HYPOTHESIS TESTING IN THE PRESENCE OF MAXWELL'S DAEMON: SIGNAL DETECTION BY UNLABELED OBSERVATIONS

S. Marano[⋆] *V. Matta*[⋆] *P. Willett*[†] *P. Braca*[‡] *R. Blum*[◇]

[⋆] DIEM, University of Salerno, Fisciano (SA), Italy, {marano,vmatta}@unisa.it
[†] ECE Department, University of Connecticut, Storrs, CT, USA, peter.willett@uconn.edu
[‡] Centre for Maritime Research and Experimentation, La Spezia, Italy, paolo.braca@cmre.nato.int
[◇] ECE Department, Lehigh University, Bethlehem, PA, USA, rblum@eecs.lehigh.edu

ABSTRACT

In modern heterogeneous sensor networks huge volumes of information rapidly flow across the system, and it is often too difficult or costly to associate data to the sensors that produced them. Then, the set of observations appears to be unlabeled: What comes from whom? We study the classical problem of detecting a known signal embedded in Gaussian noise, but under the peculiar assumption that the signal samples have been scrambled (e.g., in time or space) in an unknown way.

Our study sheds light on questions like: How much detection performance is contained in the samples' values and how much in their ordering? Are there nicely-performing detectors with affordable computational complexity?

Index Terms— Signal detection, unlabeled data, data association, scrambled signal, sample permutation.

1. INTRODUCTION

This is the time of the internet of things and big data: the objects around us – mobile phones, vehicles, buildings, surveillance cameras, environmental monitoring stations, homeland security devices, and many others – collect and deliver an enormous amount of heterogeneous data that in many cases is eventually processed to some Central Unit (CU) for inference and decision-making about some Phenomenon of Interest (PoI), examples of which are determining low-traffic routes, indicating stores for desired goods, suggesting safer places in the case of health risks, tracing escape routes in the case of security issues, and so forth. Managing such an enormous amount of data poses extremely challenging problems that often require new approaches.

Consider, in this scenario, a surveillance/monitoring system organized as a sensor network where sensors take measurements about the PoI and deliver these data to the CU which is demanded to make the final inference. One example is a network of unmanned vehicles or micro-satellites, connected to a common CU. It is often the case that the sensors (unmanned vehicles or satellites) cannot be easily synchronized, so that a sequence of data collected at the CU cannot be precisely associated to a specific sequence of time instants or geographical points.

In these and many similar applicative scenarios, when the surveillance/monitoring system is part of the more general internet of things and the global volume of data is huge, technological issues often prevent the precise association between time instants (say) and samples of the observed sequence. This inspired the present work, whose theme is as follows. We are engaged in a detection problem between two hypotheses: observations contain only noise, against the hypothesis that observations are originated by a *perfectly known* signal sequence, embedded in noise. The key point is that the known signal sequence is permuted in an unknown way.[1] What can be done in these circumstances assuming that the detector should be reasonably simple to work with huge volumes of data, which imposes severe computational and memory constraints?

The addressed scenario can be abstracted to the following model. We have a detection problem in which a known N-sequence signal $\mathbf{s} = (s_1, \ldots, s_N)^T$ is to be detected by processing a set of N observations $\mathbf{r} = (r_1, \ldots, r_N)^T$, with each r_n corresponding to a sample of the known signal, contaminated – we assume for concreteness – by additive white Gaussian noise w_n, accounting for measurement and any other source of errors. However, different from the very basic and classical problem in the detection literature, here there is uncertainly about which r_n refers to which signal sample: each observation r_n is originated by one and only one sample of $\mathbf{s}$, but all the possible associations (permutations) between sensors and samples must be accounted for.

We investigate several detectors – mainly heuristically derived – and discuss their performance. Our goal is to provide viable solutions to the detection problem, namely solutions that are reasonably simple to implement and provide acceptable performance. The next section contains standard material and serves as reference. In Sect. 3 we present the main results, followed by numerical experiments in Sect. 4, while final remarks are provided in Sect. 5.

