Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

IMPORTANCE SAMPLING TECHNIQUES FOR ESTIMATING THE BIT ERROR RATE IN DIGITAL COMMUNICATION SYSTEMS

Wheyming Song
Wenchi Chiu

Department of Industrial Engineering
National Tsing Hua University
Hsinchu, Taiwan, Republic of China

David Goldsman

School of Industrial Systems and Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

ABSTRACT

We are interested in estimating the bit error rate (BER) for signal transmission in digital communication systems. Since BERs tend to be extremely small, it is difficult to obtain precise estimators based on the use of crude Monte Carlo simulation techniques. In this paper, we review, expand upon, and evaluate a number of importance sampling variance reduction techniques for estimating the BER. We find that mixtures of certain "tailed" distributions with a uniform distribution produce estimators that are at least competitive with those in the literature. Our comparisons are based on analytical calculations and lay the groundwork for the evaluation of more-general mixture distributions.

1 INTRODUCTION

The bit error rate (BER) is defined as the probability that an error occurs in the processing of one bit of information transmission in a digital communication system. A typical communication system includes a transmitter (such as a laser or modulator), a communication channel (such as a fiber or electric cable), an erbium doped fiber, and amplifiers. Knowledge of the BER is necessary to understand the quality of service (QoS) in a digital communication system. By reducing the BER, the QoS is improved. Analytical approaches used to compute the BER are often difficult because the variance of the random noise (disturbance) associated with the system is unknown. Consequently, it may be necessary to employ another technique to estimate the BER. One such approach is Monte Carlo simulation.

We describe here what we mean by using a Monte Carlo simulation approach to estimate the BER for communication systems. We conduct a simulation experiment by utilizing a set of machines called a "BER tester", which consists of two components: (I) a pulse pattern generator; and (II) an error detector. In component (I), we use a "bit sequence generator" to generate a set of digits (called sent-digits) $\underline{s} = \{s_1, s_2, \ldots, s_n\}$, such that each $s_i \in \{s^0, s^1\}$, for example, $s^0 = 0$, $s^1 = 1$. The random noise is denoted as $\underline{Z} = \{Z_1, Z_2, \ldots, Z_n\}$, where Z_i follows a normal distribution with mean 0 and an unknown variance σ^2. In component (II), we receive the input signal $\underline{X} = \{X_1, X_2, \ldots, X_n\}$, where $\underline{X} = \underline{s} + \underline{Z}$. We then compare each X_i with a threshold T and obtain the output signal $\underline{Y} = \{Y_1, Y_2, \ldots, Y_n\}$, which is also called the received-signal,

$$Y_i = \begin{cases} s^1, & \text{if } X_i > T \\ s^0, & \text{if } X_i \le T. \end{cases}$$

Finally, by comparing the sent digits $\underline{s}$ and received digits $\underline{Y}$, we count the number of incorrect bits received in the output digits. The BER is estimated by the ratio of incorrect digits to all digits received. That is, the estimator of BER can be written as $\widehat{\text{BER}} = n^{-1}\sum_{i=1}^{n} I(s_i \ne Y_i)$, where

$$I(s_i \ne Y_i) = \begin{cases} 1, & \text{if } s_i \ne Y_i \\ 0, & \text{if } s_i = Y_i. \end{cases}$$

The above logic is illustrated in Figure 1 by the block diagram of a common type of BER tester. (The block diagram in Figure 1 is a modification of the diagram used in Lee 2004.)

The simulation approach implemented in Figure 1 is called "crude simulation". Estimation of the BER using crude simulation requires a very large number of samples, especially in the case of a low BER. For example, if a communication system has a BER $\approx 10^{-12}$, then the BER tester may need on the order of hours to obtain a reasonable BER estimate (Patrin and Li 2002). It is noteworthy that a low BER is often the case met in practice; therefore, a challenge in using a BER tester

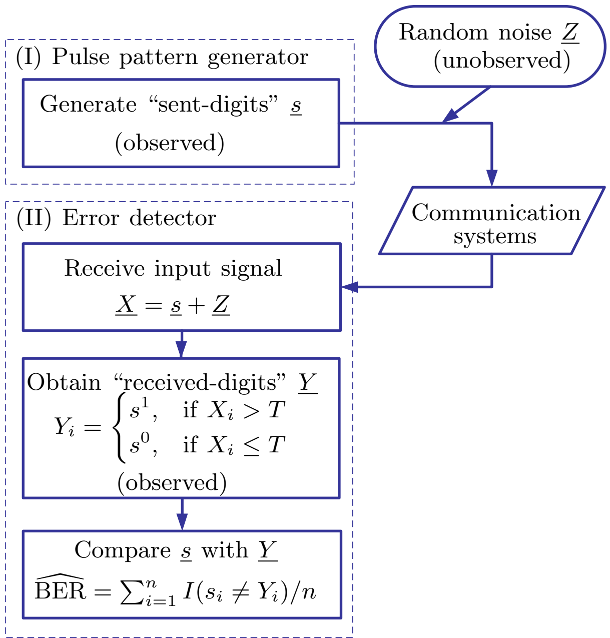

Figure 1: Block diagram of BER tester

to estimate the BER is to increase the efficiency of the crude simulation.

