Robust Detection of Signals: A Large Deviations Approach

GEORGE V. MOUSTAKIDES

Abstract—Robust detection of a signal is considered for the case of independent and identically distributed observations. Following an asymptotic but nonlocal approach, the exponential rates of decrease of the error probabilities are considered as measure of performance. Under this measure a robust detection structure for the symmetric density case is derived. This detection structure is a generalization of an existing result for the local case and is reduced to it when the signal magnitude tends to zero.

I. Introduction

Robust detection of signals in noise with uncertain statistics has been considered extensively in the literature from the local point of view. Using efficacy as a performance measure, robust detection structures have been derived for the small signal case under several noise models [1]–[5]. Clearly there is always the question of how these structures behave under nonlocal conditions and whether they remain robust. The results of [6] are applied in [1], detectors that are robust under nonlocal conditions and for the finite sample case are treated for the independent and identically distributed (i.i.d.) case. The problems with the approach in [1] are that the resulting detector is not a shift likelihood ratio for any density and that for a signal larger than a certain value the detection structure has a trivial form. Here we overcome these problems. A robust detector is found when the common noise density is symmetric. As it will be shown, there always exists a nontrivial detector that is the likelihood ratio of a least-favorable density. The result is nonlocal but asymptotic. As a measure of performance we consider the exponential rates of decrease of the error probabilities. As will be seen in Section III our result is a generalization of the local case of [1], and if we let the signal magnitude tend to zero it will yield exactly the structure that is robust under local conditions. Thus, it seems that the measure of performance introduced here is the natural extension of the efficacy in the nonlocal case.

II. Preliminaries

Let us introduce some notions from the large deviations theory for hypothesis testing. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with common density f. We would like to decide between the two hypotheses

$$\begin{aligned} H_0&: \quad X_i, \quad \text{has density } f(x) = f_0(x) \quad i = 1,2,\cdots \\ H_1&: \quad X_i, \quad \text{has density } f(x) = f_1(x) \quad i = 1,2,\cdots . \end{aligned} \tag{1}$$

We are interested in tests of the form

$$\varphi_n(\psi) = \begin{cases} 1, & \text{if } \frac{1}{n}\sum_{i=1}^{n} \psi(X_i) \geq \gamma \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\varphi_n(\psi)$ denotes the probability of deciding H_1, ψ is a function integrable with respect to $f_0(x)$ and $f_1(x)$, and γ is a threshold. It is well known that when

$$E_1\{\psi(X_1)\} > \gamma > E_0\{\psi(X_1)\} \tag{3}$$

the two error probabilities decrease exponentially to zero. Thus, it seems reasonable to use as a measure of performance the exponential rates of decrease.

Let us define $P_n^0(\psi) = E_0\{\varphi_n(\psi)\}$ and $P_n^1(\psi) = E_1\{1 - \varphi_n(\psi)\}$; i.e., $P_n^0(\psi)$ and $P_n^1(\psi)$ are the two error probabilities. Following a Neyman–Pearson type approach, let Ψ_α denote the class of all nonlinearities $\psi(x)$ that for some γ satisfy

$$-\lim_{n\to\infty} \frac{1}{n} \log P_n^0(\psi) \geq \alpha. \tag{4}$$

In other words, we consider those tests of the form of (2) that can have an exponential rate of decrease for the false-alarm probability at least equal to α. We are now interested in finding a $\psi_0 \in \Psi_\alpha$ that has the maximum possible rate for the error $P_n^1(\psi)$, i.e.,

$$-\lim_{n\to\infty} \frac{1}{n} \log P_n^1(\psi_0) \geq -\lim_{n\to\infty} \frac{1}{n} \log P_n^1(\psi) \tag{5}$$

within the constraint (4).

The parameter α is known as the *exponential level* of the test, and the rate of decrease of the probability $P_n^1(\psi)$ as the *exponential power*. It is easy to see that the exponential power can play here the same role as the efficacy in the local case. Indeed, if n_1, n_2 are the number of observations required by two different tests to reach the same power p, then if we consider the ratio of n_1 and n_2 as $p \to 1$ (which

Manuscript received November 30, 1984; revised March 18, 1985.
The author is with the Institut de Recherche en Informatique et Systemes Aleatoires, Campus Univ. de Beaulieu, 35042 Rennes Cedex, France.

0018-9448/85/1100-0822$01.00

results in $n_1, n_2 \to \infty$), we have that

$$\frac{-\lim\limits_{p\to 1} \frac{1}{n_1} \log P^1_{n_1}(\psi_1)}{-\lim\limits_{p\to 1} \frac{1}{n_2} \log P^1_{n_2}(\psi_2)} = \lim_{p\to 1} \frac{n_2}{n_1}. \tag{6}$$

We now give a lemma that specifies the optimum nonlinearity in the sense of (4) and (5).

