Estimates of Error Probability for Complex Gaussian Channels with Generalized Likelihood Ratio Detection

Michael D. DeVore, *Member*, *IEEE*

Abstract—We derive approximate expressions for the probability of error in a two-class hypothesis testing problem in which the two hypotheses are characterized by zero-mean complex Gaussian distributions. These error expressions are given in terms of the moments of the test statistic employed and we derive these moments for both the likelihood ratio test, appropriate when class densities are known, and the generalized likelihood ratio test, appropriate when class densities must be estimated from training data. These moments are functions of class distribution parameters which are generally unknown so we develop unbiased moment estimators in terms of the training data. With these, accurate estimates of probability of error can be calculated quickly for both the optimal and plug-in rules from available training data. We present a detailed example of the behavior of these estimators and demonstrate their application to common pattern recognition problems, which include quantifying the incremental value of larger training data collections, evaluating relative geometry in data fusion from multiple sensors, and selecting a good subset of available features.

Index Terms—Complex Gaussian channels, probability of error, model inaccuracy, moment estimators, Johnson's systems of distributions.

1 Introduction

It is well-known that the optimal decision rule in a simple binary detection problem is the likelihood ratio test [1]. The optimal test is defined in terms of the conditional distributions of the observation under each hypothesis, however these distributions are often incompletely known. In place of complete distributions, it is common to have only a sample of observations drawn under the two hypotheses. A decision rule based on sample observations is referred to as adaptive because its behavior depends on the particular sample that was drawn. While readily definable for nonadaptive rules, the very notion of optimality among adaptive decision rules is problematic. This is because the actual performance of any rule depends on the unknown conditional distributions [2].

One of the most common approaches is to assume some restricted family for the conditional distributions and apply a generalized likelihood ratio test, wherein the distribution that maximizes the likelihood of a training sample is used in place of the true distribution. This approach is also referred to as the "plug-in" decision rule and we will differentiate it from the ideal (that is, optimal) decision rule which is based on the true conditional distributions. We use the term "ideal" to emphasize that this rule is generally unattainable because the distributions are unknown.

Our interests lie in quantifying the class-conditional probability of error for both the ideal and plug-in decision rules and in deriving estimators for these error probabilities when the true parameter values are unknown. The focus is on hypotheses involving zero-mean complex Gaussian (Rayleigh magnitude) observation vectors with diagonal covariance matrices. In addition to communication through Rayleigh fading channels, complex Gaussian observations are often assumed when imaging rough objects with coherent radiation [3], [4], [5]. Rigorous statistical assessment of a complex Gaussian model for synthetic aperture radar (SAR) imagery has been reported by DeVore and O'Sullivan [6], and these models have been used extensively in the development of plug-in rule classification algorithms for SAR [7], [8], [9] and for high-resolution radar [10].

When data collection is expensive or when there are a large number of conditioning variables (such as object pose, articulation and other states, etc.), classifier designers must often make do with very small training samples. (For a survey of approaches to dealing with classifiers designed around small training samples, see [11]). Sample-based methods of performance estimation, such as cross validation and boot strapping, cannot give high accuracy estimates with small samples, as their precision is limited to one over the sample size. Estimates of performance obtained by substituting estimated parameters into expressions for the ideal error are overly optimistic [12]. A better approach is to consider probability of error over all possible training samples, however, closed-form solutions are hard to derive. For instance, Raudys and Pikelis [13] present integral and infinite-series expressions for the plug-in probability of error under various Gaussian assumptions and state that they "are rather complex and unsuitable for practical everyday use."

Rather than completely determining the conditional distributions necessary to exactly compute the probability of error, our approach is to find conditional moments of the likelihood and generalized likelihood ratio under each hypothesis and use these to fit an arbitrary distribution family. In most practical problems, the moments of the

- The author is with the Department of Systems and Information Engineering, University of Virginia, 102G Olsson Hall, 151 Engineer's Way, PO Box 400747, Charlottesville, VA 22904. E-mail:mdevore@virginia.edu.

Manuscript received 27 July 2004; revised 3 Jan. 2005; accepted 4 Jan. 2005; published online 11 Aug. 2005.
Recommended for acceptance by L. Kuncheva.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0388-0704.

likelihood or generalized likelihood ratio are defined in terms of unknown parameters that characterize the class-conditional distributions. We therefore derive unbiased estimators for these moments in terms of maximum-likelihood estimators for the unknown parameters. Unbiased estimators for integer moments of arbitrarily high order are obtainable by continued application of the approach described. This approach allows us to approximate the conditional probabilities of error for the ideal and plug-in tests when the true parameters are known, which might be the case with simulated data used to evaluate a recognition algorithm. It also allows us to estimate the conditional probabilities of error for the ideal and plug-in tests when only training samples are available.

A wide variety of distribution families are available for fitting with the moments of the likelihood ratio. The Gaussian family is a convenient choice, with fast and accurate routines readily available. Moreover, this family may be appropriate if the observation vectors are long and central limit theorem conditions apply at least approximately. However, a member of the Gaussian family is specified by only two moments and it may be desirable for the sake of accuracy to employ a family with a larger set of free parameters. A number of distribution systems have been developed by researchers seeking to provide approximations to a wide variety of observed sampling distributions (cf., Johnson et al. [14, Section 12.4]). In our examples, we use the test statistic moments to fit a distribution from Johnson's family [15]. The Johnson family contains the Gaussian family and three other systems which are each defined as a simple transformation of the Gaussian family. Together, these systems cover the entire space of possible values for the first four moments of any distribution and each member of the Johnson family is uniquely specified by a point in that space.

This work expands on our preliminary results [16], which presented expressions and estimators for the mean and variance of the test statistics and demonstrated fitting with the Gaussian distribution. In Sections 2 and 3, we present expressions for the first four conditional moments of the test statistics in the ideal and plug-in decision rules, respectively, for selecting between complex Gaussian hypotheses and we describe how higher order moments can be generated. In Section 4, we address unbiased estimators for those conditional moments in terms of maximum-likelihood estimates of the observation vector distributions under the two hypotheses. A brief description of the Johnson distribution family is contained in Section 5. An example demonstrating probability of error estimators constructed by fitting the Gaussian and more general Johnson families is contained in Section 6. The method is fast, accurate, gives an estimate of performance directly in terms of class-conditional probability of error, and can be used with known or estimated parameters. As a result, it has numerous applications including feature selection, quantitatively evaluating the potential benefit of larger training sets given only small training sets, and quantitatively evaluating the potential benefit of multiple sensor fusion. These applications are discussed in Section 7. Conclusions follow in Section 8.

2 Ideal Decision Rule

We first consider the ideal case in which all parameters are known. The goal is to choose between one of two complex Gaussian hypotheses

$$\begin{aligned} \mathcal{H}_1 &: \mathbf{X} \sim \mathbb{C}N(\mathbf{0}, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_K^2)) \\ \mathcal{H}_2 &: \mathbf{X} \sim \mathbb{C}N(\mathbf{0}, \mathrm{diag}(\tau_1^2, \ldots, \tau_K^2)). \end{aligned} \tag{1}$$

The ideal decision rule is

$$L(\mathbf{X}) = \sum_{k=1}^{K} \left[\frac{1}{2} \left(\frac{\sigma_k^2}{\tau_k^2} - 1 \right) \frac{2|X_k|^2}{\sigma_k^2} - \ln \frac{\sigma_k^2}{\tau_k^2} \right] \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{2}$$

where X_k is the kth component of the observation $\mathbf{X}$.

Under $\mathcal{H}_1$, the quantity $U = 2|X_k|^2/\sigma_k^2$ is a χ^2 random variable with 2 degrees of freedom (equivalently, exponential distributed with mean 2), so the conditional mean and second through fourth central moments of U given $\mathcal{H}_1$ are $2, 4, 16$, and 144, respectively. Letting L_k denote the kth summand, L is a sum of independent random variables with conditional mean and central moments

$$\mu(L_k|\mathcal{H}_1) = \frac{\sigma_k^2}{\tau_k^2} - \ln \frac{\sigma_k^2}{\tau_k^2} - 1, \tag{3}$$

$$\mu_2(L_k|\mathcal{H}_1) = \left(\frac{\sigma_k^2}{\tau_k^2} - 1 \right)^2, \tag{4}$$

$$\mu_3(L_k|\mathcal{H}_1) = 2 \left(\frac{\sigma_k^2}{\tau_k^2} - 1 \right)^3, \tag{5}$$

$$\mu_4(L_k|\mathcal{H}_1) = 9 \left(\frac{\sigma_k^2}{\tau_k^2} - 1 \right)^4. \tag{6}$$

Accordingly, the conditional mean, second, and third central moments of L are sums over K of $\mu(L_k|\mathcal{H}_1)$, $\mu_2(L_k|\mathcal{H}_1)$, and $\mu_3(L_k|\mathcal{H}_1)$, respectively. The fourth central moment of L is

$$\begin{aligned} \mu_4(L|\mathcal{H}_1) = &\sum_{k=1}^{K} 9 \left(\frac{\sigma_k^2}{\tau_k^2} - 1 \right)^4 \\ &+ 6 \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} \left(\frac{\sigma_k^2}{\tau_k^2} - 1 \right)^2 \left(\frac{\sigma_l^2}{\tau_l^2} - 1 \right)^2. \end{aligned}$$

