Adaptive Importance Sampling

J. Scott Stadler, *Student Member, IEEE,* and Sumit Roy, *Member, IEEE*

***Abstract*— Importance Sampling (IS) has been shown to be an effective method for reducing the amount of computer time necessary to accurately predict the probability of bit error (p_e) of high performance communication systems. Its widespread acceptance, however, has been limited by the complexity of selecting the IS density and determining its performance. This work makes a contribution to the theory and practice of *adaptive* importance sampling by using a class of algorithms that adapt the IS density to the system of interest during the course of the simulation. This approach enjoys the advantage of removing the burden of selecting the IS density from the system designer. The performance of two such algorithms is investigated for both linear and nonlinear systems operating in Gaussian noise. In addition, the algorithms are shown to converge to the optimum IIS density for the special case of a linear system with Gaussian noise.**

I. Introduction

Monte Carlo (MC) simulation is often used to determine the probability of bit error (p_e) in digital communication systems when its calculation is not analytically tractable. If, however, the p_e is very low (the case of interest in high performance communication systems), the amount of computer time necessary to perform the simulation can become formidably large. Importance Sampling (IS) has been used [6], [16] to reduce the simulation time of low probability events by modifying or changing the underlying probability density function (pdf) so that the event of interest is more likely to occur. The ratio of the variance of an appropriate unbiased estimate of p_e obtained from the IS density to the MC estimator of p_e serves as an indicator of the figure of merit for the IS scheme.

The choice of the IS density is, quite naturally, critical to the success of the simulation. The optimal density is well known, but it is a function of p_e and therefore cannot be used [6]. Conventional IS (CIS) uses a density that is obtained by simply increasing the variance of the underlying density [16]. The improvement in performance obtained with this technique is limited by the memory length of the system, which makes it impractical for most systems of interest. Improved Importance Sampling (IIS) uses a mean translation of the underlying density which overcomes the effects of memory [9]. In addition, the use of the tail of several different pdf's has been explored in [2], [3], and [14].

Although many IS densities have been proposed, their performance varies from system to system. This problem is further complicated by the fact that determining the suitability of an IS density to a particular communications system is similar in complexity to finding p_e itself. These problems have prevented IS from gaining wide-scale acceptance despite its promise of decreasing simulation times by several orders of magnitude.

This paper makes a contribution to the theory and practice of *adaptive* IS that has attracted recent attention [1], [4], [5] due to its potential for removing the burden of selecting a good IS density (which, as noted above, is often system-specific) from the system designer. The key to this technique is the recognition that the subset of the simulation samples that yield an error event are distributed according to the (unknown) unconstrained optimal IS density. These samples may therefore be used to estimate properties of the unconstrained optimal IS density and iteratively render the IS density closer to optimal in the sense that the measured properties (from the current simulation) are made to match those of the unconstrained optimal density. This opens a wide range of possibilities for adaptation rules, since the possible properties of interest range from the simple (e.g., the mean of the IS density) to the complex (e.g., the complete IS density). This approach has the advantage that the mechanics of the simulation remain the same for any system. This is extremely important for investigations into the sensitivity of p_e to various system parameters that is usually determined by performing a series of simulations with perturbed parameters. Other recent advances in the automatic selection of the IS density appear in [15].

The remainder of the paper is organized as follows. The next section briefly reviews relevant results on IS approaches that form a benchmark for comparison of our work. Section III is intended as a self-contained exposition on Adaptive IS (AIS) techniques—both parametric and nonparametric; in addition, results concerning the optimality of two parametric algorithms for the case of a linear system with Gaussian noise are derived. We note that the developments in this section closely parallel the presentation in [9]. Section IV presents an algorithm for parametric AIS, and Section V contains performance results of the AIS algorithms, followed by a concluding discussion.

II. Preliminaries

A generic system of interest is depicted in Fig. 1. The received data vector $\boldsymbol{r}$, consisting of the signal $\boldsymbol{s}$ corrupted by additive noise $\boldsymbol{n}$, is processed by the mapping $g(\cdot)\colon \mathcal{R}^M \mapsto \mathcal{R}^1$ and then compared against a threshold. The noise has a pdf $f_{\boldsymbol{n}}(\boldsymbol{n})$ and the memory length of the system is M. This is the same model presented in [9] using a slightly different notation.

Manuscript received February 19, 1992; revised August 1, 1992.

The authors are with the Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104.

IEEE Log Number 9207829.

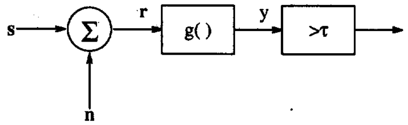


Fig. 1. Simulation model of a general system.

The probability of error can then be expressed as

$$p_e = \int_{\mathcal{R}^M} I_\tau(g(\boldsymbol{s}+\boldsymbol{n}))f_{\boldsymbol{n}}(\boldsymbol{n})f_{\boldsymbol{s}}(\boldsymbol{s})\,d\boldsymbol{s}\,d\boldsymbol{n} \tag{1}$$

where $I_\tau(x)$ is the indicator function for the set $\{x : x > \tau\}$, and $f_{\boldsymbol{s}}(\boldsymbol{s})$ is the pdf of the signal vector. Now if each of the $J = 2^{M-1}$ ISI patterns are equally likely, then (1) can be expressed as

$$p_e = \frac{1}{J}\sum_{j=1}^{J} p_e(j)$$

where

$$p_e(j) = \int_{\mathcal{R}^M} I_\tau(g(\boldsymbol{s}(j)+\boldsymbol{n}))f_{\boldsymbol{n}}(\boldsymbol{n})\,d\boldsymbol{n} \tag{2}$$

and $s(j)$ is the jth ISI pattern. In this manner, the simulation can be divided into J subsimulations, each one corresponding to a different ISI pattern. The remainder of this paper will be concerned with the evaluation of $p_e(j)$ since, clearly, p_e is obtained by repeated evaluation of $p_e(j)$.

