Defence Science Journal, Vol 49, No 1, January 1999, pp. 9-17


# Importance Sampling – the Simulation Theory of Rare Events and its Applications

Rajan Srinivasan
*Centre for Air Borne Systems, Bangalore* -560 037.

**ABSTRACT**

Importance sampling has had its origin in Monte Carlo simulation and in the last 15 years or so, it has emerged as a powerful means of analysis and design with applications in several areas of engineering and science. The principle of importance sampling is described in this paper and attention is focussed on some of the subtleties involved in its implementation. To reach a broad audience of practising engineers and scientists, mathematical rigour has been maintained at a comfortable level.

## 1. INTRODUCTION

Development of Monte Carlo (MC) simulation method of analysis of scientific phenomena and systems is attributed to the mathematician von Neumann and others. who were engaged in the Los Alamos experiment of the 1940's. Since its inception, MC simulation has found applications ranging from statistical thermodynamics in disordered systems to the design and analysis of engineering structures typically characterised by high complexity. Indeed whenever an engineering problem is analytically intractable (which is often enough) and solution by numerical techniques becomes prohibitively expensive computationally, generally the last resort to determining the input-output characteristics of or the states within a system is to carry out a simulation. Essentially, simulation is concerned with replicating or mimicking a system and its operation by mechanising the exact mathematical equations that describe the system and all its inputs using a computer. Reliability of the results of a simulation is governed primarily by the authenticity of the analytical model, i.e. by how closely the mathematical descriptions used fit the actual physical system and its environs. The accuracy is, of course, governed by precision of the computations.

In several applications, systems are driven or perturbed by stochastic inputs that may arise from natural sources or are derived from outputs of other systems. It is of interest to determine the average behaviour of the system in terms of its response or some other internal states. The MC method then uses a model of these stochastic processes to generate random numbers, and runs them through the simulated system to give rise to responses of interest. If this is carried out for a sufficiently large number of times, the law of large numbers guarantees that the averaged results approach the mean or expected behaviour of the system. Thus, analysis by MC simulation can play a very useful role in the design process of complex systems. The MC method, however, is not limited to studying systems with stochastic inputs. An early and classical use has been the evaluation of integrals of functions over complicated multidimensional

Received 11 October 1996, revised 16 May 1997

regions. Random points are generated over a simpler or more convenient region which contains the desired region of the integration. The points that fall in the latter region are then used to evaluate the integrands and the results are weighted and summed up to provide an estimate of the integral.

There are many important application areas wherein the performance of systems can be closely linked with the occurrence of certain rare phenomena or events. Importance sampling (IS) is the theory of MC simulation that deals with such situations. In digital communications, for example[1], bit error probabiiities over satellite links using error correction coding are required to be as low as $10^{-10}$. The false alarm probabilities in radar and sonar receivers[2] are usually constrained not to exceed values close to $10^{-6}$. In packet switching over telecommunication networks[3], an important parameter of performance is the probability of packet loss at a switch. These probabilities are required to be of the order $10^{-9}$. In designing fault-tolerant computers, the probability that the system will fail in a certain time is a performance index and will be very small. In these and other similar cases[4], analysis by mathematical or numerical techniques becomes very difficult owing to nonlinearities and couplings present in the systems, high dimensionality, and other such problems. Conventional MC simulation also becomes ineffective due to excessively large run times required to generate the rare events of interest in sufficiently large numbers for obtaining statistically significant results. It is here that IS has a powerful role to play. For the last 15 years, IS has been used effectively in various applications requiring the analysis of rare events. In this method of analysis, probability distributions of the underlying processes that give rise to rare events are changed or 'biased' so as to cause these events to occur more frequently, rendering them quickly countable. Each event is then weighed appropriately to provide unbiased estimates of the rare event probabilities. It turns out that if the biasing distribution is chosen carefully, the resulting estimate has markedly lower (error) variance than the conventional MC estimate. Apart from the use of IS in specific applications, an important aspect of IS research has been concerned with the search for good biasing distributions to be used in simulation.