2. BACKGROUND

2.1. Completely Known Signal

Let the $N \times 1$ vector $\mathbf{s} = (s_1, \ldots, s_N)^T$ be a known real signal, and let $\mathbf{s}^{(\pi)} = (s_{\pi(1)}, \ldots, s_{\pi(N)})^T$ be the N-sequence obtained from $\mathbf{s}$ after the sample permutation $\pi \in \Pi$, where Π is the class of all $N!$ permutations of the set $\{1, \ldots, N\}$. Let $p_s := \frac{1}{n}\sum_{n=1}^{N} s_n^2$ be the signal power (arithmetic square mean value), $\mu_s := \frac{1}{N}\sum_{n=1}^{N} s_n$ the arithmetic mean, and $\frac{1}{N}\sum_{n=1}^{N}(s_n - \mu_s)^2$ the arithmetic variance.

Consider the hypothesis test:

$$\begin{aligned} \mathcal{H}_0 :&\quad \mathbf{r} = \mathbf{w}, \\ \mathcal{H}_1 :&\quad \mathbf{r} = \mathbf{w} + \mathbf{s}^{(\pi)}, \end{aligned} \tag{1}$$

where $\mathbf{w} = (w_1, \ldots, w_N)^T$ is a $N \times 1$ vector whose entries are zero-mean Gaussian i.i.d. (independent, identically distributed) random variables with variance σ^2, and $\mathbf{r} = (r_1, \ldots, r_N)^T$. In the following we denote by $\mathbb{P}_0$ (resp. $\mathbb{P}_1$) the probability measure under hypothesis $\mathcal{H}_0$ (resp. $\mathcal{H}_1$).

Peter Willett was supported by the Naval Postgraduate School via ONR contract N00244-16-1-0017.

[1] We say that the observed samples are *unlabeled*, which we hope is not to be confused with the problem of training data without labels, arising in machine learning literature.

If the permutation π is known, the optimal detection statistic in the Neyman-Pearson sense (see e.g., [1]) is the squared norm $\|\mathbf{r}-\mathbf{s}^{(\pi)}\|^2$ or, equivalently, the scalar product $\langle \mathbf{r}, \mathbf{s}^{(\pi)} \rangle := \sum_{n=1}^{N} r_n s_{\pi(n)}$ yielding the optimal test in the form

$$\mathcal{S}_{cla} := \langle \mathbf{r}, \mathbf{s}^{(\pi)} \rangle \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{2}$$

where the threshold γ is chosen in such a way that $P_f := \mathbb{P}_0(\mathcal{S}_{cla} \geq \gamma)$ is set to a desired level. The detector in (2) will be referred to as *clairvoyant*, and the resulting detection probability is [1–3]

$$P_d = \mathbb{P}_1(\mathcal{S}_{cla} \geq \gamma) = Q(Q^{-1}(P_f) - \sqrt{N p_s/\sigma^2}), \tag{3}$$

where $Q(x) := \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\}\, dt$ is the standard Gaussian exceedance function and $Q^{-1}(\cdot)$ its inverse. A convenient performance figure is provided by the Signal-to-Noise Ratio (SNR) $\rho_s := N p_s/\sigma^2$.

2.2. Energy Detector

Suppose instead that the signal samples are completely *unknown* to the detector, i.e., $\mathbf{s}^{(\pi)}$ is a deterministic unknown vector. Then, a possible approach would be to resort to the GLRT (Generalized Likelihood Ratio Test) approach in which the unknown signal samples are replaced by their ML (Maximum Likelihood) estimate [2–5]. Since there is no statistical dependence among the samples of the observed vector $\mathbf{r}$, and the noise is Gaussian, it is easily seen (cfr., [2]) that the ML estimate of the n-th signal sample is nothing but the n-th observation itself r_n. In other words, the GLRT for the case at hand amounts to replacing $s_{\pi(n)}$ by r_n in the detection statistic of the clairvoyant detector, yielding the test in the form of an energy detector:[2]

$$\mathcal{S}_{ene} := \frac{1}{N} \sum_{n=1}^{N} r_n^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{4}$$

