Journal of ELECTRICAL ENGINEERING, VOL. 58, NO. 2, 2007, 104–108

SELF–CRITICAL PROCEDURES IN SIGNAL DETECTION PART I: STRUCTURAL CHARACTERIZATION

Ndeh Ntomambang Ningo *

This paper presents a set of new receiver structures for the detection of weak signals in non- Gaussian noise. The receiver structures result from an extension of the likelihood ratio based on the information generating function of a density. These detectors are termed self-critical in that the assumed noise probability density function serves as an adaptive factor in the information processing algorithm by being continuously used in assessing the degree of agreement between data and model. The self-criticism and the extension of the likelihood ratio depend on a parameter k which is under operator control.

K e y w o r d s: self-critical procedure, locally optimum receiver, self-critical log-likelihood ratio

1 INTRODUCTION

The likelihood ratio and functions derived from it provide a unifying theoretical foundation for the design of globally optimum receivers for the detection of signals in noise [1–4]. The resulting structures are parametric in that they depend on a precise *a priori* knowledge of the probability density function (pdf) of the corrupting noise environment.

An especially important area of signal detection within parametric procedures is concerned with the reception of weak signals, where this nomenclature means that, in comparison to the additive noise background, the signal is vanishingly small. The weak signal detection problem has received much attention [2–4] in the past because of the requirements of modern communication systems involving privacy of information, security and integrity of messages, and economic use of available power sources.

An approach to the design of detectors under weak signal conditions involves a search for a receiver structure that maximizes the slope of the detection probability under the assumption of a constant false alarm probability. Such a detector is termed a locally optimum receiver (LOR) and employs a test statistic which is a combination of a zero-memory nonlinearity (ZMNL) and the matched filter for known low-pass signals in Gaussian noise [2–5].

The above discussion has involved results that are parametric in nature in that, for their implementation, they require that the noise pdf be known precisely. Given the ever changing nature of the underlying noise background, these procedures, even though optimized for the particular signal conditions, may still be quite sensitive to departures from the assumed statistical model. In view of this and other difficulties associated with parametric procedures and in an attempt to avoid them, a concerted effort has been exerted in the recent past on developing robust detectors [6–8]. The primary motivation has been to develop detection procedures that are relatively insensitive to deviations from the assumed statistical model. For example, nonparametric detectors make no assumptions on the structure of the underlying noise environment; as such, they are intuitively appealing even though they tend to discard information that could aid the detection process.

Despite the many useful techniques developed, the lack of a unifying theoretical basis for designing robust detectors has led to an essentially *ad hoc* approach. The resulting detectors have a fundamental short-coming in that the assumed noise pdf, although initially used in developing the detection procedure, does not reenter directly and unaltered in the information processing algorithm.

An alternative approach to such standard robust methods is one in which the assumed statistical model plays a continuous and direct role in the information processing algorithm. Here, the assumed noise model is used not only for the initial development of the information processing procedure, as in parametric detectors, but also for continuously assessing the degree of agreement between the model and the available data.

In recent work on estimation theory and other statistical methods, such as sensitivity analysis, regression analysis, goodness-of-fit, *etc*, Paulson *et al* [9–11] have proposed so-called self-critical procedures defined as methods which internally and conjointly assess the self-consistency of the data and the assumed underlying structure as a single unit. This assessment of self-consistency is governed by a single parameter called a coefficient of self-criticism. These self-critical procedures represent a generalization of the maximum likelihood estimation process. They are, therefore, a generalization of the likelihood functional and, through it, a generalization of the likelihood ratio.

In this paper the results of self-critical estimation methods are extended to the design of self-critical receivers (SCR), concentrating on self-critical LOR. The resulting structure will suggest one approach to the design of suboptimum and robust detectors. In the next section a brief introduction to self-critical estimation methods is presented to serve as motivation and background for the

* Automation and Control Laboratory, Department of Electrical and Telecommunication Engineering, Ecole Nationale Supérieure Polytechnique, University of Yaoundé I, PO Box 8390, Yaoundé, Cameroon. E-mail: Nnningo@yahoo.com

ISSN 1335-3632

rest of the paper. In the sections thereafter, self-critical detection algorithms are developed.

