, Gibbs sampling does not allow for "wasted" simulated random variables. On the other hand, Gibbs sampling involves at least a secondary (or auxiliary) chain of random variables for which parametric Rao-Blackwellization may apply.

More precisely, given a target distribution $f_X(x)$, the Gibbs sampler introduces an additional random variable T and conditional densities $f_{X|T}$ and $f_{T|X}$ such that

$$f_X(x) \propto f_{X|T}(x|t)/f_{T|X}(t|x)$$

for all x and t. In terms of X, this is equivalent to the generation of a Metropolis chain with candidate (and transition) density

$$g(x|x') = \int f_{X|T}(x|t) f_{T|X}(t|x') dt,$$

for which the probability ratio ρ is always 1. The parametric Rao-Blackwellization occurs for the T chain since $\mathbb{E}[h(X)]$ can be approximated by

$$\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}[h(X)|T_i] \tag{6.1}$$

rather than by the naive average of the $h(X_i)$'s. This approximation obviously requires the conditional expectation to be available in a closed form for the function h of interest. Note also that this parametric version of Rao-Blackwellized estimator can apply for

Accept-Reject or Metropolis algorithms when they involve auxiliary generations. Alternatively, an Importance Sampling alternative can be proposed for (6.1), by using the weights $f_X(X_i)/f_{X|T}(X_i|T_i)$, proportional to $1/f_{T|X}(T_i|X_i)$.

Let us stress again that our approach, and in particular the optimizations involved in the derivation of the improvements, is statistical rather than computational. Our concern stands with efficiently using available resources (in the statistical sense) and producing estimators that are not dependent on ancillary information, rather than with the construction of optimal algorithms. To make this point clearer, consider an Accept-Reject sample with sample size t and the derivation of its Rao-Blackwellized improvement. It might be more efficient (in time) to generate an Accept-Reject sample of size t^* ($t^* > t$) such that the variance of the estimate based on this augmented sample is smaller than the variance of the Rao-Blackwellized procedure. Although such an approach may be computationally optimal, it will not be statistically optimal.

Our overall comparisons show that Rao-Blackwellization is a viable method that may yield substantial improvement in variance (mean squared error). However, we leave unanswered many other questions concerning the comparisons made in this paper. Firstly, a comparison of Accept-Reject methods with the Metropolis algorithm would be quite interesting, although more than challenging at the theoretical level, even when the same densities are used in the different approaches. (As mentioned earlier, the comparison does not makes sense otherwise.) But the result may depend on the quantity of interest (h).

More importantly, however, is the statistical comparison of Rao-Blackwellized Importance Sampling and of Rao-Blackwellized Metropolis ($\hat{\tau}_8$ and $\hat{\tau}_{10}$), since our single experiment shows a major advantage for the Importance Sampling estimate. In fact, these results cast same doubt on the value of the Metropolis estimator, since it can be so dramatically improved upon. Further work is necessary to theoretically assess these improvements, but we can already mention that the use of Importance Sampling in practical MCMC environments should bring an even greater improvement than in the above simulations. This is because the chain does not usually start from the stationary distribution, and Importance Sampling automatically corrects for the simulation from a wrong distribution. Since using Importance Sampling in a Metropolis environment does not require additional calculations, we advise the use of this estimate, either as the only estimate of the quantity of interest or at least as a control estimate which guarantees that the Metropolis estimate has actually reached stationarity. Note that this comparison could lead to further research, e.g. that a single Metropolis sample could be used in different ways (regular Metropolis, Importance Sampling, Accept-Reject, etc.) in order to increase the confidence in the reported result.

A last comment on this comparison between Metropolis and Importance Sampling is that there exist setups where Importance Sampling estimates cannot be applied because the ratio $f(y)/g(y|z)$ has infinite variance under g. In there situations we are simulating a distribution f from a distribution g with lighter tails (which happens in Gibbs sampling). While this does not (formally) prevent the corresponding Metropolis algorithm from converging, Mengersen and Tweedie (1993) have shown that convergence to the stationary distribution cannot be geometric in such cases. Hence, such schemes should only be used when better candidate densities are unavailable.

