

# A propos de bornes de Cramer-Rao pour l'estimation distribuée

Yue BI et Philippe CIBLAT

LTCI, Télécom Paris, Institut Polytechnique de Paris  
19 place Marguerite Perey, 91120 Palaiseau, France  
{bi,ciblat}@telecom-paris.fr

**Résumé** – Dans un contexte d'estimation paramétrique dont les données sont distribuées parmi un certain nombre de capteurs, nous nous intéressons aux performances ultimes au niveau d'un centre de fusion récoltant d'une certaine manière ces données. La remontée de ces données au niveau du centre de fusion est faite à travers de bits provenant d'une quantification des données brutes. Nous souhaitons savoir si il est préférable de remonter les données en une seule fois ou par petits morceaux, voire échantillon par échantillon comme cela est fait dans la littérature. Grâce à un calcul simple de borne de Cramer-Rao (BCR), nous montrons qu'il existe un découpage optimal des données qui ne correspond pas à ce qui est fait dans la littérature, lorsque la somme des bits à transmettre par tous les capteurs et le quantificateur sont prédéfinis.

**Abstract** – Assuming a parametric estimation context, we consider that the data depending on the parameter, we wish to estimate, are distributed over several sensors. A quantized versions of data are collected from a fusion center where the final estimation is performed. Actually, data are split into frames and only quantized versions of frames are sent to fusion center. The goal of this paper is to exhibit the frame size minimizing the Cramer-Rao bound of the estimation problem. When the number of bits sent by all the sensors and the quantizer are predefined, we show that the best frame size is larger than 1 (while this case is often applied in the State of the Art) and smaller than the frame encompassing the whole data.

## 1 Introduction

Dans de nombreuses applications de surveillance, des capteurs sont déposés dans des endroits différents mais cherchent néanmoins à mesurer le même phénomène et donc le même paramètre. Les capteurs n'estiment pas eux-mêmes le paramètre final mais envoient leurs données via un canal de propagation à un centre de fusion qui fera lui-même l'estimation conjointe. Ce papier se focalise sur le problème de la remontée de l'information entre les capteurs et le centre de fusion.

Dans la littérature des réseaux de capteurs, le système de remontée est en général extrêmement simple car à chaque nouvelle arrivée d'un échantillon, celui-ci est quantifié et transmis au centre de fusion et la gestion des envois entre les capteurs est fait selon une technique séquentielle de Round-Robin. Dans ce contexte, la borne de Cramer-Rao (BCR) a été calculée dans différents papiers selon différentes variantes mineures d'hypothèses [1, 2, 3]. Ainsi le nombre total de bits transmis dépend à la fois du nombre de capteurs et du nombre d'observations. Pour limiter le nombre de bits transmis, il apparaît naturel mais évidemment pas optimal de rassembler les échantillons dans des trames et de ne quantifier que ces trames. Cet aspect est manquant dans cette littérature. En revanche, dans la littérature de l'estimation statistique (notamment pour l'apprentissage), ce cas de figure est plus courant. Dans ce contexte, on considère plusieurs machines calculant en parallèle des estimateurs (et non séquentiellement mais cela ne change rien au

propos) et envoient leurs résultats quantifiés une fois le calcul fini, bref une fois toutes les données locales traitées. Finalement en mettant en série les différentes machines, on se trouve dans le contexte de l'envoi de trames quantifiées (et chaque trame contient plusieurs échantillons). Cette approche permettrait de résoudre le questionnement premier du papier. Néanmoins nous avons remarqué que les papiers provenant du calcul distribué [4, 5] offrent des bornes peu précises et ceci s'explique bien en raison des outils mathématiques employées. En effet, ils ou elles utilisent des bornes de la théorie de l'information qui s'expliquent pour tout quantificateur et de ce fait ne capturent tous les phénomènes.

Par conséquent, l'objectif de ce papier est de calculer et d'évaluer une borne de Cramer-Rao bayésienne d'un estimateur de la moyenne avec des données bruitées par un bruit gaussien pour toute taille de trame et tout nombre de bits transmis. Ensuite nous analysons numériquement plus spécifiquement le cas où le nombre total de bits et le quantificateur sont prédéfinis et nous remarquons alors que la taille de trame optimale n'est ni la trame réduite à un échantillon, ni la trame englobant toutes les données. Ainsi un compromis sur la taille de la trame est parfois nécessaire.

