Saddlepoint Approximation Applied to Fusion in Multiple Sensor and to Detection in Clutter

Christian Musso ONERA Chatillon France musso@onera.fr

Emmanuelle Jay ONERA Palaiseau France ejay@noos.fr

Jean-Philippe Ovarlez **ONERA** Palaiseau France ovarlez @onera.fr

Abstract - We propose to apply an efficient approximation called « saddlepoint approximation » to some fusion problems. The aim of this method is to approximate the law of a sum of (n) iid (independent identically distributed) variables. A lot of statistical problems (maximum likelihood estimator, hypothesis testing,...) can be formulated as such a sum. This approximation is very accurate even for small (n) and easy to implement. We describe the principle of the method and we applied it to optimal fusion in multiple sensor and to a detector in non Gaussian clutter

Keywords : Saddlepoint, classification, recognition, detection, non Gaussian clutter

1 Introduction

The saddlepoint (SP) approximation was introduced in by H. E. Daniels in 1954 [1] and has been developed and enhanced until today [2]. It consists of approximate the law of a sum of (n) iid general variables

$$Z_n = \frac{1}{n} \sum_{i=1}^n X_i \tag{1}$$

This approximation is very accurate even for small (n) $(n \ge 3)$ and has, under general conditions, the strong following property of the relative error,

$$\frac{f_n(x)}{\hat{f}_n(x)} = 1 + o(\frac{1}{n})$$
(2)

 $\hat{f}_n(x)$ is the SP approximation of the density $f_n(x)$ of Z_n . Moreover, this approximation remains accurate even for tail distribution areas wich is crucial, for example, in detection problems (probability of false alarm). For a vast family of densities (like exponential densities) relative error does not depend of x (uniform convergence). Compared with the saddlepoint approximation, the law of large numbers approximation is much poor. This later approximation works for absolute error with very large (n) and doesn't work in the tail distribution area.

A lot of statistical problems (maximum likelihood estimator, hypothesis testing,...) can be formulated as sum of (n) iid variables such the logarithm of the likelihood. For example, in chapters 3, 4 the SP approximation is applied successfully in classification, sensor fusion or detection. In these examples, there is no altenative (law of large numbers, convolution) to compute precisely and rapidly the underlying laws

2 Saddlepoint formula

Let X be a random variable following the law of the sample X_i in (1). Following [1], suppose that the moment generating function K(t) of X converges in some interval [-c1,c2],

$$K(t) = \log(E[e^{tX}]) \text{ for } -c_1 < t < c_2$$
 (3)

Then by inverse Fourier transformation, $f_n(x)$ can be expressed as,

$$f_n(x) = \frac{n}{2\pi} \int_{-\infty}^{+\infty} \exp[n(K(t) - tx]dt$$
 (4)

which can be expressed with (t) being now a complex variable,

$$f_n(x) = \frac{n}{2i\pi} \int_{\tau-\infty}^{\tau+\infty} \exp[n(K(t) - tx)]dt$$
 (5)

for every au. Let T_0 be the unique minimum of K(t) - tx on the real axis,

$$K'(T_0) - x = 0 (6)$$

then replacing τ by the root T₀ in (5), the integrand of (5) becomes negligible when (n) increases and when (t) is outside a immediate neighbourhood of T₀. T₀ is called a saddlepoint because the K(t) - tx is minimum at T_0 for real (t) and the

modulus of the integrand of (5) is maximum at T_0 Developping K(t) - tx around T_0 we obtain the SP approximation [1],

Density saddlepoint approximation

$$\hat{f}_n(x) = \sqrt{\frac{n}{2\pi K''(T_0)}} \exp(K(T_0) - T_0 x)$$
(7)

where $K''(T_0)$ is always positive [1]

Sometimes rather then estimating the density, we prefer to estimate the cumulative function $P(Z_n > \alpha)$. A very accurate SP approximation especially in the tail propability area is due to Lugananni and Rice [3]

