

Bornes d'estimation pour modèles mal spécifiés

Mouhamadou L. DIONG, Eric CHAUMETTE, François VINCENT

Isae-Supaero, Université de Toulouse,
10 av. Ed. Belin, Toulouse, France

([mouhamadou.diong, eric.chaumette, francois.vincent]@isae.fr)

Résumé – En estimation paramétrique, lorsque la densité de probabilité utilisée en estimation est différente de celle génératrice des observations, le modèle est dit mal spécifié. La plupart des résultats proposés pour l'estimation du Maximum de Vraisemblance pour les modèles mal spécifiés ne sont le plus souvent valides qu'à grand nombre d'observations indépendantes et nécessitent la résolution d'un problème d'optimisation non-linéaire particulièrement difficile en grande dimension. Pour contourner ces limitations, nous proposons de généraliser l'approche alternative adoptée par Fritsche et al. [1] à partir du formalisme proposé par Barankin [2]. Ainsi, nous montrons que la borne inférieure dérivée dans [1] fait partie d'une famille de bornes inférieures de type Barankin associées à une contrainte linéaire assimilable à une contrainte de biais mal spécifiée.

Abstract – When the assumed probability distribution of the observations differs from the true distribution, the model is said to be misspecified. The key results on maximum-likelihood estimation of misspecified models have been introduced in the limit of large sample support and depend on a parameters vector solution of a computationally expensive non-linear optimization problem. As a possible strategy to circumvent these limitations, we extend the approach lately proposed by Fritsche et al [1]. It is shown that the lower bound derived in [1] is a representative of a family of lower bounds deriving from a misspecified unbiasedness constraint leading to *generalized* Barankin-type lower bounds.

1 Introduction

Pour l'estimation de paramètres déterministes, une méthode assez répandue est l'estimateur du Maximum de Vraisemblance (MV). En effet, sous certaines conditions de régularité peu contraignantes sur le modèle d'observation[3], l'estimateur MV converge¹ à grand nombre d'observations indépendantes et admet une densité de probabilité (d.d.p.) asymptotique de forme Gaussienne. Cependant, si on considère, par exemple, T observations i.i.d. $\{\mathbf{x}_t\}_{t=1}^T$, ces différents résultats asymptotiques reposent sur l'hypothèse centrale que la d.d.p. génératrice de ces observations appartient à une famille paramétrique donnée de d.d.p. $f_\theta(\mathbf{x}_t) \triangleq f(\mathbf{x}_t|\theta)$, $\theta \in \mathbb{R}^P$. En pratique, dans la plupart des situations, il existe des écarts entre la densité $p(\mathbf{x}_t)$ génératrice des observations et le modèle supposé $f_\theta(\mathbf{x}_t)$. De ce fait, plusieurs études [4][5] se sont intéressées au comportement de l'estimateur MV lorsque le modèle d'observation est mal spécifiée. Ces études ont montré, entre autres, qu'à grand nombre d'observations indépendantes et sous des hypothèses peu contraignantes, l'estimateur dit Quasi-MV(1a), car basé sur la d.d.p. supposée $f_\theta(\mathbf{x}_t)$, admet une d.d.p.asymptotique de forme Gaussienne et converge fortement vers le vecteur de paramètres θ_f (1b) qui minimise la divergence de Kullback-Leibler (KL) [6] entre la vraie d.d.p. et la d.d.p. supposée. Ces connaissances sur les performances asymptotiques de l'estimateur Quasi-MV ont limité l'intérêt pour les bornes inférieures avec écarts de modèle à l'exception de la Borne de Cramer Rao pour les modèles mal spécifiés (MCRB) [7][8][9]. Cependant, la matrice de

covariance asymptotique et la MCRB, dépendent de θ_f . Par conséquent, leur évaluation numérique nécessite la résolution d'un problème d'optimisation non-linéaire en θ particulièrement difficile en grande dimension. En outre, ces quantités ne sont valides qu'à grand nombre d'observations indépendantes.