L'article est organisé ainsi : dans la Section 2, on introduit le modèle du système avec notamment le lien entre les données et le paramètre à estimer. A la Section 3, le calcul de la borne de Cramer-Rao de notre problème d'estimation est mené. La comparaison avec les autres formules de l'état de l'art notamment provenant du calcul distribué est effectuée en Section 4. Les résultats numériques sont discutés en Section 5.

---

Ce travail a été en partie soutenu par l'ERC "CTOCom".

## 2 Système étudié

Dans cette section, nous présentons le modèle distribué de données que nous avons considéré.

Nous considérons  $K$  capteurs qui recueillent  $N$  échantillons chacun  $\{y_k(n)\}_{n=0}^{N-1}$ . Ces échantillons seront transmis après une certaine quantification (décrite plus tard) qui conduit à  $KB$  bits émis. Nous considérons que les canaux de transmission entre les capteurs et le centre de fusion sont parfaits. Donc tout bit émis est correctement reçu.

Nous supposons que les échantillons dépendent d'un paramètre  $\theta$  unique (scalaire, pour simplifier) que nous souhaitons donc estimer. Nous supposons également que ce paramètre a une distribution *a priori*  $p_\Theta$ . Par souci de simplicité, nous choisirons une gaussienne de moyenne nulle et de variance  $\sigma_\theta^2$ . Le but de ce papier est d'obtenir des expressions d'une borne de Cramer-Rao bayésienne pour ce dit-paramètre.

Afin d'avoir des expressions de BCR exploitables, nous nous concentrons sur un modèle simple entre les échantillons et  $\theta$ . Ainsi, on a pour tout  $k \in \{1, \dots, K\}$  et  $n \in \{0, \dots, N-1\}$

$$y_k(n) = \theta + w_k(n), \quad (1)$$

avec  $w_k(n)$  une suite gaussienne indépendante et identiquement distribuée (iid) en  $k$  et  $n$  de moyenne nulle et variance connue  $\sigma_w^2$ .

Au niveau de chaque capteur, on découpe en trames le bloc de  $N$  échantillons. On considère des trames de longueur  $M$  (avec l'hypothèse que  $N$  est divisible par  $M$ ). On a ainsi  $T$  trames par capteur avec  $T = N/M$ . Pour chaque capteur  $k$ , on calcule l'estimateur optimal bayésien local de  $\theta$  associé à la trame  $t \in \{0, \dots, T-1\}$ . Ainsi, on a

$$\hat{\theta}_k(t) = c \left( \frac{1}{M} \sum_{m=0}^{M-1} y_k(tM + m) \right), \quad (2)$$

avec

$$c = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_w^2/M}. \quad (3)$$

Ensuite cet estimateur local est quantifié sur  $b$  bits via le quantificateur  $Q$  (que nous supposons identiques pour tous les capteurs car ce n'est pas ce point-là que nous voulons examiner). Finalement ces  $b$  bits sont transmis au centre de fusion. Par conséquent, le nombre de bits transmis par capteur est  $B = bT$ . Le quantificateur scalaire sur  $b$  bits est défini ainsi :

$$v = Q(u) \text{ si } u \in [\tau_v, \tau_{v+1}[ , \text{ pour } v \in \{0, \dots, L\} \quad (4)$$

où  $L = 2^b - 1$  et  $\{\tau_\ell\}_{\ell \in \{0, \dots, L+1\}}$  est une suite strictement croissante de seuil telle que  $\tau_0 = -\infty$  et  $\tau_{L+1} = +\infty$ . En pratique, on utilisera un quantificateur uniforme centré en zéro et dont les bornes inférieure et supérieure sont  $\tau_1 = -\tau$  et  $\tau_L = \tau$ . Par conséquent, on a, pour  $v \in \{1, \dots, L\}$ ,

$$\tau_v = -\tau + \frac{2\tau(v-1)}{L-1}. \quad (5)$$

Notre objectif est de calculer la borne de Cramer-Rao bayésienne sur  $\theta$  au niveau du centre de fusion qui possède pour données l'ensemble  $\{q_k(t)\}_{k \in \{1, \dots, K\}, t \in \{1, \dots, T\}}$  avec

