Robust Distributed Sequential Hypothesis Testing for Detecting a Random Signal in Non-Gaussian Noise

Mark R. Leonard and Abdelhak M. Zoubir
Signal Processing Group, Technische Universität Darmstadt
Merckstraße 25, 64283 Darmstadt, Germany
{leonard,zoubir}@spg.tu-darmstadt.de

***Abstract*—This paper addresses the problem of sequential binary hypothesis testing in a multi-agent network to detect a random signal in non-Gaussian noise. To this end, the consensus+innovations sequential probability ratio test ($\mathcal{CI}$SPRT) is generalized for arbitrary binary hypothesis tests and a robust version is developed. Simulations are performed to validate the performance of the proposed algorithms in terms of the average run length (ARL) and the error probabilities.**

I. Introduction

We study the problem of detecting a signal in non-Gaussian noise using a network of agents, each of which performs a sequential hypothesis test based on its own measurements and neighbor information. The test is sequential to make a decision as soon as enough data has been collected to guarantee a certain level of confidence [1]. It is distributed to avoid having a single point of failure and exploit the inherent scalability and fault-tolerance of a sensor network [2]. We resort to the consensus+innovations sequential probability ratio test ($\mathcal{CI}$SPRT) introduced in [3], [4] for Gaussian shift-in-mean problems and extend its concept for use in binary hypothesis tests with arbitrary noise distribution.

The contribution of this paper is threefold. First, we generalize the formulation of the $\mathcal{CI}$SPRT from [4] to be applicable to arbitrary binary hypothesis tests. Second, we derive a robust version dubbed R-$\mathcal{CI}$SPRT based on least-favorable densities (LFDs). To this end, we, third, give an approximation of the probability distribution of the log-likelihood ratio of the LFDs under the null hypothesis and the alternative and calculate its mean and variance.

The paper is structured as follows: In Section II we formulate the problem of detecting a signal with known variance in noise with a distributed sequential hypothesis test. Subsequently, we provide a generalized formulation of the $\mathcal{CI}$SPRT in Section III before deriving a robust extension in Section IV. Section V is dedicated to simulations and conclusions are drawn in Section VI.

II. Problem Formulation

Let $(X^1, \dots, X^n)$ be a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Their common distribution P is assumed to admit a density p. Consider a network of N agents, which can be modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the sets of agents $\mathcal{V}$ and edges $\mathcal{E}$. The open neighborhood of agent k is given by $\mathcal{N}_k = \{l \in \mathcal{V} \mid (k,l) \in \mathcal{E}\}$.

Each agent k sequentially performs a binary hypothesis test to decide between the null hypothesis $\mathcal{H}_0$ and the alternative $\mathcal{H}_1$, where

$$\begin{aligned}\mathcal{H}_0 &: P = P_0,\\ \mathcal{H}_1 &: P = P_1.\end{aligned}$$

We consider the case where each agent k should decide on the presence or absence of a signal $x(t) \sim \mathcal{N}(0, \sigma_x^2)$ using its measurement $y_k(t)$ at time instant t as well as information from its neighbors. The hypotheses become

$$\begin{aligned}\mathcal{H}_0 &: y_k(t) \sim \mathcal{N}(0, \sigma_n^2),\\ \mathcal{H}_1 &: y_k(t) \sim \mathcal{N}(0, \sigma_x^2 + \sigma_n^2),\end{aligned}$$

where σ_n^2 is the variance of a zero-mean white Gaussian noise process, which is independent of $x(t)$. Hence, the problem of detecting $x(t)$ boils down to a variance test.

In many practical applications there is an uncertainty about the distribution of the data, i.e., the assumption of Gaussianity in the measurement noise might be violated. By taking this uncertainty into account the test is transformed into a composite one between two disjoint probability sets $\mathcal{P}_0$ and $\mathcal{P}_1$ with

$$\begin{aligned}\mathcal{H}_0 &: P \in \mathcal{P}_0,\\ \mathcal{H}_1 &: P \in \mathcal{P}_1.\end{aligned}$$

Using the concept of robustness the test can be designed a priori to guarantee a certain reliability in terms of the probabilities of false alarm P_{FA} and misdetection P_{MD} for all possible probability pairs $(P_0, P_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ [5], [6]. The resulting test is again a likelihood ratio test but of the LFDs instead of the nominals. We will go into detail on this issue in Section IV.

