Real-time Estimation of Parameters of NonGaussian PDFs using Cumulant Generating Functions

Steven Kay
Dept. of Electrical and Computer Engineering
University of Rhode Island
Kingston, RI 02881
401-874-5804
kay@ele.uri.edu

David Middleton
127 East 91st Street
New York, NY 10128-1601
212-831-8565

Abstract — **Asymptotically optimal estimation of parameters of analytic nonGaussian probability density functions is usually accomplished using maximum likelihood estimators. However, owing to the complex analytical forms of the density functions this approach is computationally demanding and not suitable for real-time implementation. As an alternative to maximum likelihood, we propose a method based on least squares fitting of the cumulant generating function. It is shown to produce good results and is easily implemented in real-time. Application to the Class A distribution is used to illustrate the method, with its application to other parametric models planned.**

I. Introduction

Many models of nonGaussian probability density functions (PDF) have been proposed to represent noise which is encountered in radar and sonar (and other physical applications). Most of these are ad hoc, without a strong physical basis, and are chosen for their analytical simplicity rather than their physical applicability. Recently, some new models primarily based on physical-statistical approaches have been investigated, which include such features as fluctuating background noise (K-distributions), and the effects of a few large scatterers (breaking waves, bubble trains,etc.), as well as multipath and possible coherent scatter (the KA-model [1]). Here the physical-statistical approach involves independent orders of scatter expressed statistically in the form of independent Poisson processes, enabling one to calculate the effects of all levels of scatter (weak to strong). The results are expressed in terms of first and higher-order PDFs as opposed to the first few *moments* of classical theory [1]. The great advantages of this is that it is possible to obtain the optimum weak-signal detection algorithms and their associated performance measures, namely the probabilities of false alarm and correct detection, which are needed to predict performance (see *Introduction* [1]). These are analytical parametric models, which are parsimonious in the number of physical parameters, and which in turn are expressed in terms of measurable quantities.

To illustrate the role of the nonGaussian PDFs in detection let us use the case of optimum binary weak-signal operation in the coherent mode as an example. For this the threshold algorithm is the well known result [10], for additive signal and noise under $\mathcal{H}_1 : S + N$; $\mathcal{H}_0 : N$

$$\ln \Gamma_{\text{coh}} = q(\mathbf{x})^*_{\text{coh}} = \ln \mu + B^*_{\text{coh}} - \sum_{n=1}^{N} l(x_n) < s_n > \qquad (1)$$

where Γ_{coh} is the likelihood ratio, $\mathbf{x} = [x_1\, x_2 \ldots x_N]^T$, B^*_{coh} is a bias, and the last term is a discrete cross-correlation of a (zero memory) nonlinear function of the data with the (average of) the signal, $< s_n >$, $n = 1, 2, \ldots, N$. [Independent noise samples are necessarily - postulated in the general, i.e., nonGaussian cases, because the higher order PDFs are unknown.] Here we have explicitly

$$\begin{aligned} l(x_n) &= \left[\frac{d}{dx} \ln p(x|\mathcal{H}_0)\right]_{x=x_n} \\ B^*_{\text{coh}} &= -\frac{L^{(2)}}{2} \sum_{n=1}^{N} < s_n >^2 \\ L^{(2)} &= \left\langle \left(\frac{p'(x|\mathcal{H}_0)}{p(x|\mathcal{H}_0)}\right)^2 \right\rangle_{\mathcal{H}_0} \end{aligned} \qquad (2)$$

in which $< >_{\mathcal{H}_0}$ denotes the average under the null (i.e., no signal) hypothesis, and the test for signal present or absent is $\ln \Gamma_{\text{coh}} \geq K$, or $< K$, where K is a positive threshold. For our purposes here (1) and (2) show directly how the structure of the detector depends on the PDF $p(x|\mathcal{H}_0)$. Since our noise models for $p(x|\mathcal{H}_0)$ are parametric, this in turn depends on the model parameters, and consequently on our ability to measure them in practical situations, particularly where on-line operation is required. The comparatively simple noise example given here is the Class A distribution model [8]. Other widely used cases employ the Class B or impulsive model [7], and the K, and KA distributions [1].

