Estimation of False Alarm Probabilities in Cell Averaging Constant False Alarm Rate Detectors via Monte Carlo Methods

Graham V. Weinberg

Electronic Warfare and Radar Division
Systems Sciences Laboratory

DSTO–TR–1624

ABSTRACT

Monte Carlo Methods are introduced and used to estimate false alarm probabilities. The estimation of the latter is important in the context of performance analysis of Constant False Alarm Rate (CFAR) radar detection processes. A CFAR detector estimates the clutter level, producing a threshold, and a target is declared present if the statistic representing the test observation exceeds this threshold. The latter is adjusted adaptively, so that the rate of false alarms is held constant. Hence, in a radar analysis context, the performance of a CFAR process can be determined from whether it maintains a constant false alarm rate. In order to compare the performance of a number of different CFAR schemes, in a common clutter environment, we need to estimate these false alarm probabilities. This can be done quite easily using a basic Monte Carlo estimator. However, the latter may require a very large number of iterations in order to produce a reasonable estimate. To reduce this number of iterations, importance sampling techniques can be used. To illustrate these techniques, we consider the simple case of cell averaging CFAR in a Gaussian environment, with square law detection. This enables comparison of estimators with an exact result.

APPROVED FOR PUBLIC RELEASE
Estimation of False Alarm Probabilities in Cell Averaging Constant False Alarm Rate Detectors via Monte Carlo Methods

EXECUTIVE SUMMARY

Constant False Alarm Rate (CFAR) detectors use an adaptive threshold to compare the estimated clutter level with a test statistic, the latter representing the cell under test. When this test statistic exceeds the threshold, a target is declared present. The adaptive threshold is adjusted so that the probability of false alarm, that is, the probability of declaring a target present when there is actually no target, remains fixed. Performance analysis of various CFAR schemes focuses on whether the sample false alarm rate remains constant. Thus, in order to compare CFAR schemes in various clutter environments, a comparison of the false alarm probabilities is required.

Estimation of false alarm probabilities, and definite integrals in general, can be performed using a class of techniques known as Monte Carlo Methods. A key issue with such techniques is that the estimators based upon them may require a large number of iterations to produce a good estimate. In order to improve this situation, by reducing the number of iterations, Importance Sampling (IS) can be used.

This report will examine the estimation of false alarm probabilities for performance analysis of CFAR detection processes, using Monte Carlo methods. In particular, we will be interested in the performance of a number of importance sampling estimators. A key issue to be considered is whether the number of iterations needed to produce a reasonable estimate can be significantly reduced. Attention will be restricted to the case of a cell-averaging (CA) CFAR, in a Gaussian clutter environment. This enables comparison of estimated results to precise probabilities of false alarm.

This report begins with an outline of CA-CFAR techniques, and Monte Carlo Methods, followed by a derivation of an exact analytic relationship between false alarm probabilities and thresholds, in the Gaussian clutter environment. Monte Carlo IS techniques are then introduced and used to estimate false alarm probabilities. Finally, simulation studies are used to demonstrate the power of the Monte Carlo IS approach.
Graham V. Weinberg

Electronic Warfare and Radar Division

Graham V. Weinberg was educated at The University of Melbourne, graduating in 1996 with a Bachelor of Science degree with Honours in Mathematics and Statistics. Following this he was conferred in 2001 with a Doctor of Philosophy degree in Applied Probability, also at The University of Melbourne. His doctoral thesis examined probability approximations for random variables and point processes, using a technique known as the Stein–Chen Method. The latter enables a probabilistic assessment of the errors in distributional approximations of stochastic processes.

His post-doctoral work began with a 6 month contractual position with a finance group known as Webmind, investigating artificial intelligence applied to stock market prediction. This was based in the Department of Psychology at The University of Melbourne. Following this was employment as a Research Fellow in the Teletraffic Research Centre, at the University of Adelaide. The latter involved working as a telecommunications consultant to business and industry. Since early 2002 he has been employed as a Research Scientist with DSTO, examining radar performance analysis techniques for the P3 Orion maritime patrol aircraft. From mid 2004 he has changed directions in terms of research, now working on problems related to radar resource management and waveform analysis.
# Contents

1 Introduction .................................................. 1
   1.1 Monte Carlo Methods .................................... 1
   1.2 CA-CFAR Context ........................................... 3
   1.3 Monte Carlo Fundamentals .............................. 5

2 A Cell Averaging CFAR Model ................................. 8
   2.1 From Signal Return to Envelope Detection Output .......... 8
   2.2 Exponentially Distributed Post-Envelope Returns ............. 10

3 Estimation of False Alarm Probabilities ....................... 12
   3.1 Standard Monte Carlo Methods .......................... 12
   3.2 Importance Sampling Estimators .......................... 13
       3.2.1 Ad Hoc Biasing Density ............................. 14
       3.2.2 Chernoff Biasing Density ............................. 15
   3.3 The G-Function Method .................................... 17
       3.3.1 Example: CA-CFAR with Gaussian Clutter ............. 19

4 Simulation Comparison of Estimators .......................... 21

5 Conclusions .................................................. 22

References .................................................... 23

Appendix A: Figures ........................................... 25
1 Introduction

1.1 Monte Carlo Methods

In a realistic mathematical model of any real world phenomenon, it can be expected that the model complexity will often translate to difficulty in applying it to applications of interest. As an example, in the context of [Tonkin and Dolman 1990], where models are constructed for the radar signal return from a submarine periscope, parameters in the model may be difficult to estimate precisely. A sample of observations can be used, together with an appropriate estimator, to obtain an approximation to the unknown parameters. Due to their statistical nature, the estimator of a parameter will have a level of variance, and it is thus desirable to minimise this, in order to achieve a good approximation.

In many cases, these estimators are improved as the sample size increases. However, this can present a number of technical, as well as practical, problems. As an example of this, from a practical perspective, it may be difficult to obtain a large sequence of accurate observations. Consequently, an estimator may not be able to provide an acceptable level of accuracy.

The techniques known as Monte Carlo Methods provide a systematic way of estimating unknown parameters via simulation. Based upon the Law of Large Numbers, sample averages are used to estimate unknown parameters. The power of the method comes from the fact that it permits the estimation of integrals, and consequently can be used to approximate statistical expectations, and hence probabilities. The basic Monte Carlo techniques have existed for centuries, but the systematic development began during the 1940s, with the development of the atomic bomb. The name was coined by Metropolis, in the course of the Manhattan Project of the Second World War, due to the similarity between statistical simulation and games of chance. In 1948 Monte Carlo simulations were used by Fermi, Metropolis and Ulam to obtain estimates for the eigenvalues of the Schrödinger Equation.

The focus here will be on the idea of improving Monte Carlo simulations via a procedure known as Importance Sampling (IS), which from a mathematical perspective, involves a change in the underlying probability distribution associated with the estimator.

From a practical point of view, the idea is that certain observations will have more relevance to the estimation than others, and hence these “important observations” are emphasized. Additionally, since a new estimator is constructed, which focuses on only a subset of the sample points, the simulation run time should be reduced. In order to unbias this estimator, it must therefore be weighted to compensate for the reduction in sample size. It is also an objective of IS to reduce the variance of the estimator, and so IS methods are often referred to as variance reduction techniques.

As reported in [Smith, Shafi and Gao 1997], Importance Sampling (IS) began during the 1940s with the ideas of von Neumann applied in the context of particle splitting and Russian roulette. Prior to 1970 the main development of IS was in the context of nuclear
physics, with the work of [Kahn 1950, 1956] and [Kahn and Mann 1957]. Applications in the communications modelling context began to appear in the literature during the 1970s, with the work of [Balaban 1976]. An early application to the subject of false alarm probability estimation appears in [Mitchell 1981]. The latter application of IS techniques will be the main practical focus of this report. [Shanmugam and Balaban 1980] is attributed with popularising the use of IS in bit error rate estimations. [Cottrell, Fort and Malgouyres 1983] introduced the idea of using large deviation theory to construct an IS simulation of rare events. This resulted in a systematic development of the method, with applications to communications problems, with work such as [Sadowsky and Bucklew 1990]. Later developments include investigations of suboptimal biasing densities, as in [Orsak and Aazhang 1989] and [Gerlach 1999], as well as a number of different other techniques, which will be subsequently considered.

