À propos d'extensions de l'information de Fisher et de l'inégalité de Cramér-Rao

Jean-François Bercher

À propos d'extensions de l'information de Fisher et de l'inégalité de Cramér-Rao

Jean-François BERCHER,

Laboratoire d'Informatique Gaspard Monge
Université Paris-Est, ESIEE, Cité Descartes 93162 Noisy-le-Grand Cedex
Jean-Francois.Bercher@univ-paris-est.fr

Résumé – On présente quelques extensions de résultats classiques de théorie de l'estimation, conduisant à définir de nouvelles formes de l'information de Fisher et à des généralisations de la borne de Cramér-Rao. On indique que ceci suggère une variation autour du maximum de vraisemblance. Ce cadre conduit également à une nouvelle caractérisation des gaussiennes généralisées, ainsi qu'à l'extension de certaines identités (de Bruijn) et inégalités (relations d'incertitudes). On retrouve en cas particulier les résultats habituels pour la gaussienne standard.

Abstract – We present some extensions of classical results from estimation theory. This leads to new forms of Fisher information and to generalizations of the Cramér-Rao bound. We show that this suggests modifications of the maximum likelihood. This setting also yields a new characterization of generalized Gaussians, as well as the extension of some classical identity (de Bruijn) and inequalities (uncertainty relations). Classical results for the standard Gaussian are recovered as a particular case.

Dans cette communication, l'auteur souhaite présenter une synthèse de certains résultats très récents, poursuivant les travaux entamés en [1], et complétée par plusieurs résultats inédits (notamment la formulation générale (2), le résultat matriciel (4) et certaines relations d'incertitude) et une discussion.

La communication présente dans un premier temps des variations autour de la borne de Cramér-Rao pour l'estimation d'un paramètre statistique, qui est un problème central pour tout traiteur de données. On donne en particulier une formulation valable pour des normes et des ordres à peu près quelconques, ce qui permet par exemple d'étendre les résultats habituels à des normes L_p, ou à des combinaisons de normes, ou encore à des normes pondérées. D'autre part, on propose de mesurer l'erreur par rapport à une densité de probabilité générale, qui n'est pas nécessairement la densité paramétrique des données, ce qui fournit un paramètre libre permettant, par exemple, de favoriser telle caractéristiques des paramètres. Ceci nous amène alors à définir une extension de l'information de Fisher, qui intervient dans la formulation des bornes. On indique que ceci suggère alors de possibles adaptations de la technique du maximum de vraisemblance, qui conduisent à des estimateurs robustes.

Dans un second temps, on examine le cas particulier d'un paramètre de localisation, et on montre qu'en choisissant les deux densités mises en jeu comme une paire de distributions escortes, alors l'inégalité de Cramér-Rao multidimensionnelle se réduit alors à une inégalité fonctionnelle où la borne est atteinte pour des distributions gaussiennes généralisées. Ces distributions, de type Student multivariées, se réduisent à la gaussienne habituelle en cas particulier et sont importantes dans le cadre de la caractérisation de signaux et de bruits non-gaussiens, caractérisés par exemple par des lois à queues lourdes, les densités gaussiennes généralisées ont reçu une attention particulière. De tels modèles apparaissent par exemple utiles comme statistiques des coefficients d'ondelettes, sont utilisés en codage vidéo ou dans des techniques de filtrage et reconstruction d'images, ou comme modèles non gaussiens en communication. Notons encore que ces distributions apparaissent comme solutions d'équations de diffusion non linéaires, et dans des problèmes d'inégalités de Sobolev sur $\mathbb{R}^n$ $(n \geq 2)$ et bien entendu en physique statistique “nonextensive” [2] où on retrouve ce type de distributions comme solutions d'un problème de maximisation d'entropie de Rényi-Tsallis.