Lemma 1: Let $f_0(x)$ and $f_1(x)$ be two densities with the same support; then the optimum nonlinearity $\psi_0 \in \Psi_\alpha$ in the sense of (4) and (5) is given by the log-likelihood ratio

$$\psi_0(x) = \log \frac{f_1(x)}{f_0(x)}, \tag{7}$$

and γ is defined in such a way that (4) is satisfied with equality.

Proof: Actually, we can prove a much stronger result. In particular, we can prove that the test defined by Lemma 1 has the largest exponential power among all tests of exponential level α and not only among those of the form of (2). The proof is an application of the Neyman–Pearson lemma. Log-likelihood ratios maximize the power sequence $1 - P^1_n(\psi)$ for any sequence of levels $P^0_n(\psi) \le \alpha_n$. Thus they also maximize monotone transformations of the power like $-1/n \log P^1_n(\psi)$.

We now present a lemma that defines more explicitly the two rates for a test of the form of (2) in terms of the two densities and the nonlinearity ψ.

Lemma 2: Let f_0 and f_1 be two densities with the same support, and let ψ be a nonlinearity that is integrable with respect to f_0 and f_1. Also let γ be a real number that satisfies

$$E_1\{\psi(X_1)\} > \gamma > E_0\{\psi(X_1)\}. \tag{8}$$

Then we have that

$$\begin{aligned} A_0(\psi, f_0) &\triangleq -\lim_{n\to\infty} \frac{1}{n} \log P^0_n(\psi) \\ &= \min_{r\ge 0}\left[r\gamma - \log E_0\{e^{r\psi(X_1)}\}\right] \\ A_1(\psi, f_1) &\triangleq -\lim_{n\to\infty} \frac{1}{n} \log P^1_n(\psi) \\ &= \min_{r\ge 0}\left[-r\gamma + \log E_1\{e^{-r\psi(X_1)}\}\right]. \end{aligned} \tag{9}$$

Proof: This lemma is known as Cramer's theorem. (For a proof see [8].)

The threshold γ must satisfy (8) in order to have exponential decrease for the two error probabilities and the validity of (9). This requirement bounds the possible values of the exponential level. We can see from (9) that the exponential level is increasing with γ and thus the maximum value it can take is when $\gamma = E_1\{\psi(x)\}$. For this value the error probability under H_1 has rate equal to zero, i.e., we do not have exponential decrease. We now apply these results to the robust-detection theory.

III. Robust Detection

Let $N_1, N_2, \cdots$ be an i.i.d. noise sequence with common density f. We would like to decide between the two hypotheses

$$\begin{aligned} &H_0: \quad X_i = N_i + s_0, \qquad s_0 \in (-\infty, 0], \quad i = 1, 2, \cdots \\ &H_1: \quad X_i = N_i + s_1, \qquad s_1 \in [s, \infty), \quad i = 1, 2, \cdots, \end{aligned} \tag{10}$$

where $\{X_i\}$ is the observation sequence, s_0, s_1 are unknown, and $s > 0$ is known.

Let F be the class of all symmetric densities that satisfy the following ϵ-contamination model:

$$f(x) = (1-\epsilon)g(x) + \epsilon h(x), \tag{11}$$

where $0 \le \epsilon < 1$ is known, $g(x)$ is a known symmetric nowhere-vanishing density, such that $-\log g(x)$ is strictly convex. The density $h(x)$ is assumed to be symmetric but unknown. Let Ψ_α denote the class of all nonlinearities ψ for which there exist a test of the form of (2) satisfying the following for every $f \in F$:

$$A_0(\psi, f) \ge \alpha. \tag{12}$$

We would like to find a density $f_l \in F$ and a $\psi_r \in \Psi_\alpha$ such that

$$A_1(\psi_r, f) \ge A_1(\psi_r, f_l) \ge A_1(\psi, f_l) \tag{13}$$

and also

$$A_0(\psi_r, f) \ge A_0(\psi_r, f_l) = \alpha. \tag{14}$$

The right side inequality of (13) and the right side equality of (14), using Lemma 1, suggest that ψ_r is the log-likelihood ratio