In IS, a change of probability measure is performed through a biasing distribution. The choice of an appropriate biasing distribution is the main problem in the design of an effective and efficient Monte Carlo (MC) simulator. The “art” of good MC simulation via IS is to choose an appropriate biasing density for the problem under consideration. A large percentage of the IS literature after 1987 focuses on the selection of biasing densities, such as [Orsak 1993] and [Orsak and Aazhang 1989, 1991, 1993]. As will be observed, in cases of interest there exists an optimal IS biasing density, which is not suitable for use in practice because it depends on the parameter being estimated. This has resulted in designing suboptimal biasing densities, based on this ideal form. In many cases this approach has generated improved MC simulations.

As remarked previously, a biasing distribution is used to emphasize regions of importance to the estimation process. Modification of an unbiased estimator, by reducing its input sample size, will result in a biased estimate. The resulting estimator is then unbiased through pointwise weightings. The correct choice for the weighting function, to produce an unbiased estimator, turns out to be the Radon-Nikodym derivative of the original distribution, with respect to the biasing distribution. Equivalently, when densities exist, this reduces to a ratio of the original density and the biasing density.

As pointed out in [Smith, Shafi and Gao 1997], IS is not without its limitations. A key problem is designing suitable biasing densities as system complexities increase. It may in fact be very difficult to apply the method successfully to complex systems, as illustrated in [Hopmans and Kleijnen 1979]. However, this does not mean IS is completely useless in complex system modelling. An example of IS applied to a complex system, in the setting of simulating Viterbi decoders, is [Sadowsky 1990]. The relevant point to keep in mind is that the method is system dependent, and there may be successes or failures in its application, depending on the specific system under consideration.

This study will focus on IS applications to the estimation of false alarm probabilities in a signal detection context. The interest in estimating such probabilities is to facilitate the performance analysis of radar detection schemes. In particular, we are interested in the analysis of Constant False Alarm Rate (CFAR) detection schemes. Such a detector compares a test observation with an adaptive threshold, the latter produced so that the rate of false alarms remains constant. A false alarm is the declaring of a target present
when in fact there is only clutter and noise present. Two different CFAR schemes can be compared by looking at the rate of false alarms, in a common clutter environment. In this case, we need to control the clutter environment, so that a valid comparison can be performed. We will be interested in the performance of estimators of the probability of false alarm. False alarm probabilities are typically set to small numbers, so that Monte Carlo estimators may require a very large number of iterations to produce a reasonable estimate. We will be interested in examining the possibility of significantly improving this drawback, using IS.

We will restrict our attention to a Gaussian clutter and noise model. An advantage in doing this is that the exact probability of false alarm is available, providing a gauge to measure the performance of estimators.

The IS techniques considered here can be, and have been, applied to more general CFAR schemes, with non-Gaussian clutter models. See [Srinivasan 2000, 2001] for examples of this. In this report, we investigate a very simple CFAR scheme to compare these IS estimators.

We now briefly introduce CA-CFAR schemes, followed by an introduction to Monte Carlo techniques.

1.2 CA-CFAR Context

A CFAR processor is a signal processing tool which enables the automatic detection of targets in clutter and noise. It uses an adaptive threshold, where targets are declared present when a detection statistic exceeds this threshold value. This threshold is determined so that the rate of false alarms is held constant.

For a specified level of false alarm probability and noise level, a radar system must increase its transmitted power in order to increase the probability of target detection at a prescribed range. Thus we have an optimisation problem where we seek to maximise the detection probability subject to a false alarm probability constraint. In terms of producing optimum detection criteria, this suggests the usage of a Neyman-Pearson test, or Bayes decision strategy.

The Neyman-Pearson Theorem provides a mechanism for determining the form of the uniformly most powerful test between two statistical hypotheses. The following result is from [Beaumont 1980]:

**Theorem 1.1** Let $X_1, X_2, \ldots, X_n$ have joint probability density function $f(x_1, x_2, \ldots, x_n | \theta_1, \theta_2, \ldots, \theta_k)$, and suppose we want to test the null hypothesis $H_0 : \theta_i = \theta^0_i$ against the alternative $H_1 : \theta_i = \theta^1_i$, where the $\theta^0_i$ and $\theta^1_i$ are known constants for each $i \in \{1, 2, \ldots, k\}$, representative of predetermined possible states. Suppose further that $U$ is the set of points $(x_1, x_2, \ldots, x_n)$ such that:
1. \( \frac{f(x_1, x_2, \ldots, x_n|H_1)}{f(x_1, x_2, \ldots, x_n|H_0)} > \zeta \), for some \( \zeta > 0 \);
2. \( P((X_1, X_2, \ldots, X_n) \in U|H_0) = \alpha \).

Then \( U \) is a decision region of size \( \alpha \) that has the greatest power\(^1\).

The ratio of joint densities in Theorem 1.1 is a ratio of likelihood functions, and hence this test is often called a likelihood ratio criterion test. This methodology can be used to construct a standard test format in the context of interest. [Minkler and Minkler 1990]\(^2\) contains a detailed description of this, to which the reader is referred.

In the case of detection of a single target observation, within a series of clutter and noise observations, in a single radar return, we are interested in testing the hypothesis: \( H_0 : r = \eta \) against the alternative \( H_1 : r = \nu + \eta \), where \( r(t) \) denotes the radar signal return, over a time observational window \([0, \psi]\) , \( \nu = \nu(t) \) is a target observation and \( \eta = \eta(t) \) is clutter and noise. It will be necessary to assume some statistical properties of the clutter and noise are known, such as their distributions are at least partially known, and distributional parameters can be estimated from radar returns. This is in particular very important in the context of determining the adaptive threshold. There are a number of models that can be used to simulate clutter. In the work to follow, we will focus on the simple case where the clutter and noise are modelled as Gaussian. This restriction is applied because it is possible to find, in this case, an analytical relationship between the false alarm probability and threshold parameter, for a given signal to noise ratio. This allows easy performance analysis of Monte Carlo estimators. The Importance Sampling techniques considered here readily apply in the case of non-Gaussian noise and clutter models. The difficulty is that it can be hard to derive an approximate analytical solution for comparison.

Returning to the hypothesis test context, suppose we have a sample of \( m \) clutter and noise observations \( X_1, X_2, \ldots, X_m \), and based upon these we construct a test statistic \( Y = Y(X_1, X_2, \ldots, X_m) \). We assume we have a test observation \( X_0 \), which is either a target or noise/clutter observation. This observation is referred to as the statistic of the cell under test (CUT). Using the approach of Theorem 1.1, it can be shown that the form of the best test is to reject \( H_0 \) if \( X_0 > \tau Y \), where \( \tau \) is a constant. The physical interpretation of this rejection criterion is that we compare the observation under test to a normalised test statistic, the latter being based upon the clutter and noise observations. The statistic \( Y \) gives an estimate of the magnitude of the clutter. In the present context we will only be interested in the case where \( Y \) is a mean or average value, so that it gives a mean estimate of the noise and clutter. \( Y \) gives an estimate of the magnitude of the clutter. In the present context we will only be interested in the case where \( Y \) is a mean or average value, so that it gives a mean estimate of the noise and clutter. Figure\(^3\) 1 contains an illustration of this CA-CFAR process. As Figure 1 illustrates, a series of radar returns are passed to a square law detector. Although not clearly illustrated in Figure 1, the CA-CFAR process actually has a number of buffer cells separating the CUT and the observational cells.

---

\(^1\)Recall that the power of a statistical hypothesis test is the probability of rejecting the null hypothesis when the alternative is actually true. Thus it statistically measures the likelihood of making a correct decision.

\(^2\)See Chapter 3 of this book for a full mathematical treatment of the classical detection problem.

\(^3\)All figures and plots have been placed in the appendix at the end of the report.
The processed returns are then averaged, normalised and compared to the observation under test. It then outputs a detection decision. A radar detector divides a scan region up into a number of small sets of observational cells, and this process would be run a number of times over a scan region to determine the false alarm probability.

