

# Non-efficacité et non-gaussianité asymptotiques d'un estimateur du maximum de vraisemblance à fort rapport signal sur bruit

P. Forster<sup>(1)</sup>, P. Larzabal<sup>(2)</sup>, E. Boyer<sup>(2)</sup>, A. Renaux<sup>(2)</sup>

<sup>(1)</sup>GEA, IUT de Ville d'Avray, 1 Chemin Desvallières, 92410 VILLE D'AVRAY

<sup>(2)</sup>SATIE, UMR 8029, Ecole Normale Supérieure de Cachan, 61 avenue du président Wilson, 94235 CACHAN Cedex

Philippe.Forster@cva.u-paris10.fr larzabal@satie.ens-cachan.fr eric.boyer@wanadoo.fr renaux@satie.ens-cachan.fr

**Résumé** - En théorie de l'estimation, dans le cas d'observations indépendantes de mêmes densités de probabilité, l'efficacité asymptotique en le nombre  $T$  d'observations de la méthode du Maximum de Vraisemblance (MV) est un résultat bien connu qui permet d'appréhender ses performances lorsque  $T$  est grand. Dans certaines situations, le nombre d'observations peut être faible et ce résultat ne s'applique plus. Dans le cadre du traitement d'antenne et d'une modélisation stochastique des signaux émis par les sources, nous remédions à cette lacune lorsque le Rapport Signal sur Bruit (RSB) est grand. Nous montrons que dans cette situation, l'estimateur du MV est asymptotiquement (en RSB) non-efficace et non-gaussien.

**Abstract** - In estimation theory, for independent identically distributed observations, the asymptotic efficiency of the Maximum Likelihood (ML) method when the number  $T$  of observations tends to infinity is a well known result. This is useful for performance evaluation for large  $T$ . In some scenarios, the number of snapshots may be small making this result unapplicable. In the array processing framework, for Gaussian emitted signals, we fill this lack when the Signal to Noise Ratio (SNR) is high. In this situation, we show that the ML estimation is asymptotically non efficient and non Gaussian when the SNR tends to infinity.

## 1. Introduction

Les performances asymptotiques du Maximum de Vraisemblance Stochastique (MVS) en traitement d'antenne sont bien connues lorsque le nombre d'observations tend vers l'infini. En effet, en théorie de l'estimation, dans le cas d'observations indépendantes de mêmes densités de probabilité, l'efficacité et la gaussianité asymptotiques en le nombre  $T$  d'observations de la méthode du Maximum de Vraisemblance (MV) sont des résultats établis depuis longtemps [1]. Nous nous intéressons dans cet article au comportement du MVS à nombre fini d'échantillons lorsque le rapport signal à bruit tend vers l'infini : *c'est ce sens que nous donnons au qualificatif asymptotique dans cet article*. Dans ce cas à notre connaissance, il n'existe pas de résultat général fournissant la loi de l'estimateur du maximum de vraisemblance. Nous démontrons dans cet article que l'estimateur du MVS est, asymptotiquement en le rapport signal sur bruit, non gaussien et non efficace. Nous montrons aussi, à fort RSB, l'équivalence pour la détermination de la localisation des Critères du Maximum de Vraisemblance Déterministe (CMVD) et Stochastique (CMVS).

Dans le paragraphe 2 nous présentons le modèle des signaux reçus. L'équivalence entre les méthodes du MVS et MVD est démontrée dans le paragraphe 3. La loi asymptotique de l'estimateur du MVS est ensuite établie dans le paragraphe 4. Enfin les simulations sont conduites dans le paragraphe 5.

## 2. Le modèle de signal reçu

Considérons le problème classique suivant où l'on cherche à localiser  $P$  sources à l'aide d'une antenne de  $M > P$  capteurs recevant les signaux  $\mathbf{y}(t)$  :

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta}_e) \mathbf{x}(t) + \mathbf{b}(t) \quad \text{pour } t=1, \dots, T, \quad (1)$$

où :

-  $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1) \ \dots \ \mathbf{a}(\theta_P)]$  est la matrice  $M \times P$  des vecteurs directionnels ;

-  $\boldsymbol{\theta} = [\theta_1 \ \dots \ \theta_P]^T$  est le vecteur des  $P$  directions d'arrivées inconnues dont la valeur exacte est  $\boldsymbol{\theta}_e$  ;

-  $\mathbf{x}(t)$  est le vecteur  $P \times 1$  des signaux émis par les  $P$  sources ;

-  $\mathbf{b}(t)$  est le bruit supposé complexe, gaussien, circulaire, centré, temporellement et spatialement blanc, de puissance  $\sigma^2$  inconnue.