Unfortunately, the complex analytical form of these models precludes a real-time estimator of its parameters. Such an estimator is desired in order to implement the weak-signal or locally optimum detector [7,2] in practical applications. In this paper we propose a method for estimation which is useful for PDFs whose characteristic functions are readily available in analytical form. The method is not asymptotically optimal but can be made so by using more computation [3,4]. We investigate the performance of this estimator, termed the *cumulant function estimator (CFE)*, and apply it to the Middleton Class A model [8,9]. It is shown to produce reliable estimates with very little computation. It should be mentioned that the EM approach has been tried for Class A model parameter estimation [5]. The reported statistical performance is that expected of an MLE. However, the convergence to the global maximum, i.e, the true MLE, is always in question, and the procedure is computationally demanding. Furthermore, in [5] only a two-parameter estimator was implemented. We are able to implement the full three parameter estimator, as explained in the next section. Finally, it is important to note that this approach can easily be extended to other nonGaussian PDFs such as the K and KA PDFs [1].

The ultimate aim of this research is to provide on-line, i.e., real-time, estimates of the parameters of the analytic PDFs, and related functions (involving these parameters) of the detection processs. Thus, such detectors become truly adaptive, taking into account the secular changes in the noise environment and enabling the detector to remain "matched" to the noise optimally, or at least near-optimal in practical operation. The same parameter estimates may also be used in estimation problems.

II. The Cumulant Generating Function Estimator

The characteristic function of a random variable X is defined as the Fourier transform of its PDF or $\phi_X(\omega) = E(\exp(j\omega X))$. As such it is an equivalent description of the PDF. If N independent and identically distributed (IID) observations $\{x_1, x_2, \ldots, x_N\}$ are available, then one can estimate the characteristic function by its sample mean as

$$\hat{\phi}_X(\omega) = \frac{1}{N}\sum_{i=1}^{N} \exp(j\omega x_i) \quad (3)$$

or by

$$\hat{\phi}_X(\omega) = \frac{1}{N}\sum_{i=1}^{N} \cos(\omega x_i) \quad (4)$$

if the PDF is known to be symmetric, as is the case for most noise models. Assuming that the PDF depends on an unknown parameter vector $\boldsymbol{\theta}$ and hence the characteristic function depends on $\boldsymbol{\theta}$, we can estimate the parameters by minimizing the least squares error

$$J'(\boldsymbol{\theta}) = \sum_{k=1}^{M} \left(\hat{\phi}_X(\omega_k) - \phi_X(\omega_k)\right)^2$$

for some suitable sets of ω_k's. However, it is usually more advantageous to fit the estimated *cumulant generating function (CGF)* or $\ln \hat{\phi}_X(\omega)$, since it tends to be partially linear in the unknown parameters. Such is the case with the Class A PDF model. Hence, we propose to estimate the parameters of the class A model by minimizing

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{M} \left(\ln \hat{\phi}_X(\omega_k) - \ln \phi_X(\omega_k)\right)^2. \quad (5)$$

The implementation in this case is relatively simple. Note that in order to make the estimator into one that is asymptotically efficient it is only necessary to include the asymptotic covariance matrix in (5), since the characteristic function estimator is asymptotically Gaussian [3]. This implies that the estimated cumulant generating function is also asymptotically Gaussian. Thus, instead of minimizing the least squares error as is done in (5), one includes the asymptotic covariance matrix $\mathbf{C}_{\hat{\phi}}$ of the CGF estimate samples whose (k,l) element is

$$[\mathbf{C}_{\hat{\phi}}]_{kl} = E\left[\left(\ln \hat{\phi}_X(\omega_k) - E(\ln \hat{\phi}_X(\omega_k))\right)\left(\ln \hat{\phi}_X(\omega_l) - E(\ln \hat{\phi}_X(\omega_l))\right)\right]$$