From these expressions, we see that the coefficients of skewness and kurtosis for the L_k are

$$\sqrt{\beta_1(L_k)} = \frac{\mu_3(L_k)}{\mu_2^{3/2}(L_k)} = 2 \,\mathrm{sgn}(\sigma_k^2 - \tau_k^2)$$

and

$$\beta_2(L_k) = \frac{\mu_4(L_k)}{\mu_2^2(L_k)} = 9,$$

respectively. This combination of skewness and kurtosis corresponds to the S_B system within the Johnson family. It is clear that short observation vectors will not yield test statistics that follow a Gaussian distribution. For long observation vectors, however, L may be approximately Gaussian, and we can approximate the conditional probability of error under $\mathcal{H}_1$, $P_I = \Pr[L < \gamma|\mathcal{H}_1]$, as $\hat{P}_{I,G} = \Phi((\gamma - \mu(L))/\mu_2^{1/2}(L))$, where Φ is the cumulative distribution function for a standard Gaussian random variable. As will be shown in the examples, the Johnson family is more widely applicable and doesn't necessarily require long observation vectors to provide a good fit to L. Under a Johnson approximation we can approximate the conditional probability of error as $\hat{P}_{I,J} = F_J(\gamma; \mu(L), \mu_2(L), \mu_3(L), \mu_4(L))$, where F_J is the cumulative distribution function for a Johnson random variable.

3 Plug-In Decision Rule

If the true parameters σ_k^2 and τ_k^2 are unknown, then we are forced to consider alternative decision rules. One of the most popular is the generalized likelihood ratio (plug-in) test in which maximum likelihood estimates of these parameters are substituted into the optimal rule in place of the unknown values. Specifically, let

$$S_k^2 = \sum_{l=1}^{N_S} \left|V_{k,l}\right|^2 \quad \text{and} \quad T_k^2 = \sum_{l=1}^{N_T} \left|W_{k,l}\right|^2, \tag{7}$$

where the $V_{k,l}$ and $W_{k,l}$ represent the kth components of observations drawn under $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Then, $2N_S S_k^2/\sigma_k^2$ and $2N_T T_k^2/\tau_k^2$ follow χ^2 distributions with $2N_S$ and $2N_T$ degrees of freedom, respectively.

The plug-in decision rule is

$$R = \sum_{k=1}^{K} \left[\frac{1}{2} \left(\frac{\sigma_k^2}{T_k^2} - \frac{\sigma_k^2}{S_k^2} \right) \frac{2\left|X_k\right|^2}{\sigma_k^2} - \ln \frac{S_k^2}{T_k^2} \right] \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{8}$$

As before, we consider the conditional moments of the kth summand R_k over all possible observation vectors $\mathbf{X}$, but we must also account for randomness in the training samples. Because the random variables X_k, S_k^2, and T_k^2 are independent, the conditional mth raw moment of R_k can be written as

$$\mu'_m(R_k|\mathcal{H}_1) = \iiint R_k^m f_{X_k}(x) f_{S_k^2}(s^2) f_{T_k^2}(t^2)\, dx ds^2 dt^2, \tag{9}$$

where the limits of integration include the entire complex plane for x and the interval $[0, \infty)$ for s^2 and t^2. Analytical evaluation of these integrals is discussed in the Appendix. There, it is shown that the mth moment of R_k includes terms of the form $\int_0^\infty y^i e^{-y} dy$ for $i = N_S - m - 1$ and $i = N_T - m - 1$. These integrals diverge for i an integer less zero, thus, expressions for the mth moment of R_k are valid only for $N_S, N_T \geq m + 1$.

Denoting the logarithmic derivative of the gamma function as $\psi(z) = \partial \ln \Gamma(z)/\partial z$, the conditional mean and second through fourth central moments of R_k are found to be

$$\begin{aligned} \mu(R_k|\mathcal{H}_1) = {} & \frac{N_T}{N_T - 1} \frac{\sigma_k^2}{\tau_k^2} - \ln \frac{\sigma_k^2}{\tau_k^2} - \frac{N_S}{N_S - 1} + \ln \frac{N_S}{N_T} \\ & + \psi(N_T) - \psi(N_S), \quad (\text{for } N_S, N_T \geq 2), \end{aligned} \tag{10}$$

$$\begin{aligned} & \mu_2(R_k|\mathcal{H}_1) = \\ & \frac{N_T^3}{(N_T-1)^2(N_T-2)} \left(\frac{\sigma_k^2}{\tau_k^2} \right)^2 - 2 \frac{N_T(N_S N_T - 1)}{(N_T-1)^2(N_S-1)} \frac{\sigma_k^2}{\tau_k^2} \\ & + \frac{N_S(N_S^2 - 2N_S + 4)}{(N_S-1)^2(N_S-2)} + \psi'(N_S) + \psi'(N_T), \\ & (\text{for } N_S, N_T \geq 3), \end{aligned} \tag{11}$$

$$\begin{aligned} & \mu_3(R_k|\mathcal{H}_1) = \\ & \frac{2N_T^4(N_T+1)}{(N_T-1)^3(N_T-2)(N_T-3)} \left(\frac{\sigma_k^2}{\tau_k^2} \right)^3 \\ & + \frac{6N_T^2(N_T^2 - N_T - 1 + N_S - N_S N_T (N_T-1)^2)}{(N_T-1)^3(N_T-2)^2(N_S-1)} \left(\frac{\sigma_k^2}{\tau_k^2} \right)^2 \\ & + 6N_T \Big[N_S^3(N_T^2 - N_T + 1) - N_S^2(N_T^2 + N_T + 2) \\ & + N_S(2N_T^2 - 2N_T + 5) - 2 \Big] \\ & \Big/ \Big[(N_T-1)^3(N_S-1)^2(N_S-2) \Big] \left(\frac{\sigma_k^2}{\tau_k^2} \right) \\ & - \frac{2N_S(N_S^5 - 4N_S^4 + 13N_S^3 - 27N_S^2 + 39N_S - 36)}{(N_S-1)^3(N_S-2)^2(N_S-3)} \\ & - \psi''(N_S) + \psi''(N_T), \quad (\text{for } N_S, N_T \geq 4), \end{aligned} \tag{12}$$

and

$$\mu_4(R_k|\mathcal{H}_1) = \sum_{i=0}^{4} \left(\frac{\sigma_k^2}{\tau_k^2} \right)^i a_i, \quad (\text{for } N_S, N_T \geq 5), \tag{13}$$

where

$$\begin{aligned} a_0 = {} & \\ & 3N_S \big[3N_S^9 - 32N_S^8 + 179N_S^7 - 678N_S^6 + 1844N_S^5 \\ & - 3752N_S^4 + 5760N_S^3 - 6552N_S^2 + 5280N_S - 2304 \big] \\ & / \Big[(N_S-1)^4(N_S-2)^3(N_S-3)^2(N_S-4) \Big] \\ & + 3(\psi'^2(N_S) + \psi'^2(N_T)) + 6\psi'(N_T) + \psi'''(N_S) \\ & + \psi'''(N_T) + \frac{6N_S(N_S^2 - 2N_S + 4)}{(N_S-1)^2(N_S-2)} (\psi'(N_S) + \psi'(N_T)) \\ & + 6(\psi'(N_S)\psi'(N_T) - \psi'(N_T)), \end{aligned}$$

$$\begin{aligned} a_1 = {} & \\ & \frac{-12N_T N_S(3N_S^5 - 13N_S^4 + 34N_S^3 - 54N_S^2 + 44N_S - 24)}{(N_S-1)^3(N_S-2)^2(N_S-3)(N_T-1)} \\ & - \frac{12N_T N_S(3N_S^2 - 4N_S + 8)}{(N_S-1)^2(N_S-2)(N_T-1)^2} \\ & - \frac{24N_T N_S}{(N_S-1)(N_T-1)^3} - \frac{24N_T}{(N_T-1)^4} \\ & - \frac{12N_T(N_S N_T - 1)}{(N_T-1)^2(N_S-1)} (\psi'(N_S) + \psi'(N_T)), \end{aligned}$$

$$\begin{aligned} a_2 = {} & \\ & 6N_T^2 \big[N_S^3(9N_T^5 - 46N_T^4 + 89N_T^3 - 80N_T^2 + 28N_T + 4) \\ & - 2N_S^2(5N_T^5 - 14N_T^4 + 3N_T^3 + 22N_T^2 - 32N_T + 24) \\ & + 4N_S(3N_T^5 - 12N_T^4 + 19N_T^3 - 14N_T^2 - 4N_T + 13) \\ & - 8(2N_T^3 - 6N_T^2 + 4N_T + 1) \big] \\ & / \Big[(N_T-1)^4(N_T-2)^3(N_S-1)^2(N_S-2) \Big] \\ & + \frac{6N_T^3(\psi'(N_S) + \psi'(N_T))}{(N_T-1)^2(N_T-2)}, \end{aligned}$$

$$a_3 = -12N_T^3\big[3N_SN_T^5 - (16N_S+3)N_T^4 + (27N_S+12)N_T^3 - (18N_S+11)N_T^2 + (4N_S+2)N_T + 4N_S - 4\big] \Big/ \Big[(N_T^2 - 5N_T + 6)^2(N_T-1)^4(N_S-1)\Big],$$

and

$$a_4 = \frac{3N_T^5(3N_T^2 + N_T + 2)}{(N_T-1)^4(N_T-2)(N_T-3)(N_T-4)}.$$

It is easy to verify that, in the limit as $N_S, N_T \to \infty$, each of the moments above approaches the corresponding moment of the ideal test statistic L from Section 2. Moreover, it can be shown that when only one of N_S or N_T approach infinity and the other remains finite, the above moments approach those of a test statistic involving only one set of estimated parameters with the other set equal to the true values. This could be the case, for instance, in a problem of detecting the presence of a poorly known signal in a well-known channel or in problems of segmenting an object from a well-characterized background [17].