Nous noterons

$$\mathbf{B} = [\mathbf{b}(1), \mathbf{b}(2), \dots, \mathbf{b}(T)] \quad (2)$$

et

$$\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)]. \quad (3)$$

## 3. CMVD et CMVS

Deux modèles sont couramment utilisés pour les signaux sources : le modèle dit déterministe dans lequel les signaux sources émis  $\mathbf{x}(t)$  sont traités comme des paramètres inconnus, et le modèle stochastique dans lequel  $\mathbf{x}(t)$  est un vecteur aléatoire complexe, gaussien, circulaire, centré, temporellement blanc, de covariance  $\boldsymbol{\Sigma}$  inconnue.

Dans le cas du modèle déterministe, les directions d'arrivée sont obtenues par la minimisation du critère concentré :

$$C_{MVD}(\boldsymbol{\theta}) = \frac{1}{M-P} \text{Trace}(\Pi_b(\boldsymbol{\theta}) \hat{\mathbf{R}}_y), \quad (4)$$

où  $\Pi_b(\boldsymbol{\theta})$  est le projecteur candidat sur l'espace bruit et de façon générique la covariance empirique d'un vecteur  $\mathbf{v}$  est donnée par

$$\hat{\mathbf{R}}_{\mathbf{v}} = \frac{1}{T} \sum_{t=1}^T \mathbf{v}(t)\mathbf{v}(t)^H. \quad (5)$$

Dans le cas du modèle stochastique, les directions d'arrivée sont obtenues par la minimisation du critère concentré:

$$C_{MVS}(\boldsymbol{\theta}) = \left| \Pi_s(\boldsymbol{\theta}) \hat{\mathbf{R}}_{\mathbf{y}} \Pi_s(\boldsymbol{\theta}) + C_{MVD}(\boldsymbol{\theta}) \Pi_b(\boldsymbol{\theta}) \right|, \quad (6)$$

où  $\Pi_s(\boldsymbol{\theta})$  est le projecteur candidat sur l'espace signal et  $|\mathbf{F}|$  désigne le déterminant de la matrice  $\mathbf{F}$ .

### Théorème 1

Au premier ordre en  $\mathbf{B}$  les directions d'arrivée obtenues par les minimisations de  $C_{MVS}(\boldsymbol{\theta})$  et de  $C_{MVD}(\boldsymbol{\theta})$  sont les mêmes.

### Démonstration

Posons  $\mathbf{U}_s(\boldsymbol{\theta})$  et  $\mathbf{U}_b(\boldsymbol{\theta})$  les matrices  $M \times P$  et  $M \times (M - P)$  formées à partir de bases orthonormées des espaces signal et bruit et notons :  $\mathbf{U}(\boldsymbol{\theta}) = [\mathbf{U}_s(\boldsymbol{\theta}) \mathbf{U}_b(\boldsymbol{\theta})]$ .

On a d'après (6)

$$C_{MVS}(\boldsymbol{\theta}) = \left| \mathbf{U}(\boldsymbol{\theta})^H \left( \Pi_s(\boldsymbol{\theta}) \hat{\mathbf{R}}_{\mathbf{y}} \Pi_s(\boldsymbol{\theta}) + C_{MVD}(\boldsymbol{\theta}) \Pi_b(\boldsymbol{\theta}) \right) \mathbf{U}(\boldsymbol{\theta}) \right| \quad (7)$$

La matrice intervenant dans le déterminant de l'équation (7) est bloc diagonale et on vérifie alors sans peine que  $C_{MVS}(\boldsymbol{\theta})$  s'écrit de manière équivalente :

$$C_{MVS}(\boldsymbol{\theta}, \mathbf{b}) = \underbrace{\left| \mathbf{U}_s(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{\mathbf{y}}(\mathbf{b}) \mathbf{U}_s(\boldsymbol{\theta}) \right|}_{\alpha(\boldsymbol{\theta}, \mathbf{b})} C_{MVD}(\boldsymbol{\theta}, \mathbf{b})^{M-P}, \quad (8)$$

où l'on a fait apparaître explicitement la dépendance des différents termes en  $\mathbf{b} = \text{vec}(\mathbf{B})$  de façon à analyser  $\alpha(\boldsymbol{\theta}, \mathbf{b})$  et  $C_{MVD}(\boldsymbol{\theta}, \mathbf{b})$  à fort rapport signal sur bruit. Remarquons tout d'abord que  $C_{MVD}(\boldsymbol{\theta}, \mathbf{b})$  est minimal et nul en  $(\boldsymbol{\theta}, \mathbf{b}) = (\boldsymbol{\theta}_e, \mathbf{0})$ . Il en est donc de même pour  $C_{MVS}(\boldsymbol{\theta}, \mathbf{b})$  car  $\alpha(\boldsymbol{\theta}, \mathbf{b}) \geq 0$ .

