Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

OPTIMAL RESOURCE ALLOCATION IN TWO STAGE SAMPLING OF INPUT DISTRIBUTIONS

Achal Bassamboo

Kellogg School of Management
Northwestern University
Evanston, IL 60208, U.S.A.

Sandeep Juneja

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, Maharashtra 400005, INDIA

ABSTRACT

Consider a performance measure that is evaluated via Monte Carlo simulation where input distributions to the underlying model may involve *two stage sampling.* The settings of interest include the case where in the first stage physical samples from the distribution are collected. In the second stage, Monte Carlo sampling is done from the observed empirical distribution. We also consider the sampling-importance resampling (SIR) algorithm. Here it is difficult to sample directly from the desired input distribution, and these samples are generated in two stages. In the first stage, a large number of samples are generated from a distribution convenient from the sampling viewpoint. In the second stage, a resampling is done from the samples generated in the first stage so that asymptotically the new samples have the desired distribution. We discuss how to allocate computational and other effort optimally the two stages to minimize the estimator's resultant mean square error.

1 INTRODUCTION

In this paper we consider optimal allocation of computational and other resources to minimize the mean square error of the estimated performance measure when two stage sampling is involved. Two specific set-ups are considered:

1. We first consider the settings where some of the input distributions to the model are estimated from data that may be physically procured, may be extracted from a database or may be an output from another simulation model. For example, a company may run a huge simulation model of all its investments to generate samples of overall single period returns. The first stage then corresponds to this data generation. The second stage corresponds to sampling from the distribution fitted to this data. To keep the discussion simple, we assume that in the second stage, sampling is done using the empirical distribution of the data generated in the first stage. We refer to this as the data-simulation trade-off setting (DST).
2. We then consider the Sampling-Importance Resampling (SIR) algorithm where again a two stage sampling problem is encountered. The aim of SIR is to generate random samples from a target input distribution π. To achieve this, in the first stage it generates a random sample from another distribution φ that is convenient from sampling viewpoint. In the second stage it resamples from the samples generated in the first stage. The probability assigned to sample X_i (generated in the first stage) in the second stage is proportional to $\pi(X_i)/\varphi(X_i)$. It is well known that as the number of samples in the first stage increase to infinity, the distribution of the samples generated in the second stage converges to π (see, e.g., Rubin 1988, Geweke 1989, and Smith and Gelfand 1992). We refer to this as the SIR trade-off (SIRT) setting.

In this paper, in the above two settings, we assume that there are fixed per sample costs to generate data in each of the two stages. We then find the asymptotically optimal allocation of the overall budget (as it increases to infinity) to the two stages to minimize the mean square error of the resultant performance measure. In the SIRT we also identify the the density φ that asymptotically minimizes this mean square error.

To keep the analysis simple, we consider a simple performance measure. The authors will conduct a more elaborate analysis in general settings in a separate work.

In Section 2, we develop the mathematical framework and conduct analysis for DST. In Section 3, we do this for SIRT. We end with a brief conclusion and directions for further research in Section 4.

2 DATA SIMULATION TRADE-OFF

As mentioned in the previous section, in our analysis we restrict ourselves to a simple framework. We consider real valued random variables (rv) X, Y and Z defined on a probability space $(\Omega, \mathcal{F}, P)$. These rv are assumed to be independent. Their probability density functions are given by π_x, π_y and π_z, respectively. Consider a function $f : \Re^3 \to \Re$. Our aim is to estimate $\mu = \mathbb{E}[f(X,Y,Z)]$. We assume that $f \in L^2$, i.e., $\mathbb{E}[f(X,Y,Z)^2] < \infty$. We assume that π_z is known. In this section, in the DST settings, we assume that π_x and π_y are not known to us but we can gather i.i.d. samples $(X_1, X_2, \ldots)$ from distribution π_x and i.i.d. samples $(Y_1, Y_2, \ldots)$ from distribution π_y. This represents the collection of data.