## 2. PRINCIPLE OF IMPORTANCE SAMPLING

A simple way to comprehend the IS concept is to consider the estimation by simulation of the tail probability $p_t$ of an event $\{X \geq t\}$, where $X$ is a random variable with probability density function $f(x)$ assumed to exist. The value of $t$ is such that the event is rare, i.e., $p_t$ is small. The usual MC procedure is to generate a $K$-length iid sequence $\{X_i\}_1^K$ from the distribution corresponding to $f$ and count the number $k_t$ of random variables that exceed the threshold $t$. Since this can be considered as a sequence of Bernoulli trials with success probability $p_t$, the random variable $k_t$ is characterised by the binomial distribution:

$$P(k_t = k) = \binom{K}{k} p_t^k (1-p_t)^{K-k}, \quad k = 0,1,\ldots,K \tag{1}$$

An intuition-based estimate $\hat{p}_t$ of $p_t$ is, of course, $k_t/k$. More formally, the maximum likelihood estimate is obtained by setting $\partial P / \partial p_t = 0$ in Eqn (1). This yields

$$\hat{p}_t = k_t / K = \frac{1}{K}\sum_1^k 1_t(X_i) \tag{2}$$

where

$$1_t(x) = \begin{cases} 1, & x \geq t \\ 0, & x < t \end{cases}$$

is the indicator function for the event of interest. By the law of large numbers, $\hat{p}_t \xrightarrow{a.s.} p_t$. This MC estimate is unbiased. That is

$$E\{\hat{p}_t\} = \frac{1}{K}\sum_{1}^{K} E\{1_t(X_i)\}$$
$$= \frac{1}{K}\sum_{t}^{k}\int 1_t(x) f(x)\,dx$$
$$= \int_{t}^{\infty} f(x)\,dx = p_t \qquad (3)$$

by definition. The variance of $\hat{p}_t$ is then given by

$$\text{var}\,\hat{p}_t = \frac{1}{K}\ \text{var}\,1_t(X)$$
$$= \frac{1}{K}[E\{1_t^2(X)\} - p_t^2] = \frac{1}{K}(p_t - p_t^2) \qquad (4)$$

since the $X_i's$ are iid. Some observations can be made from this. From Eqns (2) and (3), it is noted that the average number of threshold crossings is $E\{k_t\} = Kp_t$. Therefore, in order to obtain, on an average, a non-zero number of crossings and hence a non-zero value for the estimate $\hat{p}_t$, it is needed to perform the simulation with sequence lengths of $K > 1/p_t$. Further, it can be shown using a central limit theorem argument that if we want to estimate $p_t$ with 95 per cent confidence of having an error not more than 20 per cent, we need to have $K \geq 100/p_t$. Thus, if $p_t = 10^{-6}$, a sequence length of at least $10^8$ is required. This places severe demands on the period length of most random number generators as well as requiring large computation times for the simulalion. Therefore, attempting to obtain a low variance estimate by increasing $K$ is clearly impractical.

Consider now some density function $f_*(x)$. From the definition of $p_t$, it can be written that

$$p_t = E\{1_t(X)\}$$
$$= \int 1_t(x)\frac{f(x)}{f_*(x)}\, f_*(x)\,dx$$
$$= E_*\{1_t(X)\, W(X)\} \qquad (5)$$

where $W(x) = f(x)/f_*(x)$ is called the weighing function and $E_*$ denotes expectation wrt density $f_*$. Based on this, an alternative estimate of $p_t$ can be defined as

$$\hat{p}_t = \frac{1}{K}\sum_{t}^{K} 1_t(X_i)\, W(X_i);\ X_i \sim f_* \qquad (6)$$

where $X_i \sim f_*$. denotes that each $X_i$ is drawn from the distribution of $f_*$. This is the IS estimator of $p_t$. By Eqn (5) also this estimator is unbiased, i.e. $E_*\{\hat{p}_t\} = p_t$.