so that the *weighted* least squares error becomes

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{M}\sum_{l=1}^{M} [\mathbf{C}_{\hat{\phi}}]^{kl} \left(\ln \hat{\phi}_X(\omega_k) - \ln \phi_X(\omega_k)\right)\left(\ln \hat{\phi}_X(\omega_l) - \ln \phi_X(\omega_l)\right). \quad (6)$$

where $[\mathbf{C}_{\hat{\phi}}]^{kl}$ denotes the (k,l) element of $\mathbf{C}_{\hat{\phi}}^{-1}$. Unfortunately, the covariance matrix depends on the unknown parameters, so that we have to resort to iterative techniques, which are not guaranteed to converge or to produce the global minimum of (6), i.e, the asymptotic MLE. Yet for large enough data records the performance may be acceptable.

III. Class A Model

The Class A model is described by the PDF

$$p_X(x; A, \Omega, \sigma_G^2) = \exp(-A) \sum_{m=0}^{\infty} \frac{A^m}{m!\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma_m^2}\right). \quad (7)$$

The parameters are defined as follows. The parameter A is a Poisson mean parameter. The parameter Ω is the nonGaussian noise power while σ_G^2 is the Gaussian noise power. The summarized parameter σ_m^2 is defined as [1]

$$\sigma_m^2 = \frac{m\Omega}{A} + \sigma_G^2. \quad (8)$$

It is sometimes more convenient and physically more meaningful to use the equivalent parameter set $\{A, \Gamma, \sigma_G^2\}$, where $\Gamma = \sigma_G^2/\Omega$ is the ratio of the Gaussian to nonGaussian noise power. We shall use this equivalent parameter set for our computer simulations in Section 4 to be consistent with previous work [1]. It should be noted that the random variable X can alternatively be described by the decomposition

$$X = \sum_{i=1}^{N} \xi_i + Z \quad (9)$$

where the ξ_i's are IID $\mathcal{N}(0, \Omega/A)$ random variables and Z is a $\mathcal{N}(0, \sigma_G^2)$ random variable that is independent of the ξ_i's. Also, N is a Poisson random variable with mean A. The first term $I = \sum_{i=1}^{N} \xi_i$ is a compound Poisson random variable which models nonGaussian noise while the second random variable Z models the ambient Gaussian noise. This representation is used to generate Class A noise samples used in the computer simulation described in Section 4. A MATLAB subprogram is listed in Appendix A for this purpose.

To show that (9) has the PDF of (7) first note that the PDF of N is Poisson with the probability mass function

$$P_N(m) = \exp(-A)\frac{A^m}{m!} \qquad m = 0, 1, \ldots \quad (10)$$

so that conditioned on $N = m$, the PDF of X is $\mathcal{N}(0, m\Omega/A + \sigma_G^2)$ or $\mathcal{N}(0, \sigma_m^2)$. The unconditional PDF is obtained by summing over the probability mass function of (10), resulting in (7). It is thus seen that the Class A model is essentially a Gaussian mixture model with a Poisson mixing PDF.

Because of the relatively simple form of (9) the CGF is easily determined. Conditioned on a fixed value of $N = m$, we

[1] In the original work, $2\hat{\sigma}_m^2 = \sigma_m^2$ of the present paper; see Eq. (6.1) of [9], for example.

have that $\sum_{i=1}^{N}\xi_i \sim \mathcal{N}(0, m\Omega/A)$ and hence $\phi_I(\omega|N=m) = \exp(-(1/2)\omega^2 m\Omega/A)$. Averaging the conditional expectation with respect to the probability mass function of (10) yields

$$\phi_I(\omega) = \exp\left[A\left(\exp[-(1/2)\omega^2\Omega/A] - 1\right)\right]$$

and finally because of the assumed independence of I and Z, we have

$$\phi_X(\omega) = \exp\left[A\left(\exp[-(1/2)\omega^2\Omega/A] - 1\right) - (1/2)\sigma_G^2\omega^2\right].$$