The conditional probability of error for the plug-in statistic is $P_P = \Pr[R < \gamma|\mathcal{H}_1]$, and we can estimate this with either $\hat{P}_{P,G} = \Phi((\gamma - \mu(R))/\mu_2^{1/2}(R))$ under a Gaussian approximation or as $\hat{P}_{P,J} = F_J(\gamma; \mu(R), \mu_2(R), \mu_3(R), \mu_4(R))$ under a Johnson approximation.

4 Unbiased Moment Estimators

The expressions for conditional moments in Sections 2 and 3 are functions of the parameters σ_k^2 and τ_k^2. While these parameters are unknown in practical recognition problems, unbiased estimates of them are generally available through (7). A direct approach would be to substitute these variance estimates into the expressions for conditional moments. However, this will yield biased moment estimators, and the resulting probability of error estimators will suffer, as is demonstrated in Section 6. As an alternative, we seek unbiased estimators for the conditional moments in terms of the unbiased parameter estimators S_k^2 and T_k^2. The conditional moments of L and R are expressed as sums of terms involving $(\sigma_k^2/\tau_k^2)^m$ and $\ln\sigma_k^2$ (similarly $\ln\tau_k^2$). Substituting unbiased estimators for these quantities into those expressions will yield unbiased estimators of the conditional moments. Employing the integrals discussed in the Appendix, we can derive the necessary expectations

$$\mathrm{E}\big[\ln S_k^2 + \ln N_S - \psi(N_S)\big] = \ln\sigma_k^2 \tag{14}$$

and

$$\mathrm{E}\left[\left(\frac{N_S}{N_T}\right)^m\left(\frac{\Gamma(N_t)}{\Gamma(N_t-m)}\right)\left(\frac{\Gamma(N_s)}{\Gamma(N_s+m)}\right)\left(\frac{S_k^2}{T_k^2}\right)^m\right] = \left(\frac{\sigma_k^2}{\tau_k^2}\right)^m \tag{15}$$

for $m > -N_S$, and $N_T - m$ neither zero nor a negative integer.

We must exercise caution since estimates of the moments $\hat{\mu}_m(L_k|\mathcal{H}_1)$ and $\hat{\mu}_m(R_k|\mathcal{H}_1)$ obtained in this way may assume invalid values. For example, the variance estimate $\hat{\mu}_2(L_k|\mathcal{H}_1)$ becomes negative if

$$\frac{S_k^2}{T_k^2} \in (c_L - d_L, c_L + d_L),$$

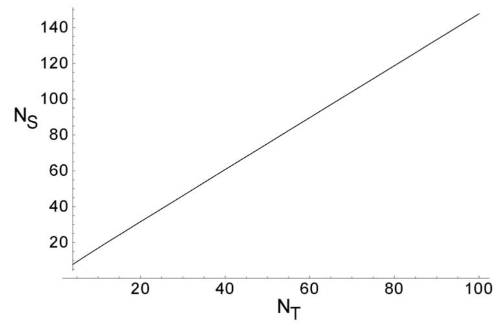


Fig. 1. Largest value of N_S to guarantee $\hat{\mu}_2(R_k) > 0$ as a function of N_T.

where

$$c_L = \frac{(N_S+1)N_T}{N_S(N_T-2)}$$
$$d_L = \frac{N_T\sqrt{(N_S+1)(N_T-1)(N_S+N_T-1)}}{N_S(N_T-1)(N_T-2)}.$$

Thus, it is always possible for this estimator to take negative values. By contrast, the variance estimate $\hat{\mu}_2(R_k|\mathcal{H}_1)$ becomes negative if

$$\frac{S_k^2}{T_k^2} \in (c_R - d_R, c_R + d_R),$$

where

$$c_R = \frac{(N_S+1)(N_SN_T-1)}{(N_S-1)N_SN_T}$$
$$d_R = \Big\{-(N_S-2)(N_S+1)(2 - N_S[N_TN_S^3 + (N_T-4)N_TN_S^2 - (6N_T^2 - 6N_T - 1)N_S + 4N_T + 1] + (N_S-2)(N_S-1)^2(N_T-1)N_SN_T[\psi'(N_S) + \psi'(N_T)])\Big\}^{1/2} / [(N_S-2)(N_S-1)N_SN_T].$$

As long as N_S is not much larger than N_T, the quantity d_R is imaginary, in which case, the variance estimate for R_k cannot be negative. Fig. 1 shows the largest possible value of N_S to ensure $\hat{\mu}_2(R_k|\mathcal{H}_1) > 0$ as a function of N_T. The estimators $\hat{\mu}_4(L)$ and $\hat{\mu}_4(R)$ can also yield negative values. In fact, these estimators can violate an even stricter inequality that must be satisfied for all distributions [18]

$$\mu_4 \geq \mu_3^2/\mu_2 + \mu_2^2. \tag{16}$$

Because $\hat{\mu}_m(L_k|\mathcal{H}_1)$ and $\hat{\mu}_m(R_k|\mathcal{H}_1)$ may violate known bounds for some values of k, the overall test statistic moments $\hat{\mu}_m(L|\mathcal{H}_1)$ and $\hat{\mu}_m(R|\mathcal{H}_1)$ may violate these bounds as well. Nevertheless, this is unlikely to occur if the number of terms, K, is not too small. In the examples of the following sections, we enforce these bounds by substituting a small positive quantity for $\hat{\mu}_2(L|\mathcal{H}_1)$ or $\hat{\mu}_2(R|\mathcal{H}_1)$ whenever the variance estimates are negative and by assigning $\hat{\mu}_4(L|\mathcal{H}_1)$ or $\hat{\mu}_4(R|\mathcal{H}_1)$ according to (16) whenever the fourth central moments are too small. An alternative approach is to substitute the biased estimators whenever the corresponding unbiased estimator yields an invalid value.

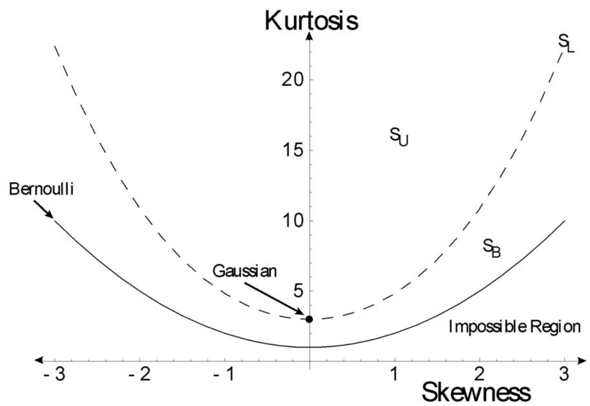


Fig. 2. Regions of the skewness-kurtosis plane that correspond to the three systems of the Johnson family of distributions.

5 Johnson's Family of Distributions

Johnson's family of probability distributions [15] is a four-parameter family and is defined in terms of three simple transformations of a Gaussian random variable. A random variable Y in the Johnson family belongs to one of three sub-families (or systems) referred to as S_U, S_L, and S_B, which are defined as

$$\text{For } S_U : Y = \xi + \lambda \sinh\left(\frac{Z-\gamma}{\delta}\right), \tag{17}$$

$$\text{For } S_L : Y = \xi + \lambda e^{\frac{Z-\gamma}{\delta}}, \quad \lambda = \pm 1, \tag{18}$$

$$\text{For } S_B : Y = \xi + \lambda\left(1 + e^{-\frac{Z-\gamma}{\delta}}\right)^{-1}, \tag{19}$$

where Z is a standard Gaussian random variable, the quantities γ, $\delta > 0$, λ, and ξ are real-valued free parameters, and by equality we mean equality in distribution. The S_U (or unbounded) system has a probability density function with infinite support. The S_L system is commonly known as the lognormal family and has a density function with semi-infinite support. Finally, the S_B (or bounded) system has a density with finite support. Calculation of the cumulative distribution and probability density functions for members in the family can be performed simply in terms of Gaussian distribution and density functions.

