Open Access

# Constrained Signals: A General Theory of Information Content and Detection

Mark M. Stecker*

*Joan C. Edwards Marshall University School of Medicine, 1600 Medical Center Blvd, Suite G500, Huntington WV, 25705, USA*

**Abstract:** In this paper, a general theory of signals characterized by probabilistic constraints is developed. As in previous work [10], the theoretical development employs Lagrange multipliers to implement the constraints and the maximum entropy principle to generate the most likely probability distribution function consistent with the constraints. The method of computing the probability distribution functions is similar to that used in computing partition functions in statistical mechanics. Simple cases in which exact analytic solutions for the maximum entropy distribution functions and entropy exist are studied and their implications discussed. The application of this technique to the problem of signal detection is explored both theoretically and with simulations. It is demonstrated that the method can readily classify signals governed by different constraint distributions as long as the mean value of the constraints for the two distributions is different. Classifying signals governed by the constraint distributions that differ in shape but not in mean value is much more difficult. Some solutions to this problem and extensions of the method are discussed.

**Keywords:** Entropy, signal detection, inequality, statistical mechanics, receiver operator characteristics.

## 1. INTRODUCTION

Making a decision as to whether a signal is more likely to fit one set of constraints or another is one of the most important problems in signal processing [1-4]. Many common detection schemes such as the classical North filter [5] or those based upon the Karhunen-Loeve expansion [6] require a knowledge of both the signal and noise distribution functions under the different hypotheses in order to estimate the performance of the detector and determine detection thresholds that will yield a given sensitivity and specificity. However, complete knowledge of both the signal and noise distributions is rarely available. One means of estimating signal and noise distributions based upon a few known characteristics of the signal and noise is the maximum entropy method [7,8]. This method allow for the calculation of the most likely distribution functions subject to certain constraints. These distributions can be used to create detectors with given specificity and sensitivity and to derive detection thresholds and generate criteria for significant changes in a time series [9]. As shown in previous work [10-12], this method yields relatively simple forms for the detection criteria for signals subject to certain equality constraints. It also allows for the derivation of approximate detection criteria for more general equality constraints. As the formal structure of this method is similar to that which underlies all of statistical mechanics [13-15] many of the techniques that have been developed to analyze problems in this field can then be applied to maximum entropy signal detection. One limitation in previous studies was the reliance on equality constraints. In this paper, the approach will be generalized to inequality constraints.

Let $\rho(\{x\})$ be the distribution function describing the probability of observing a set of measured variables: $\{x\}=\{x_1, x_2...x_n\}$. The first purpose of this paper is to use the maximum entropy approach [1,2,10] to compute $\rho(\{x\})$ and the entropy associated with a class of signals that fit a collection of probabilistic constraints. Subsequently, it will be shown how these distribution functions can be used to develop methods for detecting the difference between signals satisfying different constraints. In a previous paper [10], this problem was solved in the case in which all constraints were simple equalities. In the current paper, the more general case in which the constraints are random variables with a known distribution function is discussed and applications developed.

## BASIC THEORY

The simplest binary detection problem involves determining whether a set of measured signal values are more likely to be drawn from a system characterized by hypothesis $h_0$ or hypothesis $h_1$:

$$h_0 : x_i = s_i^0 + n_i^0$$
$$h_1 : x_i = s_i^1 + n_i^1$$

where $s_i^0, n_i^0$ are the signal and noise under hypothesis $h_0$ and $s_i^1, n_i^1$ are the signal and noise values under hypothesis $h_1$. This requires that finding a function $\Lambda(\{x\})$ that can be applied to the a set of recorded signals so that when $\Lambda(\{x\}) > \Theta(\alpha)$ this recorded signal is more likely to come from $h_1$ where $\alpha$ is the probability of a false positive detection. This is only part of what is required for the detection

*Address correspondence to this author at the Joan C. Edwards Marshall University School of Medicine, 1600 Medical Center Blvd, Suite G500, Huntington WV, 25705, USA; Tel: 1-304-526-6302; Fax: 1-304-525-4278; E-mail: mmstecker@gmail.com

1876-8253/11 2011 Bentham Open

problem. It is also necessary to know $\beta$, the probability of a false negative detection as a function of the threshold, the signal characteristics and noise characteristics under each hypothesis. If the probability distribution functions for the measured signals are well known under each hypothesis to be $\xi_0(\{x\})$ and $\xi_1(\{x\})$, one standard choice for $\Lambda(\{x\})$used in Neyman-Pearson decision theory [1,3,5] is $\Lambda(\{x\}) = \ln\left(\frac{\xi_1(\{x\})}{\xi_0(\{x\})}\right)$ although others like the Rao test [1] based upon the first and second derivatives of the probability distribution functions can be constructed. Performing these calculations is a relatively simple matter when the signal is deterministic and the noise distribution functions are known a priori. When there is no a priori knowledge of the distribution or amplitude of the noise, there are three possible approaches. First, some approximations about the signal and noise may be made that allow an estimate of the mean and variance of the noise under each hypothesis [1]. This allows for the construction of an approximate Gaussian probability distribution for the noise and application of the approach described above. Another approach would be to empirically estimate $\xi_0(\{x\})$ and $\xi_1(\{x\})$ (often by matching a known distribution function with a number of variable parameters to the data) and then make multiple measurements in order to define the probability distribution function for the measured data [16-18]. Once these three distributions are known, the Kullback-Leibler mutual entropy can be used to determine whether the measured distribution function is closer to $\xi_0(\{x\})$ or $\xi_1(\{x\})$. The problem is that estimating a complex probability distribution function from data is difficult. The third approach involves constructing maximum entropy distribution functions based upon certain constraints that the measured signal and noise are known to satisfy under each hypothesis. This is the approach to be further developed below.

Before proceeding, it is important to note that the measured variables $x_i$ could be as simple as the amplitude of a signal at a single point in time but they could also represent more general collections of data. Each value might also represent a collection of pixel values in various parts of an image at a given time or, in a system of classical particles, each $x_i$ might contain the position and moment of each particle at a given time. Each collection may also include data from a number of multiple time points. For example, if the signal is described by an autoregressive process of order $p$, then each collection contains the value of the signal at $p$ consecutive time points.

In the previous paper [10], the computation of the entropy of signals subject to constraints of the form:

$$\int \prod_{i=1}^{n} dx_i F_k\left[\{x\}\right] \rho(\{x\}) = C_k; k = 0,1...N \tag{1}$$

was explored. Although this formulation may seem unfamiliar, it becomes clear with a simple example. When $F_k\left[\{x\}\right] = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ (1) becomes a set of constraints on the mean value of the k'th moment of the signal. However, in many cases the constraint values are derived from experimental observations and can only be known to lie in certain ranges. Thus, it is important to study more general constraints with the form:

$$\int \prod_{i=1}^{n} dx_i \mu(\{x\}) F_k\left[\{x\}\right] \rho(\{x\}) = \tilde{C}_k; k = 0...N \tag{2}$$

where:

$F_k\left[\{x\}\right]$ is the k'th constraint function

$\mu(\{x\})$ is a weighting function that is either 0 or 1 depending on whether the measured values $\{x\}$ are a priori possible

$\tilde{C}_k$ is a random variable called the k'th constraint value

If each measurement is a scalar, then $dx_i$ is the differential of the value of the signal at the i'th time point however if the measurements are $q$ dimensional objects, then $dx_i$ is a shorthand notation for $d^q x_i$ expressing the differential in the $q$ dimensional signal space. $\prod_{i=1}^{n} dx_i$ is a product of the differentials of the measured signals at each time point. Thus, $\int \prod_{i=1}^{n} dx_i$ refers to an integration over all possible values of the measured signal at all time points. The weighting function $\mu(\{x\})$ is critical when not all combinations of the values of the measured data are a priori possible. This would occur, for example, in the case where digital data is being sampled and only certain values of the measured signal are possible. It would also occur in the case where each $x_i$ was a collection of measurements, some of which were also contained in $x_{i-1}$. The difference between the terms $\tilde{C}_k$ and $C_k$ is that $\tilde{C}_k$ will be used in this paper whenever the constraint values are random variables and $C_k$ will be used when they are exact values. The *k=0* constraint will always taken as the normalization constraint on the distribution function:

$$F_0\left[\{x\}\right] = 1; \tilde{C}_0 \equiv 1$$

The probability of observing the constraint values $\tilde{C}_k$ $k = 1...N$ is given by a known function $\rho^C(\{\tilde{C}_k\})$. Prior to beginning an analysis of this problem, it is useful to acknowledge the explicit dependence of the distribution function on the constraint values by considering the quantities:

$\rho(\{x\} | \{\tilde{C}_k\})$ – Conditional probability of the signal values given a specific set of constraints

$\rho(\{x\}, \{\tilde{C}_k\})$ – Joint probability of the constraint and signal values

$$\rho(\{x\}, \{\tilde{C}_k\}) = \rho(\{x\} | \{\tilde{C}_k\}) \rho^C(\{\tilde{C}_k\})$$

$$\rho^C(\{\tilde{C}_k\}) = \int \mu(\{x\}) \prod_{i=1}^{n} dx_i \rho(\{x\}, \{\tilde{C}_k\})$$

$\rho^S(\{x\}) = \int \prod_{i=1}^{N} dC_i \rho(\{x\}, \{\tilde{C}_k\})$ - Probability of observing specific signal values

(3)

With this notation, it is possible to rewrite (2) in the form:

$$\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) F_k\left[\{x\}\right]\rho\left(\{x\} \mid \{\tilde{C}_k\}\right) = \tilde{C}_k; k = 0 \ldots N \tag{4}$$

*or*

$$\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) F_k\left[\{x\}\right]\rho\left(\{x\}, \{\tilde{C}_k\}\right) = \tilde{C}_k \rho^C\left(\{\tilde{C}_k\}\right); k = 0 \ldots N$$

The introduction of the weighting function introduces another constraint associated with the fact that the distribution function must be zero at the "prohibited signal values":

$$\mu\left(\{x\}\right) = 0 \rightarrow \rho\left(\{x\}, \{\tilde{C}_k\}\right) = 0$$

This can be implemented by requiring that all distribution functions satisfy:

$$\int\prod_{i=1}^{n} dx_i \left(1 - \mu\left(\{x\}\right)\right)\rho^2\left(\{x\}, \{\tilde{C}_k\}\right) = 0 \tag{5}$$

The maximum entropy method can be used to the estimate the most likely joint distribution function using the Shannon entropy:

$$S = -\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right)\prod_{i=1}^{N} d\tilde{C}_i \rho\left(\{x\}, \{\tilde{C}_k\}\right)\ln\rho\left(\{x\}, \{\tilde{C}_k\}\right) \tag{6}$$

with the constraints:

$$\begin{aligned}
&\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right)\rho\left(\{x\}, \{\tilde{C}_k\}\right) = \rho^C\left(\{\tilde{C}_k\}\right)\\
&\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) F_k\left[\{x\}\right]\rho\left(\{x\}, \{\tilde{C}_k\}\right) = \tilde{C}_k \rho^C\left(\{\tilde{C}_k\}\right); k = 1 \ldots N\\
&\int\prod_{i=1}^{n} dx_i \left(1 - \mu\left(\{x\}\right)\right)\rho^2\left(\{x\}, \{\tilde{C}_k\}\right) = 0
\end{aligned} \tag{7}$$