The performance of the energy detector is simply found in the form:

$$P_d = \mathbb{P}_1(\mathcal{S}_{ene} \geq \gamma) = 1 - \chi_{2,N,\rho_s}\left(\chi^{-1}_{2,N,0}(1-P_f)\right), \tag{5}$$

where $\chi_{2,a,b}(\cdot)$ is the chi-square Cumulative Distribution Function (CDF) [6] with a degrees of freedom and non-centrality parameter b, and $\chi^{-1}_{2,a,b}(\cdot)$ is its inverse [7]. A convenient approximation of the above expression is [2, p. 298]:

$$P_d = Q\left(\frac{Q^{-1}(P_f) - \sqrt{\rho_s^2/(2N)}}{\sqrt{1+2\rho_s/N}}\right). \tag{6}$$

3. DETECTION WITH PERMUTED SIGNAL

Consider now the case of interest in which $\mathbf{s}$ is known, but π is not: we know the values of the signal samples, but their actual position inside the vector $\mathbf{s}^{(\pi)}$ is unknown.

[2]To avoid cumbersome notation the thresholds of the various tests will be generically denoted by γ; this does not imply that they are equal.

3.1. GLRT: Ordered Data Detector

Lacking the knowledge of the permutation $\pi \in \Pi$, one classical approach would be to resort to the GLRT: the likelihood ratio is replaced by the maximum of the likelihood ratios over the class Π of all the possible permutations. Thus, conceptually, one first decides among $N!$ mutually exclusive hypotheses (the possible permutations), and then implements the likelihood ratio test between the signal-plus-noise distribution with the selected permutation, and the only-noise distribution. It is easy to see that the GLRT detection statistic for the test in (1) amounts to

$$\mathcal{S}_{ord} := \max_{\pi \in \Pi} \langle \mathbf{r}, \mathbf{s}^{(\pi)} \rangle \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{7}$$

where the maximum is over all the $N!$ permutations. The maximization in (8) does not require that one compute the $N!$ permutations, in the light of the following simple result.

Let $\bar{\mathbf{s}} = (s_{(1)}, s_{(2)}, \ldots, s_{(N)})^T$ be the ordered version of vector $\mathbf{s}$, in increasing order, namely $i < j \Rightarrow s_{(i)} \leq s_{(j)}$ and, similarly, let $\bar{\mathbf{r}} = (r_{(1)}, r_{(2)}, \ldots, r_{(N)})^T$ be the ordered version of the observed vector $\mathbf{r}$.

LEMMA 1: $\max_{\pi \in \Pi} \langle \mathbf{r}, \mathbf{s}^{(\pi)} \rangle = \langle \bar{\mathbf{r}}, \bar{\mathbf{s}} \rangle$. •

Proof: The proof can be found in the literature, see, e.g., [8, App. I], but is given here for self-consistency. First, note that $\max_{\pi \in \Pi} \sum_{n=1}^{N} r_n s_{\pi(n)} = \max_{\pi \in \Pi} \sum_{n=1}^{N} r_{(n)} s_{\pi(n)}$, namely we can replace $\mathbf{r}$ with its ordered counterpart $\bar{\mathbf{r}}$. Consider then two addends of the sum $\sum_{n=1}^{N} r_{(n)} s_{\pi(n)}$, say $r_{(m)} s_{\pi(m)} + r_{(k)} s_{\pi(k)}$, with $m < k$. If $s_{\pi(m)} > s_{\pi(k)}$, then exchanging these two entries increases (or leaves unchanged) the value of the sum. To see this, note that $s_{\pi(m)} > s_{\pi(k)}$ implies $(r_{(m)} - r_{(k)})(s_{\pi(m)} - s_{\pi(k)}) \leq 0$, which can be rewritten as $r_{(k)} s_{\pi(k)} + r_{(m)} s_{\pi(m)} \leq r_{(k)} s_{\pi(m)} + r_{(m)} s_{\pi(k)}$. Exchanging the two entries is tantamount to selecting a different permutation of the set Π, and this procedure is repeated until no more exchanges are possible. The only permutation $\bar{\pi} \in \Pi$ with this property is when $m < k \Rightarrow s_{(m)} \leq s_{(k)}$, namely that for which the signal vector is sorted in increasing order $\mathbf{s}^{(\bar{\pi})} = \bar{\mathbf{s}}$, which proves the claim. □

Using Lemma 1, the statistical test can be written in the simple form

$$\mathcal{S}_{ord} = \langle \bar{\mathbf{r}}, \bar{\mathbf{s}} \rangle \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{8}$$

which will be referred to as the ordered-data test.

