Multidimensional Probability Density Function Approximations for Detection, Classification, and Model Order Selection

Steven M. Kay, *Fellow, IEEE*, Albert H. Nuttall, and Paul M. Baggenstoss, *Member, IEEE*

***Abstract*—This paper addresses the problem of calculating the multidimensional probability density functions (PDFs) of statistics derived from known many-to-one transformations of independent random variables (RVs) with known distributions. The statistics covered in the paper include reflection coefficients, autocorrelation estimates, cepstral coefficients, and general linear functions of independent RVs. Through PDF transformation, these results can be used for general PDF approximation, detection, classification, and model order selection. A model order selection example that shows significantly better performance than the Akaike and MDL method is included.**

***Index Terms*—Classification, class-specific features, PDF estimation, sufficient statistics.**

I. Introduction

In THIS paper, we present approximations of multidimensional probability density functions (PDFs) for statistics derived from the standard normal distribution. Let $\mathbf{z} = T(\mathbf{x})$, where $\mathbf{x}$ is a vector of independent and identically distributed (iid) samples of zero-mean Gaussian noise of unit variance. The *feature extraction* function $T(\ )$ can be any useful set of statistics. The challenge is to accurately evaluate the joint multidimensional PDF of $\mathbf{z}$. The results must be valid everywhere, including the tails of the PDF. We show that the results can be used to approximate $p(\mathbf{x}|H_1)$ for an arbitrary alternative hypothesis H_1. This approach has applications in detection, classification, and model order selection.

A. Motivation and Previous Work

The distribution of statistics derived from purely white Gaussian noise (WGN) have been studied in the past; however, applications have been limited because WGN is rarely encountered in practice. An important application of the WGN condition is as the null hypothesis in testing for colored noise. Tests for colored noise based on the periodogram [1] and serial autocorrelation function [2]–[4] have been studied.

Manuscript received September 26, 2000; revised June 14, 2001. This work was supported by the Office of Naval Research. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jian Li.

S. M. Kay is with the Department of Electrical and Computer Engineering, University of Rhode Island, Kingston, RI 02881 USA (e-mail: kay@ele.uri.edu).

A. H. Nuttall and P. M. Baggenstoss are with the Naval Undersea Warfare Center, Newport, RI 02841 USA (e-mail: p.m.baggenstoss@ieee.org).

Publisher Item Identifier S 1053-587X(01)07772-8.

Since the introduction of methods related to the class-specific method [5]–[7], the WGN hypothesis has become increasingly useful. This is because WGN is seen not as a hypothesis to be explicitly tested but, rather, as a reference hypothesis for converting likelihood tests into likelihood ratio tests in a way similar to the "dummy" hypothesis of Van Trees [8] and then taking advantage of sufficient statistics on a class-by-class basis. In particular, testing hypothesis H_1 against H_2 can be accomplished by comparing the likelihood ratios (LRs) according to

$$\frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} \underset{>}{\lessgtr} \frac{p(\mathbf{x}|H_2)}{p(\mathbf{x}|H_0)} \tag{1}$$

where $p(\ )$ denotes a PDF, and H_0 is any reference hypothesis, such as the WGN case. By finding class-specific sufficient statistics $\mathbf{z}_1 = T_1(\mathbf{x})$ and $\mathbf{z}_2 = T_2(\mathbf{x})$, (1) can be reduced to the LR comparison

$$\frac{p(\mathbf{z}_1|H_1)}{p(\mathbf{z}_1|H_0)} \underset{>}{\lessgtr} \frac{p(\mathbf{z}_2|H_2)}{p(\mathbf{z}_2|H_0)} \tag{2}$$

where $\mathbf{z}_1$ must be sufficient for H_1 versus H_0, and $\mathbf{z}_2$ must be sufficient for H_2 versus H_0. Note that only the low-dimensional numerator PDFs need to be approximated from training data. Clearly, the denominator PDFs $p(\mathbf{z}_1|H_0)$ and $p(\mathbf{z}_2|H_0)$ must be evaluated, which is the topic of this paper. The extension to M hypotheses is obvious.

In a later development, a theorem that extended the class-specific approach to the case when sufficiency of the statistics could not be guaranteed was introduced [9], [10]. This latter theorem allows the PDF of a set of statistics to be converted into a PDF of the input data. More precisely, let $\mathbf{z}_1 = T_1(\mathbf{x})$ be any multidimensional set of statistics derived from the raw data $\mathbf{x}$. Let $\hat{p}(\mathbf{z}_1|H_1)$ be an approximation to the PDF of $\mathbf{z}_1$ under hypothesis H_1. Then, the PDF of $\mathbf{x}$ under H_1 can be approximated by

$$\hat{p}(\mathbf{x}|H_1) = \left[\frac{p(\mathbf{x}|H_{0,1})}{p(\mathbf{z}_1|H_{0,1})}\right]\hat{p}(\mathbf{z}_1|H_1) \qquad \text{at } \mathbf{z}_1 = T_1(\mathbf{x}) \tag{3}$$

where $H_{0,1}$ is a fixed reference hypothesis chosen specially for H_1. According to Theorems 1 and 2 of [9], (3) is *always* a PDF; thus, it integrates to 1 over $\mathbf{x}$ for any reference hypothesis $H_{0,1}$ and any transformation $T_1(\ )$, provided $p(\mathbf{z}_1|H_0)$ meets a mild positivity requirement [9]. More precisely, we must have $p(\mathbf{z}_1|H_{0,1}) > 0$ whenever $p(\mathbf{z}_1|H_1) > 0$. While no additional requirements are needed for $\hat{p}(\mathbf{x}|H_1)$ to be a PDF, the pair $(H_{0,1}, T_1(\ ))$ should be chosen carefully so that $\hat{p}(\mathbf{x}|H_1)$ is a

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
OCT 2001		**00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multidimensional Probability Density Function Approximations for Detection, Classification, and Model Order Selection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Undersea Warfare Center,Newport,RI,02841**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**13**	

good approximation to $p(\mathbf{x}|H_1)$. In particular, if $\mathbf{z}_1 = T_1(\mathbf{x})$ is approximately sufficient for distinguishing H_1 from $H_{0,1}$, and $\hat{p}(\mathbf{z}_1|H_1) \to p(\mathbf{z}_1|H_1)$, then $\hat{p}(\mathbf{x}|H_1) \to p(\mathbf{x}|H_1)$.

Approximate sufficiency can be formally defined by the relationship

$$\frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_{0,1})} \simeq \frac{p(\mathbf{z}_1|H_1)}{p(\mathbf{z}_1|H_{0,1})}$$

although in practice, approximate sufficient statistics are obtained not always by mathematical analysis but, often, by experience and intuition. If approximate sufficient statistics $\mathbf{z}_1$ can be found, $p(\mathbf{x}|H_1)$ can be approximated simply by choosing a suitable reference hypothesis $H_{0,1}$, then approximating $p(\mathbf{z}_1|H_1)$, and finally converting this PDF into a PDF of $\mathbf{x}$ using (3). This represents a new general method for PDF approximation and statistical hypothesis testing. Using (3), a classifier may then be constructed using class-specific features

$$\left[\frac{p(\mathbf{x}|H_{0,1})}{p(\mathbf{z}_1|H_{0,1})}\right]\hat{p}(\mathbf{z}_1|H_1) \underset{>}{\overset{<}{}} \left[\frac{p(\mathbf{x}|H_{0,2})}{p(\mathbf{z}_2|H_{0,2})}\right]\hat{p}(\mathbf{z}_2|H_2) \quad (4)$$

where $\mathbf{z}_1 = T_1(\mathbf{x})$, and $\mathbf{z}_2 = T_2(\mathbf{x})$. The extension to M classes is obvious. This classifier requires PDF estimates with dimension only as high as the largest set of statistics and not as high as the total number of all statistics, as in a classical classifier.

In this paper, we limit ourselves to a particular choice of reference hypothesis, namely, the WGN case denoted by H_0. Choosing the WGN hypothesis for H_0 has many advantages. First, an analytic solution for $p(\mathbf{z}|H_0)$ is often tractable. Second, the condition of positivity is always met. Third, the sufficiency requirement against WGN is itself often a sensible requirement. Finally, the solutions for the WGN case provided herein can often be easily modified for arbitrary Gaussian-based distributions. When a common reference hypothesis is used, the classifier simplifies to

$$\left[\frac{p(\mathbf{x}|H_0)}{p(\mathbf{z}_1|H_0)}\right]\hat{p}(\mathbf{z}_1|H_1) \underset{>}{\overset{<}{}} \left[\frac{p(\mathbf{x}|H_0)}{p(\mathbf{z}_2|H_0)}\right]\hat{p}(\mathbf{z}_2|H_2). \quad (5)$$

This classifier is identical to (2), but it has a different interpretation.

B. Need for Accurate PDF Approximations in the Tails

Because H_0 is a fixed reference hypothesis determined prior to the measurement of data, it is possible that the actual data lies on the distant tails of the PDF $p(\mathbf{x}|H_0)$. This is also true at the output of the feature transformations. Thus, it is possible that both $p(\mathbf{z}_1|H_0)$ and $p(\mathbf{z}_2|H_0)$ approach zero simultaneously. Therefore, accurate tail approximations of $p(\mathbf{z}_j|H_0)$ are needed for meaningful results.

Commonly-used approximation methods such as the central limit theorem (CLT) do not provide accurate answers in the tails. In this paper, we apply the multidimensional saddlepoint approximation (SPA) [2], [11], which can provide accurate PDF tail estimates.

C. Notation

In the remainder of the paper, the raw input data $\mathbf{x} = [x_1 \cdots x_N]'$ is a set of independent real random variables (RVs) that are all Gaussian of zero mean and unit variance. The feature set whose distribution we seek is denoted $\mathbf{z} = T(\mathbf{x}) = [z_1 \cdots z_M]'$, where $M < N$ generally.

II. Some Exact Solutions

For some transformations $\mathbf{z} = T(\mathbf{x})$, the exact joint PDF of $\mathbf{z}$ can be derived. Some of these transformations can be seen as special cases of more general problems for which we have derived approximations. Therefore, they can serve as important test cases for the more general results (especially in the tails).