Pour contourner ces limitations, une approche alternative a été proposée par Fritsche et al. [1] dans le cas particulier où $p(\mathbf{x}_t|)$ et $f(\mathbf{x}_t|)$ partagent la même paramétrisation, i.e. : $p(\mathbf{x}_t) \triangleq p_\theta(\mathbf{x}_t) \triangleq p(\mathbf{x}_t|\theta)$. Dans cette approche, une borne inférieure est formulée pour les classes d'estimateurs sans biais ou avec un biais donné par rapport au modèle supposé $f_\theta(\mathbf{x}_t)$. Dans cet article, nous proposons une généralisation de la borne inférieure de Fritsche [1] pour les modèles mal spécifiés sur le même principe que les bornes de Barankin standard (Bornes de Cramer Rao, de McAulay-Seidman, etc.) [2][10] [12]. Nous avons comparé ces bornes *généralisées* à celle proposée dans [1] pour montrer que cette dernière peut être améliorée.

2 L'estimateur Quasi-MV

Plusieurs auteurs [4][5] ont montré que sous certaines hypothèses de régularité, l'estimateur MV en présence d'écarts de modèle dit Quasi-MV défini par:

$$\hat{\theta}(\bar{\mathbf{x}}) = \arg \max_{\theta} \left\{ f_\theta(\bar{\mathbf{x}}) = \prod_{t=1}^T f_\theta(\mathbf{x}_t) \right\}, \quad (1a)$$

est fortement convergeant lorsque le nombre d'observations indépendantes $T \rightarrow \infty$ et sa limite est le vecteur de paramètres

¹Si l'estimateur MV converge alors, il est asymptotiquement efficace [3].

minimisant la divergence KL:

$$\widehat{\theta}(\bar{\mathbf{x}}) \stackrel{a.s.}{\rightarrow} \theta_f = \arg \min \{D_{KL}(p \parallel f_\theta)\} \quad (1b)$$

$$D_{KL}(p \parallel f_\theta) = E_p [\ln(p(\mathbf{x}_t)) - \ln(f_\theta(\mathbf{x}_t))] \quad (1c)$$

avec $\bar{\mathbf{x}} = (\mathbf{x}_1^T, \dots, \mathbf{x}_T^T)^T$, $E_p[\mathbf{g}(\bar{\mathbf{x}})] = \int_{\mathbb{C}^{M \times T}} \mathbf{g}(\bar{\mathbf{x}}) p(\bar{\mathbf{x}}) d\bar{\mathbf{x}}$

et $p(\bar{\mathbf{x}}) = \prod_{t=1}^T p(\mathbf{x}_t)$.

En outre, la distribution asymptotique de $\widehat{\theta}(\bar{\mathbf{x}})$ est normale: $\widehat{\theta}(\bar{\mathbf{x}}) \stackrel{A}{\sim} \mathcal{N}(\theta_f, \mathbf{C}_{\widehat{\theta}})$. Sa matrice de covariance converge fortement vers la matrice "sandwich" de Huber $\mathbf{C}_{\widehat{\theta}} \stackrel{a.s.}{\rightarrow} \mathbf{C}_{HS}(\theta_f)$ défini par:

$$T\mathbf{C}_{HS}(\theta_f) = E_p \left[\frac{\partial^2 \ln f(\mathbf{x}_t | \theta_f)}{\partial \theta \partial \theta^T} \right]^{-1} \times \quad (1d)$$

$$E_p \left[\frac{\partial \ln f(\mathbf{x}_t | \theta_f)}{\partial \theta} \frac{\partial \ln f(\mathbf{x}_t | \theta_f)}{\partial \theta^T} \right] E_p \left[\frac{\partial^2 \ln f(\mathbf{x}_t | \theta_f)}{\partial \theta \partial \theta^T} \right]^{-1}$$

Toutefois, l'évaluation de la matrice "sandwich" de Huber nécessite la connaissance de θ_f . Ce qui exige la résolution d'un problème d'optimisation non linéaire multidimensionnelle (1b) pour chaque valeur de θ et donc potentiellement coûteuse en temps de calcul.

3 Bornes généralisées de type Barankin

Pour simplifier l'étude, on considère dans la suite l'estimation d'un paramètre déterministe $\theta \in \mathbb{R}$. Cependant, la démarche proposée ainsi que les résultats obtenus sont aisément applicables à l'estimation multiparamètre [12][13]. Pour la suite, on définit: $E_\theta[\mathbf{g}(\bar{\mathbf{x}})] \triangleq \int_{\mathbb{C}^{M \times T}} \mathbf{g}(\bar{\mathbf{x}}) f_\theta(\bar{\mathbf{x}}) d\bar{\mathbf{x}}$.