Dans un troisième temps, nous donnons quelques propriétés supplémentaires de l'information de Fisher généralisée, en particulier le fait qu'elle apparaisse naturellement dans une identité de type de Bruijn, pour des distributions satisfaisant une certaine équation de diffusion non-linéaire et une entropie de type Rényi-Tsallis. Enfin, nous nous tournons vers l'analyse de Fourier chère aux traiteurs de signaux et montrons que les inégalités de Cramér-Rao générales décrites ici permettent d'obtenir de nouvelles relations d'incertitude généralisant les relations d'incertitude de type Weyl-Gabor-Heisenberg.

Soient donc $f(x;\theta)$ et $g(x;\theta)$ deux densités de probabilité, avec $x \in X \subseteq \mathbb{R}^k$ et θ un paramètre de ces densités $\theta \in \mathbb{R}^n$. On s'intéresse à l'estimation d'un vecteur $h(\theta) \in \mathbb{R}^m$, et on notera alors $\hat{h}(x)$ l'estimateur correspondant. Si $\|.\|$ est une norme arbitraire sur $\mathbb{R}^m$, la norme duale $\|.\|_*$ est définie par $\|Y\|_* = \sup_{\|X\| \leq 1} X.Y$, où $X.Y$ est le produit scalaire habituel. Notamment, si $\|.\|$ est une norme L_p, alors $\|.\|_*$ est une norme L_q, avec $p^{-1}+q^{-1}=1$. L'idée de base menant à l'inégalité de Cramér-Rao généralisée est qu'il peut être utile d'évaluer des moments différents de l'erreur d'estimation par rapport à des densités différentes. Typiquement, si l'erreur est $\hat{h}(x)-h(\theta)$, alors le biais peut être évalué selon $B_f(h(\theta)) = \mathrm{E}_f\left[\hat{h}(x)-h(\theta)\right]$ alors qu'un moment général d'un norme arbitraire de l'erreur peut être calculé par rapport à une autre distribution de probabilité, $g(x;\theta)$, comme dans

$$\mathrm{E}_g\left[\left\|\hat{\theta}(x)-\theta\right\|^\beta\right] = \int_X \left\|\hat{\theta}(x)-\theta\right\|^\beta g(x;\theta)\,\mathrm{d}x.$$

Les deux distributions $f(x;\theta)$ and $g(x,\theta)$ peuvent être choisies de manière tout-à-fait arbitraire. Toutefois, on peut construire $g(x;\theta)$ comme une transformation de $f(x;\theta)$ qui mette en exergue, ou atténue, certaines caractéristiques de $f(x;\theta)$. À titre d'exemples, $g(x;\theta)$ peut être une version pondérée de $f(x;\theta)$, i.e. $g(x;\theta)=h(x;\theta)f(x;\theta)$, ou encore une version quantifiée $g(x;\theta)=[f(x;\theta)]$. Enfin, un cas de figure important est le cas où $f(x;\theta)$ et $g(x;\theta)$ forment une paire de distributions escortes d'ordre q, où q est un paramètre réel positif, selon

$$f(x;\theta)=\frac{g(x;\theta)^q}{\int g(x;\theta)^q\mathrm{d}x} \quad \text{et} \quad g(x;\theta)=\frac{f(x;\theta)^{\bar{q}}}{\int f(x;\theta)^{\bar{q}}\mathrm{d}x}, \tag{1}$$

avec $\bar{q}=1/q$, et pourvu bien entendu que les intégrales mises en jeu soient finies. Ces distributions escortes sont très utilisées en physique statistique non extensive, sont apparues dans le domaine des multifractales et ont reçu des applications jusque dans des problèmes de codage de source. Dans ces conditions, et sous un certain nombre d'hypothèses techniques, on aura :

Proposition 1. *[Inégalité de type Cramér-Rao, partiellement dans [3]] Pour tout estimateur $\hat{h}(x)$ de $h(\theta)$,*

$$E_g\left[\left\|\hat{h}(x)-h(\theta)\right\|^\alpha\right]^{\frac{1}{\alpha}} E_g\left[\left\|H\,\frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}\right\|_*^\beta\right]^{\frac{1}{\beta}} \geq \left|m+\nabla_{h(\theta)}.\,B_f(h(\theta))\right| \tag{2}$$

où α et β sont conjugués de Hölder , i.e. $\alpha^{-1}+\beta^{-1}=1$, $\alpha\geq 1$, avec $H_{ij}=\partial\theta_j/\partial h(\theta)_i$, et où le second facteur du membre de gauche correspond à une information de Fisher étendue qu'on notera $I_\beta[f_\theta|g_\theta;h(\theta)]^{\frac{1}{\beta}}$.