We now introduce the basic ideas in Monte Carlo estimation of probabilities and integrals.

1.3 Monte Carlo Fundamentals

Monte Carlo Methods use simulation techniques to estimate integrals. An integral can be estimated via a summation of functions evaluated at randomly generated numbers. Since a probability or an expectation can be written as an integral, the techniques can be applied in this context too.

Monte Carlo estimation is based upon the Strong Law of Large Numbers (SLLN):

**Theorem 1.2** Suppose \( \eta_1, \eta_2, \ldots, \eta_m \) is a sequence of independent and identically distributed random variables with finite mean \( E[\eta] \). Then

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} \eta_j = E[\eta] \quad \text{almost surely.} \tag{1}
\]

In a probability context, the phrase *almost surely* (a.s.) means that the result holds except on a set of (probability) measure zero. Thus, for sufficiently large \( m \), the average of the random variables in (1) can be approximated by the mean value. To clarify this, consider the integral \( I = \int_{\Omega} w(x)f(x)dx \), where \( f \) is a density on \( \Omega \), and \( w \) is a deterministic function. Here \( \Omega \) may be a vector space, so that \( x \) may be a vector. Hence integral \( I \) is an expectation: suppose \( \eta \) is a random variable on \( \Omega \) with density \( f \). Then we can write \( I = E[w(\eta)] \). Now suppose \( \eta_1, \eta_2, \ldots, \eta_m \) is an independent and identically distributed (IID) sequence of random variables. Then the SLLN (1) implies that

\[
\lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} w(\eta_j) = E[w(\eta)] = I, \quad \text{a.s.} \tag{2}
\]

Thus it follows that we have the approximation

\[
I = \int_{\Omega} w(x)f(x)dx \approx \frac{1}{m} \sum_{j=1}^{m} w(z_j), \tag{3}
\]

where the sequence \( z_1, z_2, \ldots, z_m \) consists of realisations of the variables \( \eta_1, \eta_2, \ldots, \eta_m \).

The approximation in (3) is called a *Monte Carlo estimate*. This can be used to estimate integrals in a number of different contexts. This method can be used to easily estimate
analytically difficult integrals. As an example, consider the integral \( I = \int_0^\infty e^{-x^2} dx \). Since we can write the integrand as \( e^{-x^2 + x}e^{-x} \), and the function \( f(x) = e^{-x} \) is a density on \( \Omega = [0, \infty) \), it follows that the integral can be estimated by \( \frac{1}{m} \sum_{j=1}^{m} e^{\frac{1}{\log(r_j)^2}} \), where for each \( j \), \( r_j \) are realisations of a uniform random variable on the unit interval \([0, 1]\). A simulation of 1000 results gave \( I \approx 0.8710 \) while a sample of size 10,000 yielded \( I \approx 0.8854 \). Almost a quarter of a million simulations produced the estimate \( I \approx 0.8865 \). This compares to the exact value of \( I = 0.5\sqrt{\pi} \approx 0.8862 \). While it may take a large number of simulations to produce an accurate result, the technique is quite easy to implement, and can be used to estimate analytically intractable integrals.

The Monte Carlo estimate in (3) can also be used to estimate probabilities, with the choice of \( w(x) = I[x \in A] \), where \( I \) is an indicator function, defined by

\[
I[x \in A] = \begin{cases} 
1 & \text{if } x \in A; \\
0 & \text{otherwise.}
\end{cases}
\]

Let \( X \) be a random variable on \( \Omega \) with density \( f \). Suppose we want to estimate the probability \( P(X \in A) \), for some set \( A \). Then an application of (3) yields

\[
P(X \in A) = \int_\Omega I[x \in A] f(x)dx
\]

\[
\approx \frac{\sum_{j=1}^{m} I[x_j \in A]}{m},
\]

where the sample \( x_1, x_2, \ldots, x_m \) is IID generated from \( X \). To illustrate, consider the case where \( X \) is an exponential random variable with parameter 1, so \( f(x) = e^{-x}, x \geq 0 \). Suppose we want to estimate \( P(X \leq 1) \). It is not difficult to show that \( P(X \leq 1) = 1 - e^{-1} \approx 0.6321 \). Now \( A = [0, 1] \) and \( \Omega = [0, \infty) \), so a Monte Carlo estimator of this probability is \( p_{MC1} = \frac{1}{m} \sum_{j=1}^{m} I[x_j \leq 1] \), where the \( x_j \) are generated from the exponential distribution. However, there is an alternative estimator, since \( P(X \leq 1) = \int_0^1 e^{-x}1dx = 1 \) is a density on the unit interval \([0, 1]\). Thus an alternative estimator is \( p_{MC2} = \frac{1}{m} \sum_{j=1}^{m} e^{-x_j} \), where the \( x_j \) are random numbers in the interval \([0, 1]\).

A difficulty with the Monte Carlo approach is that it can take a considerable number of simulations to achieve a reasonable level of accuracy. To illustrate this, recall that the Central Limit Theorem (CLT) states that for a series \( \{\eta_j\} \) of IID random variables with finite first and second moments,

\[
\lim_{m \to \infty} \frac{\frac{1}{m} \sum_{j=1}^{m} \eta_j - \mu}{\frac{1}{\sqrt{m}}} \overset{d}{=} N(0, 1),
\]

where \( \mu = E[\eta] \) and \( \sigma^2 = V[\eta] \). Using a moment generating function expansion of the sum of random variables in (5), it can be shown that the rate of convergence is roughly

\footnote{This is because an exponential random variable with parameter 1 can be simulated by \(-\log(r)\), with \( r \) a random number between 0 and 1.}
of the order $\frac{1}{\sqrt{m}}$. Hence, for an accuracy of $10^{-6}$, the number of Monte Carlo runs can be roughly of the order of $10^{12}$. This can result in considerably long run times. To counteract this, the techniques of Importance Sampling (IS) are used. This involves generating the random variables in (4) from a biasing distribution, $f_*$. The latter is chosen to emphasize points in the estimation that are more important to the simulation. The effect of this is that the simulation is using less points than the standard Monte Carlo estimator. To balance the estimation, and to produce an unbiased estimate, the estimator is weighted at each point. Thus, an IS estimate of the probability in (4) is given by

$$p_{IS} = \frac{1}{m} \sum_{j=1}^{m} I[z_j \in A]W(z_j),$$

(6)

where the $z_j$ are generated from a distribution with biasing density $f_*$, and $W$ is a weighting function. To produce an unbiased estimator, it can be shown that $W(z) = \frac{f(z)}{f_*(z)}$. As an example, consider the problem again of estimating $P(X \leq 1)$ where $X$ is an exponential random variable with parameter 1. Choose as biasing distribution a uniform random variable on the interval [0, 1]. Now $f(z) = e^{-z}$ and $f_*(z) = 1$ on the unit interval, and so the IS estimator is $p_{IS} = \frac{1}{m} \sum_{j=1}^{m} e^{-z_j}$. This is the same as the second Monte Carlo estimator $p_{MC2}$ considered previously.

Before considering IS estimation in more detail, we firstly derive an analytical expression for the probability of false alarm in a CA-CFAR, in the case of Gaussian clutter and noise.
2 A Cell Averaging CFAR Model

In this section we will construct a model for a CA-CFAR detector operating in the presence of Gaussian clutter and noise. A useful relationship between the CA-CFAR probability of false alarm, and the threshold multiplier will be derived. The latter will provide a useful mechanism for testing the efficiency and accuracy of Monte Carlo estimators considered in this report.

2.1 From Signal Return to Envelope Detection Output

To give a context to the analysis to follow, we briefly consider the radar signal return and its characteristics under some assumptions. [Levanon 1988] and [Minkler and Minkler 1990] both contain good discussions on the processing of the signal return before it is passed into the CFAR detector. In a simulation of a CA-CFAR system, one can generate results from the point of view of simulating the process after square law detection, since the statistics introduced previously were defined in this part of the CFAR model.

A case in which the threshold parameter $\tau$ can be obtained analytically in terms of the false alarm probability, is when the clutter/noise is assumed to be Gaussian. In this case, the resulting random variables from the square law detector are exponential. We now demonstrate this, based upon the approach of [Levanon 1988], and also derive an expression for the probability of false alarm as a function of the threshold parameter.

