19th European Signal Processing Conference (EUSIPCO 2011) Barcelona, Spain, August 29 - September 2, 2011

QUICKEST DISTRIBUTED DETECTION VIA RUNNING CONSENSUS

Paolo Braca†, *Stefano Marano*†, *Vincenzo Matta*†, *Peter Willett**

† Department of Electronic & Computer Engineering (DIEII), University of Salerno
via Ponte don Melillo I-84084, Fisciano (SA), Italy
e-mail: {pbraca, marano, vmatta}@unisa.it

* Electrical & Computer Engineering Department (ECE)
University of Connecticut, Storrs, CT 06269 USA
e-mail: willett@engr.uconn.edu

ABSTRACT

Running consensus is a recently proposed distributed strategy for fostering agreement among sensors of fully flat networks, by interleaving the two stages of measurements and node-to-node communications. Quickest detection is a well-established technique for discovering abrupt changes (if any) in the statistical distribution of the observed data.

In this paper we tailor the running consensus idea to the quickest detection problem, to address change-detection issues in distributed inference systems with random and time-varying sensors' connections, in architectures without fusion center. Performance benchmarks are expressed in terms of detection delay and false alarm rate, for which closed form approximations are derived, yielding a simple analytical expression of the operational characteristic of the detector. The proposed system is tested on typical signal processing problems by means of numerical simulations that validate the theoretical analysis.

Keywords—Quickest detection, Page's test, Running consensus.

1. INTRODUCTION

Running consensus is a gossip paradigm for sensor networks, originally proposed in [1], whose main feature is the simultaneous managing of the acquisition and data-exchange stages, that allows reaching agreement at node level by elaborating on the *time-varying* dataset collected by the network, elaborating on classical gossip protocols [2–5]. Running consensus has been recently recognized as an efficient way to perform distributed detection in non-parallel, fully flat architectures, that is, when no fusion center is available [6, 7]. Extensions and applications to random and time-varying networks have been proposed in [8].

The typical way of operation for such decentralized detectors prescribes that the sensors acquire data, exchange their local information, that, suitably processed, lead to agreement about a final decision that is (asymptotically with time) common to all nodes.

In this paper, we focus on detecting abrupt changes in the data distribution, usually referred to as quickest detection. This is a classical problem, that emerges in practical scenarios where a sudden change in the state of the nature is to be reported as soon as possible. Page's test is a well-established signal processing technique for quickest change detection, relying upon the so-called CUSUM statistic [9–11].

Departing from the classical centralized application of Page's tests, the investigation has been extended in several directions, including transient changes, partially unspecified statistical models, different optimization criteria constraints, quantized data. Useful entry points for these topics can be, among many others, [10–12].

However, as far as we can tell, quickest change detection for fully distributed detection in sensor networks has not received the same degree of attention in the topical literature. This motivates us in pursuing the basic idea behind this work, that of merging the running consensus update rule with Page's test recursion.

The remainder of this paper is so organized. In Sect. 2 we pose and formalize the problem, also including the basic relevant facts about Page's test. Section 3 contains the main results of the proposed strategy. Section 4 collects a summary of the results of the numerical simulations used for testing the algorithm, in the context of typical change detection problems. Conclusive remarks are provided in Sect. 5.

2. PROBLEM FORMALIZATION

The basic change detection problem considered in this work is now formalized according to a very classical setup [9, 10]. In the following, the index $j \in \{1, 2, \dots, M\}$ identifies a specific sensor, while $n \geq 1$ is the (discrete) time index. The $n-$th observation $x_{n,j}$ collected by the $j-$th node follows the null-hypothesis distribution $f_0(x)$ until a deterministic but *unknown* time n_0. From n_0 (included) on, the distribution for all j suddenly *changes* to $f_1(x)$.