The IS estimate of $p_e(j)$ is given by

$$\hat{p}_e(j) = \frac{J}{N^*}\sum_{i=1}^{N^*/J} I_\tau(g(\boldsymbol{s}(j)+\boldsymbol{n}_i))w(\boldsymbol{n}_i) \tag{3}$$

where N^* is the total number of simulation samples, chosen such that N^*/J is an integer, and

$$w(\boldsymbol{n}) = \frac{f_{\boldsymbol{n}}(\boldsymbol{n})}{f_{\boldsymbol{n}^*}(\boldsymbol{n})}$$

is the IS weighting function when $f_{\boldsymbol{n}^*}(\boldsymbol{n})$ is the IS density. The variance of the resulting estimate $\hat{p}_e$ was shown in [9] to be

$$\mathrm{Var}\{\hat{p}_e\} = \frac{1}{JN^*}\sum_{j=1}^{J}\int_{\mathcal{R}^M} I_\tau(g(\boldsymbol{s}(j)+\boldsymbol{n}))f_{\boldsymbol{n}}(\boldsymbol{n}) [w(\boldsymbol{n}) - p_e(j)]\,d\boldsymbol{n} \tag{4}$$

which can be estimated from simulation experiments as

$$\mathrm{Var}\{\hat{p}_e\} = \frac{1}{N^{*2}}\sum_{j=1}^{J}\sum_{i=1}^{N^*/J} I_\tau(g(\boldsymbol{s}(j)+\boldsymbol{n}_i))w^2(\boldsymbol{n}_i) - \frac{\hat{p}_e(j)^2}{N^*}. \tag{5}$$

Although (5) is of little practical use in evaluating the effectiveness of a proposed IS density, accurate estimates are nevertheless obtained if it is known *a priori* that $f_{\boldsymbol{n}^*}(\boldsymbol{n})$ is a “*good*” IS density.

The usual approach to IS is to choose a parametric form of $f_{\boldsymbol{n}^*}(\boldsymbol{n}|\boldsymbol{\Theta})$ and then to minimize (4) with respect to the parameter vector $\boldsymbol{\Theta}$. Unfortunately, for all but the simplest systems, the optimization problem is usually of the same complexity as calculating the p_e directly. The following sections describe an *adaptive* approach that obviates a direct optimization and, instead, approaches the optimal IS density iteratively.

III. Overview of Adaptive Importance Sampling

This section introduces the notion of adaptive importance samplig (AIS) algorithms and describes how they can be used to evaluate integrals of the form (2). These algorithms appear to have been originally developed and applied in the area of systems reliability to determine the failure rate of mechanical structures [4], [5], [7], but they can be applied to the evaluation of p_e in digital communications systems as well.

In order to obtain a fundamental understanding of AIS techniques, it is necessary to first look at the unconstrained optimal IS density. It is well known that if

$$f_{\boldsymbol{n}^*}(\boldsymbol{n}) = f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n}) = \frac{f_{\boldsymbol{n}}(\boldsymbol{n})}{p_e(j)} I_\tau(g(\boldsymbol{s}(j)+\boldsymbol{n})) \tag{6}$$

then the variance of $\hat{p}_e(j)$ is zero, implying that a perfect estimate of $p_e(j)$ can be obtained with a single simulation sample. It is obvious, however, that the unconstrained optimal IS density cannot be directly used in a simulation because it is dependent on the parameter of interest p_e which is unknown *a priori*. The following development is intended to show that although initially the optimal IS density may be unknown, useful estimates of its properties can be obtained (and subsequently used) in the course of the simulation.

AIS relies on the fact that the pdf of the noise samples conditioned on the failure domain are distributed according to the unconstrained optimal IS density, i.e.,

$$f_{\boldsymbol{n}}(\boldsymbol{n}|g(\boldsymbol{s}(j)+\boldsymbol{n}) > \tau) = f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n}) = \frac{f_{\boldsymbol{n}}(\boldsymbol{n})}{p_e(j)} I_\tau(g(\boldsymbol{s}(j)+\boldsymbol{n})).$$

This means that the set of samples $\boldsymbol{n}_i$ obtained during the simulation that result in errors are distributed according to $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$, and can therefore be used to estimate its properties. The properties of interest can range from a simple parameter such as the mean $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$ to the estimation of the entire pdf.

An AIS algorithm consists of several short simulation runs. During each run, $\hat{p}_e(j)$ and the properties of $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$ are estimated. The IS density is then modified such that its properties match the estimated properties of $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$ for use in the subsequent simulation run. In this manner, as successive simulation runs are performed, the IS density becomes more like the unconstrained optimal density, and $\hat{p}_e(j)$ becomes more accurate.

In Section III, an overview of the AIS algorithms described in the systems reliability literature is given. Our contribution to this section is the analysis which reveals both the optimality (at convergence) and the deficiencies of the various algorithms. A new, modified AIS algorithm is then developed in Section IV that alleviates these deficiencies. A catalog of representative experimental results using both linear systems with memory and nonlinear systems is presented in Section V.

A. Parametric AIS

In the parametric case, a finite vector of parameters that characterize $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$ is estimated. These types of algorithms are exemplified by [4], [11] where the conditional mean and covariance are estimated in the former, and the conditional mode in the latter. It will be shown that for the case of a linear system with Gaussian noise, both algorithms will yield an IS density that corresponds to the optimal IIS density derived in [9].

1) Conditional Mean and Covariance Algorithm: In [4], the conditional mean and covariance were estimated and used as the parameters of a Gaussian IS density. The algorithm is as follows.

1) An initial shift vector μ_{init} is found (such as by using a sensitivity analysis—see [4] for details).

2) Run a simulation using the noise density shifted to μ_{unit} as the IS density.

3) Using the samples that resulted in errors, estimate the conditional mean $\hat{\mu}$, and the conditional covariance matrix $\widehat{\boldsymbol{K_n}}$.

4) Run a simulation using $\hat{\mu}$ and $\widehat{\boldsymbol{K_n}}$ as the parameters of a Gaussian IS density and calculate $\hat{p}_e$.

5) Repeat steps 3) and 4) until the desired accuracy is achieved.

In [4], the form of the estimates for $\hat{\mu}$ and $\widehat{\boldsymbol{K_n}}$ were not explicitly specified. It is worthwhile noting that if standard IS mean and covariance estimates are used, they yield biased estimates of μ and $\boldsymbol{K_n}$, respectively. To see this in the case of the conditional mean, note that

$$\begin{aligned}\mathcal{E}\{\hat{\mu}\} &= \mathcal{E}_*\left\{\frac{1}{p\sum_{i=1}^{p}\boldsymbol{n}_i w(\boldsymbol{n}_i)}\right\}\\ &= \int_{\boldsymbol{E}} \boldsymbol{n}_i w(\boldsymbol{n}_i)\frac{f^*_{\boldsymbol{n}}(\boldsymbol{n}_i)}{p^*_e(j)}d\boldsymbol{n}_i\\ &= \frac{p_e(j)}{p^*_e(j)}\mathcal{E}\{\boldsymbol{n}|\boldsymbol{n}\in\boldsymbol{E}\} = \frac{p_e(j)}{p^*_e(j)}\mu\end{aligned}$$

where $\boldsymbol{E} = \{\boldsymbol{n} : g(\boldsymbol{s}(j)+\boldsymbol{n}) > \tau\}$ is the error region, $\boldsymbol{n}_i$, $i = 1, \cdots, p$ are the p samples that caused errors, $\mathcal{E}_*$ denotes expectation with respect to the IS density, and $p^*_e(j)$ is the probability of error for the jth ISI pattern when the noise is distributed according to the IS density. Consequently, $\hat{\mu}$ is not a consistent estimate of μ. This has some serious consequences for the convergence of the algorithm as μ cannot be a stable point. This can be seen by considering an initial shift that is close to the optimum shift. In this case, $p^*_e(j) \gg p_e(j)$ because the error rate of the system with the IS density will be much higher than the error rate of the system with the underlying density. This means that $\hat{\mu} \ll \mu_{\text{opt}}$ and the algorithm will *diverge* from the optimal solution. In Section IV, a consistent estimate of the conditional mean will be derived and used to develop a new AIS algorithm.