Lastly, the idea of a complete decision-theoretic assessment of simulation methods is suggested but untouched by this paper. If such a structure could be set up in an implementable way, comparisons between the various methods introduced above could be undertaken as well as derivations of uniform improvements. For instance, the domintion of Importance Sampling over Metropolis suggested by the simulations could be studied more thoroughly. Moreover, the incorporation of computing costs in loss functions could reunite statistical and computational issues into the decision-theoretic structure.

References

Devroye, L. (1985) *Non-uniform random variate generation.* Springer-Verlag, New York.

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

Geyer, C.J. (1992) Practical Monte Carlo Markov Chain (with discussion). *Statist. Science* **7**, 473-511.

Hastings, W.K. (1970) Monte-Carlo sampling methods using Markov chains and their application. *Biometrika* **57**, 97-109.

Lehmann, E. L. (1983) *Theory of Point Estimation.* Pacific Grove: Wadsworth.

Liu, J.S., Wong, W.H., and Kong, A. (1993) Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and sampling schemes. Tech. report 299, Dept. of Statistics, Harvard University.

Mengersen, K.L. and Tweedie, R.L. (1993) Exact convergence rates for the Hasting and Metropolis sampling algorithms. Tech. report, Dept. of Statistics, Colorado State University, Fort Collins.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-1092.

Tierney, L. (1991) Markov chains for exploring posterior distributions. *Computer Sciences and Statistics: Proc. 23d Symp. Interface*, 563-570.

Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.* (to appear).

Table 2.1 – Estimation of a gamma mean c hosen to be 1/2 using the Accept-Reject Algorithm, based on 7,500 simulations.

a. Accept-Reject algorithm with acceptance rate .9

AR Sample Size	AR Estimate $\hat{\tau}_1$	RB Estimate $\hat{\tau}_2$	AR MSE	Percent Decrease in MSE
10	.5002	.5007	.0100	15.83
25	.5001	.4999	.0041	19.66
50	.4996	.4997	.0020	20.70
100	.4996	.4997	.0010	22.73

b. Accept-Reject algorithm with acceptance rate .3

AR Sample Size	AR Estimate $\hat{\tau}_1$	RB Estimate $\hat{\tau}_2$	AR MSE	Percent Decrease in MSE
10	.5005	.5004	.0012	53.28
25	.4997	.5000	.0005	59.28
50	.4998	.5001	.0002	59.28
100	.4995	.5001	.0001	60.46

Table 2.2 – Estimation of a gamma tail probability c hosen to be .05 using the Accept-Reject Algorithm, based on 7, 500 simulations.

a. Accept-Reject algorithm with acceptance rate .9

AR Sample Size	AR Estimate $\hat{\tau}_1$	RB Estimate $\hat{\tau}_2$	AR MSE	Percent Decrease in MSE
10	.0512	.0503	.0049	20.45
25	.0508	.0501	.0019	22.37
50	.0506	.0507	.0009	21.06
100	.0504	.0503	.0005	21.14

b. Accept-Reject algorithm with acceptance rate .3

AR Sample Size	AR Estimate $\hat{\tau}_1$	RB Estimate $\hat{\tau}_2$	AR MSE	Percent Decrease in MSE
10	.0495	.0505	.0048	63.02
25	.0499	.0506	.0019	69.80
50	.0491	.0498	.0009	72.17
100	.0487	.0498	.0005	73.77

Table 3.1 – Estimation of the mean 0 of a Student's t distribution with 3 degrees of freedom based on an independent Metropolis Sample using a Cauchy distribution 7,500 simulations.

Sample Size	Metrop. Estimate $\hat{\tau}_4$	RB Metrop. $\hat{\tau}_5$	Import. Sampling $\hat{\tau}_6$	MSE Metrop.	MSE Decrease $\hat{\tau}_5$ over $\hat{\tau}_4$	MSE Decrease $\hat{\tau}_6$ over $\hat{\tau}_4$
10	-.0035	-.0058	-.0064	.3622	50.11	47.80
25	-.0037	-.0015	-.0023	.1468	49.39	51.05
50	-.0020	.0012	.0004	.0724	48.27	51.59
100	-.0027	-.0007	-.0009	.0361	46.68	51.19

Table 3.2 – Estimation of a tail probability c hosen to be .05 of a Student's t distribution with 3 degrees of freedom based on an independent Metropolis Sample using a Cauchy distribution 7,500 simulations.