The family contains a unique distribution for every possible combination of moments 1 through 4 and, thus, is useful for fitting distributions to the moments derived in previous sections. Fig. 2 shows the regions of the skewness-kurtosis $(\sqrt{\beta_1}, \beta_2)$ plane corresponding to each of the three systems. The solid curve at the bottom of the figure is given by $\beta_2 = \beta_1 + 1$, and it represents the smallest possible β_2 for any distribution as a function of $\sqrt{\beta_1}$. This curve, which corresponds to a discrete distribution with all probability mass concentrated on two real values, is contained within the S_B system. The dashed curve represents the S_L system, which is a limiting form for the S_B system below and the S_U system above. The S_L curve is given by the parametric relations $\beta_1 = (\omega - 1)(\omega + 2)^2$ and $\beta_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3$, for $\omega \geq 1$. The Gaussian family is represented by the point $(0, 3)$ and is a limiting distribution for all three systems. Algorithms for calculating moments from distribution parameters and vice versa are available [19], [20].

To approximate the conditional probability of error, we will evaluate the tail probabilities of Johnson distributions which are fitted to a distribution involving χ^2 random variables. It is instructive to compare the tail probabilities of a χ^2 random variable to the best-fit Johnson distribution. Let z_l and z_r denote the values of a χ^2 random variable with N degrees of freedom that yield left and right tail probabilities equal to α. That is, $F_{\chi^2_N}(z_l) = \alpha$ and $1 - F_{\chi^2_N}(z_r) = \alpha$, where $F_{\chi^2_N}$ is the cumulative distribution of a χ^2 random variable with N degrees of freedom. The probability in the corresponding tails of an approximating distribution is $\hat{\alpha}_l = F_J(z_l)$ and $\hat{\alpha}_r = 1 - F_J(z_r)$. Plots of the relative error in the tail probabilities, $(\alpha - \alpha_l)/\alpha$ and $(\alpha - \alpha_r)/\alpha$, versus N are shown in Fig. 3 for several values of α. The figure shows that there is significant error in the left tail (near zero) for small N. The fit becomes exact with increasing N, and the figure suggests that convergence is quite fast. The reason for the poor fit of the left tail can be seen in Fig. 4, which shows that the best fit Johnson distribution can take negative values (≈ 0.05 probability

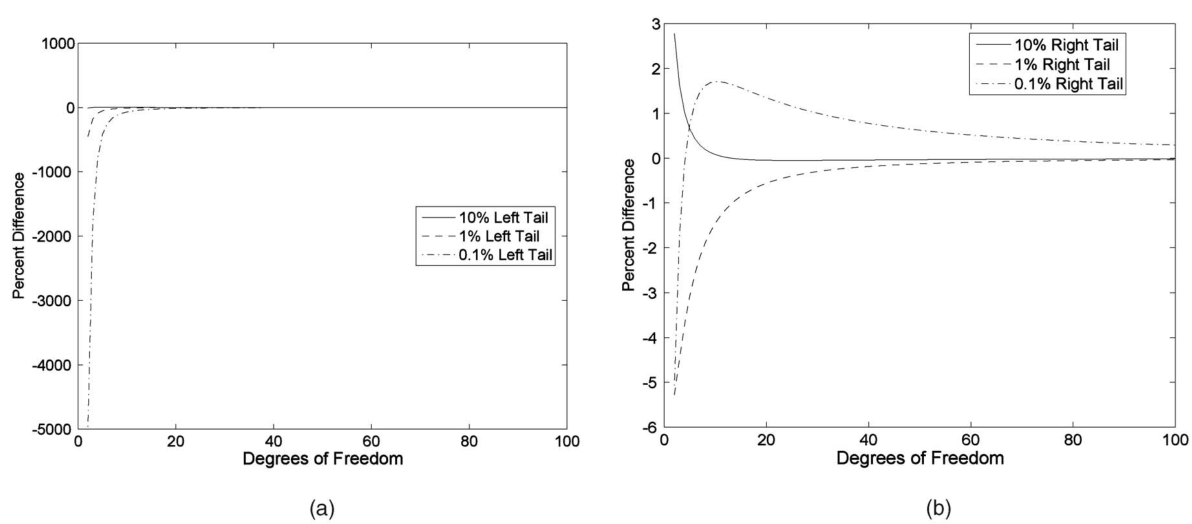


Fig. 3. Relative error in the left and right tails with a Johnson approximation to a χ^2 distribution. (a) Left tail. (b) Right tail.

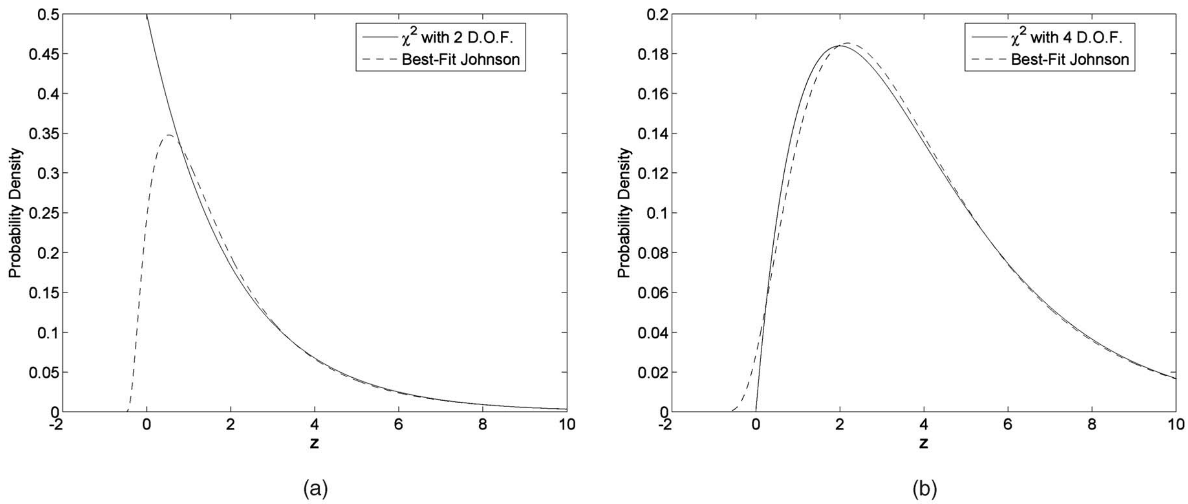


Fig. 4. Comparison of χ^2 probability density functions with the best-fit Johnson approximation. (a) χ^2 with 2 degrees of freedom. (b) χ^2 with 4 degrees of freedom.

mass is left of zero). However, for greater degrees of freedom, this lack of fit becomes negligible, as the right panel of the figure suggests.

The lack of fit for small N does not suggest that the Johnson approximation is inappropriate for our application. On the contrary, the Johnson family is used to fit combinations of χ^2 random variables as in (2) and (8), and these results suggest that our error predictions can be accurate as long as there are more than a couple of terms in the sum. Note that a χ^2 random variable with $2N$ degrees of freedom, to which we have compared, arises as the sum of N independent χ^2 random variables with 2 degrees of freedom. Arbitrary linear combinations of χ^2 random variables have been studied extensively (cf., Mathai and Provost [21]), so a corresponding analysis could be conducted to explore the Johnson approximation for any particular combination of σ_k^2 and τ_k^2. The resulting probability density functions have a somewhat complicated form, which may make established numerical approximation methods, such as the saddle-point approximation [22], attractive (cf., Kay et al. [23] for an application to linear combinations of χ^2 random variables).

6 Example

In this section, we demonstrate the approximating distributions, conditional moment estimators, and conditional probability of error estimators on a simple pair of complex Gaussian hypotheses characterized by

$$\sigma_k^2 = \begin{cases} 1 & \text{for } k \text{ odd} \\ 2 & \text{for } k \text{ even} \end{cases}$$

$$\tau_k^2 = \begin{cases} 2 & \text{for } k \text{ odd} \\ 1 & \text{for } k \text{ even} \end{cases}$$

for $k = 1, 2, \ldots, K$. The numerical results were produced by conducting 100,000 independent experiments in which training samples of size N were drawn according to the two hypotheses and used to construct a plug-in decision rule. In each experiment, an additional observation was drawn according to $\mathcal{H}_1$, and both the ideal and plug-in test statistics were computed from this observation. The experiments were performed for vector lengths $K = 1, 20,$ and 100, and for $N = 7$ and 15 training samples. These combinations were chosen because they reveal key properties of the various estimators, and these are emphasized below.

Fig. 5 demonstrates that for observations with more than a few components, the Johnson family together with the theoretical moments previously derived is a good fit to the test statistics L and R. The figure shows the sample probability density function for the test statistics in the plug-in and ideal decision rules for three combinations of K and N. Overlaid on each plot are the closest distributions from the Gaussian and Johnson families, which were obtained by fitting the theoretical moments from Sections 2 and 3 (not by fitting the sample moments). The Johnson family is consistently a superior fit, but the Gaussian family becomes adequate for long observation vectors. Except when short vectors and small training samples are used, the Johnson and sample PDFs are nearly indistinguishable. As the observation vectors get longer, the test statistics become less skewed and have lower excess kurtosis, so beyond $K = 100$, the Gaussian approximation is quite reasonable.