Although setting the mean and covariance of the IS density equal to the mean and covariance of $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$ is a reasonable criterion in its own right, it will now be shown that for the case of a linear system with Gaussian noise, the mean shift obtained in this manner corresponds to that of the optimal IIS density.

First, consider a system with no memory ($M = 1$); the

$$\mathcal{E}\{n|n \in \boldsymbol{E}\} = \frac{1}{\displaystyle\int_\tau^\infty \frac{1}{\sigma\sqrt{2\pi}}e^{-(\boldsymbol{n}^2/2\sigma^2)}\,dn}\int_\tau^\infty n\frac{1}{\sigma\sqrt{2\pi}}e^{-(n^2/2\sigma^2)}\,dn = \frac{\dfrac{\sigma}{\sqrt{2\pi}}e^{-(\tau^2/2\sigma^2)}}{Q\left(\dfrac{\tau}{\sigma}\right)} \tag{7}$$

where

$$Q(t) = \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-(t^2/2)}\,dt.$$

Using the approximation

$$Q(t) \approx \frac{1}{t\sqrt{2\pi}}e^{-(t^2/2)} \qquad \text{for large } t \tag{8}$$

yields

$$\mathcal{E}\{n|n \in \boldsymbol{E}\} \approx \tau$$

which corresponds to the optimum shift derived in [9, equation 4.22]. A similar calculation for the variance can be performed by evaluating

$$\mathcal{E}\{n^2|n \in \boldsymbol{E}\} = \frac{1}{\displaystyle\int_\tau^\infty \frac{1}{\sigma\sqrt{2\pi}}e^{-(n^2/2\sigma^2)}\,dn}\cdot\int_\tau^\infty n^2\frac{1}{\sigma\sqrt{2\pi}}e^{-(n^2/2\sigma^2)}\,dn. \tag{9}$$

Integrating once by parts reduces the integral to

$$\mathcal{E}\{n^2|n \in \boldsymbol{E}\} = \frac{\dfrac{1}{\sqrt{2\pi}}\sigma\tau e^{-(\tau^2/2\sigma^2)}}{Q\left(\dfrac{\tau}{\sigma}\right)} + \sigma^2.$$

Using the approximation for $Q(t)$ and $\mathcal{E}\{n|n \in \boldsymbol{E}\}$ from above yields

$$\text{Var}\,\{n|n \in \boldsymbol{E}\} \approx \sigma^2.$$

This agrees with the conclusions in [13] and [14] that, for a linear system, the mean should be shifted and the variance left unchanged.

Now if the output of a finite impulse response (FIR) linear system is modeled as

$$y = \sum_{i=1}^{M} b_i n_i$$

where b_i, $i = 1, \cdots, M$ are the coefficients, and we let τ_j be the threshold corresponding to the jth ISI realization as in [9] (note: the notation used here differs slightly from that in [9] where b_{M-i} is the coefficient corresponding to n_i), then

$$\mathcal{E}\{n_l|\boldsymbol{n} \in \boldsymbol{E}\} = \frac{1}{\displaystyle\int_{\tau_j}^{\infty} \frac{\exp - \dfrac{n^2}{2\sigma^2 \sum_{i=1}^{M} b_i^2}}{\sqrt{2\pi\sigma^2 \sum_{i=1}^{M} b_i^2}}\, dn} \cdot \int_{-\infty}^{\infty} n_l \frac{1}{\sigma\sqrt{2\pi}} e^{-(n_l^2/2\sigma^2)} \cdot \int_{\tau_j - b_l n_l}^{\infty} \frac{1}{\sigma\sqrt{2\pi\tilde{b}^2}} e^{-(n_*^2/2\sigma^2\tilde{b}^2)}\, dn_* dn_l \tag{10}$$

where

$$n_* = \sum_{\substack{i=1\\ i\neq l}}^{M} b_i n_i \sim N(0, \sigma^2\tilde{b}^2), \quad \text{and} \quad \tilde{b}^2 = \sum_{\substack{i=1\\ i\neq l}}^{M} b_i^2.$$

The variable n_* is equivalent to $\tilde{b}\tilde{n}$ where $\tilde{n} \sim N(0, \sigma^2)$. Therefore, (10) can be written as

$$\begin{aligned}\mathcal{E}\{n_l|\boldsymbol{n} \in \boldsymbol{E}\} &= \frac{1}{\mathcal{Q}\left(\dfrac{\tau_j}{\sigma\sqrt{b_l^2 + \tilde{b}^2}}\right)} \int_{-\infty}^{\infty} n_l \frac{1}{\sigma\sqrt{2\pi}} e^{-(n_l^2/2\sigma^2)} \\ &\quad \cdot \int_{\tau_j - b_l n_l/\tilde{b}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(\tilde{n}^2/2\sigma^2)}\, d\tilde{n}\, dn_l \\ &= \frac{1}{\mathcal{Q}\left(\dfrac{\tau_j}{\sigma\sqrt{b_l^2 + \tilde{b}^2}}\right)} \cdot \int_{-\infty}^{\infty} n_l \frac{1}{\sigma\sqrt{2\pi}} e^{-(n_l^2/2\sigma^2)} \mathcal{Q}\left(\frac{\tau_j - b_l n_l}{\sigma\tilde{b}}\right) dn_l \\ &= \frac{1}{\mathcal{Q}\left(\dfrac{\tau_j}{\sigma\sqrt{b_l^2 + \tilde{b}^2}}\right)} \frac{b_l\sigma}{\sqrt{2\pi(b_l^2 + \tilde{b}^2)}} \cdot e^{-(\tau_j^2/2\sigma^2(b_l^2+\tilde{b}^2))}.\end{aligned}$$

Now using the approximation for $\mathcal{Q}(t)$ yield

$$\mathcal{E}\{n_l|\boldsymbol{n} \in \boldsymbol{E}\} \approx \frac{\tau_j b_l}{\sum_{i=1}^{M} b_i^2}, \tag{11}$$

which agrees with the shift parameters of [9, equations (4.35) and (4.41)]. Therefore, for the case of a linear system with Gaussian noise, the conditional mean algorithm yields the optimal IIS density.