Consider the following strategy to estimate $\mu = \mathbb{E}[f(X,Y,Z)]$. We first procure samples

$$(X_1, X_2, \ldots, X_{n_x})$$

and

$$(Y_1, Y_2, \ldots, Y_{n_y}).$$

We then generate via simulation i.i.d. samples $(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n)$ using the empirical distribution associated with $(X_1, X_2, \ldots, X_{n_x})$ so that each $P(\tilde{X}_i = X_j) = 1/n_x$ for each $i \leq n$ and each $j \leq n_x$. Similarly, we generate $(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n)$, where we have suppressed the dependence of $\tilde{X}_i$ and $\tilde{Y}_j$ on the generated data $(X_1, X_2, \ldots, X_{n_x})$ and $(Y_1, Y_2, \ldots, Y_{n_y})$ for notational convenience. We also generate samples i.i.d. $(Z_1, Z_2, \ldots, Z_n)$ using π_z. The resultant estimator for μ equals

$$\hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} f(\tilde{X}_i, \tilde{Y}_i, Z_i).$$

Suppose that it costs c_x to generate a single data point X_i, c_y to generate a single data point Y_i, and c to generate a sample of $f(\tilde{X}_i, \tilde{Y}_i, Z_i)$. Note that cost of generating a sample of X_i or Y_i may be known in monetary value or maybe in terms of computer time. The cost of generating $f(\tilde{X}_i, \tilde{Y}_i, Z_i)$ may be estimated in terms of computer time. We are assuming that all these costs can be measured in same units (monetary or computer time). Further suppose that total budget available to us is C. Note that $E(\hat{\alpha}_n - \mu)^2$ denotes the mean square error of *rv* $\hat{\alpha}_n$. Then our optimization problem reduces to

$$\min E(\hat{\alpha}_n - \mu)^2$$

subject to

$$c_x n_x + c_y n_y + cn = C,$$

where n_x, n_y and n are non-negative. We analyze this problem asymptotically as $C \to \infty$.

We set $n_x = \beta_x C$, $n_y = \beta_y C$ and $n = \beta C$ for positive values of β_x, β_y and β. We then show that

$$\lim_{C \to \infty} C \times E(\hat{\alpha}_{\beta C} - \mu)^2 = g(\beta_x, \beta_y, \beta)$$

for a convex function g that we identify in our analysis. Asymptotically, then our optimization problem reduces to

$$\min g(\beta_x, \beta_y, \beta)$$

subject to

$$c_x \beta_x + c_y \beta_y + c\beta = 1$$

where β_x, β_y and β are non-negative. We refer to this optimization problem as **O1**. As we observe in the next subsection, this is easily solved.

2.1 Evaluating $g(\beta_x, \beta_y, \beta)$

Some additional notation is presented below.

- $\mathcal{F}_\infty$: sigma field $\sigma((X_1, Y_1), (X_2, Y_2), \ldots)$;
- $\sigma^2(f)$: variance of $f(X,Y,Z)$;
- Random variables (X', Y'): independent rv with same distribution as (X,Y) and independent of (X,Y);
- $h(X,Y)$: $\mathbb{E}[f(X,Y,Z)|\sigma(X,Y)]$;
- $\sigma_{X,Y'}$: covariance of $h(X,Y)$ and $h(X,Y')$;
- $\sigma_{X',Y}$: covariance of $h(X,Y)$ and $h(X',Y)$.

Note that $\sigma_{X,Y'}$ and $\sigma_{X',Y}$ are non-negative. To see this for $\sigma_{X,Y'}$ note that

$$\sigma_{X,Y'} = \mathbb{E}\left[\mathbb{E}[f(X,Y,Z)|\mathcal{F}_\infty]\mathbb{E}[f(X,Y',Z)|\mathcal{F}_\infty]\right] - (Ef)^2.$$

The first term may be re-expressed as

$$\mathbb{E}\left[\mathbb{E}[f(X,Y,Z)|\mathcal{F}_\infty^X]^2\right]$$

where $\mathcal{F}_\infty^X$ denotes the sigma-field generated by $(X_1, X_2, \ldots)$. This obviously dominates $(Ef)^2$.

Proposition 1 *For $f \in L^2$,*

$$g(\beta_x, \beta_y, \beta) = \frac{\sigma^2(f)}{\beta} + \frac{\sigma_{X,Y'}}{\beta_y} + \frac{\sigma_{X',Y}}{\beta_x},$$

so that g is a convex function. Furthermore, the solution to the convex optimization problem **O1** *is:*

$$\beta_x = \frac{\sqrt{\sigma_{X',Y}/c_x}}{\sqrt{\sigma_{X',Y}c_x}+\sqrt{\sigma_{X,Y'}c_y}+\sqrt{\sigma^2(f)c}},$$

$$\beta_y = \frac{\sqrt{\sigma_{X,Y'}/c_y}}{\sqrt{\sigma_{X',Y}c_x}+\sqrt{\sigma_{X,Y'}c_y}+\sqrt{\sigma^2(f)c}},$$

and

$$\beta = \frac{\sqrt{\sigma^2(f)/c}}{\sqrt{\sigma_{X',Y}c_x}+\sqrt{\sigma_{X,Y'}c_y}+\sqrt{\sigma^2(f)c}}.$$

The results of the proposition are quite intuitive. For instance, all else being equal, β is large if the variance of the function f is large and if the cost of generating a sample c is small.