A variance reduction technique, such as importance sampling (IS) can be used to estimate a low BER efficiently. The key idea behind the application of IS in estimating the BER lies in transforming the input signal $\underline{X}$ into $\underline{X}^* = \{X_i^*\}_{i=1}^n$, where each X_i^* follows a distribution with a special "biasing" probability density function (p.d.f.) $f_{X^*}(x)$. Then the "received-digits" become $\underline{Y}^* = \{Y_i^*\}_{i=1}^n$, where

$$Y_i^* = \begin{cases} s^1, & \text{if } X_i^* > T \\ s^0, & \text{if } X_i^* \leq T. \end{cases}$$

The estimator of the BER is then given by $\widehat{\text{BER}} \equiv wn^{-1}\sum_{i=1}^n I(s_i \neq Y_i^*)$, where w is an adjustment factor, which depends on $f_{X^*}(x)$. Two major issues related to using IS in a BER tester for the estimation of BER are: (i) How can one determine the "optimal biasing p.d.f." $f_{X^*}(x)$, and (ii) how can one transfer $\underline{X}$ into $\underline{X}^*$ such that the p.d.f. of $\underline{X}^*$ satisfies the optimal biasing p.d.f. determined as a result of issue (i). To our knowledge, all existing papers on the estimation of BER using IS address this issue. For general articles discussing this issue, the reader can consult Jeruchim (1984), Wang and Bhargava (1987), and Wang and Lu (1993), Smith (1997), Chen (2002), and Lee (2004).

Because the distribution of error occurrence is a tail distribution, many researchers consider the use of such distributions as the Gaussian tail, Rayleigh tail, and exponential tail as the biasing p.d.f. Shih and Song (1995) combines a Gaussian tail and a uniform distribution as a "mixed" biasing p.d.f., and they show that the mixed performs better than does the Gaussian tail distribution alone. This paper reviews, expands upon, and evaluates the three well-known biasing p.d.f.s for estimating the BER, including some simple "fixes" to the mixed biasing p.d.f., e.g., the Gaussian tail and the uniform. Our results disprove a conjecture from Shih and Song (1995) that a mixed biasing p.d.f. of any tail-like biasing distribution with a uniform distribution should always be more robust than the initial tail-like biasing distribution alone.

2 PROBLEM STATEMENT

The BER can be written as the sum of two types of error probabilities, where a Type I error occurs when s^0 is sent but a disturbance causes $Z_i + s^0$ to exceed the threshold T; and a Type II error occurs when s^1 is sent but a disturbance causes $Z_i + s^1$ to fall below T. That is, when the threshold setting is constant, the BER can be expressed by

$$\begin{aligned}\text{BER} &= \text{P}(s^1 \text{ is received}|s^0 \text{ is sent})\text{P}(s^0 \text{ is sent}) \\ &\quad + \text{P}(s^0 \text{ is received}|s^1 \text{ is sent})\text{P}(s^1 \text{ is sent}) \\ &= \text{P}(s^0 + Z > T|s^0 \text{ is sent})\text{P}(s^0 \text{ is sent}) \\ &\quad + \text{P}(s^0 + Z \leq T|s^1 \text{ is sent})\text{P}(s^1 \text{ is sent}) \\ &= \text{P}(X > T|s^0 \text{ is sent})\text{P}(s^0 \text{ is sent}) \\ &\quad + \text{P}(X \leq T|s^1 \text{ is sent})\text{P}(s^1 \text{ is sent}) \\ &= \text{P}(s^0 \text{ is sent}) \int_T^\infty f_0(x)\,dx \\ &\quad + \text{P}(s^1 \text{ is sent}) \int_{-\infty}^T f_1(x)\,dx,\end{aligned}$$

where f_0 and f_1 are the distributions of X when the signals s^0 and s^1 are sent, respectively. The densities can be sketched as in Figure 2.

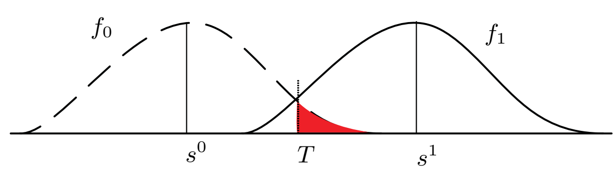

Figure 2: Densities of signal X: f_0 when s^0 is sent, f_1 when s^1 is sent

Because the estimation of the probabilities of Type I and Type II errors is similar, in this paper we assume that $\mathrm{P}(s^0 \text{ is sent}) = 1$; and therefore we estimate the Type I error probability, defined as

$$p = \mathrm{P}(X > T | s^0 \text{ is sent}) = \int_T^\infty f_X(x)\, dx,$$

where $f_X(x)$ is used to replace $f_0(x)$ in Figure 2.