. The CGF is then

$$K_X(\omega) = A\left(\exp[-(1/2)\omega^2\Omega/A] - 1\right) - (1/2)\sigma_G^2\omega^2. \quad (11)$$

To simplify the estimation procedure we let $K = \Gamma A = A\sigma_G^2/\Omega$ as was done in [5], and also $\alpha = \sigma_G^2/K$, so that we have

$$K_X(\omega) = A\left[\exp(-(1/2)\alpha\omega^2) - 1\right] - (1/2)K\alpha\omega^2. \quad (12)$$

Note that the CGF is linear in the parameters A and K and nonlinear in α. We can set up the partially linear model as

$$\underbrace{\begin{bmatrix} \hat{K}_X(\omega_1) \\ \hat{K}_X(\omega_2) \\ \vdots \\ \hat{K}_X(\omega_M) \end{bmatrix}}_{\hat{\mathbf{K}}} =$$

$$\underbrace{\begin{bmatrix} \exp[-(1/2)\alpha\omega_1^2] - 1 & -(1/2)\alpha\omega_1^2 \\ \exp[-(1/2)\alpha\omega_2^2] - 1 & -(1/2)\alpha\omega_2^2 \\ \vdots & \vdots \\ \exp[-(1/2)\alpha\omega_M^2] - 1 & -(1/2)\alpha\omega_M^2 \end{bmatrix}}_{\mathbf{H}(\alpha)} \underbrace{\begin{bmatrix} A \\ K \end{bmatrix}}_{\boldsymbol{\theta}} + \mathbf{w} \quad (13)$$

where $\hat{K}_X(\omega)$ represents an estimate of the CGF and $\mathbf{w}$ represents a noise vector which models the statistical estimation error in $\hat{K}_X(\omega)$. The points used for ω in the least squares estimator are chosen to be equally spaced. The least squares estimator for $\boldsymbol{\theta}$ is found by minimizing the least squares error, which in matrix form is [6]

$$J(\boldsymbol{\theta}, \alpha) = \left(\hat{\mathbf{K}} - \mathbf{H}(\alpha)\boldsymbol{\theta}\right)^T \left(\hat{\mathbf{K}} - \mathbf{H}(\alpha)\boldsymbol{\theta}\right). \quad (14)$$

Expanding this into

$$J(\boldsymbol{\theta}, \alpha) = \hat{\mathbf{K}}^T\hat{\mathbf{K}} - 2\hat{\mathbf{K}}^T\mathbf{H}(\alpha)\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{H}(\alpha)^T\mathbf{H}(\alpha)\boldsymbol{\theta}$$

and taking the gradient, using the formulas

$$\frac{\partial \mathbf{a}^T\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} = \mathbf{a}$$

$$\frac{\partial \boldsymbol{\theta}^T\mathbf{A}\boldsymbol{\theta}}{\partial\boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta} \qquad (\text{assuming } \mathbf{A}^T = \mathbf{A})$$

and setting equal to zero and solving, produces the result

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T(\alpha)\mathbf{H}(\alpha))^{-1}\mathbf{H}^T(\alpha)\hat{\mathbf{K}}. \quad (15)$$

After substituting this in the least squares error of (14) we obtain

$$J(\hat{\boldsymbol{\theta}}, \alpha) = \hat{\mathbf{K}}^T\hat{\mathbf{K}} - \hat{\mathbf{K}}^T\mathbf{H}(\alpha)(\mathbf{H}^T(\alpha)\mathbf{H}(\alpha))^{-1}\mathbf{H}^T(\alpha)\hat{\mathbf{K}} \quad (16)$$

which needs to be minimized over α. Equivalently, we must maximize

$$J(\alpha) = \hat{\mathbf{K}}^T\mathbf{H}(\alpha)(\mathbf{H}^T(\alpha)\mathbf{H}(\alpha))^{-1}\mathbf{H}^T(\alpha)\hat{\mathbf{K}} \quad (17)$$