3.1 La borne de Barankin standard

Soit $\widehat{\theta}^0$ un estimateur d'une valeur particulière θ^0 du paramètre $\theta \in \Theta$. En l'absence d'erreurs de modèle, la borne de Barankin est l'Erreur Quadratique Moyenne (EQM) minimale pour la classe d'estimateurs uniformément sans biais. Elle est obtenue par la résolution du problème d'optimisation suivant:

$$\min \left\{ EQM_{\theta^0}[\widehat{\theta}^0] = E_{\theta^0} \left[\left| \widehat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \right|^2 \right] \right\} \quad (2)$$

sous contrainte d'absence uniforme de biais:

$$\forall \theta \in \Theta : E_\theta[\widehat{\theta}^0(\bar{\mathbf{x}})] = \theta. \quad (3)$$

En considérant le produit scalaire suivant $\langle u(\bar{\mathbf{x}}) | v(\bar{\mathbf{x}}) \rangle_\theta = E_\theta[u(\bar{\mathbf{x}})v(\bar{\mathbf{x}})]$ et en supposant que le support de $f_\theta(\bar{\mathbf{x}})$ est indépendant de θ , on peut réécrire le problème sous la forme[2]:

$$\min \left\{ \left\| \widehat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \right\|_{\theta^0}^2 \right\} \text{ s.c.}$$

$$\forall \theta \in \Theta : \left\langle \widehat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \mid \nu_{\theta^0}(\bar{\mathbf{x}}; \theta) \right\rangle_{\theta^0} = \theta - \theta^0. \quad (4)$$

où $\nu_{\theta^0}(\bar{\mathbf{x}}; \theta) = f_\theta(\bar{\mathbf{x}})/f_{\theta^0}(\bar{\mathbf{x}})$ désigne le Rapport de Vraisemblance (RV). Dans la suite du document, nous reprenons cette formulation.

3.2 Généralisation aux modèles mal spécifiés

Sous des conditions générales peu contraignantes sur $f_\theta(\bar{\mathbf{x}})$ [3], l'estimateur MV $\widehat{\theta}_{MV}^0$ de θ^0 converge à grand nombre d'observations indépendantes, et est uniformément sans biais par rapport à $f_\theta(\bar{\mathbf{x}})$:

$$\forall \theta \in \Theta : E_\theta \left[\widehat{\theta}_{MV}^0(\bar{\mathbf{x}}) \right] = \theta. \quad (5a)$$

Cependant, si $f(\bar{\mathbf{x}}|\theta)$ n'est pas la vraie d.d.p. des observations, alors l'estimateur devient celui Quasi-MV qui vérifie non plus une contrainte de biais uniforme (3), mais une contrainte linéaire donnée:

$$\int \widehat{\theta}_{MV}^0(\bar{\mathbf{x}}) f_\theta(\bar{\mathbf{x}}) d\bar{\mathbf{x}} = \theta, \quad \int f_\theta(\bar{\mathbf{x}}) d\bar{\mathbf{x}} = 1, \quad (5b)$$

où le vecteur contrainte $f_\theta(\bar{\mathbf{x}})$ est d'intégrale normalisée. En fait, puisque $f_\theta(\bar{\mathbf{x}})$ est une d.d.p., cette contrainte linéaire (5a-5b) peut être vue comme une contrainte de biais mal spécifiée. En outre, si $f_\theta(\bar{\mathbf{x}})$ et la vraie d.d.p. $p(\bar{\mathbf{x}})$ partagent le même support, alors tout estimateur $\widehat{\theta}^0$ vérifiant la contrainte linéaire (5a) vérifie aussi:

$$\forall \theta \in \Theta : E_p \left[\left(\widehat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \right) \omega_p(\bar{\mathbf{x}}; \theta) \right] = \theta - \theta^0, \quad (5c)$$