Suppose in some "ideal" situation, the "nominal" variance σ^2 of the p.d.f. $f_X(x)$ is known. Then the BER p can be computed analytically or approximated numerically. We call such p the nominal BER.

We can express p as an expectation of a certain random variable as follows. Define

$$g(X) = \begin{cases} 1, & \text{if } X > T \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Then we have

$$\begin{aligned} \mathrm{E}(g(X)) &= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \\ &= \int_T^\infty f_X(x)\, dx \\ &= p. \end{aligned}$$

Therefore, we can use $\hat{\mathrm{E}}(g(X))$, the sample average of $g(X)$, to estimate p. This approach is called crude simulation, and the corresponding estimator is denoted as $\hat{p}^{(\mathrm{c})}$.

$$\hat{p}^{(\mathrm{c})} = \frac{1}{n} \sum_{i=1}^{n} g(X_i). \tag{2}$$

Recall that T is the threshold of the receiver, n is the number of trials, and x_i, $i = 1, \ldots, n$, is a sample of size n following the p.d.f. $f_X(x)$, the underlying distribution for the input signal X. The estimator $\hat{p}^{(\mathrm{c})}$ obtained via Equation (2) can then be compared with the nominal BER with nominal variance σ^2. If the estimator $\hat{p}^{(\mathrm{c})}$ provides a higher BER than the nominal BER, we then conclude that the true variance of the noise in the underlying communication system is likely being underestimated.

Crude simulation is time consuming when the BER is very low. Importance sampling is a variance reduction technique used in simulation experiments to increase the simulation's efficiency. The key concept behind IS is to generate data from a so-called *biasing distribution* or *biasing p.d.f.* (denoted as f_{X^*}) instead of from the original underlying distribution (f_X). A good biasing p.d.f. will mimic the underlying p.d.f. closely and will be more likely to indicate the location of the error region (which is the so-called "importance region").

We rewrite $\mathrm{E}(g(X))$ as follows:

$$\begin{aligned} \mathrm{E}(g(X)) &= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} \left[\frac{f_X(x)}{f_{X^*}(x)} g(x) \right] f_{X^*}(x)\, dx \\ &= \int_{-\infty}^{\infty} \left[w(x) g(x) \right] f_{X^*}(x)\, dx \\ &= \mathrm{E}(g^*(X^*)), \end{aligned}$$

where

$$w(\cdot) = \frac{f_X(\cdot)}{f_{X^*}(\cdot)} \tag{3}$$

is called an adjustment factor or weight factor, and $g^*(X^*) = w(\cdot) g(X^*)$ is a function of the random variable X^* which follows p.d.f. $f_{X^*}(\cdot)$ instead of $f_X(\cdot)$.

The corresponding IS estimator of p is denoted as

$$\hat{p}^{(\mathrm{IS})} = \frac{1}{n} \sum_{i=1}^{n} g^*(X_i^*) = \frac{1}{n} \sum_{i=1}^{n} w(X_i^*) g(X_i^*),$$

where

$$g(X_i^*) = \begin{cases} 1, & \text{if } X_i^* > T \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

which is the same as that used in Equation (1). The weight $w(\cdot)$ is obtained by plugging the nominal variance σ^2 and the corresponding nominal variance for the biasing p.d.f. into $f_X(x)$ and $f_{X^*}(x)$, respectively. The estimator $\hat{p}^{(\mathrm{IS})}$ obtained via Equation (4) can then be compared with the nominal BER with variance σ^2.

When T is constant, the Gaussian tail distribution is known as the optimal biasing distribution (Kahn and Marshall 1953, I and Lusignan 1986), and is defined by

$$f_{\mathrm{g}}(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi} Q(T/\sigma)} e^{-x^2/(2\sigma^2)}, & x \geq T \\ 0, & x < T, \end{cases} \tag{5}$$

where $Q(a) \equiv \mathrm{P}(Z \geq a) = \int_a^\infty (2\pi)^{-1/2} e^{-z^2/2}\, dz$, and where Z follows the standard normal distribution.