over α, since $\hat{\mathbf{K}}^T\hat{\mathbf{K}}$ does not depend on α. Once that is done via numerical means to obtain $\hat{\alpha}$, $\hat{\boldsymbol{\theta}}$ is found from (15) by replacing α by $\hat{\alpha}$. Note that the estimator $\hat{\mathbf{K}}$ is defined as the sample mean estimator

$$[\hat{\mathbf{K}}]_k = \ln\frac{1}{N}\sum_{i=1}^{N}\cos(\omega_k x_i) \qquad k = 1, 2, \ldots, M \quad (18)$$

since the Class A PDF is symmetric. In the next section we give some results.

IV. Computer Simulation Results

In the first simulation example the parameter values were $\Gamma = 0.5 \times 10^{-3}$, $A = 0.35$, and $\sigma_G^2 = 1$. Using $N = 10,000$ data samples generated according to (9) (see Appendix A), the estimated CGF and true CGF are shown in Figure 1. Note the good agreement, which is clearly due to the consistency of the sample characteristic function estimator. The estimated values for this simulation were $\hat{\Gamma} = 0.5242 \times 10^{-3}$, $\hat{A} = 0.3511$, and $\hat{\sigma}_G^2 = 1.0140$. Note that we used $M = 100$ equally spaced ω values on the interval [0,1]. Also, the true value of α is $\alpha = \sigma_G^2/(A\Gamma) = 5714$ and so we maximized the function (17) by searching over the interval $[0, 2\alpha]$ in one increment steps. For shorter data records the results are still reasonable. As an example, for $N = 500$ the estimated CGF is shown in Figure 2 with the parameters estimated as $\hat{\Gamma} = 0.4949 \times 10^{-3}$, $\hat{A} = 0.3221$, and $\hat{\sigma}_G^2 = 0.9314$.

V. Conclusions

A new method for the real-time estimation of the parameters of a nonGaussian PDF has been proposed. This method enables us to evaluate the parameters of an analytic PDF on-line, and thus obtain a full representation of the distribution. Our method is applied here to the first-order case, but is capable of extension to the second-order, from which, on the assumption of (simple) Markovian properties (see Sec. 1.4–3 of [7]) we can establish the nth order cases. Current work is ongoing to compare the statistical performance of the proposed estimator in terms of its mean, variance, and relative efficiency, including applications to the K and KA-distributions. In addition, application to real nonGaussian sonar reverberation data is in progress.

Acknowledgments

This work was supported by the Office of Naval Research under contracts No. N00014-00-1-0871 and No. N00014-01-1-0845.

References

1. D. Middleton, "New Physical-Statistical Methods and Models for Clutter and Reveberation: The KA-Distribution and Related Probability Structures", *IEEE Journal of Oceanic Engineering*, Vol. 24, No. 3, July 1999, pp. 261–284; Errata, ibid, pp. 300–301, Vol. 26, No. 3, April 2001.
2. S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, Upper Saddle River, NJ, 1998.

3. A. Feuerverger, P. McDunnough, "On the Efficiency of Empirical Characteristic Function Procedures", *Journal Royal Statistical Society B.*, 1981, pp. 20–27.

4. A. Feuerverger, P. McDunnough, "On Some Fourier Methods for Inference", *Journal American Statistical Association*, June 1981, pp. 379–387.

5. S.M. Zabin, H.V. Poor, "Efficient Estimation of Class A Noise Parameters via the EM Algorithm", *IEEE Trans. on Information Theory*, Jan. 1991, pp.60–72.

6. S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

7. D. Middleton, *Introduction to Statistical Communication Theory*, Classic Edition, IEEE Press, Piscataway, NJ, 1996.

8. ———, "Canonical and Quasi-Canonical Probability Models of Class A Interference", *IEEE Trans. on Electromagnetic Compat.*, *EMC-25*, pp. 76–106, May, 1983.