Note that the solution to **O1** follows from the first order conditions once the form of g is established. We now focus our efforts on identifying g. Set

$$\hat{\mu}_C = \mathbb{E}[f(\tilde{X}_i,\tilde{Y}_j,Z)|\mathcal{F}_\infty].$$

It follows that

$$\hat{\mu}_C = \frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f(X_i,Y_j,Z)|\mathcal{F}_\infty].$$

We may re-express the mean square error

$$\begin{aligned}
\mathbb{E}[(\widehat{\alpha}_{\beta C}-\mu)^2] &= \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C}-\mu)^2|\mathcal{F}_\infty]] \\
&= \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C}-\hat{\mu}_C+\hat{\mu}_C-\mu)^2|\mathcal{F}_\infty]] \\
&= \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C}-\hat{\mu}_C)^2|\mathcal{F}_\infty]] \\
&\quad +\mathbb{E}[(\hat{\mu}_C-\mu)^2] \qquad (1)
\end{aligned}$$

Next we analyze the two terms in (1) separately to identify g.

Consider the term $\mathbb{E}[(\widehat{\alpha}_{\beta C}-\hat{\mu}_C)^2|\mathcal{F}_\infty]$. This is simply conditional variance of $\widehat{\alpha}_{\beta C}$ that conditioned on $\mathcal{F}_\infty$ is simply an average of βC iid terms. It can be seen to equal

$$\frac{1}{\beta C}\left[\frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f(X_i,Y_j,Z)^2|\mathcal{F}_\infty] - \left(\frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f(X_i,Y_j,Z)|\mathcal{F}_\infty]\right)^2\right].$$

Let W denote the rv within the large round brackets in the above equation and let $\sigma^2(W)$ denote its variance. Further note that $\mathbb{E}[W]=\mathbb{E}[f]$. Then the expected value of above rv simplifies to

$$\frac{1}{\beta C}(\sigma^2(f))-\sigma^2(W)).$$

It is easy to see that $\sigma^2(W)$ is $O(1/C)$, hence

$$\lim_{C\to\infty} C\times\mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C}-\hat{\mu}_C)^2|\mathcal{F}_\infty]] = \sigma^2(f)/\beta.$$

Now consider the term $\mathbb{E}[(\hat{\mu}_C-\mu)^2]$. Recall that $h(X,Y)=\mathbb{E}[f(X,Y,Z)|\sigma(X,Y)]$. Then,

$$\mathbb{E}[(\hat{\mu}_C-\mu)^2] = \mathbb{E}\left[\frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}(h(X_i,Y_j)-\mu)^2\right].$$

It is easy to see through expanding the square term inside the expectation in the RHS above that $\mathbb{E}[(\hat{\mu}_C-\mu)^2]$ equals

$$\frac{\sigma^2(h(X,Y))}{\beta_x\beta_y C^2}+\frac{(\beta_x C-1)\sigma_{X,Y'}}{\beta_x\beta_y C^2}+\frac{(\beta_y C-1)\sigma_{X',Y}}{\beta_x\beta_y C^2}.$$

In particular, it follows that

$$\lim_{C\to\infty} C\times\mathbb{E}[(\hat{\mu}_C-\mu)^2] = \frac{\sigma_{X,Y'}}{\beta_y}+\frac{\sigma_{X',Y}}{\beta_x}$$

and the form of the function g in the proposition follows.

3 SAMPLE IMPORTANCE RESAMPLING ALGORITHM TRADE-OFF

Next we study the use of SIR algorithm in the same setting as described in Section 2. We assume that all the three probability density functions are known. In this section, in the SIRT settings, we assume that it is difficult to generate from the density π_x and π_y so SIR is used to generate samples approximately from distributions, whereas we can generate easily from π_z.