The following is for single pulse detection. The transmitted signal is a sine wave of duration $\psi$ and frequency $\omega_C$. The returned signal will be a phased shifted version of the originally transmitted signal, except with the addition of noise and clutter. The radar return is passed to a narrow bandpass filter, with centre frequency $\omega_C$. We assume this filter has a rectangular response with bandwidth $f_B$. Then assuming that $f_B > \frac{1}{\psi}$, the returned pure signal can be described by

$$s_0(t) = A\cos(\omega_C t - \phi_S) = a\cos(\omega_C t) + b\sin(\omega_C t), \quad (7)$$

where $\phi_S = \arctan \left( \frac{b}{a} \right)$ is the phase shift of the signal, and the amplitude $A = \sqrt{a^2 + b^2}$.

When Gaussian noise is passed through a narrow bandpass filter, the output can be written as

$$n_0(t) = X(t)\cos(\omega_C t) + Y(t)\sin(\omega_C t), \quad (8)$$

where $X(t)$ and $Y(t)$ are both independent Gaussian random variables with mean 0 and equal variance, $\sigma^2$. By combining (7) and (8), the radar signal return at the detector can then be written as

$$\zeta(t) = s_0(t) + n_0(t)$$

$$= (a + X(t))\cos(\omega_C t) + (b + Y(t))\sin(\omega_C t)$$

$$= R(t)\cos(\omega_C t + \Phi(t)), \quad (9)$$
where

\[ R(t) = \sqrt{(a + X(t))^2 + (b + Y(t))^2} \quad \text{and} \quad \Phi(t) = \arctan \left[ \frac{b + Y(t)}{a + X(t)} \right]. \tag{10} \]

It is desirable to obtain an understanding of the distribution of the amplitude \( R(t) \) in equation (10), since amplitude characteristics of return signals are used in the detection of targets. Note that the square of \( R(t) \) is a sum of the squares of two Gaussian random variables: let \( X_1 = a + X(t) \) and \( Y_1 = b + Y(t) \). Using the joint distribution of \((X_1, Y_1)\), and changing to polar coordinates \((R, \Phi)\) defined through the transformations implied by (10), it can be easily shown that the transformed joint density is

\[ f_{(R, \Phi)}(r, \phi) = \frac{r}{2\pi \sigma^2} \exp \left[ -\frac{r^2 + a^2 + b^2 - 2ra \cos \phi - 2rb \sin \phi}{2\sigma^2} \right], \tag{11} \]

where \( \sigma^2 \) is the common variance of \( X_1 \) and \( Y_1 \). To obtain the probability density function (pdf) of the amplitude \( R(t) \), we integrate the density (11) over all phases to obtain

\[ f_R(r) = \int_0^{2\pi} f_{(R, \Phi)}(r, \phi) \, d\phi = \frac{r}{\sigma^2} \exp \left[ -\frac{r^2 + 2rA}{2\sigma^2} \right] I_0 \left[ \frac{rA}{\sigma^2} \right], \tag{12} \]

where \( I_0 \) is the modified Bessel function of order zero\(^5\). Note from (12) that in the case of no signal present in the return, so that \( A = 0 \), the pdf of the amplitude of the return is Rayleigh with parameter \( \sigma \).

The amplitude is then passed through a square law detector, so we are interested in the resulting distribution from squaring \( R \). Consider the normalised variable \( Z = R^2 \). It can be shown\(^6\) that its pdf is given by

\[ f_Z(z) = \frac{1}{2\sigma^2} \frac{1}{\left(1 + \frac{A^2}{\sigma^2}\right)} \exp \left[ -\frac{z}{2\sigma^2} \left(1 + \frac{A^2}{\sigma^2}\right) \right]. \tag{13} \]

Hence the radar returns, processed through the square law detector, are exponentially distributed. The ratio \( \frac{A^2}{\sigma^2} := S \) is a measure of the signal to noise strength. Thus, based upon (13), we can simulate the CFAR system with Gaussian noise and clutter, by simulating exponentially distributed variables. In the case of no target in the CUT, the statistic for it will have an exponential distribution with parameter \( \frac{1}{2\sigma^2} \), since the amplitude \( A = 0 \). When there is a target in the CUT, this distribution is the same, except the parameter becomes \( \frac{1}{2\sigma^2(1+S)} \), with the parameter \( S \) controlling the significance of the target return.

In the next section we use this result to formulate a statistical test for the presence of a target in Gaussian noise.

\(^5\)See [Levanon 1988] for an integral expression for this. This integral depends on the initial phase \( \phi_0 \).

\(^6\)See [Levanon 1988], problem 3.1 for the details.
2.2 Exponentially Distributed Post-Envelope Returns

To illustrate the CFAR detection scheme for a single pulse, we consider the simple case formulated in [Gandhi and Kassam 1994]. In the latter, it is shown that the CA-CFAR detector is optimal for the detection of Swerling 1 type targets. This is in the case where Gaussian noise is passed into a square law envelope detector, equivalent to homogeneous exponential clutter. The set up of the previous Subsection provides the framework for this approach. Here we suppose the radar signal returns, already processed by a square law detector, are IID exponential random variables, motivated from the above analysis.

Suppose \( X_1, X_2, \ldots, X_m \) is a sample of such a return of pure clutter/noise observations, so that each \( X_j \) has an exponential distribution, with parameter \( \frac{1}{\mu} \). This implies the mean of this distribution is \( \mu \). With reference to (13), and the comments proceeding (12), it follows that \( \mu = 2\sigma^2 \) (since \( A = 0 \)). The latter is the mean noise plus clutter power. The CUT statistic \( X_0 \) is either noise and clutter, and so is identically distributed to the sample above, or a radar target return in the presence of noise and clutter. We assume in the latter case that \( X_0 \) has an exponential distribution with parameter \( \frac{1}{\mu(1+S)} \), where \( S \) is the signal to noise ratio of a Swerling 1 target, as defined previously, and motivated from (13).

Hence, to test for the presence of a target, we perform a test of \( H_0 : \rho = \mu \) against the alternative \( H_1 : \rho = \mu(1 + S) \), assuming the test observation \( X_0 \) has an exponential distribution with parameter \( \frac{1}{\rho} \). The form of the test is to reject \( H_0 \) if

\[
\frac{X_0}{\sum_{j=1}^{m} X_j} > \frac{\tau}{m},
\]

where \( \tau \) is a threshold multiplier. Thus the adaptive threshold for this CA-CFAR scheme is \( \frac{\tau}{m} \sum_{j=1}^{m} x_j \), based upon a realised sample \( x_1, x_2, \ldots, x_m \). The probability of false alarm \( P_{FA} \) can be determined through

\[
P_{FA} = P(\text{reject } H_0 | H_0 \text{ is true})
\]

\[
= P \left( X_0 > \frac{\tau}{m} \sum_{j=1}^{m} X_j \right | H_0 \text{ is true} \).
\]

Under \( H_0 \), \( X_0 \) has an exponential distribution with parameter \( \frac{1}{\mu} \). The sum \( \sum_{j=1}^{m} X_j \) of similarly defined independent exponential random variables has a Gamma Distribution\(^7\) \( \gamma \left( m, \frac{1}{\mu} \right) \). Such a distribution has density

\[
f_Z(z) = \frac{1}{\mu\Gamma(m)} \left( \frac{z}{\mu} \right)^{m-1} e^{-\frac{z}{\mu}},
\]

where \( \Gamma \) is the Gamma function, defined as the integral

\[
\Gamma(m) = \int_{0}^{\infty} x^{m-1} e^{-x} dx = (m-1)!,
\]

\(^7\)See [Beaumont 1980], Appendix 1, noting that an exponential variable is also Gamma.
where the latter equality holds only if \( m \) is a nonnegative integer. Here the random variable
\[
Z = \sum_{j=1}^{m} X_j.
\]
Hence, applying (15) to (14), we obtain
\[
P_{FA} = P \left( X_0 > \frac{\tau}{m} Z \middle| H_0 \right)
\]
\[
= \int_0^\infty P \left( X_0 > \frac{\tau}{m} z \middle| H_0 \right) f_Z(z) \, dz
\]
\[
= \frac{1}{\mu \Gamma(m)} \int_0^\infty \left( \frac{z}{\mu} \right)^{m-1} e^{-\frac{z}{\mu}(1+\frac{\tau}{m})} \, dz
\]
\[
= \frac{1}{(1 + \frac{\tau}{m})^m}. \tag{16}
\]
Thus, for a given threshold multiplier \( \tau \) and number of cells \( m \), the probability of false alarm can be obtained exactly, as given by (16). This result provides a means of testing the performance of Monte Carlo estimators for both the probability of false alarm and the threshold, in the case of Gaussian clutter and noise.