III. A General Formulation of the $\mathcal{CI}$SPRT

In [3], [4] the authors propose a distributed sequential detector called *$\mathcal{CI}$SPRT* based on the *consensus+innovations* approach [7]. Analogous to Wald's centralized SPRT [1], each agent k in the $\mathcal{CI}$SPRT compares its test statistic $S_k(t)$ at time

ISBN 978-0-9928626-7-1 © EURASIP 2017

instant t with an upper and a lower threshold to either decide for one of the two hypotheses if the respective threshold is crossed, or continue the test. $S_k(t)$ is recursively calculated as [3], [4]

$$S_k(t) = \sum_{l \in \mathcal{N}_k \cup \{k\}} w_{kl} \left(S_l(t-1) + \eta_l(t) \right), \tag{1}$$

with w_{kl} denoting appropriate combination weights that sum up to one and

$$\eta_k(t) = \log\left(\frac{p_1(y_k(t))}{p_0(y_k(t))} \right) \tag{2}$$

being the log-likelihood ratio of agent k at time instant t. Since the decision thresholds in [4] only hold for symmetric Gaussian shift-in-mean hypothesis tests, we generalize them for use in arbitrary binary hypothesis tests. In the following, we, first, give expressions for the mean and the variance of the test statistic under both hypotheses. Subsequently, we formulate expressions for the two decision thresholds depending on the mean and the variance of the log-likelihood ratio.

A. Mean and Variance of the Test Statistic

In $\mathcal{CI}$SPRT, the distributed test statistic $S_k(t)$ at agent k and time instant t is given by [3], [4]

$$S_k(t) = \sum_{j=1}^{t} \boldsymbol{e}_k^\top \boldsymbol{W}^{t+1-j} \boldsymbol{\eta}(j), \tag{3}$$

with $\boldsymbol{e}_k$ denoting the kth column of the identity matrix of size N. Furthermore, $\boldsymbol{W}$ is the $N \times N$ matrix of combination weights and $\boldsymbol{\eta}(j) = [\eta_1(j), \ldots, \eta_N(j)]^\top$ denotes the vector of the log-likelihood ratios of all agents at time instant j. Assuming that $\eta_k(t)$ is Gaussian distributed under $\mathcal{H}_i \in \{0, 1\}$ with mean $\mu_{\eta,i}$ and variance $\sigma^2_{\eta,i}$, i.e., $\eta_k(t) \sim \mathcal{N}(\mu_{\eta,i}, \sigma^2_{\eta,i})$, the expected value of the test statistic under $\mathcal{H}_i$ is given by

$$E_i\{S_k(t)\} = \sum_{j=1}^{t} \boldsymbol{e}_k^\top \boldsymbol{W}^{t+1-j} E_i\{\boldsymbol{\eta}(j)\} = \mu_{\eta,i} t. \tag{4}$$

With $\boldsymbol{C}_{\boldsymbol{\eta}} = \sigma^2_{\eta,i} \boldsymbol{I}$, $\boldsymbol{I}$ being the identity matrix of size N, and following the derivation in [4], the variance of the test statistic under hypothesis $\mathcal{H}_i$ is upper bounded by

$$\begin{aligned} &\mathrm{Var}_i\{S_k(t)\} = E_i\{(S_k(t) - \mu_\eta t)^2\} \\ &= \sum_{j=1}^{t} \boldsymbol{e}_k^\top \boldsymbol{W}^{t+1-j} \boldsymbol{C}_{\boldsymbol{\eta}} \boldsymbol{W}^{t+1-j} \boldsymbol{e}_k \\ &\leq \sigma^2_{\eta,i} \left(\frac{t}{N} + \frac{1 - r^{2t}}{1 - r^2} \right) \leq \frac{\sigma^2_{\eta,i} t}{N} (m+1), \end{aligned} \tag{5}$$

with $m = Nr^2$ and $r = \|\boldsymbol{W} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\|$ being the rate of information flow in the network, $\|\cdot\|$ denoting the Euclidean norm and $\mathbf{1}$ the one-vector of length N.

