La borne bayésienne de Schützenberger-van Trees : Un principe d'incertitude sur l'a posteriori

Olivier RIOUL[1] Alexandre RENAUX[2]

[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, France

[2]Université Paris-Saclay, CNRS, CentraleSupelec, Laboratoire des Signaux et Systèmes, France

Résumé – Cette communication propose une perspective historique sur la borne de Cramér-Rao bayésienne (BCRB), généralement attribuée à van Trees qui l'a découverte en 1968. Selon la loi de Stigler sur l'éponymie, aucune découverte scientifique ne porte le nom de son premier découvreur. C'est non seulement le cas de la borne Cramér-Rao elle-même – due notamment aux mathématiciens français Fréchet et Darmois – mais aussi de l'inégalité de van Trees. En effet, le médecin, généticien, épidémiologiste et mathématicien français Marcel-Paul (Marco) Schützenberger, dans un petit article d'une quinzaine de lignes seulement, écrit en 1956 – plus d'une décennie avant van Trees – avait non seulement démontré la BCRB, mais comme le montre une lecture approfondie de sa preuve, l'avait fait avec une démarche très originale en la reliant au principe d'incertitude de Weyl-Heisenberg sur l'a posteriori. Nous passons en revue et comparons les contributions de Schützenberger et de van Trees ainsi que celles de Gart en 1959. L'équivalence générale entre BCRB et principe d'incertitude sur l'a posteriori ouvre de nouvelles perspectives.

Abstract – This paper offers a historical perspective on the Bayesian Cramér-Rao Bound (BCRB), generally attributed to van Trees, who discovered it in 1968. According to Stigler's law of eponymy, no scientific discovery is named after its first discoverer. This is the case not only for the Cramér-Rao bound itself—due in particular to the French mathematicians Fréchet and Darmois—but also for the van Trees inequality: The French physician, geneticist, epidemiologist and mathematician Marcel-Paul (Marco) Schützenberger, in a paper of just fifteen lines written in 1956—more than a decade before van Trees—had not only demonstrated the BCRB but, as a close examination of his proof shows, used a very original approach based on the Weyl-Heisenberg uncertainty principle on the posterior. We review and compare the contributions of Schützenberger and van Trees, as well as those of Gart in 1959. The general equivalence between BCRB and the uncertainty principle opens up new perspectives.

1 Introduction

1.1 L'inégalité dite de Cramér-Rao

La borne de Cramér-Rao est un outil important en traitement du signal car elle permet, sous un modèle d'observation paramétrique, d'évaluer facilement les meilleures performances possibles en termes d'erreur quadratique moyenne (EQM), pouvant être atteintes exactement ou asymptotiquement (si on y ajoute une preuve d'efficacité). Dans sa forme la plus simple, pour un estimateur non biaisé $\hat{\theta}(x)$ du paramètre $\theta \in \mathbb{R}$ calculé sur les observations $x = (x_1, x_2, \dots, x_n)$ — réalisations de l'échantillon $X = (X_1, X_2, \dots, X_n)$ — et sous certaines hypothèses de régularité du modèle souvent vérifiées, l'EQM (égale ici à la variance de $\hat{\theta}(X)$) admet la borne

$$\mathrm{EQM}_\theta = \mathbb{V}(\hat{\theta}(X)) = \mathbb{E}\{(\hat{\theta}(X)-\theta)^2\} \geqslant \frac{1}{J_\theta} \tag{1}$$

donnée par l'inverse de l'*information de Fisher* définie par

$$J_\theta = \mathbb{E}\Big\{\Big(\frac{\partial \log p_\theta(X)}{\partial\theta}\Big)^2\Big\} \tag{2}$$

où $p_\theta(x) = p_\theta(x_1, x_2, \dots, x_n)$ décrit le modèle paramétrique de la loi conjointe des n observations. Un calcul classique donne (toujours sous les conditions de régularité habituelles)

$$J_\theta = -\,\mathbb{E}\Big\{\frac{\partial^2 \log p_\theta(X)}{\partial\theta^2}\Big\}. \tag{3}$$

Pour des observations i.i.d. on a simplement $J_\theta = nJ_{\theta,1}$ où $J_{\theta,1}$ est l'information de Fisher pour une seule observation.

C'est un fait connu aujourd'hui des chercheurs du domaine – en tout cas en France [4] – que l'inégalité (1) fut établie par le mathématicien français Maurice Fréchet dès l'hiver 1939 dans son cours à l'Institut Henri Poincaré, et publiée en 1943 dans le cas i.i.d. [10][1]. L'extension au cas non i.i.d. et au cas vectoriel de plusieurs paramètres $\theta \in \mathbb{R}^d$ est faite par son collègue Georges Darmois en 1945 [6][2]. La même borne (dans le cas scalaire et vectoriel) est établie indépendamment par C. R. Rao [13] en 1945 et par Cramér dans son excellent livre [5, Chap. 32] en 1946, et s'appelle aujourd'hui la borne de Cramér-Rao. La preuve est identique pour tous ces auteurs : elle se fonde sur l'inégalité de Cauchy-Schwarz :