<u>Cumulative probability function saddlepoint</u> <u>approximation</u>

$$\hat{P}(Z_n > \alpha) = 1 - \Phi(y) + \phi(y)(t^{-1} - y^{-1})$$
(8)

where ϕ is the standard Gaussian distribution and Φ its cumulative distribution function,

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) \quad \Phi(y) = \int_{-\infty}^{y} \phi(t) dt \tag{9}$$
$$t = T_0 [nK''(T_0)]^{1/2} \quad y = \pm \sqrt{2n} (T_0 \alpha - K(T_0))^{1/2}$$

and $K'(T_0) - \alpha = 0$, y is chosen such that it has the same sign as (t)

The root T_0 can be computed with a newton algorithm like,

$$t_{k+1} = t_k - K'(t_k) / K''(t_k)$$
(10)

with t_0 chosen close to $-c_1$ or close to c_2

We can easily generalize the SP approximation when X_i are idenpendent but not identically distributed by computing each moment generating function,

$$K(t) = \frac{1}{n} \sum_{i=1}^{n} K_i(t) = \frac{1}{n} \sum_{i=1}^{n} \log(E[e^{tX_i}])$$
(11)

<u>Note</u>: In practice the training densities of X are not available. But they are not required, it is sufficient to evaluate the empirical moment generation functions in order to compute completely the saddlepoint approximations and its derivatives,

$$\hat{K}(t) = \log(\frac{1}{n} \sum_{j=1}^{n} e^{iX_j})$$
(12)

3 Sensors fusion.

We compute the optimal threshold estimation with saddlepoint approximation for the optimal fusion in multiple sensor. The SP approximation can be performed also with discrete variables (decentralized fusion [8]) : each sensor gives its local decision

3.1 Modelisation

Consider (k) sensors observintarget in classification context (2 hypothesis H_0 and H_1). The sensor (i) measures some characteristic Y_*^i of the unknown target class. Y_*^i has one distribution under H_0 (f_0^i) and one under H_1 (f_1^i) (see figures 1, 2). The goal is to classify by fusionning the measurements Y_*^i (here taken one-dimentional) collected by each sensor.



Figure1 Densities of the classifier for sensor 1 and 2



Figure 2 Densities of the classifier for sensor 3 and 4

The number of the measurements (n_i) of the i-th sensor $(Y_j^1,...,Y_{n(i)}^i)$ is not fixed in advance and therefore the law of the fusions cannot be computed offline. They have to be rapidly and precisely estimated in order to decide optimally $(\delta = H_i)$

$$\prod_{i=1}^{k} \prod_{j=1}^{n_i} \frac{f_1^i(Y_j^i)}{f_0^i(Y_j^i)} > s \Longrightarrow \quad \delta = H_1$$
(13)

with a false decision controlled by the threshold (t)

$$P_{fa} = P(Z > \alpha = \frac{\log(s)}{\sum_{i=1}^{k} n_i} / H_0) = \varepsilon$$

where P_{fa} is the probability of false alarm Let Z be,

$$Z = (1 / \sum_{i=1}^{k} n_i) \sum_{i=1}^{k} \sum_{j=1}^{n_i} \log[\frac{f_1^i}{f_0^i}(Y_j^i)]$$
(14)

The goal is to compute the law of Z to determine this threshold.. Z is a complex mixture of densities, its law can not be analytically expressed.

3.2 Saddlepoint derivation

The fusionned likelihood ratio Z (14) can be written as

$$Z = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{n(i)} \sum_{j=1}^{n(i)} X_{j}^{i}$$
(15)

where

$$X_{j}^{i} = \log(\frac{\sigma_{10}}{\sigma_{1i}}) + \frac{1}{2\sigma_{0i}^{2}}(Y_{j}^{i} - m_{0i})^{2} - \frac{1}{2\sigma_{1i}^{2}}(Y_{j}^{i} - m_{1i})^{2}$$

