Two results in statistical decision theory for detecting signals with unknown distributions and priors in white Gaussian noise.

Dominique Pastor

GET - ENST Bretagne,
CNRS UMR 2872 TAMCIC,
Technopôle de Brest Iroise, CS 83818,
29238 BREST Cedex, France
(e-mail: dominique.pastor@enst-bretagne.fr)

Abstract. Two recent advances in statistical decision and estimation theory are presented. These results concern the detection of signals whose amplitudes are above or equal to some bound and that are less present than absent in a background of white Gaussian noise. The first result describes the non parametric detection of such signals when the noise standard deviation is known whereas the second result affords to perform the detection when this standard deviation is unknown. For both results, the role played by thresholding tests on the observation norms is crucial. The detection of radar targets is a typical field of application of these results.
Keywords: Estimation theory, Likelihood theory, Limit theorem, Non parametric decision, Thresholding test.

1 A sharp upper-bound for the probability of error of the MPE decision scheme and the MPE suboptimal test.

Albeit simple, a reasonable model for observations performed by sensors is that of signals randomly present or absent in additive and independent white Gaussian noise (WGN). In contrast with the simplicity of this model, the detection of such signals on the basis of a set of observations can be intricate. Actually, in many applications of most importance, very little is known about the observations or most of their parameters ([Kailath and Poor, 1998, section I]). In such situations, the detection of signals of interest cannot be achieved by standard likelihood theory based on the usual Bayes, minimax and Neyman-Pearson criteria for these ones require full knowledge of the signal distributions. Nonparametric and robust detection ([Poor, 1994, section III.E]), as well as Generalized Likelihood Ratio Tests ([Kay, 1998]), are then alternative formulations affording to deal with such cases. For instance, Constant False Alarm Rate (CFAR) systems standardly used in radar processing for detecting targets with a specified false alarm rate typically derive from such alternative approaches ([Minkler and Minkler, 1990]).

In [Pastor *et al.*, 2002], we investigate how far we can get if we assume only two hypotheses on the signal to detect. First, the signal is supposed to be less present than absent in the sense that its prior is less than or equal to one half; second, the norm of this same signal is assumed to be larger than or equal to some positive real number A. The purpose of such assumptions is to bound our lack of prior knowledge. The following theorem is then established in [Pastor *et al.*, 2002]. In the statement of this theorem, $\mathbf{I}_n$ stands for the identity matrix with size $n \times n$; by thresholding test with threshold height T, we mean the binary hypothesis test whose decision is that some signal is present if the observation norm exceeds T and whose decision is that noise only is present otherwise; finally, we remind the reader that the so-called MPE decision scheme is basically the likelihood ratio test that yields the least probability of error amongst all possible binary hypothesis tests ([Poor, 1994]).

Theorem 1 *Let $U, \Lambda, X : \Omega \to \mathbf{R}^n$ be three random vectors and let $\epsilon : \Omega \to \{0,1\}$ be a random variable defined on the same probability space $(\Omega, \mathcal{B}, P)$ such that Λ, X and ϵ are independent, $X \in \mathcal{N}(0, \sigma_0^2 I_n)$ and $U = \epsilon\Lambda + X$.*

Let $V(\rho)$ be the function of the positive real variable ρ

$$V(\rho) = \frac{e^{-\rho^2/2}}{2^{n/2}\Gamma(n/2)} \int_0^{\xi(\rho)} e^{-t^2/2} t^{n-1} {}_0F_1(n/2\,;\rho^2 t^2/4) dt + \frac{1}{2}\left[1 - \frac{2^{1-n/2}}{\Gamma(n/2)} \int_0^{\xi(\rho)} e^{-t^2/2} t^{n-1} dt\right]. \tag{1}$$

where $\xi(\rho)$ is the unique positive solution for x in the equation

$${}_0F_1(n/2; \rho^2 x^2/4) = e^{\rho^2/2}. \tag{2}$$

Then, given any positive real number $A > 0$, for any Λ less present than absent with norm almost surely larger than or equal to A, $V(A/\sigma_0)$ is an upper-bound for the probability of error of both the MPE decision scheme and the threshold test with threshold height $\sigma_0\xi(A/\sigma_0)$. This bound is reached by both tests when the prior $P(\{\varepsilon = 1\})$ equals $1/2$ and Λ is uniformly distributed on the sphere with radius A centred at the origin.

The thresholding test with threshold height $\sigma_0\xi(A/\sigma_0)$ is hereafter called the MPE suboptimal test. It is basically nonparametric in the sense given by [Poor, 1994] since $V(A/\sigma_0)$ is the constant performance measurement this test guarantees over the whole class of those signals less present than absent with norms larger than or equal to A.