The first of these constraints simply reproduces the definition of the relationship between the joint distribution function of both signal and constraint values and the marginal distribution of constraint values. The second relates to the constraints on signal values themselves. The third is the prohibition on signal values that are not possible as determined by the weighting function. The constraints (7) must apply for every possible set of constraint values $\{\tilde{C}_k\}$. For this reason, the Lagrange multipliers will be designated $\{\lambda_k\left(\{\tilde{C}_k\}\right)\}$ and $\zeta\left(\{\tilde{C}_k\}\right)$. Using these Lagrange multipliers, the maximum entropy distribution function is that which maximizes:

$$\begin{aligned}
I = &-\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right)\prod_{i=1}^{N} d\tilde{C}_i \rho\left(\{x\}, \{\tilde{C}_k\}\right)\ln\rho\left(\{x\}, \{\tilde{C}_k\}\right) +\\
&\int\prod_{i=1}^{N} d\tilde{C}_i \lambda_0\left(\{\tilde{C}_k\}\right)\left[\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right)\rho\left(\{x\}, \{\tilde{C}_k\}\right) - \rho^C\left(\{\tilde{C}_k\}\right)\right] +\\
&\sum_{k=1}^{N}\int\prod_{i=1}^{N} d\tilde{C}_i \lambda_k\left(\{\tilde{C}_k\}\right)\left[\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) F_k\left[\{x\}\right]\rho\left(\{x\}, \{\tilde{C}_k\}\right) - \tilde{C}_k \rho^C\left(\{\tilde{C}_k\}\right)\right]\\
&\int\prod_{i=1}^{N} d\tilde{C}_i \zeta\left(\{\tilde{C}_k\}\right)\int\prod_{i=1}^{n} dx_i \left(1 - \mu\left(\{x\}\right)\right)\rho^2\left(\{x\}, \{\tilde{C}_k\}\right) = 0
\end{aligned} \tag{8}$$

Thus the distribution function is determined from $\dfrac{\partial I}{\partial\rho\left(\{x\}, \{\tilde{C}_k\}\right)} = 0$ or:

$$\begin{aligned}
0 = &-\mu\left(\{x\}\right)\left[\ln\rho\left(\{x\}, \{\tilde{C}_k\}\right) + 1\right] + \mu\left(\{x\}\right)\lambda_0\left(\{\tilde{C}_k\}\right) + \mu\left(\{x\}\right)\sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right]\\
&+ 2\zeta\left(\{\tilde{C}_k\}\right)\left(1 - \mu\left(\{x\}\right)\right)\rho\left(\{x\}, \{\tilde{C}_k\}\right)
\end{aligned} \tag{9}$$

so that the maximum entropy distribution function is given by:

$$\rho\left(\{x\}, \{\tilde{C}_k\}\right) = \mu\left(\{x\}\right) e^{\lambda_0\left(\{\tilde{C}_k\}\right) + \sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right] - 1} \tag{10}$$

The values of the Lagrange multipliers are chosen so that the constraints resulting from $\dfrac{\partial I}{\partial\lambda_k\left(\{\tilde{C}_k\}\right)} = 0$ are satisfied:

$$\begin{aligned}
&\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) e^{\lambda_0\left(\{\tilde{C}_k\}\right) + \sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right] - 1} = \rho^C\left(\{\tilde{C}_k\}\right)\\
&\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) F_k\left[\{x\}\right] e^{\lambda_0\left(\{\tilde{C}_k\}\right) + \sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right] - 1} = \tilde{C}_k \rho^C\left(\{\tilde{C}_k\}\right); k = 1 \ldots N
\end{aligned} \tag{11}$$

Note that with the distribution function (10), the third constraint in (7) is automatically satisfied and the value of $\zeta$ is not important in determining the distribution function. By taking the ratio of these two expressions in (11), it can be seen that:

$$\frac{\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) F_k\left[\{x\}\right] e^{\sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right]}}{\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) e^{\sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right]}} = \tilde{C}_k; k = 1 \ldots N \tag{12}$$

These relations can be written more compactly by defining:

$$A\left(\left\{\lambda_k\left(\{\tilde{C}_k\}\right)\right\}\right) = \ln\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) e^{\sum_{k=1}^{N}\lambda_k\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right]} \tag{13}$$

from which it follows that the specific values of the Lagrange multipliers satisfying the constraints $\left\{\lambda_k^*\left(\{\tilde{C}_k\}\right)\right\}$ can be determined from:

$$\left.\frac{\partial A\left(\left\{\lambda_k\left(\{\tilde{C}_k\}\right)\right\}\right)}{\partial\lambda_k\left(\{\tilde{C}_k\}\right)}\right|_{\left\{\lambda_k^*\left(\{\tilde{C}_k\}\right)\right\}} = \tilde{C}_k; k = 1 \ldots N \tag{14}$$

It is helpful to define:

$$A^*\left(\{\tilde{C}_k\}\right) = \ln\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) e^{\sum_{k=1}^{N}\lambda_k^*\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right]} \tag{15}$$

Noting that the normalization (*k=0*) constraint can be rewritten as:

$$e^{\lambda_0^*\left(\{\tilde{C}_k\}\right) - 1}\int\prod_{i=1}^{n} dx_i \mu\left(\{x\}\right) e^{\sum_{k=1}^{N}\lambda_k^*\left(\{\tilde{C}_k\}\right) F_k\left[\{x\}\right]} = \rho^C\left(\{\tilde{C}_k\}\right) \tag{16}$$

makes it possible to write:

$$\lambda_0^*\left(\{\tilde{C}_k\}\right) - 1 + A^*\left(\{\tilde{C}_k\}\right) = \ln\rho^C\left(\{\tilde{C}_k\}\right) \tag{17}$$

Substituting this into the definition of the joint distribution function (8) yields:

$$\rho\left(\{x\},\{\tilde{C}_k\}\right)=\rho^C\left(\{\tilde{C}_k\}\right)e^{-A^*\left(\{\tilde{C}_k\}\right)}e^{\sum_{k=1}^{N}\lambda_k^*\left(\{\tilde{C}_k\}\right)F_k[\{x\}]} \tag{18}$$

Although these manipulations may seem to be purely formal, they are important in making the connection between the detection/information content and thermodynamics/statistical mechanics which guides these developments. In particular, $A^*\left(\{\tilde{C}_k\}\right)$ is the analog of the Helmholtz free energy [15].

The entropy of the system (within the maximum entropy approximation) is given by:

$$\begin{aligned}S^* &= -\int\prod_{i=1}^{n}dx_i\mu(\{x\})\prod_{i=1}^{N}d\tilde{C}_i\rho\left(\{x\},\{\tilde{C}_k\}\right)\ln\rho\left(\{x\},\{\tilde{C}_k\}\right)\\ &= -\int\prod_{i=1}^{N}d\tilde{C}_i\rho^C\left(\{\tilde{C}_k\}\right)\ln\rho^C\left(\{\tilde{C}_k\}\right)+\int\prod_{i=1}^{N}d\tilde{C}_i\rho^C\left(\{\tilde{C}_k\}\right)A^*\left(\{\tilde{C}_k\}\right)-\sum_{k=1}^{N}\int\prod_{i=1}^{N}d\tilde{C}_k\rho^C\left(\{\tilde{C}_k\}\right)\lambda_k^*\left(\{\tilde{C}_k\}\right)\tilde{C}_k\end{aligned} \tag{19}$$

This expression can be simplified considerably by noting that the entropy of signals associated with a specific set of constraint values $\{C_k\}$ is:

$$\begin{aligned}S^*\left(\{C_k\}\right) &= -\int\prod_{i=1}^{n}dx_i\mu(\{x\})\rho\left(\{x\}|\{C_k\}\right)\ln\rho\left(\{x\}|\{C_k\}\right)\\ &= A^*\left(\{\tilde{C}_k\}\right)-\sum_{k=1}^{N}\lambda_k^*\left(\{\tilde{C}_k\}\right)\tilde{C}_k\end{aligned} \tag{20}$$

so that (19) can be rewritten as:

$$S^*=-\int\prod_{i=1}^{N}d\tilde{C}_i\rho^C\left(\{\tilde{C}_k\}\right)\ln\rho^C\left(\{\tilde{C}_k\}\right)+\int\prod_{i=1}^{N}d\tilde{C}_i\rho^C\left(\{\tilde{C}_k\}\right)S^*\left(\{\tilde{C}_k\}\right) \tag{21}$$

As might be expected, the first term is the entropy associated with the uncertainty in the value of the constraints. The second term is the weighted average of the entropy associated with any single constraint value. This expression is critical because it allows for an immediate connection with the results obtained for equality constraints in the previous paper [10].

## APPLICATION: ENTROPY CHANGE DUE TO UNCERTAINTIES IN CONSTRAINT VALUES

Within the theoretical framework developed above, it is possible and useful to address the change in entropy (and corresponding loss of information content) when the constraint values are not known precisely. Consider the case in which the constraint values $\{\tilde{C}_k\}$ are known to be close to the values $\{C_k^0\}$ and are independently Gaussian distributed around $\{C_k^0\}$:

$$\rho^C\left(\{\tilde{C}_k\}\right)=\prod_{i=1}^{N}\left[\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{\left(\tilde{C}_k-C_k^0\right)^2}{2\sigma_i^2}}\right] \tag{22}$$

If the constraint values do not vary much from the $\{C_k^0\}$, a Taylor series expansion can be used to write:

$$S^*\left(\{\tilde{C}_k\}\right)\approx S^*\left(\{\tilde{C}_k^0\}\right)+\sum_{k=1}^{N}\left.\frac{\partial S^*}{\partial C_k}\right|_{\{\tilde{C}_k^0\}}\left(\tilde{C}_k-C_k^0\right)+\frac{1}{2}\sum_{kk'=1}^{N}\left.\frac{\partial^2 S^*}{\partial C_k\partial C_k}\right|_{\{\tilde{C}_k^0\}}\left(\tilde{C}_k-C_k^0\right)\left(\tilde{C}_j-C_j^0\right) \tag{23}$$

Substituting this into (21) results in the expression:

$$\begin{aligned}S^* &= -\int\prod_{i=1}^{N}d\tilde{C}_i\rho^C\left(\{\tilde{C}_k\}\right)\ln\rho^C\left(\{\tilde{C}_k\}\right)+\int\prod_{i=1}^{N}d\tilde{C}_i\rho^C\left(\{\tilde{C}_k\}\right)S^*\left(\{\tilde{C}_k\}\right)\\ &= S_0+\frac{1}{2}+\sum_{k=1}^{N}\ln\sqrt{2\pi\sigma_k^2}+\frac{1}{2}\sum_{k=1}^{N}\left.\frac{\partial^2 S^*}{\partial C_k^2}\right|_{\{\tilde{C}_k^0\}}\sigma_k^2\end{aligned} \tag{24}$$

In [4] it was shown, using a multidimensional cumulant expansion, that if:

$$\begin{aligned}\left\langle F_i\left[\{x\}\right]\right\rangle_0 &= \int\prod_{i=1}^{n}dx_iF_i\left[\{x\}\right]\rho\left(\{x\}|\{C_k^0\}\right)\\ M_{ii'} &= \left\langle F_i\left[\{x\}\right]F_{i'}\left[\{x\}\right]\right\rangle_0-\left\langle F_i\left[\{x\}\right]\right\rangle_0\left\langle F_{i'}\left[\{x\}\right]\right\rangle_0\end{aligned} \tag{25}$$

then:

$$\frac{\partial^2 S^*}{\partial C_k\partial C_{k'}}\approx -M_{kk'}^{-1} \tag{26}$$

Important insights can be obtained by considering the special case where the matrix $M_{kk'}$ is diagonal. In this case, since each $M_{ii}$ must be positive by the Cauchy-Schwartz inequality, it is clear that $S^*$ is an increasing function of the $\{\sigma_k\}$ for small values (due to the first factor in (21)) and eventually declines as the $\{\sigma_k\}$ increase further. If no information about the uncertainties in the constraint values were available, then the best guess for those variations would be that which is associated with the largest value of the entropy $S^*$. Some simple algebra demonstrates that this occurs when:

$$\sigma_k=\sqrt{M_{kk}} \tag{27}$$

The associated maximum value of the entropy is:

$$S_{\max}^*=S_0^*+N\ln\sqrt{2\pi}+\frac{1}{2}\sum_{i=1}^{N}\ln M_{kk} \tag{28}$$

This is an important point because it indicates that**,** if no additional information is provided, it is most likely that the constraints are NOT constants but stochastic variables. This indicates the importance of understanding stochastic constraints of the form (2) in addition to the equality constraints of (1). It also points out the relationship between the second moments of the constraint functions and the likely variability in the constraint values. This is a familiar relationship in thermodynamics [14.15].