De nombreuses conséquences peuvent être obtenues à partir de là. Si par exemple on choisit $h(\theta)=\theta$, et un estimateur $\hat{h}(x)=\hat{\theta}(x)$ sans biais, alors l'inégalité précédente se réduit à

$$E_g\left[\left\|\hat{\theta}(x)-\theta\right\|^\alpha\right]^{\frac{1}{\alpha}} E_g\left[\left\|\frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}\right\|_*^\beta\right]^{\frac{1}{\beta}} \geq n. \tag{3}$$

Si on prend $f=g$, on obtient une extension de l'inégalité de Cramér-Rao habituelle pour une norme quelconque et une puissance arbitraire ; dans le cas scalaire, l'inégalité avec une puissance quelconque, bien que peu connue, a été donnée par [4, 5]. Lorsque $h(\theta)$ est à valeurs scalaires, on peut établir la variante suivante, qui fait apparaître, malheureusement implicitement, une matrice d'information de Fisher :

Proposition 2. *Soit $\eta(\theta)=E_f[\hat{h}(X)]$. Alors, pour toute matrice $A>0$*

$$E_g\left[\left|\hat{h}(X)-h(\theta)\right|^\alpha\right]^{\frac{1}{\alpha}} \geq \sup_{A>0} \frac{\dot{\eta}(\theta)^T A\,\dot{\eta}(\theta)}{E_g\left[\left|\dot{\eta}(\theta)^T A\,\psi_g(X;\theta)\right|^\beta\right]^{\frac{1}{\beta}}}. \tag{4}$$

où $\psi_g=\frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}$ est une fonction score, et avec égalité si et seulement si

$$\dot{\eta}(\theta)^T A\,\psi_g(x;\theta)=c(\theta)\mathrm{sign}(\hat{h}(x)-h(\theta))\left|\hat{h}(x)-h(\theta)\right|^{\alpha-1},$$

avec $c(\theta)>0$.

La démonstration de cette proposition est assez simple et donnée ci-dessous.

Démonstration. Posons $\eta(\theta)=E_f[\hat{h}(X)]$. Dans la suite, on fait l'hypothèse classique qu'il est possible de permuter les opérations d'intégration et de dérivation, ce qui impose des des conditions de régularité sur f et sur g. On observe alors que la moyenne de la fonction score est nulle : $E_g[\psi_g(x;\theta)] = \frac{\mathrm{d}}{\mathrm{d}\theta}\int_X f(x;\theta)\,\mathrm{d}x=0$. En dérivant $\eta(\theta)=E_f[\hat{h}(X)]$ par rapport à chacune de ses composantes θ_i, on obtient alors que

$$\begin{aligned}\dot{\eta}(\theta)=\nabla_\theta\eta(\theta) &= \nabla_\theta\int_X \hat{h}(x)\,f(x;\theta)\,\mathrm{d}x \\ &= \int_X \hat{h}(x)\,\frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}\,g(x;\theta)\,\mathrm{d}x \\ &= \int_X \left(\hat{h}(x)-h(\theta)\right)\,\psi_g(x;\theta)\,g(x;\theta)\,\mathrm{d}x.\end{aligned}$$

Pour toute matrice définie positive A, la multiplication à gauche par $\dot{\eta}(\theta)^T A$ fournit

$$\dot{\eta}(\theta)^T A\,\dot{\eta}(\theta)=\int_X\left(\hat{h}(x)-h(\theta)\right)\,\dot{\eta}(\theta)^T A\,\psi_g(x;\theta)\,g(x;\theta)\,\mathrm{d}x.$$