$$q_k(t) = Q(\hat{\theta}_k(t)). \quad (6)$$

## 3 Expressions de la BCR

Dans cette section, on calcule la BCR pour le problème d'estimation posé. On notera par  $\hat{\theta}$  un estimateur de  $\theta$  au niveau du centre de fusion. Par conséquent selon [6], on a l'inégalité suivante pour l'erreur quadratique moyenne (EQM) d'estimation

$$\mathbb{E}[(\hat{\theta} - \theta)^2] \geq \text{BCR} = \frac{1}{F}, \quad (7)$$

avec  $F$  la matrice (ici juste un scalaire) d'information de Fisher. Comme les données  $\{y_k(t)\}_{k \in \{1, \dots, K\}, t \in \{1, \dots, T\}}$  et finalement aussi  $\{q_k(t)\}_{k \in \{1, \dots, K\}, t \in \{1, \dots, T\}}$  sont iid, la matrice d'information de Fisher se décompose de la manière suivante

$$F = KTF_d + F_a \quad (8)$$

avec  $F_d$  Fisher d'une donnée d'un capteur reçue au centre de fusion (donc  $q_k(t)$ ) et  $F_a$  le Fisher de l'a priori.

Le second terme lié à l'a priori est défini de la manière suivante

$$F_a = \mathbb{E}_\theta \left[ \left( \frac{\partial \ln p_\Theta(\theta)}{\partial \theta} \right)^2 \right]. \quad (9)$$

En raison de l'hypothèse gaussienne de l'a priori, nous obtenons classiquement

$$F_a = \frac{1}{\sigma_\theta^2}. \quad (10)$$

Concentrons-nous maintenant sur  $F_d$ . Comme les données sont iid, on peut calculer cette information sur Fisher à partir de n'importe quel échantillon  $q_k(t)$ . Nous choisirons  $q = q_1(1)$ . De ce fait, on a

$$F_d = \mathbb{E}_{Q, \Theta} \left[ \left( \frac{\partial \log p_{Q, \Theta}(q, \theta)}{\partial \theta} \right)^2 \right] \quad (11)$$

$$= \mathbb{E}_\Theta \left[ \mathbb{E}_{Q|\Theta} \left[ \left( \frac{\partial \log p_{Q|\Theta}(q|\theta)}{\partial \theta} \right)^2 \right] \right] \quad (12)$$

avec  $\mathbb{E}_X$  l'espérance mathématique agissant sur la variable aléatoire  $X$ . Dans la suite, on posera

$$G_d(\theta) = \mathbb{E}_{Q|\Theta} \left[ \left( \frac{\partial \log p_{Q|\Theta}(q|\theta)}{\partial \theta} \right)^2 \right]. \quad (13)$$

Un simple développement de l'espérance mathématique (effectué dans [3]) permet de trouver que

$$G_d(\theta) = \sum_{i=0}^L \frac{1}{p_{Q|\Theta}(q=i|\theta)} \cdot \left( \frac{\partial p_{Q|\Theta}(q=i|\theta)}{\partial \theta} \right)^2 \quad (14)$$

Il convient maintenant de calculer le terme  $p_{Q|\Theta}(q=i|\theta)$ . Pour cela, on remarque que la valeur de  $q$  ne dépend que de la valeur d'un  $\hat{\theta}$  (on omet maintenant les indices  $k$  et  $t$  car on ne se focalise sur un d'entre eux). Ainsi

$$p_{Q|\Theta}(q=i|\theta) = \int p_{Q|\hat{\Theta}}(q=i|\hat{\theta}) p_{\hat{\Theta}|\Theta}(\hat{\theta}|\theta) d\hat{\theta} \quad (15)$$

Etant donné le modèle (1), nous savons que la variable aléatoire  $\hat{\theta}|\theta$  est gaussienne de moyenne  $\theta$  et de variance  $c^2 \sigma_w^2/M$ .