B. Decision Thresholds

The test can easily be shown to terminate almost surely at finite stopping time T. Hence, $S_k(T)$ is well-defined and the probability of false alarm can be written as [4]

$$\begin{aligned} P_{\mathrm{FA}} &= P_0(S_k(T) \geq \gamma_u) \\ &\leq \sum_{t=1}^{\infty} P_0(S_k(t) \geq \gamma_u) \\ &\leq \sum_{t=1}^{\infty} \mathcal{Q}\left(\frac{\gamma_u - \mu_{\eta,0} t}{\sigma_{\eta,0} \sqrt{\frac{t}{N}(m+1)}} \right). \end{aligned} \tag{6}$$

Using the property $\mathcal{Q}(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}}$ and following the derivation in [4], we obtain

$$\begin{aligned} P_{\mathrm{FA}} &\leq \frac{1}{2} \sum_{t=1}^{\infty} e^{\frac{-N\gamma_u^2 - N\mu^2_{\eta,0} t^2 + 2N\gamma_u \mu_{\eta,0} t}{\sigma^2_{\eta,0} t(m+1)}} \\ &\leq \frac{2e^{\frac{7N\gamma_u}{4(m+1)} \frac{\mu_{\eta,0}}{\sigma^2_{\eta,0}}}}{1 - e^{-\frac{N}{2(m+1)} \frac{\mu^2_{\eta,0}}{\sigma^2_{\eta,0}}}}. \end{aligned} \tag{7}$$

Requiring $P_{\mathrm{FA}} \leq \alpha$ and solving for γ_u yields the upper threshold

$$\gamma_u \geq \frac{4(m+1)}{7N} \frac{\sigma^2_{\eta,0}}{\mu_{\eta,0}} \left[\log\left(\frac{\alpha}{2} \right) + \log\left(1 - e^{-\frac{N}{2(m+1)} \frac{\mu^2_{\eta,0}}{\sigma^2_{\eta,0}}} \right) \right]. \tag{8}$$

The lower threshold can be found similarly in terms of the maximally allowed probability of misdetection $P_{\mathrm{MD}} \leq \beta$ as

$$\gamma_l \leq \frac{4(m+1)}{7N} \frac{\sigma^2_{\eta,1}}{\mu_{\eta,1}} \left[\log\left(\frac{\beta}{2} \right) + \log\left(1 - e^{-\frac{N}{2(m+1)} \frac{\mu^2_{\eta,1}}{\sigma^2_{\eta,1}}} \right) \right]. \tag{9}$$

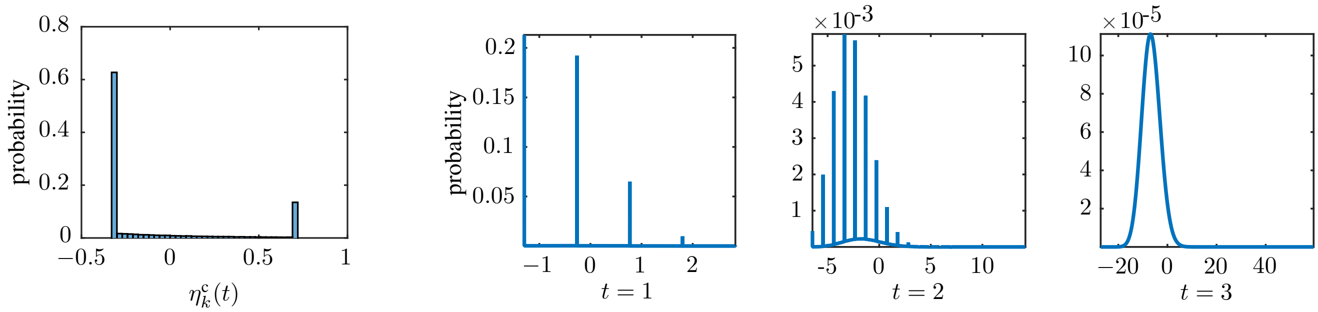
As mentioned in [4], tighter thresholds can be obtained by numerically solving