Par l'inégalité de Hölder, on obtient alors l'inégalité

$$E_g\left\|\hat{h}(x)-h(\theta)\right|^\alpha\Big]^{\frac{1}{\alpha}} E_g\left[\left|\dot{\eta}(\theta)^T A\,\psi_g(x;\theta)\right|^\beta\right]^{\frac{1}{\beta}} \geq \dot{\eta}(\theta)^T A\,\dot{\eta}(\theta),$$

avec égalité si et seulement si $\left(\hat{h}(x)-h(\theta)\right)\,\dot{\eta}(\theta)^T A\,\psi_g(x;\theta) > 0$ d'une part, et $\left|\hat{h}(x)-h(\theta)\right|^\alpha = k(\theta)\left|\dot{\eta}(\theta)^T A\,\psi_g(x;\theta)\right|^\beta$, $k(\theta)>0$, d'autre part. Cette inégalité nous fournit ainsi une borne inférieure (4) pour un moment d'ordre $\alpha>1$ quelconque de l'erreur d'estimation, ce moment étant calculé sous une distribution g quelconque. □

On définit alors comme matrice d'information de Fisher l'inverse de la matrice A qui maximise le membre de droite. Dans le cas où $\alpha=\beta=2$, et en utilisant l'inégalité $(x^t x)^2 \leq (x^t Bx)\,(x^t B^{-1}x)$ valide pour toute matrice B définie non négative, on peut vérifier que le maximum est atteint pour $A^{-1}=I_{2,g}(\theta)=E_g[\psi_g\psi_g^t]$, qui a la même structure que la matrice d'information de Fisher habituelle et que l'inégalité se réduit alors au résultat $E_g\left[\left|\hat{h}(X)-h(\theta)\right|^2\right] \geq \dot{\eta}(\theta)^T I_{2,g}(\theta)^{-1}\,\dot{\eta}(\theta)$.

Dans le cas scalaire, on obtient immédiatement le cas particulier suivant en simplifiant par A et $\dot{\eta}$ (le résultat s'obtient également à partir de (2))

Corollaire 3. *Dans le cas scalaire, (ou en ne considérant qu'une seule composante de θ), on a*

$$E_g\left[\left|\hat{h}(X)-h(\theta)\right|^\alpha\right]^{\frac{1}{\alpha}} \geq \frac{|\dot{\eta}(\theta)|}{E_g\left[|\psi_g(X;\theta)|^\beta\right]^{\frac{1}{\beta}}}, \qquad (5)$$

avec égalité si et seulement si $\psi_g(x;\theta) = c(\theta)\mathrm{sign}(\hat{h}(x) - h(\theta))\left|\hat{h}(x)-h(\theta)\right|^{\alpha-1}$.

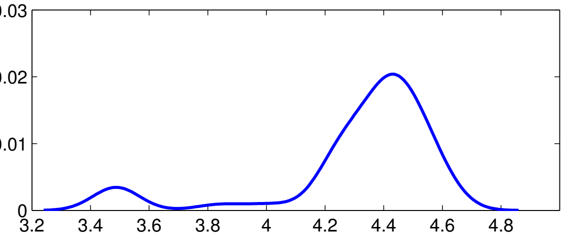
Dans le cas où f et g sont deux distributions escortes liées par (1), la fonction score se réduit à $\psi_g(x;\theta) = d(\theta)\, f(x;\theta)^{-\bar{q}}\partial_\theta f$, où ∂_θ désigne la dérivée partielle par rapport à θ et $d(\theta) > 0$; et la condition d'égalité conduit alors, avec $q \neq 1$ et $K(\theta) > 0$, à

$$\nabla_\theta f(x;\theta)^{1-\bar{q}} = K(\theta)(1-\bar{q})\mathrm{sign}(\hat{h}(x)-h(\theta))\left|\hat{h}(x)-h(\theta)\right|^{\alpha-1}$$