avec $\omega_p(\bar{\mathbf{x}}; \theta) = f_\theta(\bar{\mathbf{x}})/p(\bar{\mathbf{x}})$. En substituant cette contrainte à celle (3) dans le problème de minimisation (4):

$$\min \left\{ MSE_p[\widehat{\theta}^0] = \left\| \widehat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \right\|_p^2 \right\} \text{ s.c.}$$

$$\forall \theta \in \Theta : \left\langle \widehat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \mid \omega_p(\bar{\mathbf{x}}; \theta) \right\rangle_p = \theta - \theta^0, \quad (5d)$$

avec $\langle u(\bar{\mathbf{x}}) | v(\bar{\mathbf{x}}) \rangle_p = E_p[u(\bar{\mathbf{x}})v(\bar{\mathbf{x}})]$ et $\|\cdot\|_p$ la norme associée à $\langle \cdot | \cdot \rangle_p$, on obtient une généralisation de la borne de Barankin pour les estimateurs vérifiant la contrainte linéaire (5a-5b), mais appliquée à des observations générées par $p(\bar{\mathbf{x}})$ et donc une borne pour l'estimateur Quasi-MV. Cependant, les bornes de Barankin généralisée et standard présentent l'inconvénient suivant: lorsque Θ contient un sous-ensemble continu de \mathbb{R} , les contraintes des équations (4, 5d) aboutissent à des équations intégrales sans solution analytique en général [12]. De ce fait, à partir des travaux de Barankin, plusieurs études [10][13] ont proposé d'autres bornes inférieures comme approximations de celle de Barankin standard basées sur des transformations linéaires de la contrainte de "Barankin" (4) et dérivées à partir du lemme de minimisation de la norme [12]. Nous proposons une approche similaire pour déterminer des approximations de la Borne de Barankin généralisée.

• Lemme de minimisation de la norme

Soit \mathbb{U} un espace vectoriel réel euclidien muni d'un produit scalaire $\langle \cdot | \cdot \rangle$. Soit $(\mathbf{c}_1, \dots, \mathbf{c}_K)$ une famille libre de K vecteurs de \mathbb{U} et $\mathbf{v} \in \mathbb{R}^K$. Le lemme de minimisation de la norme stipule que la solution au problème de minimisation $\|\mathbf{u}\|^2$ sous K contraintes linéaires $\langle \mathbf{u} | \mathbf{c}_k \rangle = v_k, k \in [1, K]$, est la suivante:

$$\min \left\{ \|\mathbf{u}\|^2 \right\} = \mathbf{v}^T \mathbf{R}^{-1} \mathbf{v}, \quad \mathbf{R}_{n,k} = \langle \mathbf{c}_k | \mathbf{c}_n \rangle. \quad (6)$$

• Borne de McAulay-Seidman généralisée (BMSG)

La BMSG d'ordre N s'obtient à partir d'une relaxation de la con-trainte dans (5d) en la discrétisant en N valeurs particulières de θ dit points test : $\boldsymbol{\theta}^N = (\theta^1, \dots, \theta^N)^T \in \Theta^N$.

Posons $\boldsymbol{\xi}^N = (\theta^1 - \theta^0, \dots, \theta^N - \theta^0)^T$. Tout estimateur $\hat{\theta}^0$ vérifiant la contrainte (5c) doit satisfaire :

$$E_p \left[\left(\hat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \right) \omega_p(\bar{\mathbf{x}}; \theta) \right] = \boldsymbol{\xi}^N, \quad (7a)$$

avec $\omega_p(\bar{\mathbf{x}}; \boldsymbol{\theta}^N) = (\omega_p(\bar{\mathbf{x}}; \theta^1), \dots, \omega_p(\bar{\mathbf{x}}; \theta^N))^T$. Le lemme de minimisation de la norme (6) permet d'associer à ce jeu de contraintes une borne inférieure de l'EQM:

$$EQM_p \left[\hat{\theta}^0 \right] \geq \left(\boldsymbol{\xi}^N \right)^T \mathbf{R}_{\omega_p}^{-1} \boldsymbol{\xi}^N, \quad (7b)$$

$$\left(\mathbf{R}_{\omega_p} \right)_{n,n'} = E_p \left[\omega_p(\bar{\mathbf{x}}; \theta^{n'}) \omega_p(\bar{\mathbf{x}}; \theta^n) \right]. \quad (7c)$$