From the CA-CFAR detector point of view, a target in the CUT is declared present if
\[
x_0 > \left( P_{FA}^{-\frac{1}{m}} - 1 \right) \sum_{j=1}^{m} x_j.
\]
It is also interesting to note that the probability of detection, \( P_D \), can be found in a similar manner. Noting that this is the probability that \( H_0 \) is rejected when \( H_1 \) is true, it follows that
\[
P_D = P \left( X_0 > \frac{\tau}{m} \sum_{j=1}^{m} X_j \middle| X_0 \overset{d}{=} \text{Exp} \left( \frac{1}{\mu(1+S)} \right) \right)
\]
\[
= \frac{1}{\mu^m \Gamma(m)} \int_0^\infty \frac{1}{\Gamma(m)} e^{-\frac{\tau}{m(1+S)} z^{m-1} e^{-\frac{z}{\mu}}} \, dz
\]
\[
= \frac{1}{(1 + \frac{\tau}{m(1+S)})^m}. \tag{17}
\]
Observe that both the probability of false alarm and detection are independent of the mean clutter and noise parameter \( \mu \). The probability of detection (17) has been included for completeness. We will not investigate it further in this report.

We now turn to Monte Carlo techniques for the estimation of false alarm probabilities in CA-CFAR detectors.
3 Estimation of False Alarm Probabilities

The purpose of this Section is to illustrate the application of Monte Carlo methods to the estimation of false alarm probabilities, in a CA-CFAR context. The problem is formulated as a statistical test, and standard Monte Carlo and Importance Sampling estimators are introduced. Two biasing densities are then considered, as well as simulation gain analysis. Finally, an alternative method of performing IS simulations is introduced. Simulation studies of the performance of these estimators will be the subject of the final Section.

3.1 Standard Monte Carlo Methods

Suppose we have \(m\) observational cells, with observational cell statistics \(x_1, x_2, \ldots, x_m\). Let \(x_0\) be the statistic for the cell under test (CUT). For ease of notation, let \(x = (x_0, x_1, \ldots, x_m)\). Additionally, let \(\tau_0 = \frac{1}{m}\). We perform a test of \(H_0:\) no target in the CUT, against the alternative \(H_1:\) there is a target in the CUT. The test statistic is \(D(x) = x_0 - \tau_0 \sum_{j=1}^{m} x_j\). Based upon the analysis in Subsection 1.2, the form of the test is to reject \(H_0\) if \(D(x) > 0\). The region where we reject \(H_0\) is called the critical region. The CA-CFAR adaptive threshold is \(\tau_0 \sum_{j=1}^{m} x_j\), with \(\tau_0\) called a threshold multiplier. The threshold multiplier is used to control the false alarm probability, so that the rate of false alarms is constant in the CFAR process.

The probability of false alarm \(P_{FA}\) is the probability of declaring a target present in the CUT, when there is actually no target present. Mathematically, this can be written

\[
P_{FA} = P(\text{reject } H_0 | H_0 \text{ is true } ) = P(D(X) > 0 | H_0 \text{ is true } ),
\]

where \(X = (X_0, X_1, \ldots, X_m)\). Assume \(X\) has density \(f\) on \(\Omega\) under \(H_0\). Typically, for the purposes of modelling, we assume the observational cell statistics are independent and identically distributed. Hence, under \(H_0\), the marginal statistics in vector \(X\) are IID clutter statistics. Thus the probability of false alarm can be written

\[
P_{FA} = \int_{\Omega} I[D(x) > 0] f(x) dx
\]

where

\[
\mu(x) = I[D(x) > 0] = \begin{cases} 1 & \text{if } D(x) > 0; \\ 0 & \text{otherwise.} \end{cases}
\]

and is called an indicator function. Hence, applying the Strong Law of Large Numbers (1), and using the approximation (3), a Monte Carlo estimate of the probability of false alarm\(^8\)

\[
\hat{p}_{MC} = \frac{1}{N} \sum_{j=1}^{N} \mu(x_j),
\]

\(^8\)Throughout, a hat over a variable will indicate that it is an estimate or estimator. Thus, \(\hat{x}\) is an estimate, and \(\hat{X}\) is an estimator.
where \( x_j = (x_0^j, x_1^j, \ldots, x_m^j) \) is generated from a distribution with density \( f \). Here, the notation \( x_i^j \) refers to the \( i \)th element of the \( j \)th simulation vector. \( N \) is the number of simulations/recursions in the Monte Carlo estimate. The estimator version of this is

\[
\hat{P}_{MC} = \frac{1}{N} \sum_{j=1}^{N} \mu(X_j),
\]

with \( X_j \) a random vector version of \( x_j \). It is not difficult to show that \( \hat{P}_{MC} \) is an unbiased estimator of \( P_{FA} \), with variance that decreases to zero as the sample size increases:

\[
E[\hat{P}_{MC}] = P_{FA} \quad \text{and} \quad V[\hat{P}_{MC}] = \frac{P_{FA}(1 - P_{FA})}{N}.
\]

As pointed out previously, the difficulty with this estimator is its rate of convergence. Thus we look at alternative ways of estimating false alarm probabilities.

### 3.2 Importance Sampling Estimators

From a purely mechanical perspective, the idea of importance sampling is to generate the sample vectors \( x_j \) in the estimate (19) from a different density. In a simulation, certain values of the input variables will have more importance to the estimation than others. In IS, one chooses a new density to emphasize these important points. Since one is then sampling from a smaller population, it is possible to gain a reduction in variance.

To clarify, consider the problem of integral estimation via Monte Carlo methods, as outlined in Section 1.3. In many cases, a function of interest has most of its mass located within a certain region, such as the function \( f(x) = e^{-x^2} \), for \( x \in \mathbb{R} \). In this particular case, the function is bell shaped with most of its mass lying within an interval symmetric about the origin. Importance sampling attempts to increase the density of points within these important regions.

The effect on the estimator (20), when simulating from a different density, is that it is no longer an unbiased estimator for the probability of false alarm. To address this issue, each sample point in the series of (20) is weighted to produce a centered estimate. The new distribution used for simulation is called a biasing distribution, and its density is referred to as a biasing density. We now turn to a mathematical treatment of these ideas.

Suppose we sample from a biasing distribution with density \( f_* \). Then the modified estimator for probability of false alarm is

\[
\hat{P}_{IS} = \frac{1}{M} \sum_{j=1}^{M} \mu(Z_j)W(Z_j),
\]

where \( Z_j = (Z_0^j, Z_1^j, \ldots, Z_m^j) \) is generated from a distribution with density \( f_* \), and \( W(z) \) is a weighting function. In order to produce an unbiased estimator for \( P_{FA} \), it can be shown that

\[
W(z) = \frac{f(z)}{f_*(z)}.
\]
With this choice, the variance of estimator (22) is

$$V_* (\hat{P}_{IS}) = \frac{1}{M} \left[ \int_{\Omega} \mu(x) \frac{f^2(x)}{f_*(x)} dx - P_{FA} \right].$$  \hspace{1cm} (23)

Observe that in the case where the biasing density is

$$f_*(x) = P_{FA}^{-1} \mu(x) f(x),$$ \hspace{1cm} (24)

the variance in (23) is reduced to zero. This is known as the optimal solution. It is, however, not implementable because it depends on the unknown probability of false alarm. However, its form indicates the characteristics of a good suboptimal biasing density. It suggests that a suboptimal biasing density should be proportional to the original density $f$. Additionally, all of its mass is concentrated on the critical region. A function constructed according to these specifications must then be weighted by a constant to produce a probability density.