$$\psi_r(x) = \log \frac{f_l(x+s_1)}{f_l(x+s_0)} \tag{15}$$

for some s_1 and s_0. We now define the density f_l by

$$f_l(x) = \begin{cases} (1-\epsilon)g(x), & \text{for } 0 \le x \le x_0 \\ \dfrac{(1-\epsilon)}{k^n} g(x - ns), & \text{for } x_0 + (n-1)s \\ & \quad \le x \le x_0 + ns, \\ & \quad n = 1, 2, \cdots, \end{cases} \tag{16}$$

where $k = g(x_0 - s)/g(x_0)$ and $x_0 > s/2$ is selected in order to have

$$\int_0^\infty f_l(x)\,dx = (1-\epsilon)\left[\int_0^{x_0} g(x)\,dx + \frac{g(x_0)}{g(x_0 - s) - g(x_0)} \int_{x_0 - s}^{x_0} g(x)\,dx\right] = 0.5. \tag{17}$$

A typical form of $f_l(x)$ is given in Fig. 1. In the Appendix it is shown that an x_0 always exists and that it is unique and also that $f_l \in F$. From the definition in (16) notice that $\epsilon h_l(x) = f_l(x) - (1-\epsilon)g(x)$ puts all its mass outside the interval $(-x_0, x_0)$. Let us now see that the form of ψ_r defined by (15) when f_l is defined by (16) and $s_0 = 0$,

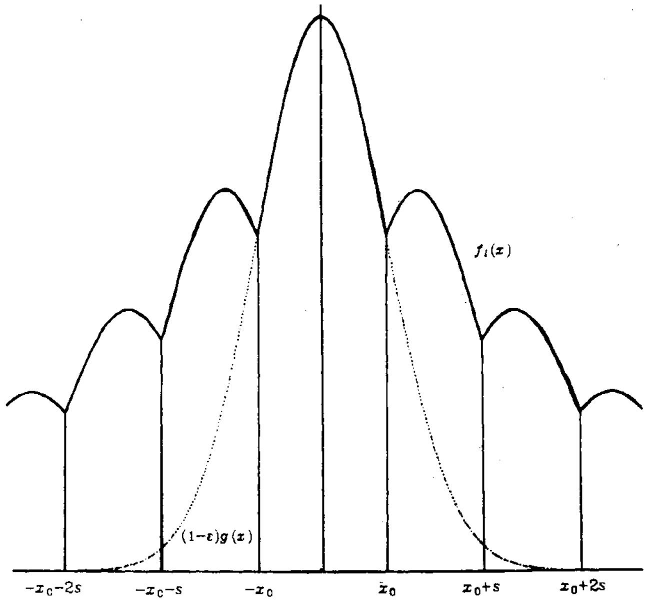

Fig. 1. Typical form of least favorable density.

$s_1 = s$ is

$$\psi_r(x) = \begin{cases} \log \dfrac{g(x_0 - s)}{g(x_0)}, & \text{for } x \geq x_0 \\ \log \dfrac{g(x-s)}{g(x)}, & \text{for } x_0 \geq x \geq -x_0 + s. \\ -\log \dfrac{g(x_0 - s)}{g(x_0)}, & \text{for } -x_0 + s \geq x \end{cases} \tag{18}$$

The following theorem proves that f_l and ψ_r is the pair that satisfies (13) and (14).

Theorem 1: When f_l and ψ_r are defined by (16) and (18), they satisfy (13) and (14).

The proof is given in the Appendix.

As we can see in Fig. 1, the least favorable density f_l repeats a piece of length s of the density g after dividing it every time with the constant k. By taking $s \to 0$, (17) reduces to

$$\int_0^\infty f_l(x)\,dx = (1-\epsilon)\left[\int_0^{x_0} g(x)\,dx - \frac{g(x_0)^2}{g'(x_0)}\right] = 0.5, \tag{19}$$

which is the equation for determining the point x_0 for the local case [1], [7]. Since multiplying ψ_r by a positive constant does not change the performance of the test, if we also multiply the threshold by the same constant, we have that $1/s\psi_r(x)$ is also optimum in our sense. If we now take $s \to 0$, we can see that we recover the locally robust nonlinearity defined in [1] and [7].