A. Order Statistics

Order statistics are important features in classification. Examples of order statistics are the three largest FFT bin magnitudes or the median of N sample values. The joint distribution of a collection of order statistics is easily found for i.i.d. RVs. [12]. Consider N iid samples derived from $\mathbf{x}$ using a transformation $w_n = T(x_n)$. Define $\mathbf{w} \triangleq [w_1 \cdots w_N]'$. Let $\mathbf{y} \triangleq [y_1 \cdots y_N]'$ be obtained from $\mathbf{w}$ by sorting into ascending order. Let $\mathbf{z} \triangleq [z_1 \cdots z_M]'$ be a subset of $\mathbf{y}$. Specifically, $z_1 = y_{n_1}$, $z_2 = y_{n_2}$, $\ldots$, $z_M = y_{n_M}$, where $1 \le n_1 < n_2 < \cdots < n_M \le N$.

Let $P_w(w)$ and $p_w(w)$ be the cumulative distribution function (CDF) and PDF of w_n, respectively. Then, for $M = 2$, for example, the joint PDF is

$$\begin{aligned} p_z(z_1, z_2) &= \frac{N!}{(n_1-1)!(n_2-n_1-1)!(N-n_2)!}[P_w(z_1)]^{(n_1-1)} \\ &\quad \cdot p_w(z_1)[P_w(z_2)-P_w(z_1)]^{(n_2-n_1-1)}p_w(z_2) \\ &\quad \cdot [1-P_w(z_2)]^{(N-n_2)}. \end{aligned}$$

Extending this to arbitrary order M, consider the n_1th, n_2th, $\cdots$, n_Mth-order statistics, where

$$1 \le n_1 < n_2 < \cdots < n_M \le N$$

follows the joint PDF

$$\begin{aligned} p_z(z_1, z_2 \cdots z_M) &= \frac{N!}{(n_1-1)!(n_2-n_1-1)!(n_3-n_2-1)! \cdots (N-n_M)!} \\ &\quad \cdot [P_w(z_1)]^{(n_1-1)}p_w(z_1)[P_w(z_2)-P_w(z_1)]^{(n_2-n_1-1)} \\ &\quad \cdot p_w(z_2)[P_w(z_3)-P_w(z_2)]^{(n_3-n_2-1)}p_w(z_3) \cdots \\ &\quad \cdot [1-P_w(z_M)]^{(N-n_M)}p_w(z_M). \end{aligned} \quad (6)$$

We consider two cases.

1) When $\{x_n\}$ are real Gaussian $N(0, 1)$ RVs and $w_n = |x_n|$, then $\{w_n\}$ are Chi(1) (chi-distributed with 1 degree of freedom), and we have $p_w(w) = (2/\sqrt{2\pi})e^{-w^2/2}$ and $P_w(w) = \text{erf}(w/\sqrt{2})$ for $w > 0$.
2) When $\{x_n\}$ are complex Gaussian $CN(0, 2)$ RVs and $w_n = |x_n|^2$, then $\{w_n\}$ are Chi-squared(2) (chi-squared with 2 degrees of freedom), and we have $p_w(w) = (1/2)e^{-w/2}$ and $P_w(w) = 1 - e^{-w/2}$ for $w > 0$.

Using these forms for $P_w(w)$ and $p_w(w)$ and (6), we can evaluate the desired PDFs for $\mathbf{z}$. Calculation involves some delicate

numerical problems, especially in the tails; however, these can be successfully dealt with, often by resorting to exceedance distributions instead of cumulative distributions.

B. Autocorrelation Function and Reflection Coefficients

A widely used model for signal processing applications is the autoregressive (AR) filter driven by white Gaussian noise [13]. The infinite length autocorrelation function (ACF) completely describes such processes. However, practical signals have an autocorrelation function that either decays to zero or is periodic. Thus, a finite number of autocorrelation samples often provide adequate information to characterize the process. In short, a set of autocorrelation samples can be approximately *sufficient* for testing statistical hypotheses concerning the process.

There are many ways to compute an autocorrelation estimate [13]. These methods are asympotically equivalent for large N but differ significantly for small N. We will concern ourselves with one particular ACF estimator because an exact formula can be found for its PDF, which is due to Watson [14]. In particular, we use the normalized *circular* autocorrelation samples $\tilde{r}_k$. Let

$$\tilde{r}_k = \frac{r_k}{r_0}, \qquad k = 1 \cdots P \tag{7}$$

where

$$r_k = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(x_{i-k} - \overline{x})$$

and

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and we define $x_{N+i} = x_i$. Assume that N is odd so that $n = (N-1)/2$ is an integer. Then, the exact joint PDF of $\{\tilde{r}_1, \tilde{r}_2 \cdots \tilde{r}_P\}$ is

$$p(\tilde{r}_1, \tilde{r}_2 \cdots \tilde{r}_P) = \frac{\Gamma(n)}{\Gamma(n-P)} \underbrace{\sum\sum\cdots\sum}_{(j_1, j_2 \cdots j_P) \in S_P(\tilde{r}_1, \tilde{r}_2 \cdots \tilde{r}_P)} \cdot q(j_1, j_2 \cdots j_P) \tag{8}$$

where

$$q(j_1, j_2 \cdots j_P) = \frac{\operatorname{sgn}[\det(\mathbf{B}(j_1, j_2 \cdots j_P))][\det(\mathbf{A}(j_1, j_2 \cdots j_P))]^{n-P-1}}{\det(\mathbf{C}(j_1, j_2 \cdots j_P))}$$

and $\operatorname{sgn}(x) = 1$ for $x > 0$, $= -1$ for $x < 0$, and $= 0$ for $x = 0$, and the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are of dimension $(P+1)\times(P+1)$, $P \times P$, and $(P+1)\times(P+1)$, respectively. They are defined as

$$\mathbf{A}(j_1, j_2 \cdots j_P) = \begin{bmatrix} 1 & \tilde{r}_1 & \tilde{r}_2 & \cdots & \tilde{r}_P \\ 1 & \gamma_{j_1}^{(1)} & \gamma_{j_1}^{(2)} & \cdots & \gamma_{j_1}^{(P)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \gamma_{j_P}^{(1)} & \gamma_{j_P}^{(2)} & \cdots & \gamma_{j_P}^{(P)} \end{bmatrix}$$

$$\mathbf{B}(j_1, j_2 \cdots j_P) = \begin{bmatrix} \gamma_{j_1}^{(1)} & \gamma_{j_1}^{(2)} & \cdots & \gamma_{j_1}^{(P)} \\ \gamma_{j_2}^{(1)} & \gamma_{j_2}^{(2)} & \cdots & \gamma_{j_2}^{(P)} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{j_P}^{(1)} & \gamma_{j_P}^{(2)} & \cdots & \gamma_{j_P}^{(P)} \end{bmatrix}$$

$$\mathbf{C}(j_1, j_2 \cdots j_P) = \prod_{j \neq j_1, j_2 \cdots j_P} \begin{bmatrix} 1 & \gamma_j^{(1)} & \gamma_j^{(2)} & \cdots & \gamma_j^{(P)} \\ 1 & \gamma_{j_1}^{(1)} & \gamma_{j_1}^{(2)} & \cdots & \gamma_{j_1}^{(P)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \gamma_{j_P}^{(1)} & \gamma_{j_P}^{(2)} & \cdots & \gamma_{j_P}^{(P)} \end{bmatrix}$$

where

$$\gamma_j^{(k)} = \cos\left(\frac{2\pi}{N} jk\right) \qquad j = 1, 2 \cdots n;\ k = 1, 2 \cdots P.$$

The region $S_P(\tilde{r}_1, \tilde{r}_2 \cdots \tilde{r}_P)$ is the convex polyhedron in R^P formed from the vectors

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \gamma_{j_1}^{(1)} \\ \gamma_{j_1}^{(2)} \\ \vdots \\ \gamma_{j_1}^{(P)} \end{bmatrix} \cdots \begin{bmatrix} \gamma_{j_P}^{(1)} \\ \gamma_{j_P}^{(2)} \\ \vdots \\ \gamma_{j_P}^{(P)} \end{bmatrix}$$

or the convex polyhedron of the vectors $\{\mathbf{x}_1, \mathbf{x}_2 \cdots \mathbf{x}_{P+1}\}$ is the convex set formed by the linear combination

$$\sum_{i=1}^{P+1} \alpha_i \mathbf{x}_i$$

where $\alpha_i \geq 0$, and $\sum_{i=1}^{P+1} \alpha_i \leq 1$. A MATLAB implementation of (8) is provided in Appendix A.

1) Experimental Verification: To validate the analysis, it is useful to compare the derived PDF with experimental values. The above analysis can be verified experimentally by comparing a scatter plot of the first two normalized circular ACF estimates with an intensity image of the PDF obtained from the program. Fig. 1 shows such a comparison for $N = 9$ samples. The figure illustrates that for small N, the range of possible ACF values occupy regions with linear boundaries in the plane. The shape of the PDF itself at $N = 9$ is a tetrahedron.