2) Conditional Mode Algorithm: Another parametric approach is described in [11] where, instead of using an estimate of the conditional mean of $\boldsymbol{n}$, an estimate of the *mode* of the conditional density is used, i.e., the mean of $f_{\boldsymbol{n}}(\boldsymbol{n})$ is shifted to

$$\hat{\mu} = \arg\max_i f_{\boldsymbol{n}}(\boldsymbol{n}_i)$$

where arg max f is the value of $\boldsymbol{n}_i$ that maximizes $f(\cdot)$. The primary intuition behind this algorithm is that under the assumption that the maximization has a unique solution (as is the case for a linear system with Gaussian noise), the IS density will be centered at the point in the failure domain where an error is most likely to occur (and, consequently, samples from this region are most likely to cause errors). The algorithm is as follows.

1) Initialize by finding an $\boldsymbol{n}$ that causes an error. This will be the mean shift of the initial IS density.

2) Run a short simulation using the IS density from 1) and calculate $\hat{\mu} = \arg\max_i f_{\boldsymbol{n}}(\boldsymbol{n}_i)$.

3) Run a simulation with $\hat{\mu}$ as the shift parameter of the IS density and calculate $\hat{p}_e$.

It will now be shown that for a linear system with Gaussian noise, shifting the IS density to the mode of $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$ also yields the optimal IIS density. This result establishes the asymptotic equivalence of the algorithms proposed in [4] and [11] for this case.

The mode can be found by considering the constrained optimization problem

$$\max_{\boldsymbol{n}\in\boldsymbol{E}} \frac{1}{M\sigma\sqrt{2\pi}} e^{-(\boldsymbol{n}^t\boldsymbol{n}/2(M\sigma)^2)}.$$

The log-concavity of the Guassian density allows the optimization to be expressed as the minimization of a convex function over a convex set which implies that a unique optimum exists [10]. Assuming that the origin is not contained in the error domain, the maximum will belong to the set $\{\boldsymbol{n} : \boldsymbol{n}^t\boldsymbol{b} = \tau\}$, which is the boundary of the error domain. The optimization can proceed using Lagrange multipliers. The objective function can be written as

$$\Phi(\boldsymbol{n}) = \boldsymbol{n}^t\boldsymbol{n} - \lambda(\boldsymbol{b}^t\boldsymbol{n} - \tau).$$

Differentiating with respect to $\boldsymbol{n}$ and setting the result equal to zero yields

$$2\boldsymbol{n} - \lambda\boldsymbol{b} = 0.$$

The constraint requires $\lambda = 2\tau/\boldsymbol{b}^t\boldsymbol{b}$ which gives

$$\arg\max f_{\boldsymbol{n}}(\boldsymbol{n}) = \frac{\boldsymbol{b}\tau}{\boldsymbol{b}^t\boldsymbol{b}} \tag{12}$$

which is identical to the optimal shift given by (11). The fact that the optimization can be expressed as the minimization of a convex function over a convex set guarantees the convergence of the conditional mode algorithm if step 2) is repeated in an iterative manner. Note that although the mean and the mode of the Gaussian density are equal, the conditional mean and the conditional mode are not. Consequently, the equality of the shift parameters obtained via the two algorithms is only asymptotic and is due to the high SNR assumption inherent in the use of the approximation for $Q(t)$ in (8).

Parametric AIS algorithms are suitable for many systems for which useful parametric models are available, but there are instances where such parametric techniques are not appropriate. For example, the class of IS distributions obtained by increasing the variance of the underlying distribution performs poorly for linear systems with memory [16]. This failure of the parametric approach is due to the inability of the scaled IS density to approximate the unconstrained optimal IS density over large areas of the error region. Therefore, if it is not known *a priori* that a particular parametric class of IS densities is appropriate for a given system, a nonparametric approach may be more useful. An overview of nonparametric IS techniques is given below.

B. Nonparametric AIS

In the nonparametric case, instead of trying to estimate properties of the unconstrained optimal IS density, the entire density is estimated. The density estimate is then used as the IS density in a simulation to estimate $\hat{p}_e$. This approach was suggested in [1] where the kernel density estimate with a simple rectangular kernel was used to estimate the failure probability of a system with no memory. The kernel density estimate in this case becomes

$$\hat{f}_n(x) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{w} K\left(\frac{x - n_i}{w}\right) \tag{13}$$

where n_i is the ith sample that caused an error, and w is a constant [18]. The kernel $K(\cdot)$ was chosen to be

$$K(y) = \begin{cases} \frac{1}{2} & \text{if } |y| < 1 \\ 0 & \text{otherwise} \end{cases}$$

which estimates the density by constructing a histogram with a bin width of w. The algorithm is as follows.

1) Run a simulation and form an estimate of $f_{n^*_{\text{opt}}}(n)$.

2) Run a simulation with $\hat{f}_{n^*_{\text{opt}}}(n)$ obtained above as the IS density, and calculate $\hat{p}_e$.

Although good numerical results were obtained for the examples considered, there are several potential problems with the algorithm that require attention. First, the number of errors that must be observed to obtain a good estimate of $f_{n^*_{\text{opt}}}(n)$ using the histogram kernel is much larger than the number of samples that are needed to estimate p_e. In addition, the choice of a kernel with finite support will lead to a large variance, as there are sections of the error domain where $w(n)$ will be unbounded. The authors modified the above algorithm by using a Gaussian kernel, and an IS simulation for the estimate of $f_{\boldsymbol{n}^*_{\text{opt}}}(\boldsymbol{n})$. Good results were obtained for systems with short memory lengths ($M < 10$), but the slow convergence of the kernel estimate for pdf's of high dimensionality prevented the algorithm from being useful for systems with large memory lengths.

Another nonparametric approach is described in [7], with associated algorithms appearing in [5] and [11]. In this case, several different modes representing regions of importance in the error domain are identified during the course of a preliminary simulation. The IS density is then formed by replicating the original density at each mode in proportion to the importance of that mode. The IS density is thus defined as

$$f_{\boldsymbol{n}^*}(\boldsymbol{n}) = \frac{1}{\sum_{i=1}^{k} f_{\boldsymbol{n}}(\boldsymbol{m}_i)} \sum_{i=1}^{k} f_{\boldsymbol{n}}(\boldsymbol{m}_i) f_{\boldsymbol{n}}(\boldsymbol{n} - \boldsymbol{m}_i) \tag{14}$$

where $\boldsymbol{m}_i$ is the point chosen in the preliminary simulation to represent the ith mode. The algorithm is as follows.