La Borne de McAulay-Seidman généralisée (BMSG) est le supremum de (7b) par rapport au vecteur point-test $\boldsymbol{\theta}^N$. Un cas particulier intéressant consiste à inclure θ^0 dans les points-test i.e. $(\boldsymbol{\theta}^{N+1})^T = \left(\theta^0, (\boldsymbol{\theta}^N)^T \right)$. Alors, en posant $(\boldsymbol{\xi}^{N+1})^T = \left(0, (\boldsymbol{\xi}^N)^T \right)$ et en effectuant une inversion par bloc, la BMSG devient :

$$BMSG = \sup_{\boldsymbol{\xi}} \left\{ \left(\boldsymbol{\xi}^N \right)^T \left(\mathbf{R}_{\omega_p} - \frac{\mathbf{e}_p^N (\mathbf{e}_p^N)^T}{E_p \left[\omega_p^2(\bar{\mathbf{x}}; \theta^0) \right]} \right)^{-1} \boldsymbol{\xi}^N \right\} \quad (7d)$$

avec $\mathbf{e}_p^N = E_p \left[\omega_p(\bar{\mathbf{x}}; \theta^0) \omega_p(\bar{\mathbf{x}}; \boldsymbol{\theta}^N) \right]$. La BMSG se réduit à la BMS[11, (3)] : $\sup_{\boldsymbol{\xi}} \left\{ \left(\boldsymbol{\xi}^N \right)^T \left(\mathbf{R}_{\nu_{\theta^0}} - \mathbf{1}_N \mathbf{1}_N^T \right)^{-1} \boldsymbol{\xi}^N \right\}$ lors-que $p(\bar{\mathbf{x}}) \triangleq p_{\theta}(\bar{\mathbf{x}}) \triangleq f_{\theta}(\bar{\mathbf{x}})$.

• Borne de Cramér-Rao généralisée (BCRG)

La BCRG est construite par passage à la limite de l'expression (7a) pour les 2 points-test suivants:

$$E_p \left[\left(\hat{\theta}^0(\bar{\mathbf{x}}) - \theta^0 \right) \begin{pmatrix} \omega_p(\bar{\mathbf{x}}; \theta^0) \\ \omega_p(\bar{\mathbf{x}}; \theta^0 + d\theta) \end{pmatrix} \right] = \begin{pmatrix} 0 \\ d\theta \end{pmatrix} \quad (8a)$$

Lorsque $d\theta \rightarrow 0$, on peut montrer à partir de (7d) que:

$$BCRG(\theta^0) = \left(F_p(\theta^0) - \frac{r_{21}^2(\theta^0)}{r_{11}(\theta^0)} \right)^{-1} \quad (9)$$

avec $F_p(\theta) = E_p \left[\left(\frac{\partial \omega_p(\bar{\mathbf{x}}; \theta)}{\partial \theta} \right)^2 \right]$, $r_{11}(\theta) = E_p \left[\omega_p(\bar{\mathbf{x}}; \theta)^2 \right]$ et $r_{21}(\theta) = E_p \left[\frac{\partial \omega_p(\bar{\mathbf{x}}; \theta)}{\partial \theta} \omega_p(\bar{\mathbf{x}}; \theta) \right]$.

En outre, si $p(\bar{\mathbf{x}}) \triangleq p_{\theta}(\bar{\mathbf{x}}) \triangleq f_{\theta}(\bar{\mathbf{x}})$, alors $\frac{\partial \omega_p(\bar{\mathbf{x}}; \theta^0)}{\partial \theta} = \frac{\partial \ln p_{\theta^0}(\bar{\mathbf{x}})}{\partial \theta}$ et $r_{21}(\theta^0) = E_{\theta^0} \left[\frac{\partial \ln p_{\theta^0}(\bar{\mathbf{x}})}{\partial \theta} \right] = 0$ et par conséquent :

$$BCRG(\theta^0) = F_p(\theta^0)^{-1} = F(\theta^0)^{-1} = BCR(\theta^0) \quad (10)$$

où $F(\theta) = E_{\theta} \left[\left(\frac{\partial \ln p_{\theta}(\bar{\mathbf{x}})}{\partial \theta} \right)^2 \right]$ désigne la Matrice d'Information de Fisher. De ce fait, on peut identifier $F_p(\theta)$ à une Matrice d'Information de Fisher généralisée.