Using these considerations, a number of authors have produced suboptimal biasing densities [Gerlach 1999, Orsak 1993 and Orsak and Aazhang 1989]. Two such densities are now considered, in the context of the CA-CFAR model we are examining.

### 3.2.1 Ad Hoc Biasing Density

A simple and widely used biasing density is the so-called Ad Hoc or variance scaled density. It is often used because of its simplicity, but does not always result in consistently improved results. Scaling the input vector of samples by a constant produces the biasing density. Although not attributed to a single author, it is discussed in a number of publications [Gerlach 1999, Smith, Shafi and Gao 1997, and Srinivasan 2000]. In view of Section 2, assume the clutter observations are IID exponential with parameter $\frac{1}{\mu}$ for some fixed $\mu$. This means the clutter mean value is also $\mu$. Hence, the joint density of $(X_0, X_1, \ldots, X_m)$ under $H_0$ is

$$f(x) = \prod_{j=0}^{m} f_j(x_j) = \frac{1}{\mu^{m+1}} e^{-\frac{1}{\mu} \sum_{j=0}^{m} x_j}. \hspace{1cm} (25)$$

The biasing density, for some $\alpha < 1$, is

$$f_*(x) = \alpha^{m+1} f(\alpha x) = \prod_{j=0}^{m} \left( \frac{\alpha}{\mu} \right) e^{-\frac{\alpha}{\mu} x_j}. \hspace{1cm} (26)$$

Hence the biasing distribution has exponential marginals, with parameter $\frac{\alpha}{\mu}$, due to the fact that it is a product of independent exponential random variable densities. The weighting function is

$$W(x) = \frac{f(x)}{f_*(x)} = \left( \frac{1}{\alpha} \right)^{m+1} e^{-\left[ \frac{1-\alpha}{\mu} \right] \sum_{j=0}^{m} x_j}. \hspace{1cm} (27)$$

Now it can be shown, by applying (25) and (26) to (23), that for this Ad Hoc biasing density,

$$V_* (\hat{P}_{IS}) = \frac{1}{M} \left( \frac{1}{\alpha(2 - \alpha)} \right)^{m+1} P_{FA} - P_{FA}^2. \hspace{1cm} (28)$$
Since $\alpha < 1$, $MV_\alpha (P_{IS}) > P_{FA} - P_{FA}^2 = NV (P_{MC})$, so that for the same number of simulations, the IS estimator has larger variance than the standard MC estimator.

The simulation gain $\Gamma$ gives a measure of the number of standard Monte Carlo simulations that are required to perform with the same level of variation as an IS estimator. We are thus looking for cases where the gain is large, since this will imply the IS estimator will perform, with the same level of accuracy as a standard Monte Carlo estimator, for less simulation runs. Mathematically, the gain is obtained by equating the variance expressions in (21) and (23), and solving for the ratio of simulation runs, namely $\frac{N}{M}$. For the same level of variance, the simulation gain in the current context is

$$\Gamma = \frac{N}{M} = \frac{1 - P_{FA}}{\left[\frac{1}{\alpha (2 - \alpha)}\right]^{m+1} - P_{FA}}, \quad (29)$$

and since $0 < \alpha < 1$, it follows that $\Gamma < 1$. Thus, for the same level of variation, the IS Ad Hoc approach does not reduce the number of simulation runs. This does not mean that this approach is entirely useless. In some cases, an Ad Hoc biasing density has been found to provide improvements in estimation, when compared to the standard Monte Carlo estimator.

### 3.2.2 Chernoff Biasing Density

The Chernoff IS method was introduced in [Gerlach 1999]. In this case, the indicator function in the optimal solution (24) is replaced by an exponential function of the detection statistic. It then turns out that the simulation gain is inversely proportional to the simulation reduction factor. The latter is referred to as a Chernoff-like bound [see Van Trees 1971]. In some cases, this can result in quite substantial simulation savings.

Observe that, for a fixed $\lambda > 0$, $e^{\lambda D(x)} > \mu(x)$. Thus we introduce a biasing density

$$f_*(x) = \frac{1}{P_C(\lambda)} f(x) e^{\lambda D(x)}, \quad (30)$$

for some $\lambda > 0$, such that $P_C(\lambda) = \int \Gamma f(x) e^{\lambda D(x)} dx < \infty$. It is not difficult to see that $P_{FA} < P_C(\lambda)$. Hence we choose a $\lambda$ to minimise $P_C(\lambda)$. Assuming the same clutter model (25), and using the fact that $D(x) = x_0 - \tau_0 \sum_{j=1}^{m} x_j$, it can be shown that

$$\frac{1}{P_C(\lambda)} = \mu^{m+1} \left[\frac{1}{\mu} - \lambda \right] \left[\frac{1}{\mu} + \lambda \tau_0 \right]^m \quad (31)$$

and

$$f_*(x) = \left(\left[\frac{1}{\mu} - \lambda \right] e^{-\left[\frac{1}{\mu} - \lambda \right] x_0} \prod_{j=1}^{m} \left[\frac{1}{\mu} + \lambda \tau_0 \right] e^{-\left[\frac{1}{\mu} + \lambda \tau_0 \right] x_j} \right). \quad (32)$$

Hence the biasing distribution of $X = (X_0, X_1, \ldots, X_m)$ has IID marginals, with $X_0$ having an exponential distribution with parameter $\frac{1}{\mu} - \lambda$, and each $X_j$ having an exponential distribution also, but with parameter $\frac{1}{\mu} + \lambda \tau_0$, for $j \in \{1, 2, \ldots, m\}$.
The weight function for the IS estimator can be shown to be

$$ W(x) = e^{\lambda \left[ -x_0 + \tau_0 \sum_{j=1}^{m} x_j \right]} \frac{1}{[1 - \mu \lambda][1 + \mu \lambda \tau_0]^m}. \tag{33} $$

By applying (30) to (23), it can be shown that

$$ V_*(\hat{P}_{IS}) \leq \frac{1}{M} \left[ P_{FA} P_C(\lambda) - P_{FA}^2 \right], $$

and so for the same level of variance as the standard MC estimator, we have

$$ \Gamma = \frac{N}{M} \approx \frac{1}{P_C(\lambda)}. \tag{34} $$

Thus, depending on whether $P_C(\lambda)$ is sufficiently small or not, there is a possibility of improved performance with this IS biasing density. [Gerlach 1999] points out that, in some cases, $P_C(\lambda)$ is within a few orders of magnitude of $P_{FA}$, and so the simulation gain (34) can be large. We now demonstrate this for the particular case of CA-CFAR under consideration.

In view of (34), we want to minimise $P_C(\lambda)$. It can be shown a local minimum exists and occurs at

$$ \lambda_{opt} = \frac{m \tau_0 - 1}{\mu \tau_0 [m + 1]} \tag{35} $$

Substituting (35) into the Chernoff factor (31), we obtain

$$ P_C(\lambda_{opt}) = (\tau_0 + 1)^{-m} \left[ 1 + \frac{1}{m} \right]^m \left[ \tau_0 (m + 1) \right] \frac{\tau_0 (m + 1)}{\tau_0 + 1}. \tag{36} $$

Observe that (36) is independent of the clutter parameter $\mu$. Hence, with reference to (16), we can write (36) as

$$ P_C(\lambda_{opt}) = P_{FA} \left( 1 - P_{FA}^m \right) (m + 1) \left[ 1 + \frac{1}{m} \right]^m, \tag{37} $$

where we have used the fact\(^9\) that $\tau_0 = \frac{\tau}{m}$. An application of (37) to (34), together with a geometric series expansion\(^10\), establishes that the simulation gain will be

$$ \Gamma = P_{FA}^{-1} \left( 1 - P_{FA}^m \right)^{-1} \frac{1}{m + 1} \left[ 1 + \frac{1}{m} \right]^{-m} $$

$$ = P_{FA}^{-1} \left( 1 + P_{FA}^m + P_{FA}^{2m} + \ldots \right) \frac{1}{m + 1} \left[ 1 + \frac{1}{m} \right]^{-m} \tag{38} $$

Thus the gain can be of the order of the reciprocal of the probability of false alarm, and hence can be quite large. Although the gain (38) can provide substantial simulation savings, it turns out that an alternative scheme provides a larger gain over all other IS estimators.