Fig. 6 demonstrates reasonable behavior in the unbiased estimators of Section 4 and suggests their superiority over the biased estimators for the plug-in rule. The figure shows sample probability density functions for both the biased and unbiased moment estimators of Section 4 normalized by the corresponding actual moment for the $K = 20$, $N = 15$ case. For the unbiased estimators, the sample average of each of these ratios is indeed approximately one and it can be seen that moment estimators for the ideal test statistic are more heavily skewed than those for the plug-in test statistic. As previously discussed, the lower bound in (16) can be violated by the unbiased moment estimators, particularly for short observation vectors and/or small training samples. Table 1 shows that, for short observation vectors, most of the estimates $\hat{\mu}_4(L|\mathcal{H}_1)$ had to be corrected because they violated this lower bound. This was rarely a problem with long observation vectors or with the estimates $\hat{\mu}_4(R|\mathcal{H}_1)$. Such corrections are not necessary with the biased moment estimators, which for the ideal test statistic appear reasonable,

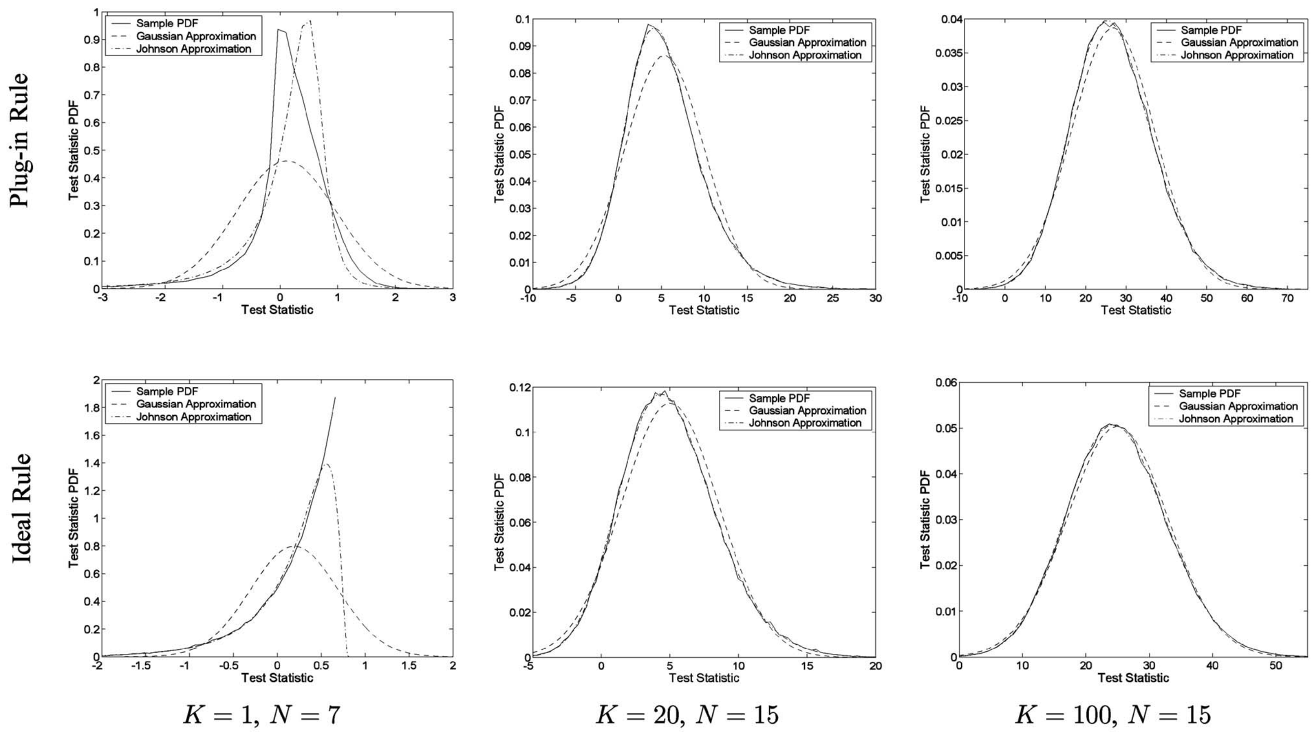


Fig. 5. Sample distributions of plug-in and ideal test statistics under $\mathcal{H}_1$ with best fit members from the Gaussian and Johnson families.

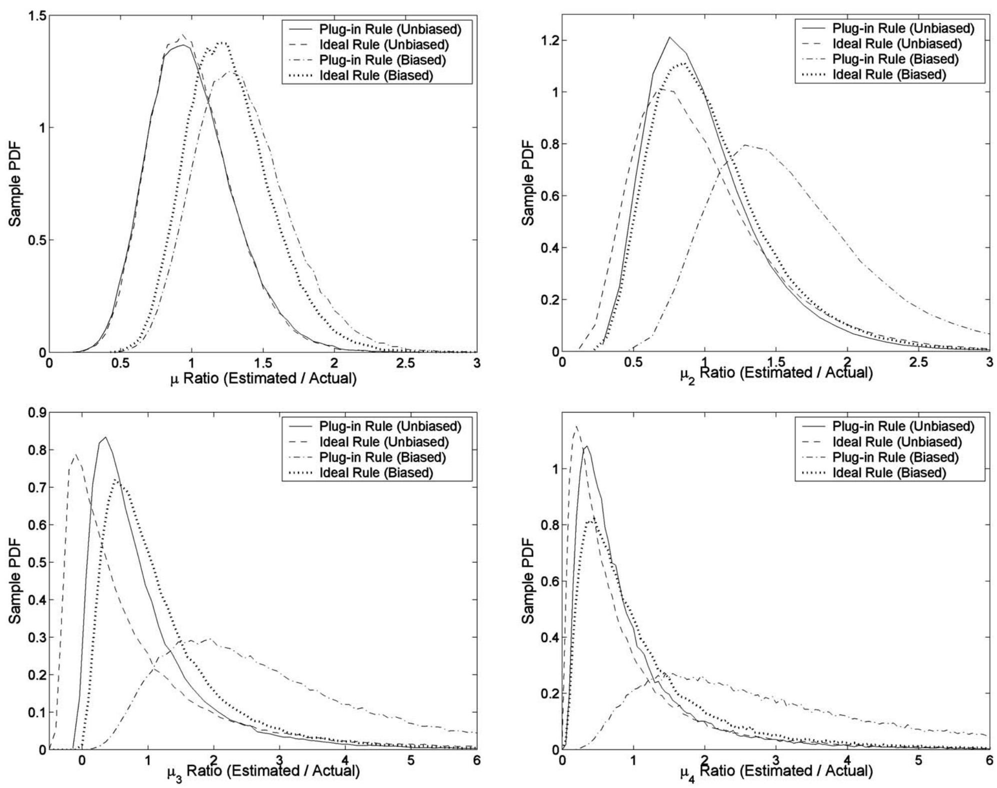


Fig. 6. Sample distribution of estimated moments for the plug-in and ideal test statistics under $\mathcal{H}_1$ using length 20 observation vectors and training samples of size 15.

though there is a slight tendency to produce estimates that are too large. However, the biased moment estimator for the plug-in test statistic is comparatively quite poor, with a significant positive bias and extremely long tails.

Fig. 7 shows that the use of biased moment estimators results in error estimates that are generally optimistic, while the use of unbiased estimators, though resulting in a larger variance, yields, error estimates that are approximately

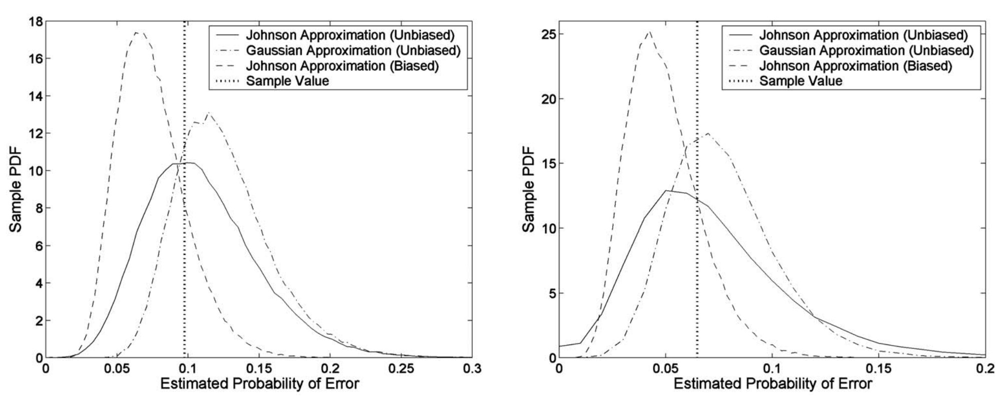
TABLE 1
Percentage of Experiments in which Unbiased Estimates $\hat{\mu}_4$ Violated Known Bounds

	Plug-In Rule		Ideal Rule	
	$N=7$	$N=15$	$N=7$	$N=15$
$K=1$	4.6%	3.6%	62%	51%
$K=20$	0%	0%	72%	0.15%
$K=100$	0%	0%	0.21%	0%

unbiased. The figure shows the sample distribution of estimated conditional probability of error under both the Johnson and Gaussian approximations with unbiased moment estimators. Also shown are the Johnson approximation to probability of error using biased moment estimators and the sample probabilities of error estimated through Monte Carlo simulation. For both approximations with unbiased moment estimators, the estimated error rates are approximately centered on the sample error rate. In this example, the estimates obtained under a Johnson approximation have a smaller bias but a larger variance than those obtained under a Gaussian approximation. This is a case of approximation error versus estimation error tradeoff and is a reflection of the fact that the Johnson family is a better fit, but the required third and fourth-order moments have high variability. When biased moment estimators are used, the probability of error tends to be significantly underestimated, though there is smaller variance in the estimates.