As in Eqn (4), the variance of the estimator is given by

$$\text{var}_*\,\hat{p}_t = \frac{1}{K}[E_*\{1_t^2(X)W^2(X)\} - p_t^2] \qquad (7)$$

The IS problem then centres around attempting to determine a biasing density $f_*$ such that $\text{var}_*\,\hat{p}_t$ in Eqn (7) is less than $\text{var}\,\hat{p}_t$ in Eqn (4) for the same value of $K$ used in both estimations. Alternatively, for equal estimator variances, the simulation length $K_I$, required for IS estimator should be less than the length $K_{MC}$, for MC estimator. This permits the definition of a performance measure $\Gamma$, called the IS gain, as $\Gamma = K_{MC}/K_I$, the ratio of sample sizes required for $\text{var}_*\,\hat{p}_t = \text{var}\,\hat{p}_t$. From Eqns (4) and (7) one gets:

$$\Gamma = \frac{p_t - p_t^2}{E_*\{1_t^2(X)W^2(X)\} - p_t^2} \qquad (8)$$

as the IS gain resulting from the use of simulation density $f_*$. For IS to provide appreciable improvement over MC simulation, the gain should satisfy $\Gamma >> 1$.

Before considering specific examples, a result that provides some insight into the selection of good biasing distributions is given. There exists a biasing density that reduces the estimator variance in Eqn (7) to zero, that is, it provides an exact estimate of $p_t$. Consider the expression in square brackets in the RHS of Eqn (7). This can be written, using Eqn (6), as

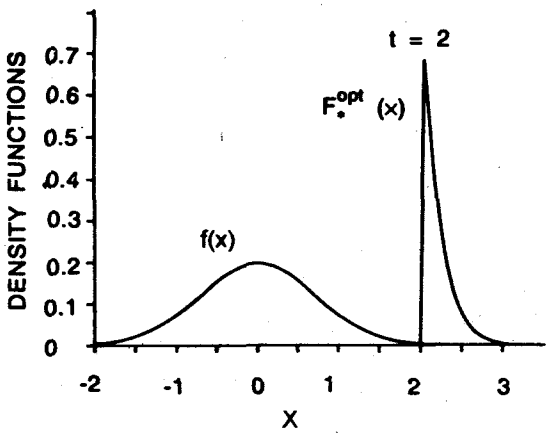


**Figure 1. Unconstrained optimal biasing density**

$$\begin{aligned} \text{var}_* 1_t(X)W(X) &= E_*\{1_t^2(X)\,W^2(X)\} - p_t^2 \\ &= E_*\{1_t^2(X)\,W^2(X)\} \\ &\quad - E_*^2\{1_t(X)\,W(X)\} \end{aligned} \tag{9}$$

because of unbiasedness. It is obvious that this can be made zero if one chooses an $f_*$ that satisfies

$$1_t(X)\,W(X) = p_t \tag{10}$$

Denoting this $f_*$ as $f_*^{opt}$ and using the definition of $(x)$ in Eqn (10), one gets

$$f_*^{opt}(x) = \frac{1}{p_t} 1_t(x) f(x) \tag{11}$$

which is known as the unconstrained optimal biasing density that provides a zero variance eslimate. Two observations can be made from Eqn (11). Firstly, the optimal density is in general unrealizable because it depends on $p_t$, the unknown quantity that is to be estimated. Indeed, if $f_*^{opt}$ could be found, then simulation is not needed as $p_t$ would be known. Secondly, presence of the indicator function implies that the entire probability mass of $f_*^{opt}$ is placed in the region $x \geq t$, the event region of interest (Fig. 1). This provides a useful guideline to the selection of $f_*$. That is, a good simulation density attempts to increase the probability mass in the event region. This is, in essence, the concept of IS.

One may ask why IS is needed at all if it is only to estimate the tail probability of a random variable whose density function is known, as this probability can be obtained analytically, or at most, by a numerical integration over the event region. The real power of IS lies in its ability to precisely estimate the rare event probabilities of a random variable which is a function of several random variables. Such situations arise in applications and instances of functions include iid and non-iid sums, and other nonlinear transformations encountered, for example, in signal processing and communications.

## 3. METHODS OF IMPORTANCE SAMPLING

Of the available methods of choosing $f_*$, two methods that are popular, easy to implement, and frequently used in applications are discussed:

### 3.1 Variance Scaling

Shilting probability mass into the event region $\{X \geq t\}$ by scaling results in heavier tails of the density function and is one of the earliest IS methods used in practice.