2) Reflection Coefficients: Due to the one-to-one transformation linking the normalized autocorrelation coefficients with the reflection coefficients [13], it is possible to utilize the above results to find the exact distribution of reflection coefficients; however, this applies only when the *circular* autocorrelation coefficients (7) are used. Consider the transformation

$$[K_1 \cdots K_P]' = T_1(\tilde{r}_1 \cdots \tilde{r}_P)$$

where $\{K_i\}$ are the reflection coefficients. The joint PDF $p(K_1 \cdots K_P)$ requires knowing the Jacobian of the transformation $T_1(\ )$, which is

$$\log J_1 = \sum_{i=1}^{P-1} (P-i)\log(1-K_i^2).$$

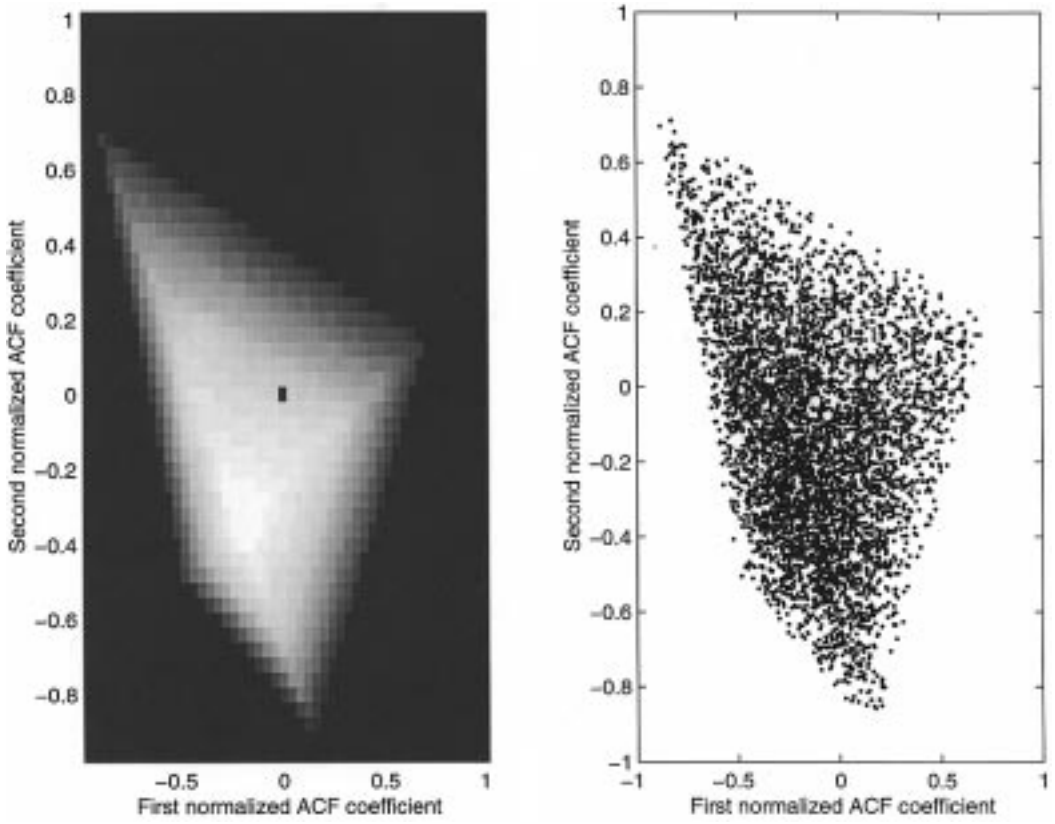

Fig. 1. (Left) Comparison of (8) with (right) a scatter plot of experimental data points. Autocorrelation estimates were computed from $N = 9$ samples. Notice the "hole " at (0,0), where the exact PDF cannot be computed.

Thus

$$\log p(K_1 \cdots K_P) = \log p(\tilde{r}_1 \cdots \tilde{r}_P) + \log J_1.$$

3) Log-Transformed Reflection Coefficients: Because the reflection coefficients are subject to the constraint $|K_i| < 1$, the joint PDF of $p(K_1 \cdots K_P)$ is discontinuous and difficult to approximate using standard techniques such as Gaussian mixtures [15]. A better-behaved set of features is obtained by the one-to-one mapping

$$K_k' = \log\left(\frac{1 - K_k}{1 + K_k}\right).$$

The log of the Jacobian of this transformation is

$$\log J_2 = -\sum_{k=1}^{P} \log\left(\frac{2}{1 - K_k^2}\right).$$

Thus

$$\log p(K_1' \cdots K_P') = \log p(K_1 \cdots K_P) + \log J_2.$$

III. Saddlepoint Approximation (Theory)

The class of statistics for which the exact PDF is known is relatively limited. For a broader class of statistics, the exact moment generating function (MGF) is often known. The problem is that inversion of the MGF transformation to find the exact PDF may not be possible in closed form. For these cases, we can use an approximation that provides accurate tail PDF estimates.

Let $\mathbf{z} = [z_1 \cdots z_M]'$ be an M-dimensional real random vector with joint MGF $g_z(\boldsymbol{\lambda}) = E\{\exp(\boldsymbol{\lambda}'\mathbf{z})\}$, where vector $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_M]'$. Then, the joint PDF of $\mathbf{z}$ at the M-dimensional point $\mathbf{u} = [u_1 \cdots u_M]'$ is given by the Mth order contour integral

$$p_z(\mathbf{u}) = \frac{1}{(i2\pi)^M} \int_C \exp(-\boldsymbol{\lambda}'\mathbf{u}) g_z(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \tag{9}$$

where $i = \sqrt{-1}$, and the M-dimensional contour C is parallel to the imaginary axis in each of the M dimensions of $\boldsymbol{\lambda}$. The joint cumulant generating function (CGF) of $\mathbf{z}$ is $c_z(\boldsymbol{\lambda}) = \log g_z(\boldsymbol{\lambda})$. The most useful M-dimensional saddlepoint (SP) of the integrand of (9) is that real point $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\mathbf{u})$ in M-dimensional space where all M partial derivatives satisfy

$$\left.\frac{\partial c_z(\boldsymbol{\lambda})}{\partial \lambda_m}\right|_{\hat{\boldsymbol{\lambda}}} = u_m \qquad \text{for } 1 \le m \le M. \tag{10}$$

When the contour C is moved in M dimensions to go through the real SP $\hat{\boldsymbol{\lambda}}$ and the change of variable

$$\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}} + i\mathbf{t}, \qquad \mathbf{t} = [t_1 \cdots t_M]' \tag{11}$$

is made, (9) becomes

$$p_z(\mathbf{u}) = (2\pi)^{-M} \exp(-\hat{\boldsymbol{\lambda}}'\mathbf{u}) \int \exp(-i\mathbf{t}'\mathbf{u}) g_z(\hat{\boldsymbol{\lambda}} + i\mathbf{t})\, d\mathbf{t} \tag{12}$$

where the new M-dimensional contour passes through the SP at $\mathbf{t} = \mathbf{0}$.

The logarithm of the integrand of (12) can be expanded in a power series about the origin $t = 0$ according to

$$\begin{aligned}&\log\{\exp(-i\mathbf{t}'\mathbf{u})g_z(\hat{\boldsymbol{\lambda}}+i\mathbf{t})\}\\&\quad=-i\mathbf{t}'\mathbf{u}+c_z(\hat{\boldsymbol{\lambda}}+i\mathbf{t})\\&\quad\simeq-i\mathbf{t}'\mathbf{u}+c_z(\hat{\boldsymbol{\lambda}})+i\mathbf{t}'\nabla c_z(\hat{\boldsymbol{\lambda}})-\tfrac{1}{2}\,\mathbf{t}'C_z(\hat{\boldsymbol{\lambda}})\mathbf{t}\end{aligned}\quad(13)$$

where the $M \times M$ matrix

$$C_z(\boldsymbol{\lambda}) \triangleq \left[\frac{\partial^2 c_z(\boldsymbol{\lambda})}{\partial\lambda_m\partial\lambda_{m_1}}\right] \quad (14)$$

is symmetric in m and m_1 for all $\boldsymbol{\lambda}$. Thus, using (10) and (13), the integrand of (12) can be approximated as

$$\exp(-i\mathbf{t}'\mathbf{u})g_z(\hat{\boldsymbol{\lambda}}+i\mathbf{t}) \simeq \exp\left[c_z(\hat{\boldsymbol{\lambda}})-\tfrac{1}{2}\,\mathbf{t}'C_z(\hat{\boldsymbol{\lambda}})\mathbf{t}\right] \quad (15)$$

for small $|\mathbf{t}|$. If this approximation is now extrapolated to *all* $\mathbf{t}$ and substituted in (12), there follows the usual saddlepoint (or tilted Edgeworth) approximation (SPA) in M dimensions [11]

$$\begin{aligned}p_z(\mathbf{u}) &\simeq \frac{\exp\left[c_z(\hat{\boldsymbol{\lambda}})-\hat{\boldsymbol{\lambda}}'\mathbf{u}\right]}{(2\pi)^M}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\mathbf{t}'C_z(\hat{\boldsymbol{\lambda}})\mathbf{t}\right]d\mathbf{t}\\&=\frac{\exp\left[c_z(\hat{\boldsymbol{\lambda}})-\hat{\boldsymbol{\lambda}}'\mathbf{u}\right]}{(2\pi)^{M/2}\left[\det(C_z(\hat{\boldsymbol{\lambda}}))\right]^{1/2}};\qquad \hat{\boldsymbol{\lambda}}=\hat{\boldsymbol{\lambda}}(\mathbf{u}).\end{aligned}\quad(16)$$

A. Finding the Saddlepoint $\hat{\boldsymbol{\lambda}}(\mathbf{u})$

To obtain the SPA (16), the real SP $\hat{\boldsymbol{\lambda}}(\mathbf{u})$ satisfying (10) must be found. The SP may be found using the Newton–Raphson iteration

$$\boldsymbol{\lambda}_{n+1} = \boldsymbol{\lambda}_n + \kappa C_z^{-1}(\boldsymbol{\lambda}_n)(\mathbf{u} - c_z'(\boldsymbol{\lambda}_n)) \quad (17)$$

where $0 < \kappa \leq 1$ is a step-size parameter. The zero vector $\boldsymbol{\lambda} = [0, 0 \cdots 0]'$ can always serve as a starting point because it is always in the ROC. At each iteration, the new value of $\boldsymbol{\lambda}_n$ must be tested to see if it is still in the ROC and modified if necessary.

If the search for the SP is confined to the real axes in multidimensional (MD) lambda space, it can be shown that the Hessian matrix of the joint cumulative generating function (CGF) is positive definite for all real lambda inside the MD region of definition of the joint MGF. This means that the MD integrand has a bowl-like behavior with a single minimum in this real MD space. Thus, the Newton–Raphson search procedure will always find the single minimum if conducted in an "appropriately slow" fashion and if the search is constantly confined to the region of definition.

B. Accuracy of the SPA

The accuracy of PDF approximations can be experimentally determined in the tails if an exact formula is available for some special case. We use this approach whenever possible in what follows.