---

\(^9\)See Section 3.1.

\(^10\)Recall that $1 + r + r^2 + r^3 + \ldots = \frac{1}{1-r}$, provided $|r| < 1$. 

---

DSTO–TR–1624
3.3 The G-Function Method

We now consider an alternative approach to Importance Sampling. A number of authors have suggested modification of the form of the IS estimator (22). [Smith and Orsak 1995 and Srinivasan 2000] are examples, and here we will focus on the G-Function method of [Srinivasan 2000]. This approach involves biasing only the observational cell statistics, and then estimating a mean value of a function of a conditional cumulative distribution function. Although this method may seem somewhat complicated, its appeal is that it can be shown that for the same number of simulation trials, it will always perform better than any other IS technique. [Srinivasan 2000] contains a mathematical proof of this.

We begin by returning to the basic expression for the probability of false alarm. Let

\[ Y = \sum_{j=1}^{m} X_j. \]

Then observe that

\[
P_{FA} = P(X_0 > \tau_0 Y \mid H_0) = 1 - E(I[X_0 \leq \tau_0 Y] \mid H_0, Y) = 1 - E_0(I[X_0 \leq \tau_0 Y] \mid H_0, Y),
\]

where \( E_0 \) is expectation under \( H_0 \). Define the function \( g(x) = 1 - F_{X_0\mid H_0, Y}(x) \), where \( F_{X_0\mid H_0, Y}(x) \) is the conditional cumulative distribution function of the statistic of the CUT, given \( H_0 \) is true and given \( Y \). Hence

\[
P_{FA} = E_0[g(\tau_0 Y)],
\]

suggesting that IS estimation can be performed by biasing \( Y \) and estimating the expectation of \( g(\tau_0 Y) \).

The corresponding IS estimator is

\[
\hat{P}_{IS} = \frac{1}{M} \sum_{j=1}^{M} g(\tau_0 Y_j) W\left((Y_1^j, \ldots, Y_m^j)\right),
\]

where \( Y_j = \sum_{i=1}^{m} Y_i^j \), and \( (Y_1^j, \ldots, Y_m^j) \) is a vector of length \( m \) generated from a biasing distribution with density \( f_\theta \). The weight function in (40) is the same as before, except it is now a function of the observational cell statistics only. The function \( f \) is also now a density of \( m \) variables, since the CUT is not biased.

The variance of this estimator is given by

\[
V_*(\hat{P}_{FA}) = \frac{1}{M} \left[ E_\theta[g^2(\tau_0 Y) W((Y_1, \ldots, Y_m))] - P_{FA}^2 \right].
\]

Assume a biasing density that involves a single biasing parameter \( \theta \). As an example, this could be an Ad Hoc biasing density, as considered previously. Let

\[
I_g(\theta) = E_\theta[g^2(\tau_0 Y) W((Y_1, \ldots, Y_m), \theta)].
\]
In view of (41) and (42), we want to find a \( \theta \) that minimises \( I_g(\theta) \). Hence want to solve for \( \frac{d}{d\theta} I_g(\theta) = 0 \). This will often be quite difficult to do analytically. Hence, a Newton-Raphson scheme can be implemented, so that starting with an initial approximation \( \theta_0 \),

\[
\theta_{j+1} = \theta_j - \zeta \frac{d}{d\theta} I_g(\theta_j), \quad \text{for } j \in \{1, 2, \ldots\}. \tag{43}
\]

The parameter \( \zeta \) is used to speed up the convergence, but can be set to 1. In practice, we will need to estimate \( I_g(\theta) \) and its derivatives. This can also be done using Monte Carlo simulations. Note that

\[
I_g(\theta) = E \left[ g^2(\tau_0 y) W^2((y_1, y_2, \ldots, y_m), \theta) f_*(y) dy \right] = E \left[ g^2(\tau_0 y) W((y_1, y_2, \ldots, y_m), \theta) f(y) dy \right] = E[g^2(\tau_0 y) W((y_1, y_2, \ldots, y_m), \theta)]. \tag{44}
\]

Hence, differentiating the last equality in (44), and changing variables in terms of the expectation,

\[
\frac{d}{d\theta} I_g(\theta) = E \left[ g^2(\tau_0 y) \frac{\partial}{\partial \theta} W((y_1, y_2, \ldots, y_m), \theta) \right] = \int_{\Gamma} g^2(\tau_0 y) \left\{ \frac{\partial}{\partial \theta} W((y_1, y_2, \ldots, y_m, \theta) f_*(y) dy \right\} W((y_1, \ldots, y_m), \theta) = E_* \left[ g^2(\tau_0 y) \left\{ \frac{\partial}{\partial \theta} W((y_1, y_2, \ldots, y_m, \theta) \right\} W((y_1, y_2, \ldots, y_m), \theta) \right]. \tag{45}
\]

Similarly, it can be shown that

\[
\frac{d^2}{d\theta^2} I_g(\theta) = E_* \left[ g^2(\tau_0 y) \left\{ \frac{\partial^2}{\partial \theta^2} W((y_1, y_2, \ldots, y_m, \theta) \right\} W((y_1, \ldots, y_m), \theta) \right]. \tag{46}
\]

Monte Carlo techniques can be applied to estimate (45) and (46). Simulations are based upon

\[
\frac{d}{d\theta} I_g(\theta) = \frac{1}{\kappa_1} \sum_{j=1}^{\kappa_1} g^2(\tau_0 y_j) \left\{ \frac{\partial}{\partial \theta} W((y_1^j, y_2^j, \ldots, y_m^j, \theta) \right\} W((y_1^j, \ldots, y_m^j), \theta)
\]

\[
\frac{d^2}{d\theta^2} I_g(\theta) = \frac{1}{\kappa_2} \sum_{j=1}^{\kappa_2} g^2(\tau_0 z^j) \left\{ \frac{\partial^2}{\partial \theta^2} W((z_1^j, z_2^j, \ldots, z_m^j, \theta) \right\} W((z_1^j, \ldots, z_m^j), \theta) \tag{47}
\]

where, for each \( j \), the vectors \((y_1^j, \ldots, y_m^j)\) and \((z_1^j, \ldots, z_m^j)\) are independently generated from a distribution with the biasing density \( f_* \). These are standard Monte Carlo estimates, even though they are based upon the biasing density, due to the fact that they are estimating an expectation with respect to this density. In principle, IS techniques could also be employed here, but it makes the simulation much more complicated.
Thus, an optimum $\theta$ is estimated via the recursion
\[
\theta_{j+1} = \theta_j - \zeta \frac{d}{d\theta} I_g(\theta_j), \quad \text{for } j \in \{1, 2, \ldots\}.
\]

Figure 2 in the Appendix contains plots of successive recursions of the scheme (48). It is for the case of Gaussian clutter, with square law detection. The three cases considered are $m = 4$ and $P_{FA} = 10^{-4}$, $m = 5$ and $P_{FA} = 10^{-6}$, and $m = 6$ and $P_{FA} = 10^{-8}$. The parameter $\zeta = 1$, and the number of recursions refers to the parameter $j$ in (48). The derivatives in (48) were estimated using (47). The plots show that after approximately 10 recursions, a reasonable estimate of $\theta$ is obtained in each case.

The simulation gain is
\[
\Gamma = \frac{P_{FA} - P_{FA}^2}{I_g(\theta) - P_{FA}^2},
\]
and as before, we can estimate $I_g(\theta)$ via a Monte Carlo simulation:
\[
\hat{I}_g(\theta) = \frac{1}{K_3} \sum_{j=1}^{K_3} g^2(\tau_0 w^j) W^2((w^j_1, \ldots, w^j_m), \theta),
\]
with $(w^j_1, \ldots, w^j_m)$ generated from $f_*$. Thus the G-Function method is quite involved, but the level of simulation improvement can be quite considerable, as will be demonstrated in Section 4.