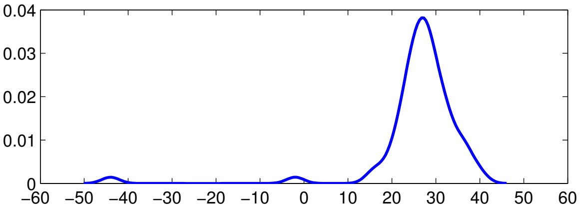
ce qui fournit après intégration la forme générique des densités qui permettent d'atteindre la borne (5). Cependant, ce résultat ne fournit pas directement d'algorithme, où de forme d'estimateur, qui permettrait d'atteindre la borne. On peut toutefois noter que si la borne est atteinte, alors l'égalité précédente suggère de maximiser la distribution escorte f de g. En effet, si on maximise $f(x;\theta)$ en un certain $\theta^*(x)$, alors $\partial_\theta f_\theta = 0$, et par suite $\hat{h}(x) = h(\theta^*(x))$. Ainsi, la procédure suggère donc de rechercher les paramètres comme maximisant la distribution escorte f d'ordre q de la vraisemblance g. Bien entendu, dans le cas $q = 1$, on retrouve le maximum de vraisemblance habituel. Ce type de procédure a été proposé récemment, avec une argumentation différente, dans [6]. Il a été en particulier montré par Ferrari *et al.* que ceci conduit à des estimateurs robustes à des contaminations des données. Nous illustrerons ceci par deux exemples.

Les données présentées correspondent d'une part (a) aux mesures du logarithme de la température effective à la surface des différentes étoiles de l'amas d'étoiles CYG OB1, qui contient 47 étoiles dans la direction de Cygnus (ces données sont issues de [7, p. 27, table 3], et d'autre part (b) à des mesures de temps de passage de la lumière, mesures utilisées afin de d'estimer la vitesse de la lumière [8]. Les densités de probabilité, estimées par noyaux gaussiens sont données sur la figure 1 ; elles font apparaître des modes de faible amplitude, liés à la présence de quelques données aberrantes (*outliers*).

On estime les paramètres de localisation et d'écart-type en maximisant la distribution compagne, pour différentes valeurs du paramètre q. Dans notre cas, la distribution est assimilée à une Gaussienne avec contamination, et les deux estimateurs sont pris comme les estimateurs standards pour la Gaussienne. On tient en outre compte du fait que la variance estimée doit être corrigée d'un facteur q. Les résultats sont donnés sur le tableau 1, qui montre clairement que les estimateurs ainsi définis affichent une robustesse vis-à-vis des données aberrantes, ce qui était effectivement le comportement recherché, et se rapprochent des estimées que l'on obtient en supprimant manuellement les données identifiées comme « aberrantes » (*trimming*).



(a)



(b)

FIGURE 1 – Densité de probabilité (a) de la température effective pour les étoiles de CYG OB1 [7] (b) de mesures de temps de passage de la lumière [8]

q	1	1.4	2	4
m	4.31	4.36	4.39	4.42
σ	0.298	0.27	0.22	0.15

(a)

q	1	1.4	2	4
m	26.21	27.18	27.37	27.12
σ	10.85	8.04	5.83	4.6

(b)

TABLE 1 – Estimées de la moyenne et de l'écart type pour différentes valeurs du paramètre q

——————o0o——————

Tournons nous maintenant vers le cas particulier d'un paramètre de localisation, qui nous conduit d'une part à une famille d'inégalités fonctionnelles, et d'autre part à une nouvelle caractérisation des gaussiennes généralisées. Soit donc un paramètre de localisation $\theta \in \mathbb{R}^n$ pour une famille $f(x;\theta) = f(x-\theta)$. Dans ce cas, on a $\nabla_\theta f(x;\theta) = -\nabla_x f(x-\theta)$. On suppose en outre que $f(x)$ est de moyenne nulle. Dans ces conditions, en prenant $\hat{\theta}(x) = x$, l'estimateur est non-biaisé. En prenant enfin $\theta = 0$, la relation (3) conduit au résultat suivant.