Because the SPA is based on a series expansion of the MGF at the SP, its accuracy depends on the shape of the MGF at this point. Experience has shown that even in the tails, the errors tends to be in the "mantissa" rather than in the "exponent."

There is no reason why additional terms in the expansion cannot be obtained. In fact, additional terms of the expansion have been derived for some important cases including linear functions of independent RVs and are available in an NUWC technical report [16]. These terms can be used not only to provide additional accuracy but as an indication of the validity of the first-order SPA as well. Issues of SPA accuracy are treated in more detail in the technical report.

C. Linear Functions of Independent RVs

In many signal processing applications, linear transformations are made on a set of non-Gaussian but independent RVs. Examples include Fourier analysis of the squared magnitudes of a set of real or complex time samples, autocorrelation estimates (by the FFT method), and cepstrum estimates. These problems can be posed in the form $\mathbf{z} = A\mathbf{y}$, where $y_n = T_n(x_n)$, and A is an $M \times N$ matrix. Note that RVs $\{y_n\}$ are independent but not necessarily identically distributed. The output vector $\mathbf{z}$ is of length M, where $M < N$; thus, the transformation is not one-to-one. Let the MGFs and CGFs of RVs $\{y_n\}$ be denoted $\{g_n(v)\}$ and $\{c_n(v)\}$, respectively. That is

$$\begin{aligned}g_n(v) = \mathrm{E}\{\exp(vy_n)\},\qquad c_n(v) = \log g_n(v)\\ \text{for } 1 \leq n \leq N\end{aligned}\quad(18)$$

where $\mathrm{E}\{\ \}$ denotes an ensemble average. The weighted sum of independent RVs of interest is given by

$$z_m = \sum_{n=1}^{N} a_{mn}y_n \qquad \text{for } 1 \leq m \leq M \quad (19)$$

where $M \leq N$, and the $M \times N$ real matrix $[a_{mn}]$ is arbitrary, except that it must have rank M.

The joint MGF of RVs $\{z_m\}$ is, upon use of (18) and (19) and the independence of RVs $\{y_n\}$

$$\begin{aligned}g_z(\lambda_1 \cdots \lambda_M) &\triangleq \mathrm{E}\left\{\exp\left(\sum_{m=1}^{M}\lambda_m z_m\right)\right\}\\&=\mathrm{E}\left\{\exp\left(\sum_{m=1}^{M}\lambda_m\sum_{n=1}^{N}a_{mn}y_n\right)\right\}\\&=\prod_{n=1}^{N}\mathrm{E}\left\{\exp\left(y_n\sum_{m=1}^{M}a_{mn}\lambda_m\right)\right\}\\&=\prod_{n=1}^{N}g_n(b_n(\boldsymbol{\lambda}))\end{aligned}\quad(20)$$

where $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_M]'$, and

$$b_n(\boldsymbol{\lambda}) \triangleq \sum_{m=1}^{M} a_{mn}\lambda_m \qquad \text{for } 1 \leq n \leq N. \quad (21)$$

That is

$$g_z(\boldsymbol{\lambda}) = \prod_{n=1}^{N} g_n(b_n(\boldsymbol{\lambda})). \quad (22)$$

The joint CGF of RVs $\{z_m\}$ is, using (18)

$$c_z(\boldsymbol{\lambda}) = \log g_z(\boldsymbol{\lambda}) = \sum_{n=1}^{N} c_n(b_n(\boldsymbol{\lambda})). \tag{23}$$

There follows, by reference to (21)

$$\frac{\partial c_z(\boldsymbol{\lambda})}{\partial \lambda_m} = \sum_{n=1}^{N} a_{mn} c_n'(b_n(\boldsymbol{\lambda})) \qquad \text{for } 1 \le m \le M. \tag{24}$$

Finally, the M simultaneous equations that must be solved for the M-dimensional saddlepoint $\hat{\boldsymbol{\lambda}}$ are

$$\sum_{n=1}^{N} a_{mn} c_n'(b_n(\hat{\boldsymbol{\lambda}})) = z_m \qquad \text{for } 1 \le m \le M \tag{25}$$

where $\mathbf{z} = [z_1 \cdots z_M]'$ are the particular values at which the joint PDF of RVs $\{z_m\}$ in (19) is to be evaluated. The second-order partial derivatives required to obtain the SPA follow from (21) and (24) as

$$\frac{\partial^2 c_z(\boldsymbol{\lambda})}{\partial \lambda_m \partial \lambda_{m^*}} = \sum_{n=1}^{N} a_{mn} a_{m^* n} c_n''(b_n(\boldsymbol{\lambda})) \quad \text{for } 1 \le m, m^* \le M. \tag{26}$$

Once saddlepoint $\hat{\boldsymbol{\lambda}}$ is found from (25), it can be substituted in (26), and the $M \times M$ Hessian matrix can be evaluated at the saddlepoint, namely, $C_z(\hat{\boldsymbol{\lambda}})$.

IV. Applications of the Saddlepoint Approximation

A. Linear Sums of Magnitude-Squared Gaussian Random Variables

An important set of statistics are weighted sums of Chi-squared RVs. Specifically, statistics of the form

$$z_m = \sum_{n=1}^{N} a_{mn} y_n, \qquad \text{for } 1 \le m \le M \tag{27}$$

where $y_n = |x_n|^2$ and where x_n is a real or complex Gaussian RV, are frequently encountered in signal processing. Included are least-squares polynomial approximations of magnitude-squared time series, Fourier analysis of squared time series or FFT output bins, and two-dimensional (2-D) Fourier analysis of images or spectrograms. Such statistics are widely used in feature-based classification problems. We consider both the case when RVs $\{x_n\}$ are real and when $\{x_n\}$ are complex.

1) Complex RVs: Assume that $\{x_n\}$ are complex zero-mean Gaussian CN(0,1) RVs and $y_n = |x_n|^2$. Then, y_n are exponentially distributed $p_y(y) = e^{-y}$. Alternatively, we may say that $u = 2y$ has the Chi-squared distribution with 2 degrees of freedom, which are denoted $p_u(u) = \chi^2(u, 2)$. Thus, $p_y(y) = 2p_u(2y)$.

Since MGF $g_n(\lambda)$ is not a function of n, we write $g_n(\lambda) = g_y(\lambda) = 1/(1-\lambda)$, where $\lambda < 1$. We have $c_y(\lambda) = \log g_y(\lambda) = -\log(1-\lambda)$ for $\lambda < 1$. Substituting this into (23)

$$c_z(\boldsymbol{\lambda}) = \log g_z(\boldsymbol{\lambda}) = -\sum_{n=1}^{N} \log\left(1 - \sum_{m=1}^{M} \lambda_m a_{mn}\right) \tag{28}$$

where the ROC is defined by

$$\sum_{m=1}^{M} \lambda_m a_{mn} < 1, \qquad \text{for } n = 1 \cdots N.$$

Taking derivatives

$$\frac{\partial c_z(\boldsymbol{\lambda})}{\partial \lambda_m} = \sum_{n=1}^{N} a_{mn} \left(1 - \sum_{l=1}^{M} \lambda_l a_{ln}\right)^{-1} \quad \text{for } m = 1 \cdots M.$$

The second derivatives are

$$\frac{\partial^2 c_z(\boldsymbol{\lambda})}{\partial \lambda_m \partial \lambda_p} = \sum_{n=1}^{N} \left(1 - \sum_{l=1}^{M} \lambda_l a_{ln}\right)^{-2} (a_{pn} a_{mn}) \quad \text{for } 1 \le m, p \le M \tag{29}$$

from which we can construct $M \times M$ Hessian matrix $C_z(\boldsymbol{\lambda})$. Finally, we use $C_z(\boldsymbol{\lambda})$, $c_z(\boldsymbol{\lambda})$ in (16).

2) Real RVs: When $\{x_n\}$ are real zero-mean Gaussian $N(0,1)$ RVs, the distribution of $y_n = |x_n|^2$ is χ^2 with 1 degree of freedom; thus, $p_y(y) = \chi^2(y, 1)$. Thus, $g_y(\lambda) = 1/\sqrt{1-2\lambda}$, where $\lambda < 0.5$. Similar to the complex case, we have $g_n(\lambda) = g_y(\lambda) = 1/\sqrt{1-2\lambda}$, where $\lambda < 0.5$. We have $c_y(\lambda) = \log g_y(\lambda) = -(1/2)\log(1-2\lambda)$ for $\lambda < 0.5$. Substituting this into (23)

$$c_z(\boldsymbol{\lambda}) = \log g_z(\boldsymbol{\lambda}) = -\sum_{n=1}^{N} \log\left(1 - 2\sum_{m=1}^{M} \lambda_m a_{mn}\right)$$

where the ROC is defined by

$$\sum_{m=1}^{M} \lambda_m a_{mn} < 0.5, \qquad \text{for } n = 1 \cdots N.$$

Taking derivatives

$$\frac{\partial c_z(\boldsymbol{\lambda})}{\partial \lambda_m} = 2\sum_{n=1}^{N} a_{mn} \left(1 - 2\sum_{l=1}^{M} \lambda_l a_{ln}\right)^{-1} \quad \text{for } 1 \le m \le M.$$

The second derivatives are

$$\frac{\partial^2 c_z(\boldsymbol{\lambda})}{\partial \lambda_m \partial \lambda_p} = \sum_{n=1}^{N} \left(1 - 2\sum_{l=1}^{M} \lambda_l a_{ln}\right)^{-2} (4a_{pn} a_{mn}) \quad \text{for } 1 \le m, p \le M$$

from which we can construct Hessian matrix $C_z(\boldsymbol{\lambda})$.