### 3.3.1 Example: CA-CFAR with Gaussian Clutter

We now turn to the CA-CFAR model under investigation, where we assume the clutter and noise is IID exponentially distributed with parameter $\frac{1}{\mu}$. Since we are looking at developing techniques to compare CFAR schemes in known clutter conditions, we assume this parameter is known and fixed.

The clutter density is
\[
f(x) = \prod_{j=1}^{m} f_j(x_j) = \frac{1}{\mu^m} e^{-\frac{1}{\mu} \sum_{j=1}^{m} x_j}.
\]
[Srinivasan 2000] suggests using a single parameter Ad Hoc biasing density: the marginal density will be
\[
f_*^j(z) = \frac{1}{\mu \theta} e^{-\frac{z}{\mu \theta}}, \quad \text{with } 0 < \theta < 1,
\]
and so the joint density is
\[
f_*(y, \theta) = \prod_{j=1}^{m} f_*^j(y^j, \theta) = \frac{1}{(\mu \theta)^m} e^{-\frac{\sum_{j=1}^{m} y_j}{\mu \theta}}.
\]
Since we are assuming IID clutter observations, independent of the CUT, the $g$-function is simply
\[
g(x) = 1 - F_{X_0|Y,H_0}(x) = e^{-\frac{x}{\mu}}.
\]
The weight function can be shown to be
\[ W((y_1, y_2, \ldots, y_m), \theta) = \theta^m e^{-[1-\theta] \frac{1}{\mu} \sum_{j=1}^{m} y_j}. \] (54)

Finally, the simulation gain is
\[ \Gamma = \frac{P_{F_A} - P_{F_A}^2}{\left[ \frac{\theta}{2(m+1) - \theta} \right]^m - P_{F_A}^2}. \] (55)

Figure 3 in the Appendix shows the simulation gain (55), as a function of \( \theta \), in a typical scenario. There is clearly a \( \theta \) that will provide a huge simulation gain. With the corresponding choice for \( \theta \), this implies that for the same level of variation, this IS estimator will require significantly less simulation runs.

The next section contains an analysis of some simulations of the IS estimators introduced in this Section.
4 Simulation Comparison of Estimators

In this section we briefly examine four simulations of the estimators introduced in Section 3. Mathematical analysis implies that the G-function estimator will have superior performance in the context of interest. Hence we expect it to provide good estimates of the false alarm probability for less simulation recursions.

Each G-function estimate has been produced with an independent estimate of the biasing parameter \( \theta \). In the Newton-Raphson scheme (48), \( \zeta \) has been set to 1, and the derivative estimates (47) have been generated with 100 recursions. That is, \( \kappa_1 = \kappa_2 = 100 \).

Figures 4–7 in the Appendix contain a sample of simulations of the three estimators. Figure 4 contains the result of 15 simulations of each estimator. In each case, 10 recursions are used to generate the estimate. The Ad Hoc density parameter was chosen to be \( \alpha = 0.05 \). Also, \( m = 5 \) and \( P_{FA} = 10^{-6} \), and the latter is plotted on the graph at each simulation point, to provide a comparison of the estimates. The Ad Hoc estimates are all zero, because 10 recursions are not sufficient for this estimator. Simulation experiments showed that it can take an enormous number of simulations for this estimator to produce a non-zero estimate. The Chernoff estimator performs quite well, especially given the small recursion size used. However, the G-function estimator has superior performance.

Figure 5 is a comparison of the estimators in the case where \( m = 4 \) and \( P_{FA} = 10^{-4} \). The Ad Hoc density parameter is \( \alpha = 0.9 \), and again, 10 recursions are used to generate the estimates. As in the previous simulation, we see the same behaviour in the estimators.

Figure 6 is a comparison of estimators for larger recursions. In this case, the number of recursions used to generate the estimates is 100. In this case, we set \( m = 5 \), \( P_{FA} = 10^{-3} \) and \( \alpha = 0.05 \). The same behaviour in the estimators is repeated.

The final simulation is in Figure 7, where \( m = 5 \) and \( P_{FA} = 10^{-3} \). 10 simulation recursions are used to generate the estimates, and the Ad Hoc estimate is not included. The G-function estimator again has the expected superior performance.

An interesting observation is that the Ad Hoc biasing density has performed poorly when used in a standard IS scheme, but when coupled with the G-function technique, it has had superior performance.
5 Conclusions

This report has analysed Monte Carlo methods used for the estimation of false alarm probabilities in a simple CA-CFAR radar detection scheme. Performance analysis of radar detection schemes requires the estimation of such probabilities. Since the false alarm probabilities are typically very small, standard Monte Carlo estimators require a very large number of runs to produce a reasonable estimate. Hence methods that can provide a high level of accuracy on small sample sizes are desired. Importance Sampling techniques permit such estimations to be made. Three IS estimators were introduced: Ad Hoc, Chernoff and G-Function approaches. Both the Chernoff and G-Function methods perform extremely well. The G-Function method clearly has superior performance, and can provide accurate estimates on very small sample sizes.

Although attention was restricted to the simple case of Gaussian noise and clutter, it is important to reiterate that these techniques can be applied in cases where the clutter and noise is non-Gaussian. The IS estimation techniques are also not restricted to CA-CFAR, but can be applied to other CFAR schemes. This will be investigated in future work.

Acknowledgements

I would like to thank Paul Berry, Brett Haywood, Stephen Howard and Andrew Shaw for advice and comments on this report. Thanks are also due to Yunhan Dong, who vetted the report.
References


Appendix A: Figures

Figure 1: A standard CA-CFAR Detector. The input signal is passed through the square law detector, and the processed return is separated into three components. The cell under test (CUT), and two sets of observations (Test cells A and B), from which a clutter measure is extracted. A CFAR process is then performed on these test cells, with the result passed into a third process that combined these to produce an average clutter measure. In the case illustrated, this is summing. The result is then multiplied by a threshold constant, and this is compared to the CUT statistic. The output is the declaring of a target present, or not present, in the CUT.
Figure 2: Convergence of $\theta$, as estimated from the Newton-Raphson scheme (48), for three particular cases. As can be observed, a reasonable estimate is obtained after approximately 10 recursions.
Figure 3: Simulation gain for the G-Function estimator, as a function of \( \theta \). There is a \( \theta \) that will provide a gain of the order \( 10^7 \), which will result in huge simulation savings.
Figure 4: A comparison of the three IS estimators considered in this report. In each case 15 simulation runs are plotted for each estimator, with each IS estimator using 10 recursions to generate the estimate. The ad hoc parameter was chosen to be $\alpha = 0.05$. Additionally, $m = 5$ and $P_{FA} = 10^{-6}$. The exact false alarm probability is also plotted at each simulation point, for comparison.
Figure 5: Comparison of the three estimators in the case where $m = 4$, $P_{FA} = 10^{-4}$ and $\alpha = 0.9$. Each estimator uses 10 recursions to produce the estimate.
Figure 6: A comparison of the estimators for larger recursions. Here $m = 5$, $P_{FA} = 10^{-3}$, $\alpha = 0.05$ and the number of simulation recursions is 100.
Figure 7: This is a comparison of the G-function and Chernoff IS estimators, for the case where $m = 5$ and $P_{FA} = 10^{-3}$, with 10 recursions used to generate the estimates.
Monte Carlo Methods are introduced and used to estimate false alarm probabilities. The estimation of the latter is important in the context of performance analysis of Constant False Alarm Rate (CFAR) radar detection processes. A CFAR detector estimates the clutter level, producing a threshold, and a target is declared present if the statistic representing the test observation exceeds this threshold. The latter is adjusted adaptively, so that the rate of false alarms is held constant. Hence, in a radar analysis context, the performance of a CFAR process can be determined from whether it maintains a constant false alarm rate. In order to compare the performance of a number of different CFAR schemes, in a common clutter environment, we need to estimate these false alarm probabilities. This can be done quite easily using a basic Monte Carlo estimator. However, the latter may require a very large number of iterations in order to produce a reasonable estimate. To reduce this number of iterations, importance sampling techniques can be used. To illustrate these techniques, we consider the simple case of cell averaging CFAR in a Gaussian environment, with square law detection. This enables comparison of estimators with an exact result.