2 SELF–CRITICAL ESTIMATION METHODS AND THE SELF–CRITICAL LIKELIHOOD FUNCTION

Self-critical procedures were first developed in the field of estimation on the one hand through a process of construction based on a generalization of the maximum likelihood estimation procedure [9–11] and on the other hand through information-theoretic arguments related to maximum likelihood (ML) estimation [10, 11]. Briefly, suppose $x_1, x_2, \dots, x_N$ is a random sample from an absolutely continuous density $f(x \mid \theta)$ that depends on the parameter θ whose value is to be estimated from the sample. Under rather mild regularity conditions on the density $f(x \mid \theta)$, there exists a function $Q(\theta, k)$ defined as follows

$$Q(\theta,k) = \int_{-\infty}^{\infty} f^{1+k}(x \mid \theta)\mathrm{d}x \tag{1}$$

where $k > 0$ is a parameter under operator control. $Q(\theta, k)$ is the information generating function (IGF) [12] of the density $f(x \mid \theta)$. Note that for $k = 0$, $Q(x, \theta) = 1$. Then it is clear that

$$\int_{-\infty}^{\infty} \frac{f^{1+k}(x \mid \theta)}{Q(\theta,k)}\mathrm{d}x = 1\,. \tag{2}$$

Considering the properties of a probability density function, (2) satisfies one of them. Using an approach analogous to that for deriving the maximum likelihood estimation procedure, it is shown in [9] and [11] that from (2) the estimation equation for θ is obtained as

$$\sum_{i=1}^{N} f^k(x_i \,|\, \theta)\Big[(1+k)\frac{\partial \log f(x_i \,|\, \theta)}{\partial\theta} - \frac{\partial \log Q(\theta,k)}{\partial\theta}\Big] = 0 \tag{3}$$

from which it can be noted that for $k \to 0$ the estimators reduce to the standard ML estimator. Further, these estimators are robust for any given value of k. Equation (3) is therefore a generalization of the ML estimation procedure, where the generalization is dependent on the parameter k. The factor $f^k(x)$ is a weighting factor which attributes relative importance to each data point in accordance with its deviation from the assumed model. Since the statistical model enters directly in the estimation process, the method is considered self-critical or model critical.

An alternative interpretation is that $f^k(x)$ is acting as a filter which internally assesses the self-consistency of the data and the underlying model represented by $f(x)$. If the data and model are internally consistent, then the estimated values of the parameter will not vary significantly from each other as k takes on different values. On the other hand, if the data and model lack internal consistency, the estimated values will be significantly sensitive to changes in k as the aberrant data values are filtered out to varying degrees by the factor $f^k(x)$.

It is well-known that an alternative approach to the ML estimation procedure begins with the log-likelihood function [1]. As such, to obtain the log-likelihood function corresponding to (3), the expected value of each summand in (3) must be equal to zero. The solution of the resulting differential equation is the following objective function, called a self-critical log-likelihood function [9], [10] (See Appendix A for the derivation of self-critical procedures).

$$l_k(\mathbf{x} \mid \theta, k) = \frac{1}{k}\sum_{i=1}^{N}\Big[\frac{f^k(x_i \mid \theta)}{Q^a(\theta,k)} - 1\Big] \tag{4}$$

where

$$a = \frac{k}{1+k}\,. \tag{5}$$

The corresponding self-critical likelihood functional obtained from (4) is

$$p_k(\mathbf{x} \mid \theta, k) = \prod_{i=1}^{N}\exp\Big[\frac{1}{k}\Big[\frac{f^k(x_i \mid \theta)}{Q^a(\theta,k)} - 1\Big]\Big]. \tag{6}$$

It has the structure of a joint pdf of N independent and identically distributed (iid) random variables with common univariate density given by

$$p_k(x_i \mid \theta, k) = \exp\Big[\frac{1}{k}\Big[\frac{f^k(x_i \mid \theta)}{Q^a(\theta,k)} - 1\Big]\Big]. \tag{7}$$

It is easily shown (see Appendix B) that as $k \to 0$ (4), (6) and (7) reduce to well-known results for N independent and identically distributed random variables. It must be emphasized that this apparent property of (6) and (7) does not mean that a new density has been described. Rather, it must be viewed as a new method of processing information based on the parameter k and the assumed noise pdf $f(x)$. Equations (4)–(7) are the main results and will be pivotal in the sequel on self-critical detection structures.