$$\begin{aligned} \frac{1}{2} \sum_{t=1}^{\infty} e^{\frac{-N\gamma_l^2 - N\mu^2_{\eta,1} t^2 + 2N\gamma_l \mu_{\eta,1} t}{2\sigma^2_{\eta,1} t(m+1)}} &= \beta \\ \frac{1}{2} \sum_{t=1}^{\infty} e^{\frac{-N\gamma_u^2 - N\mu^2_{\eta,0} t^2 + 2N\gamma_u \mu_{\eta,0} t}{2\sigma^2_{\eta,0} t(m+1)}} &= \alpha. \end{aligned} \tag{10}$$

IV. Robust Distributed Sequential Detection using the R-$\mathcal{CI}$SPRT

Having found a general formulation of the $\mathcal{CI}$SPRT for arbitrary binary hypothesis tests, we will now extend the concept to deal with composite hypotheses arising from distributional uncertainties.

ISBN 978-0-9928626-7-1 © EURASIP 2017



(a) Histogram of $\eta_k^c(t)$

(b) Probability density of $S_k^{\text{robust}}(t)$

Fig. 1: (a) Exemplary histogram of the clipped log-likelihood ratio $\eta_k^c(t)$ and (b) probability density of the robust test statistic $S_k^{\text{robust}}(t)$ of an agent with 3 neighbors at different time instances

A. Least-favorable Densities

We characterize the set of possible probabilities $\mathcal{P}_i$ under hypothesis $\mathcal{H}_i$ using Kassam's band model [8], [9] as

$$\mathcal{P}_i = \left\{ P_i \;\middle|\; p_i' \leq p_i \leq p_i'' \right\}, \tag{11}$$

such that the true density p_i is assumed to lie within a band specified by the bounds p_i' and p_i''. Using the construction algorithm stated in [9, Table 1], we can iteratively calculate the corresponding LFDs as

$$\begin{aligned} q_0 &= \min\left\{p_0'', \max\left\{c_0(\nu q_0 + q_1), p_0'\right\}\right\} \\ q_1 &= \min\left\{p_1'', \max\left\{c_1(q_0 + \nu q_1), p_1'\right\}\right\} \end{aligned} \tag{12}$$

for some $\nu \geq 0$ and some $c_0, c_1 \in (0, \frac{1}{\nu}]$. In this paper, we assume uncertainties of the ε-contamination type [10], i.e.,

$$\begin{aligned} p_i &= (1-\varepsilon)p_i^0 + \varepsilon h_i \\ &= p_i' + \varepsilon h_i, \end{aligned} \tag{13}$$

where ε is the contamination factor, and p_i^0 and h_i denote the density of the nominal and the contaminating distribution under $\mathcal{H}_i$, respectively. Furthermore, h_i is assumed to have a κ-times higher variance than p_i^0. ε-contamination is captured by the band model by setting $p_0'' = p_1'' = \infty$ and $\nu = 0$ [9]. Thus, (12) reduces to

$$\begin{aligned} q_0 &= \max\left\{c_0 q_1, p_0'\right\} \\ q_1 &= \max\left\{c_1 q_0, p_1'\right\}, \end{aligned} \tag{14}$$

which corresponds to the LFDs of Huber's famous clipped likelihood ratio test [5], [6].