What if the threshold setting T is random instead of a constant? Consider the situation in which the threshold of the receiver decision device is distorted by some random noise or the underlying noise departs slightly from the original Gaussian distribution. In both cases, the threshold value is no longer a constant. So let us assume that the threshold is a random variable which follows some p.d.f. $f_T(t)$ bounded within a certain interval (c_1, c_2). Then BER can be written as

$$\begin{aligned} p &= \mathrm{P}(X > T | s^0 \text{ is sent}) \\ &= \mathrm{P}(X > T | T = t) f_T(t) \\ &= \int_{c_1}^{c_2} \int_t^{\infty} f_X(x) f_T(t)\, dx\, dt. \end{aligned} \tag{6}$$

Equation (6) indicates that p is the volume under a two-dimensional integral. In summary, the problem we address in this paper is to estimate p defined in Equation (6), where the threshold setting T is a random variable. In this paper, we consider the case when the real threshold T follows a uniform distribution between c_1 and c_2.

3 TAIL BIASING DISTRIBUTIONS

In this section, we review and expand upon the Gaussian tail (f_{g}), Rayleigh tail (f_{r}), and exponential tail (f_{e}) biasing distributions. The corresponding IS estimators using f_{g}, f_{r}, f_{e} as the biasing distributions are denoted as $\hat{p}_{\mathrm{g}}$, $\hat{p}_{\mathrm{r}}$, $\hat{p}_{\mathrm{e}}$, respectively. In the following discussion, the parameter c is denoted as the truncation point of each tail biasing p.d.f. We use the mean squared error (mse), which is equal to bias^2 + variance, as the criterion for comparison.

3.1 Gaussian Tail Biasing Distribution

The Gaussian tail biasing distribution (I and Lusignan 1986) is defined as

$$f_{\mathrm{g}}(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}Q(c/\sigma)} e^{-x^2/(2\sigma^2)}, & \text{if } x \geq c \\ 0, & \text{if } x < c. \end{cases}$$

The mse, derived in Shih and Song (1995), has the form

$$\begin{aligned} &(c_2 - c_1)\,\mathrm{mse}(\hat{p}_{\mathrm{g}}) \\ &\quad = \int_{c_1}^{c_2} \left\{ \mathrm{bias}^2(\hat{p}_{\mathrm{g}} | T = t) + \mathrm{var}(\hat{p}_{\mathrm{g}} | T = t) \right\} dt, \end{aligned}$$

where the conditional bias and variance of $\hat{p}_{\mathrm{g}}$ are

$$\mathrm{bias}(\hat{p}_{\mathrm{g}} | T = t) = \begin{cases} Q(t/\sigma) - Q(c/\sigma), & \text{if } c \geq t \\ 0, & \text{if } c < t, \end{cases}$$

$$\begin{aligned} &\mathrm{var}(\hat{p}_{\mathrm{g}} | T = t) \\ &\quad = \begin{cases} 0, & \text{if } c \geq t \\ Q(c/\sigma)Q(t/\sigma) - Q^2(t/\sigma), & \text{if } c < t. \end{cases} \end{aligned}$$

3.2 Rayleigh Tail Biasing Distribution

The Rayleigh tail biasing distribution, proposed in Beaulieu (1990) and Wang and Lu (1993), is defined as

$$f_{\mathrm{r}}(x) = \begin{cases} 2\alpha x e^{\alpha(c^2 - x^2)}, & \text{if } x \geq c \\ 0, & \text{if } x < c. \end{cases}$$

We can derive the mse of $\hat{p}_{\mathrm{r}}$ as follows.

$$\begin{aligned} &(c_2 - c_1)\,\mathrm{mse}(\hat{p}_{\mathrm{r}}) \\ &\quad = \int_{c_1}^{c_2} \left\{ \mathrm{bias}^2(\hat{p}_{\mathrm{r}} | T = t) + \mathrm{var}(\hat{p}_{\mathrm{r}} | T = t) \right\} dt, \end{aligned}$$

where the conditional bias and variance are

$$\mathrm{bias}(\hat{p}_{\mathrm{r}} | T = t) = \begin{cases} Q(t/\sigma) - Q(c/\sigma), & \text{if } c \geq t \\ 0, & \text{if } c < t, \end{cases}$$

$$\begin{aligned} &\mathrm{var}(\hat{p}_{\mathrm{r}} | T = t) \\ &\quad = \begin{cases} 0, & \text{if } c \geq t \\ \dfrac{e^{-\alpha t^2} E_1[t^2(1/\sigma^2 - \alpha)]}{8\pi\sigma^2\alpha} - Q^2(t/\sigma), & \text{if } c < t, \end{cases} \end{aligned}$$

where $E_1(s) = \int_s^{\infty} e^{-y}/y\, dy$ is the exponential integral.

Beaulieu (1990) compares the performance of BER for Rayleigh with $\alpha = 1/(2\sigma^2)$ and Gaussian tail distributions assuming that the threshold is unknown. His results show that Rayleigh with $\alpha = 1/(2\sigma^2)$ performs better than does the Gaussian tail distribution.