2 Detection of relatively big signals in WGN with unknown level: the Essential Supremum Test.

The thresholding test introduced by theorem 1 is workable in practice only if the noise standard deviation is known. On the basis of theorem 2 stated in subsection 2.2 below, subsection 2.3 then introduces an algorithm named the Essential Supremum Test (EST) and aimed at detecting signals of interest even if the noise standard deviation is unknown. Beforehand, we need some appropriate notations and pieces of terminology.

2.1 Some notations.

The characteristic function of a set K will be denoted by $\mathcal{X}_K$: $\mathcal{X}_K(x) = 1$ if $x \in K$ and $\mathcal{X}_K(x) = 0$ otherwise. A real number x (resp. an integer k) is said to be positive if $x > 0$ (resp. $k > 0$). The real number x (resp. the integer k) is said to be non negative if $x \geq 0$ (resp. $k \geq 0$).

Only one probability space $(\Omega, \mathcal{M}, P)$ is considered in what follows.

Given any positive integer n, $\|\cdot\| : \mathbf{R}^n \to [0, \infty)$ will stand for the usual euclidean norm on $\mathbf{R}^n$. Given any n-dimensional random vector $Y : \Omega \to \mathbf{R}^n$, $\|Y\|$ will stand for the random variable $\|Y\| : \Omega \to [0, \infty)$ that assigns the non negative real number $\|Y(\omega)\|$ to every given $\omega \in \Omega$.

Let $\mathbf{S}$ henceforth stands for the set of all the sequences of n-dimensional real random vectors defined on Ω. Given some positive real number σ_0 and some natural number n, an element $X = (X_k)_{k \in \mathbf{N}}$ of $\mathbf{S}$ will be called an *n-dimensional WGN with standard deviation σ_0* if the random vectors X_k, $k \in \mathbf{N}$, are mutually independent and identically Gaussian distributed with null mean vector and covariance matrix $\sigma_0^2 \mathbf{I}_n$ $(X_k \mathrm{sim} \mathcal{N}(0, \sigma_0^2 \mathbf{I}_n))$.

As usual, we denote by $L^2(\Omega, \mathbf{R}^n)$ the Hilbert space of those n-dimensional real random vectors $Y : \Omega \to \mathbf{R}^n$ such that $E[\|Y\|^2] < \infty$. We will hereafter deal with the set of those elements $\Lambda = (\Lambda_k)_{k \in \mathbf{N}}$ of $\mathbf{S}$ such that $\Lambda_k \in L^2(\Omega, \mathbf{R}^n)$ for every $k \in \mathbf{N}$ and $\sup_{k \in \mathbf{N}} E[\|\Lambda_k\|^2]$ is finite. According to standard notations, we denote this subset of $\mathbf{S}$ by $\ell^\infty(\mathbf{N}, L^2(\Omega, \mathbf{R}^n))$.

2.2 A limit theorem

The subsequent theorem derives from a more general result established in [Pastor, 2004] and suffices for achieving our purpose, that is introducing the EST.

Theorem 2 *Let $U = (U_k)_{k \in \mathbf{N}}$ be some element of $\mathbf{S}$ such that $U = \varepsilon \Lambda + X$ where $\Lambda = (\Lambda_k)_{k \in \mathbf{N}}$, $X = (X_k)_{k \in \mathbf{N}}$ and $\varepsilon = (\varepsilon_k)_{k \in \mathbf{N}}$ are respectively an element of $\mathbf{S}$, some n-dimensional WGN with standard deviation σ_0 and a sequence of random variables valued in $\{0, 1\}$.*

Assume that

(H1) *for every $k \in \mathbf{N}$, Λ_k, X_k and ε_k are mutually independent;*

(H2) *the random vectors U_k, $k \in \mathbf{N}$, are mutually independent;*

(H3) *the set of priors $\{\{P(\{\varepsilon_k = 1\}) : k \in \mathbf{N}\}$ has a maximum p in $[0, 1)$ and the random variables ε_k, $k \in \mathbf{N}$, are mutually independent;*

(H4) *$\Lambda \in \ell^\infty(\mathbf{N}, L^2(\Omega, \mathbf{R}^n))$ and there exists $A \in (0, \infty)$ such that, for every $k \in \mathbf{N}$, $\|\Lambda_k\| \geq A$ almost surely.*