Consider the following strategy to estimate $\mu = \mathbb{E}[f(X,Y,Z)]$. In the first stage, we generate samples

$$(X_1',X_2',\ldots,X_{n_x}')$$

and

$$(Y_1',Y_2',\ldots,Y_{n_y}')$$

under the probability density function φ_x and φ_y. In the second stage we generate i.i.d. samples $(\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_n)$ using the first stage sample $(X_1', X_2', \dots, X_{n_x}')$ so that each

$$P(\tilde{X}_i = X_j') = \frac{\pi_x(X_j')/\varphi_x(X_j')}{\sum_{j=1}^{n_x}(\pi_x(X_j')/\varphi_x(X_j'))}$$

for each $i \leq n$ and each $j \leq n_x$. Similarly, we generate $(\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n)$. Further for future use, we define $k_x = \pi_x/\varphi_x$ and $k_y = \pi_y/\varphi_y$. We also generate samples i.i.d. $(Z_1, Z_2, \dots, Z_n)$ using π_z. The resultant estimator for μ equals

$$\hat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} f(\tilde{X}_i, \tilde{Y}_i, Z_i).$$

As in Section 2, suppose that it costs c_x to generate a single data point X_i', c_y to generate a single data point Y_i', and c to generate a sample of $f(\tilde{X}_i, \tilde{Y}_i, Z_i)$. Again suppose that total budget available to us is C. Note that $E(\hat{\alpha}_n - \mu)^2$ denotes the mean square error of *rv* $\hat{\alpha}_n$. Then our optimization problem reduces to

$$\min E(\hat{\alpha}_n - \mu)^2$$

subject to

$$c_x n_x + c_y n_y + cn = C,$$

where n_x, n_y and n are non-negative. To analyze this problem asymptotically as $C \to \infty$, as in Section 2, set $n_x = \beta_x C$, $n_y = \beta_y C$ and $n = \beta C$ for positive values of β_x, β_y and β. We then show that

$$\lim_{C\to\infty} C \times E(\hat{\alpha}_{\beta C} - \mu)^2 = \widetilde{g}(\beta_x, \beta_y, \beta)$$

for a convex function $\widetilde{g}$ that we identify in our analysis. Asymptotically, then our optimization problem reduces to

$$\min \quad \widetilde{g}(\beta_x, \beta_y, \beta)$$

subject to

$$c_x\beta_x + c_y\beta_y + c\beta = 1$$

where β_x, β_y and β are non-negative. We refer to this optimization problem as **O2**.

3.1 Evaluating $\widetilde{g}(\beta_x, \beta_y, \beta)$

We require the following additional notation

- $\mathcal{F}_\infty$: sigma field $\sigma((X_1', Y_1'), (X_2', Y_2'), \dots)$.
- $\widetilde{h}(X', Y')$: $\mathbb{E}[f(X', Y', Z)|\sigma(X', Y')]$.
- $\phi_x(X')$: $\mathbb{E}[(h(X', Y)|\sigma(X')]$ where Y is a rv with probability density function π_y and is independent of X'.
- $\phi_y(Y')$: $\mathbb{E}[(h(X, Y')|\sigma(Y')]$ where X is a rv with probability density function π_x and is independent of Y'.

Recall that $\sigma^2(f)$ represents the variance of $f(X, Y, Z)$, where X, Y and Z have probability density function π_x, π_y and π_z, respectively.

We now show that under the assumption of uniform integrability for certain sequences of rv,

$$\widetilde{g}(\beta_x, \beta_y, \beta) = \frac{\sigma^2(f)}{\beta} + \frac{M_x}{\beta_y} + \frac{M_y}{\beta_x}, \tag{2}$$

where

$$M_x = \mathbb{E}[(\phi_x(X_1') - \mu)^2 k_x^2(X_1')] \tag{3}$$
$$M_y = \mathbb{E}[(\phi_y(Y_1') - \mu)^2 k_y^2(Y_1')], \tag{4}$$

so that $\widetilde{g}$ is a convex function. Furthermore, the solution to the convex optimization problem **O2** is:

$$\beta_x = \frac{\sqrt{M_x/c_x}}{\sqrt{M_x c_x} + \sqrt{M_y c_y} + \sqrt{\sigma^2(f)c}},$$

$$\beta_y = \frac{\sqrt{M_y/c_y}}{\sqrt{M_x c_x} + \sqrt{M_y c_y} + \sqrt{\sigma^2(f)c}},$$

and

$$\beta = \frac{\sqrt{\sigma^2(f)/c}}{\sqrt{M_x c_x} + \sqrt{M_y c_y} + \sqrt{\sigma^2(f)c}}.$$