In general, the GLRT is not optimal. The GLRT is expected to perform well when the unknown parameter can be efficiently estimated as happens, e.g., when the signal-to-noise ratio is large and/or when the estimation quality improves as the number of samples N grows. Unfortunately, in our setting, increasing N corresponds to increasing the number of unknowns, namely, the number of possible configurations (permutations) of the known signal samples. As already observed, the energy detector is also a GLRT, which estimates the underlying signal blindly, namely, without any prior knowledge of the signal samples. In this connection, it is very interesting to evaluate the effective information gain that can be achieved from the knowledge of the *scrambled* signal sequence. In other words, it is legitimate to ask whether: i) the ordered-data detector really takes advantage of the scrambled signal sequence, or ii) the difficulty of the estimation task, especially for large N, makes the ordered-data detector almost equivalent to the energy detector. Such aspects will be addressed in due detail when illustrating the experimental results.

3.2. Linear Approach: Mean Detector

In order to handle the lack of knowledge of the permutation $\pi \in \Pi$, another possible approach is to look for some linear transformation of the data that is invariant with respect to π. This approach is close in spirit to the synthetic discriminant function approach used in pattern recognition [9]. In formulas, we seek a matrix $A \in \Re^{M \times N}$ such that $\mathbf{y} := A\,\mathbf{r}$ is invariant. Since $\mathbf{r} = \mathbf{w} + \mathbf{s}^{(\pi)}$ this translates to requiring that $A\,\mathbf{s}^{(\pi_1)} = A\,\mathbf{s}^{(\pi_2)}$ for any $\pi_1, \pi_2 \in \Pi$. The statistical test (1) then becomes

$$\begin{array}{ll} \mathcal{H}_0: & \mathbf{y} = A\,\mathbf{w}, \\ \mathcal{H}_1: & \mathbf{y} = A\,(\mathbf{w} + \mathbf{s}^{(\pi)}). \end{array} \tag{9}$$

A little thought reveals that the aforementioned requirement is met if the rows of A have constant entries, namely:

$$A = \begin{pmatrix} \alpha_1 & \alpha_1 & \dots & \alpha_1 \\ \alpha_2 & \alpha_2 & \dots & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_M & \alpha_M & \dots & \alpha_M \end{pmatrix}, \tag{10}$$

in which case we have $y_m = \alpha_m N \mu_s$, for $m = 1, \dots, M$. Similarly, the m-th entry of the noise vector $A\mathbf{w}$ results in $\alpha_m \sum_{n=1}^{N} w_n$. This implies that test (9), when using matrix A in (10), is equivalent to the test made on a single entry, e.g., the m-th, of the observed $\mathbf{y}$. Since in that case the constant α_m can be set to $1/N$ with no loss of generality, the statistical test reduces to the following, with the scalar observation y:

$$\begin{array}{ll} \mathcal{H}_0: & y = \frac{1}{N}\sum_{n=1}^{N} w_n, \\ \mathcal{H}_1: & y = \mu_s + \frac{1}{N}\sum_{n=1}^{N} w_n, \end{array} \tag{11}$$

where $\frac{1}{N}\sum_{n=1}^{N} w_n \sim \mathcal{N}(0, \sigma^2/N)$, which is a shortcut to stating that $\sum_{n=1}^{N} w_n$ is Gaussian with mean 0 and variance σ^2/N. In terms of the original observation vector $\mathbf{r}$, the optimal solution for test (9) is

$$\mathcal{S}_{mean} := \frac{1}{N}\sum_{n=1}^{N} r_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \tag{12}$$

for $\mu_s \geq 0$, while the inequalities are reversed for $\mu_s < 0$. The resulting performance is

$$P_d = Q(Q^{-1}(P_f) - \sqrt{N\mu_s^2/\sigma^2}). \tag{13}$$

The SNR loss incurred by this "mean" detector with respect to the clairvoyant, using for the latter the approximation (6), is

$$\mathcal{L}_{mean/cla} := 1 - \frac{\mu_s^2}{p_s}. \tag{14}$$

LEMMA 2: *$\mathcal{L}_{mean/cla} \in [0, 1]$. $\mathcal{L}_{mean/cla}$ is equal to 0 (no loss) if and only if $\mathbf{s}$ is a constant signal (all entries are equal), and $\mathcal{L}_{mean/cla}$ is equal to 1 (maximum loss) if and only if $\mu_s = 0$.* •