En découpant l'intégrale présente dans (15) sur les intervalles de quantificateur, on obtient immédiatement que

$$p_{Q|\Theta}(q = i|\theta) = \Phi\left(\frac{\tau_i - c\theta}{c\sigma_w/\sqrt{M}}\right) - \Phi\left(\frac{\tau_{i-1} - c\theta}{c\sigma_w/\sqrt{M}}\right) \quad (16)$$

avec  $\Phi$  la fonction de répartition d'une gaussienne centrée unitaire. Par convention, on pose  $\Phi(-\infty) = 0$  et  $\Phi(+\infty) = 1$ . Par conséquent, on obtient

$$\frac{\partial p_{Q|\Theta}(q = i|\theta)}{\partial \theta} = \frac{\sqrt{M}}{\sigma_w} \left[ \phi\left(\frac{\tau_{i-1} - c\theta}{c\sigma_w/\sqrt{M}}\right) - \phi\left(\frac{\tau_i - c\theta}{c\sigma_w/\sqrt{M}}\right) \right] \quad (17)$$

avec  $\phi$  la densité de probabilité d'une gaussienne centrée unitaire.

Dans notre cas, on analysera les BCR avec le quantificateur uniforme optimal défini de la manière suivante

$$\tau^* = \arg \max_{\tau} F_d. \quad (18)$$

Ainsi le pas de quantificateur s'adaptera à la qualité moyenne de l'estimation et on évitera le phénomène de résonance [7]. Cette équation sera résolue par une recherche mono-dimensionnelle. Notez que (18) a également été résolue numériquement dans le cadre d'un quantificateur non-uniforme que très récemment [8].

## 4 Lien avec d'autres BCR

Dans la littérature, les autres de BCR sont soit calculées selon la même approche qu'à la section précédente avec soit  $K = 1$  soit  $T = 1$ . Comme dit en introduction, une autre approche est aussi présente basée sur la théorie de l'information et agnostique au quantificateur. Dans la suite, nous présentons les deux bornes existantes et nous les adaptons à notre contexte et à nos notations. Les bornes obtenues sont pour une approche fréquentiste et non bayésienne du problème d'estimation et donc sans a priori sur le paramètre recherché.

Dans [4], la borne développée qui nous est utile est globalement donnée dans le premier théorème du papier. Néanmoins une relecture attentive des preuves est nécessaire pour enlever tout ce qui est inutile à notre cas. Ainsi on a

$$\mathbb{E}[(\hat{\theta} - \theta)^2] \geq c_1 \min\{1, \text{BCR}_1, \text{BCR}_2\} \quad (19)$$

avec

$$\text{BCR}_1 = \frac{\sigma_w^2}{c_2 M \log(KT)}$$

$$\text{BCR}_2 = \frac{\sigma_w^2}{c_3 K T M \min\{c_4 b \log(KT), 1\}}$$

et  $c_1 = 0.25$ ,  $c_2 = 16^2$ ,  $c_3 = 6^2$  et  $c_4 = 2^{11}$ . Comme on a toujours  $c_4 b \log(KT) \geq 1$ , on obtient une version plus simple

$$\mathbb{E}[(\hat{\theta} - \theta)^2] \geq c_1 \min\left\{1, \frac{\sigma_w^2}{c_2 M \log(KT)}, \frac{\sigma_w^2}{c_3 K T M}\right\}. \quad (20)$$

On remarque que le produit  $KT$  joue un rôle et donc les capteurs et les trames ont le même impact. Ensuite la taille de la

trame joue un rôle. Quand  $KT > \mathcal{W}_{-1}(-9/64)/(-9/64) \approx 21.97$  avec  $\mathcal{W}_{-1}$  la fonction  $W$  de Lambert, la borne est alors proportionnelle à  $\sigma_w^2/(KTM)$  avec un facteur de proportionnalité de 1/144. C'est donc la borne du problème centralisé à ce facteur près et donc l'influence du quantificateur n'apparaît plus ce qui est dommage et de plus la borne n'est pas précise en raison de ce facteur égal à deux ordres de grandeur.