1) Choose a point $\boldsymbol{n}_{\text{init}}$ in the error domain, set $f_{\boldsymbol{n}^*}(\boldsymbol{n}) = f_{\boldsymbol{n}}(\boldsymbol{n} - \boldsymbol{n}_{\text{init}})$, and choose a cluster radius d.

2) Perform an IS simulation using the noise density determined in the previous step and save the points $\boldsymbol{n}_i$ that cause errors to occur.

3) Calculate $\hat{p}_e$.

4) Form $f_{\boldsymbol{n}^*}(\boldsymbol{n})$ for the next iteration in the following manner.

a) Set $\boldsymbol{m}_1 = \arg\max_i f_{\boldsymbol{n}}(\boldsymbol{n}_i)$.

b) Delete all points $\boldsymbol{n}_i$, whose distance to $\boldsymbol{m}_1$ is less than d.

c) In a similar manner, select $\boldsymbol{m}_2$ from the remaining $\boldsymbol{n}_i$.

d) Continue this process until all $\boldsymbol{n}_i$ have been deleted.

5) Decrease d and repeat steps 2)–4) until the sample variance of $\hat{p}_e$ has been reduced to an acceptable level.

The important modes are identified by performing the first few iterations with a large variance to obtain points throughout the error domain. With each iteration, the variance of the IS density and the cluster radius d is decreased. As the number of modes increase and d gets infinitesimally small, (14) converges to the unconstrained optimal IS density [7].

The algorithms described above represent only a small fraction of the possibilities that can be developed in the framework outlined in the Introduction. The next section analyzes the convergence properties of a modified parametric AIS algorithm.

IV. A New AIS Algorithm

In this section, a new parametric AIS algorithm that is a modified version of the one in [4] is developed. This is based on the estimate of

$$\mu = \mathcal{E}\{\boldsymbol{n} | \boldsymbol{n} \in \boldsymbol{E}\} = \int_{\boldsymbol{E}} \boldsymbol{n} \frac{f_{\boldsymbol{n}}(\boldsymbol{n})}{p_e(j)} d\boldsymbol{n} \tag{15}$$

as the shift parameter for the IS density, which can be developed by rewriting (15) as

$$\begin{aligned} \mathcal{E}\{\boldsymbol{n} | \boldsymbol{n} \in \boldsymbol{E}\} &= \frac{p_e^*(j)}{p_e(j)} \int_{\boldsymbol{E}} \boldsymbol{n} \frac{f_{\boldsymbol{n}}(\boldsymbol{n})}{f_{\boldsymbol{n}}^*(\boldsymbol{n})} \frac{f_{\boldsymbol{n}}^*(\boldsymbol{n})}{p_e^*(j)} d\boldsymbol{n} \\ &= \frac{p_e^*(j)}{p_e(j)} \mathcal{E}_*\{\boldsymbol{n} w(\boldsymbol{n}) | \boldsymbol{n} \in \boldsymbol{E}\} = \frac{p_e^*(j)}{p_e(j)} \bar{\boldsymbol{n}} \end{aligned} \tag{16}$$

where $p_e^*(j)$ is the probability of error when the IS density is used and $\bar{\boldsymbol{n}} = \mathcal{E}_*\{\boldsymbol{n}w(\boldsymbol{n})|\boldsymbol{n} \in \boldsymbol{E}\}$. Replacing $p_e(j)$ with $\hat{p}_e(j)$ from (3), $p_e^*(j)$ with its MC estimate, $\hat{p}_e^*(j) = pJ/N^*$, and $\bar{\boldsymbol{n}}$ with its MC estimate

$$\hat{\bar{\boldsymbol{n}}} = \frac{1}{p}\sum_{i=1}^{p} \boldsymbol{n}_i w(\boldsymbol{n}_i)$$

yields

$$\hat{\mu} = \frac{1}{\sum_{i=1}^{p} w(\boldsymbol{n}_i)} \sum_{i=1}^{p} \boldsymbol{n}_i w(\boldsymbol{n}_i) \tag{17}$$

which is a (weakly) consistent estimator of μ. To establish this, first consider the component estimators in $\hat{\mu}$. Individually, $\hat{p}_e(j), \hat{\bar{\boldsymbol{n}}}$, and $\hat{p}_e^*(j)$ are consistent because they are unbiased (i.e., they are either MC or IS etimates) and their variances approach zero as the appropriate number of samples approaches infinity [8]. The variance of $\hat{p}_e(j) \approx c/N^*$, where $c < \infty$ is a constant determined from (4). Furthermore, invoking the well-known fact that the variance of the average of n i.i.d. random variables $\approx o(1/n)$, it follows that the variances of $\hat{p}_e^*(j)$ and $\hat{\bar{\boldsymbol{n}}}$ decrease as $1/N^*$ and $1/p$, respectively. This establishes consistency of the component estimates in $\hat{\mu}$. Now recalling that if a_i and b_i are sequences of random variables that satisfy

$$plim_{i\to\infty} a_i = a, \qquad plim_{i\to\infty} b_i = b$$

it follows that [12]

$$plim_{i\to\infty} a_i b_i = ab, \qquad plim_{i\to\infty} \frac{a_i}{b_i} = \frac{a}{b}$$

where $plim_{i\to\infty} x_i = x$ is short-hand for the usual weak convergence (or convergence in probability), i.e.,

$$\lim_{i\to\infty} x_i = x \equiv \lim_{i\to\infty} \mathcal{P}\{|x_i - x| < \epsilon\} = 1.$$

Therefore,

$$plim_{p\to\infty} \frac{\hat{p}_e^*(j)\hat{\bar{\boldsymbol{n}}}}{\hat{p}_e(j)} = \frac{p_e^*(j)\bar{\boldsymbol{n}}}{p_e(j)} = \mu$$

implying that $\hat{\mu}$ is a (weakly) consistent estimator of μ.

The AIS algorithm can now be described in the following three steps.

1) Initialize by finding an $\boldsymbol{n}$ that causes an error. This will be the mean shift of the initial IS density.

2) Run a short simulation using the IS density from the previous step and calculate $\hat{\mu}$ using (17).

3) Run a simulation with $\hat{\mu}$ as the shift parameter of the IS density and calculate $\hat{p}_e(j)$ using (3).

Note that if the initial mean shift is far from the optimal, it may be necessary to repeat step 3) several times before convergence of $\hat{\mu}$ is achieved. This can be done by calculating $\hat{\mu}$ using the samples obtained in step 3) and then using this estimate as the shift for the next iteration.