= $a_{i} (Y_{j}^{i})^{2} + b_{i} Y_{j}^{i} + c_{i}$

under H₀, Y_j^i follows the Gaussian law $f_0^i = N(m_{0i}, \sigma_{0i})$. The corresponding moment generation function is

$$K(t) = \frac{1}{k} \sum_{i=1}^{k} K_i(t)$$

where,

$$K_i(t) = \log(\mathrm{E}[\mathrm{e}^{\mathrm{t}X_*^{i}}/H_0])$$

is straightforwardly computed as,

$$K_{i}(t) = -\frac{1}{2\sigma_{0i}^{2}}m_{0i}^{2} + c_{i}t + \frac{(m_{0i}\sigma_{0i}^{2} + b_{i}\sigma_{0i}^{2}t)^{2}}{\sigma_{0i}^{2}(2 - 4a_{i}\sigma_{0i}^{4})} - \frac{1}{2}\log(1 - 2a_{i}\sigma_{0i}^{2}t)$$
(16)

K(t) is defined (3) if and only if,

$$-c_1 = \max_{a_{i \le 0}} \frac{1}{2a_i \sigma_{0i}^2} \le t \le \min_{a_{i \ge 0}} \frac{1}{2a_i \sigma_{0i}^2} = c_2$$

The root T_0 (6) is computed with the algorithm (10) with $t_0 = -c_1 + \varepsilon$. On average, this algorithm

converges after 6 Newton iterations. SP approximation is performed with (8).

3.3 Simulation Results

For simplicity and without loss of generality we suppose that each sensor has collected the same number of measurements in range (n/s=2, 4, 8). The number of sensors is in range (k=2,4). The law of the optimal fusion $P(Z>t/H_0)$ is computed by 10^6

Monte Carlo trials for a P_{fa} less than 0.5 (see Figure 3). This law is supposed to be exact, it is the reference for the next approximations. SP approximation is applied to estimate P(Z>t/H₀) for $P_{fa} \in [10^{-3} \ 0.5]$



Figure 3. Law of the fusionned lassifier P(Z>t/H₀) for $P_{fa} \in [10^{-3} \ 0.5]$

We show the results of the SP approximation $\hat{P}(Z \ge t/H_0)$ by computing the relative error

$$RE(t) = P(Z \ge t/H_0)/\hat{P}(Z \ge H_0)$$
⁽¹⁷⁾

As previewed by the theory we observe (see Figure 4) that the RE is close to one, especially when the number of measurements increases (2, 4, 8).

Moreover, the RE is nearly independent of the threshold. Even for very few measurements and even in the tail area distribution, the saddlepoint approximation is accurate (unlike the law of large numbers). Note that the gaussian choice of the underlying densities $f_{0,1}^i$ is not relevant, the densities can be much more general.



Figure 4. Relative Error for the SP approximation for $P_{fa} \in [10^{-3} \ 0.5]$

4 Detection in non Gaussian SIRP clutter

Coherent radar detection against non-Gaussian clutter has gained many interests in the radar community since experimental clutter measurements made by organizations like MIT [4] showed to fit non-Gaussian statistical models. One of the most tractable and elegant non-Gaussian model results in the so-called Spherically Invariant Random Process (SIRP) theory which states that some non-Gaussian random processes are the product of a Gaussian random process - called speckle - with a non-negative random variable (r.v.)- called texture. Using this model many results arised. For example in Gini and al.'s works [5] the optimum detector in the presence of composite disturbance of known statistics modeled as SIRP is derived. In previous Jay and al.'s works [6,7], a bayesian approach was proposed to determine the PDF of the texture (the characteristic PDF of the SIRP) from n reference clutter cells. Although SIRP model allows to derive optimum non-Gaussian detectors, it is quite difficult to determine their analytical expression. In this section, we propose to apply the Saddlepoint approximation to estimate the performance of a detector called Asymptotical BORD which was derived in [6,7] and which is an asymptotical result of the so-called BORD.