Then, σ_0 is the only strictly positive real number σ, such that, for every $\beta_0 \in (0, 1]$,

$$\lim_{A \to \infty} \left\| \overline{\lim_{m}} \left| \frac{\displaystyle\sum_{k=1}^{m} \|U_k\| \mathcal{X}_{[0, \beta\sigma\xi(A/\sigma)]}(\|U_k\|)}{\displaystyle\sum_{k=1}^{m} \mathcal{X}_{[0, \beta\sigma\xi(A/\sigma)]}(\|U_k\|)} - \sigma G_n\left(\beta\xi(A/\sigma)\right) \right| \right\|_\infty = 0 \qquad (3)$$

uniformly in $\beta \in [\beta_0, 1]$ where, for every non negative real value x,

$$G_n(x) = \frac{\displaystyle\int_0^x t^n e^{-t^2/2} dt}{\displaystyle\int_0^x t^{n-1} e^{-t^2/2} dt}.$$

2.3 The Essential Supremum Test

Let L be some natural number and set $\beta_\ell = \ell/L$ for every $\ell \in \{1, \ldots, L\}$. On the basis of theorem 2, given some elementary event $\omega \in \Omega$ and m vectors $U_1(\omega), \ldots, U_m(\omega)$, the idea is then to estimate σ_0 by an eventually local minimum $\hat{\sigma}_0(m, \omega)$ of

$$\sup_{\ell \in \{1, \ldots, L\}} \left\{ \left| \frac{\displaystyle\sum_{k=1}^{m} \|U_k(\omega)\| \mathcal{X}_{[0, \beta_\ell\sigma\xi(A/\sigma)]}(\|U_k(\omega)\|)}{\displaystyle\sum_{k=1}^{m} \mathcal{X}_{[0, \beta_\ell\sigma\xi(A/\sigma)]}(\|U_k(\omega)\|)} - \sigma G_n\left(\beta_\ell\xi(A/\sigma)\right) \right| \right\}, \quad (4)$$

when σ runs through the search interval $(0, \sigma_{max}(m, \omega)]$ where

$$\sigma_{\max}(m, \omega) = \sup_{k \in \{1, \ldots, m\}} \{\|U_k(\omega)\|\}/\sqrt{n}.$$

When σ runs through the search interval proposed above, the discrete cost (4) is a scalar bounded nonlinear function of σ. We thus seek an eventual local minimum of the discrete cost (4) by means of a standard minimization routine such as the golden section search and parabolic interpolation ([Forsythe *et al.*, 1976] and [Press *et al.*, 1992]). Given $k \in \mathbf{N}$, the decision on the value of ε_k is then achieved by replacing, in the expression of the MPE suboptimal test, the exact value of σ_0 by its estimate. The resulting binary hypothesis test is then the map of Ω into $\{0, 1\}$ defined by $\hat{\mathcal{T}}_k = \mathcal{X}_{[0,\infty)}\left(\|U_k\| - \hat{\sigma}_0(m, \omega)\xi(A/\hat{\sigma}_0(m, \omega))\right)$.

Our choice for the search interval upper bound is then justified as follows. If $\hat{\sigma}_0(m,\omega)$ might be larger than $\sigma_{\max}(m,\omega)$, we would take the risk to get an estimate larger than every ratio $\|U_k(\omega)\|/\xi(A/\sigma_{\max}(m,\omega))$, when $k \in \{1,\ldots,m\}$. Indeed, $\xi(\rho) \geq \sqrt{n}$ ([Pastor *et al.*, 2002]) for all non negative real value ρ. Thereby, the outcome of the test $\hat{\mathcal{T}}_k$ could be that no signal is present whereas the full absence of signals of interest amongst m observations is hardly probable when m is large.

2.4 Some experimental results

The performance of the EST should be less than that of the MPE suboptimal test. However, when m and A increase, we also can expect that the performance measurements of these two tests become close to each other. If so, above which values for m and A can the essential supremum test be considered as workable in practice? Till now, we have no theoretical answer to this question and it seems hardly feasible to get an experimental answer to it because we simply do not known which priors and distributions to choose for such experiments? Therefore, in this section, we will be satisfied with some experimental results concerning the following basic case.