The goal of the network is to discover the change as soon as possible, with a constraint on the average time between false alarms. Throughout the paper, we make the basic assumption of statistical independence across time and across sensors. We have, for all j,

$$\begin{array}{lll} f_0(x) : x_{1,j}, x_{2,j}, \dots, x_{n_0-1,j} & \searrow & \\ f_1(x) : & & x_{n_0,j}, x_{n_0+1,j}, \dots \end{array}$$

Note that, at each time slot n, the network globally collects a vector of observations:

$$\mathbf{x}_n = [x_{n,1}, x_{n,2}, \dots, x_{n,M}].$$

2.1 Classical parallel architecture

If a fusion center is available, the quickest detection problem can be addressed by means of the well-known Page's test [9], which is basically made of the following three elements.

- The CUSUM log-likelihood of the data is

$$\sum_{i=1}^{n}\sum_{j=1}^{M}\log\frac{f_1(x_{i,j})}{f_0(x_{i,j})}. \quad (1)$$

- Page's recursion rule is defined as

$$S_n = \max\left\{0, S_{n-1} + \sum_{j=1}^{M}\log\frac{f_1(x_{n,j})}{f_0(x_{n,j})}\right\}, \quad (2)$$

  where $S_0 = 0$. We explicitly note that the log-likelihood resets each time it falls below zero, which is thus the point from which Page's test restarts.

- A decision rule prescribing that a change is declared as soon as a threshold γ is crossed, implicitly defining the test stopping time as

$$N = \arg\min_{n}\{S_n \geq \gamma\}. \quad (3)$$

The usual optimality criterion for assessing the test performance is that of imposing a constraint on the false alarm rate, and accordingly minimizing the detection delay. The former is defined as the reciprocal of the average sample size under the null hypothesis, $1/\mathrm{E}_0[N]$, where $\mathrm{E}_{0,1}[\cdot]$ denotes expectation computed under distribution $f_{0,1}(x)$. The latter is approximated by $\mathrm{E}_1[N]$, which is in fact an upper bound on the real delay, corresponding to the assumption that the CUSUM is exactly zero at time n_0. The precise computation of $\mathrm{E}_1[N]$ would instead require knowledge of the exact value of the CUSUM statistic at n_0, and it is usually intractable [9, 10, 13].

The above key quantities admit closed form approximations mainly relying upon neglecting the excess over the threshold of the test statistic at the stopping time [9, 10, 13]. Specifically, the false alarm rate and detection delay of the centralized system (suffix c consistently appended) are related to the detection threshold via

$$R_c(\gamma) \approx \frac{M\,\Delta_{01}}{e^{\gamma}-\gamma-1}, \quad (4)$$

$$D_c(\gamma) \approx \frac{\gamma+e^{-\gamma}-1}{M\,\Delta_{10}}, \quad (5)$$

where Δ_{01} is the Kullback-Leibler divergence [10] from $f_0(x)$ to $f_1(x)$, and Δ_{10} is similarly defined. By combining (4) and (5) the basic operational curve $D_c(R)$ of the detector, that expresses the detection delay as a function of a prescribed false alarm rate R, can be obtained. In the regime of large γ (corresponding to small false alarm rates), the operational curve can be conveniently approximated by the following closed form

$$D_c(R) \approx \frac{\log\left(M\,\Delta_{01}/R\right)}{M\,\Delta_{10}}. \quad (6)$$

Note that the overall divergence pertaining to a single time slot is $M\,\Delta$, accounting for the fact that, at each time slot, M independent observations are collected.

3. RUNNING CONSENSUS FOR QUICKEST DETECTION

As already anticipated, the main strategy proposed in this work for quickest distributed detection in fully flat networks relies upon the running consensus algorithm. Details about this latter can be found in [1, 6] and will not be repeated here for space reasons. In the following we limit ourselves to report the basic elements in order to make the paper self-contained.

The network topology is formalized by an undirected graph[1] $(\mathcal{V}, \mathcal{E}_n)$ where $\mathcal{V} = \{1, 2, \ldots, M\}$ is the vertex set (sensors) and $\mathcal{E}_n$ the edge set that describes sensors' connections. To address the general problem of random and time-varying sensors' connections, we allow $\mathcal{E}_n$ to be random and dependent upon the time slot n. Accordingly, at each n, M data are collected by the network and a realization of $\mathcal{E}_n$ is drawn, meaning that some subset of $\mathcal{V}$ is selected, and the corresponding nodes share their states according to a standard consensus algorithm [2]. The exchanged data are not simply the measurements, but rather the suitable detection statistics computed by the nodes, summarized in the state variables $S_{n,j}$.