3 SELF–CRITICAL RECEIVER STRUCTURES

We consider the following detection problem posed as a binary hypothesis test. Given a set of N observations $\boldsymbol{r}$, we shall assume that, under the null hypothesis H_0, it consists of noise samples only while under the alternative hypothesis H_1, the observations consist of a known signal additively embedded in noise, possibly with unknown non-Gaussian density $f(x)$. We are to choose between the two hypotheses

$$\begin{aligned} H_0 &: \quad \boldsymbol{r} = \boldsymbol{n}\,, \\ H_1 &: \quad \boldsymbol{r} = \theta\boldsymbol{s} + \boldsymbol{n} \end{aligned} \tag{8}$$

where $\boldsymbol{s}$ is a vector of known signal observations, θ is a common signal strength parameter, and $\boldsymbol{n}$ is a vector of noise samples, which are assumed independent and identically distributed (iid).

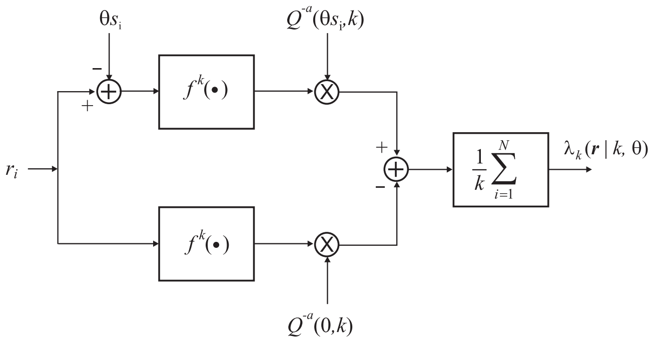

Fig. 1. General self-critical receiver structure for known signals in non-Gaussian noise.

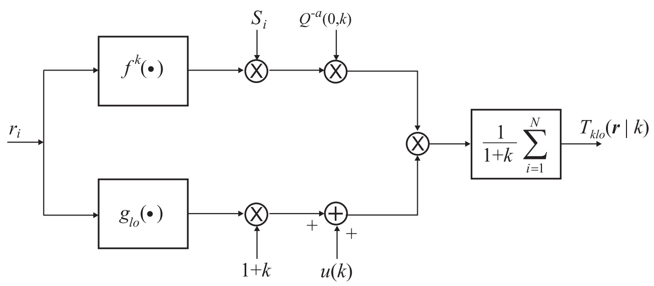

Fig. 2. Self-critical locally optimum receiver structure for known signals in non-Gaussian noise

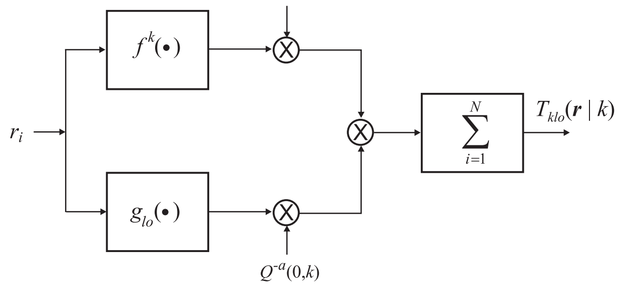

Fig. 3. Self-critical locally optimum receiver structure for known signals in symmetric non-Gaussian noise.

From (4) the self-critical log-likelihood function of the observations under each hypothesis is

$$l_{k|H_0}(\mathbf{r} \mid k) = \frac{1}{k}\sum_{i=1}^{N}\Big[\frac{f^k(r_i)}{Q^a(0,k)} - 1\Big] \tag{9}$$

and

$$l_{k|H_1}(\mathbf{r} \mid \theta, k) = \frac{1}{k}\sum_{i=1}^{N}\Big[\frac{f^k(r_i - \theta s_i)}{Q^a(\theta s_i, k)} - 1\Big]\,. \tag{10}$$

If the detection strategy is based on a threshold test of the self-critical log-likelihood ratio, then the following detection statistic is obtained

$$\lambda_k(\mathbf{r} \mid \theta, k) = l_{k|H_1}(\mathbf{r} \mid \theta, k) - l_{k|H_0}(\mathbf{r} \mid k) > \lambda_0 \tag{11}$$

where $\lambda_k(\mathbf{r} \mid \theta, k)$ is the self-critical log-likelihood ratio and λ_0 is a threshold. Using (9) and (10) in (11) and ignoring the comparison to the threshold, the test statistic is obtained as

$$\lambda_k(\mathbf{r} \mid \theta, k) = \frac{1}{k}\sum_{i=1}^{N}\Big[Q^{-a}(\theta s_i, k) f^k(r_i - \theta s_i) - Q^{-a}(0,k) f^k(r_i)\Big] \tag{12}$$

and is illustrated in Fig. 1. Eq. (12) is the self-critical log-likelihood ratio and reduces to the standard log-likelihood ratio as k approaches zero.