Similar to the solution of **O1**, the solution to **O2** follows from the first order conditions once the form of $\widetilde{g}$ is established. We now focus our efforts on identifying $\widetilde{g}$. Let

$$\hat{\mu}_C = \mathbb{E}[f(\tilde{X}_i, \tilde{Y}_j, Z)|\mathcal{F}_\infty].$$

It follows that

$$\hat{\mu}_C = \frac{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C} \mathbb{E}[f(X_i', Y_j', Z)|\mathcal{F}_\infty] k_x(X_i') k_y(Y_j')}{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C} k_x(X_i') k_y(Y_j')}.$$

We may re-express the mean square error

$$\begin{aligned}
\mathbb{E}[(\widehat{\alpha}_{\beta C} - \mu)^2] &= \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C} - \mu)^2|\mathcal{F}_\infty]] \\
&= \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C} - \hat{\mu}_C + \hat{\mu}_C - \mu)^2|\mathcal{F}_\infty]] \\
&= \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C} - \hat{\mu}_C)^2|\mathcal{F}_\infty]] \\
&\quad + \mathbb{E}[(\hat{\mu}_C - \mu)^2] \qquad (5)
\end{aligned}$$

Next we analyze the two terms in (5) separately to identify $\widetilde{g}$.

Consider the term $\mathbb{E}[(\widehat{\alpha}_{\beta C} - \hat{\mu}_C)^2|\mathcal{F}_\infty]$. This is simply conditional variance of $\widehat{\alpha}_{\beta C}$ that conditioned on $\mathcal{F}_\infty$ is simply an average of βC iid terms. It can be seen to equal

$$\frac{1}{\beta C}\left[\frac{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f^2(X_i',Y_j',Z)|\mathcal{F}_\infty]k_x(X_i')k_y(Y_j')}{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}k_x(X_i')k_y(Y_j')} - \left(\frac{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f(X_i',Y_j',Z)|\mathcal{F}_\infty]k_x(X_i')k_y(Y_j')}{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}k_x(X_i')k_y(Y_j')}\right)^2\right].$$

Using SLLN we have

$$\begin{aligned}
&\frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f^2(X_i',Y_j',Z)|\mathcal{F}_\infty]k_x(X_i')k_y(Y_j') \\
&\to \mathbb{E}[f^2(X_1',Y_1',Z_1)k_x(X_1')k_y(Y_1')] \quad \text{a.s.} \\
&= \mathbb{E}[f^2(X,Y,Z)], \qquad (6) \\
&\frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}\mathbb{E}[f(X_i',Y_j',Z)|\mathcal{F}_\infty]k_x(X_i')k_y(Y_j') \\
&\to \mathbb{E}[f(X_1',Y_1',Z_1)k_x(X_1')k_y(Y_1')] \quad \text{a.s.}, \\
&= \mathbb{E}[f(X,Y,Z)], \qquad (7) \\
&\frac{1}{\beta_x\beta_y C^2}\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}k_x(X_i')k_y(Y_j') \\
&\to \mathbb{E}[k_x(X_1')k_y(Y_1')] \quad \text{a.s.} \\
&= 1, \qquad (8)
\end{aligned}$$

as $C \to \infty$.

Hence under suitable uniform integrability conditions we have

$$\lim_{C\to\infty} C \times \mathbb{E}[\mathbb{E}[(\widehat{\alpha}_{\beta C} - \hat{\mu}_C)^2|\mathcal{F}_\infty]] = \sigma^2(f)/\beta.$$

Now consider the term $\mathbb{E}[(\hat{\mu}_C - \mu)^2]$. Recall that

$$\widetilde{h}(X',Y') = \mathbb{E}[f(X',Y',Z)|\sigma(X',Y')].$$

Then,

$$\hat{\mu}_C - \mu = \frac{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}(h(X_i',Y_j') - \mu)k_x(X_i')k_y(Y_j')}{\sum_{i=1}^{\beta_x C}\sum_{j=1}^{\beta_y C}k_x(X_i')k_y(Y_j')}.$$

Let us denote the numerator of the above expression by A_C and denominator by B_C. Then it is easy to through expanding that we have

$$\mathbb{E}\left[\left(\frac{A_C}{\beta_x\beta_y C^2}\right)^2\right] = \frac{\mathbb{E}[(h(X_1',Y_1') - \mu)^2 k_x^2(X_1')k_y(Y_1')]}{\beta_x\beta_y C^2} + \frac{(\beta_x C - 1)M_x}{\beta_x\beta_y C^2} + \frac{(\beta_y C - 1)M_y}{\beta_x\beta_y C^2}.$$

Let the RHS of the above equation be represented by M_C. Using Theorem 12.6 from van der Vaart (1998), we have that

$$\sqrt{C}\frac{A_C}{\sqrt{M_C}\beta_x\beta_y C^2} \Rightarrow Z, \quad \text{as } C \to \infty,$$

where Z is standard normal rv. Here we use '$\Rightarrow$' to represent weak convergence. Also by SLLN we have that B_C satisfies (8). Thus, under suitable uniform integrability conditions we have

$$\lim_{C\to\infty} C\mathbb{E}[(\hat{\mu}_C - \mu)^2] = \frac{M_x}{\beta_y} + \frac{M_y}{\beta_x}$$

and the form of the function $\widetilde{g}$ in (2) follows.