Dans [5], les auteurs ont fait les calculs avec  $M = 1$  et alors l'estimateur  $\hat{\theta}_k(t) = y_k(t)$ . On peut facilement adapter leur approche avec  $M > 1$  et alors

$$\hat{\theta}_k(t) = \frac{1}{M} \sum_{m=1}^M y_k(tM + m) \quad (21)$$

qui est l'estimateur du maximum de vraisemblance (dans l'approche fréquentiste, d'où l'absence du coefficient  $c$ ). La variance de cet estimateur étant  $\sigma_w^2/M$ , la borne de [5] (via une bon assemblage des premier et deuxième corollaires du papier cité) vaut

$$\mathbb{E}[(\hat{\theta} - \theta)^2] \geq \frac{\sigma_w^2/M}{KT \min\{1, \frac{32}{3}b\}} = \frac{\sigma_w^2}{KTM}. \quad (22)$$

De nouveau, on remarque que la borne obtenue appliquée à notre contexte conduit à effacer complètement l'impact du quantificateur ce qui induit que cette borne est trop optimiste car égale à celle du cas centralisé.

## 5 Résultats numériques

Dans cette section, par simplicité, nous considérons  $K = 1$ . Nous avons posé  $\sigma_w^2 = 0.01$  et  $\sigma_\theta^2 = 0.01$  ce qui implique que l'a priori est faible et donc les comparaisons entre les BCR fréquentistes et les BCR bayésiennes auront un sens dès que le nombre d'échantillons n'est pas trop petit. Enfin nous allons étudier deux configurations.

**Configuration 1 :** On considère que le nombre de bits par trame ( $b$ ) est fixé. Par conséquent, le nombre total de bits échangé est  $bT$  et évolue avec  $T$ .

Sur la figure 1, on trace la BCR en fonction de  $M$  avec  $N = 50$  pour différentes valeurs de  $b$ . On remarque que l'on a intérêt à envoyer des bits à chaque échantillon de chaque capteur. Ainsi l'agrégation n'est pas une bonne option dans ce cas. Elle s'avère même catastrophique quand le nombre de bits est faible car alors malgré l'optimisation du quantificateur, l'erreur de quantification domine. Bref dans ce cas, la taille de trame optimale est 1 et  $N$  bits seront envoyés au total.

Sur la figure 2, on trace la BCR en fonction de  $N$  avec une trame englobant toutes les données ( $N = M$  ou  $T = 1$ ) pour différentes valeurs de  $b$  et aussi les BCR obtenues dans la littérature (cf. (20) et (22)). On remarque que les bornes de la littérature ne sont pas très précises quand le nombre de données augmente car elles n'intègrent pas bien le quantificateur. A très faible nombre d'échantillons, le croisement est dû au terme associé à l'a priori pour les bornes calculées dans ce papier. Par

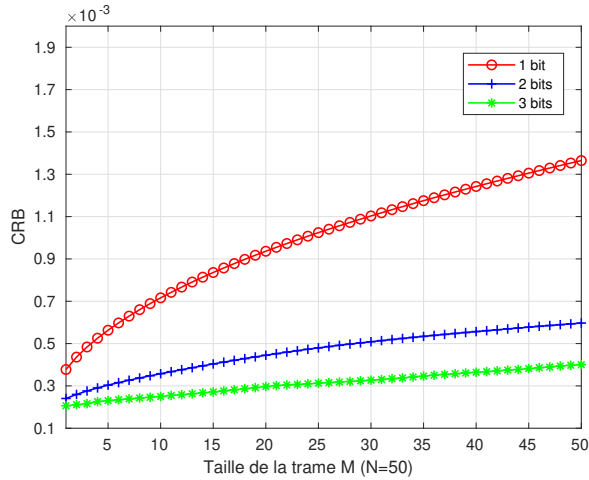


FIGURE 1 – BCR en fonction de  $M$  (config 1).

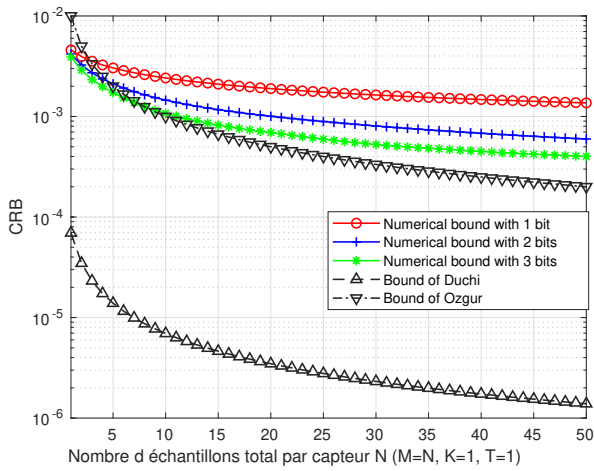


FIGURE 2 – BCR en fonction de  $N$  (avec  $M = N$  et config 1).

conséquent, l'approche suivie qui consiste à calculer des BCR dépendant du quantificateur choisi nous paraît plus judicieuse.