The basic problem of detecting the presence (H_1) or absence (H_0) of a complex signal *s* in a set of *n* measurements of m-complex (or real) vectors y corrupted by a sum *c* of independent additive complex (or real) noises (noises + clutter) can be described in terms of a statistical hypothesis test, $H_0: y = c$ and $H_1: y = s + c$. The target signal *s* corresponds to a modified version of the perfectly known transmitted signal and can be rewritten as s = Ap, where *A* is the complex target amplitude. The observed vector *y* is used to form the well-known Likelihood Ratio Test (LRT) $\Lambda(y)$ which is compared with a threshold η set to a desired false alarm probability (P_{fa}) computed as follows.

$$P_{fa} = P(\Lambda(y) > \eta/H_0)$$
(18)

4.1 SIRV and Bayesian Optimum Radar Detector

SIRV (Vector)model interprets each element of the clutter vector c as the product of a m-complex (or real) Gaussian vector X with a positive random variable (r.v.) τ the so-called *texture*, that is $c = X\sqrt{\tau}$. Applied to the detection problem the LRT becomes:

$$\int_{0}^{\infty} \left[\exp(-\frac{q_{1}(y)}{2\tau}) - \exp(\lambda - \frac{q_{0}(y)}{2\tau}) \right] \frac{p(\tau)}{\tau^{m}} d\tau \underset{H_{0}}{\overset{H_{1}}{>}} 0 \quad (19)$$

where $q_0 = y^H M^{-1}y$ and $q_1 = q_0(y - s)$ for a known signal *s* and $\lambda = \log(\eta)$. From *n* clutter reference cells of size *m*, $R = [r_1, ..., r_n]^T$ where $r_i = [r_i(1), ..., r_i(m)]^T$, a bayesian estimator $\hat{p}_n(\tau)$ of $p(\tau)$ can be computed and the so-called BORD expression is given by [6,7]:

$$\Lambda_{n}(y) = \frac{\sum_{i=1}^{n} \left(\frac{r_{i}^{H} M^{-1} r_{i}}{(q_{1}(y) + r_{i}^{H} M^{-1} r)^{2}}\right)^{m}}{\sum_{i=1}^{n} \left(\frac{r_{i}^{H} M^{-1} r_{i}}{(q_{0}(y) + r_{i}^{H} M^{-1} r)^{2}}\right)^{m}} \overset{=}{\underset{H_{0}}{\overset{<}{\longrightarrow}}} \lambda.$$
(20)

Asymptotically, BORD converges in law when $n \rightarrow +\infty$ to the Asymptotical BORD :

$$1 - (\Lambda_n(y))^m \approx \frac{|p^H M^{-1} y|^2}{p^H M^{-1} p y^H M^{-1} y} = \frac{Q_1}{Q_2} = U$$
(21)

for which it is possible to obtain analytically the PDF [6]. When y is a complex vector, we have,

$$p_U(u) = (n-1)(1-u)^{n-2}$$
(22)

and when y is a real vector,

$$p_U(u) = u^{-1/2} (1-u)^{(n-3)/2} \left[B((n-1)/2, 1/2) \right]^{-1}$$
(23)

where B((n-1)/2,1/2) is the Beta function. Then the detection threshold λ can be computed as follows for a fixed P_{fa} value $P_{fa} = P(U > \alpha/H_0) = (1-\alpha)^{n-1}$ with

 $\alpha = 1 - (\sqrt[q]{\lambda})^{-1}$ and for a complex vector y. When y is real valued, we have to compute the following expression,

$$P_{fa} = \left[B((n-1)/2, 1/2)\right]^{-1} \int_{\alpha}^{1} u^{-1/2} (1-u)^{(n-3)/2} du \qquad (24)$$

whose result is related to the ${}_{1}F_{2}$ hypergeometric function. Generally, and for example when we have to estimate the correlation matrix M from the received data, these formulas become no more valid and it is quite heavy to compute the detection threshold via Monte Carlo runs especially when P_{fa} is low like 10^{-6} or less. So, in the next section, the Saddle-point Approximation is presented and applied to this problem. We will see that this method allows to compute the PDF of variables expressed in a more general form than (21).