The Johnson approximation becomes clearly superior to the Gaussian approximation as either the length of observation vectors or the training sample size grow. This is illustrated in Tables 2 and 3 which summarize the behavior of the unbiased estimators for both the plug-in and ideal decision rules. These tables show the sample mean and standard deviation both of normalized moment estimators and of the difference between estimated and sample probability of error, all conditioned on $\mathcal{H}_1$. The tables suggest that the estimator for plug-in test statistic mean has a higher variance than that for the ideal test statistic. For higher-order moments, however, estimators for the plug-in rule generally have lower bias and lower variance than estimators for the ideal rule. The fourth-order moment estimator for the ideal test statistic exhibits extreme variability when observation vectors are short and training samples are small. The observed bias is due to the large number of estimates that were beneath the known lower bound and required upward correction, as previously discussed. Note that the catastrophic behavior of this estimator ceases when the vector length and training sample size increase to moderate values.

The last two lines of each table summarize the difference between estimated and observed conditional probabilities of error. The subscripts P and I indicate plug-in and ideal decision rules, respectively, and the subscripts G and J indicate Gaussian and Johnson approximations, respectively. The results suggest that unbiased moment estimators do not lead to unbiased error probability estimators, as both approximations demonstrate a positive bias. However, both become quite accurate with increases in observation vector



(a) (b)

Fig. 7. Sample distribution of estimated conditional probability of error given $\mathcal{H}_1$ for (a) plug-in and (b) ideal rules under Gaussian and Johnson approximations using length $K=20$ observation vectors and training samples of size $N=15$.

TABLE 2
Summary of Plug-In Decision Rule Unbiased Estimates by Observation Vector Length and Training Sample Size N

	Length 1				Length 20				Length 100			
	$N=7$		$N=15$		$N=7$		$N=15$		$N=7$		$N=15$	
	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.
$\hat{\mu}(R)/\mu(R)$	0.9998	2.4598	0.9959	1.1489	1.0003	0.5309	1.0006	0.3031	0.9998	0.2370	0.9998	0.1347
$\hat{\mu}_2(R)/\mu_2(R)$	0.9993	0.6384	0.9979	0.5094	0.9992	0.7947	1.0013	0.4294	0.9990	0.3574	1.0000	0.1892
$\hat{\mu}_3(R)/\mu_3(R)$	0.9976	1.4425	0.9978	0.5786	0.9966	3.8219	1.0050	1.2193	0.9956	1.8652	0.9996	0.5184
$\hat{\mu}_4(R)/\mu_4(R)$	1.007	2.4626	1.0017	0.6315	0.9862	7.8110	1.0067	1.4101	0.9954	2.0413	0.9995	0.4204
$\hat{P}_{P,G}-P_P$	0.1553	0.1124	0.1534	0.0973	0.0437	0.0744	0.0274	0.0339	0.0108	0.0141	0.0030	0.0024
$\hat{P}_{P,J}-P_P$	0.0760	0.2368	0.0586	0.1609	0.0241	0.0795	0.0122	0.0399	0.0022	0.0138	-9.08E-5	0.0016

The indicated moments and probabilities of error are conditional on $\mathcal{H}_1$.

TABLE 3
Summary of Ideal Decision Rule Unbiased Estimates by Observation Vector Length and Training Sample Size N

	Length 1				Length 20				Length 100			
	$N=7$		$N=15$		$N=7$		$N=15$		$N=7$		$N=15$	
	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.	mean	std.
$\hat{\mu}(L)/\mu(L)$	0.9991	1.5624	0.9965	0.9970	1.0009	0.5083	1.0005	0.2960	0.9999	0.2269	0.9998	0.1316
$\hat{\mu}_2(L)/\mu_2(L)$	1.1494	1.0615	1.0280	0.6968	1.0001	1.2376	1.0015	0.5258	0.9985	0.5572	1.0000	0.2315
$\hat{\mu}_3(L)/\mu_3(L)$	0.9950	2.8208	0.9947	1.1690	0.9918	10.914	1.0088	2.0809	0.9884	5.4335	0.9991	0.8747
$\hat{\mu}_4(L)/\mu_4(L)$	2.50E14	8.69E14	1.27E13	4.18E13	4.52E12	4.32E13	1.0088	1.7661	0.9963	2.3572	0.9993	0.5064
$\hat{P}_{I,G} - P_I$	0.2495	0.2867	0.1844	0.2145	-0.0032	0.0549	0.0121	0.0245	0.0007	0.0017	0.0005	0.0006
$\hat{P}_{I,J} - P_I$	0.2378	0.3656	0.1530	0.3114	-0.0152	0.1014	0.0076	0.0389	-0.0003	0.0003	-0.0002	0.0002

The indicated moments and probabilities of error are conditional on $\mathcal{H}_1$.

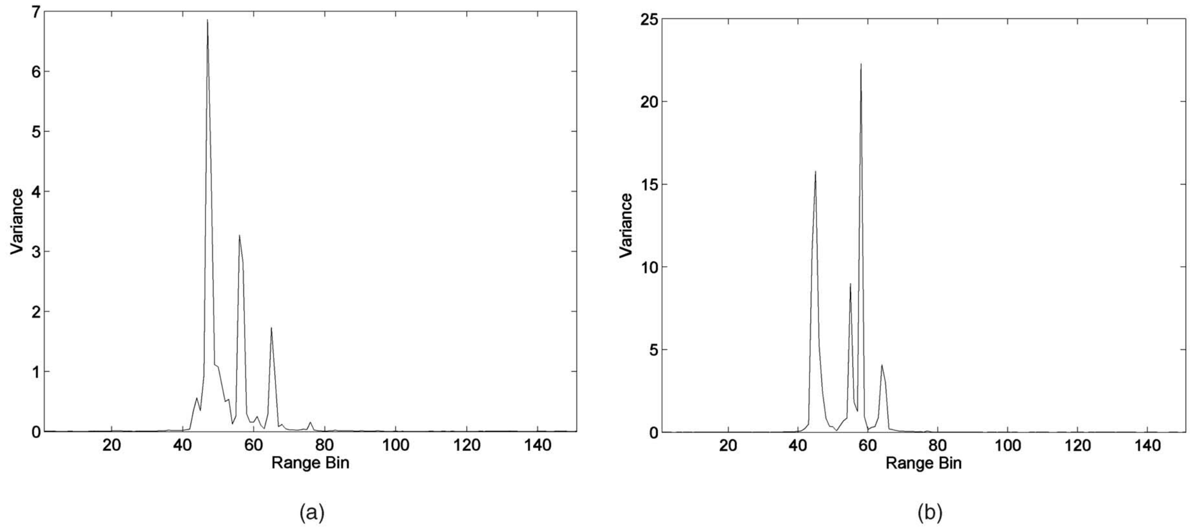


Fig. 8. Variance of the range profiles as a function of range bin for (a) T1 and (b) M1 tanks. Observation vectors are dominated by a small number of components.

length and/or training sample size. For extremely short observations, the estimators are not particularly valuable unless very large training samples are available. As previously discussed, the conditional distributions of L and R are decidedly non-Gaussian when the observations vectors are short. Thus, the bias in probability of error estimates under a Gaussian approximation is consistently higher than under the Johnson approximation, even when there is huge bias in the fourth-order moment estimate. Beyond length 20 observations, however, the estimators perform quite well, with the Johnson approximation delivering lower bias and lower variance than the Gaussian.

7 Applications

The ability to produce estimates of the conditional probability of error directly from trained parameters can be exploited in a number of applications, examples of which we consider in this section. The data used are simulated high resolution radar profiles of the vehicles "T1 tank" and "M1 tank" from the University Research Initiative Synthetic Dataset (URISD) [24]. The data are simulated from CAD vehicle models using the signature prediction tool XPATCH. Signatures used in the experiment are vectors of $K = 151$ elements and represent UHF band measurements from a 10 degree elevation angle with the object nominally facing the radar.

Fig. 8 shows variance values estimated from 20 observations spanning a 6 degree range in azimuth for both vehicles. The variance profiles of Fig. 8 are taken as the "true" variance parameters σ_k^2 and τ_k^2 in a test of $\mathcal{H}_1$, an assertion that an observation represents the T1 tank, versus $\mathcal{H}_2$, an assertion that an observation represents the M1 tank. The applications we address are concerned with the unconditional probability of error, which can be estimated on the basis of two conditional error probabilities of the form explored in the previous section. We use the minimum probability of error criterion with equal prior probabilities on the two hypotheses.