Sample Size	Metrop. Estimate $\hat{\tau}_4$	RB Metrop. $\hat{\tau}_5$	Import. Sampling $\hat{\tau}_6$	MSE Metrop.	MSE Decrease $\hat{\tau}_5$ over $\hat{\tau}_4$	MSE Decrease $\hat{\tau}_6$ over $\hat{\tau}_4$
10	.0490	.0487	.0505	.0056	42.20	38.91
25	.0486	.0490	.0498	.0024	44.75	45.90
50	.0488	.0491	.0497	.0012	45.44	48.68
100	.0494	.0496	.0498	.0006	44.57	49.16

Table 5.1 – Estimation of the mean 0 of a Student's t distribution with 3 degrees of freedom based on a dependent Metropolis Sample using a Cauchy distribution 5 0,000 simulations.

a. Acceptance rate .327

Sample Size	Metrop. Estimate $\hat{\tau}_7$	RB Metrop. $\hat{\tau}_8$	RB Import. Sampling $\hat{\tau}_{10}$	MSE Metrop.	MSE Decrease $\hat{\tau}_8$ over $\hat{\tau}_7$	MSE Decrease $\hat{\tau}_{10}$ over $\hat{\tau}_7$
10	-.0002	-.0016	-.0028	1.517	10.71	87.17
25	-.0012	-.0004	-.0008	.9841	8.78	92.02
50	.0032	.0021	-.0004	.6252	7.68	93.63
100	.0012	.0004	.0005	.3002	7.89	93.40

b. Acceptance rate .749

Sample Size	Metrop. Estimate $\hat{\tau}_7$	RB Metrop. $\hat{\tau}_8$	RB Import. Sampling $\hat{\tau}_{10}$	MSE Metrop.	MSE Decrease $\hat{\tau}_8$ over $\hat{\tau}_7$	MSE Decrease $\hat{\tau}_{10}$ over $\hat{\tau}_7$
10	.0014	.0001	.0005	2.2849	.1751	77.93
25	.0003	.0030	.0008	1.7698	.1526	85.98
50	.0017	.0017	.0018	1.3066	.1071	90.31
100	.0014	.0014	.0003	.8681	.0691	92.85

Table 5.2 – Estimation of a tail probability c hosan to be .05 of a Student's t distribution with 3 degrees of freedom based on a dependent Metropolis Sample using a Cauchy distribution 5 0,000 simulations.

a. Acceptance rate .327

Sample Size	Metrop. Estimate $\hat{\tau}_7$	RB Metrop. $\hat{\tau}_8$	RB Import. Sampling $\hat{\tau}_{10}$	MSE Metrop.	MSE Decrease $\hat{\tau}_8$ over $\hat{\tau}_7$	MSE Decrease $\hat{\tau}_{10}$ over $\hat{\tau}_7$
10	.0500	.0499	.0499	.0203	23.64	88.66
25	.0504	.0503	.0501	.0111	25.22	92.70
50	.0505	.0504	.0500	.0062	25.80	93.87
100	.0501	.0502	.0500	.0032	25.00	94.37

b. Acceptance rate .749

Sample Size	Metrop. Estimate $\hat{\tau}_7$	RB Metrop. $\hat{\tau}_8$	RB Import. Sampling $\hat{\tau}_{10}$	MSE Metrop.	MSE Decrease $\hat{\tau}_8$ over $\hat{\tau}_7$	MSE Decrease $\hat{\tau}_{10}$ over $\hat{\tau}_7$
10	.0501	.0503	.0502	.0301	.9967	75.08
25	.0505	.0505	.0500	.0212	.9433	84.43
50	.0503	.0503	.0501	.0140	.7142	87.86
100	.0500	.0501	.0498	.0084	1.190	90.24