## SIGNAL DETECTION THEORY

One important application of the theory developed above is in signal detection. Consideration will be given to the case in which a decision must be made as to whether a measured signal belongs to the class of signals associated with one of two different constraint distributions. Specifically, a decision must be made as to whether hypothesis

***H*₀** or hypothesis ***H*₁** is the best descriptor of the test signal:

$$\boldsymbol{H_0} : \rho_0^C\left(\left\{\tilde{C}_k\right\}\right) \text{ is the distribution of constraints} \qquad (29)$$
$$\boldsymbol{H_1} : \rho_1^C\left(\left\{\tilde{C}_k\right\}\right) \text{ is the distribution of constraints}$$

In order to make the decision as to which is hypothesis is most likely based upon a set of observed data, a test statistic is needed that is both simple to calculate and takes on very different values depending on which hypothesis is true. One such statistic is the following:

$$\Lambda\left(\{x\}\right) = \int \prod_{k=1}^{N} d\tilde{C}_k d\tilde{C}_k{}' \rho_1^C\left(\left\{\tilde{C}_k\right\}\right)\rho_0^C\left(\left\{\tilde{C}_k{}'\right\}\right)\ln\left(\frac{\rho_1\left(\{x\},\left\{\tilde{C}_i\right\}\right)}{\rho_0\left(\{x\},\left\{\tilde{C}_i{}'\right\}\right)}\right) \qquad (30)$$

where $\rho_1\left(\{x\},\left\{\tilde{C}_i\right\}\right)$ is the joint probability distribution under $H_1$ and $\rho_0\left(\{x\},\left\{\tilde{C}_i\right\}\right)$ is the joint probability distribution under $H_0$, This statistic has the property that if $\boldsymbol{H_1}$ is more likely than $\boldsymbol{H_0}$, averaged over the possible sets of constraints, then $\Lambda\left(\{x\}\right) > 0$. Using (18) and (20) it is possible to simplify (30) to

$$\begin{aligned}\Lambda\left(\{x\}\right) &= S_0^* - S_1^* \\ &+ \int \prod_{i=1}^{N} d\tilde{C}_i \prod_{i=1}^{N} d\tilde{C}_i{}' \rho_1^C\left(\left\{\tilde{C}_k\right\}\right)\rho_0^C\left(\left\{\tilde{C}_k{}'\right\}\right)\left[\sum_{k=1}^{N}\lambda_k^*\left(\left\{\tilde{C}_k\right\}\right)\left(F_k\left[\{x\}\right]-\tilde{C}_k\right)-\sum_{k=1}^{N}\lambda_k^*\left(\left\{\tilde{C}_k{}'\right\}\right)\left(F_k\left[\{x\}\right]-\tilde{C}_k{}'\right)\right]\end{aligned} \qquad (31)$$

where $S_0^*$ is the entropy under hypothesis $\boldsymbol{H_0}$ and $S_1^*$ is the entropy under hypothesis $\boldsymbol{H_1}$. The first term in this expression demonstrates the importance of the entropy difference between the signals that fit the two hypotheses in the detection problem. In order to determine detection criteria, it is necessary to estimate both the expectation value and the variance of this test statistic under the two hypotheses. As a first step, note that:

$$\begin{aligned}\int \prod_{i=1}^{n} dx_i \rho_0^S\left(\{x\}\right)F_k\left[\{x\}\right] &= \int \prod_{i=1}^{n} d\tilde{C}_i \rho_0^C\left(\left\{\tilde{C}_k\right\}\right)\tilde{C}_k = \overline{C}_{k|0} \\ \int \prod_{i=1}^{n} dx_i \rho_1^S\left(\{x\}\right)F_k\left[\{x\}\right] &= \int \prod_{i=1}^{n} d\tilde{C}_i \rho_1^C\left(\left\{\tilde{C}_k\right\}\right)\tilde{C}_k = \overline{C}_{k|1}\end{aligned} \qquad (32)$$

where $\overline{C}_{k|0}, \overline{C}_{k|1}$ are the expectation values of the constraint values under the two hypotheses. The expressions (32) can be used to write:

$$\begin{aligned}\left\langle \Lambda\left(\{x\}\right)\right\rangle_{(0,1)} &= \int \prod_{i=1}^{n} dx_i \rho_{(0,1)}^S\left(\{x\}\right)\Lambda\left(\{x\}\right) \\ &= S_0^* - S_1^* + \int \prod_{i=1}^{N} d\tilde{C}_i \prod_{i=1}^{N} d\tilde{C}_i{}' \rho_1^C\left(\left\{\tilde{C}_k\right\}\right)\rho_0^C\left(\left\{\tilde{C}_k{}'\right\}\right)\left[\sum_{k=1}^{N}\lambda_k^*\left(\left\{\tilde{C}_k\right\}\right)\left(\overline{C}_{k|(0,1)}-\tilde{C}_k\right)-\sum_{k=1}^{N}\lambda_k^*\left(\left\{\tilde{C}_k{}'\right\}\right)\left(\overline{C}_{k|(0,1)}-\tilde{C}_k{}'\right)\right] \\ &= S_0^* - S_1^* + \left[\sum_{k=1}^{N}\overline{\lambda}_{k|1}\left(\overline{C}_{k|(0,1)}-\overline{\overline{C}}_{k|1}\right)-\sum_{k=1}^{N}\overline{\lambda}_{k|0}\left(\overline{C}_{k|(0,1)}-\overline{\overline{C}}_{k|0}\right)\right]\end{aligned} \qquad (33)$$

Where:

$$\begin{aligned}\overline{\lambda}_{k|l} &= \int \prod_{i=1}^{N} d\tilde{C}_i \rho_l^C\left(\left\{\tilde{C}_k\right\}\right)\lambda_k^*\left(\left\{\tilde{C}_k\right\}\right) \\ \overline{\overline{C}}_{k|l} &= \frac{\int \prod_{i=1}^{N} d\tilde{C}_i \rho_l^C\left(\left\{\tilde{C}_k\right\}\right)\lambda_k^*\left(\left\{\tilde{C}_k\right\}\right)\tilde{C}_k}{\int \prod_{i=1}^{N} d\tilde{C}_i \rho_l^C\left(\left\{\tilde{C}_k\right\}\right)\lambda_k^*\left(\left\{\tilde{C}_k\right\}\right)}\end{aligned} \qquad (34)$$

The notation here is that whenever a parenthesis appears in a subscript it indicates a choice between different hypotheses. For example $\left\langle \Lambda\left(\{x\}\right)\right\rangle_{(0,1)}$ would refer to the average of the test statistic under either hypothesis 0 or hypothesis 1. Also, as indicated above, when a vertical slash appears in a subscript, it indicates which hypothesis is used to compute the relevant average.

Equations (34) and (35) imply that:

$$\left\langle \Lambda\left(\{x\}\right)\right\rangle_0 - \left\langle \Lambda\left(\{x\}\right)\right\rangle_1 = \sum_{k=1}^{N}\left(\overline{\lambda}_{k|1}-\overline{\lambda}_{k|0}\right)\left(\overline{C}_{k|0}-\overline{C}_{k|1}\right) \qquad (35)$$

As in the case of equality constraints [4], the difference in the averaged test statistic in the two hypotheses can be considered as the output of a simple filter. If the difference in the average values of the Lagrange multipliers is considered one vector and the difference in the average constraint values another, then (35) is the scalar product of these two vectors. The structure of the detection statistic (35) is the same as that used in the classical matched filter detection algorithms and is hence a generalization of these techniques [3]. The structure of (35) indicates that the ability of the test statistic to classify signals is proportional to the difference in the Lagrange multipliers associated with each class of signals. Signals are best detected by the differences in mean values of constraint functions that are associated with the largest difference in Lagrange multipliers. This is very analogous to the situation in statistical mechanics where the inverse temperature (which is essentially the Lagrange multiplier that appears as a result of the constraint of total energy) sets the scale for which energy differences will be associated with significant changes in the probability distribution function [14]. A somewhat similar idea has been used in the maximum entropy approach to statistical inference [19,20] and in the use of moments to distinguish between different time series [21].

In order to determine the receiver operating characteristics of this detector, it is important to compute the variance of the test statistic under the different hypotheses. Some calculations will show that:

$$\Lambda\left(\{x\}\right) = S_0 - S_1 + \left[\sum_{k=1}^{N}\overline{\lambda}_{k|1}\left(F_k\left[\{x\}\right]-\overline{\overline{C}}_{k|1}\right)-\sum_{k=1}^{N}\overline{\lambda}_{k|0}\left(F_k\left[\{x\}\right]-\overline{\overline{C}}_{k|0}\right)\right]$$

$$\begin{aligned}\Lambda\left(\{x\}\right)-\left\langle \Lambda\left(\{x\}\right)\right\rangle_{(0,1)} &= \sum_{k=1}^{N}\left(\overline{\lambda}_{k,1}^*-\overline{\lambda}_{k,0}^*\right)\left(F_k\left[\{x\}\right]-\overline{C}_{k|(0,1)}\right) \\ \sigma_{(0,1)}^2 = \left\langle\left(\Lambda\left(\{x\}\right)-\left\langle \Lambda\left(\{x\}\right)\right\rangle_{(0,1)}\right)^2\right\rangle_{(0,1)} &= \sum_{\substack{k=1\\k'=1}}^{N}\left(\overline{\lambda}_{k|1}^*-\overline{\lambda}_{k|0}^*\right)\left(\overline{\lambda}_{k'|1}^*-\overline{\lambda}_{k'|0}^*\right)M_{kk'}^{(0,1)} \\ M_{kk'}^{(0,1)} &= \left\langle\left(F_k\left[\{x\}\right]-\overline{C}_{k|(0,1)}\right)\left(F_{k'}\left[\{x\}\right]-\overline{C}_{k'|(0,1)}\right)\right\rangle_{(0,1)}\end{aligned} \qquad (36)$$

Within the context of Neymann-Pearson decision theory, $\boldsymbol{H_1}$ will be chosen if:

$$\Lambda\left(\{x\}\right) > \Theta \qquad (37)$$

for some threshold $\Theta$. In order to choose the optimal threshold, it is necessary to estimate the probability distribution functions of the test statistic under each hypothesis. The simplest assumption is that this distribution of the test statistic is approximately Gaussian:

$$p_0(\Lambda) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\Lambda - \langle\Lambda\rangle_0)^2}{2\sigma_0^2}}$$
$$p_1(\Lambda) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(\Lambda - \langle\Lambda\rangle_1)^2}{2\sigma_1^2}} \quad (38)$$

We now need the false positive rate as a function of the threshold $\Theta$.

$$\alpha = \int_{\Theta}^{\infty} p_0(\Lambda')\, d\Lambda' \quad (39)$$

It is also important to calculate the false negative probability, $\beta$. This is the probability that the signal is classified as coming from a distribution characterized by the $\boldsymbol{H_0}$ constraints when it truly comes from $\boldsymbol{H_1}$.

$$\beta = \int_{-\infty}^{\Theta} p_1(\Lambda')\, d\Lambda' \quad (40)$$

Now, the value of the threshold can be computed as a function of the desired false positive rate:

$$\alpha = \int_{\Theta}^{\infty} p_0(\Lambda')\, d\Lambda' = \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_{\Theta - \langle\Lambda\rangle_0}^{\infty} e^{-\frac{(\Lambda')^2}{2\sigma_0^2}} d\Lambda'$$
$$= \frac{1}{2} erfc\left( \frac{\Theta - \langle\Lambda\rangle_0}{\sqrt{2\sigma_0^2}} \right) \quad (41)$$

and

$$\beta = \int_{-\infty}^{\Theta} p_1(\Lambda')\, d\Lambda' = \frac{1}{\sqrt{2\pi\sigma_1^2}} \int_{-\infty}^{\Theta - \langle\Lambda\rangle_0} e^{-\frac{(\Lambda' - \upsilon)^2}{2\sigma_1^2}} d\Lambda'$$
$$= 1 - \frac{1}{2} erfc\left( \frac{\Theta - \langle\Lambda\rangle_0 - \upsilon}{\sqrt{2\sigma_1^2}} \right) \quad (42)$$

where:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$$
$$\upsilon = \langle\Lambda\rangle_1 - \langle\Lambda\rangle_0 = -\sum_{k=1}^{N} \left(\bar{\lambda}_{k|1} - \bar{\lambda}_{k|0}\right)\left(\bar{C}_{k|0} - \bar{C}_{k|1}\right) \quad (43)$$

Now,

$$\Theta = \sqrt{2\sigma_0^2}\, erfc^{-1}(2\alpha) + \langle\Lambda\rangle_0 \quad (44)$$

and so:

$$\beta = 1 - \frac{1}{2} erfc\left( \frac{\sqrt{2\sigma_0^2}\, erfc^{-1}(2\alpha) - \upsilon}{\sqrt{2\sigma_1^2}} \right)$$
$$\beta = 1 - \frac{1}{2} erfc\left( \frac{\sigma_0}{\sigma_1} erfc^{-1}(2\alpha) - \frac{\upsilon}{\sqrt{2\sigma_1^2}} \right) \quad (45)$$

This is the expression that determines the receiver operating characteristics of the detector.