4.2 Saddlepoint approximation

This approximation be applied to general Gaussian quadratic forms ratios like,

$$\mu = P\left[\frac{Q_1}{Q_2} = \frac{X^T B X}{X^T A X} > \alpha\right]$$
(25)

where the square matrices B and A are (nxn) positive symetrical (B can be singular) and X is a centered Gaussian vector. μ can be expressed as,

$$\mu = P[\sum_{i=1}^{n} \lambda_i(\alpha) Y_i^2 > 0]$$
(26)

where Y_i are iid standard Gaussians, $Y_i \sim N(0,1)$ and $\lambda_i(\alpha)$ the eigen values of the symetrical matrix (B - α A). Indeed,

$$\mu = \int_{X^T BX \atop X^T AX} \phi(X) dX = \int_{X^T (B - \alpha A) X > 0} \phi(X) dX.$$

Changing variables, X = UY where U is an orthogonal matrix of the eigen value decomposition of the symetrical matrix $(B - \alpha A)$, such that $(B - \alpha A) = UAU^{T}$ then,





Figure 5. True P_{fa} vs α for the detector

$$P[Z = \frac{1}{n} \sum_{i=1}^{n} X_i > 0]$$
(27)

Now the formulation (26) allows to use the saddlepoint approximation (8). The computation of the moment generating function and its derivatives is straightforwardly,

$$K(t) = \sum_{i=1}^{n} K_i(t) = \sum_{i=1}^{n} \int \exp[t\lambda_i(\alpha)x^2]\phi(x)dx = -\frac{1}{2}\sum_{i=1}^{n} \log(1 - 2t\lambda_i(\alpha))$$
(28)

which is defined when (3) is verified :

$$-c_1 = \max_{a_{i \le 0}} \frac{1}{2\lambda_i(\alpha)} \le t \le \min_{a_{i \ge 0}} \frac{1}{2\lambda_i(\alpha)} = c_2$$

4.3 Simulation Results

Applied to the detection problem and after a simple variable change $(q = M^{-1/2}p)$, $q = M^{-1/2}p$), the saddlepoint approximation can be computed for (21), with $B = q(q^Hq)^{-1}q$ and $A = Id_n$ (25) with n=10.

Figure 5 represents the P_{fa} computed with the true value (24) and Figure 6 shows the relative error (RE) of the SP approximation for a P_{fa} varying between 10^{-3} and 0.5. Again, we observe the good behavior of the SP approximation.

$$RE = \frac{P_{fa}}{\hat{P}_{fa}}(\alpha) = P(\frac{Q_1}{Q_2} > \alpha/H_0)$$



Figure 6. Relative Error vs α for the SP approximation

5 Conclusions

Modified saddlepoint approximations have been proposed and applied to optimal fusion in multiple sensor and to detection in non Gaussian clutter. This approximation is very accurate even for a small sample size and even for tail area of the distribution. Morover this method is easy to implement and requires a little of computation effort.

References

[1] H. E. Daniels. « Saddlepoint approximations in statistic ». Ann. Math. Statis. 5 631-650

[2] O. E. Barndorff-Nielsen, C. Klüppelberg « Tail exactness of multivariate saddlepoint approximations ». Scand. J. Statist. 26 (1999), 253-264

[3] R. Lugannani, S. Rice (1980). « Saddlepoint approximation for the distribution of the sum of

independent random variables. Adv. IN Appl. Probab 12 475-490

[4] J.B.Billingsley, « Low-angle radar land clutter. Measurements and empirical models », SciTech Publishing, February 2002.

[5] F.Gini, M.V.Greco, A.Farina and P.Lombardo, « Optimum and mismatched detection against Kdistributed plus Gaussian clutter », IEEE Trans.-AES, Vol.34, No.3, pp.860-876, (July 1998)

[6] E.Jay, « Détection en environnement non-Gaussien », Ph.D. Thesis, Université de Cergy-Pontoise / Onera, France, (Juin 2002)

[7] E.Jay, J.P.Ovarlez, D.Declercq, P.Duvaut, « BORD : Bayesian Optimum Radar Detector », Proc.-IEEE-ICASSP02, Vol.II, pp.1289-1292, Orlando, FL, USA, (May 2002)

[8] C. A. S. Thomopoulos, R. Viswanathan. « Optimal Decision Fusion in Multiple Sensor Systems ». IEEE Trans. Aero. Elec. Sys. AES 23 no 5. September 1987