Notons que la borne inférieure proposée par [1, (8)] est :

$$FB(\theta^0) = F_p(\theta^0)^{-1} \quad (11)$$

ne correspond pas exactement à la BCRG (9). La différence provient du fait que les auteurs dans [1] n'exploitent que la première partie de la contrainte (8a). Ainsi:

$$FB(\theta^0) \leq BCRG(\theta^0) \quad (12)$$

4 Modèles linéaires Gaussiens

En reprenant un exemple similaire à celui traité par Fritsche [1], on considère le cas où le vrai modèle d'observation est linéaire gaussien de la forme : $\mathbf{x} = \mathbf{d}_p \theta + \mathbf{n}$, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_p)$, avec $\mathbf{d}_p \in \mathbb{R}^M$ et $\mathbf{C}_p \in \mathbb{R}^{M \times M}$. Pour toute valeur donnée θ^0 du paramètre θ , la vraie d.d.p. est : $p(\mathbf{x}) \triangleq p_{\theta^0}(\mathbf{x}) \triangleq p_{\mathcal{N}}(\mathbf{x} | \mathbf{d}_p \theta^0, \mathbf{C}_p)$. En pratique, s'il est possible de bien identifier la structure linéaire Gaussienne du modèle, une connaissance imparfaite des coefficients \mathbf{d}_p et \mathbf{C}_p peut conduire à l'utilisation de valeurs approchées \mathbf{d}_f et \mathbf{C}_f ; la d.d.p. utilisée pour l'estimation est alors la suivante : $f_{\theta}(\mathbf{x}) \triangleq p_{\mathcal{N}}(\mathbf{x} | \mathbf{d}_f \theta, \mathbf{C}_f)$. Afin d'illustrer les différents résultats théoriques, nous proposons de comparer l'EQM de l'estimateur MV :

$$\hat{\theta}^0(\mathbf{x}) = \mathbf{w}_f^T \mathbf{x}, \quad \mathbf{w}_f = \mathbf{C}_f^{-1} \mathbf{d}_f / \mathbf{d}_f^T \mathbf{C}_f^{-1} \mathbf{d}_f, \quad (13)$$

à "l'EQM de Huber, la BMSG (7d) pour 2 points-test, la BCRG (9), la borne FB (11) proposée par Fritsche et al. [1] et les BCRs standards (10) associées à la vraie d.d.p. et à celle sup-posée. Les résultats obtenus sont présentés Fig. 1.

• L'EQM associée aux résultats de Huber

Pour deux d.d.p.s Gaussiens $p_{\theta^0}(\mathbf{x})$ and $f_{\theta}(\mathbf{x})$, la divergence KL (1b) est:

$$D_{KL} \propto (\mathbf{d}_f \theta - \mathbf{d}_p \theta^0)^T \mathbf{C}_f^{-1} (\mathbf{d}_f \theta - \mathbf{d}_p \theta^0) + \text{cte} \quad (14a)$$

Par conséquent :

$$\theta_f^0 = \mathbf{w}_f^T \mathbf{d}_p \theta_0 = E_p \left[\hat{\theta}^0(\mathbf{x}) \right]. \quad (14b)$$

Ensuite, du fait que $\frac{\partial \ln f_{\theta}(\mathbf{x})}{\partial \theta} = \mathbf{d}_f^T \mathbf{C}_f^{-1} (\mathbf{x} - \mathbf{d}_f \theta)$, on a :

$$\mathbf{C}_{HS}(\theta_f) = \frac{\mathbf{d}_f^T \mathbf{C}_f^{-1} \mathbf{C}_p \mathbf{C}_f^{-1} \mathbf{d}_f}{(\mathbf{d}_f^T \mathbf{C}_f^{-1} \mathbf{d}_f)^2} = \text{Var}_p \left[\hat{\theta}^0 \right], \quad (14c)$$

L'EQM de Huber: $EQM_{Huber} = (\theta^0 - \theta_f^0)^2 + \mathbf{C}_{HS}(\theta_f)$ coïncide exactement avec l'EQM de l'estimateur MV représentée par la courbe "MML" dans Fig. 1.