Greater insight into the detector described above can be obtained by considering the situation in which there is a single constraint other than the normalization constraint (i.e. *N=1*):

$$\upsilon = \left\langle \Lambda(\{x\}) \right\rangle_1 - \left\langle \Lambda(\{x\}) \right\rangle_0 = \left(\bar{\lambda}_{1|1} - \bar{\lambda}_{1|0}\right)\left(\bar{C}_{1|1} - \bar{C}_{1|0}\right)$$
$$\sigma_{0,1}^2 = \left(\bar{\lambda}_{1|1} - \bar{\lambda}_{1|0}\right)^2 M_{11}^{(0,1)}$$
$$\frac{\upsilon}{\sqrt{2\sigma_1^2}} = \frac{\left(\bar{C}_{1|1} - \bar{C}_{1|0}\right)}{\sqrt{2M_{11}^{(1)}}} \operatorname{sgn}\left(\bar{\lambda}_{1|1} - \bar{\lambda}_{1|0}\right) \quad (46)$$
$$\frac{\sigma_0}{\sigma_1} = \sqrt{\frac{M_{11}^{(0)}}{M_{11}^{(1)}}}$$

This formula again points out the role of the variance in the constraint functions as the effective noise level in the system under consideration as was suggested earlier. It is also clear that the sensitivity of this detector is mainly determined by the difference in the mean value of the constraints under the two conditions only to a lesser extent the detailed structure of the distribution function for the constraint values.

Appendices A and B describe the application of the principles developed above to the case in which the constraint functions are quadratic functions of the signal values. In addition to illustrating the mathematical analysis, it demonstrates the critical importance of the ratio of the difference between the mean constraint values in the two different conditions to the square root of the variance of the signal in determining detectability. It does also illustrate the point evident in (46) that although the detection statistic (30) is sensitive to the difference in the shapes of the distributions under the two hypotheses, it is more sensitive to differences in the mean constraint values in the different distributions because the latter contribution scales with the square root of the number of data points.

## DISCUSSION

There are many ways to compute the information content of a signal and create schemes to detect specific signals. The specific advantage of using the constrained signal approach is that it can be applied even when very little information is known about the signals. This is particularly important when dealing with biological signals where it is difficult to distinguish signal from “noise” and to define many details of the signal and noise characteristics. However, the human eye can distinguish certain patterns such as a seizure on an electroencephalogram quite easily even when they are low in amplitude compared to other types of brain electrical activity. It is possible that this process used by the eye/brain is better described by a constraint type analysis since it relies only on a few specific signal details of interest rather than a complex analysis of many signal characteristics. This is the basis of the perceptron detection algorithms [22,23]. Additional studies would need to be performed in order to answer these questions.

In any case, the general theory of constrained signals provides important means for computing the information content or entropy in a class of signals. It allows for the solution of some problems whose solution would be difficult by other means. One example is the computation of the effect that specifying a signal's power and bispectrum has on entropy [11]. Another example is the detection of signals that have been distorted by processing through a non-linear filter [11]. This method also subsumes other more general problems in signal detection and does lead, in the field of signal detection, to a generalization of the correlation/matched filter detector.

There remain some difficulties using the maximum entropy approach. The first difficulty is that it is only possible to compute the function $A(\{\lambda_k\})$ analytically for a few types of constraint some of which have been discussed in this and previous papers [10,11]. However, this does not limit the applicability of the technique. The similarity between $A(\{\lambda_k\})$ and the Helmholtz free energy from statistical mechanics makes it possible to directly apply techniques already developed in statistical mechanics over the last 100 years to problems in signal processing and detection. In particular, there are a larger number of exactly solvable models from statistical mechanics that could be useful in studying certain signal types. Even in the case where exact solutions are not available, useful approximation methods involving perturbation expansions and the asymptotic expansion are available [10]. Other approximations developed originally from the field of statistical mechanics such as the mean field approximation [11] or the Kirkwood superposition approximation [24] have been shown useful in computing the entropy of a signal. When analytic methods cannot be used, there are well known methods in statistical mechanics such as the Metropolis algorithm [25] for numerical computation of partition functions that could be adapted for use in the analysis of constrained signals.

This paper has developed a formalism to distinguish signals constrained by different probabilistic constraints. The classification scheme presented in this paper using the test statistic (30) is only one possible means of classifying signals and using other test statistics may be desirable under certain circumstances. Detection methods based upon the simple test statistic (30) readily distinguish between signals when the mean value of the constraints is different in the various classes. However, they perform less well in classifying signals associated with the same mean constraint values that have different shapes of the constraint distribution function. This is a difficult problem but it is likely that there are solutions within the maximum entropy approach outlined above. First, it is possible that breaking the observed signal into subsignals and testing each of the subsignals may provide a better overall classification. Second, multiple test statistics can be used which place different weights on different observed signal values $\{x\}$. A third possibility would be to apply a non-linear filter to the signal prior to implementing the detector. In particular, if two signals differ in the degree of variance around a mean value, simple threshold detections based upon the signal value will be problematic. Transforming the signal into the square of the distance between its value and the mean value and using a threshold detection on this transformed signal will result in improved detectability.

## APPENDIX A. BASIC THEORY OF GENERAL QUADRATIC CONSTRAINTS

The general theory of detection for inequality constraints discussed in the main text simplifies dramatically when the constraint functions are quadratic functions of the signal values:

$$F_k\left[\{x\}\right] = \sum_{j=1}^{n} D_j^{(k)} x_j + \sum_{ij=1}^{n} E_{ij}^{(k)} x_i x_j; i,\ j = 1...n; k = 1...N \tag{A1}$$

where the $D^{(k)}$ are known constant vectors and $E^{(k)}$ are known constant matrices. In this case:

$$A(\{C_k\}) = \ln \int \prod_{i=1}^{n} dx_i e^{\sum_{k=1}^{N} \lambda_k (\{C_k\}) F_k\left[\{x\}\right]} \tag{A2}$$

This can be simplified by writing:

$$\sum_{k=1}^{N} \lambda_k\ (\{C_k\}) F_k\left[\{x\}\right] = \sum_{j=1}^{n} D_j x_j + \sum_{ij=1}^{n} E_{ij} x_i x_j \tag{A3}$$

where:

$$\begin{aligned} D_j &= \sum_{k=1}^{N} \lambda_k\ (\{C_k\}) D_j^{(k)} \\ E_{ij} &= -\sum_{k=1}^{N} \lambda_k\ (\{C_k\}) E_{ij}^{(k)} \end{aligned} \tag{A4}$$

It is then possible to write:

$$A(\{C_k\}) = \ln \int e^{-\sum_{ij=1}^{n} E_{ij} x_i x_j + \sum_{i=1}^{N} D_i x_i} d^n x = \ln\left[\sqrt{\frac{\pi^n}{\det E}} e^{\frac{1}{4} D^T E^{-1} D}\right] = \frac{n}{2} \ln \pi - \frac{1}{2} \ln \det E + \frac{1}{4} D^T E^{-1} D \tag{A5}$$

This is true only when the integral converges. One criterion for this is $\det E > 0$, another criterion is that if E is symmetric, all of its eigenvalues must be positive. The Lagrange multipliers are then determined by:

$$\left.\frac{\partial A\left(\left\{\lambda_k\left(C_k\right)\right\}\right)}{\partial \lambda_k}\right|_{\left\{\lambda_k^*\left(C_k\right)\right\}} = C_k ; k = 1 \ldots N \tag{A6}$$

And the entropy for a single set of constraints is determined from:

$$S^*\left(\left\{C_k\right\}\right) = A^*\left(\left\{C_k\right\}\right) - \sum_{k=1}^{N} \lambda_k^*\left(\left\{C_k\right\}\right) C_k \tag{A7}$$

In the special case where E is a diagonal matrix, the above equations simplify greatly. In that case:

$$\begin{aligned} A\left(\left\{C_k\right\}\right) &= \frac{n}{2}\ln\pi - \frac{1}{2}\sum_{i=1}^{n}\ln\left(E_{ii}\right) + \frac{1}{4}\sum_{i=1}^{n}\frac{D_i^2}{E_{ii}} \\ \frac{\partial A}{\partial \lambda_k} &= \sum_{i=1}^{n}\frac{dE_{ii}}{d\lambda_k}\frac{\partial A}{\partial E_{ii}} + \sum_{i=1}^{n}\frac{dD_i}{d\lambda_k}\frac{\partial A}{\partial D_i} = -\frac{1}{2}\sum_{i=1}^{n}\frac{d\ln\left(E_{ii}\right)}{dE_{ii}}\frac{dE_{ii}}{d\lambda_k} - \frac{1}{4}\sum_{i=1}^{n}\frac{D_i^2}{E_{ii}^2}\frac{dE_{ii}}{d\lambda_k} + \frac{1}{2}\sum_{i=1}^{n}\frac{D_i}{E_{ii}}\frac{dD_i}{d\lambda_k} \\ &= \frac{1}{2}\left[\sum_i E_{ii}^{(k)}\left(\frac{1}{E_{ii}} + \frac{1}{2}\frac{D_i^2}{E_{ii}^2}\right) + \sum_{i=1}\frac{D_i}{E_{ii}}D_i^{(k)}\right] \end{aligned} \tag{A8}$$

The equations that determine the values of the Lagrange multipliers are then:

$$\frac{1}{2}\left[\sum_i E_{ii}^{(k)}\left(\frac{1}{E_{ii}} + \frac{1}{2}\frac{D_i^2}{E_{ii}^2}\right) + \sum_{i=1}\frac{D_i}{E_{ii}}D_i^{(k)}\right] = C_k \tag{A9}$$

Multiplying both sides of (A9) by $\lambda_k^*$ and summing over k yields :

$$\frac{1}{2}\left[-n + \frac{1}{2}\sum_{i=1}\frac{D_i^2}{E_{ii}}\right] = \sum_k C_k \lambda_k^* \tag{A10}$$

so that it is possible to write:

$$\begin{aligned} A^*\left(\left\{C_k\right\}\right) &= \frac{n}{2}\ln\pi - \frac{1}{2}\sum_{i=1}\ln\left(E_{ii}\right) + \frac{n}{2} + \sum_k C_k\lambda_k^* \\ S^*\left(\left\{C_k\right\}\right) &= \frac{n}{2}\ln e\pi - \frac{1}{2}\sum_{i=1}\ln\left(E_{ii}\right) \end{aligned} \tag{A11}$$