3) Experimental Validation: To validate the above results, especially in the tails, it is necessary to find a case for which the exact result is known. Notice that (27) may be written in the matrix form

$$\mathbf{z} = \mathbf{A}\mathbf{y}$$

where $\mathbf{A}$ is a general full-rank $M \times N$ matrix. A class of matrices $\mathbf{A}$ for which the exact joint PDF of $\mathbf{z}$ can be calculated is

described by Nuttall [17]. A special case of this more general class is the following. Let

$$\mathbf{u} = \mathbf{A}_1\mathbf{y}$$

where

$$\mathbf{A}_1 = \begin{bmatrix} 2 & 0 & 2 & 0 & \cdots & 2 & 0 \\ 0 & 2 & 0 & 2 & \cdots & 0 & 2 \end{bmatrix}.$$

Clearly then, u_1 and u_2 are independent. Furthermore, for the case of Section IV-A1, u_1 and u_2 are chi-square RVs with N degrees of freedom. Thus, it is straightforward to write down the PDF $p_u(\mathbf{u})$. The above can be generalized if we linearly transform $\mathbf{u}$ using a general full-rank two-by-two matrix $\mathbf{A}_2$. Let

$$\mathbf{z} = \mathbf{A}_2\mathbf{A}_1\mathbf{y} = \mathbf{A}_2\mathbf{u}.$$

Then, the joint PDF of $\mathbf{z}$ will equal

$$p_z(\mathbf{z}) = p_u(\mathbf{A}_2^{-1}\mathbf{z})|\det(\mathbf{A}_2)|^{-1}.$$

We tried this approach with $N = 128$ and

$$\mathbf{A}_2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

and $\{y_i\}$ derived from complex-valued data, as in Section IV-A1.

The experiment was designed to probe the tails of the PDF. This was accomplished by generating data using a variance differing from one (the assumed PDF). In the experiment, 1000 data samples of $\mathbf{z}$ were generated using $x_i \sim CN(0, \sigma^2)$, where σ^2 was chosen randomly according to $\sigma^2 = w^2$ and where $w \sim N(0, 100)$. The SPA error was determined by comparing with the exact expression. The results are shown in Fig. 2. The error was approximately $+0.0026$ for all samples except one, for which the error was -0.0026. It is unclear why this pattern occurred. Notice, however, that this very low approximation error occurred in the deep tails where $\log p_z(\mathbf{z})$ is as low as $-140\,000$. The accuracy that is obtained by the SPA method depends ultimately on the accuracy of the integral approximation (15) and the accuracy to which the saddlepoint itself is determined. These, in turn, depend on the matrix $\mathbf{A}$ as well as the univariate MGFs $g_y(y)$. Issues of the approximation accuracy have been studied by Nuttall [16]. In particular, additional terms of the power series expansion (13) can be obtained for the case of the linear function of *iid* RVs.

B. Cepstrum and Autocorrelation Estimates (Noncontiguous)

It is possible to represent the cepstrum and autocorrelation estimates as linear functions of a set of independent (but not identically distributed) RVs, allowing the results of Section III-C to be used. Recall that samples $\{x_t\}$, $0 \le t \le N-1$ are independent identically distributed (iid) real Gaussian RVs with zero mean and unit variance. The corresponding complex Fourier coefficients are defined as

$$X_k = \left(\frac{2}{N}\right)^{1/2} \sum_{t=0}^{N-1} x_t \exp(-i2\pi kt/N) \quad \text{for } 0 \le k \le N-1. \tag{30}$$

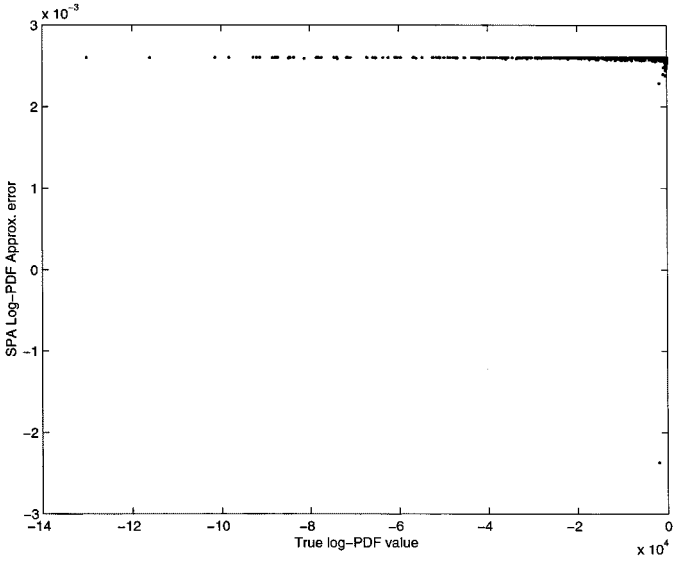

Fig. 2. PDF estimation error for SPA in the far tails using data generated with random variance. The vertical axis shows the difference between the log-PDF values of the SPA and the exact expression. The error was $+2.6$e-3 most of the time, except for one occurrence of -2.6e-3.

Since $\{x_t\}$ are real, it follows that $X_{N-k} = X_k^*$ for $1 \le k \le N-1$; however, X_0 and $X_{N/2}$ are always real.

Define the set of N real spectral quantities

$$\begin{aligned} &\text{Case I (Autocorrelation):} \quad Y_k = |X_k|^2 \\ &\text{Case II (Cepstrum):} \quad Y_k = \log|X_k|^2 \end{aligned} \tag{31}$$

for $0 \le k \le N-1$ and N even. Then, $Y_{N-k} = Y_k$ for $1 \le k \le N-1$. Finally, perform an inverse FFT back into the time domain to get the ACF or cepstrum estimates, respectively, according to

$$w_t = \sum_{k=0}^{N-1} Y_k \exp(i2\pi kt/N), \qquad \text{for } 0 \le t \le N-1. \tag{32}$$

For Case I, w_t are autocorrelation estimates (scaled by $2N$), and for Case II, they are cepstrum estimates (also with a special scaling).

Expression (32) can be written purely in terms of real variables

$$w_t = \sum_{k=0}^{N/2} \epsilon_k \cos(2\pi kt/N) Y_k, \qquad \text{for } 0 \le t \le N/2 \tag{33}$$

where

$$\epsilon_k \triangleq \left\{ \begin{array}{ll} 1, & \text{for } k = 0 \text{ and } N/2 \\ 2, & \text{for } 1 \le k \le N/2 - 1 \end{array} \right\}. \tag{34}$$

The remainder of RVs $\{w_t\}$ can be found, if desired, from the relation $w_{N-t} = w_t$ for $1 \le t \le N-1$.

Complex RVs $\{X_k\}$ in (30), for $0 \le k \le N/2$, are all independent of each other, due to the white Gaussian statistics of $\{x_t\}$ that were assumed. We let

$$X_k = U_k + jV_k \qquad \text{for } 0 \le k \le N/2. \tag{35}$$

Then, $V_0 = 0$, $V_{N/2} = 0$, and

$$\mathrm{E}(U_0^2) = 2, \quad \mathrm{E}(U_{N/2}^2) = 2, \quad \mathrm{E}(U_k^2) = \mathrm{E}(V_k^2) = 1 \tag{36}$$

for $1 \le k \le N/2 - 1$. All these $\{U_k\}$ and $\{V_k\}$ RVs in (35) are independent of each other and are Gaussian zero mean. We express (33) in the compact form

$$w_t = \sum_{k=0}^{N/2} a_{tk} Y_k, \qquad \text{for } 0 \le t \le N/2 \tag{37}$$

where

$$a_{tk} \triangleq \epsilon_k \cos(2\pi kt/N), \qquad \text{for } 0 \le t, k \le N/2. \tag{38}$$

In (37), all of the $\{Y_k\}$ RVs for $0 \le k \le N/2$ are independent of each other because all of the RVs $\{X_k\}$ for $0 \le k \le N/2$ are independent of each other.

1) Case I—Autocorrelation Estimates: From the information above, it is seen that RVs U_0^2 and $U_{N/2}^2$, if scaled by $1/2$, have a chi-squared distribution with 1 degree of freedom. The $\chi^2(1)$ PDF is $p(u) = (1/2\sqrt{2\pi})(u/2)^{-1/2}\exp(-u/4)$ for $u > 0$, with MGF $g(v) = 1/\sqrt{1-2v}$ for $v < 0.5$. It follows that without the scaling, the MGF is

$$g_0(v) = g(2v) = \frac{1}{\sqrt{1-4v}}, \qquad \text{for } v < 0.25. \tag{39}$$

In addition, RVs $(U_k^2 + V_k^2)$, for $1 \le k \le N/2 - 1$, have PDF $\exp(-u/2)/2$ for $u > 0$, with MGF

$$g_1(v) = \frac{1}{1-2v}, \qquad \text{for } v < 0.5. \tag{40}$$

In compact notation, define the MGFs of all the $\{Y_k\}$ in (37)

$$g_k^*(v) \triangleq \left\{ \begin{array}{ll} g_0(v), & \text{for} \quad k = 0 \text{ and } N/2 \\ g_1(v), & \text{for} \quad 1 \le k \le N/2 - 1 \end{array} \right\}. \tag{41}$$

Then, the corresponding CGFs for the $\{Y_k\}$ in (37) are

$$\begin{aligned} c_k^*(v) &= \log {g^*}_k(v) \\ &= \left\{ \begin{array}{ll} c_0(v) = \log g_0(v), & \text{for} \quad k = 0 \text{ and } N/2 \\ c_1(v) = \log g_1(v), & \text{for} \quad 1 \le k \le N/2 - 1 \end{array} \right\}. \end{aligned} \tag{42}$$

Now, we will consider a subset of all the N $\{w_t\}$ RVs originally defined in (32) and then manipulated into forms (33) and (37). In particular, consider only the set

$$w_t = \sum_{k=0}^{N/2} a_{tk} Y_k, \qquad \text{for } t = t_1 \cdots t_M,\ M \le N/2 + 1 \tag{43}$$

where $\{t_m\}$ are M arbitrary *distinct* time instants in the interval $[0, N/2]$. That is

$$w_{t_m} = \sum_{k=0}^{N/2} a_{t_m k} Y_k \qquad \text{for } 1 \le m \le M. \tag{44}$$

Write this expression as

$$z_m = \sum_{k=0}^{N/2} b_{mk} Y_k \qquad \text{for } 1 \le m \le M \tag{45}$$

where

$$\begin{aligned} z_m &= w_{t_m}, \qquad b_{mk} = a_{t_m k} \\ &\qquad \text{for } 1 \le m \le M,\ 0 \le k \le N/2. \end{aligned} \tag{46}$$