*Case* 1

Consider a random variable $X$ with one-sided Rayleigh density function $f(x)=2xe^{-x^2}$, $x \geq 0$ for which an estimate of $p_t$ is required. We obtain by direct integration that $p_t = e^{-t^2}$. Let us scale the *rv* $X$ by **a** > 1, so that $f_*$ is the density function of $aX$. From elementary transformation theory it follows that

$$f_*(x) = \frac{2x}{a^2} e^{-x^2/a^2}, \quad x \geq 0$$

By the definition of $W$, one obtains

$$W(x) = a^2 e^{-(1-1/a^2)x^2}$$

and using Eqn (6), the IS estimator is given by

$$\hat{p}_t = \frac{1}{K}\sum_{1}^{K} 1_t(X_i)a^2 e^{-(1-1/a^2)X_i^2},\ X_i \sim f_*$$

$$= \frac{1}{K}\sum_{1}^{K} 1_t(aX_i)a^2 e^{-(a^2-1)X_i^2},\ X_i \sim f \qquad (12)$$

which can be implemented easily. Of course, a good value of the scaling parameter a has to be chosen. That is, a should be selected to minimise the estimator variance in Eqn (7). Denoting by $I$, the expectation in the RHS of Eqn (7), one has

$$I(a) = \int 1_t(x)\, W^2(x) f_*(x)\, dx$$

$$= \int 1_t(x)\, W(x) f(x)\, dx$$

$$= 2a^2 \int_t^{\infty} x e^{-(2-1/a^2)x^2}\, dx = \frac{a^4}{2a^2-1} e^{-(2-1/a^2)t^2}$$

This can be minimised wrt a by setting $I'(a)=0$ and confirming that $I''(a)>0$ at the solution. One obtains for the optimum scaling parameter

$$a_{opt} = \left( \left( t^2 + 1 + (t^4+1)^{1/2} \right) \Big/ 2 \right)^{1/2} \approx t \quad \text{for } t >> 1$$

The resulting IS gain is obtained from Eqn (8) as

$$\Gamma_{opt} = \frac{p_t - p_t^2}{I(a_{opt}) - p_t^2}$$

Suppose $p_t = 10^{-6}$, then with $t = 3.7169$, one obtains $a_{opt} = 3.7859$ and $\Gamma_{opt} = 54145$. That is, approx 1850 samples from $f_*$ would yield the same estimator variance as $10^8$ samples from $f$.

Some observations can be made from this that are of help while implementing IS. In any application of IS, once a biasing scheme is chosen, it should be optimised to provide minimum variance. For the above case, an optimum value was selected for the scaling parameter. In actual application, it will seldom be possible to optimise the biasing scheme so easily. Usually one has to set up a sample or 'running' estimate of the variance or the $I$ function and then attempt to minimise it algorithmically[5] to extract the best estimator performance. Another point of importance is that excessive IS leads to degradation of performance, with the result that IS estimator variance can, in fact, exceed the MC estimator variance. Therefore, pushing too much probability mass into the event region is undesirable despite the fact that the unconstrained optimal biasing density places all mass in this region. This can be seen by letting $a \to \infty$ in the expression for $I(a)$ above. Clearly, $I(a) \to \infty$, implying that $\text{var}_*\, \hat{p}_t \to \infty$. Associated with excessive IS is another deleterious effect called underestimation. Consider the second expression in Eqn. (12) and let $a \to \infty$ for a fixed value of $K$. It follows that $\hat{p}_t \to 0$ with probability 1. That is, the estimator always underestimates $p_t$.