B. The Robust Test Static and its Density

To design a robust version of the $\mathcal{CI}$SPRT, we replace $\eta_k(t)$ in (1) by the corresponding clipped log-likelihood ratio

$$\eta_k^c(t) = \log\left(\frac{q_1(y_k(t))}{q_0(y_k(t))}\right) \tag{15}$$

to obtain a robust test statistic $S_k^{\text{robust}}(t)$. A histogram of the probability density of $\eta_k^c(t)$ under the LFD Q_0 is shown in Fig. 1(a). It corresponds to the scaled nominal density clipped at $C_0 = -\log(c_0)$ and $C_1 = \log(c_1)$ with the excess probability being accumulated at the clipping points as

$$\begin{aligned} A_{0,i} &= Q_i(\eta_k^c(t) \leq C_0) \\ &= (1-\varepsilon)P_i^0(\eta_k^c(t) \leq C_0) + i\varepsilon \\ &= (1-\varepsilon)\mathcal{Q}\left(-\frac{C_0 - \mu_{p_i^0}}{\sigma^2_{p_i^0}}\right) + i\varepsilon \end{aligned} \tag{16}$$

$$\begin{aligned} A_{1,i} &= Q_i(\eta_k^c(t) \geq C_1) \\ &= (1-\varepsilon)P_i^0(\eta_k^c(t) \geq C_1) + (1-i)\varepsilon \\ &= (1-\varepsilon)\mathcal{Q}\left(\frac{C_1 - \mu_{p_i^0}}{\sigma^2_{p_i^0}}\right) + (1-i)\varepsilon, \end{aligned} \tag{17}$$

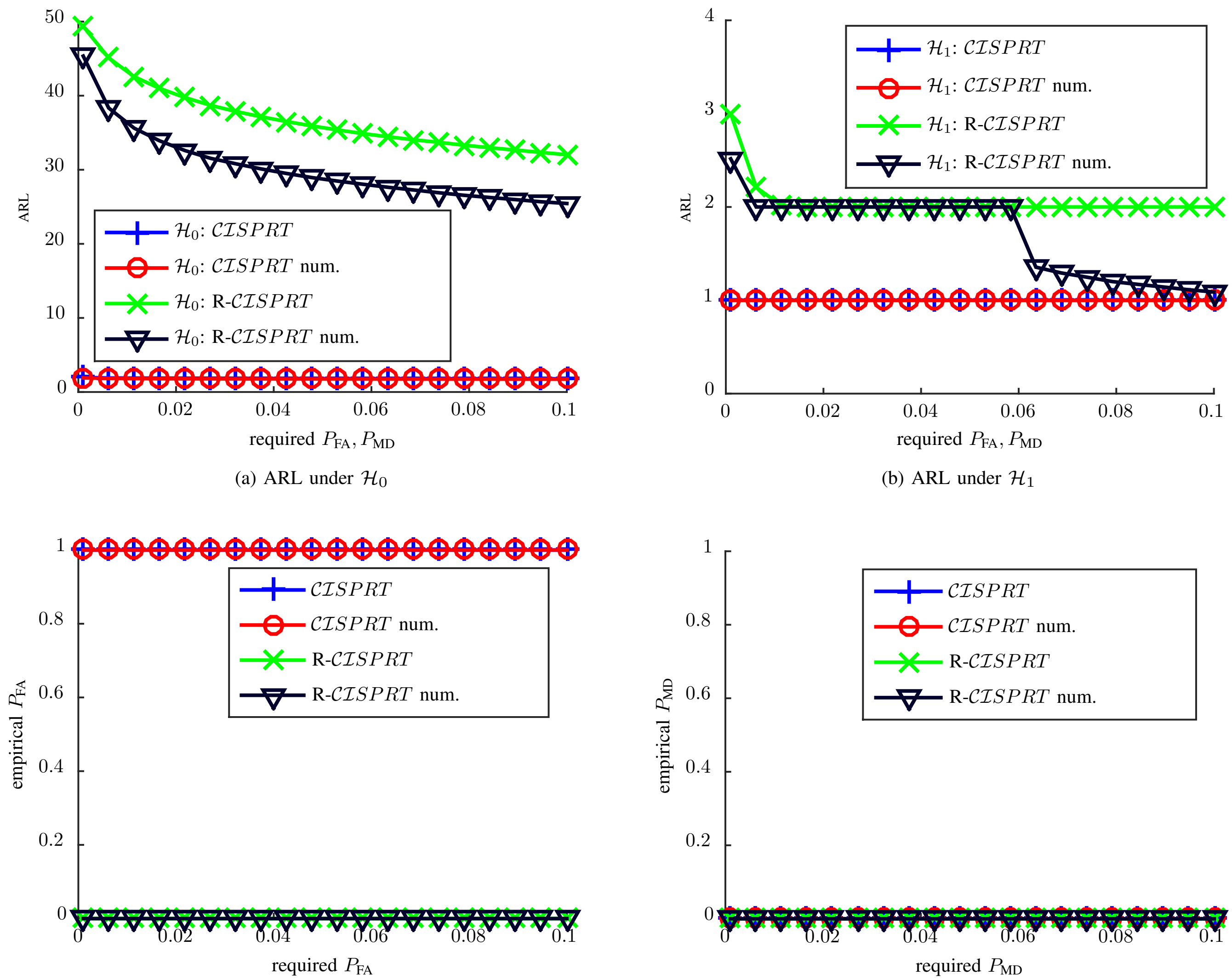
with $\mu_{p_i^0}$ and $\sigma^2_{p_i^0}$ denoting the mean and the variance of the nominal distribution under hypothesis $\mathcal{H}_i$. The probability of drawing an outlier is also placed at C_{1-i} since an outlier under $\mathcal{H}_0$ ($\mathcal{H}_1$) causes a large (small) value of $\eta_k^c(t)$.