Although this expression for the entropy is simple, the problem is that it is in general difficult to solve (A9) analytically to obtain the value of the Lagrange multipliers. However, in the special case where:

$$E_{ij}^{(k)} = 0; k \geq 2$$

$$D_i^{(1)} = 0$$

it is possible to find a simple solution. In this case:

$$\begin{aligned} \left.\frac{\partial A}{\partial \lambda_k}\right|_{\left\{\lambda_k^*\right\}} &= \frac{1}{2}\left[\sum_{i=1}^{n} E_{ii}^{(k)}\left(\frac{1}{E_{ii}} + \frac{1}{2}\frac{D_i^2}{E_{ii}^2}\right) + \sum_{i=1}^{n}\frac{D_i}{E_{ii}}D_i^{(k)}\right] \\ \left.\frac{\partial A}{\partial \lambda_1}\right|_{\left\{\lambda_k^*\right\}} &= \frac{1}{2}\left[\sum_{i=1}^{n} E_{ii}^{(1)}\left(\frac{1}{E_{ii}} + \frac{1}{2}\frac{D_i^2}{E_{ii}^2}\right)\right] = \frac{1}{2}\left[-\frac{n}{\lambda_1^*} + \frac{n}{2\lambda_1^{*2}}\sum_{kk'=2}^{N}\lambda_k^*\lambda_{k'}^* F_{kk'}\right] \\ \left.\frac{\partial A}{\partial \lambda_k}\right|_{\left\{\lambda_k^*\right\}} &= -\frac{1}{2}\left[\frac{n}{\lambda_1}\sum_{k'=2}^{N}\lambda_{k'}^* F_{kk'}\right]; k \geq 2 \end{aligned} \tag{A12}$$

where:

$$F_{kk'} = \frac{1}{n}\sum_{i=1}^{n}\frac{D_i^{(k)}D_i^{(k')}}{E_{ii}^{(1)}};k,k'=2...N \tag{A13}$$

Defining the inverse matrix to *F*, $F^{-1}$, if it exists:

$$\sum_{l=2}^{N} F_{kl}F_{lk'}^{-1} = \delta_{kk'};k,k'=2...N$$

enables a solution for all of the Lagrange multipliers $k \geq 2$ in terms of $\lambda_1$:

$$-\frac{n}{2\lambda_1}\sum_{k'=2}^{N} F_{kk'}\lambda_{k'}^{*} = C_k;k\geq 2 \Rightarrow \lambda_k^{*} = -2\frac{\lambda_1^{*}}{n}\sum_{k'=2}^{N} F_{kk'}^{-1}C_{k'} \tag{A14}$$

with $\lambda_1^{*}$ determined from:

$$\frac{1}{2}\left[-\left(\frac{n}{\lambda_1^{*}} - \frac{2}{n}\sum_{\substack{k=2\\k'=2}}^{N} C_k C_{k'} F_{kk'}^{-1}\right)\right] = C_1 \tag{A15}$$

Thus, the explicit solutions for the Lagrange multipliers are:

$$\lambda_1^{*} = \frac{1}{2}\frac{n}{\frac{1}{n}\sum_{\substack{k=2\\k'=2}}^{N} C_k C_{k'} F_{kk'}^{-1} - C_1}$$

$$\lambda_k^{*} = -\frac{n}{\frac{1}{n}\sum_{\substack{k=2\\k'=2}}^{N} C_k C_{k'} F_{kk'}^{-1} - C_1}\frac{1}{n}\sum_{k'=2} F_{kk'}^{-1}C_{k'} \tag{A16}$$

This allows the entropy in the case of equality constraints to be written as:

$$S^{*}\left(\{C_k\}\right) = \frac{n}{2}\ln e\pi - \frac{1}{2}\sum_{i=1}^{n}\ln\left(E_{ii}^{(1)}\right) - \frac{n}{2}\ln\left(-\frac{1}{2}\frac{n}{\frac{1}{n}\sum_{\substack{k=2\\k'=2}}^{N} C_k C_{k'} F_{kk'}^{-1} - C_1}\right) \tag{A17}$$

and for a general distribution of constraints:

$$S^{*} = \frac{n}{2}\ln e\pi - \frac{1}{2}\sum_{i=1}^{n}\ln\left(E_{ii}^{(1)}\right) - \frac{n}{2}\int\prod_{l=1}^{N} d\tilde{C}_l\,\rho\left(\{\tilde{C}_k\}\right)\ln\left(-\frac{1}{2}\frac{n}{\frac{1}{n}\sum_{\substack{k=2\\k'=2}} \tilde{C}_k \tilde{C}_{k'} F_{kk'}^{-1} - \tilde{C}_1}\right) \tag{A18}$$

Note that if $E_{ii}^{(1)} > 0$ then $\lambda_1^{*}$ must be negative. Hence any set of possible constraints must satisfy $\frac{1}{n}\sum_{\substack{k=2\\k'=2}}^{N} C_k C_{k'} F_{kk'}^{-1} < C_1$.

Under the above assumptions, it is also possible to compute the values of $\upsilon, \sigma_{(0,1)}^{2}$ required to evaluate the performance of detectors based upon the test statistic (30). Appendix B shows the detailed but straightforward analytic calculations in a simple case in which there is one linear and one quadratic constraint. This is accomplished both in the case of uniform distributions and discrete distributions. This case is particularly enlightening because it demonstrates the expected effect that the difference in mean values of the constraint under the two hypotheses has on the sensitivity and specificity of the detection algorithm. As expected and shown in Fig. (**1**), the larger the difference in the mean values in the two distributions, the better the test statistic (30) can discriminate between signals. The efficacy of the detector is a function of the variable $y = \frac{C_2}{\sqrt{C_1 n F_{22}}}$ which, as defined in Appendix B, is a normalized ratio of the *k=2* constraint value to the square root of the *k=1* constraint value (and is thus a "signal to noise ratio"). It is important to note that both $C_1$ and $C_2$ will increase linearly with the number of sample points in the signal so that y will be independent of the number of sample points in the signal. Fig. (**2**) shows the effect of increasing the number of data points, n, on the value of $\beta$ when the detection threshold is set to keep the value of $\alpha$ constant. This figure demonstrates that the number of false negatives decreases quickly as n increases as expected. On the other hand, the ability of the test statistic (30) to classify a signal as belonging to one or another class of signals characterized by different shapes but the same mean values of the constraint values is much poorer. Fig. (**3**) shows the ROC curve obtained using the test statistic to determine whether a signal comes from one of two distributions with the same mean constraint values. Distribution A has the constraint values all near the value of 0.5 and distribution B has half of the constraint values near 0.1 and half near 0.9. There are two reasons for this lack of power in the detection of constraint distribution shape. The first is that the ratio of the two variances of the test statistic under the two hypotheses which occurs in (45) is one of the main channels through which changes in the constraint shape affect the ROC curves. However, this ratio cannot depend on n while the factor $\frac{\upsilon}{\sqrt{2\sigma_1^2}}$ which determines how the changes in the mean value of the test statistic affect the ROC curve increase with $\sqrt{n}$ as demonstrated in Appendix B. Thus, detectability of the changes in mean values improves quickly with n while the detectability of changes based upon shape changes does not. In addition, Figs. (**4** and **5**) demonstrate that the entropy of the constrained signals changes much more rapidly with changes in the distribution mean value than with changes in the width of the distribution.

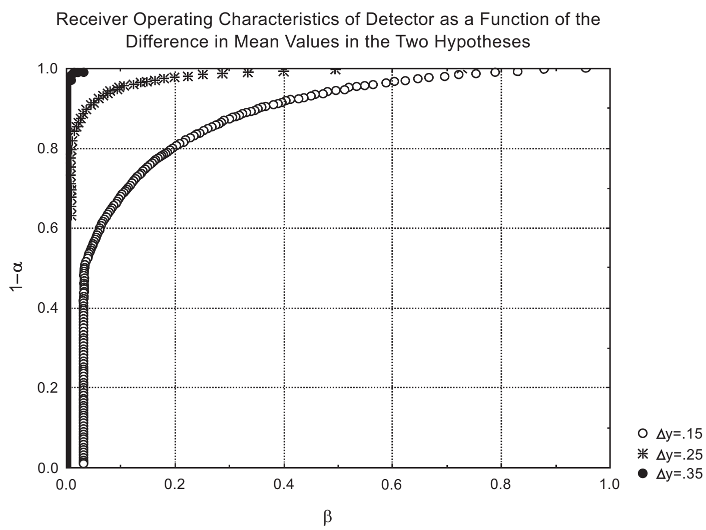



**Fig. (1).** Receiver Operating Characteristics of the detector based upon the test statistic (30) for signals with a constant quadratic constraint and a linear constraint (measured by the normalized variable y) that is different between the two hypotheses. Δy is a measure of the difference in the linear constraint values between the two hypotheses.

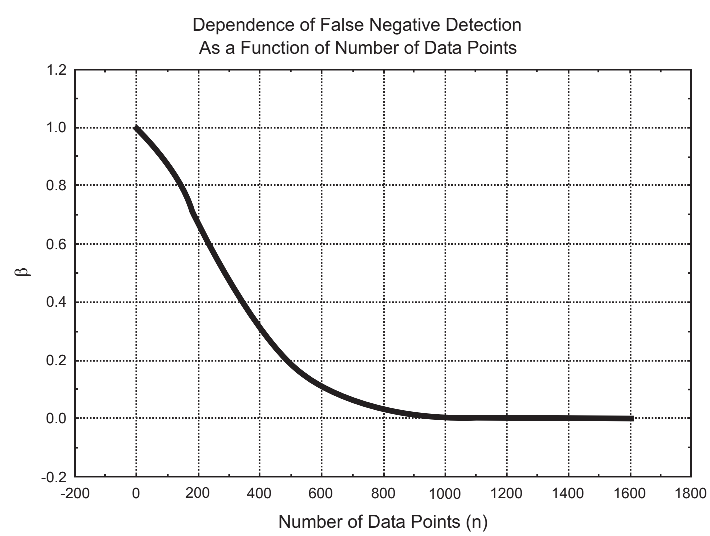



**Fig. (2).** As a function of the number of points in the signal, n, the detection threshold is set to maintain a constant probability of 0.01 of false positive detections. The probability of false negative detections declines rapidly with increasing number of data points. Δy=0.1.

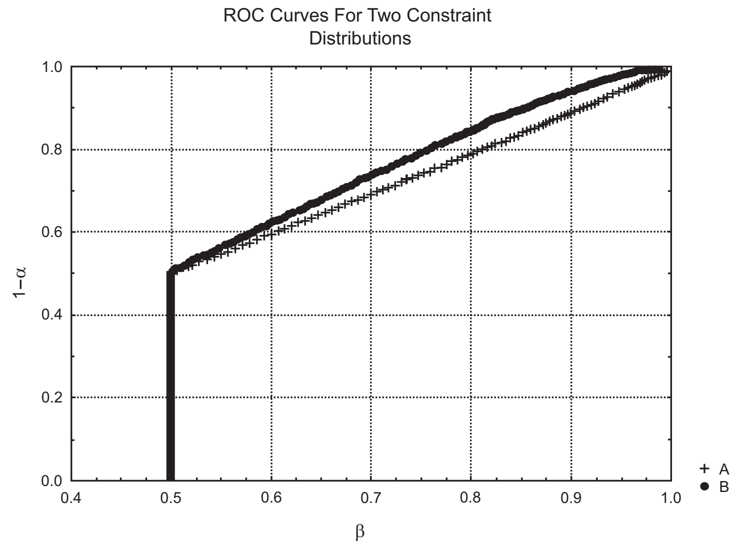



**Fig. (3).** ROC curve where under the two hypotheses, the constraint values have the same mean but the different shapes shown as A and B. It can be seen that the ability to reliably detect a signal under these circumstances is low.