Let vector $\boldsymbol{\lambda} = [\lambda_1 \cdots \lambda_M]'$ and random vector $\mathbf{z} = [z_1 \cdots z_M]'$. The joint MGF of the M-dimensional vector $\mathbf{z}$ is

$$\begin{aligned} g_z(\boldsymbol{\lambda}) &= \mathrm{E}\left\{\exp\left(\boldsymbol{\lambda}'\mathbf{z}\right)\right\} \\ &= \mathrm{E}\left\{\exp\left(\sum_{m=1}^{M} \lambda_m z_m\right)\right\} \\ &= \mathrm{E}\left\{\exp\left(\sum_{m=1}^{M} \lambda_m \sum_{k=0}^{N/2} b_{mk} Y_k\right)\right\} \\ &= \prod_{k=0}^{N/2} \mathrm{E}\left\{\exp\left(Y_k \sum_{m=1}^{M} b_{mk}\lambda_m\right)\right\} \\ &= \prod_{k=0}^{N/2} g_k^*(d_k(\boldsymbol{\lambda})) \end{aligned} \tag{47}$$

where we used the independence of RVs $\{Y_k\}$ [see (39)–(41)] and defined

$$d_k(\boldsymbol{\lambda}) \triangleq \sum_{m=1}^{M} b_{mk}\lambda_m, \qquad \text{for } 0 \le k \le N/2. \tag{48}$$

The joint CGF of $\mathbf{z}$ follows, from (42) and (47), as

$$c_z(\boldsymbol{\lambda}) = \log g_z(\boldsymbol{\lambda}) = \sum_{k=0}^{N/2} c_k^*(d_k(\boldsymbol{\lambda})). \tag{49}$$

The partial derivatives of interest are

$$\frac{\partial}{\partial \lambda_m} c_z(\boldsymbol{\lambda}) = \sum_{k=0}^{N/2} b_{mk} c_k^{*\prime}(d_k(\boldsymbol{\lambda})), \qquad \text{for } 1 \le m \le M \tag{50}$$

where we used (48) to determine

$$\frac{\partial}{\partial \lambda_m} d_k(\boldsymbol{\lambda}) = b_{mk}, \qquad \text{for } 1 \le m \le M, 0 \le k \le N/2. \tag{51}$$

The M simultaneous equations that must be solved for saddlepoint $\hat{\boldsymbol{\lambda}}$ are, from (46) and (50)

$$\sum_{k=0}^{N/2} a_{t_m k} c_k^{*\prime}\left(d_k\left(\hat{\boldsymbol{\lambda}}\right)\right) = w_{t_m}, \quad \text{for } 1 \le m \le M,\ M \le N/2+1 \tag{52}$$

where $\{w_{t_m}\}$ are the particular M values of RVs $\{w_t\}$ and where the joint PDF of the ACF estimates is of interest. We also have, from (50) and (51)

$$\begin{aligned} \frac{\partial^2}{\partial \lambda_{m_1} \partial \lambda_{m_2}} c_z(\boldsymbol{\lambda}) &= \sum_{k=0}^{N/2} b_{m_1 k} b_{m_2 k} c_k^{*\prime\prime}(d_k(\boldsymbol{\lambda})) \\ &\qquad \text{for } 1 \le m_1, m_2 \le M. \end{aligned} \tag{53}$$

The MGFs for $\{Y_k\}$ were presented in (39) and (40). The functions required in the SPA are

$$\begin{aligned} g_0(v) &= 1/\sqrt{1-4v} \\ c_0(v) &= -0.5\log(1-4v) \\ c_0'(v) &= \frac{2}{1-4v} \\ c_0''(v) &= \frac{8}{(1-4v)^2} \end{aligned} \tag{54}$$

and

$$\begin{aligned} g_1(v) &= 1/(1-2v) \\ c_1(v) &= -\log(1-2v) \\ c_1'(v) &= \frac{2}{1-2v} \\ c_1''(v) &= \frac{4}{(1-2v)^2}. \end{aligned} \tag{55}$$

2) Case II: Cepstrum Estimates: From the information above, it is seen that RVs U_0^2 and $U_{N/2}^2$ have PDF $\exp(-u/4)/(4\pi u)^{1/2}$ for $u > 0$. In additon, RVs $(U_k^2 + V_k^2)$, for $1 \le k \le N/2-1$, have PDF $\exp(-u/2)/2$ for $u > 0$. It follows that RVs Y_0 and $Y_{N/2}$ have PDF $\exp(u/2 - e^u/4)/(4\pi)^{1/2}$ for all u, whereas RVs Y_k, for $1 \le k \le N/2-1$, have PDF $\exp(u - e^u/2)/2$ for all u. This immediately leads to the common MGF for Y_0 and $Y_{N/2}$ in the form

$$g_0(v) \triangleq 4^v\Gamma(v+1/2)/\sqrt{\pi} \qquad \text{for } -1/2 < v \tag{56}$$

whereas the common MGF for Y_k, $1 \le k \le N/2-1$ is

$$g_1(v) \triangleq 2^v\Gamma(v+1) \qquad \text{for } -1 < v. \tag{57}$$

The functions required for the SPA for the cepstrum estimates are

$$\begin{aligned} g_0(v) &= 4^v\Gamma(v+1/2)/\sqrt{\pi} \\ c_0(v) &= v\log(4) + \log\Gamma(v+1/2) - (1/2)\log(\pi) \\ c_0'(v) &= \log(4) + \psi(v+1/2) \\ c_0''(v) &= \psi'(v+1/2) \\ c_0'''(v) &= \psi''(v+1/2) \\ c_0''''(v) &= \psi'''(v+1/2) \end{aligned} \tag{58}$$

and

$$\begin{aligned} g_1(v) &= 2^v\Gamma(v+1) \\ c_1(v) &= v\log(2) + \log\Gamma(v+1) \\ c_1'(v) &= \log(2) + \psi(v+1) \\ c_1''(v) &= \psi'(v+1) \\ c_1'''(v) &= \psi''(v+1) \\ c_1''''(v) &= \psi'''(v+1) \end{aligned} \tag{59}$$

where ψ is the psi function (see [18, Sec. 6.3 and 6.4].

3) Experimental Validation: Exact solutions to validate the unnormalized ACF estimates (Section IV-B1) and Cepstrum estimates (Section IV-B-2) have not yet been worked out; therefore, approximation error cannot yet be determined. Note, however, that the basic approach is the same as Section IV-A1, which has been validated experimentally. The approaches differ in the choice of matrix $\mathbf{A}$ and the fact that RVs $\{y_n\}$ are not all identically distributed.

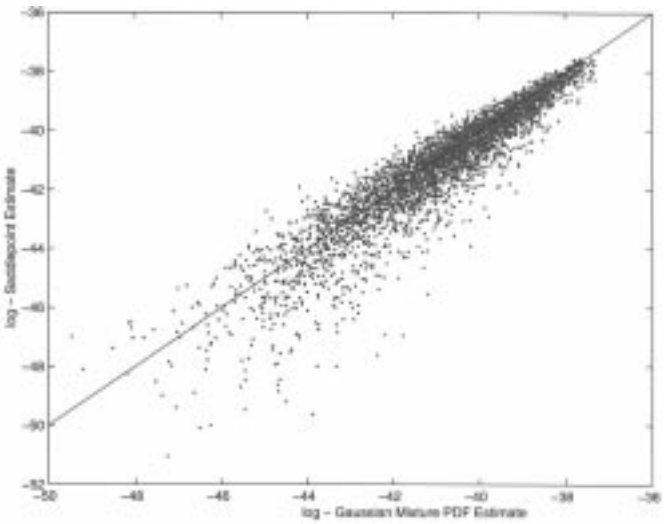

Fig. 3. Comparison of the method of Section IV-B2 with a Gaussian mixture approximation with standard normal input data.

In spite of that, it is still a good idea to validate the analysis with another entirely different approach. To do this, we used a Gaussian mixture (GM) PDF approximation using simulated data. While a PDF approximation obtained from simulated data cannot test errors in the tails, it can at least validate the PDF in the neighborhood of the peak. With this in mind, the method of Section IV-B2 was tested using a noncontiguous set of cepstrum coefficients. We took 4000 independent samples of a set of eight cepstrum outputs from a size-1024 cepstrum. The cepstrum indexes were $\{t_i\} = \{2, 5, 8, 11, 12, 13, 14, 15\}$. The cepstrum processor was excited with RVs $\{x_i\}$ from the standard normal distribution. The 4000 eight-tuples were used as training data for a GM PDF estimator [15]. The same data was used to produce Fig. 3. In this plot, we evaluated the log-PDF for each data sample. The log-PDF from the GM approximation is plotted on the x axis, and the value from the method of Section IV-B2 is plotted on the y axis. Ideally, all data points should fall on the $x = y$ line. The plot shows good agreement, considering the fact that the PDF approximation is in a relatively high dimension. The errors increase in the tails; however, this is due to the mixture approximation and not the SPA.

C. Other Applications

The SPA is applicable whenever the MGF or CGF can be derived. The SPA has been derived for additional statistics including correlated and non-Gaussian statistics [16], [19].