C. Mean and Expected Value of the Robust Test Statistic

In order to calculate the mean and the variance of the robust test statistic, we have to find expressions for the mean $\mu_{\eta^c,i}$ and the variance $\sigma^2_{\eta^c,i}$ of the clipped log-likelihood ratio under $\mathcal{H}_i$. Note that the superscript k has been dropped since the distribution is equal for all agents. We approximate the probability density shown in Fig. 1(a) by two weighted Kronecker deltas at C_0 and C_1 and a weighted uniform distribution in between. With $A_{2,i} = \frac{(1-A_{0,i}-A_{1,i})}{C_1 - C_0}$, $\mu_{\eta^c,i}$ and $\sigma^2_{\eta^c,i}$ can be calculated as

$$\begin{aligned} \mu_{\eta^c,i} &= E_i\{\eta^c\} = \int_\Omega p_{\eta^c,i}(x)x\,dx \\ &= \int_{C_0}^{C_1} \left(A_{0,i}\delta(x-C_0) + A_{1,i}\delta(x-C_1) + A_{2,i}\right)x\,dx \\ &= A_{0,i}C_0 + A_{1,i}C_1 + A_{2,i}\frac{C_1^2 - C_0^2}{2} \end{aligned} \tag{18}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017



(a) ARL under $\mathcal{H}_0$ (b) ARL under $\mathcal{H}_1$ (c) Empirical P_{FA} (d) Empirical P_{MD}

Fig. 2: Results for detecting a signal in noise using $\mathcal{CI}$SPRT and R-$\mathcal{CI}$SPRT under $\mathcal{H}_0$ and $\mathcal{H}_1$ with different thresholds

$$\begin{aligned}\sigma^2_{\eta^c,i} &= E_i\{(\eta^c)^2\} - \mu^2_{\eta^c,i} = \int_\Omega p_{\eta^c,i}(x)x^2\,dx - \mu^2_{\eta^c,i} \\ &= A_{0,i}C_0^2 + A_{1,i}C_1^2 + A_{2,i}\frac{C_1^3 - C_0^3}{3} - \mu^2_{\eta^c,i}.\end{aligned} \tag{19}$$

D. Robust Decision Thresholds

Contrary to the assumption in Section III-A, the clipped log-likelihood ratio is not normally distributed. However, due to the central limit theorem we can assume the distribution of the robust test statistic $S_k^{\text{robust}}(t)$ to be approximately normal [11], [12]. Figure 1(b) shows the density of $S_k^{\text{robust}}(t)$ for an agent with three neighbors at different time instances. Due to the data exchange over the neighborhood the density becomes approximately normal already after the first few time instants. In a denser network convergence is even faster. Hence, we calculate the mean and variance of $S_k^{\text{robust}}(t)$ by replacing $\mu_{\eta,i}$ and $\sigma^2_{\eta,i}$ in (4) and (5) with their robust counterparts. Plugging the results into (8) and (9) yields the robust decision thresholds γ_l^{robust} and γ_u^{robust}.