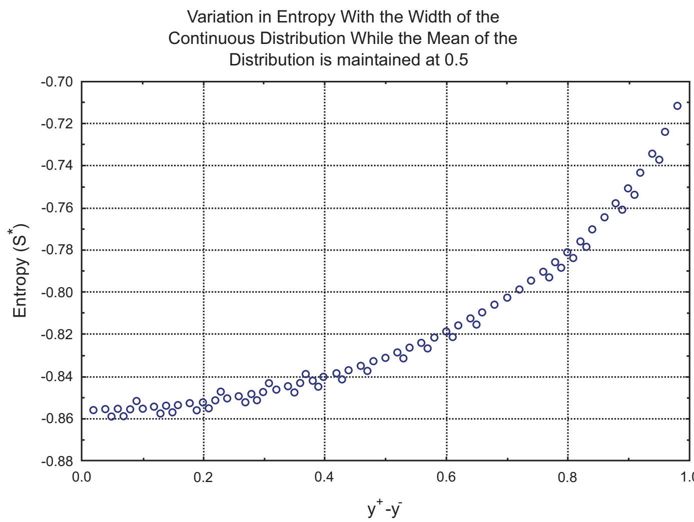



**Fig. (4).** The change in entropy when a uniform distribution of constraint values between $y^+$ and $y^-$ is chosen and the mean of the distribution remains at y=0.5. This demonstrates the minimal effects of the constraint distribution width on the entropy.

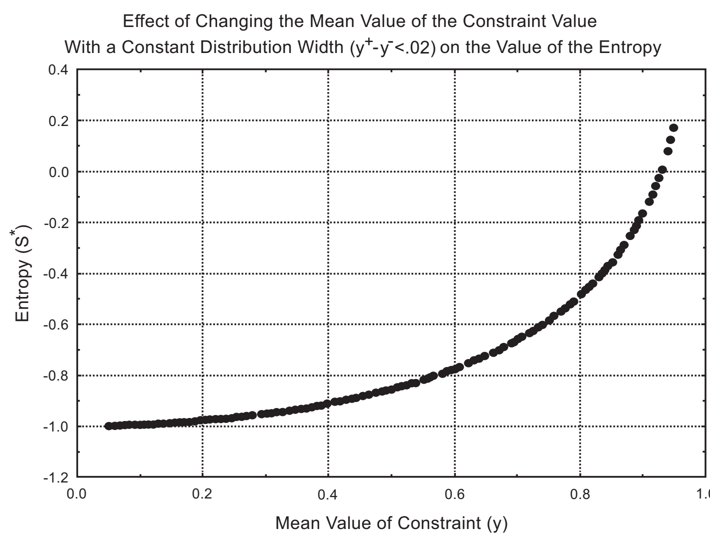



**Fig. (5).** In comparision to Fig. **4**, this figure shows the larger effect that changes in the mean value of the constraint has when the width of the distribution is held constant.

**APPENDIX B: COMPUTATION OF THE DETECTION TEST STATISTIC IN THE CASE OF SIMPLE QUADRATIC CONSTRAINTS.**

The goal of this appendix is to illustrate the computation of the detection test statistic parameters:

$$\upsilon = -\sum_{k=1}^{N}\left(\bar{\lambda}_{k|1}-\bar{\lambda}_{k|0}\right)\left(\bar{C}_{k|0}-\bar{C}_{k|1}\right)$$
$$\sigma^2_{(0,1)} = \sum_{\substack{k=1\\k'=1}}^{N}\left(\bar{\lambda}_{k|1}-\bar{\lambda}_{k|0}\right)\left(\bar{\lambda}_{k'|1}-\bar{\lambda}_{k'|0}\right)M^{(0,1)}_{kk'} \quad \text{(B1)}$$

in the very simple case where there is one linear and one quadratic constraint and the linear constraint value is differently distributed under the two hypotheses $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$:

$$F_1\left[\{x\}\right] = \sum_{ij=1}^{n} e_{ij}x_i x_j = C_1$$
$$F_2\left[\{x\}\right] = \sum_{j=1}^{n} d_j x_j = \tilde{C}_2 \quad \text{(B2)}$$

where:

$$\mathsf{H}_0 : \rho_0\left(\tilde{C}_2\right) = \begin{Bmatrix} \dfrac{1}{\mu_0^+ - \mu_0^-}; \mu_0^- \le \tilde{C}_2 \le \mu_0^+ \\ 0; Otherwise \end{Bmatrix}$$
$$\mathsf{H}_1 : \rho_1\left(\tilde{C}_2\right) = \begin{Bmatrix} \dfrac{1}{\mu_1^+ - \mu_1^-}; \mu_1^- \le \tilde{C}_2 \le \mu_1^+ \\ 0; Otherwise \end{Bmatrix} \quad \text{(B3)}$$

Since the value of $C_1$ is the same under both hypotheses, (B1) simplifies to:

$$\upsilon = -\left(\bar{\lambda}_{2|1}-\bar{\lambda}_{2|0}\right)\left(\bar{C}_{2|0}-\bar{C}_{2|1}\right); \quad \left(\bar{C}_{2|0}-\bar{C}_{2|1}\right) = \frac{\mu_0^+ + \mu_0^-}{2} - \frac{\mu_1^+ + \mu_1^-}{2}$$
$$\sigma^2_{(0,1)} = \left(\bar{\lambda}_{1|1}-\bar{\lambda}_{1|0}\right)^2 M^{(0,1)}_{1,1} + 2\left(\bar{\lambda}_{2|1}-\bar{\lambda}_{2|0}\right)\left(\bar{\lambda}_{1|1}-\bar{\lambda}_{1|0}\right)M^{(0,1)}_{1,2} + \left(\bar{\lambda}_{2|1}-\bar{\lambda}_{2|0}\right)^2 M^{(0,1)}_{2,2} \quad \text{(B4)}$$

The first step in evaluating these quantities is computing the mean value of the Lagrange multipliers under each hypothesis:

$$\bar{\lambda}_{1|0} = \frac{n}{2}\int \rho_0\left(C_2\right)\frac{1}{\frac{C_2^2}{nF}-C_1}; \bar{\lambda}_{1|1} = \frac{n}{2}\int \rho_1\left(C_2\right)\frac{1}{\frac{C_2^2}{nF}-C_1}$$
$$\bar{\lambda}_{2|0} = -\int \rho_0\left(C_2\right)\frac{n}{\frac{C_2^2}{nF}-C_1}\frac{C_2}{nF}; \bar{\lambda}_{2|1} = -\int \rho_1\left(C_2\right)\frac{n}{\frac{C_2^2}{nF}-C_1}\frac{C_2}{nF} \quad \text{(B5)}$$

where:

$$F = \frac{1}{n}\sum_{i=1}^{n}\frac{d_i^2}{e_{ii}} \quad \text{(B6)}$$

Carrying out the integrations:

$$\bar{\lambda}_{1|0} = \frac{n}{2}\frac{1}{\mu_0^+ - \mu_0^-}\int_{\mu_0^-}^{\mu_0^+} dC_2 \frac{1}{\frac{C_2^2}{nF} - C_1}; x = \frac{C_2}{\sqrt{nF}}; c = \sqrt{C_1}$$

$$= -\frac{n}{2}\frac{\sqrt{nF}}{\mu_0^+ - \mu_0^-}\int_{\frac{\mu_0^-}{\sqrt{nF}}}^{\frac{\mu_0^+}{\sqrt{nF}}} dx \frac{1}{c^2 - x^2} = -\frac{n}{4c}\frac{1}{\frac{\mu_0^+}{\sqrt{nF}} - \frac{\mu_0^-}{\sqrt{nF}}}\left\{\ln\left[\frac{c + \frac{\mu_0^+}{\sqrt{nF}}}{c - \frac{\mu_0^+}{\sqrt{nF}}}\right] - \ln\left[\frac{c + \frac{\mu_0^-}{\sqrt{nF}}}{c - \frac{\mu_0^-}{\sqrt{nF}}}\right]\right\} \tag{B7}$$

$$\bar{\lambda}_{2|0} = \int \rho_0(C_2)\frac{n}{\frac{C_2^2}{nF} - C_1}\frac{C_2}{nF} = -\frac{n}{(\mu_0^+ - \mu_0^-)\sqrt{nF}}\int_{\mu_0^-}^{\mu_0^+} dC_2 \frac{1}{\frac{C_2^2}{nF} - C_1}\frac{C_2}{\sqrt{nF}}$$

$$= \frac{n}{(\mu_0^+ - \mu_0^-)}\int_{\frac{\mu_0^-}{\sqrt{nF}}}^{\frac{\mu_0^+}{\sqrt{nF}}} dx \frac{x}{c^2 - x^2} = -\frac{n}{2(\mu_0^+ - \mu_0^-)}\ln\left[\frac{c^2 - \left(\frac{\mu_0^+}{\sqrt{nF}}\right)^2}{c^2 - \left(\frac{\mu_0^-}{\sqrt{nF}}\right)^2}\right]$$

or in terms of the "dimensionless" variables:

$$y_0^+ = \frac{\mu_0^+}{c\sqrt{nF}}; y_0^- = \frac{\mu_0^-}{c\sqrt{nF}}$$

$$\bar{\lambda}_{1|0} = -\frac{n}{4c^2}\frac{1}{y_0^+ - y_0^-}\left\{\ln\left[\frac{1 + y_0^+}{1 - y_0^+}\right] - \ln\left[\frac{1 + y_0^-}{1 - y_0^-}\right]\right\} \tag{B8}$$

$$\bar{\lambda}_{2|0} = -\frac{n}{2(y_0^+ - y_0^-)c\sqrt{nF}}\ln\left[\frac{1 - (y_0^+)^2}{1 - (y_0^-)^2}\right]$$

and so:

$$\upsilon = n\left[\frac{y_0^+ + y_0^-}{2} - \frac{y_1^+ + y_1^-}{2}\right]\left[\frac{1}{2(y_1^+ - y_1^-)}\ln\left[\frac{1 - (y_1^+)^2}{1 - (y_1^-)^2}\right] - \frac{1}{2(y_0^+ - y_0^-)}\ln\left[\frac{1 - (y_0^+)^2}{1 - (y_0^-)^2}\right]\right] \tag{B9}$$

In order to compute the variance of the test statistic, it is critical to know the value of the expectations of the second moments of the constraint functions:

$$M_{kk'}^{(0,1)} = \left\langle \left(F_k\left[\{x\}\right] - \bar{C}_{k|(0,1)}\right)\left(F_{k'}\left[\{x\}\right] - \bar{C}_{k'|(0,1)}\right)\right\rangle_{(0,1)} \tag{B10}$$

and in terms of the function A:

$$M_{kk'}^{(0,1)} = \int \prod_{i=1}^{N} d\tilde{C}_i \rho_{(0,1)}\left(\{\tilde{C}_k\}\right)\frac{\partial^2 A}{\partial\lambda_k \partial\lambda_{k'}}\bigg|_{\{\lambda_k^*\}} \tag{B11}$$

Since:

$$\frac{\partial^2 A}{\partial\lambda_1 \partial\lambda_1} = -\frac{C_1}{\lambda_1} - \frac{1}{n\lambda_1}\sum_{\substack{l=2\\l'=2}}^{N} C_l C_{l'} F_{ll'}^{-1}$$

$$\frac{\partial^2 A}{\partial\lambda_k \partial\lambda_1} = -\frac{C_k}{\lambda_1} \tag{B12}$$

$$\frac{\partial^2 A}{\partial\lambda_k \partial\lambda_{k'}} = -\frac{n}{2\lambda_1}F_{kk'}$$