• BMSG, BCRG et FB

Si $\theta^N \triangleq \theta^1 = \theta^0 + d\theta$, alors, en appliquant (7d):

$$BMSG = \sup_{d\theta} \frac{d\theta^2 R_{\theta^0, \theta^0}}{R_{\theta^0 + d\theta, \theta^0 + d\theta} R_{\theta^0, \theta^0} - R_{\theta^0 + d\theta, \theta^0}^2}, \quad (15a)$$

avec :

$$\begin{aligned} R_{\theta^0, \theta^1} &= E_p [\omega_p(\mathbf{x}; \theta^0) \omega_p(\mathbf{x}; \theta^1)] \\ &= \frac{\sqrt{|\bar{\mathbf{C}}| |\mathbf{C}_p|}}{|\mathbf{C}_f|} e^{\frac{1}{2}(\mathbf{m}_{0,1}^T \bar{\mathbf{C}} \mathbf{m}_{0,1} - \delta_{\theta^0, \theta^1})} \end{aligned} \quad (15b)$$

$$\begin{aligned} \bar{\mathbf{C}} &= (2\mathbf{C}_f^{-1} - \mathbf{C}_p^{-1})^{-1}, \quad \mathbf{m}_{0,1} = \mathbf{C}_f^{-1} \mathbf{d}_f (\theta^0 + \theta^1) - \mathbf{C}_p^{-1} \mathbf{d}_p \theta^0 \\ \delta_{\theta^0, \theta^1} &= ((\theta^0)^2 + (\theta^1)^2) \mathbf{d}_f^T \mathbf{C}_f^{-1} \mathbf{d}_f - (\theta^0)^2 \mathbf{d}_p^T \mathbf{C}_p^{-1} \mathbf{d}_p. \end{aligned}$$

Ces résultats ne sont valides que si $\mathbf{C}_p > \frac{1}{2} \mathbf{C}_f$ [12, (33-34)]. La BCRG (9) correspond au cas limite de la BMSG (15a) lorsque $d\theta \rightarrow 0$, i.e. :

$$BCRG(\theta_0) = \frac{e^{-\frac{1}{2}(\mathbf{m}_{\theta^0, \theta^0}^T \bar{\mathbf{C}} \mathbf{m}_{\theta^0, \theta^0} - \delta_{\theta^0, \theta^0})}}{\mathbf{d}_f^T \mathbf{C}_f^{-1} \bar{\mathbf{C}} \mathbf{C}_f^{-1} \mathbf{d}_f}. \quad (16)$$

Enfin, la borne FB (11) est obtenue à partir de [1, (10)] tandis que les BCRs standards sont : $BCR_f = (\mathbf{d}_f^T \mathbf{C}_f^{-1} \mathbf{d}_f)^{-1}$, $BCR_p = (\mathbf{d}_p^T \mathbf{C}_p^{-1} \mathbf{d}_p)^{-1}$.

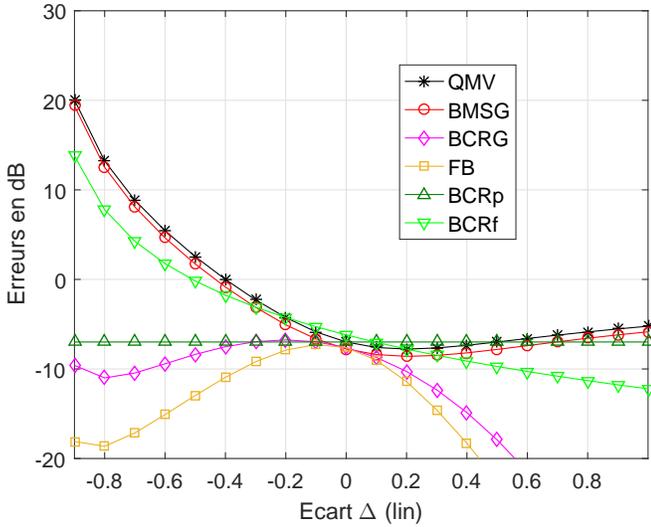


FIG. 1: Bornes et EQM en fonction de Δ , $M = 5$, $\mathbf{C}_f = 1.2\mathbf{C}_p$, $\mathbf{C}_p = \mathbf{I}_M$, $\theta^0 = 1$.