V. Other Asymptotic Methods

A. Reflection Coefficients and Autocorrelation Estimates (Contiguous)

For large N, an asymptotic form is available for the reflection coefficients, which is due to Daniels [2]. Let $[K_1 \cdots K_P]'$ be the reflection coefficients derived from the Levinson recursion

on $[\tilde{r}_1 \cdots \tilde{r}_P]'$ [13]. The general form for the asymptotic (large N) distribution of $[K_1 \cdots K_P]'$ is the following. Let

$$c_o = \log\Gamma\left(\frac{N}{2}\right) - \frac{\log\pi}{2} - \log\Gamma\left(\frac{N-1}{2}\right)$$

$$c_e = \log\Gamma(N+1) - N\log 2 - \log\Gamma\left(\frac{N-1}{2}\right) - \log\Gamma\left(\frac{N+3}{2}\right).$$

Let n_e be the largest integer less than or equal to $P/2$, and let n_o be the smallest integer greater than or equal to $P/2$. Then

$$\log p(K_1 \cdots K_P) = \sum_{i=1}^{n_e}\left\{c_e + 2\log(1+K_{2i}) + \frac{N-3}{2}\log(1-K_{2i}^2)\right\} + \sum_{i=1}^{n_o}\left\{c_o + \log(1+K_{2i-1}) + \frac{N-3}{2}\log(1-K_{2i-1}^2)\right\}. \quad (60)$$

As explained in Section II-B2, the normalized autocorrelation function estimates are related to the reflection coefficients by a one-to-one transformation [13]. The PDF $p(\tilde{r}_1 \cdots \tilde{r}_P)$ requires knowing the Jacobian of the transformation from ACF to reflection coefficients, namely

$$\log J = -\sum_{i=1}^{P-1}(P-i)\log(1-K_i^2).$$

Thus

$$\log p(\tilde{r}_1 \cdots \tilde{r}_P) = \log J + \log p(K_1, K_2 \cdots K_P). \quad (61)$$

B. Experimental Validation

We can compare the approximation for $p(\tilde{r}_1, \tilde{r}_2 \cdots \tilde{r}_P)$ from Section V-A with the exact expression from Section II-B. In order to evaluate tail accuracy, independent samples were passed through an AR filter of order 4 to make them correlated. The AR filter coefficients were chosen at random by selecting reflection coefficients from a uniform distribution in the range [−1,1] and then transforming to AR coefficients and ACF samples. Two-hundred independent trials were computed. The log-PDF value from the exact method of Section II-B was plotted on the X axis of Fig. 4, and the difference between the log-PDF from (61) and the exact expression is plotted on the Y axis. The error was quite small (less than 1 in magnitude) for samples with log-PDF values above -10 but increased in the tails. This could reflect errors due to an inherent assumption of independence. Note, however, that the errors are acceptable even at very low values of log-PDF.

VI. Application to Autoregressive (AR) Model-Order Selection

Consider the problem of determining the order P of an autoregressive process, which is denoted AR(P). The MAP rule

$$\arg\max_P p(H_P|\mathbf{x}) = \max_P p(\mathbf{x}|H_P)p(H_P) \quad (62)$$

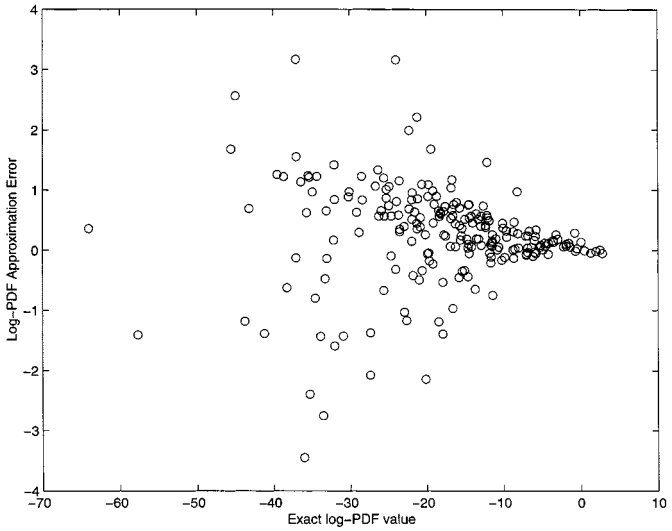

Fig. 4. Comparison of theoretical and approximate PDF for normalized ACF estimates.

where H_P is the hypothesis corresponding to order P, may be implemented using the class-specific approach. Applying (3), we have

$$\arg\max_P \frac{p(\mathbf{z}_P|H_P)}{p(\mathbf{z}_P|H_0)} \quad (63)$$

where we have assumed that $p(H_P)$ is a constant, H_0 corresponds to the case of iid Gaussian noise, and $\mathbf{z}_P$ is a sufficient statistic for the AR(P) process. Of course, this assumes we have the prior knowledge of $p(\mathbf{z}_P|H_P)$. It is interesting to note, however, that the denominator term in (63) usually has a dominant effect on the decision. In fact, Kay [6] has shown that omission of the numerator in (63) and using the rule

$$\arg\max_P \frac{1}{p(\mathbf{z}_P|H_0)} \quad (64)$$

implements the conditional model estimator (CME), which has been shown to outperform the minimum description length (MDL) rule [7]. However, (63) should be an upper bound in performance due to knowledge of the numerator PDF (which we call the *a priori* PDF of $\mathbf{z}_P$).

We now test formulas (63) and (64) and compare with the Akaike and MDL method [13]. As an approximate sufficient statistic for an AR(P) process, we use the vector of circular ACF estimates (7), which is denoted

$$\hat{\tilde{\mathbf{r}}}^P = [\hat{r}_1/\hat{r}_0, \hat{r}_2/\hat{r}_0 \cdots \hat{r}_P/\hat{r}_0]'.$$

In the experiment, we generate an AR process with known P in the range $1 \le P \le 4$. We utilize odd data record lengths of $N =$ 11, 13, 15, 17, 21, 127, and 255 with 1000 independent trials for each combination of P, N. In each trial, the AR process is determined by randomly selecting the reflection coefficients (RCs) from a uniform distribution on [−1,1]. The only restriction was that the Pth RC had to be greater than 0.2 in magnitude (to make sure the data model was truly of order P). The RCs were then converted into AR coefficients, and an AR process was created

by filtering independent Gaussian noise. The model order selection was accomplished by determining the best fit of the given approach over $1 \leq P \leq 5$. We compared the following approaches:

- *CS-RC: Class-Specific Using RC:* Because the RC estimates are related to $\hat{\hat{\mathbf{r}}}^P$ by a one-to-one transformation, they are equivalent from a sufficiency point of view. We may approximate the PDF of the RC estimates $p(\hat{\mathbf{K}}^P|H_P)$ by the PDF of the true RCs used in the experiment. This approach should provide somewhat of an upper bound of performance since it makes use of knowledge not available to the other methods. Let $\hat{\mathbf{K}}^P \triangleq [\hat{K}_1, \hat{K}_2 \cdots \hat{K}_P]'$ be the vector of RC estimates. Equation (63) becomes

$$\max_P \frac{p(\hat{\mathbf{K}}^P|H_P)}{p(\hat{\mathbf{K}}^P|H_0)} \tag{65}$$

  with (61) used to approximate the denominator PDF. We used the prior density of $p(\hat{\mathbf{K}}^P|H_P) = 2^{-P}$, which is the density used for the true RCs in the experiment. We ignored the effect of the constraint $|K_P| > 0.2$ and the fact that the RC estimates $\hat{\mathbf{K}}$ do not in fact have the same density as the true RCs.
- *CME-ACF—CME Approach Using ACF:* Implementing CME using $\hat{\hat{\mathbf{r}}}^P$, (63) becomes

$$\max_P \frac{1}{p\left(\hat{\hat{\mathbf{r}}}^P|H_0\right)} \tag{66}$$

  with (61) used to approximate the denominator PDF.
- *Akaike Method (Circular ACF):* The Akaike method is

$$\min_P \left\{N \log \hat{\sigma}_P^2 + 2P\right\} \tag{67}$$

  where $\hat{\sigma}_P^2$ is the estimate of the power of the white noise driving sequence computed from the ACF estimates $\hat{\mathbf{r}}$ using the Levinson algorithm [13]. The *circular* ACF estimates were used.
- *Minimum Description Length (MDL):* The MDL method is

$$\min_P \left\{N \log \hat{\sigma}_P^2 + P \log N\right\}. \tag{68}$$

Results of the experiment are provided in Tables I–IV for true value of P ranging from 1 to 4. It is difficult to compare the performance of the various approaches based on just one value of true P. This is due to biases that cause a given approach to perform better at a particular P and worse at another. For example, the Akaike method is biased toward a higher value of P and thus appears better at $P = 4$ for low values of N. Average performance (averaged over P) is plotted in Fig. 5. The results show that at low N, the CS-RC and CME-ACF methods outperformed both the MDL and Akaike methods consistently by about 3%. At high N, all methods were similar, except the Akaike method, which is known to be an inconsistent estimator of model order. The advantage of the known prior in the CS-RC approach was not significant compared with CME-ACF, although it appears to always outperform the CME-ACF method by a small amount. The CS-RC approach, which has some prior knowledge of the distribution of the RCs, provided better performance most of the time than the other approaches; however, it appears to fall

TABLE I
RESULTS FOR $P = 1$. PROBABILITY OF CORRECT MODEL ORDER IN PERCENT

N	CS-RC	CME-ACF	Akaike	MDL
11	85.3	65.6	60.5	68.5
13	87.4	69.5	62.1	72.8
15	87.0	71.2	66.6	78.9
17	88.4	74.5	68.4	81.7
21	90.2	78.4	70.0	85.8
127	96.4	94.2	74.4	96.7
255	97.0	96.5	75.1	97.6

TABLE II
RESULTS FOR $P = 2$. PROBABILITY OF CORRECT MODEL ORDER IN PERCENT

N	CS-RC	CME-ACF	Akaike	MDL
11	28.2	38.2	27.8	26.9
13	32.4	40.8	30.4	30.3
15	40.4	47.3	35.6	36.8
17	41.9	48.1	38.7	38.0
21	53.7	57.0	48.3	48.8
127	86.3	86.5	67.8	86.0
255	93.8	95.1	70.8	94.1

TABLE III
RESULTS FOR $P = 3$. PROBABILITY OF CORRECT MODEL ORDER IN PERCENT

N	CS-RC	CME-ACF	Akaike	MDL
11	9.2	17.7	11.5	10.0
13	15.0	20.3	16.3	14.7
15	21.0	27.9	23.4	21.0
17	24.7	29.9	27.5	22.3
21	33.3	36.9	33.6	29.4
127	82.9	85.9	64.7	83.2
255	86.6	89.8	65.2	87.3