(B11) can be rewritten as:

$$
\begin{aligned}
M_{11}^{(0,1)} &= \frac{C_1^2}{y_{(0,1)}^+ - y_{(0,1)}^-}\frac{2}{n}\left[\left(\left(y_{(0,1)}^+ - y_{(0,1)}^-\right) - \frac{1}{5}\left({y_{(0,1)}^+}^5 - {y_{(0,1)}^-}^5\right)\right)\right] \\
M_{12}^{(0,1)} &= -\frac{C_1 c\sqrt{nF}}{y_{(0,1)}^+ - y_{(0,1)}^-}\frac{2}{n}\left[\left(\frac{1}{4}\left({y_{(0,1)}^+}^4 - {y_{(0,1)}^-}^4\right) - \frac{1}{2}\left({y_{(0,1)}^+}^2 - {y_{(0,1)}^-}^2\right)\right)\right] \\
M_{22}^{(0,1)} &= -\frac{C_1 F}{y_{(0,1)}^+ - y_{(0,1)}^-}\left[\left(\frac{1}{3}\left({y_{(0,1)}^+}^3 - {y_{(0,1)}^-}^3\right) - \left(y_{(0,1)}^+ - y_{(0,1)}^-\right)\right)\right]
\end{aligned}
\tag{B13}
$$

and so:

$$\sigma_{(0,1)}^2 = n f_{(0,1)} \tag{B14}$$

where:

$$
\begin{aligned}
g\left(y_1^+, y_1^-\right) &= \frac{1}{y_1^+ - y_1^-}\left\{\ln\left[\frac{1+y_1^+}{1-y_1^+}\right] - \ln\left[\frac{1+y_1^-}{1-y_1^-}\right]\right\} \\
h\left(y_1^+, y_1^-\right) &= \frac{1}{\left(y_1^+ - y_1^-\right)}\ln\left[\frac{1-\left(y_1^+\right)^2}{1-\left(y_1^-\right)^2}\right] \\
f_{(0,1)} &= \frac{1}{8}\frac{1}{y_{(0,1)}^+ - y_{(0,1)}^-}\left[\left[\left(\left(y_{(0,1)}^+ - y_{(0,1)}^-\right) - \frac{1}{5}\left({y_{(0,1)}^+}^5 - {y_{(0,1)}^-}^5\right)\right)\right]\right]\left[g\left(y_1^+, y_1^-\right) - g\left(y_0^+, y_0^-\right)\right]^2 \\
&- \frac{1}{2}\frac{1}{y_{(0,1)}^+ - y_{(0,1)}^-}\left[\left(\frac{1}{4}\left({y_{(0,1)}^+}^4 - {y_{(0,1)}^-}^4\right) - \frac{1}{2}\left({y_{(0,1)}^+}^2 - {y_{(0,1)}^-}^2\right)\right)\right]\left[g\left(y_1^+, y_1^-\right) - g\left(y_0^+, y_0^-\right)\right]\left[h\left(y_1^+, y_1^-\right) - h\left(y_0^+, y_0^-\right)\right] \\
&\frac{1}{y_{(0,1)}^+ - y_{(0,1)}^-}\left(\frac{1}{4}\right)\left[\left(\left(y_{(0,1)}^+ - y_{(0,1)}^-\right) - \frac{1}{3}\left({y_{(0,1)}^+}^3 - {y_{(0,1)}^-}^3\right)\right)\right]\left[h\left(y_1^+, y_1^-\right) - h\left(y_0^+, y_0^-\right)\right]^2
\end{aligned}
\tag{B15}
$$

The quantities that determine the receiver operating characteristics are:

$$
\frac{\upsilon}{\sqrt{2\sigma_1^2}} = \frac{\sqrt{n}\left[\frac{y_0^+ + y_0^-}{2} - \frac{y_1^+ + y_1^-}{2}\right]\left[\frac{1}{2\left(y_1^+ - y_1^-\right)}\ln\left[\frac{1-\left(y_1^+\right)^2}{1-\left(y_1^-\right)^2}\right] - \frac{1}{2\left(y_0^+ - y_0^-\right)}\ln\left[\frac{1-\left(y_0^+\right)^2}{1-\left(y_0^-\right)^2}\right]\right]}{\sqrt{2 f_1}} \tag{B16}
$$

$$\frac{\sigma_0}{\sigma_1} = \sqrt{\frac{f_0}{f_1}}$$

It is important to note, as expected, that the detectability for small changes increases with the square root of the number of data points in the signal since the quantities y should have only minimal dependence on n. Fig. (**1**) shows the receiver operating characteristics as a function of the difference between the mean values of y under the two hypotheses. Fig. (**2**) shows the changes in false positive rate as a function of n. This illustrates the significant influence of this factor on detectability. However, the ability to detect differences based upon the distribution is very poor. In fact, with this simple model and the uniform distribution functions, it is impossible to use the test statistic (30) to discriminate between two uniform distributions with equal means and different variance. This is in part due to the fact that the variances in both hypotheses scale similarly with n and hence their ratio which is what determines the ROC curves is independent of *n*.

In order to probe the limits of the ability of the proposed test statistic to detect signals constrained by different constraint distributions but the same mean values. Consider the simple case in which the hypotheses about the constraint functions are

$$
\begin{aligned}
H_0 &: \rho_0\left(\tilde{C}_2\right) = \delta\left(\tilde{C}_2 - \frac{1}{2}c\sqrt{nF}\right) \\
H_1 &: \rho_1\left(\tilde{C}_2\right) = \frac{1}{2}\delta\left(\tilde{C}_2 - (1-\eta)c\sqrt{nF}\right) + \frac{1}{2}\delta\left(\tilde{C}_2 - \eta c\sqrt{nF}\right)
\end{aligned}
\tag{B17}
$$

for some value $0 \le \eta < 1$. Both of these distributions have the same mean value and hence $\upsilon = 0$ but they have very different shapes. It is clear that the Lagrange multipliers have the

$$\bar{\lambda}_{1|0} = \frac{n}{2c^2}\left[\frac{1}{\frac{1}{4}-1}\right] = -\frac{2n}{3c^2};\bar{\lambda}_{1|1} = \frac{n}{4c^2}\left[\frac{1}{(1-\eta)^2-1}+\frac{1}{\eta^2-1}\right]$$
$$\bar{\lambda}_{2|0} = -\frac{n}{c\sqrt{nF}}\left[\frac{\frac{1}{2}}{\frac{1}{4}-1}\right] = \frac{2n}{3c\sqrt{nF}} =;\bar{\lambda}_{2|1} = -\frac{n}{2c\sqrt{nF}}\left[\frac{(1-\eta)}{(1-\eta)^2-1}+\frac{\eta}{\eta^2-1}\right] \quad \text{(B18)}$$

Now, the expressions for the variance are:

$$M_{11}^{(0)} = \frac{30C_1^2}{16n}$$
$$M_{11}^{(1)} = -\frac{C_1^2}{n}\left[(\eta^4-1)+((1-\eta)^4-1)\right]$$
$$M_{12}^{(0)} = \frac{6C_1c}{8n}\sqrt{nF}$$
$$M_{12}^{(1)} = -\frac{C_1c}{n}\sqrt{nF}\left[(\eta^3-\eta)+((1-\eta)^3-(1-\eta))\right]$$
$$M_{22}^{(0)} = \frac{3C_1F}{4}$$
$$M_{22}^{(1)} = -\frac{C_1F}{2}\left[(\eta^2-1)+((1-\eta)^2-1)\right] \quad \text{(B19)}$$

and so the expressions for the variance of the test statistic under each hypothesis are:

$$\frac{\sigma_0^2}{n} = \frac{30}{256}\left[\frac{1}{(1-\eta)^2-1}+\frac{1}{\eta^2-1}+\frac{8}{3}\right]^2 - \frac{6}{32}\left[\frac{1}{(1-\eta)^2-1}+\frac{1}{\eta^2-1}+\frac{8}{3}\right]\left[\frac{(1-\eta)}{(1-\eta)^2-1}+\frac{\eta}{\eta^2-1}+\frac{4}{3}\right]$$
$$+\frac{3}{16}\left[\frac{(1-\eta)}{(1-\eta)^2-1}+\frac{\eta}{\eta^2-1}+\frac{4}{3}\right]^2 \quad \text{(B20)}$$

$$\frac{\sigma_1^2}{n} = -\frac{1}{16}\left[(\eta^4-1)+((1-\eta)^4-1)\right]\left[\frac{1}{(1-\eta)^2-1}+\frac{1}{\eta^2-1}+\frac{8}{3}\right]^2$$
$$\frac{1}{4}\left[(\eta^3-\eta)+((1-\eta)^3-(1-\eta))\right]\left[\frac{1}{(1-\eta)^2-1}+\frac{1}{\eta^2-1}+\frac{8}{3}\right]\left[\frac{(1-\eta)}{(1-\eta)^2-1}+\frac{\eta}{\eta^2-1}+\frac{4}{3}\right]$$
$$-\frac{1}{8}\left[(\eta^2-1)+((1-\eta)^2-1)\right]\left[\frac{(1-\eta)}{(1-\eta)^2-1}+\frac{\eta}{\eta^2-1}+\frac{4}{3}\right]^2$$

Fig. (**3**) shows an example of the receiver operating characteristics of the test statistic when $\eta = 0.9$. It can be seen that even in this extreme case, the discrimination of the test statistic is poor when applied to a single signal. Since this result is not strongly dependent on the number of data points analyzed, it is likely that breaking the signal into subsignals and application of the test statistic to each will improve the ability to detect differences in the distributions.

In a previous paper [4], detectors of the class discussed in this paper were called "entropy detectors". With this in mind it is important to understand how the entropy changes with the shape and mean value of the constraint distribution. In the case of the uniform distribution discussed above, the entropy is given by:

$$S^* = \frac{n}{2}\ln e\pi - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\frac{ne_{ii}}{2C_1}\right) - \frac{n}{2(y^+-y^-)}\int_{y^-}^{y^+}dy\ln\left(\frac{1}{1-y^2}\right)$$
$$S^* = \frac{n}{2}\ln\left(\frac{\pi}{e}\right) - \frac{1}{2}\sum_{i=1}^{n}\ln\left(\frac{ne_{ii}}{2C_1}\right)$$
$$-\frac{n\left[-(1-y^-)\ln(1-y^-)+(1-y^+)\ln(1-y^+)-(1+y^+)\ln(1+y^+)+(1+y^-)\ln(1+y^-)\right]}{2(y^+-y^-)} \quad \text{(B21)}$$

Evaluating this function, as shown in Figs. (**4** and **5**), the variation in the entropy when the mean value of the distribution $\frac{y^{+}+y^{-}}{2}$ changes (Fig. **4**) is more than ten times greater than the variation in entropy when the width of the distribution $y^{+}-y^{-}$ is varied and the mean value is kept constant (Fig. **5**). This parallels the poor ability to detect signal with different distribution shapes.