Self-critical locally optimum receiver

The result in (12) is true for all signal strengths. Except for rather simple pdf's, it can be rather mathematically complicated and uneconomical to implement. Because of its critical importance in many applications, attention will be directed to an area of signal detection that has received considerable attention lately and concerns the reception of weak signals [2–5] for which the detectors are termed locally optimum receivers (LOR).

It can be shown [13] that, for the detection problem posed in (8), the LOR is given by the derivative of the likelihood (or log- likelihood) ratio with respect to θ evaluated at $\theta = 0$; that is, the LOR statistic is

$$T_{klo}(\mathbf{r}) = \frac{\partial \lambda_k(\mathbf{r} \mid \theta, k)}{\partial \theta}\Big|_{\theta=0} \tag{13}$$

where $\lambda_k(\mathbf{r} \mid \theta, k)$ is the log-likelihood ratio. Applying this result to (12), the corresponding self-critical LOR statistic is

$$T_{klo}(\mathbf{r} \mid k) = \frac{1}{k+1}\sum_{i=1}^{N} s_i Q^{-a}(0,k) f^k(r_i)[(1+k)g_{lo}(r_i) + u(k)] \tag{14}$$

which can be more compactly written as

$$T_{klo}(\mathbf{r} \mid k) = \frac{1}{1+k}\sum_{i=1}^{N} s_i W(r_i, k)[(1+k)g_{lo}(r_i) + u(k)] \tag{15}$$

where $W(x,k)$ is a weighting factor defined as

$$W(x,k) = Q^{-a}(0,k) f^k(x)\,. \tag{16}$$

The quantities $u(k)$ and $g_{lo}(x)$ in (14) and (15) are defined respectively as

$$u(k) = -\frac{\partial \log Q(\theta s, k)}{\partial \theta}\Big|_{\theta=0} \tag{17}$$

and

$$g_{lo}(x) = -\frac{\partial \log f(x)}{\partial x} \tag{18}$$

is the normal locally optimum zero-memory nonlinearity (ZMNL) [2]. Eq. (14) is illustrated in Fig. 2.

An important case results from considering symmetric noise pdf's. In this case $Q(\theta,k)$ is independent of θ; therefore,

$$\frac{\partial Q(\theta,k)}{\partial\theta}=0 \tag{19}$$

The self-critical LOR statistic becomes

$$T_{klo}(\mathbf{r}\mid k)=\sum_{i=1}^{N}s_iQ^{-a}(0,k)f^k(r_i)g_{lo}(r_i)\,. \tag{20}$$

Eq. (20) is illustrated in Fig. 3.

By comparison the conventional LOR statistic is given by [2]

$$T_{lo}(\mathbf{r})=\sum_{i=1}^{N}s_ig_{lo}(r_i) \tag{21}$$

In the light of (21), the self-critical nature of (20) is now clear. The quantity $f^k(r)$ is acting as a weighting factor and is used to assess the degree of agreement and self-consistency between data and model as represented by $f(x\mid\theta)$. It assigns a weight to each observation depending on the degree of discordance between sample and model. Since the statistical model enters directly in the detection procedure, the method is called self-critical. This detection strategy stands in sharp contrast to conventional schemes, which, lacking the feature of self criticism, attach equal weights to all observations. The parameter k is called the coefficient of self–riticism. As an illustrative example, the linear receiver, with statistic

$$T_l(\mathbf{r})=\sum_{i=1}^{N}s_ir_i \tag{22}$$

makes no further use of the Gaussianity of the noise model and, therefore, attaches equal weight of unity to all observations including those that could clearly be non-Gaussian in nature. Another interpretation of self-critical detectors is that they can be viewed as adaptive receivers where the adaptation is achieved through the factor $f^k(x)$. Unlike other adaptive detector schemes in which the adaptability is ad hoc, the adaptability of self-critical detectors is determined by the assumed noise model.

4 CONCLUSION

A set of robust receivers has been introduced and characterized. The receivers result from a generalization of the likelihood ratio based on the information generating function of the noise density. The detectors are termed self-critical in that, in determining their robustness, the assumed noise model is used continuously in assessing the degree of agreement between the model and the received observations, thereby deemphasizing any samples that deviate significantly from the assumed model. They can also be considered as adaptive receivers. The self-criticism and generalization of the likelihood ratio depend on a parameter under operator control.