V. Simulations

A. Simulation Setup

We consider a network of $N = 20$ agents with uniformly distributed x- and y-coordinates on the interval $[0, 1]$. Agents within a radius of $g = 0.6$ are neighbors. The required false alarm and misdetection probabilities are assumed to be equal, ranging from 10^{-3} to 10^{-1}. The maximum run length is $N_{\text{max}} = 50$. We test for a known speech signal sampled at frequency $f_s = 16$ kHz with a variance of $\sigma_x^2 = 4.5$ in the considered interval. The measurement noise is ε-contaminated with $\sigma_n^2 = 1.5$, $\varepsilon = 0.1$ and $\kappa = 10$.

To assess the performance of the R-$\mathcal{CI}$SPRT vs. the regular $\mathcal{CI}$SPRT, we evaluate the average run length (ARL), as well

ISBN 978-0-9928626-7-1 © EURASIP 2017

as the probabilities of false alarm and misdetection, P_{FA} and P_{MD}, respectively. The results are averaged over $N_{\text{MC}} = 1000$ Monte Carlo runs.

B. Simulation Results

Figures 2(a) and (b) show the ARL of the $\mathcal{CI}$SPRT and the R-$\mathcal{CI}$SPRT when $\mathcal{H}_0$ or $\mathcal{H}_1$ is true, Figures 2(c) and (d) depict the corresponding false alarm and misdetection probabilities. We use the decision thresholds calculated according to (8) and (9) as well as thresholds obtained by numerical evaluation of (10). We observe that the robust algorithm always achieves $P_{\text{FA}} = P_{\text{MD}} = 0$ while the non-robust algorithm has a false alarm probability of 1, i.e., it always assumes the signal to be present and thus fails under $\mathcal{H}_0$. The trade-off of having a robust detector is a higher ARL as can be seen in Figures 2(a) and (b). However, the ARL can be reduced by using the tighter decision thresholds.

VI. Conclusion

In this paper we generalized the $\mathcal{CI}$SPRT to be applicable to arbitrary binary hypothesis tests and developed the robust R-$\mathcal{CI}$SPRT. Our simulations showed that the robust algorithm can deal with non-Gaussian noise at the cost of a longer testing time while the non-robust algorithm breaks down.

References

[1] A. Wald, *Sequential Analysis*. New York City, New York, USA: Wiley, 1947.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, Aug 2002.

[3] A. K. Sahu and S. Kar, "Distributed sequential detection for Gaussian binary hypothesis testing: Heterogeneous networks," in *48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 723–727.

[4] ——, "Distributed sequential detection for Gaussian shift-in-mean hypothesis testing," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 89–103, Jan 2016.

[5] P. J. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965.

[6] ——, *Robust Statistics*. Hoboken, New Jersey, USA: Wiley, 1981.

[7] S. Kar and J. M. F. Moura, "Consensus + innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 99–109, May 2013.

[8] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Transactions on Information Theory*, vol. 27, pp. 242–247, 1981.

[9] M. Fauß and A. M. Zoubir, "Old bands, new tracks—Revisiting the band model for robust hypothesis testing," 2016, accepted for publication in the *IEEE Transactions on Signal Processing*. [Online]. Available: http://arxiv.org/abs/1510.04524

[10] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, July 2012.

[11] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York City, New York, USA: Springer, 2005.

[12] S. Kay, *Intuitive Probability and Random Processes Using MATLAB®*. New York City, New York, USA: Springer, 2006.

ISBN 978-0-9928626-7-1 © EURASIP 2017