7.1 Value of Additional Training Data

It is often desirable to understand the potential value of increasing the training sample size on the basis of existing samples. This is true, for example, when data collection is expensive or time consuming. Because the plug-in decision rule is consistent when the underlying probability families are chosen correctly, the ideal probability of error represents the limiting plug-in error rate as ever more training data is collected. That is, the difference between the plug-in probability of error and the ideal probability of error represents the maximum possible benefit from data collection. The development in previous sections has provided estimators for both these quantities in terms of available sample data, and their difference represents an estimate of the best-case value of additional data collection. Fig. 9 shows the approximate decrease in probability of error as a function of training set size for the two-class radar profile problem. The

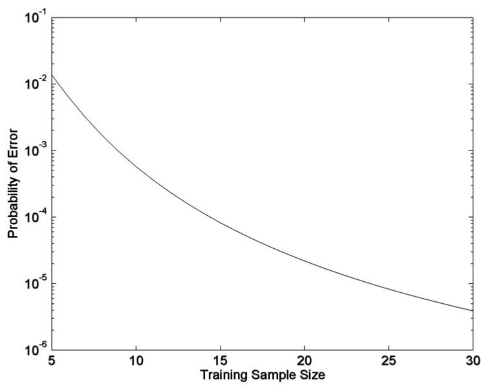


Fig. 9. Probability of error as a function of training sample size.

curve illustrates the kind of diminishing return that can be expected in this problem.

7.2 Value of Fused Observations

The variance profiles of Fig. 8 were estimated from only a small interval of azimuth angles between the vehicles and the radar platform. It is often instructive to evaluate the extent to which error probability depends on azimuth angle. In particular, one may be concerned with whether there are some angles for which performance will be especially poor and how that situation might be mitigated. Fig. 10a shows the approximate probability of error using a Johnson approximation as a function of target azimuth relative to the radar platform (at 0 degrees, the target is facing the radar platform). The solid line shows the probability of error when a single observation is collected and indicates that several azimuth angles are expected to be problematic, including 45 degrees, 90 degrees, 225 degrees, and 270 degrees.

If the indicated error rates at these angles is unacceptable, a possible solution is to collect two independent observations before classifying the vehicle. The approximate error rate for two observations is indicated by the dashed line in the left panel. It indicates the extent of error reduction, which represents the incremental value of the second observation. If two observations are to be collected, it is reasonable to ask whether it is preferable on average to collect them from along two different azimuth angles. The right panel shows the average probability of error over all azimuth angles as a function of the angular separation between the two observations. The graph indicates that, for distinguishing between these two vehicles, two observations from the same position will be better by approximately $1/10$ of a percentage point. While this seems counterintuitive, closer inspection of the results indicates that, for example, a 90 degree separation frequently results in pairs of angles (such as (0 degree, 270 degree)) that are individually problematic and together yield much higher error rates than other combinations.

7.3 Feature Selection

It is well-known that pattern recognition systems based on training data frequently exhibit what is known as the "peaking phenomenon," a tendency for recognition accuracy to degrade as more features are included in observation vectors. Because they are simple to compute, the Gaussian or Johnson approximations can be used to determine the subset of features that will minimize the estimated probability of error. In general, this requires an exhaustive search [25], which may not be unreasonable for short observation vectors but quickly becomes prohibitive as the number of components involved increases. Because of the exponential number of possible subsets, a wide variety of fast but suboptimal feature selection algorithms have been developed [26]. These generally operate on the basis of metrics which, while not directly related to the probability of error, suggest an approximate relative value of feature subsets. Moment-based error estimation can be used together with suboptimal feature selection algorithms to estimate the probability of error they will deliver. They can even be used directly in such an algorithm as the objective function to be minimized, eliminating the need for alternate metrics.

One of the simplest such algorithms is a "greedy" search that, in the first step, selects the single best feature and, in subsequent steps, adds to the set of selected features the single feature which, in combination with those already selected, yields the lowest estimated probability of error. The order in which features were selected is recorded and the first K of them are used in the pattern recognition algorithm, where K is the number that gave the lowest estimated probability of error. While there are clearly more sophisticated algorithms, we choose it for demonstration because of its simplicity.

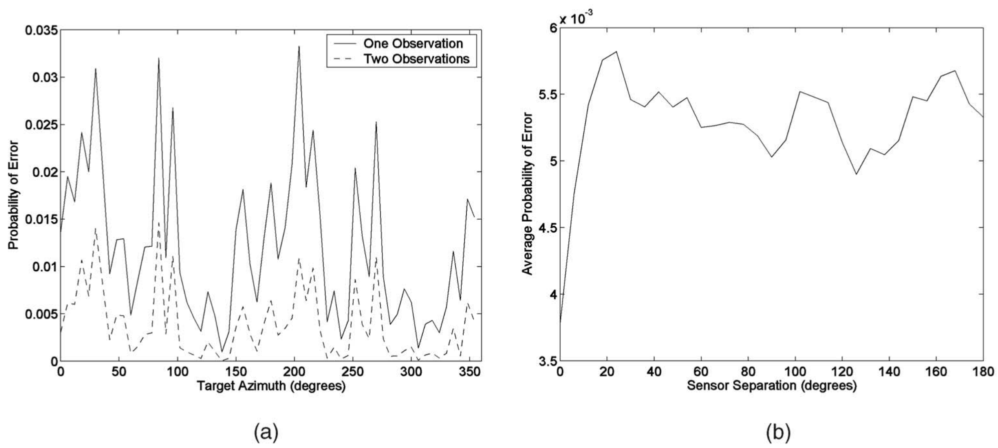


Fig. 10. (a) Angle dependence. Dependence of probability of error on target azimuth angle for one and two observations. (b) Separation dependence. Average probability of error as a function of the separation between two sensor observations.

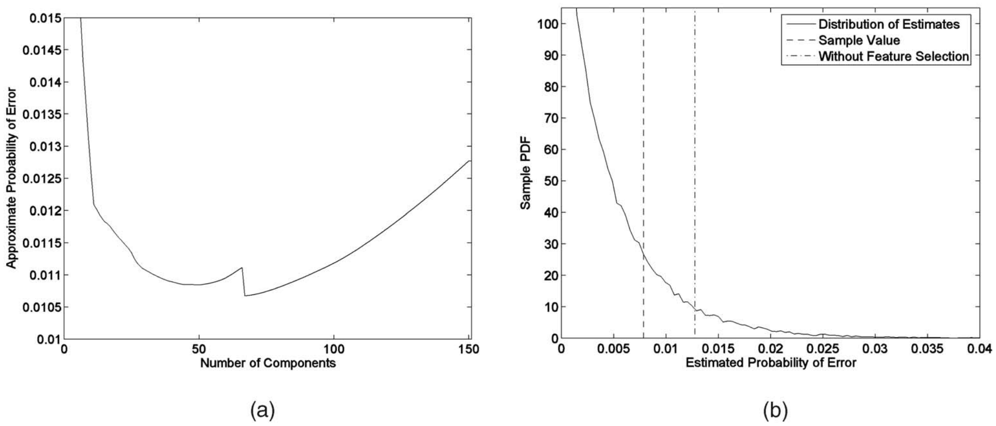


Fig. 11. (a) Peaking phenomenon. Plug-in probability of error as a function of feature vector length with components selected by a greedy algorithm exploiting actual variance parameters. (b) Adaptive feature selection. Sample distribution of estimated error with feature selection. Also shown are the sample error rate with feature selection and the error rate when no feature selection is employed.

The left panel of Fig. 11 shows the unconditional probability of error achievable by the plug-in decision rule as a function of feature vector length obtained by the greedy search method. Features were ordered using actual (not estimated) variance parameters and the minimum probability of error along the curve is 0.0107 which occurs with 67 components. The slight aberration near the bottom of the curve results from nonoptimality of the greedy algorithm. Fig. 11a shows the sample distribution of estimated probability of error when feature selection is based on five training samples. The dashed vertical line represents the sample probability of error in 100,000 trials incorporating the feature selection algorithm. The dash-dotted line to the right indicates the probability of error when no feature selection is employed. The results indicate that feature selection based on probability of error estimates from plug-in statistics can significantly improve recognition accuracy.

8 Conclusions

Our focus has been on the classification accuracy of recognition algorithms for observations from complex Gaussian hypotheses, which are assumed in a variety of communication and imaging problems. We have derived expressions for the conditional moments of ideal (optimal) and plug-in decision rules and have developed unbiased estimators for those moments in terms of training samples. With these moments, we form estimators for probability of classification error that are, in most cases, quite accurate and which can be computed much easier than exact probability of error expressions. Two pairs of probability of error estimators were considered, one based on approximating the test statistics with a distribution from the Johnson family and the other based on a simpler Gaussian family approximation.

We have performed a series of experiments on a simple classification problem to study estimator behavior under varying observation vector lengths and training sample sizes. We demonstrated that the Gaussian distribution is not a good fit to the test statistic distribution when observation vectors are short, regardless of the quantity of training data available. By contrast, the Johnson distribution is a good fit even when the observation vectors have only a few components. However, large training samples are required for such short vectors because of large variability in the required estimates of third and fourth-order moments. For moderate observation vector lengths, the Johnson distribution together with estimated test statistic moments provides a good fit and accurate probability of error estimates can result. For long observation vectors, both Johnson and Gaussian distributions provide an excellent fit.