Dans l'exemple présenté, on suppose que $\mathbf{d}_p = \mathbf{1}_M$, $\mathbf{d}_f = (1 + \Delta) \mathbf{d}_p$ avec $\Delta \in [-1, 1]$, $\theta^0 = 1$, et $\mathbf{C}_p = \mathbf{I}_M$ (matrice identité), $M = 5$, et $\mathbf{C}_f = 1.2 \times \mathbf{C}_p$. La figure obtenue montre bien (conformément aux résultats dans [1]) que les BCRs standard, BCR_f et BCR_p , ne sont plus nécessairement des bornes pour l'EQM. En outre, la borne FB est comme prévue moins performante que la BCRG. Enfin, notons que les bornes généralisées pour les "Small Errors" (FB et BCRG) sont peu informatives lorsque les écarts aux modèles (Δ , \mathbf{C}_f) sont importants puisque celles-ci deviennent trop optimistes dans leur prédiction de l'EQM. En revanche, les bornes généralisées "Large Errors" (BMSG) constituent un prédicteur fiable de l'EQM sur un large domaine pour Δ .

5 Conclusion

L'examen des bornes *généralisées* pour le cas linéaire montre un potentiel important dans leur capacité de prédiction de l'EQM de l'estimateur Quasi-MV. L'objectif futur est l'extension de cette analyse aux cas non linéaires, cas pour lequel l'EQM de Huber ne coïncide plus nécessairement avec l'EQM de l'estimateur Quasi-MV sur les 3 zones de fonctionnement connues : a priori, de transition et asymptotique.

References

- [1] C. Fritsche et al, "On the Cramér-Rao lower bound under model mismatch", in Proceedings of ICASSP 2015, pp 3986-3990
- [2] E.W. Barankin, "Locally best unbiased estimates", Ann. Math. Stat., 20(4): 477-501, 1949.
- [3] E. L. Lehmann and G. Casella, *Theory of Point Estimation (2nd ed.)*. Springer, 1998
- [4] P. J. Huber, "The behavior of maximum likelihood estimates under nonstandard conditions," in Proc. Fifth Berkeley Symp. on Math. Statist. and Prob. Berkeley, CA, USA: Univ. of California, 1967.
- [5] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, pp. 1-25, Jan. 1982.
- [6] S. Kullback and A. Leibler, "On information and sufficiency", *Annals of Mathematical Statistics*, 22: 79-86, 1951
- [7] Q. H. Vuong, "Cramér-Rao bounds for misspecified models," Div. of the Humanities and Social Sci., Caltech, Pasadena, CA, USA, Working Paper 652, Oct. 1986
- [8] C.D. Richmond and L.L. Horowitz, "Parameter Bounds on Estimation Accuracy Under Model Misspecification", *IEEE Trans. on SP*, 63(9): 2263-2278, 2015
- [9] S. Fortunati, F. Gini, and M. S. Greco, "The misspecified Cramér-Rao bound and its application to the scatter matrix estimation in complex elliptically symmetric distributions," in *IEEE Trans. on SP*, 64(9): 2387-2399, 2016
- [10] R. McAulay and L.P. Seidman, "A useful form of the Barankin lower bound and its application to PPM threshold analysis", *IEEE Trans. on IT*, 15(2): 273-279, 1969
- [11] I. Reuven and H. Messer, "A Barankin-type lower bound on the estimation error of a hybrid parameter vector," *IEEE Trans. on IT*, 43(3): 1084-1093, 1997
- [12] E. Chaumette, J. Galy, A. Quinlan, P. Larzabal, "A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum-Likelihood Estimators", *IEEE Trans. on SP*, 56(11):5319-5333, 2008
- [13] K. Todros and J. Tabrikian, "General Classes of Performance Lower Bounds for Parameter Estimation-Part I: Non-Bayesian Bounds for Unbiased Estimators", *IEEE Trans. on IT*, 56(10): 5064-5082, 2010