### 3.2 Density or Mean Translation

Another simple and very effective biasing technique employs translation of the density function (and hence random variable) so as to place more of its mass in the event region. In biasing by translation, the simulation density is given by $f_*(x) = f(x-c)$, where $c > 0$ is the amount of shift and is to be chosen to minimise $\text{var}_*\, \hat{p}_t$, or equivalently $I(c)$. If $X$ is a non-negative random variable with density $f(x)$, $x \geq 0$, the biasing density is $f_*(x) = f(x-c), x \geq c$, and the IS estimate is given by

$$\hat{p}_t = \frac{1}{K}\sum_{1}^{K} 1_t(X_i)\frac{f(X_i)}{f(X_i - c)};\ X_i \geq c,\ X_i \sim f_*$$

The mean value of $\hat{p}_t$ is

$$E_*\{\hat{p}_t\} = \frac{1}{K}\sum_{1}^{K} \int_c^{\infty} 1_t(x)\frac{f(x)}{f(x-c)} f(x-c)\, dx$$

$$= p_c \leq p_t \quad \text{for } c > t$$

That is, the estimate is biased if the original density is translated beyond t. Hence, translations for one-sided densities are restricted to $c \leq t$.

*Case* 2

Let $X$ be $N(0,1)$.

That is $f(x) = (1/\sqrt{2\pi})\exp(-x^2/2), -\infty < x < \infty$

Then, with $W(x) = f(x)/f(x-c)$, one has $W(x) = \exp(-cx + x^2)$. Substituting this in the expectation in Eqn (7) and simplifying, one obtains $I(c) = Q(c+t)\exp(c^2/2)$

where

$$Q(x) = \int_x^\infty e^{-y^2/2}dy/\sqrt{2\pi}$$

Using the approximation $Q(x) \approx (1/x\sqrt{2\pi})\exp(-x^2/2)$ for large $x$ in $I(c)$ and minimising on $c$ yields $c_{opt} = (1+t^2)^{1/2} \cong t$ for $t >> 1$. For $p_t = 10^{-6}$, one knows from error function tables that $t \cong 4.7534$ Then $c_{opt} = 4.8574$. Using this in $I(c)$, the gain evaluates to $\Gamma_{opt} = 184310$, so that approximately only 540 samples from $f_*$ provide the same estimator quality as $10^8$ from $f$. Note that if this same example is worked for IS by scaling, then approximately 5100 samples from (scaled) $f_*$ are required, nearly a 10-fold increase.

## 4. VARIANCE ESTIMATION & MINIMISATION

It is a rule of thumb that anything which can be placed in the form of an expectation can be estimated by simulation. To estimate $\text{var}_*$, $\hat{p}_t$ in Eqn (7) for the purpose of minimisation, one needs only estimate

$$I = E_* \{1_t^2(X)W^2(X)\} \tag{13}$$

since $K$ and $p_t$, are unaffected constants. One has, therefore,

$$\hat{I} = \frac{1}{K}\sum_1^K 1_t(X_i)\, W^2(X_i);\ X_i \sim f_* \tag{14}$$

as an unbiased estimate of $I$, which can be implemented alongwith $\hat{p}_t$. Assuming that $I$ has a unique minimum wrt to some parameter in a chosen biasing scheme, one needs to find the root of $\hat{I}' = 0$. If, for example, the chosen scheme is scaling, then an estimate of the optimum scaling parameter $a_{opt}$ can be found by the recursion

$$a_{m+1} = a_m - \delta \frac{\hat{I}'(a_m)}{\hat{I}''(a_m)} \tag{15}$$

where $m$ is the recursion index. This is just the Newton root finding formula wherein a rate parameter $\delta$ has been included to control convergence. The quantities $\hat{I}'$ and $\hat{I}''$ can be determined by successively differentiating $I$ in Eqn (13) and setting up their sample estimates. Owing to the stochastic nature of the algorithm, convergence is characterised by a small random vibration around the optimum value. This is typical of stochastic approximation procedures. It has been found in practice that the simple Newton formula yields good results when IS by scaling is applied to the analysis of CFAR detectors used in radar and sonar.

To evaluate the performance of an IS implementation, it is necessary to estimate the gain $\Gamma$ in Eqn (8). An estimate generally used is

$$\hat{\Gamma} = \frac{\hat{p}_t - \hat{p}_t^2}{\hat{I} - \hat{p}_t^2} \tag{16}$$

which should also be implemented along with the IS estimator.