Proof: This follows straightforwardly from definition (14), by exploiting the fact that the arithmetic variance is nonnegative: $0 \leq \frac{1}{N}\sum_{n=1}^{N}(s_n - \mu_s)^2 = p_s - \mu_s^2$, with equality if and only if $s_n = \mu_s$ for all $i = 1, \dots, N$. □

We see that the detector using the statistic $\mathcal{S}_{mean}$ performs as well as the optimum only when the signal is constant, which is obvious because in that case the permutation has no effect at all, and the two detectors are actually the same. To the other extreme, a signal with arithmetic mean equal to zero cannot be detected using the mean detector. In general, the performance of the mean detector is expected to be reasonable when μ_s^2 is close enough to the signal power p_s.

3.3. CDF-Based Detector

The bottom line of the analysis conducted in the previous two sections is that: i) the GLRT is expected to be neither optimal, nor necessarily well-suited to our problem, especially for large number of samples, and ii) the linear approach works for nonzero-mean signals, which appears to be an unnecessary restriction to a particular class of signals. These considerations motivate us to seek alternative solutions. In designing a novel detector, we are primarily guided by the following observations.

• There is no particular reason to focus only on the first moment (i.e., the mean) of the signal when designing a detector. Some useful information could also be contained in the higher-order moments.

• The detector must depend upon the signal sequence, but we seek a detector which is *invariant* to permutations of the signal sequence.

• Finding alternatives to a GLRT approach when the dimensionality of the unknown parameter is large, and/or increases with N, is a challenging problem. Notable examples that have been considered in the literature include the case of transient detection [10–16], and fall under the umbrella of the general problem of simultaneous detection and estimation [17–20]. One appealing strategy is that of constructing a suitable prior distribution for the signal, and designing the corresponding (Bayes/minimax-optimal) detector. In doing so, one should cleverly include the available knowledge about the signal in the prior model.

• Given the lack of knowledge about the true signal ordering, knowing the signal samples essentially corresponds to knowing the empirical CDF of the signal sequence. Note that the empirical CDF intuitively matches our idea of encompassing the information about higher-order moments.

• Let us interpret the entries in the sequence $\mathbf{s}$ as i.i.d. realizations drawn from the empirical Probability Density Function (PDF) [6]:

$$\frac{1}{N}\sum_{n=1}^{N} \delta(s - s_n), \tag{15}$$

where s is the independent variable, the s_n's are the known signal samples, and $\delta(s)$ is the Dirac delta generalized function. This PDF is perhaps the most straightforward solution to select a prior on the signal. (Elaborating on the empirical distribution is reminiscent of the popular method of types for finite alphabets, see, e.g., [21].)

In the light of the above observations, we construct the Neyman-Pearson detector corresponding to the case that, under $\mathcal{H}_1$, the signal samples are drawn from the PDF in (15). Accordingly, the distribution of the data under $\mathcal{H}_1$ results in the convolution of the Gaussian PDF of w_n with the PDF in (15), yielding:

$$\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = N^{-N}\prod_{i=1}^{N}\sum_{n=1}^{N} e^{s_n(r_i - s_n/2)/\sigma^2}. \tag{16}$$

Taking the logarithm and incorporating constant terms into the threshold, we finally get:

$$\mathcal{S}_{cdf} := \frac{1}{N}\sum_{i=1}^{N} \log \sum_{n=1}^{N} e^{s_n(r_i - s_n/2)/\sigma^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{17}$$

which will be referred to as the CDF-based detector.

A remark is now in order. At least for sufficiently large N, it might be expected that knowledge of the signal sequence (not of the ordering) does not add much more information with respect to treating the signal sequence as a *random* sequence. However, we exclude availability of a model for the distribution of the signal samples, while we assume availability of the scrambled signal sequence. Such signal sequence is exploited by the CDF-based detector to compute an empirical distribution of the samples.