9. ———, "Non-Gaussian Noise Models in Signal Processing for Telecommunications", *IEEE Trans. on Information Theory*, Vol. 46, No. 4, May 1999, pp. 1129–1149.

10. ———,"Threshold Detection in Non-Gaussian Interference Environments. Exposition and Interpretation of New Results for EMC Applications", *IEEE Trans. Electromag. Compat.*, Vol. EMC-26, pp. 14–28, Feb. 1984.

A. MATLAB Code

Generation of Class A Noise Samples

```
% classAgenerate.m
%  Input parameters:
%    nreal  - number of samples of class A
%             noise desired
%    A      - mean of Poisson random variable
%    Gamma  - ratio of Gaussian to nonGaussian
%             noise power
%    varg   - Gaussian noise power
%
%  Output parameters:
%    x      - real array of dimension nreal x 1
%             of class A noise samples
function x=classAgenerate(nreal,A,Gamma,varg)
x=zeros(nreal,1);m=x;
for i=1:nreal
m(i,1)=pois(A);
var=varg*((m(i)+Gamma*A)/(Gamma*A));
x(i,1)=sqrt(var)*randn(1,1);
end
% pois.m
%
%  This subprogram generates a realization of
%  Poisson random variable with mean lambda
function k=pois(lambda)
prod=1;k=0;
while prod>exp(-lambda)
   prod=prod*rand(1,1);
   k=k+1;
end
k=k-1;
```

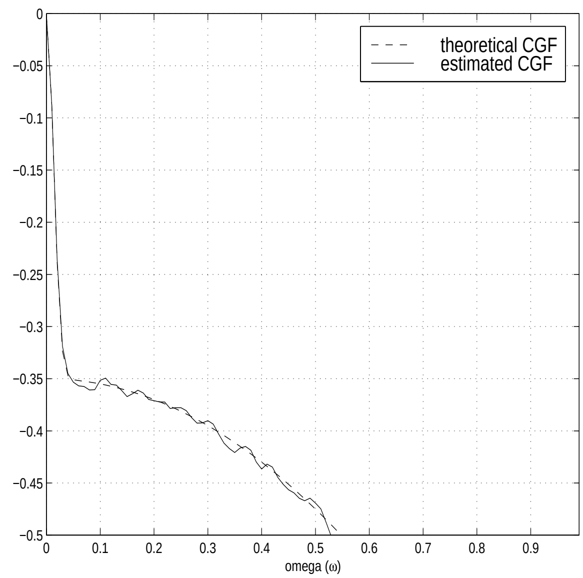

Fig. 1: Theoretical and Estimated Cumulant Generating Function for Class A Noise - $N = 10,000$

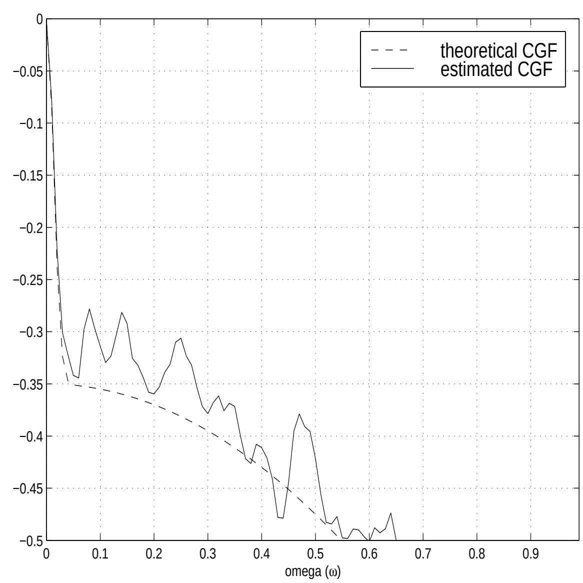

Fig. 2: Theoretical and Estimated Cumulant Generating Function for Class A Noise - $N = 500$