With the same notations as those used so far, we suppose that for every given $k \in \mathbf{N}$, U_k, Λ_k and X_k are two-dimensional random vectors ($n = 2$) where Λ_k is uniformly distributed on the circle centred at the origin with radius A. We further assume that $P(\varepsilon_k = 1\}) = 1/2$. Given $k \in \mathbf{N}$, the two components of Λ_k can be regarded as the in-phase and quadrature components of a sinusoidal carrier with amplitude A and phase uniformly distributed in $[0, 2\pi]$. Thereby, deciding whether ε_k equals 0 or 1 is the standard "Non coherent Detection of a Modulated Sinusoidal Carrier" problem ([Poor, 1994, Example III.B.5, p. 65]). The MPE decision scheme for making a decision on the value of ε_k is the thresholding test whose threshold height is the unique solution in x to the equation $I_0\left(A/\sigma_0 x\right) = e^{A^2/2\sigma_0^2}$, where I_0 is the zeroth order modified Bessel function of the first kind ([Poor, 1994, Example II.E.1, p. 34]). Since $I_0(x) = {}_0F_1(1; x^2/4)$, the reader will easily verify that the result is also a straighforward consequence of theorem 1.

Suppose now that the noise standard deviation is unknown. If we dispose of m observations U_k, $k = 1, \ldots, m$, we can estimate this standard deviation by minimizing the discrete cost (4) on the basis of those m references. This estimate can then be used for tuning the EST and, on the basis of the $(m+1)$th observation U_{m+1}, make a decision on the value of ε_{k+1}. This decision making has a certain probability of error $\hat{V}_m(A/\sigma_0)$. If m and A are large enough, $\hat{V}_m(A/\sigma_0)$ and $V(A/\sigma_0)$ are expected to draw near to each other. In other words, when m and $\rho \in (0,\infty)$ are large enough, $\hat{V}_m(\rho)$ and $V(\rho)$ should be close to each other. We thus carry out simulations so as to experimentally verify this intuitive claim.

In these simulations, $\sigma_0 = 1$ for this choice induces no loss of generality; given $\rho \in (0,\infty)$, $\hat{V}_m(\rho)$ is computed by choosing signals uniformly distributed on the sphere centred at the origin with radius ρ. We minimize the discrete cost (4) with $L = m$ as a trade-off between accuracy of the estimate and computational cost. Given $m \in \mathbf{N}$ and $\rho \in (0, \infty)$, we approximate $V_m(\rho)$ by the EST Binary Error Rate (BER), computed as follows. Given $j \in \mathbf{N}$, the EST estimates σ_0 on the basis of the m observations $U_{(j-1)(m+1)+k}$, $k = 1, 2, \ldots, m$ and makes a decision on the value of $\varepsilon_{j(m+1)}$. If I_j stands for the indicator variable defined by $I_j = 1$ if the EST makes the wrong decision on the value of $\varepsilon_{j(m+1)}$ and by $I_j = 0$ otherwise, the random variables I_j, $j \in \mathbf{N}$, are mutually independent because of the mutual independence of the trials. It turns out that estimating $\hat{V}_m(\rho)$ by the sample proportion S_k/k, where $S_k = \sum_{j=1}^{k} I_j$ and k is some specified number of trials, is not suitable with respect to our purpose. Indeed, $\hat{V}_m(\rho)$ is expected to approximate reasonably well $V(\rho)$ for large values of m and ρ; now, $V(\rho)$ rapidly decreases with ρ; hence, the accuracy of the sample proportion S_k/k may significantly depend on the value of $\hat{V}_m(\rho)$. Thence, we resort to inverse binomial sampling as practitioners in telecommunication systems usually do since error probabilities also decrease rapidly with input signal to noise ratios. The BER is thus defined as the ratio i/K where $K = \inf\{k \in \mathbf{N} : S_k = i\}$ is the minimum number of trials experimentally required for achieving a predefined number of errors equal to i.

Figures 1 to 3 present experimental results obtained for different values for m. Each figure displays $V(\rho)$ and the BERs of the EST for $\rho = 0.5, 1, 1.5, \ldots, 5$ and a pre-specified number of errors i equal to 400, which is a reasonable choice according to practitioners in telecommunication systems. As expected, the larger m and ρ, the closer $V(\rho)$ and $\hat{V}_m(\rho)$.

Consider now the case of sinusoidal carriers with amplitudes all equal to $C\rho$ with $C > 1$. According to theorem 2, the least we can expect is that the larger C, the better the performance of the EST. For instance, the results displayed in figure 4 were obtained for $m = 300$ and signals of interest with amplitude A one dB larger than the value ρ, that is $A = 1.2589\rho$. These results strongly suggest that the asymptotic conditions of theorem 2 are not so constraining in practice and can probably be relaxed.

3 Perspectives and extensions

Forthcoming work should address the respective influence of the EST various parameters, analyse how the asymptotic conditions of theorem 2 can actually be relaxed and assess the quality of EST estimate of the noise standard deviation.