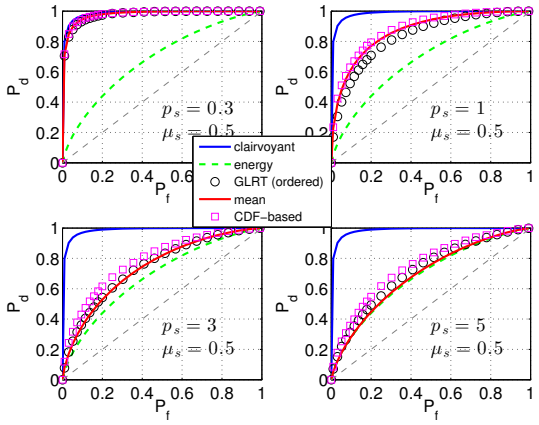

Fig. 1: ROC of different detectors with $\rho_s = 10$ dB, $N = 100$. The signal is a linearly increasing sequence made of N equally spaced samples with the shown values of the parameters p_s and μ_s.

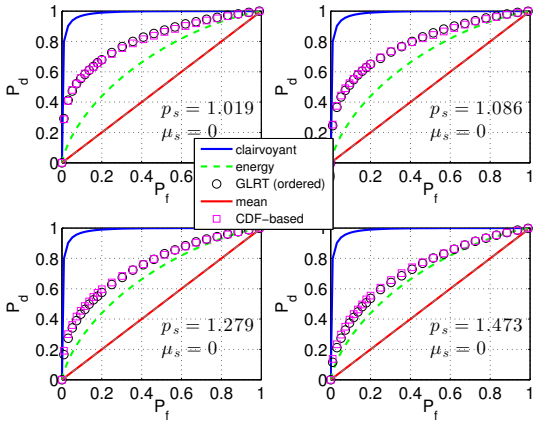

Fig. 2: ROC of different detectors with $\rho_s = 10$ dB, $N = 100$. The signal is a zero-mean spiky sequence (see main text), with the shown values of p_s.

4. NUMERICAL EXPERIMENTS

We run computer experiments to simulate the detection system under $\mathcal{H}_0$ and under $\mathcal{H}_1$. The signal $\mathbf{s}$ is a linearly increasing sequence made of N equally spaced samples from $s_1 = \min(s_1, \ldots, s_N)$ to $s_N = \max(s_1, \ldots, s_N)$, where s_1 and s_N are chosen to ensure that the signal $\mathbf{s}$ has prescribed values of arithmetic mean value μ_s and power p_s. Figure 1 refers to $\mu_s = 0.5$ and various values of p_s. We assume $N = 100$ signal samples and the SNR is $\rho_s = 10$. The results of the numerical experiments are reported in Fig. 1 in the form of ROC (Receiver Operating Characteristic) curves, i.e., by plotting P_d versus P_f. The theoretical formulas (solid lines) refer to eq. (3) for the clairvoyant, eq. (5) for the energy detector, and eq. (13) for the mean detector. The symbols refer to the ordered-data and to the CDF-based detectors, respectively, and are obtained by means of computer simulations based on 10^4 Monte Carlo runs.

In the four panels of Fig. 1 the ROC curves of the clairvoyant and of the energy detector are the same, as their performance only depends on ρ_s and N. In the upper-left panel we see that the mean, the ordered-data, and the CDF-based detectors perform similarly, and close to the clairvoyant. In the remaining three panels we see a certain superiority of the CDF-based detector, which seems to exploit better the information contained in the data. Conversely, the ordered-data detector first is outperformed by (upper-right) and then outperforms (lower-right) the mean detector, and they are essentially equivalent in the case addressed in the lower-left panel.

Next, we consider a *spiky* signal obtained by first defining a linearly increasing sequence made of N equally spaced samples (in the shown example from $s_1 = -0.297$ to $s_{100} = 0.297$), then adding a spike at the first sample ($s_1 = 10$) and finally subtracting the mean value of the resulting signal in order to impose zero mean. With the signal sequence so constructed, we obtain Fig. 2. The mean detector, needless to say, is useless. The ordered-data and the CDF-based perform similarly and significantly better than the energy detector, showing remarkable ability to exploit the information contained in the unordered samples.