Notons  $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_e$ . En effectuant un développement en série de  $\alpha(\boldsymbol{\theta}, \mathbf{b})$  et de  $C_{MVD}(\boldsymbol{\theta}, \mathbf{b})$  au voisinage de  $\boldsymbol{\theta} = \boldsymbol{\theta}_e$  et  $\mathbf{b} = \mathbf{0}$  et en ne mentionnant que les premiers termes non nuls on a :

$$\begin{aligned} \alpha(\boldsymbol{\theta}, \mathbf{b}) &= \alpha(\boldsymbol{\theta}_e, \mathbf{0}) + \dots \\ C_{MVD}(\boldsymbol{\theta}, \mathbf{b}) &= [\Delta\boldsymbol{\theta}^T \mathbf{b}^T] \mathbf{H} [\Delta\boldsymbol{\theta}^T \mathbf{b}^T]^T + \dots, \end{aligned} \quad (9)$$

où  $\mathbf{H}$  est le Hessien du critère du MVD et  $\alpha(\boldsymbol{\theta}_e, \mathbf{0}) = \left| \mathbf{U}_s(\boldsymbol{\theta}_e)^H \mathbf{A}(\boldsymbol{\theta}_e) \hat{\mathbf{R}}_{\mathbf{x}} \mathbf{A}(\boldsymbol{\theta}_e)^H \mathbf{U}_s(\boldsymbol{\theta}_e) \right| \neq 0$ . Par conséquent le critère du  $C_{MVS}$  est équivalent, à fort rapport signal sur bruit, à  $\alpha(\boldsymbol{\theta}_e, \mathbf{0}) C_{MVD}(\boldsymbol{\theta}, \mathbf{b})^{M-P}$ , ce qui montre

que les minimisations de  $C_{MVS}$  et  $C_{MVD}$  sont alors équivalentes.

## 4. Loi asymptotique du MVS

Nous nous plaçons à partir de maintenant dans le cadre du modèle stochastique des signaux sources précisé au début du paragraphe 3 et nous notons  $\hat{\boldsymbol{\theta}} = \arg \min C_{MVS}(\boldsymbol{\theta})$  l'estimateur du MVS. Le théorème suivant établit la loi asymptotique de  $\hat{\boldsymbol{\theta}}$  et démontre par la même sa non gaussianité.

### Théorème 2

Lorsque  $\sigma \rightarrow 0$ ,  $\tilde{\boldsymbol{\theta}} = \frac{1}{\sigma} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_e)$  converge en loi vers  $\mathbf{C}\mathbf{y}$

où :

- $\mathbf{y}$  est un vecteur  $P \times 1$  aléatoire gaussien centré de covariance  $\mathbf{I}$  ;
- $\mathbf{C}$  est une matrice  $P \times P$  aléatoire indépendante de  $\mathbf{y}$  dont la loi est détaillée ci-dessous.

La matrice  $\mathbf{C}$  est telle que

$$\mathbf{C}\mathbf{C}^T = \frac{1}{2T} \left\{ \text{Re}[\mathbf{H} \otimes \mathbf{W}] \right\}^{-1}, \quad (10)$$

où :

- $\otimes$  désigne le produit d'Hadamard.
- $\mathbf{W}$  est une matrice qui suit une loi de Wishart complexe à  $T$  degrés de liberté, de matrice paramètre la covariance  $\boldsymbol{\Sigma}$  des signaux sources ;
- $\mathbf{H} = \mathbf{D}^H [\mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H] \mathbf{D}$  ;

$$\mathbf{D} = \begin{bmatrix} \left( \frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}} \right)_{\theta_1} & \dots & \left( \frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}} \right)_{\theta_P} \end{bmatrix}. \quad (12)$$

### Démonstration du corollaire 2

D'après le théorème 1, à faible bruit, les critères du MVS et du MVD sont équivalents : nous considérons donc que  $\hat{\boldsymbol{\theta}}$  est obtenu par la minimisation du critère du MVD donné par l'équation (4) bien que les signaux sources soient stochastiques.