Stressing on the flat architecture of the system, we would like to achieve the following goals.

- Each sensor implements its own test by comparing the local statistic $S_{n,j}$ to a detection threshold γ. The $j-$th test accordingly stops at a random time

$$N_j = \arg\min_{n}\{S_{n,j} \geq \gamma\}. \quad (7)$$

- No post-detection fusion of the local decisions is allowed, the data fusion being instead embodied in the running consensus protocol.
- The decision taken by *any* of the sensors must be representative of the (unavailable) global, centralized decision. Accordingly, it must be possible to retrieve a reliable decision by querying an arbitrary node in the network.

These design goals basically require asymptotic (with n) similarity of $S_{n,j}$ with the centralized detection statistic S_n, for all j. To this aim, we propose the following update rule, that is essentially borrowed from the running consensus data-exchange protocol [1, 6]:

$$\begin{pmatrix} S_{n,1} \\ S_{n,2} \\ \vdots \\ S_{n,M} \end{pmatrix} = \mathbf{W}_n \begin{pmatrix} S_{n-1,1} \\ S_{n-1,2} \\ \vdots \\ S_{n-1,M} \end{pmatrix} + M\,\mathbf{W}_n \begin{pmatrix} \log\frac{f_1(x_{n,1})}{f_0(x_{n,1})} \\ \log\frac{f_1(x_{n,2})}{f_0(x_{n,2})} \\ \vdots \\ \log\frac{f_1(x_{n,M})}{f_0(x_{n,M})} \end{pmatrix}$$

or, in a more compact form

$$S_{n,j} = \mathcal{U}(\{S_{n-1,j}\}_{j=1}^{M}). \quad (8)$$

[1]Unlike the case of the classic consensus, the literature on running consensus [1,6] only deals with undirected graph network; we comply with such assumption.

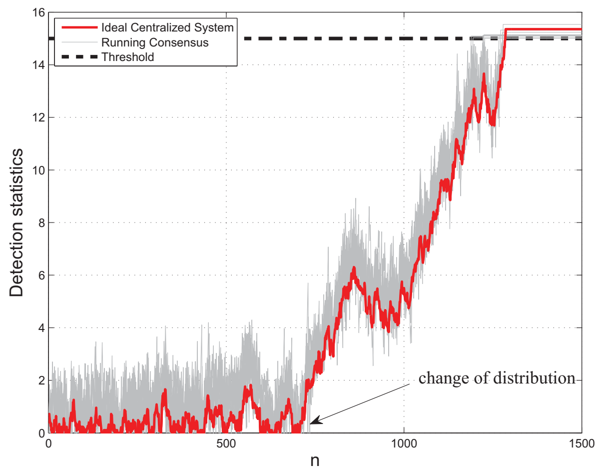

Figure 1: Empirical realizations of the running consensus statistics, in a fully connected network made of 10 sensors, operating under the pairwise averaging algorithm. The distributions under the two hypotheses are Bernoulli, $x_{n,j} \in \{0,1\}$. Under the null hypothesis the measurements are equiprobable, while after the change (at $n_0 = 750$) the probability of 1 modifies to 0.51. Thinner gray lines refer to sensors (almost superimposed to each other), the bold red line refers to the centralized system

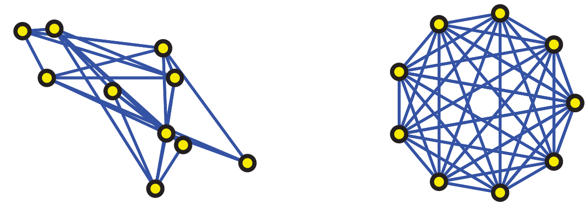

Figure 2: Network topologies for the two examples considered in Sect. 4. The circles represent the vertex set $\mathcal{V}$, while the random edge set $\mathcal{E}_n$ is selected among the possible connections shown by lines between the vertices.