TABLE IV
RESULTS FOR $P = 4$. PROBABILITY OF CORRECT MODEL ORDER IN PERCENT

N	CS-RC	CME-ACF	Akaike	MDL
11	6.6	8.5	15.5	12.2
13	7.9	10.5	17.4	13.6
15	10.6	13.3	20.4	14.6
17	16.2	16.6	23.6	17.4
21	24.1	25.9	31.9	25.6
127	72.7	74.7	63.8	72.8
255	80.4	84.1	65.2	81.0

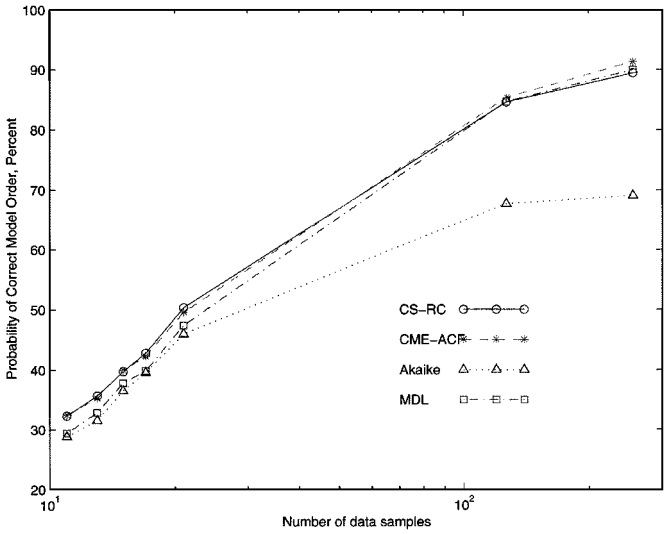

Fig. 5. Model order selection results averaged over $P = 1$–4.

below CME at high N. This can be explained by the fact that the CS-RC approach, while it has prior knowledge, is only an approximate MAP implementation.

VII. Conclusions

In this paper, we have provided approximations to the joint multidimensional PDFs of some important statistics in signal processing, including autocorrelation estimates, Cepstrum estimates, and general linear functions of independent RVs. Although the approaches we have used can, in principle, be used for any input data hypothesis, we have provided examples in which the input data is assumed to be iid samples of Gaussian noise. Because these approximations are valid in the tails of the PDF, they can be used in conjunction with the PDF projection theorem (3) to provide PDF estimates of the input raw data for real-world statistical hypotheses. An application of the method has provided an AR model-order selection approach that outperforms the MDL and Akaike methods. The model selection approach is quite general and can, in principle, be applied to any model-order selection problem, provided there exists a well-defined approximate sufficient statistic or approximate sufficient statistic for each model order.

Appendix

A. MATLAB Implementation of Equation (8)

The function `corrpdf(r,N)` below returns the PDF value for an input vector $\mathtt{r} = [\tilde{r}_1 \cdots \tilde{r}_P]'$, which has been obtained from N input data samples. The program requires N to be odd. The algorithm becomes numerically unstable for N greater than about 30, and it is slow. A faster version, using arbitrary precision arithmetic, has been written in C, which extends its usefulness to higher N (as high as 500 has been tried successfully); however, it is still somewhat impractical. The usefulness of these programs is in the fact that they are able to validate the asymptotic form to be presented in Section III.

```
function pdf = corrpdf(r, N)
%%%%%%%%%%% CORRPDF.M %%%%%%%%%%%%%%%%%%%%%
% function pdf = corrpdf(r, N)               %
% Code developed by S. M. Kay, 1998          %
% Modified by P. M. Baggenstoss              %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
r = r(:);
M = length(r);
if round((N-1)/2) = (N-1)/2,
   error("N must be odd");
end
n = (N-1)/2;
lambda = zeros(n, M);
for j = 1: n
lambda(j, :) = cos(2*pi*j*[1:M]/N);
end
sm = 0;
for j1 = 1: n
  sm = corrloop(j1, r, lambda, sm);
end;
pdf=sm*prod(n - [1: M]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% subroutine corrloop                        %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function term = corrloop(idx, r, lambda, term);
[n, M] = size(lambda);
m = length(idx);
if(m >= M),
   B = lambda(idx, 1: M);
     alpha = B'\r;
     if min (alpha) > 0 & sum(alpha) < 1,
A = [1 r'; ones(M, 1) B];
da = det(A);
C = 1;
for j = 1: n
   if sum(j == idx) == 0,
     D = [1 lambda(j, 1: M); ones(M, 1) B];
     C = C*(det (D)/da);
   end
end
term = term+det(A)^(-1)*sign(det(B))/C;
     end
else,
   for i = idx(m) + 1: n,
      term = corrloop([idx; i], r, lambda, term);
   end;
end;
```

References

[1] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Applications*. San Francisco, CA: Holden-Day, 1968.

[2] H. Daniels, "The approximate distribution of serial correlation coefficients," *Biometrika*, pp. 169–185, 1956.

[3] M. H. Quenouille, "The joint distribution of serial correlation coefficients," *Ann. Math. Stat.*, vol. 20, pp. 561–571, 1949.

[4] E. J. Hannan, *Multiple Time Series*. New York: Wiley, 1970.

[5] P. M. Baggenstoss, "Class-specific features in classification," *IEEE Trans Signal Processing*, vol. 47, pp. 3428–3432, Dec. 1999.

[6] S. Kay, "Sufficiency, classification, and the class-specific feature theorem," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1654–1658, July 2000.

[7] ——, "Conditional model estimation," *IEEE Trans. Signal Processing*, vol. 49, pp. 1910–1917, Sept. 2001.
[8] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part 3, Radar–Sonar Signal Processing and Gaussian Signals in Noise*. New York: Wiley, 1971.
[9] P. M. Baggenstoss, "A modified Baum–Welch algorithm for hidden Markov models with multiple observation spaces," *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 411–416, May 2001.
[10] ——, "A theoretically optimum approach to classification using class-specific features," in *Proc. ICPR*, Barcelona, Spain, 2000.
[11] O. E. Barndorff-Nielsen and D. R. Cox, *Asymptotic Techniques for Use in Statistics*. London, U.K.: Chapman & Hall, 1989.
[12] H. A. David, *Order Statistics*. New York: Wiley, 1981.
[13] S. Kay, *Modern Spectral Estimation: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[14] G. Watson, "On the joint distribution of the circular serial correlation coefficients," *Biometrika*, vol. 43, pp. 161–168, 1956.
[15] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
[16] A. H. Nuttall, "Saddlepoint approximation and first-order correction term to the joint probability density function of M quadratic and linear forms in K Gaussian random variables with arbitrary means and covariances,", Newport, RI, Naval Underwater Warfare Cent. Tech. Rep. 11 262, Dec. 2000.
[17] ——, "A class of linear transformations of independent RV's for which the exact PDF can be calculated," memorandum, 2000.
[18] *Handbook of Mathematical Functions*, ser. Applied Math. Series 55. Washington, DC: Nat. Bur. Stand., U.S. Govt. Printing Office, June 1964.
[19] A. H. Nuttall, "Saddlepoint approximations for various statistics of dependent non-Gaussian random variables; Applications to the maximum variate and the range variate,", Newport, RI, Naval Underwater Warfare Cent. Tech. Rep. 11 280, Apr. 2001.

Steven M. Kay (F'89) was born in Newark, NJ, on April 5, 1951. He received the B.E. degree from Stevens Institute of Technology, Hoboken, NJ, in 1972, the M.E. degree from Columbia University, New York, NY, in 1973, and the Ph.D. degree from the Georgia Institute of Technology (Georgia Tech), Atlanta, in 1980, all in electrical engineering.

From 1972 to 1975, he was with Bell Laboratories, Holmdel, NJ, where he was involved with transmission planning for speech communications and simulation and subjective testing of speech processing algorithms. From 1975 to 1977, he attended Georgia Tech to study comunications theory and digital signal processing. From 1977 to 1980, he was with the Submarine Signal Division, Raytheon Corporation, Portsmouth, RI, where he engaged in research on autoregressive spectral estimation and design of sonar systems. He is currently Professor of electrical engineering at the University of Rhode Island, Kingston, and a consultant to industry and the United States Navy. He has written numerous papers, many of which have been reprinted in the IEEE Press book *Modern Spectral Analysis II*. He is a contributor to several edited books on spectral estimation and is the author of *Modern Spectral Estimation: Theory and Application* (Englewood Cliffs, NJ: Prentice-Hall, 1993). He conducts research in mathematical statistics with applications to digital signal processing. This includes the theory of detection, estimation, time series, and spectral analysis with applications to radar, sonar, communications, image processing, speech processing, biomedical signal processing, vibration, and financial data analysis.

Dr. Kay is a member of Tau Beta Pi and Sigma Xi. He has served on the IEEE Acoustics, Speech, and Signal Processing Committee on Spectral Estimation and Modeling.

Albert H. Nuttall received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1954, 1955, and 1958, respectively.

He was with MIT as an Assistant Professor until 1959. From 1957 to 1960, he was with Melpar, and from 1960 to 1968, he was with Litton Industries. He was with the Naval Underwater Systems Center (NUSC), New London, CT, where his interest was in statistical communication theory. He is now with the Naval Undersea Warfare Center, Newport, RI.

Dr. Nuttall received the NUSC Distinguished Chair in Signal Processing from the Navy in April 1987.

Paul M. Baggenstoss (S'82–M'82) was born in Gastonia, NC, in 1957. He received the B.S.E.E. degree in 1979 and the M.S.E.E. degree in 1982 from Rensselear Polytechnic Institute (RPI), Troy, NY. He received the Ph.D. degree in statistical signal processing from the University of Rhode Island, Kingston, in 1990, under the supervision of Prof. S. Kay.

From 1979 to 1996, he was with Raytheon Company, Waltham, MA, and joined the Naval Undersea Warfare Center, Newport, RI, in August 1996. Since then, he has been involved with classification and pattern recognition. He was an Adjunct Professor with the University of Connecticut, Storrs, where he taught detection theory and digital signal processing. In February 2000, he began a joint research effort with the Pattern Recognition Chair, University of Erlangen, Erlangen, Germany.