Etudions tout d'abord, à fort rapport signal à bruit ( $\sigma$  petit), la loi  $f(\tilde{\boldsymbol{\theta}} | \mathbf{X})$  de  $\tilde{\boldsymbol{\theta}}$  conditionnellement à une réalisation  $\mathbf{X}$  (équation (3)) des signaux sources. Cette loi asymptotique est gaussienne, de covariance donnée par la borne de Cramer Rao déterministe notée  $\mathbf{B}_D$

$$\mathbf{B}_D = \frac{1}{2T} \left\{ \text{Re}[\mathbf{H} \otimes \hat{\mathbf{R}}_{\mathbf{x}}^T] \right\}^{-1}. \quad (13)$$

$\hat{\mathbf{R}}_{\mathbf{x}}$  suit une loi de Wishart complexe à  $T$  degrés de liberté, de matrice paramètre la covariance  $\boldsymbol{\Sigma}$  des signaux sources, et il en est de même pour  $\hat{\mathbf{R}}_{\mathbf{x}}^T$ . Soit  $\mathbf{B}_D = \mathbf{C}\mathbf{C}^T$  la factorisation de Cholesky de la covariance de  $\tilde{\boldsymbol{\theta}}$  conditionnellement à  $\mathbf{X}$ .  $\tilde{\boldsymbol{\theta}}$  a donc pour loi asymptotique conditionnelle à  $\mathbf{X}$  celle de  $\mathbf{C}\mathbf{y}$  où  $\mathbf{y}$  est un vecteur aléatoire gaussien complexe centré de covariance unité. En levant le conditionnement sur les

signaux sources  $\mathbf{X}$  on aboutit au résultat énoncé où  $\mathbf{W} = \hat{\mathbf{R}}_{\mathbf{X}}^T$

### Corollaire 1

Avec les notations du théorème 2, la covariance asymptotique de  $\tilde{\boldsymbol{\theta}}$  est donnée par

$$\text{cov}(\tilde{\boldsymbol{\theta}}) = \frac{1}{2T} E \left[ \left( \text{Re}[\mathbf{H} \otimes \mathbf{W}] \right)^{-1} \right]. \quad (14)$$

### Corollaire 2

Dans le cas d'une seule source, pour lequel l'estimateur du MV consiste en une simple formation de voie,  $\tilde{\theta}$  est distribué asymptotiquement selon la loi  $kS_{2T}$  où  $S_{2T}$  est une variable aléatoire de Student avec  $2T$  degrés de liberté et où  $k$  est donné par

$$k^2 = \lim_{\sigma \rightarrow 0} \frac{1}{\sigma^2} B_S, \quad (15)$$

où  $B_S$  est la borne de Cramer Rao stochastique. La variance asymptotique de  $\tilde{\theta}$  est donnée par

$$\text{var}_{as}(\tilde{\theta}) = k^2 \frac{T}{T-1}, \quad (16)$$

et l'estimateur du maximum de vraisemblance n'est pas efficace.

### Démonstration

Dans le cas d'une seule source de puissance  $\Sigma$ , le théorème 2 montre que

$$\tilde{\theta} \xrightarrow{\text{en loi}} cy, \quad (17)$$

où  $y$  suit une loi gaussienne réduite et

$$c = \left( 2h \sum_{t=1}^T |x(t)|^2 \right)^{-1/2} \quad (18)$$

où

$$h = \left\| \left( \frac{\partial \mathbf{a}}{\partial \theta} \right)_{\theta_1} \right\|^2 - \frac{1}{\|\mathbf{a}(\theta_1)\|^2} \left| \mathbf{a}(\theta_1)^H \left( \frac{\partial \mathbf{a}}{\partial \theta} \right)_{\theta_1} \right|^2. \quad (19)$$

Dans le cas d'une seule source la borne de Cramer Rao stochastique s'écrit [2]

$$B_S = \frac{\sigma^2}{2T\Sigma h} \frac{\sigma^2 + \Sigma \|\mathbf{a}(\theta_1)\|^2}{\Sigma \|\mathbf{a}(\theta_1)\|^2} \quad (20)$$

$$\approx \frac{\sigma^2}{2T\Sigma h} \quad (\text{pour } \sigma \text{ tendant vers } 0)$$

De plus

$$\frac{2}{\Sigma} \sum_{t=1}^T |x(t)|^2 \sim \chi_{2T}^2, \quad (21)$$

où  $\chi_{2T}^2$  est une variable aléatoire du Chi2 à  $2T$  degrés de liberté. On sait que lorsque  $X$  suit une loi normale réduite et que  $\chi_N^2$  est indépendant de  $X$ , le rapport  $X/\sqrt{\chi_N^2/N}$  suit une loi de Student à  $N$  degrés de liberté. Dans ces conditions le report de (20) et (21) dans (18) et (17) démontre (15).