TABLE I
Performance of Conditional Mean Algorithm for System With no Signal, $M = 1$

N^*	$\hat{p}_e$	$\hat{\mu}$	$\widehat{var}$	var	$\gamma_{MC/AIS}$
1000	$1.03\text{x}10^{-5}$	4.48	$6.4\text{x}10^{-13}$	$5.5\text{x}10^{-13}$	18,361
2000	$9.85\text{x}10^{-6}$	4.45	$2.4\text{x}10^{-13}$	$2.4\text{x}10^{-13}$	20,735
3000	$9.49\text{x}10^{-6}$	4.48	$1.6\text{x}10^{-13}$	$1.6\text{x}10^{-13}$	21,439
4000	$9.49\text{x}10^{-6}$	4.46	$1.2\text{x}10^{-13}$	$1.2\text{x}10^{-13}$	21,862
5000	$1.02\text{x}10^{-5}$	4.49	$7.9\text{x}10^{-14}$	$9.1\text{x}10^{-14}$	22,007

TABLE II
Performance of Conditional Mode Algorithm for System With no Signal, $M = 1$

N^*	$\hat{p}_e$	$\hat{\mu}$	$\widehat{var}$	var	$\gamma_{MC/AIS}$
1000	$9.43\text{x}10^{-6}$	4.28	$5.5\text{x}10^{-13}$	$5.5\text{x}10^{-13}$	18,366
2000	$1.06\text{x}10^{-5}$	4.42	$2.8\text{x}10^{-13}$	$2.4\text{x}10^{-13}$	20,769
3000	$9.73\text{x}10^{-6}$	4.30	$1.6\text{x}10^{-13}$	$1.6\text{x}10^{-13}$	21,469
4000	$1.04\text{x}10^{-5}$	4.34	$1.3\text{x}10^{-13}$	$1.2\text{x}10^{-13}$	21,915
5000	$9.80\text{x}10^{-6}$	4.28	$9.8\text{x}10^{-14}$	$9.1\text{x}10^{-14}$	22,039

V. Performance Evaluation

This section gives experimental results using the conditional mean algorithm described in Section III-A-1 with the estimate of μ from Section IV and the conditional mode algorithm decsribed in Section III-A-2. All the algorithms were tested on a linear system with no signals and no memory (i.e., $g(x) = x$), a linear system with $M = 3$ and ISI, and a nonlinear system with ISI. Note that in the context of the conditional mean algorithm, $\hat{\mu} = 1/\Sigma_{i=1}^{p} w(\boldsymbol{n}_i) \, \Sigma_{i=1}^{p} \boldsymbol{n}_i w(\boldsymbol{n}_i)$, while in the context of the conditional mode algorithm, $\hat{\mu} = \arg\max_i \, f_{\boldsymbol{n}}(\boldsymbol{n}_i)$.

A. Linear Systems

The results for the system with no memory appear in Table I for the conditional mean algorithm, and in Table II for the conditional mode algorithm. For this case, $\tau = 4.2667$ and $n \sim N(0, 1)$ which yields $p_e = 1.0041 \times 10^{-5}$. The total number of samples, N^*, ranged from 1000 to 5000, of which 200 were used to estimate μ. Both algorithms produced an estimate of μ that is close to the optimal shift of $\mu_{\text{opt}} = 4.38$, and an estimate of p_e that is very close to the true value of p_e. In addition, the estimated variance of $\hat{p}_e$ is very close to the true variance which was calculated using [9, eq. (4.20)]

$$\text{Var}\{\hat{p}_e\} \approx \frac{p_e^2}{N^*}\left[\sqrt{2\pi}\frac{\tau^2}{\sigma(\tau+\mu)} e^{(\tau-\mu)^2/2\sigma^2} - 1\right].$$

This agrees with the theoretical results presented in Section III-A, that both the conditional mean and the conditional mode should converge to the optimal shift parameter.

The convergence of $\hat{\mu}$ as a function of N appears in Fig. 2 for the conditional mean algorithm, and in Fig. 3 for the conditional mode algorithm. Similar results have been obtained for systems with large memories (i.e., $M \geq 50$), which implies that the rate of convergence is not sensitive to M. The estimate of μ was very robust with respect to the shift of the initial IS density. Even when the initial shift was much greater than the optimal, $\hat{\mu}$ was closer to the optimal than the initial shift, suggesting that an iterative algorithm would eventually converge to the optimal μ regardless of the initial conditions.

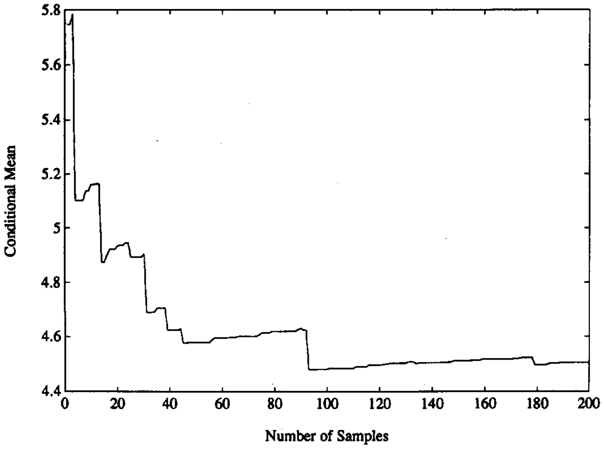


Fig. 2. Convergence of $\hat{\mu}$ as a function of N for the conditional mean algorithm.

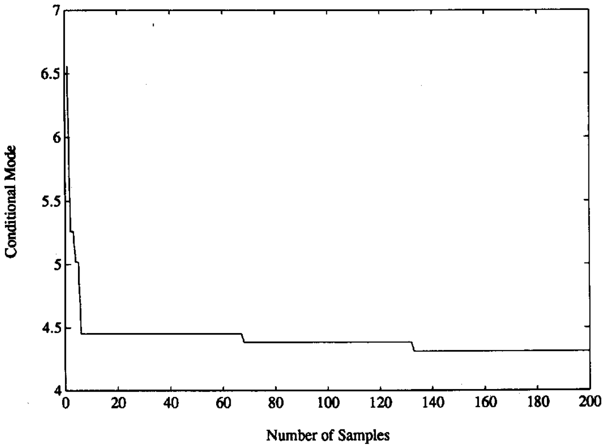


Fig. 3. Convergence of $\hat{\mu}$ as a function of N for the conditional mode algorithm.