A natural application of the approach presented in this paper is the design of Constant False Alarm Rate (CFAR) systems used in radar processing for detecting targets. Our intention is then to study to what extent theorems 1 and 2 are complementary to standard results and algorithms such as those described in [Minkler and Minkler, 1990].

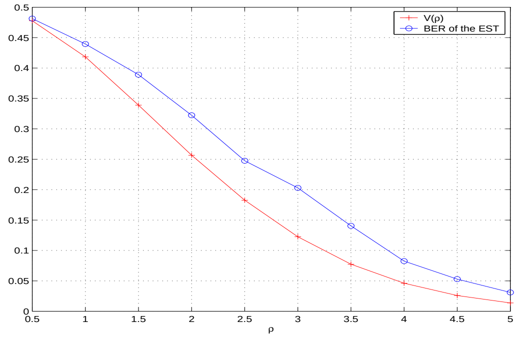

Fig. 1. Performance of the EST with $L = 100$ and $m = 100$ references for the non coherent detection of modulated sinusoidal carriers .

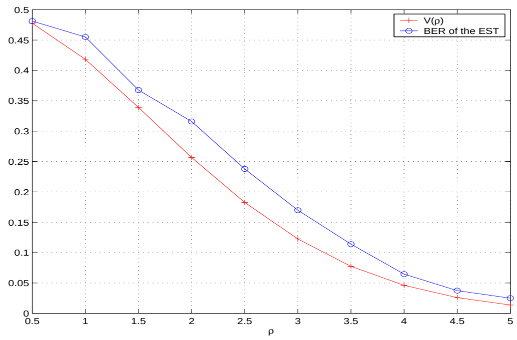

Fig. 2. Performance of the EST with $L = 200$ and $m = 200$ references for the non coherent detection of modulated sinusoidal carriers.

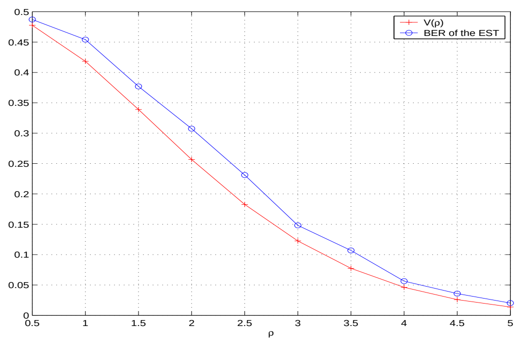

Fig. 3. Performance of the EST with $L = 300$ and $m = 300$ references for the non coherent detection of modulated sinusoidal carriers.

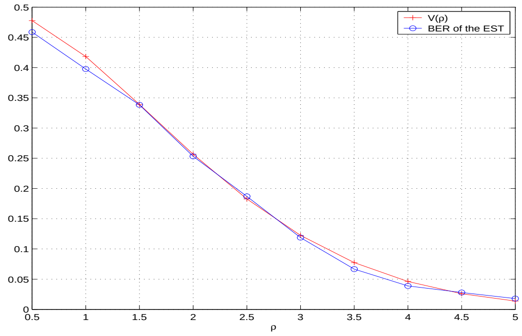

Fig. 4. Performance of the EST with $L = 300$ and $m = 300$ references for the non coherent detection of modulated sinusoidal carriers with amplitudes $A[dB]$ equal to $\rho[dB] + 1$.

References

[Forsythe *et al.*, 1976]G. E. Forsythe, M. A. Malcolm, and C. B. Moler. *Computer Methods for Mathematical Computations*. Prentice-Hall, 1976.

[Kailath and Poor, 1998]T. Kailath and H. V. Poor. Detection of stochastic processes. *IEEE Transactions On Information Theory*, pages 2230–2259, 1998.

[Kay, 1998]S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice Hall, Upper Saddle River, 1998.

[Minkler and Minkler, 1990]G. Minkler and J. Minkler. *The Principles of Automatic Radar Detection In Clutter, CFAR*. Magellan Book Company, Baltimore, 1990.

[Pastor *et al.*, 2002]D. Pastor, R. Gay, and A. Groenenboom. A sharp upper-bound for the probability of error of the likelihood ratio test for detecting signals in white gaussian noise. *IEEE Transactions On Information Theory*, pages 228–238, 2002.

[Pastor, 2004]D. Pastor. *A limit theorem for sequences of independent random vectors with unknown distributions and its application to nonparametric detection.* GET/ENST-Bretagne Internal Report, Paris, 2004.

[Poor, 1994]H. V. Poor. *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1994.

[Press *et al.*, 1992]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C, The art of Scientific Computing*. Cambridge University Press, 1992.