De plus, sachant que  $\text{var}(S_N) = N/(N-2)$ , la variance asymptotique de  $\tilde{\theta}$  s'obtient facilement : elle vaut  $k^2 T/T-1$ . Comme  $T/T-1 > 1$ , l'estimateur du MV n'est pas asymptotiquement efficace.

## 5. Simulations

On considère une antenne linéaire uniforme qui comporte 4 capteurs équirépartis d'une demi longueur d'onde. Les sources sont repérées par rapport au travers de l'antenne. Nous allons tout d'abord développer le cas d'une source pour lequel nous avons précédemment établi la loi des directions estimées. Le cas de plusieurs sources est ensuite étudié.

### 5.1 Cas d'une source

Dans ce cas on montre facilement que les critères du maximum de vraisemblance stochastique et déterministe fournissent les mêmes estimées quel que soit le rapport signal sur bruit. Dans le cas d'un vecteur source de norme indépendante de la direction d'arrivée ceux ci se réduisent par ailleurs à une simple formation de voies et il s'agit ici d'un cas particulier du théorème 1 pour lequel les deux critères sont rigoureusement identiques.

Nous nous intéressons à la localisation d'une source située en 0 degré à l'aide de  $T=2$  snapshots. La figure 1 représente la variance empirique de l'estimateur du maximum de vraisemblance (10000 Monte Carlo), la borne de Cramer Rao, et la variance asymptotique théorique donnée par l'équation (16) et qui vaut ici 2 fois la borne de Cramer Rao. On note une parfaite adéquation entre les résultats théoriques et les résultats de simulations.

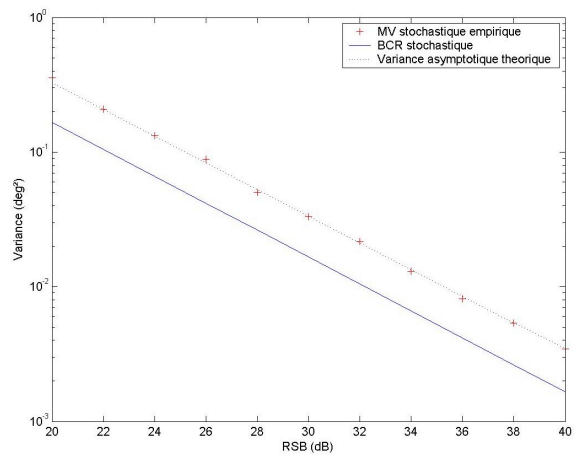


Fig 1 Evolution des variances (deg²) en fonction du RSB (dB)

La figure 2 présente un histogramme des directions d'arrivée estimées par le MV. Cet histogramme a été obtenu à l'aide de 50000 Monte Carlo, pour un RSB de 30dB. Il est en très bonne adéquation avec la loi de Student théorique que nous

avons établie : pour comparaison, la figure 2 montre aussi la densité de probabilité d'une loi normale de même variance qui ne correspond clairement pas à la distribution des estimées.

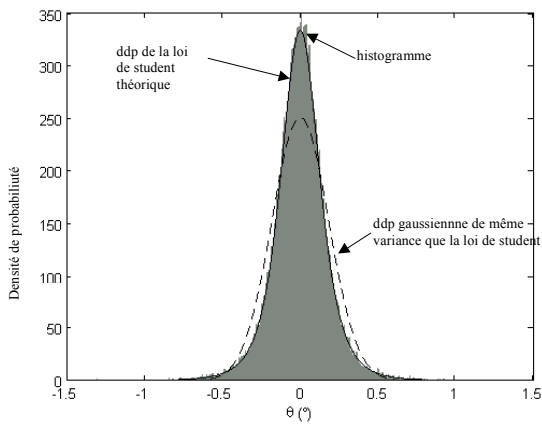


Fig 2 : Comparaison de l'histogramme des directions d'arrivée avec la loi théorique de Student, et une loi gaussienne de même variance.