IV. Examples

As an example, we present the Gaussian nominal case. Clearly the robust nonlinearity ψ_r will be linear inside the interval $[-x_0 + s, x_0]$. In Table I values of x_0 are given

TABLE I
Values of x_0 for Different Values of ϵ and s

s \ ϵ	0.001	0.01	0.1	0.5
0.1	2.682	1.994	1.190	0.486
0.5	2.856	2.175	1.378	0.682
1.0	3.032	2.369	1.595	0.919
2.0	3.297	2.684	1.980	1.374
5.0	3.973	3.536	3.080	2.717
10.0	5.760	5.529	5.294	5.110

TABLE II
Values of the Two Rates for $s = 1$ and Different Values of ϵ and γ

γ \ ϵ	0.001	0.01	0.1	0.5
0.0	0.000	0.000	0.000	0.000
	0.492	0.459	0.318	0.078
0.1	0.005	0.005	0.003	0.001
	0.391	0.372	0.258	0.063
0.2	0.020	0.019	0.013	0.003
	0.315	0.295	0.204	0.050
0.3	0.045	0.042	0.030	0.007
	0.241	0.226	0.157	0.038
0.4	0.079	0.075	0.052	0.013
	0.178	0.167	0.116	0.028
0.5	0.124	0.116	0.081	0.020
	0.124	0.116	0.08	0.020

for different values of the contamination ϵ and the signal s. Table II contains the exponential level and the worse-case exponential power for the case $s = 1$ and for different values of γ and ϵ. It is assumed that the mean of $\psi_r(X_1)$ has been normalized to zero under H_0 and to unity under H_1, and that γ takes values in the interval $[0, 0.5]$. For values of γ in the interval $[0.5, 1]$, the table is symmetric in that the exponential level at $\gamma > 0.5$ is equal to the worst exponential power at $1 - \gamma$, and the worst exponential power is equal to the exponential level.

V. Conclusion

We have presented a detection structure that is robust to partial knowledge of the signal magnitude and of the noise distribution function. The result is asymptotic but nonlocal. The advantage of this approach is that the robust detector is a likelihood ratio for a specific density and is always nontrivial, something which is not true for all existing approaches. It will be interesting to see if this approach also applies to the case where the densities are symmetric only inside an interval around the origin, thus generalizing the result in [3].

Appendix

Proof that $f_l \in F$

We first prove existence of an $x_0 > s/2$ that satisfies (17). Define as $B(x_0)$ the function

$$\begin{aligned} B(x_0) &= \frac{1}{1-\epsilon}\int_0^\infty f_l(x)\,dx \\ &= \int_0^{x_0} g(x)\,dx + \frac{1}{l(x_0) - 1}\int_{x_0-s}^{x_0} g(x)\,dx, \end{aligned} \tag{20}$$

where $l(x) = g(x-s)/g(x)$. Because the function $-\log g(x)$ is convex, the function l is strictly increasing and thus $l(x) > 1$ for $x > s/2$. Notice that

$$\lim_{x_0 \to (s/2)+} B(x_0) = \infty > \frac{1}{2(1-\epsilon)}$$
$$\lim_{x_0 \to \infty} B(x_0) = \frac{1}{2} \le \frac{1}{2(1-\epsilon)}. \quad (21)$$

Using continuity arguments there exists an x_0 that satisfies $B(x_0) = 1/2(1-\epsilon)$. The uniqueness can be easily deduced by taking the derivative of $B(x_0)$; this derivative is always negative for $x_0 > s/2$.

To prove now that f_l belongs to the class F, it is enough to show that

$$f_l(x) \ge (1-\epsilon) g(x). \quad (22)$$

This inequality is trivial for the case $0 \le x \le x_0$. For the case $x_0 + (n-1)s \le x \le x_0 + ns$ it is equivalent, using (16), to

$$\frac{-\log g(x_0 - s) + \log g(x_0)}{s} \ge \frac{-\log g(x - ns) + \log g(x)}{ns}, \quad (23)$$

and since $-\log g(x)$ is convex and $x - ns \ge x_0 - s$, the inequality in (23) is true.

Proof of Theorem 1

Before proving that f_l and ψ_r satisfy (13) and (14), we first prove a lemma.

Lemma 3: Let ω be a nondecreasing function such that $\omega(x) + \omega(-x)$ is also nondecreasing for $x \ge 0$. If $f \in F$, $s_0 \le 0$, and $s_1 \ge s$, then

$$\int_{-\infty}^{\infty} \omega(\psi_r(x)) f_l(x)\, dx \ge \int_{-\infty}^{\infty} \omega(\psi_r(x)) f(x - s_0)\, dx \quad (24a)$$
$$\int_{-\infty}^{\infty} \omega(\psi_r(x)) f_l(x - s)\, dx \le \int_{-\infty}^{\infty} \omega(\psi_r(x)) f(x - s_1)\, dx. \quad (24b)$$