## 5. APPLICATION TO CFAR DETECTION

How the IS concept can be applied for estimating false alarm probabilities in CFAR detectors[6,7] has been indicated briefly. Virtually the entire class of CFAR processors used in radar and sonar receivers employ target detection algorithms of the form $X \gtrless tY$, where the upper event indicates a target-present decision. The random variable $X$ represents the content of a range or Doppler cell that is tested for a target, and $Y$ is some function of the contents of a set of cells called the CFAR window that surrounds the test cell. Such a simple

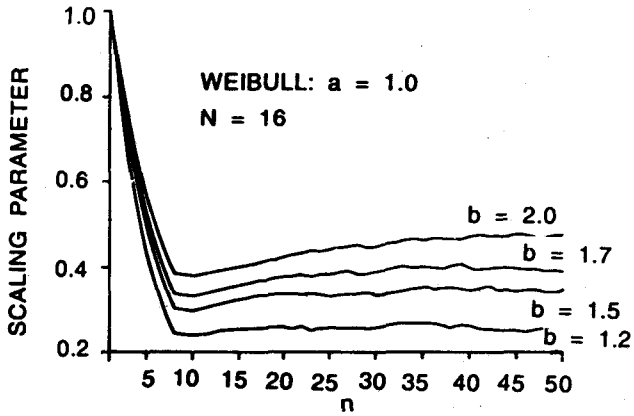


**Figure 2. Convergence of scaling parameter algorithm for $p_t = 10^{-6}$.**

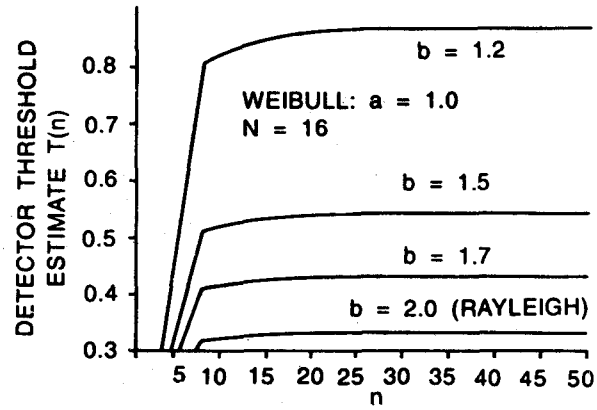


**Figure 3. Convergence of threshold finding algorithm for $p_t = 10^{-6}$.**

detection structure is a consequence of a requirement for maintaining $p_t = P(X \geq tY)$ at a specified value in the face of variations in the statistical properties of $X$ when it is known that there is no target in the test cell, coupled with certain model assumptions regarding the detection environment. Then $p_t$ represents the false alarm probability of the detector. Although it does not appear directly as a tail probability, the event $\{X \geq tY\}$ is a rare event. The random variable $Y$ could be an iid sum or a more complicated processing of the cells. Since cell outputs are usually the result of energy maximisation processing, it can be assumed that $X$ and $Y$ have one-sided densities on $[0, \infty]$.

The task is to perform IS on $X$ and $Y$ to estimate $p_t = P(Z \equiv X - tY \geq 0)$. The optimal biasing density for $Z$ places all its mass in the region $\{Z \geq 0\}$. It is clear that a biasing density for $X$ should shift its probability mass to the right on the line $[0,\infty]$. It also turns out that biasing for $Y$ must shift mass to the left towards zero, effecting a compression of the density function tail. This can be seen by rewriting $p_t$ as

$$\begin{aligned} p_t &= P(Y \leq X/t) \\ &= \int P(Y \leq x/t) f(x)\, dx. \end{aligned}$$

Since a biasing for $Y$ must result in an increase in $p_t$, it is clear that there must be an increase in $P(Y \leq x/t)$ for each $x$, implying a compression of the density of $Y$ toward zero. Therefore, for IS analysis by scaling of CFAR detectors, the cell under test must be scaled up and the random threshold scaled down.