APPENDIX A: Derivation of self-critical methods

Self-critical procedures are a generalization of maximum likelihood estimation methods. To demonstrate this generalization, it is necessary to review maximum likelihood estimation. Suppose $x_1,x_2,\dots,x_N$ is a random sample from the probability density function $f(x\mid\theta)$ where θ is a parameter to be estimated. The likelihood function of the sample is given by

$$L(\mathbf{x}\mid\theta)=\prod_{i=1}^{N}f(x_i\mid\theta)\,. \tag{A1}$$

The log-likelihood is obtained from (A1) as

$$l(\mathbf{x}\mid\theta)=\sum_{i=1}^{N}\log f(x_i\mid\theta)\,. \tag{A2}$$

For each x_i, $i=1,2,\dots,N$

$$\int_{-\infty}^{\infty}f(x_i\mid\theta)\mathrm{d}x_i=1\,. \tag{A3}$$

Taking partial derivative of (A3) with respect to θ, the result can be written as

$$\int_{-\infty}^{\infty}\Big\{\frac{\partial\log f(x_i|\theta}{\partial\theta}\Big\}f(x_i\mid\theta)\mathrm{d}x_i=0\,. \tag{A4}$$

Eq. (A4) means that the expected value of the term in brackets $\{\cdot\}$ is zero; that is

$$E\Big\{\frac{\partial\log f(x_i\mid\theta)}{\partial\theta}\Big\}=0\,. \tag{A5}$$

The estimate of θ can be obtained from the following equation, which results either from taking the partial derivative of (A2) with respect to θ and setting the result to zero or from summing the term in brackets in (A4) or (A5) and setting the result to zero.

$$\sum_{i=1}^{N}\Big\{\frac{\partial\log f(x_i\mid\theta)}{\partial\theta}\Big\}=0\,. \tag{A6}$$

The development of (A1) to (A6) provides an approach for the self-critical procedures. Given the probability density function $f(x_i\mid\theta)$, its information generating function (IGF) [12] is defined as

$$Q(\theta,k)=\int_{-\infty}^{\infty}f^{1+k}(x_i\mid\theta)\mathrm{d}x_i\,. \tag{A7}$$

Using (A7), an expression analogous to (A3) can be defined as [9], [10]

$$\int_{-\infty}^{\infty}\frac{f^{1+k}(x_i\mid\theta)}{Q(\theta,k)}\mathrm{d}x_i=1\,. \tag{A8}$$

Following the same line of reasoning as in (A3) to (A6), the self-critical estimation of can be derived from (A8). The self-critical equation corresponding to (A4) is obtained by taking partial derivatives of (A8) with respect to θ and setting to zero; that is

$$\int_{-\infty}^{\infty}\Big\{f^k(x_i \mid \theta)\Big[(1+k)\frac{\partial \log f(x_i \mid \theta)}{\partial\theta} - \frac{\partial \log Q(\theta,k)}{\partial\theta}\Big]\Big\}f(x_i \mid \theta)\mathrm{d}x_i = 0\,. \quad \text{(A9)}$$

Eq. (A9) implies that the expected value of the term in brackets is zero.

$$E\Big\{f^k(x_i \mid \theta)\Big[(1+k)\frac{\partial \log f(x_i \mid \theta)}{\partial\theta} - \frac{\partial \log Q(\theta,k)}{\partial\theta}\Big]\Big\} = 0\,. \quad \text{(A10)}$$

Eq. (A10) is the self-critical analogue of (A5). Returning to the two methods of deriving (A6), one approach of deriving the self-critical estimate of θ is by summing the term in brackets in (A9) or (A10) and setting the result to zero.

$$\sum_{i=1}^{N} f^k(x_i \mid \theta)\Big[(1+k)\frac{\partial \log f(x_i \mid \theta)}{\partial\theta} - \frac{\partial \log Q(\theta,k)}{\partial\theta}\Big] = 0\,. \quad \text{(A11)}$$

Another approach to the derivation of (A11) is through the self-critical log-likelihood function. Accordingly, to complete the development of the self-critical procedures, it remains to determine the self-critical log-likelihood function which will yield (A11) through partial differentiation with respect to θ and setting the result to zero. If (A11) is considered as a set of differential equations, the solution is the following self-critical log-likelihood function [9], [10] in (4)

$$l(\mathbf{x} \mid \theta,k) = \frac{1}{k}\sum_{i=1}^{N}\Big\{\frac{f^k(x_i \mid \theta)}{Q^a(\theta,k)} - 1\Big\} \quad \text{(A12)}$$

where

$$a = \frac{k}{1+k}\,. \quad \text{(A13)}$$

APPENDIX B: Limiting behaviour of self-critical likelihood function

To demonstrate that (4) and (6) reduce to standard results as $k \to 0$, it suffices to show that (7), which is reproduced below, reduces to the pdf $f(x_i \mid \theta)$.