3.3 Exponential Tail Biasing Distribution

The exponential tail biasing, proposed in Wang and Lu (1993), is defined as

$$f_{\mathrm{e}}(x) = \begin{cases} \alpha e^{\alpha(c - x)}, & \text{if } x \geq c, \\ 0, & \text{otherwise}, \end{cases}$$

where $0 < \alpha < 1$.

We can derive the mse of $\hat{p}_{\mathrm{e}}$ as follows.

$$\begin{aligned} &(c_2 - c_1)\,\mathrm{mse}(\hat{p}_{\mathrm{e}}) \\ &\quad = \int_{c_1}^{c_2} \left\{ \mathrm{bias}^2(\hat{p}_{\mathrm{e}} | T = t) + \mathrm{var}(\hat{p}_{\mathrm{e}} | T = t) \right\} dt, \end{aligned}$$

where the conditional bias and variance are

$$\text{bias}(\hat{p}_e|T=t) = \begin{cases} Q(t/\sigma) - Q(c/\sigma), & \text{if } c \geq t \\ 0, & \text{if } c < t, \end{cases}$$

$$\text{var}(\hat{p}_e|T=t)$$

$$= \begin{cases} 0, & c \geq t \\ \dfrac{\exp(-\alpha c + \frac{\alpha^2\sigma^2}{4})}{2\alpha\sqrt{\pi}\sigma} Q\left(\dfrac{t - \alpha\sigma^2/2}{\sigma/\sqrt{2}}\right) - Q^2(\frac{t}{\sigma}), & c < t. \end{cases}$$

4 MIXED BIASING DISTRIBUTIONS

Shih and Song (1995) proposed the Gaussian tail and uniform mixed distribution and showed that this distribution performs better than the Gaussian tail distribution alone when the signal or the threshold setting follows the uniform distribution. In this section, we investigate three examples of mixed biasing distributions. The term "mixed" is chosen because the biasing p.d.f.s combine two types of distributions:

1. the Gaussian tail and uniform distributions, denoted as f_{gu},
2. the Rayleigh tail and uniform distributions, denoted as f_{ru}, and
3. the exponential tail and uniform distributions, denoted as f_{eu}.

The general forms of the mixed biasing p.d.f.s are similar for the three examples and can be illustrated in Figure 3. The three parameters in Figure 3 are the height h, $0 \leq hk \leq 1$, and two widths a and b, where $0 \leq a \leq b \leq c_2 - c_1$.

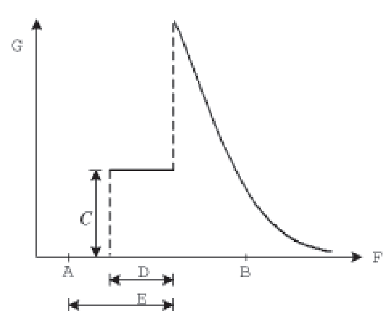

Figure 3: Tail p.d.f. + uniform p.d.f.

We use $\hat{p}_{\text{gu}}$, $\hat{p}_{\text{ru}}$, $\hat{p}_{\text{eu}}$ to denote the IS estimator of the BER p obtained by using f_{gu}, f_{ru}, f_{eu} as the biasing distributions, respectively. We derive mean squared error results in Sections 4.1–4.3, followed by corresponding numerical results. The derivations and additional details regarding the propositions can be found in Song, Chiu, and Goldsman (2005).

4.1 Gaussian Tail + Uniform Distribution

The p.d.f. of the Gaussian tail and uniform distribution proposed by Shih and Song (1995) is as follows:

$$f_{\text{gu}}(x) = \begin{cases} h, & x \in [c_1 + b - a, c_1 + b] \\ \dfrac{(1-ha)e^{-\frac{x^2}{2\sigma^2}}}{Q\left((c_1+b)/\sigma\right)\sqrt{2\pi}\sigma}, & x > c_1 + b \\ 0, & \text{otherwise.} \end{cases}$$

The mse of $\hat{p}_{\text{gu}}$ is given in Proposition 1.