**Configuration 2 :** On considère que le nombre total de bits échangés ( $B$ ) est fixé. Par conséquent, le nombre de bits par trame vaut  $B/T$  et varie en fonction du nombre de trames considéré. Par souci de simplicité, on choisit aussi  $N = B$ . Donc pour chaque trame émise de taille  $M (= N/T)$ , on envoie  $N/T$  (et donc  $M$ ) bits.

Sur la figure 3, on trace la BCR en fonction de  $M$  (avec  $N = B = 10$ ) pour différents quantificateurs : le quantificateur uniforme optimal, le quantificateur uniforme optimal pour  $M = 2$  appliqué aux autres valeurs de  $M$  et l'absence de quantificateur. On remarque que dans ce cas, la trame doit être longue et en fait cela conduit à émettre une seule trame ce qui est le résultat opposé de la configuration 1. Quand le quantificateur n'est pas optimalement adapté, alors il y a même une valeur optimale de la taille de la trame qui n'est ni de taille 1

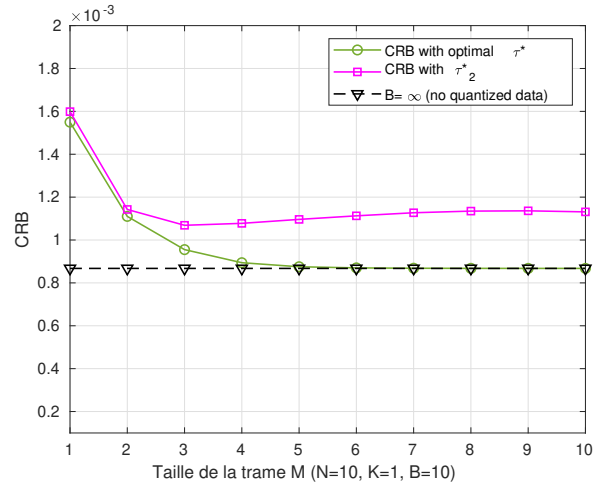


FIGURE 3 – BCR en fonction de  $M$  (avec  $N = 10$  et config 2).

ni de taille  $N$ . De plus la borne sans quantificateur est peu ou prou atteinte quand la trame correspond à la moitié des données ( $M = 5$ ).

## Références

- [1] A. Ribeiro, G.B. Giannakis, "Bandwidth-Constrained distributed estimation for wireless sensor networks - part 1 : Gaussian case", *IEEE Trans. on Signal Processing*, vol. 54, no. 3, pp. 1131-1143, Mar 2006.
- [2] P. Stoica, X. Shang, Y. Cheng, "The Cramer-Rao Bound for Signal Parameter estimation from quantized data", *IEEE Signal Processing Magazine*, pp. 118-125, Jan 2022.
- [3] A. Sani, A. Vosoughi, "On the performance limit for bandwidth-constrained distributed parameter estimation systems", *Conference on Information Sciences and Systems (CISS)*, Mar. 2022.
- [4] J. Duchi, M. Jordan, M. Wainwright, Y. Zhang, "Optimally guaranteed for distributed statistical estimation", *Arxiv preprint*, Jun 2014.
- [5] L. Barnes, Y. Han, A. Özgür, "Lower bounds for learning distributions under communications constraints via Fisher Information", *Journal of Machine Learning Research*, vol. 21, pp. 1-30, 2020.
- [6] H. Van Trees, "Detection, Estimation and Modulation Theory", Wiley, 1968.
- [7] A. Mezghani, F. Antreich, J. Nosske, "Multiple parameter estimation with quantized channel output", *Workshop on Smart Antennas (WSA)*, Oct. 2010.
- [8] Y. Cheng, X. Shang, J. Li, P. Stoica, "Interval design for signal parameter estimation from quantized data", *IEEE Trans. on Signal Processing*, pp. 6011-6020, Dec. 2022.