TABLE III
PERFORMANCE OF CONDITIONAL MEAN ALGORITHM FOR A SYSTEM WITH ISI, $M = 3$

ISI	$\hat{p}_e(j)$	$p_e(j)$	$\widehat{var}$	$\hat{\mu}$	μ_{opt}
$+A,+A$	3.85x10^{-5}	4.09x10^{-5}	9.3x10^{-12}	[4.0, 0.91, 0.54]	[4.24, 0.85, 0.42]
$+A,-A$	1.98x10^{-7}	2.05x10^{-7}	2.9x10^{-16}	[5.1, 0.81, 0.46]	[4.94, 0.99, 0.49]
$-A,+A$	3.27x10^{-10}	3.02x10^{-10}	8.6x10^{-22}	[6.2, 1.41, 0.49]	[6.04, 1.20, 0.60]
$-A,-A$	1.29x10^{-13}	1.26x10^{-13}	2.0x10^{-28}	[7.3, 1.13, 0.98]	[7.14, 1.42, 0.71]

TABLE IV
PERFORMANCE OF CONDITIONAL MODE ALGORITHM FOR A SYSTEM WITH ISI, $M = 3$

ISI	$\hat{p}_e(j)$	$p_e(j)$	$\widehat{var}$	$\hat{\mu}$	μ_{opt}
$+A,+A$	3.58x10^{-5}	4.09x10^{-5}	1.1x10^{-11}	[3.79, 1.24, 0.71]	[4.24, 0.85, 0.42]
$+A,-A$	2.02x10^{-7}	2.05x10^{-7}	3.8x10^{-16}	[4.96, 1.11, 0.09]	[4.94, 0.99, 0.49]
$-A,+A$	2.99x10^{-10}	3.02x10^{-10}	8.0x10^{-22}	[6.1, 1.09, 0.70]	[6.04, 1.20, 0.60]
$-A,-A$	1.16x10^{-13}	1.26x10^{-13}	1.5x10^{-28}	[7.3, 1.58, 0.56]	[7.14, 1.42, 0.71]

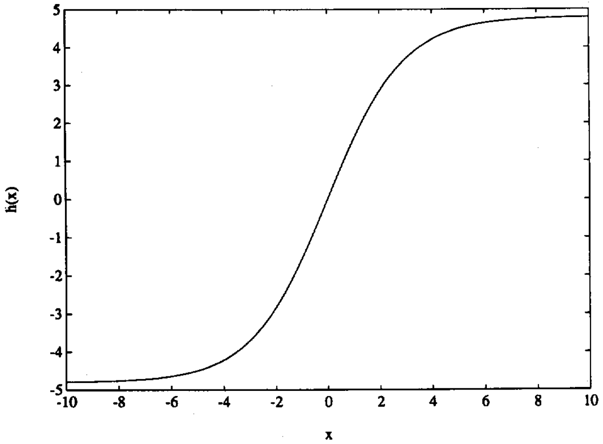


Fig. 4. Saturating nonlinearity.

It is interesting to note that as the number of samples is increased (i.e., as $\hat{p}_e$ becomes more accurate), $\gamma_{MC/AIS}$, the improvement factor of the adaptive scheme over standard MC increases. This is due to the fact that 200 samples were used to find $\hat{\mu}$ in all the cases considered, and as N^* gets larger, the percentage of samples that are used for the adaptation process gets smaller. The improvement ratio for the optimal mean shift is 23,222, which means that the overhead for both of the adaptive algorithms is relatively small.

The results for the system with memory and ISI appear in Tables III and IV. In this case, $M = 3, \boldsymbol{b} = [1.0, 0.2, 0.1]^T, n \sim N(0,1)$, the signal amplitude is $A = 5.77$, and $-A$ is transmitted, which yields $p_e = 1.0265 \times 10^{-5}$. Each ISI pattern was simulated using 200 samples to determine $\hat{\mu}$ and 800 samples to calculate $\hat{p}_e$. The conditional mean algorithm produced an estimate of $\hat{p}_e = 9.67 \times 10^{-6}$ with $\gamma_{MC/AIS} = 1077$, while the conditional mode algorithm produced an estimate of $\hat{p}_e = 9.00 \times 10^{-6}$ with $\gamma_{MC/AIS} = 934$. It is interesting to note that even with a small amount of ISI, p_e is dominated by the worst-case ISI pattern, and that a better improvement ratio could be obtained by dedicating more samples to its simulation. This is evident in the high degree of accuracy that was achieved in the estimation of the ISI patterns with negligible values of $p_e(j)$. The improvement ratio for the conditional mean algorithm could also be improved for this case by recognizing that the optimal shift is $c\boldsymbol{b}$, where $c = \tau_j/\boldsymbol{b}^t\boldsymbol{b}$ was obtained in (11), and then estimating the scalar c instead of the vector parameter μ. While this would provide a variance reduction of $1/M$ in the estimate $\hat{\mu}$ for linear systems with Gaussian noise, we prefer to use the vector version in view of its applicability to nonlinear systems.

B. Nonlinear Systems

This subsection gives an example of the performance of the above algorithms in a nonlinear system with ISI. The system is modeled as a memoryless saturating nonlinearity followed by the linear system described above. The nonlinearity (Fig. 4) is described by $h(x) = (2A/(1+e^{bx})) - A$, with $b = 3.272/A$ chosen such that the nonlinearity is close to linear over $[-0.9A, 0.9A]$. This is meant to model the characteristics of a saturating amplifier. The signal amplitude was $A = 4.8$, where, once again, the transmitted signal was $-A$ and $n \sim N(0,1)$,

TABLE V
CONDITIONAL MEAN ALGORITHM FOR A SYSTEM WITH NONLINEARITY AND ISI, $M = 3$

ISI	$\hat{p}_e(j)$	$p_e(j)$	$\widehat{var}$
$+A,+A$	$3.53\text{x}10^{-5}$	$3.5\text{x}10^{-5}$	$8.4\text{x}10^{-12}$
$+A,-A$	$2.70\text{x}10^{-6}$	$3.0\text{x}10^{-6}$	$5.6\text{x}10^{-14}$
$-A,+A$	$1.97\text{x}10^{-7}$	$2.0\text{x}10^{-7}$	$3.33\text{x}10^{-16}$
$-A,-A$	$1.05\text{x}10^{-8}$	$9.7\text{x}10^{-9}$	$1.9\text{x}10^{-18}$