### 5.2 Cas de 2 sources

Nous considérons maintenant le cas de deux sources de même puissance situées à  $-7.5$  deg et  $7.5$  deg. La localisation par la méthode du Maximum de vraisemblance est effectuée à l'aide d'un algorithme de Gauss Newton opérant sur  $T=3$  snapshots. La figure 3 donne, pour chacune des deux sources, l'évolution de la variance empirique (10000 Monte Carlo), de la variance théorique (cf eq. 14), et de la borne de Cramer Rao. Dans l'expression (14), à défaut d'une expression analytique, l'espérance a été estimée par le moyennage de 10000 tirages de Monte Carlo. On remarque à nouveau la bonne adéquation des simulations avec les résultats théoriques, ainsi que la non efficacité du MVS à fort RSB.

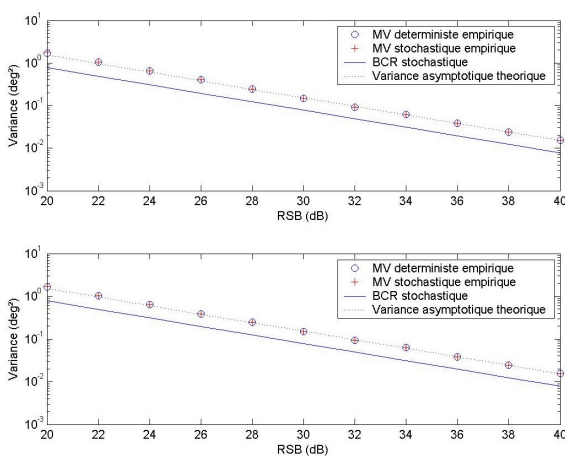


Fig 3 Evolution des variances ( $\text{deg}^2$ ) pour chacune des deux sources en fonction du RSB (dB)

Sur la figure 4 nous avons reporté les deux histogrammes des directions d'arrivée correspondant au cas précédemment décrit. Les histogrammes ont été calculés à partir de 10000 Monte Carlo pour deux sources de mêmes puissance avec un  $RSB=30\text{dB}$ . On constate qu'ici aussi la densité de probabilité des estimées n'est pas une gaussienne.

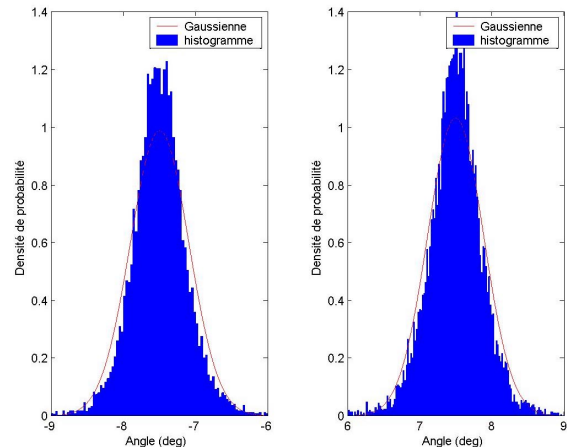


Fig 4 : Comparaison de l'histogramme des directions d'arrivée pour chacune des deux sources avec une loi gaussienne de variance égale à la variance empirique.

## 6. Conclusion

Nous avons montré dans cet article que l'estimateur du Maximum de Vraisemblance est asymptotiquement (en RSB) non-efficace et non-gaussien. Les calculs ont été entièrement effectués dans le cas d'une unique source, ils sont actuellement en cours dans le cas de plusieurs sources et feront l'objet d'une publication ultérieure.

## Références

- [1] H. Cramer. *Mathematical methods of statistics*. Princeton University Press, New Jersey, 1946,
- [2] B. Ottersten, M. Viberg, P. Stoica and A. Nehorai, *Exact and large sample Maximum likelihood techniques for parameter estimation and detection in array processing*, Berlin : in radar array processing, S. Haykin and J. Litva and T. J. Shepherd, Springer-Verlag, 1993.
- [3] Performance Study of Conditional and Unconditional DOA Estimation, P. Stoica, A. Nehorai, *IEEE Trans. on ASSP*, vol.38, no.10, pp.1783-1795, October 1990.
- [4] Second-Order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios, A. Eriksson, P. Stoica, T. Söderström, *Radar and Signal Processing, IEE PROCEEDINGS-F*, vol.140, no.4, pp.266-272, August 1993.