Proposition 4. *[Inégalité fonctionnelle de Cramér-Rao] Pour toute paire de densités de probabilités de probabilité et sous quelques conditions techniques,*

$$\left(\int_X \|x\|^\alpha\, g(x)\ dx\right)^{\frac{1}{\alpha}} \left(\int_X \left\|\frac{\nabla_x f(x)}{g(x)}\right\|_*^\beta g(x)\ dx\right)^{\frac{1}{\beta}} \geq n, \qquad (6)$$

avec égalité si $\nabla_x f(x) = -K\, g(x)\|x\|^{\alpha-1}\nabla_x\|x\|$.

On peut dès lors en déduire une nouvelle caractérisation très intéressante des distributions gaussiennes généralisées, qui fait intervenir l'information de Fisher étendue. En effet, en prenant $f(x)$ et $g(x)$ comme une paire de distributions escortes,

Proposition 5. *[Inégalité de q-Cramér-Rao [3]] Pour toute densité g, si $m_\alpha[g] = E\left[\|x\|^\alpha\right]$ désigne le moment d'ordre α de la norme de x, et si*

$$I_{\beta,q}\left[g\right] = \left(q/M_q\left[g\right]\right)^\beta \; E_g\left[g(x)^{\beta(q-1)} \left\|\nabla_x \ln g(x)\right\|_*^\beta\right] \quad (7)$$

désigne l'information de Fisher d'ordre (β, q), avec $M_q[g] = \int g(x)^q dx$, alors

$$m_\alpha[g]^{\frac{1}{\alpha}} \, I_{\beta,q}\left[g\right]^{\frac{1}{\beta}} \geq n \quad (8)$$

avec égalité si et seulement si $g(x)$ est une "gaussienne généralisée" de la forme

$$g(x) \propto \left(1 - \gamma(q-1)\|x\|^\alpha\right)_+^{\frac{1}{q-1}} \quad (9)$$

Notons simplement ici que $g(x)$ est aussi appelée une q-gaussienne étirée, devient une gaussienne étirée lorsque $q = 1$, et une gaussienne standard lorsqu'en outre $\alpha = 2$. L'importance et les domaines d'apparition de ces gaussiennes généralisées a été évoquée plus haut. Il est bien connu que ces gaussiennes généralisées peuvent être caractérisées comme les maximisantes d'une q-entropie sous contrainte de moment. L'inégalité précédente fournit une caractérisation complémentaire : en effet, elle étend à l'information de Fisher généralisée et aux gaussiennes généralisées le résultat bien connu selon lequel le minimum de l'information de Fisher à moment d'ordre 2 fixé est atteint pour la gaussienne. Notons encore que l'inégalité (8) est similaire (mais non identique) à une inégalité donnée par Lutwak et al [9] et également saturée par les gaussiennes généralisées (9).

L'information de Fisher généralisée (7) surgit également dans une extension de l'identité de de Bruijn. Celle-ci est en général obtenue pour les solutions d'une équation de la chaleur. Une extension est obtenue en considérant les solutions d'une équation de diffusion doublement non linéaire [10] (cette équation de la chaleur non linéaire est utilisée en traitement d'images)

$$\frac{\partial}{\partial t} f = \Delta_\beta f^m = \operatorname{div}\left(\left|\nabla f^m\right|^{\beta-2} \nabla f^m\right). \quad (10)$$

En calculant ainsi la dérivée de l'entropie de Tsallis pour les solutions de cette équation de diffusion non linéaire, et en utilisant l'identité de Green, on arrive au résultat suivant.

Proposition 6. *[Identité de de Bruijn étendue [11]] Pour $q = m + 1 - \frac{\alpha}{\beta}$, $M_q[f] = \int f^q$ et $S_q[f] = \frac{1}{1-q}\left(M_q[f] - 1\right)$ l'entropie de Tsallis, on a*

$$\frac{d}{dt} S_q[f] = \left(\frac{m}{q}\right)^{\beta-1} M_q[f]^\beta \, I_{\beta,q}[f]. \quad (11)$$

Bien entendu, l'identité de Bruijn habituelle est retrouvée dans le cas particulier $\alpha = \beta = 2$, et $q = m = 1$.