Proposition 1.

$$(c_2 - c_1)\,\text{mse}(\hat{p}_{\text{gu}}) = \int_{c_1}^{c_2} \left\{\text{bias}^2(\hat{p}_{\text{gu}}|T=t) + \text{var}(\hat{p}_{\text{gu}}|T=t)\right\} dt,$$

where the conditional bias and variance of $\hat{p}_{\text{gu}}$, derived in Shih and Song (1995), are

$$\text{bias}(\hat{p}_{\text{gu}}|T=t) = \begin{cases} Q(t/\sigma) - Q(c/\sigma), \text{if } c_1 + b - a \geq t \\ 0, \quad \text{otherwise,} \end{cases}$$

$$\text{var}(\hat{p}_{\text{gu}}|T=t)$$

$$= \begin{cases} 0, & \text{if } t \in [c_1, c_1 + b - a] \\ \dfrac{Q(\sqrt{2}t/\sigma) - Q(\sqrt{2}(c_1+b)/\sigma)}{2h\sigma\sqrt{\pi}} - Q^2(t/\sigma) & \\ \quad + \dfrac{Q^2((c_1+b)/\sigma)}{1-ha}, & \text{if } t \in [c_1 + b, c_1 + b - a] \\ \dfrac{Q((c_1+b)/\sigma)Q(t/\sigma)}{1-ha} - Q^2(t/\sigma), & \text{if} t \in [c_1 + b, c_2]. \end{cases}$$

The results in this paper correct some typos in Shih and Song (1995).

4.2 Rayleigh Tail + Uniform Distribution

The p.d.f. of the Rayleigh tail and uniform distribution is

$$f_{\text{ru}}(x) = \begin{cases} h, & x \in [c_1 + b - a, c_1 + b] \\ (1-ha)2\alpha x e^{\alpha((c_1+b)^2 - x^2)}, & x > c_1 + b \\ 0, & \text{otherwise,} \end{cases}$$

where $0 < \alpha < 1$.

The mse of $\hat{p}_{\text{ru}}$ is given in Proposition 2.

Proposition 2.

$$(c_2-c_1)\,\mathrm{mse}(\hat{p}_{\mathrm{ru}}) = \int_{c_1}^{c_2} \left\{\mathrm{bias}^2(\hat{p}_{\mathrm{ru}}|T=t) + \mathrm{var}(\hat{p}_{\mathrm{ru}}|T=t)\right\} dt,$$

where the conditional bias and variance are

$$\mathrm{bias}(\hat{p}_{\mathrm{ru}}|T=t) = \begin{cases} Q(t/\sigma) - Q(c/\sigma), & \text{if } c_1+b-a \geq t \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathrm{var}(\hat{p}_{\mathrm{ru}}|T=t) = \begin{cases} 0, & \text{if } t \in [c_1, c_1+b-a] \\ \dfrac{Q(\sqrt{2}t/\sigma) - Q(\sqrt{2}(c_1+b)/\sigma)}{2h\sigma\sqrt{\pi}} - Q^2(t/\sigma) \\ \quad + \dfrac{e^{-\alpha(c_1+b)^2}}{8\pi\sigma^2\alpha}\left[\dfrac{E_1\left[(c_1+b)^2(1/\sigma^2-\alpha)\right]}{1-ha}\right], & \text{if } t \in [c_1+b, c_1+b-a] \\ \dfrac{e^{-\alpha t^2}}{8\pi\sigma^2\alpha}\left\{\dfrac{E_1\left[t^2(1/\sigma^2-\alpha)\right]}{1-ha} - Q^2(t/\sigma)\right\}, & \text{if } t \in [c_1+b, c_2]. \end{cases}$$

4.3 Exponential Tail + Uniform Distribution

The p.d.f. of the exponential tail and uniform distribution is defined as

$$f_{\mathrm{eu}}(x) = \begin{cases} h, & x \in [c_1+b-a, c_1+b] \\ (1-ha)\alpha e^{\alpha((c_1+b)-x)}, & x > c_1+b \\ 0, & \text{otherwise.} \end{cases}$$

The mse of is given in Proposition 3.

Proposition 3.

$$\mathrm{mse}(\hat{p}_{\mathrm{eu}}) = \frac{\int_{c_1}^{c_2}\left\{\mathrm{bias}^2(\hat{p}_{\mathrm{eu}}|T=t) + \mathrm{var}(\hat{p}_{\mathrm{eu}}|T=t)\right\} dt}{c_2-c_1},$$

where the conditional bias and variance are

$$\mathrm{bias}(\hat{p}_{\mathrm{eu}}|T=t) = \begin{cases} Q(t/\sigma) - Q(c/\sigma), & \text{if } c_1+b-a \geq t \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathrm{var}(\hat{p}_{\mathrm{eu}}|T=t) = \begin{cases} 0, & \text{if } g \in [c_1, c_1+b-a] \\ \dfrac{Q(\sqrt{2}t/\sigma) - Q(\sqrt{2}(c_1+b)/\sigma)}{2h\sigma\sqrt{\pi}} - Q^2(t/\sigma) \\ \quad + \dfrac{\exp(-\alpha t + \frac{\alpha^2\sigma^2}{4})}{2\alpha\sqrt{\pi}\sigma(1-ha)} \cdot Q\left(\dfrac{c_1+b-\alpha\sigma^2/2}{\sigma\sqrt{2}}\right), & \text{if } t \in [c_1+b, c_1+b-a] \\ \dfrac{\exp(-\alpha t + \frac{\alpha^2\sigma^2}{4})}{2\alpha\sqrt{\pi}\sigma(1-ha)} Q\left(\dfrac{t-\alpha\sigma^2/2}{\sigma/\sqrt{2}}\right) - Q^2(t/\sigma), & \text{if } t \in [c_1+b, c_2]. \end{cases}$$