Proof: We only prove the first inequality since in a similar way we can prove the second. Notice first some important properties of the function ψ_r defined in (18). It is nondecreasing with x, the function $\psi_r(x + s/2)$ is odd symmetric nondecreasing and, for $x \ge 0$, it is nonnegative. Notice also that the density $h_l(x)$ puts all its mass on points where $\psi_r(x)$ is maximum. Since ψ_r and ω are nondecreasing their composition is also nondecreasing. Thus

$$\int_{-\infty}^{\infty} \omega(\psi_r(x)) f(x - s_0)\, dx = \int_{-\infty}^{\infty} \omega(\psi_r(x + s_0)) f(x)\, dx \le \int_{-\infty}^{\infty} \omega(\psi_r(x)) f(x)\, dx. \quad (25)$$

Using (25) in order to prove (24a) it is enough to prove that

$$\int_{-\infty}^{\infty} \omega(\psi_r(x)) f_l(x)\, dx \ge \int_{-\infty}^{\infty} \omega(\psi_r(x)) f(x)\, dx, \quad (26)$$

or, by eliminating common terms,

$$\int_{-\infty}^{\infty} \omega(\psi_r(x)) h_l(x)\, dx \ge \int_{-\infty}^{\infty} \omega(\psi_r(x)) h(x)\, dx. \quad (27)$$

Since $h_l(x)$ puts its mass on points where $|\psi_r(x)|$ is maximum, we can see that (27) is equivalent to

$$\frac{\omega(M) + \omega(-M)}{2} \ge \int_{-\infty}^{\infty} \omega(\psi_r(x)) h(x)\, dx, \quad (28)$$

where M is the maximum value of $\psi_r(x)$. Notice now that

$$\int_{-\infty}^{\infty} \omega(\psi_r(x)) h(x)\, dx \le \int_{-\infty}^{\infty} \omega\left(\psi_r\left(x + \frac{s}{2}\right)\right) h(x)\, dx = \int_0^{\infty} \left[\omega\left(\psi_r\left(x + \frac{s}{2}\right)\right) + \omega\left(-\psi_r\left(x + \frac{s}{2}\right)\right)\right] h(x)\, dx \le \frac{\omega(M) + \omega(-M)}{2}. \quad (29)$$

The last equality comes from the fact that $\psi_r(x + s/2)$ is odd symmetric. The last inequality is true because $\omega(x) + \omega(-x)$ is by assumption nondecreasing for $x \ge 0$ and $\psi_r(x + s/2)$ is nonnegative for $x \ge 0$. Thus (28) is true.

To prove the theorem we apply Lemma 3. Selecting $\omega(x) = x$ we have from (24a) and (24b) that $\psi_r(X_1)$ has the maximum mean for f_l under H_0 and the minimum under H_1. This is important because if we take the threshold γ between these two means, then we are assured that we will have exponential decrease for both errors for any density $f \in F$. To show now the inequalities in (13) and (14), we first show that they are equivalent. Notice that to show any of the two, using (9), it is enough to show that for any $r \ge 0$ we have

$$\int_{-\infty}^{\infty} e^{-r\psi_r(x)} f(x - s_1)\, dx \le \int_{-\infty}^{\infty} e^{-r\psi_r(x)} f_l(x - s)\, dx$$
$$\int_{-\infty}^{\infty} e^{r\psi_r(x)} f(x - s_0)\, dx \le \int_{-\infty}^{\infty} e^{r\psi_r(x)} f_l(x)\, dx. \quad (30)$$

By change of variables and using the symmetries of ψ_r and f, we can see that

$$\int_{-\infty}^{\infty} e^{-r\psi_r(x)} f(x - s_1)\, dx = \int_{-\infty}^{\infty} e^{r\psi_r(x)} f(x - s_0')\, dx, \quad (31)$$

where $s_0' = s - s_1 \le 0$. Thus the first inequality is equivalent to the second. The second inequality is true by a simple application of Lemma 3 with $\omega(x) = e^{rx}$. This concludes the proof.

References

[1] R. D. Martin and S. C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50–56, Jan. 1971.

[2] S. A. Kassam and J. B. Thomas, "Asymptotically robust detection of a known signal in contaminated non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 22–26, Jan. 1976.

[3] S. A. Kassam, G. V. Moustakides, and J. G. Shin, "Robust detection of known signals in asymmetric noise," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 84–91, Jan. 1982.

[4] H. V. Poor, "Signal detection in the presence of weakly dependent noise—Part II: Robust detection," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 744–752, Sept. 1982.

[5] G. V. Moustakides and J. B. Thomas, "Min-max detection of weak signals in φ-mixing noise," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 529–537, May 1984.

[6] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753–1758, Dec. 1965.

[7] ——, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, Mar. 1964.

[8] D. W. Stroock, *An Introduction to the Theory of Large Deviations.* New York: Springer-Verlag, 1984.