Studying the ability of this method to detect simulated signals sheds more light on its advantages and disadvantages. The goal of these simulation studies is to determine whether a test signal is better described by one or another set of constraints (***H₀*** and ***H₁***). In each simulation the constraint functions will be those of equation (B2) with the simplified form:

$$e_{ij}=\begin{Bmatrix} 1; i=j \\ 0; i\neq j \end{Bmatrix}$$
$$d_j=1$$

In the first set of simulations, the mean square of the test signals and the comparison signals is $\frac{1}{n}\sum_{i=1}^{n}x_i^2=\frac{C_1}{n}=1$; however, the distribution of the mean signal value $\frac{1}{n}\sum_{j=1}^{n}x_j=\frac{C_2}{n}$ is different for the test signal, and the two comparison signals (described by the distributions $\rho_0\left(C_2\right)$ and $\rho_1\left(C_2\right)$).

$$\begin{aligned}
&\mathrm{H}_0:\rho_0\left(\tilde{C}_2\right)=\begin{Bmatrix} \frac{1}{0.4}; -0.3\leq\frac{\tilde{C}_2}{n}\leq 0.1 \\ 0; Otherwise \end{Bmatrix} (\text{mean}= -0.1)\\
&\mathrm{H}_1:\rho_1\left(\tilde{C}_2\right)=\begin{Bmatrix} \frac{1}{0.4}; 0.1\leq\frac{\tilde{C}_2}{n}\leq 0.5 \\ 0; Otherwise \end{Bmatrix} (\text{mean}= 0.3)\\
&Test:\rho_t\left(\tilde{C}_2\right)=\begin{Bmatrix} \frac{1}{2\Delta}; \mu-\Delta\leq\frac{\tilde{C}_2}{n}\leq \mu+\Delta \\ 0; Otherwise \end{Bmatrix}
\end{aligned} \tag{B22}$$

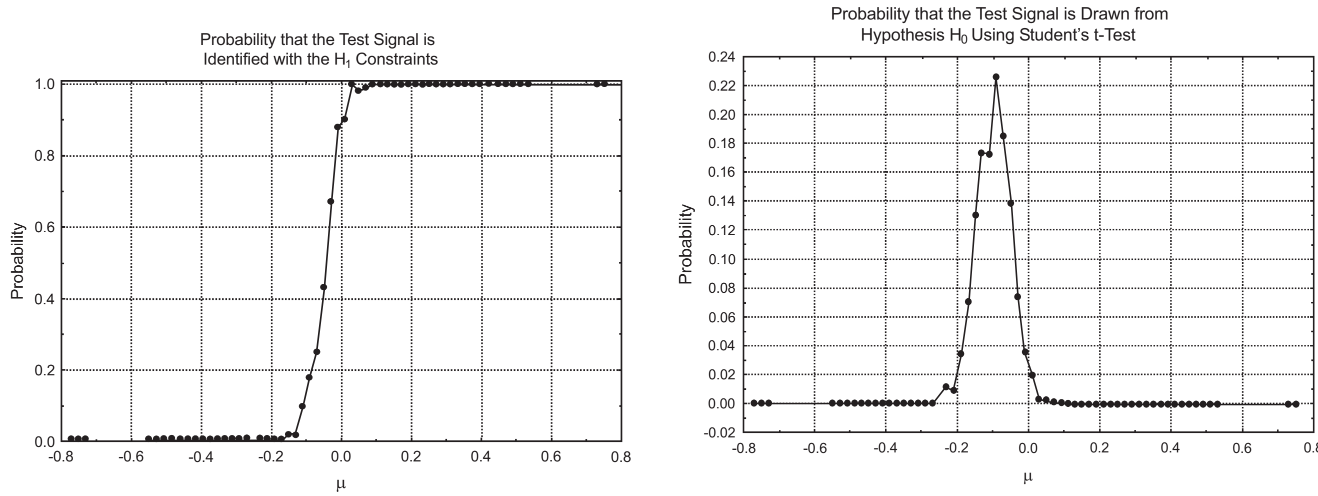
In each simulation, the test signal was 1000 samples in size with each sample point chosen randomly from the maximum entropy distribution associated with the constraints given by $\frac{C_1}{n}=1; C_2$. The Neyman-Pearson decision test with $\alpha$=0.05 was applied to each signal. 100 different signals with $C_2$ chosen as a random variable from the above distribution were analyzed and the percentage of signals classified as more likely to fit the constraints $\mathbf{H}_0$ as a function of the mean value $\mu$ when $\Delta$=0.1. In addition, Student's t-test is used to compare the test signal to the mean value expected under $\mathbf{H}_0$ and the average probability of the test signal having this mean value is computed over all of the test signals. Fig. (**6a**) shows that when the mean value of the test signal is more than 0.0 the probability of assigning the signal to $H_0$ is less than 10%. Fig. (**6b**) demonstrates that this is roughly the point at which Student's t-test indicates a less than 0.05 probability that the test signal has a mean value equal to that under $\mathbf{H}_0$. In this case, the traditional t-test performs similarly to the detector described above.

A second simulation reveals the flexibility of the detector described above. Consider the case in which the distributions of the two comparison signals are given by

$$\begin{aligned}
&\mathrm{H}_0:\rho_0\left(\tilde{C}_2\right)=\begin{Bmatrix} \frac{1}{0.4}; -0.3\leq\frac{\tilde{C}_2}{n}\leq 0.1 \\ 0; Otherwise \end{Bmatrix}; \frac{C_1}{n}=1\\
&\mathrm{H}_1:\rho_1\left(\tilde{C}_2\right)=\begin{Bmatrix} \frac{1}{0.4}; -0.3\leq\frac{\tilde{C}_2}{n}\leq 0.1 \\ 0; Otherwise \end{Bmatrix}; \frac{C_1}{n}=2
\end{aligned} \tag{B23}$$

And the test signal is governed by the distribution:

$$Test:\rho_t\left(\tilde{C}_2\right)=\begin{cases}\frac{1}{0.16};\mu-0.08\le\tilde{C}_2\le\mu+0.08\\0;Otherwise\end{cases} \tag{B24}$$




**a** **b**

**Fig. (6). a.** The results of a simulation study with the distributions characterizing hypotheses $\boldsymbol{H}_0$ (mean=-0.1) and $\boldsymbol{H}_1$ (mean=0.3) as shown in equation (B22). μ is the mean value of the test signal. Probability is the chance that the Neyman-Pearson decision criterion (37) identifies the test signal as coming from the constraints in $\boldsymbol{H}_1$. This shows that the algorithm detects the signal as not belonging to $\boldsymbol{H}_0$ when the mean of the test signal significantly exceeds the mean of $\boldsymbol{H}_0$ as expected. The mean value of the square of the signal is the same in all cases.

**b.** The results of a simulation study with the distributions characterizing hypotheses $\boldsymbol{H}_0$ (mean=-0.1) and $\boldsymbol{H}_1$ (mean=0.3) as shown in equation (B22). μ is the mean value of the test signal. Probability is the chance computed from the Student's t-test that the test distribution comes from distribution $\boldsymbol{H}_0$. It is important to note that as μ increases past the mean value of the signal in $\boldsymbol{H}_0$, the point at which the probability that this signal belongs to $\boldsymbol{H}_0$ becomes less than 0.05 is the same value at which the maximum entropy detector begins to assign most of the test signals to $\boldsymbol{H}_1$. The mean value of the square of the signal is the same in all cases.

while $C_1$ takes on varying values. Fig. (**7**) demonstrates that the test signal is reliably detected as not from $\mathbf{H}_0$ whenever $C_1$ is greater than 1.1 for any mean value. All of these results could not be explained by use of the Student's t-test because this test is only sensitive to changes in the mean values of the time series. The maximum entropy detector is sensitive to both changes in mean and variance.

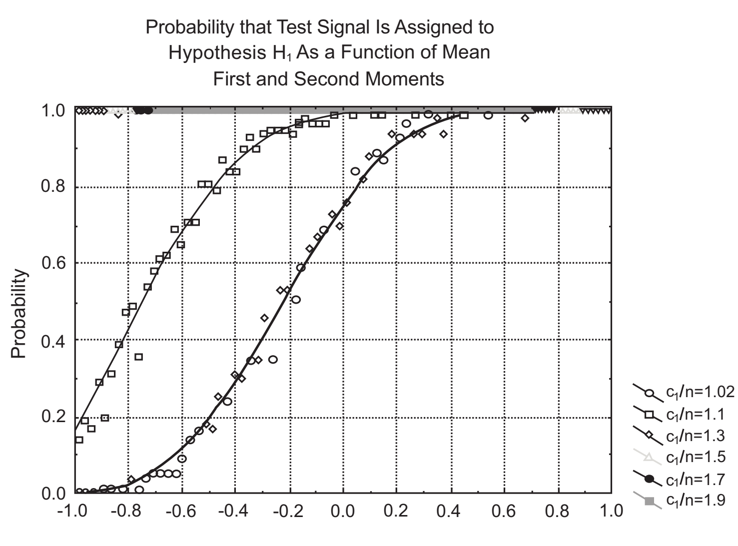



**Fig. (7).** In this simulation, probability is the chance that the Neyman-Pearson detection criteria selects the test signal as more likely coming from H1 than H0. This simulation differs from the simulation used in Figs. **6a** and **6b** in that the mean square of the signals differs in $\boldsymbol{H}_1$ and $\boldsymbol{H}_0$ and the test signal is drawn from the distribution (B24) in which the mean and mean square of the signal is allowed to vary. This demonstrates that small variations in the square of the signal value produce large changes in the distribution.

## REFERENCES


[1] S. M. Kay, *Fundamentals of statistical signal processing: volume ii detection theory*, Upper Saddle River, New Jersey, Prentice Hall, 1998.

[2] H.V. Poor. *An introduction to signal detection and estimation,* New York, Springer-Verlag, 1988.

[3] R.N. McDonough and A.D. Whalen, *Detection of signals in noise*, New York, Academic Press, 1995.

[4] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Transations on Information Theory,* vol. 44, pp. 2230-2259, 1998.

[5] R.E. Blahut, *Principles and practice of information theory*, Boston Mas, Addison-Wesley, 1987.

[6] J Ruiz-Moling, J Navarro-Morena and A. Oya. "Signal detection using approximate karhunen-loeve expansions," *IEEE Transactions On Information Theory* , vol. 47, pp. 1672-1680, 2001.

[7] C. R. Smith and W.T. Grandy *Maximum entropy and Bayesian methods in inverse problems,* Berlin Germany, Springer-Verlag, 1985.

[8] N. Wu, *The maximum entropy metho,* Berlin Germany, Springer-Verlag, 1997.

[9] J.F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Transaction on Signal Processing,* vol. 48, pp.1687-1694, 2000.

[10] M.M. Stecker, "The entropy of a constrained signal: a maximum entropy approach with applications," *Signal Processing*, vol. 88, pp.639-669, 2008.

[11] M.M. Stecker, "Approximate information content of a signal with bispectral constratints," *IEEE Transactions on Signal Processing*, vol. 51, pp. 52-63, 2003.

[12] B. Chen,. J. Hu and Y. Zhu, "Computing maximum entropy densities: a hybrid approach," *Signal Processing: An International Journal (SPIJ)*, vol. 4, pp.114-122, 2010.

[13] E. T. Jaynes, Information theory and statistical mechanics, *Physical Reviews,* vol. 106, pp. 620-630, 1957.

[14] H.B. Callen, *Thermodynamics*, New York, NY, Wiley, 1960.

[15] T.L. Hill, *Statistical mechanics*: *principles and selected applications*, New York, NY, Dover, 1956.

[16] K.P. Burnham and D.R. Anderson, "Kullback-Leibler information as a basis for strong inference in ecological studies," Wildlife *Research*, vol. 28, pp. 111-119, 2001.

[17] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic critieria: general asymptotic performance analysis," *IEEE Transactions on Signal Processing,* vol. 50, pp. 1027-1036, 2002.

[18] J. Ramirez, J. C. Segura, C. Benitez, A. de la Torre, and A. J. Rubio, "A new Kullback-Leibler VAD for speech recognition in noise," *IEEE Signal Processing Letters*, vol. 11, pp. 266-269, 2004.

[19] L. Yan and D.J. Miller, "General statistical inference for discrete and mixed spaces by an appropriate application of the maximum entropy principle," *IEEE Transactions on Neural Networks,* vol. 11, pp. 558-573, 2000.

[20] A. Ben Hamza, H. Krim, and G.B. Unal, "Unifying probabilistic and variational estimation," *IEEE Signal Processing Magazine*, vol. 19, pp. 37-47, 2002.

[21] S. Colonnese and G. Scarano, "Transient signal detection using higher order moments," *IEEE Transactions on Signal Processing,* vol. 47, pp. 515-520, 1999.

[22] D. Miller, A. Rao, K. Rose and A. Gersho, "A maximum entropy approach for optimal statistical classification," In: *Neural Networks for Signal Processing V. Proceedings of the 1005 IEEE Workshop*, 1995, pp. 58-66,.

[23] H.D. Block, "The perceptron: a model for brain functioning," *Reviews of Modern Physics*, vol. 34, pp.123-135, 1963.

[24] P. Attard, O.G. Jepps, and S. Marcelja, "Information content of signals using correlation function expansions of the entropy," *Physical Reviews E* , vol. 56, pp. 4052-4067, 1997.

[25] N. Metropolis, A.W. Rosenbluth, M.D. Rosenbluth, A.H. Teller, E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics,* vol. 21, pp 1087-1092, 1953.


Received: October 14, 2010 Revised: January 03, 2011 Accepted: January 22, 2011