However, a more effective IS method has also been developed by the author which is applicable in several radar clutter and sonar reverberation situations. Assuming, for the purpose of illustration, that the density function of $X$ is integrable analytically, i.e., in closed form, then one can write

$$\begin{aligned} p_t &= P(X \geq tY) \\ &= \int P(X \geq ty) f_Y(y)\, dy \\ &= \int g_t(y) f_Y(y)\, dy = E_Y\{g_t(Y)\} \end{aligned} \tag{17}$$

where $g_t(y) \equiv P(X \geq ty)$ and $E_Y$ denotes expectation over the density of $Y$. Clearly, to estimate $p_t$ by IS, one needs to bias only the density of Y. So, one has

$$\hat{p}_t = \frac{1}{K}\sum_{l}^{K} g_t(Y_i) W(Y_i);\ Y_i \sim f_{Y^*} \tag{18}$$

where $W$ is defined as usual and $f_{Y^*}$ is some compressing biasing density for $Y$. The results of applying this method to cell averaging CFAR detection preceded by an envelope detector in the presence of Weibull clutter are shown in

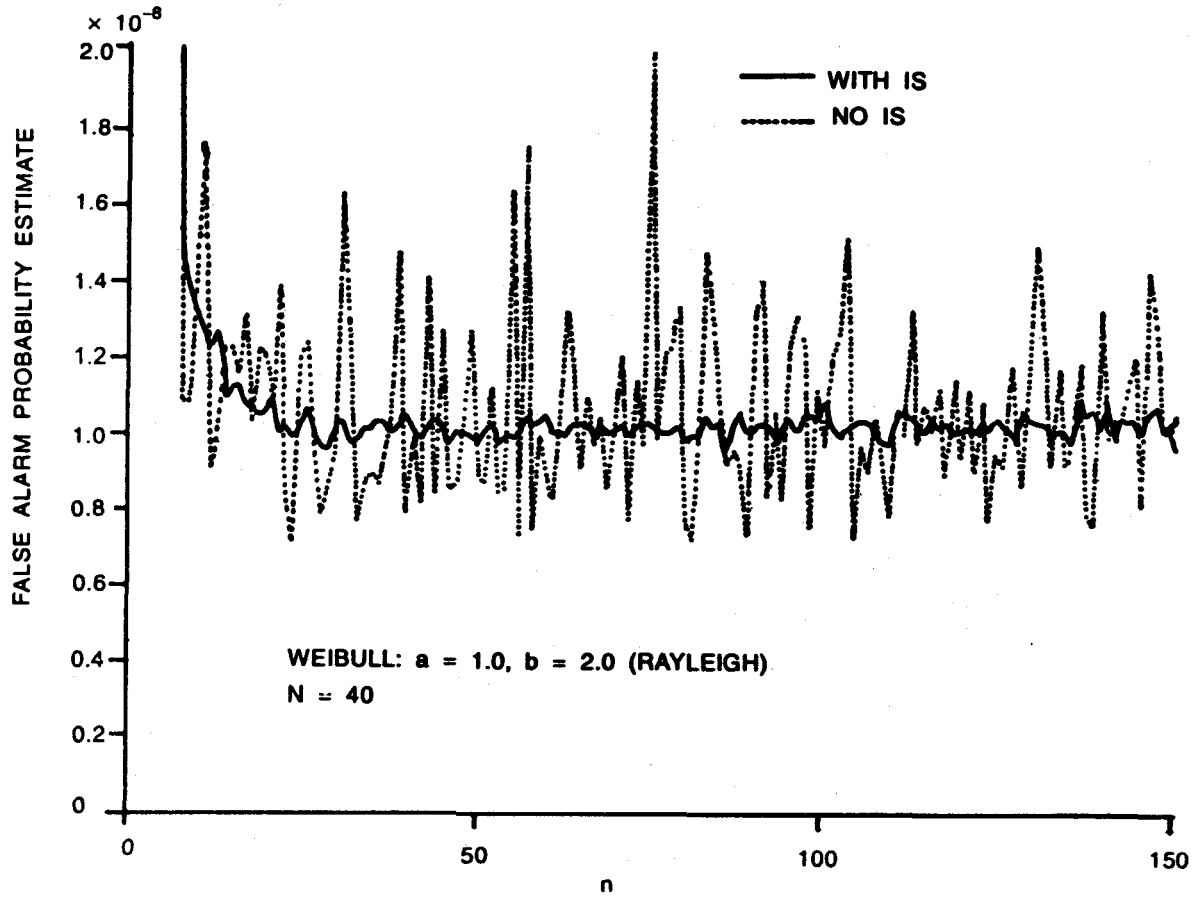


**Figure 4. False alarm probability estimates with and without IS for desired $p_t = 10^{-6}$**

Figs (2)-(4). This CFAR problem is intractable analytically and numerically very tedious, and IS provides the only quick and accurate technique for its solution. The convergence of the scaling parameter $a_m$ for a target $p_t = 10^{-6}$ with a CFAR window length of $N = 16$ and various Weibull clutter shape parameters, is shown in Fig. 2.