The M by M consensus matrices $\mathbf{W}_n$, $n = 1, 2, \ldots$, are iid (independent identically distributed) and doubly stochastic. To better highlight the physical meaning of the above matrices, let us consider the classical example of a pairwise averaging algorithm, according to which, at time n, a pair (h, k) of sensors is selected uniformly at random. The corresponding realization of $\mathbf{W}_n$ is

$$\mathbf{W}_n = \mathbf{I} - \frac{(\mathbf{u}_k - \mathbf{u}_h)(\mathbf{u}_k - \mathbf{u}_h)^T}{2}, \tag{9}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{u}_k$ is a vector of all zeros, but for the $k-$th entry which is unity. Using this matrix into the update equation simply amounts to let sensors h and k replace their state by the corresponding arithmetic averages. Formally, in this case $\mathcal{E}_n = \{h, k\}$.

Our solution for quickest detection via running consensus is finally obtained by merging the update rule (8) to the classical Page's recursion (2), the overall recursion (at the $j-$th node) becoming[2]

$$S_{n,j} = \max\{0, \mathcal{U}(S_{n-1,j})\}. \tag{10}$$

Before going into the details of performance evaluation, it is instructive to start from empirical evidences. Figure 1, obtained by computer experiments, displays the behavior of the ideal centralized statistic S_n (bold red curve), along with the locally computed sensor statistics $S_{n,j}$ (tiny gray curves) of the running consensus Page's detectors. A general trend is observed: in a first portion of the time axis, the statistics often reset to zero; once that the change in distribution takes place ($n_0 = 750$), they tend to grow up to eventually cross the detection threshold. As a matter of fact, the different running consensus statistics always behave quite similarly, and, in addition, closely track the statistic of the centralized system. This in turn implies that the instants of detection events, i.e., the times at which the curves cross the positive threshold, are almost the same for the different statistics, leaving hope that the performance of the running consensus quickest detectors may approach the theoretical limit represented by the performance of the centralized system.

This behavior can be explained as follows. Running consensus introduces strong dependencies among nodes by continuously propagating information across the network, and this implies that the change is detected at almost equal times at different sensors. As time elapses, the effect is emphasized and the statistics $S_{n,j}$ at different j become closer and closer each other.

As a consequence, provided that the algorithm evolves for a sufficiently long time, a reliable estimate of the instant at which the distribution-change took place can be obtained by querying any of the M nodes, and the performance of the running consensus scheme can be computed with reference to any of the sensors, according to the genuinely flat nature of the system.

3.1 Performance evaluation

A complete derivation of the performance formulas is not reported with all the details here; we refer the reader to [7]. The arguments below, however, are sufficient for a complete understanding of the main ideas behind the formal derivations.

It is convenient to regard the local detection statistic as $S_{n,j} = S_n + e_{n,j}$, where the difference between the current state $S_{n,j}$ and its centralized counterpart S_n is measured by an error term, that is assumed for now to be bounded, $|e_{n,j}| \leq \epsilon$, $\forall n$ and $\forall j$.

Sensors initially acquire data following the distribution $f_0(x)$. Until a threshold crossing occurs (either because a real change happened, or because a false alarm is going to be declared), the $j-$th sensor may have experienced a certain number of resets. This number, however, does not depend only upon S_n, but it is also determined by the behavior of the error term $e_{n,j}$. On the other hand, it is reasonable to assume that, for $\gamma \gg \epsilon$,

[2] While the update rule $\mathcal{U}$ is linear, the addition of Page's reset rule introduce a nonlinear effect, which is not present in the classical gossip algorithms.