4.4 Numerical Comparison

In this section we compare the "optimal" mixed biasing p.d.f.s discussed in Sections 4.1–4.3 with the corresponding original "optimal" biasing p.d.f.s. Consider $\hat{p}_{\mathrm{gu}}$ as an example. The optimal mixed biasing pdf can be obtained by using the optimal parameters a^*, b^*, and h^* such that the $\mathrm{mse}(\hat{p}_{\mathrm{gu}}(a^*, b^*, h^*)) \leq \mathrm{mse}(\hat{p}_{\mathrm{gu}}(a, b, h))$ for all $0 \leq h \leq 1$, and $0 \leq a \leq b \leq c_2 - c_1$. The optimal values a^*, b^*, and h^* are obtained via a numerical (grid serach) method.

The results are summarized in Tables 1–3. We consider a variety of choices for c_1 and c_2, for which the BERs range from moderate to small values. We assume that the variance of the Gaussian noise is $\sigma^2 = 1$.

Table 1 shows that for three choices of (c_1, c_2), the IS estimator with the mixed Gaussian biasing p.d.f. yields smaller mses than does the Gaussian tail p.d.f. when the threshold follows a uniform distribution. Table 1 corrects some typos in Shih and Song (1995).

Table 1: Gaussian tail and the mixed Gaussian

(c_1, c_2)	(2,5)	(3,5)	(4,5)
Gaussian tail	$b^* = 0.3$	$b^* = 0.2$	$b^* = 0.2$
mse	1.32E-5	4.33E-8	3.63E-11
mixed (Gaussian +uniform)	$b^* = 0.8$ $h^* = 1.2$ $a^* = 0.5$	$b^* = 0.6$ $h^* = 1.5$ $a^* = 0.4$	$b^* = 0.5$ $h^* = 1.9$ $a^* = 0.3$
mse	8.06E-6	3.19E-8	2.96E-11
mse-reduction	39%	26%	18%

Table 2 shows that for the three choices of (c_1, c_2), the IS estimator with the mixed Rayleigh biasing p.d.f. yields essentially the same mse compared to the Rayleigh tail p.d.f. when the threshold follows a uniform distribution. First, we can see that the optimal setting occurs when $\alpha \approx 1/(2\sigma^2)$, which is consistent with the results in Beaulieu (1990). Second, the results between the Rayleigh tail and the mixed Rayleigh + uniform disprove the conjecture from Shih and Song (1995) that a mixed biasing p.d.f. of any tail-like biasing distribution with a uniform distribution should always be more robust than the initial tail-like biasing distribution.

Table 2: Rayleigh tail and the mixed Rayleigh

(c_1, c_2)	(2,5)	(3,5)	(4,5)
Rayleigh tail	$\alpha^* = 0.55$	$\alpha^* = 0.55$	$\alpha^* = 0.50$
	$b^* = 0.05$	$b^* = 0.05$	$b^* = 0.05$
mse	6.42E−8	4.01E−10	7.44E−13
	$\alpha^* = 0.55$	$\alpha^* = 0.55$	$\alpha^* = 0.50$
mixed	$b^* = 0.05$	$b^* = 0.05$	$b^* = 0.05$
(Rayleigh	$h^* = 0.0$	$h^* = 0.0$	$h^* = 0.0$
+uniform)	$a^* = 0.0$	$a^* = 0.0$	$a^* = 0.0$
mse	6.42E-8	4.01E-10	7.44E-13
mse-reduction	0%	0%	0%

Table 3 shows that for two of the cases, the IS estimator with the mixed exponential biasing distribution yields a smaller mse than those with the exponential tail p.d.f. alone when the threshold follows a uniform distribution with smaller range.

Table 3: Exponential tail and the mixed exponential

(c_1, c_2)	(2,5)	(3,5)	(4,5)
Exp tail	$\alpha^* = 1.0$	$\alpha^* = 1.0$	$\alpha^* = 1.0$
	$b^* = 0.3$	$b^* = 0.25$	$b^* = 0.25$
mse	9.52E−6	4.73E−8	4.83E−11
	$\alpha^* = 1.0$	$\alpha^* = 1.0$	$\alpha^* = 1.0$
mixed	$b^* = 0.3$	$b^* = 0.9$	$b^* = 0.8$
(Exp	$h^* = 0.0$	$h^* = 1.2$	$h^* = 1.4$
+uniform)	$a^* = 0.0$	$a^* = 0.7$	$a^* = 0.6$
mse	9.52E-6	4.16E-8	3.77E-11
mse-reduction	0%	12%	22%

5 CONCLUSION

This paper investigates the idea of using a mixed biasing distribution to estimate the bit error rate for a linear binary digital communication system with a threshold logic receiver. The performance measure used in this paper is the mean squared error. The statistical performance of the mixed distribution is superior to that of the corresponding initial tail-like biasing distribution if the initial biasing distribution is a Gaussian tail or an exponential tail. The mixed distribution performs essentially the same as the initial tail-like biasing distribution if the initial biasing distribution is a Rayleigh tail.