3.2 Optimal Choice of φ_x and φ_y

Next we wish to study the impact of the first stage distribution φ_x and φ_y on the mean square error. To this end we tag the function $\widetilde{g}$ and constants M_x and M_y with φ to explicitly show this dependence. Thus, $\widetilde{g}_\varphi$ represents the asymptotic mean square error. We can now state the following result.

Proposition 2 *For the class of φ_x and φ_y for which (2) holds, we have*

$$\widetilde{g}_\varphi(\beta,\beta_x,\beta_y) \geq \widetilde{g}_{\varphi^*}(\beta,\beta_x,\beta_y),$$

for all $\beta,\beta_x,\beta_y \geq 0$, where

$$\varphi_x^*(\cdot) = K_x|\phi(\cdot) - \mu|\pi_x(\cdot) \qquad (9)$$
$$\varphi_y^*(\cdot) = K_y|\phi(\cdot) - \mu|\pi_y(\cdot) \qquad (10)$$

and K_x and K_y are normalizing constants.

To prove the above proposition we note that to minimize $\widetilde{g}_\varphi$ over φ_x and φ_y is equivalent to minimizing $M_{x,\varphi}$ over φ_x and $M_{y,\varphi}$ over φ_y. We first consider M_x. Using the definition, we have

$$M_{x,\varphi} = \mathbb{E}[(\phi(X') - \mu)^2 k_x^2(X')].$$

Using Jensen's inequality we have

$$M_{x,\varphi} \geq \left(\mathbb{E}\left[|\phi(X') - \mu|k_x(X')\right]\right)^2 = \left(\mathbb{E}\left[|\phi(X) - \mu|\right]\right)^2,$$

where X is rv with probability density function π_x. It can be easily verified that M_{x,φ^*} where φ_x^* is as defined in (9) achieves this lower bound. We can prove the optimality of φ_y^* in an analogous manner.

It is noteworthy that Hesterberg (1995) showed that a similar density function optimal in one-dimensional setting for conducting certain weighted importance sampling.

4 CONCLUSION AND FURTHER RESEARCH DIRECTIONS

In this paper we outlined a methodology for optimal allocation of resources when sampling for input distribution involved two stages. We illustrated our results in a simple setting. In our ongoing research we generalize this analysis to more complex and realistic performance measures.

In the SIRT settings, we identified the optimal first stage distribution. An interesting research direction may be to develop approximations to this to improve performance of the SIR algorithm. One potential application of SIR may be to efficiently generate samples of some practically important random variates, e.g., those with Normal or Gamma distribution at computationally cheaper rates compared to existing algorithms.

REFERENCES

Geweke, J. 1989. Bayesian inference in econometirc models using Monte Carlo integration. *Econometrica* 57:1317–1339.

Hesterberg, T. 1995. Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37:185–194.

Rubin, D. B. 1988. Using the SIR algorithm to simulate posterior distributions. In *Bayseyian Statitics 3*, ed. J. Bernando, M. Degroot, D. Lindley, and A. Smith. Oxford University Press, Inc.

Smith, A. F., and Gelfand A. E. 1992. Bayesian statistics without tears: a sampling-resampling perspective. *American Statistician* 46:84–88.

van der Vaart, A. W. 1997. *Asymptotic statistics*. Cambridge University Press.

AUTHOR BIOGRAPHIES

ACHAL BASSAMBOO is Assistant Professor in managerial economics and decision sciences at Kellogg School of Management. His research interests are stochastic systems, revenue management, and rare event simulation. His e-mail address is <a-bassamboo@northwestern.edu>.

SANDEEP JUNEJA is an Associate Professor in the School of Technology and Computer Science at the Tata Institute of Fundamental Research in India. His research interests include performance analysis of stochastic systems via simulation. His e-mail address is <juneja@tifr.res.in> and his web page is <www.tcs.tifr.res.in/~sandeepj>.