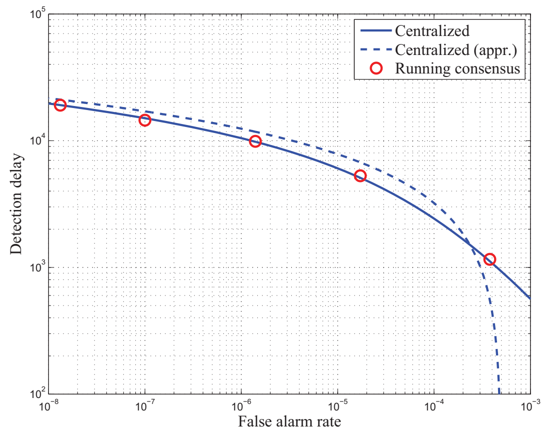

Figure 3: Operational curve of the running-consensus quickest detector. The detection delay is shown as function of the false alarm rate for the Bernoulli example considered in Sect. 4. Dots refer to simulations of the detector, while the continuous, see eqs. (4)-(5), and dashed curves, see eq. (6), show the operational characteristic of the ideal centralized system.

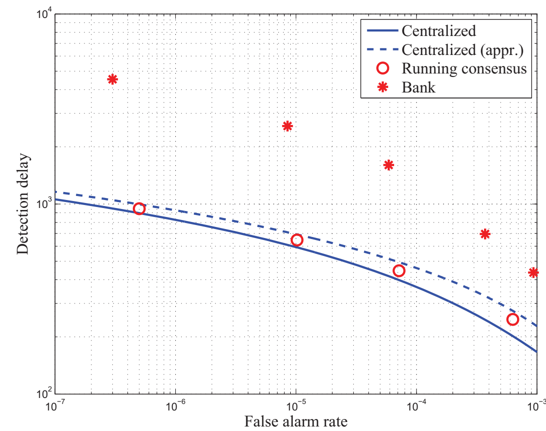

Figure 4: Operational curve of the running-consensus quickest detector. The detection delay is shown as function of the false alarm rate for the Gaussian example of Sect. 4. Dots refer to simulations of the running consensus strategy, while the continuous and dashed curves pertain to the ideal centralized system. Asterisks show the operational curve of a bank of Page's detectors.

the role of the centralized statistic S_n as to the threshold crossing will be predominant. Formally we have the following: Let us define

$$\underline{N} = \arg\min_n \{S_n > \gamma - \epsilon\},$$
$$\overline{N} = \arg\min_n \{S_n > \gamma + \epsilon\}, \qquad (11)$$

that are nothing but the stopping times pertaining to a centralized Page's test with modified thresholds. Obviously, we have $\mathrm{E}_{0,1}[\underline{N}] \leq \mathrm{E}_{0,1}[N_j] \leq \mathrm{E}_{0,1}[\overline{N}]$. Let us focus on $f_1(x)$. Applying the last inequality, we have, for the detection delay at the $j-th$ sensor:

$$D_c(\gamma - \epsilon) \leq D_j \leq D_c(\gamma + \epsilon).$$

In the regime of large γ (i.e., of small false alarm rate), we can neglect the effect of ϵ (which is $\ll \gamma$), and obtain the approximate operational characteristic of the running scheme:

$$D_r(R) \approx \frac{\log (M\,\Delta_{01}/R)}{M\,\Delta_{10}}. \qquad (12)$$

We have assumed so far that the error is bounded. Such assumption is usually made in sequential analysis for managing the errors due to the excesses over the thresholds, and provides simple refinements of Wald's approximations, see, e.g. [13]. We would like to mention that an extension of these results to the case that the errors are bounded only on the average can also be pursued, but this would require rather advanced mathematical tools [14].

4. NUMERICAL EXPERIMENTS

For the classic running consensus schemes it is well-known that the performance depends on the network connectivity: if the adjacency matrix of the graph is sparse (especially when the number of nodes is large), then the rate of convergence towards the optimal statistic slows down. We expect the same behavior for our decentralized Page's test. While exhaustive investigations in this direction are out of the scope of this paper, we next present a summary of the results obtained by Monte Carlo simulations for two typical setups, where the nodes communicate only with their direct (single hop) neighbors. The two considered examples are schematically displayed in Fig. 2, where neighboring sensors are connected by straight lines. The symmetric topology shown in the right plot serves as an ideal benchmark, while the left plot shows a more realistic scenario. Note that both the network graphs are connected.