Admittedly, the number of signal samples considered in our numerical experiments ($N = 100$) is rather small compared to the typical numbers arising in the applications mentioned in Sect.1. Simulations on much larger scale and/or comparison with available datasets used as benchmark are certainly desirable, and left for future studies.

5. FINAL REMARKS

The scenario with a known signal embedded in white Gaussian noise is among the most classical and widely known examples of signal detection problems, whose Neyman-Pearson solution is the so-called matched filter (the "clairvoyant," in this work). The modern ubiquity of sensor networks and similar kinds of distributed inference systems operating under the emerging paradigms of the internet of things and big data motivated us to study the aforementioned detection problem by assuming that the signal samples are *unlabeled*. How much "information for detection" is contained in the unknown signal ordering? How much is contained in the sample *values* and is therefore retained? How to exploit that by a computationally affordable detector?

Consider the upper-right panel in Fig. 1: the performance of energy detector represents what can be done without knowledge of the signal values. The detectors investigated in this work show a remarkable performance gain with respect to the energy detector, and the performance gap shown in the figure quantifies what it can be gained by the knowledge of the signal values. Similarly, the gap between the performance of these detectors and that of the clairvoyant one quantifies the additional information that could be retrieved by knowing who comes from whom.

Our studies reveal a certain superiority of the CDF-based detector over the GLRT and the mean detectors, which makes sense in the light of the arguments discussed in this paper. But we also see that the CDF-based and the GLRT perform similarly when the signal is zero mean (Fig. 2). In addition, for zero-mean signals which are less "spiky" than those considered in Fig. 2, computer experiments not reported here reveal that the performance of these two detectors is close to that of the energy-based one. This suggests further investigations.

Other lines for further studies include other forms of time (or space) uncertainty, including, e.g., possible repetitions of the same sample, "local" scrambling of nearby samples, and so forth. The present study can be also extended to the detection of signals defined over graphs, where the time or space dependence is replaced by the more structured association between samples and network nodes, assuming (partially) unknown labeling. We plan to address some of these issues in an extended version of this work, which is in preparation.

6. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I*. New York: John Wiley & Sons, Inc., 1968 (reprinted, 2001).

[2] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1998.

[3] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.

[4] E. L. Lehmann, *Testing Statistical Hypotheses*. Springer, 2005.

[5] E. L. Lehmann and G. Casella, *Theory of Point Estimation*. Springer, 1998.

[6] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

[7] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*. New York: John Wiley & Sons, 1994.

[8] D. Slepian, "Permutation codes," *Proc. IEEE*, vol. 53, pp. 228–236, Mar. 1965.

[9] D. Casasent, "Unified synthetic discriminant function computational formulation," *Applied Optics*, vol. 23, no. 10, pp. 1620–1627, May 1984.

[10] C. Han, P. Willett, B. Chen, and D. Abraham, "A detection optimal min-max test for transient signals," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 866–869, Mar. 1998.

[11] R. L. Streit and P. K. Willett, "Detection of random transient signals via hyperparameter estimation," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1823–1834, Jul. 1999.

[12] Z. Wang and P. Willett, "A performance study of some transient detectors," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2682–2685, Sep. 2000.

[13] B. Chen and P. Willett, "Detection of hidden markov model transient signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1253–1268, Oct. 2000.

[14] ——, "Superimposed HMM transient detection via target tracking ideas," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 3, pp. 946–956, Jul. 2001.

[15] Z. Wang and P. Willett, "All-purpose and plug-in power-law detectors for transient signals," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2545–2466, Nov. 2001.

[16] Z. J. Wang and P. Willett, "A variable threshold Page procedure for detection of transient signals," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4397–4402, Nov. 2005.

[17] B. Baygun and A. O. Hero, "An order selection criterion via simultaneous estimation/detection theory," in *Proc. of Fifth ASSP Workshop on Spectrum Estimation and Modeling*, 14-17 Apr. 1990.

[18] ——, "Tradeoffs between detection and estimation for multiple signals," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1991)*, Toronto, Ontario, Canada, 10-12 Oct. 1991.

[19] ——, "Further results on tradeoffs between detection and estimation," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 1992)*, San Francisco, CA, USA, 23-26 March 1992.

[20] ——, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 688–703, May 1995.

[21] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.