On termine cet article en indiquant qu'il est possible d'obtenir de nouvelles relations d'incertitude à partir de l'inégalité de Cramér-Rao générale (8). Ces inégalités mettent en jeu des moments calculés par rapport à des distributions escortes telles que (1). On note ainsi $E_q[.]$ une moyenne calculée par rapport à une escorte d'ordre q. Si ψ est une fonction d'onde, x et ξ deux variables conjuguées au sens de la transformée de Fourier, alors on obtient par exemple

Proposition 7. *[Relations d'incertitude] Pour $k = 2/\left(2q-1\right)$, $\lambda = n(q-1)+1$, et $\gamma \geq 2$, $\theta \geq 2$,*

$$\frac{M_{\frac{k}{2}}\left[|\psi|^2\right]^{\frac{1}{2}}}{M_{\frac{kq}{2}}\left[|\psi|^2\right]} \; E_{\frac{k}{2}}\left[\|x\|_2^\gamma\right]^{\frac{1}{\gamma}} \, E\left[\|\xi\|_2^\theta\right]^{\frac{1}{\theta}} \geq \frac{n}{2\pi kq}, \quad (12)$$

Pour $\gamma = \theta = 2$, la borne inférieure est atteinte uniquement si $|\psi|$ est sous forme d'une gaussienne généralisée. Pour $\gamma = \theta = 2$, $q = 1$, cette inégalité fournit une version multidimensionnelle du principe d'incertitude de Weyl-Heisenberg :

$$E\left[\|x\|_2^2\right] E\left[\|\xi\|_2^2\right] \geq \left(\frac{n}{2\pi}\right)^2, \quad (13)$$

Pour $\frac{3}{2} - \frac{1}{\beta} > q$, on a également une inégalité du type

$$\left(E_{\frac{k}{2}}\left[\|x\|_2^\gamma\right]\right)^{\frac{1}{\gamma}} \left(E\left[\|\xi\|_2^\theta\right]\right)^{\frac{1}{\theta\lambda}} > K. \quad (14)$$

Références

[1] J.-F. Bercher, "Quelques inégalités caractérisant les gaussiennes généralisées," in *30th Conference Gretsi*, Bordeaux, France, 2011. [Online]. Available : http://hal-univ-mlv.archives-ouvertes.fr/hal-00766555

[2] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics*. Springer, Apr. 2009.

[3] J.-F. Bercher, "On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q-Gaussian distributions," *Journal of Physics A : Mathematical and Theoretical*, vol. 46, no. 9, p. 095303, Feb. 2013. [Online]. Available : http://hal-upec-upem.archives-ouvertes.fr/hal-00766695

[4] I. Vajda, "χ^α-divergence and generalized Fisher information," in *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, 1973, pp. 873–886.

[5] E. Weinstein and A. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 338–342, 1988.

[6] D. Ferrari, "Maximum lq-likelihood estimation," *The Annals of Statistics*, vol. 38, no. 2, pp. 753–783, Apr. 2010.

[7] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*, ser. Wiley Series in Probability and Statistics. Wiley, 2005.

[8] S. M. Stigler, "Do robust estimators work with real data ?" *The Annals of Statistics*, pp. 1055–1098, 1977.

[9] E. Lutwak, S. Lv, D. Yang, and G. Zhang, "Extensions of Fisher information and Stam's inequality," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1319 –1327, Mar. 2012.

[10] J. L. Vázquez, *Smoothing and Decay Estimates for Nonlinear Diffusion Equations : Equations of Porous Medium Type*. Oxford University Press, USA, Oct. 2006.

[11] J.-F. Bercher, "Some properties of generalized Fisher information in the context of nonextensive thermostatistics," *Physica A : Statistical Mechanics and its Applications*, vol. 392, no. 15, pp. 3140–3154, Aug. 2013. [Online]. Available : http://hal-upec-upem.archives-ouvertes.fr/hal-00766699