In the first example, we assume the measurements taken by the sensors as iid binary variables taking value in $\{0,1\}$, drawn from a Bernoulli distribution. The network topology is depicted in the left plot of Fig. 2. Initially, the outcomes are equiprobable, while after the change the probability of 1 slightly modifies to 0.505. Note that the two hypotheses are "quite close", thus leading to a challenging detection task. It is also assumed that at each time step v pairs of neighboring sensors are selected to average their own states. In the following examples we use $v = 5$.

The results of 10^4 Monte Carlo simulations, with $M = 10$ sensors, are shown in Fig. 3, where the empirical operational characteristic of the detector is compared to the operational curve $D_c(R)$ obtained by combining (4) and (5); also shown is the closed form approximation in eq. (6), valid for tight false alarm rates R. As it can be seen, the match with $D_c(R)$ in this case is excellent, and quite accurate is also the match with (6) in the regime of interest.

Consider now a second case study, namely, the classical change detection problem of zero-mean Gaussian observations with different variances. Without loss of generality, we assume that the variance under the null hypothesis is set to 1, and that pertaining to the distribution after the change is σ^2. As communication strategy, we adopt here the same repeated pairwise averaging ($v = 5$), but with the topology shown in the right plot of Fig. 2.

In Fig. 4 we report the results from 10^4 Monte Carlo iterations with $M = 10$ sensors, and a value of $\sigma = 1.032$, that is, a value very close to 1 that again leads to a difficult detection task. Comments similar to those of the previous example apply, and the match appears to be satisfying for any practical purposes.

To further highlight the benefits of the (pre-detection) data fusion achieved via running consensus, let us consider a simpler detection scheme working in flat architectures: a bank of Page's detectors that independently process the locally observed data, without any form of on-the-fly cooperation. In this case, as soon as one of these filters declares a change is the distribution of the monitored phenomenon, a broadcast message is sent to the whole system to halt the detection task, and the decision of the quickest sensor is taken as the global decision of the bank. As seen in fig. 4, the running consensus scheme largely outperforms the bank in terms of detection performance, even though this should be expected to be paid in the coin of communication burden.

5. SUMMARY

We address the problem of detecting a sudden change in the distribution of the data collected by a sensor network. With the further constraint that the network is fully flat, we propose the novel paradigm of running consensus for reaching agreement about a final decision.

We show that the test performance, *at any node*, is asymptotically (with time) equivalent to that of an ideal parallel architecture with fusion center. In addition, the running consensus allows each sensor to declare a detection at approximately the same time instant, allowing retrieval of a fully distributed and reliable estimate of the change in the state of nature from any of the deployed nodes. These characteristics are particularly attractive in applications where the network operates in dangerous environments, or when the physical topology of the system is time-varying and makes large part of the nodes unaccessible most of time.

REFERENCES

[1] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, 2008.

[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[3] R. Olfati-Saber, A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[4] J. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Trans. Autom. Control*, vol. 29, pp. 42–50, Jan. 1984.

[5] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. of the 44th IEEE Conference on Decision & Control*, 2005, pp. 8179–8184.

[6] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing statistical hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, 2010.

[7] ——, "Consensus-based Page's test in sensor networks," *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.

[8] D. Bajovic, D. Jakovetic, B. Sinopoli, J. Xavier, and J. M. F. Moura, "Distributed detection over time varying networks: Large deviations analysis," in *Proc. of the 48th Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Sept. 29 - Oct. 1 2010.

[9] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, Jan. 1954.

[10] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, N.J: Prentice-Hall, 1993.

[11] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge University Press, 2008.

[12] V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[13] A. Wald, *Sequential Analysis*. New York: Dover, 1947.

[14] A. G. Tartakovsky and I. A. Ivanova, "Approximations in sequential rules for testing composite hypotheses and their accuracy in the problem of signal detection from post-detector data," *Prob. Inform. Transm.*, vol. 28, no. 1, pp. 55–64, Jan.-Mar. 1992.