TABLE VI
CONDITIONAL MODE ALGORITHM FOR A SYSTEM WITH NONLINEARITY AND ISI, $M = 3$

ISI	$\hat{p}_e(j)$	$p_e(j)$	$\widehat{var}$
$+A,+A$	$3.46\text{x}10^{-5}$	$3.5\text{x}10^{-5}$	$1.0\text{x}10^{-11}$
$+A,-A$	$2.68\text{x}10^{-6}$	$3.0\text{x}10^{-6}$	$8.9\text{x}10^{-14}$
$-A,+A$	$1.82\text{x}10^{-7}$	$2.0\text{x}10^{-7}$	$5.0\text{x}10^{-16}$
$-A,-A$	$1.11\text{x}10^{-8}$	$9.7\text{x}10^{-9}$	$2.8\text{x}10^{-18}$

which gives $p_e = 9.59 \times 10^{-6}$, where the $p_e(j)$ were found using IIS with $N^* = 500,000$ to ensure good estimates. The results are presented in Tables V and VI. Each ISI pattern was again simulated using 200 samples to determine $\hat{\mu}$, and 800 samples to calculate $\hat{p}_e(j)$. Despite the nonlinear nature of the system, the conditional mean algorithm provided an estimate of $\hat{p}_e = 9.55 \times 10^{-6}$ with $\gamma_{MC/AIS} = 1115$, and the conditional mode algorithm provided an estimate of $\hat{p}_e = 9.37 \times 10^{-6}$ with $\gamma_{MC/AIS} = 955$ where $\gamma_{MC/AIS}$ was calculated using the variance estimate from (5).

VI. CONCLUDING REMARKS

This paper has presented a parametric AIS algorithm, and demonstrated it in the simulation of linear and nonlinear systems. In addition, strong theoretical justification was given for its use in a linear system with Gaussian noise. It is anticipated that by using the ideas contained in Section III, many other AIS algorithms can be developed. Furthermore, nonparametric AIS techniques that are particularly appealing, in that they yield estimates of the unconstrained optimal IS density regardless of the form of the system, need to be further investigated. The difficulty with the nonparametric approach lies in the estimation of functionals of many variables with small sample sizes. This problem is exemplified in [17], where it is shown that the number of samples required to estimate an M-dimensional pdf to a given accuracy using kernel estimation techniques increases exponentially with M.

The experience with parametric AIS in this work suggests that a well-chosen adaptation scheme enables the IS density parameters to converge to their optimum value. Furthermore, at convergence, the sample variance yields a good estimate of the true variance (see Tables I and II).

In fact, in all of the experiments run by the authors, the 95% confidence intervals determined by $[\hat{p}_e \pm Z_{\alpha/2}\sqrt{\widehat{\sigma^2}/N^*}]$ contained the true value of p_e, where $Z_{\alpha/2}$ is chosen such that $\int_{Z_{\alpha/2}}^{\infty} 1/\sqrt{2\pi}e^{-(x^2/2)}\,dx = \alpha/2$ for the $(1-\alpha)$ 100% confidence interval. This additional property of such AIS techniques is significant in view of the fact that calculation of the estimator variance is in itself often a difficult problem.

REFERENCES

[1] G. L. Ang, "Kernel method in importance sampling density estimation," in *Proc. ICOSSAR '89, 5th Int. Conf. Structural Safety Reliability,* Aug. 1989, San Francisco, CA, 1989, pp. 1193–1200.

[2] N. C. Beaulieu, "A composite importance sampling technique for digital communication system simulation," *IEEE Trans. Commun.,* vol. 38, pp. 393–396, Apr. 1990.

[3] N. C. Beaulieu, "An investigation of Gaussian tail and Rayleigh tail density functions for importance sampling digital communication system simulation," *IEEE Trans. Commun.,* vol. 38, pp. 1288–1292, Sept. 1990.

[4] C. G. Bucher, "Adaptive sampling—An iterative fast Monte Carlo procedure," *Structural Safety,* vol. 5, pp. 119–126, June 1988.

[5] Y. Ibrahim, "An adaptive importance sampling strategy," *Mech. Comput. in 1990's,* pp. 1288–1292, 1990.

[6] M. C. Jeruchim, "Techniques for estimating the bit error rate in the simulation of digital communication systems," *IEEE J. Select. Areas Commun.,* vol. SAC-2, pp. 153–170, Jan. 1984.

[7] A. Karamchandani, "Adaptive importance sampling," in *Proc. ICOSSAR (Int. Conf. Structural Safety Reliability),* Aug. 1989, pp. 855–862.

[8] E. L. Lehmann, *Theory of Point Estimation.* New York: Wiley, 1983.

[9] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE J. Select. Areas Commun.,* vol. 6, pp. 67–75, Jan. 1988.

[10] D. G. Luenberger, *Linear and Nonlinear Programming.* Addison-Wesley, 1984.

[11] R. E. Melchers, "Search-based importance sampling," *Structural Safety,* vol. 9, pp. 117–128, Dec. 1990.

[12] J. M. Mendel, *Lessons in Digital Estimation Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[13] S. S. Sadowsky and J. A. Bucklew, "Large deviations theory and asymptotically efficient Monte Carlo estimation," *IEEE Trans. Inform. Theory,* vol. 36, pp. 579–588, May 1990.

[14] H. J. Schlebusch, "Nonlinear importance sampling techniques for efficient simulation of communication systems," in *Proc. Int. Conf. Commun.,* Apr. 1990, pp. 631–635.

[15] M. Devetsikiotis and J. K. Townsend, "An algorithmic approach to the optimization of importance sampling parameters in digital communication system simulation," *IEEE Trans. Commun.,*1991.

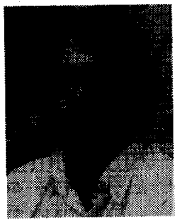
[16] K. S. Shanmugam and P. Balaban, "A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.,* vol. 28, pp. 1916–1924, Nov. 1980.

[17] B. W. Silverman, *Density Estimation for Statistical Data Analysis.* London: Chapman and Hall, 1986.

[18] E. Parzen, "On the estimation of a probability density function and mode," *Ann. Math Stat.,* vol. 33, pp. 1065–1076, 1962.

J. Scott Stadler (S'91) received the B.S.E.E. degree from Worcester Polytechnic Institute in 1987, and the M.S.E.E. degree from the University of Southern California in 1989.

From 1987 to 1990, he was employed as a Communication Systems Engineer at TRW Inc., Redondo Beach, CA. He is currently a candidate for the Ph.D. degree in electrical engineering at the University of Pennsylvania. His current research interests include detection theory, adaptive systems, and fast simulation techniques.



Sumit Roy received the B.Tech. degree from the Indian Institute of Technology in 1983, and the M.S. and Ph.D. degrees from the University of California at Santa Barbara in 1985 and 1988, respectively, all in electrical engineering, as well as the M.A. degree in applied probability and statistics in 1988.

Currently, he is an Assistant Professor at the Moore School of Electrical Engineering, University of Pennsylvania, engaged in research activities spanning the general area of statistical signal processing with a particular emphasis on communication systems.