Because of their accuracy and ease of use, these probability of error estimators can be used effectively in a variety of applications. We have demonstrated their use in quantifying the benefit of collecting additional training data, forming recommended collection geometries in data fusion problems, and selection of features to address the so-called "peaking phenomenon."

Appendix

Evaluation of Test Statistic Moments

Here, we outline the solution to the triple integral (9). The integrand consists of four multiplicands:

$$R_k^m = \left[\frac{1}{2}\left(\frac{\sigma_k^2}{t^2}\frac{\sigma_k^2}{s^2}\right)u - \ln\frac{s^2}{t^2}\right]^m,$$

$$f_U(u) = \frac{1}{2}e^{-\frac{u}{2}}, \text{for } u \geq 0,$$

$$f_{S_k^2}(s^2) = \left(\frac{N_s}{\sigma_k^2}\right)^{N_s}\frac{s^{2(N_s-1)}}{\Gamma(N_s)}e^{-\frac{N_s s^2}{\sigma_k^2}},$$

$$f_{T_k^2}(t^2) = \left(\frac{N_t}{\tau_k^2}\right)^{N_t}\frac{t^{2(N_t-1)}}{\Gamma(N_t)}e^{-\frac{N_t t^2}{\tau_k^2}},$$

where we have made the change of variable $U = 2|X_k|^2/\sigma_k^2$. With the subsequent change of variables $s^2 \rightarrow \frac{\sigma_k^2}{N_s}s^2$, $t^2 \rightarrow \frac{\tau_k^2}{N_t}t^2$, and $u \rightarrow 2u$, the integrand can be expanded into a series with terms of the form

$$C \cdot (s^2)^{i_1}\ln^{j_1}(s^2)e^{-s^2} \cdot (t^2)^{i_2}\ln^{j_2}(t^2)e^{-t^2} \cdot u^{i_3}e^{-u},$$

where C is not a function of $s^2, t^2,$ or u, i_1 is an integer value in the range from $N_s - 1 - m$ to $N_s - 1$, i_2 is an integer value in the range $N_t - 1 - m$ to $N_t - 1$, and $j_1, j_2,$ and i_3 are integer values in the range from 0 to m. These terms can be integrated by noting that integrals

$$\int_0^\infty y^i \ln^j(y) e^{-y} dy \tag{20}$$

are solved by taking the jth logarithmic derivative of the gamma function. This yields, successively,

$$\Gamma(n+1) = \int_0^\infty y^n e^{-y} dy$$
$$\Gamma(n+1)\psi(n+1) = \int_0^\infty y^n \ln(y) e^{-y} dy$$
$$\Gamma(n+1)(\psi^2(n+1) + \psi'(n+1)) = \int_0^\infty y^n \ln^2(y) e^{-y} dy$$
$$\vdots,$$

where ψ and ψ' denote the digamma and trigamma functions, respectively. Software implementations of arbitrary polygamma functions are readily available.

Acknowledgments

This work was supported by the US Office of Naval Research grant N00014-03-1-0110.

References

[1] H.L. Van Trees, *Detection, Estimation, and Modulation Theory,* vol. I, John Wiley & Sons, 1968.

[2] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag, 1996.

[3] D.L. Fried, "Statistics of the Laser Radar Cross Section of a Randomly Rough Target," *J. Optical Soc. Am.,* vol. 66, no. 11, pp. 1150-1160, Nov. 1976.

[4] J.W. Goodman, "Some Fundamental Properties of Speckle," *J. Optical Soc. Am.,* vol. 66, no. 11, pp. 1145-1150, Nov. 1976.

[5] H.L. Van Trees, *Detection, Estimation, and Modulation Theory. Radar-Sonar Signal Processing and Gaussian Signals in Noise,* vol. III, John Wiley & Sons, 1971.

[6] M.D. DeVore and J.A. O'Sullivan, "Quantitative Statistical Assessment of Conditional Models for Synthetic Aperture Radar," *IEEE Trans. Image Processing,* vol. 13, no. 2, pp. 113-125, Feb. 2003.

[7] J.A. O'Sullivan, M.D. DeVore, V. Kedia, and M.I. Miller, "Automatic Target Recognition Performance for SAR Imagery Using a Conditionally Gaussian Model," *IEEE Trans. Aerospace and Electronic Systems,* vol. 37, no. 1, pp. 91-108, Jan. 2001.

[8] M.D. DeVore and J.A. O'Sullivan, "A Performance-Complexity Study of Several Approaches to Automatic Target Recognition from SAR Images," *IEEE Trans. Aerospace and Electronic Systems,* vol. 38, no. 2, pp. 632-648, Apr. 2000.

[9] M.D. DeVore and J.A. O'Sullivan, "Target-Centered Models and Information-Theoretic Segmentation for Automatic Target Recognition," *Multidimensional Systems and Signal Processing,* vol. 14, pp. 139-159, Jan. 2003.

[10] S.P. Jacobs and J.A. O'Sullivan, "Automatic Target Recognition Using Sequences of High Resolution Radar Range-Profiles," *IEEE Trans. Aerospace and Electronic Systems,* vol. 36, no. 2, pp. 364-382, Apr. 2000.

[11] S.J. Raudys and A.K. Jain, "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 3, pp. 252-264, Mar. 1991.

[12] N. Glick, "Sample-Based Classification Procedures Derived from Density Estimators," *J. Am. Statistical Assoc.,* vol. 67, no. 337, pp. 116-122, Mar. 1972.

[13] S.J. Raudys and V. Pikelis, "On Dimensionality, Sample Size, Classification Error, and Complexity of Classification Algorithm in Pattern Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 2, no. 3, pp. 242-252, May 1980.

[14] N.L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions,* vol. 1, John Wiley & Sons, 1994.

[15] N.L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika,* vol. 36, no. 1/2, pp. 149-176, June 1949.

[16] M.D. DeVore, "Analytical Performance Evaluation of SAR ATR with Inaccurate or Estimated Models," *Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery XI,* E.G. Zelnio and F.D. Garber, eds., vol. 5427, pp. 407-417, 2004.

[17] N.A. Schmid and J.A. O'Sullivan, "Thresholding Method for Reduction of Dimensionality," *IEEE Trans. Information Theory,* vol. 47, no. 7, pp. 2903-2920, Nov. 2001.

[18] V.K. Rohatgi and G.J. Székely, "Sharp Inequalities between Skewness and Kurtosis," *Statistics and Probability Letters,* vol. 8, pp. 297-299, Sept. 1989.

[19] I.D. Hill, R. Hill, and R.L. Holder, "Algorithm AS 99: Fitting Johnson Curves by Moments," *Applied Statistics,* vol. 25, no. 2, pp. 180-189, 1976.

[20] I.D. Hill, "Algorithm AS 100: Normal-Johnson and Johnson-Normal Transformations," *Applied Statistics,* vol. 25, no. 2, pp. 190-192, 1976.

[21] A.M. Mathai and S.B. Provost, *Quadratic Forms in Random Variables: Theory and Applications.* Marcel Dekker, 1992.

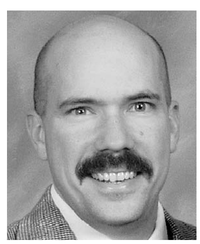
[22] O. Barndorff-Nielsen and D.R. Cox, "Edgeworth and Saddle-Point Approximations with Statistical Applications," *J. Royal Statistical Soc., Series B,* vol. 41, no. 3, pp. 279-312, 1979.

[23] S.M. Kay, A.H. Nuttall, and P.M. Baggenstoss, "Multidimensional Probability Density Function Approximations for Detection, Classification, and Model Order Selection," *IEEE Trans. Signal Processing,* vol. 49, no. 10, pp. 2240-2252, Oct. 2001.

[24] "University Research Initiative Synthetic Dataset (HRR)," Johns Hopkins Univ., Center for Imaging Science, http://cis.jhu.edu/data.sets/urisd, 2005.

[25] T.M. Cover and J.M. Van Campenhout, "On the Possible Orderings in the Measurement Selection Problem," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 7, no. 9, pp. 657-661, Sept. 1977.

[26] A. Jain and D. Zongker, "Feature Selection: Evaluation, Application, and Small Sample Performance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 2, pp. 153-158, Feb. 1997.



Michael D. DeVore received BS degrees in electrical engineering, computer engineering, and mathematics from the University of Missouri at Columbia in 1991 and the MS degree in electrical engineering in 1993. From 1993 through 1998, he was responsible for system integrity and integration and for operations support at Amdocs, Inc., a solutions provider to the telecommunications industry. In 2001, he received the DSc degree in electrical engineering from Washington University in St. Louis, Missouri, with dissertation research in automatic target recognition from synthetic aperture radar data. Dr. DeVore is currently an assistant professor in the Systems and Information Engineering Department at the University of Virginia. His research interests include target recognition, information systems for computer vision, shape-based object recognition, statistical model assessment, model representation and storage, and performance of recognition systems under dynamic resource constraints. Dr. DeVore is a member of IEEE, the IEEE Computer Society, SPIE, and the Optical Society of America.