$$p_k(x_i \mid \theta,k) = \exp\Big[\frac{1}{k}\Big[\frac{f^k(x_i \mid \theta)}{Q^a(\theta,k)} - 1\Big]\Big] \quad \text{(B1)}$$

To begin, note that for the information generating function (IGF) $Q(\theta,k)$

$$\lim_{k\to 0} Q(\theta,k) = 1\,. \quad \text{(B2)}$$

The factor $f^k(x_i \mid \theta)$ in (B1) can be rewritten as $\exp\{k \ln f(x_i \mid \theta)\}$ and expanded in a Taylor series as

$$\exp[k \ln f(x_i \mid \theta)] = 1 + k \ln f(x_i \mid \theta) + o(k^2) + \dots \quad \text{(B3)}$$

Using (B2) and (B3) in (B1), we obtain

$$p_k(x_i \mid \theta) = f(x_i \mid \theta)\,. \quad \text{(B4)}$$

as $k \to 0$.

References

[1] van TREES, H. L.: Detection, Estimation, and Modulation Theory, Part I, Wiley, New York, 1968.

[2] MILLER, J. H.—THOMAS, J. B.: Detectors for Discrete-time Signals in non-Gaussian Noise, IEEE Trans. Inform. Theory **IT-18** (Mar. 1972), 241-250.

[3] MODESTINO, J. W.—NINGO, A. Y.: Detection of Weak Signals in Narrowband Non-Gaussian Noise, IEEE Trans. Inform. Theory **IT-25** No. 5 (Sept. 1979), 592–600.

[4] LU, N. H.—EISENSTEIN, B. A.: Detection of Weak Signals in non-Gaussian Noise, IEEE Trans. Inform. Theory **IT-27** (Nov. 1981), 755–772.

[5] NIRENBERG, L. M.: Low SNR Digital Communication over Certain Additive Non-Gaussian Channels, IEEE Trans. Comm. **COM-23** (Mar. 1975), 332-340.

[6] KASSAM, S. A.—POOR, H. V.: Robust Techniques for Signal Processing: A Survey, Proc. IEEE **73** No. 3 (Mar. 1985), 433–481.

[7] GIBBONS, J. D.—MELSA, J. L.: Introduction to Nonparametric Detection with Applications, Academic Press, New York, 1975.

[8] KASSAM, S. A.: A Bibliography on Nonparametric Detection, IEEE Trans. Inform. Theory **IT-26** No. 5 (Sept. 1980), 595-602.

[9] PAULSON, A. S.—PRESSER, M. A.—LAWRENCE, C. E.: Self-critical and Robust Procedures for the Analysis of Univariate Complete Data, Unpublished report, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, 1982, Troy, New York.

[10] DELEHANTY, T. A.: Sensitivity Analysis in Statistics Based on Self-Critical Estimation, Ph.D. Dissertation, Rensselaer Polytechnic Institute, 1983, Troy, New York.

[11] PAULSON, A. S.—DELEHANTY, T. A.: Sensitivity Analysis in Experimental Design, Computer Science and Statistics: Proc. of th 14th. Symposium on the Interface, 1983, pp. 52–57.

[12] GOLOMB, S. W.: The Information Generating Function of a Probability Distribution, IEEE Trans. Inform. Theory **IT-12** No. 1 (Jan. 1966), 75-77.

[13] FERGUSON, T. S.: Mathematical Statistics, Academic Press, New York, 1967.

Received 7 March 2006

Ndeh Ntomambang Ningo (PhD) is senior lecturer in the Department of Electrical and Telecommunications Engineering at Ecole Nationale Supérieure Polytechnique of the University of Yaoundé I in Cameroon. He was born in Cameroon in 1949 and obtained BS and M Eng degrees in Electrical Engineering in 1973 and a PhD in Computer and Systems Engineering in 1982, all from Rensselaer Polytechnic Institute, Troy, New York, USA where he held teaching and research assistantships and was a lecturer. His research interests are in signal detection and computer networks.