The theoretical mse expressions provide a reasonable foundation to search for the optimal biasing distribution when the threshold T is a random variable; and the three mixed biasing p.d.f.s studied in this paper lay the groundwork for the evaluation of more-general mixture distributions.

REFERENCES

Beaulieu, N. C. 1990. A composite importance sampling technique for digital communication systems simulation. *IEEE Trans. Commun*, 38: 393–396.

Chen, S. 2002. Importance sampling simulation for evaluating the lower-bound BER of the Bayesian DFE, *IEEE Trans. Commun.*, 50(2): 179–182.

I, C. L. and B. B. Lusignan. 1986. The ‘optimum’ BER estimator for digital satellite communication systems, in *AIAA 11th Commun. Satellite Syst. Conf.* San Diego, CA, Mar. 1986, 144–151.

Jeruchim, M. C. 1984. Techniques for estimating the bit error rate in the simulation of digital communication systems, *IEEE J. Select. Areas, Commun.*, 2(1): 153–170.

Kahn, H. and A. W. Marshall. 1953. Methods of reducing sample size in Monte Carlo computations. *J. Oper. Res. Soc. Amer.* 1: 263–271.

Lee, B. 2004. Sequential Bayesian bit error rate measurement, *IEEE Transactions on Instrumentation and Measurement*, 53(4): 947–954.

Patrin, J. and M. Li. 2002. Comparison and correlation of signal integrity measurement techniques, in *Proc. DesignCon Jan. 2002 High Performance System Design Conf.*, Santa Clara, CA.

Shih, N. H., and W. T. Song. 1995. Robust consideration of importance sampling in digital communication system, *Journal of the Chinese Institute of Engineers*, 18(5): 713–718.

Smith, P. J. 1997. Quick simulation: A review of importance sampling techniques in communications systems, *IEEE J. Select. Areas, Commun.*, 15(4): 597–613.

Song, W. T., W. Chiu, and D. Goldsman. 2005. An importance sampling technique for estimating bit error rate in digital communication systems, Technical Report, Tsing Hua University, Taiwan, R.O.C.

Wang, Q. and V. K. Bhargava. 1987. On the application of importance sampling to BER estimation in the simulation of digital communication systems, *IEEE Trans. Commun.*, 35: 1231–1233.

Wang, Y. Y. and C. C. Lu. 1993. Using Weibull density function in importance sampling for digital communication system simulation, *IEEE J. Select. Areas, Commun.*, 140(2): 163–168.

AUTHOR BIOGRAPHIES

WHEYMING T. SONG is a professor in the Department of Industrial Engineering at the National Tsing Hua University in Taiwan. She received master degrees in applied mathematics in 1983 and industrial engineering in 1984, both from the University of Pittsburgh. She then received her Ph.D. from the School of Industrial Engineering at Purdue University in 1989. She joined Tsing Hua in 1990 after spending one year as a visiting assistant professor at Purdue IE. Professor Song received the outstanding university-wide teaching award from the National Tsing Hua University in 1993. Her research honors include the 1988 Omega Rho student paper award, the 1990 IIE Doctoral Dissertation award, and the 1996 distinguished research award from the National Science Council of the Republic of China. Her research interests are applied operations research, probability and statistics, and the statistical aspects of stochastic simulation. Her e-mail address is `<wheyming@ie.nthu.edu.tw>`.

WENCHI CHIU received a B.S. degree in transportation engineering and management from the National Chiao Tung University in Taiwan in June 1997. She then received her masters degree in transportation engineering and management from the National Chiao Tung University in June 1999. She is now a Ph.D. candidate at the National Tsing Hua University. Her e-mail address is `<d897802@oz.nthu.edu.tw>`.

DAVID GOLDSMAN is a Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. He received his Ph.D. in Operations Research and Industrial Engineering from Cornell University. His research interests include simulation output analysis and ranking and selection. He is an active participant in the Winter Simulation Conference, having been Program Chair in 1995, and having served on the WSC Board of Directors since 2002. His e-mail address is `<sman@isye.gatech.edu>`, and his web page is `<www.isye.gatech.edu/~sman/>`.