## 6. INVERSE IS PROBLEM

In practical applications, it is often necessary to determine the value of $t$ that can provide a specified rare event probability $p_t$. In CFAR detectors, for example, this means finding the value of threshold multiplier that results in a desired false alarm probability. We shall refer to this as the inverse IS problem[6]. There is a powerful feature in the IS method described just above which permits solution of this problem in an easy and accurate manner. Form the stochastic objective function $J(t) = (\hat{p}_t - \alpha_0)^2$, where $\alpha_0$ is the desired false alarm probability in a CFAR detector. Our aim is to find a $t = t_0$, such that $p_{t_0} = \alpha_0$. By minimising $J(t)$ what one can find is a $t = \hat{t}_0$, such that $J(\hat{t}_0) = 0$. It can be proved that $J(t)$ has a unique minimum and hence it is simple to seek its minimum using a descent method. Algorithms based on descent methods require computation of gradients of the objective function $J(t)$. But this is easily done, since $\hat{p}_t$, in Eqn (18) is differentiable wrt $t$ by virtue of existence of the density function $f$ of $X$. The results of using such a threshold finding technique together with the adaptive scaling parameter algorithm of Eqn (15) are shown in Fig. 3. The false alarm probability estimates for a $N = 40$ CA-CFAR detector in Rayleigh clutter for $p_t = 10^{-6}$ are shown in Fig. 4. The advantage of using this IS technique is evident. The results are based on an IS sample size of 1000.

## 7. DISCUSSION

The concept of IS and indications of its use in applications have been introduced in this paper. Admittedly, we have only scratched the surface of the subject in the depth of the theory as well as in the extent of its usage. On the practical side, there are undoubtedly a number of applications where IS theory can be used to solve very difficult problems through simulation. In our opinion, the theoretical analysis and mathematical manipulations required to implement a good IS scheme in application are far less burdensome and computationally demanding than a brute force analytical-numerical attack.

Apart from the initial descriptions of IS and scaling and translation techniques therein, the methods and detection applications presented here are part of more extensive results on IS obtained by the author recently.

## ACKNOWLEDGEMENT

The author is grateful to Dr K Ramchand, Director, Centre for Air Borne Systems (CABS) Bangalore, for his encouragement and support to carry out this study.

## REFERENCES

1. Chen, J.; Lu, D.; Sadowsky, J. S. & Yao, K. On importance sampling in digital communication-Part I: Fundamentals. *IEEE. J. Sel. Areas Commun.*, 1993, **11**, 289 - 99.

2. Mitchell, R. L. Importance sampling applied to simulation of false alarm statistics. *IEEE Trans. Aerosp. Electron. Syst.*, 1981, **17**, 15-24.

3. Heidelberger, P. Fast simulation of rare events in queuing and reliability models. *ACM Trans. Modelling Comp. Simu.*, 1995, **5**, 43-85.

4. Remondo, D.; Srinivasan, R.; Nicola, N.; van Etten, W.; & Tattije, H. Impact of crosstalk on the performance of an optical cross-connect considering finite extinction ratio. Proceedings of the International Conference on Applications of Photonic Technology 98, July 1998, Ottawa, Canada.

5. Devetsikiotis, M. & Townsend, J. K. An algorithmic approach to the optimisation of importance sampling parameters in digital communication system simulation. *IEEE Trans. Commun.*, 1993, **41**, 1464-73.

6. Srinivasan, R. Some results in importance sampling and an application to detection. *Signal Processing*, 1998, **65** (1), 73-88.

7. Srinivasan, R. Estimation and approximation of densities of iid sums via importance sampling. *Signal Processing*, 1998, **71** (3).