$$\mathbb{E}\big\{(\hat{\theta}(X)-\theta)^2\big\}\,\mathbb{E}\Big\{\Big(\frac{\partial \log p_\theta(X)}{\partial\theta}\Big)^2\Big\} \geqslant \Big[\mathbb{E}\Big\{(\hat{\theta}(X)-\theta)\frac{\partial \log p_\theta(X)}{\partial\theta}\Big\}\Big]^2 \tag{4}$$

où un calcul simple montre que la covariance au second membre se réduit simplement à $\mathbb{E}\big(\hat{\theta}(X)\frac{\frac{\partial p_\theta}{\partial\theta}(X)}{p_\theta(X)}\big) = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}(\hat{\theta}(X)) = \frac{\mathrm{d}\theta}{\mathrm{d}\theta} = 1$ puisque l'estimateur est non biaisé. Fréchet [10] et Cramér [5] ont également étendu cette inégalité au cas d'un estimateur *biaisé*, de biais $B_{\hat{\theta}}(\theta) = \mathbb{E}(\hat{\theta}(X)) - \theta$ non nul, auquel cas $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}(\hat{\theta}(X)) = \frac{\mathrm{d}}{\mathrm{d}\theta}(\theta + B_{\hat{\theta}}(\theta))$, d'où la

[1]Des considérations similaires avaient étudiées auparavant (notamment par Pearson, Edgeworth, Fisher…) dans un cadre asympotique ou normal.

[2]Contrairement à ce qui est parfois affirmé, leur étudiant Daniel Dugué n'avait pas, semble-t-il, déjà établi cette inégalité dans sa thèse de 1937 [8].

borne (1) modifiée avec la dérivée du biais :

$$\mathrm{EQM}_\theta = \mathbb{V}(\hat{\theta}(X)) + B^2_{\hat{\theta}}(\theta) \geqslant \frac{(1 + B'_{\hat{\theta}}(\theta))^2}{J_\theta} + B^2_{\hat{\theta}}(\theta). \quad (5)$$

L'intérêt d'une telle borne est limitée puisqu'elle dépend en général de l'estimateur lui-même *via* son biais. Par la suite, de nombreuses autres versions de la borne de Cramér-Rao ont permis de prendre en compte des conditions de régularité différentes du cadre classique [3], des contraintes dans le vecteur de paramètres [12], une contrainte de périodicité inhérente à certain problèmes d'estimation [18], et plus récemment la structure géométrique des paramètres [22]. La borne de Cramér-Rao a connu pléthore d'applications à des problèmes d'ingénierie, parfois à la limite de l'abus [23].

1.2 L'inégalité dite de van Trees

Une des extensions les plus importantes de la borne de Cramér-Rao est la *borne de Cramér-Rao bayésienne* (BCRB) dans un contexte bayésien où le paramètre d'intérêt θ est supposé aléatoire et muni d'une loi connue *a priori* $p(\theta)$. Le modèle de la loi des données $p(x|\theta)$ dépend de la variable θ et l'EQM n'est plus calculé pour une vraie valeur de θ fixée, mais moyennée sur l'*a priori* :

$$\mathrm{EQM} = \mathbb{E}_{X,\theta}\big((\hat{\theta}(X) - \theta)^2\big) \quad (6)$$

Le changement notable par rapport au cas classique précédent est que la loi sur laquelle on prend l'espérance est maintenant la loi *conjointe* $p(x, \theta) = p(x|\theta)p(\theta)$ du couple (X, θ) (plutôt que $p_\theta(x) = p(x|\theta)$ pour θ fixé).

La BCRB est quasiment toujours attribuée à van Trees qui l'a démontré dans son ouvrage de référence [24] publié en 1968, aussi bien dans le cas scalaire $\theta \in \mathbb{R}$ que vectoriel $\theta \in \mathbb{R}^d$ où l'information de Fisher devient une matrice de covariance de taille $d \times d$. Dans le cas scalaire, elle s'écrit

$$\mathrm{EQM} \geqslant \frac{1}{J} = \frac{1}{\tilde{J} + \mathbb{E}_\theta\, J_\theta} \quad (7)$$

où J est l'information de Fisher *conjointe* ou complète :

$$J = \mathbb{E}_{X,\theta}\Big\{\Big(\frac{\partial \log p(X, \theta)}{\partial \theta}\Big)^2\Big\}. \quad (8)$$

Un calcul identique à celui qui démontre (3) à partir de (2) donne[3] :

$$J = -\,\mathbb{E}_{X,\theta}\Big\{\frac{\partial^2 \log p(X, \theta)}{\partial \theta^2}\Big\}. \quad (9)$$

Puisque $p(x, \theta) = p(\theta)p(x|\theta)$ sous le logarithme, on a la relation $J = \tilde{J} + \mathbb{E}_\theta\, J_\theta$ où

$$\tilde{J} = \mathbb{E}_\theta\Big\{\Big(\frac{\mathrm{d} \log p(\theta)}{\mathrm{d}\theta}\Big)^2\Big\} = -\,\mathbb{E}_\theta\Big\{\frac{\mathrm{d}^2 \log p(\theta)}{\mathrm{d}\theta^2}\Big\}. \quad (10)$$

est appelé information de Fisher de l'*a priori*, et où J_θ est l'information de Fisher classique donné par (2)-(3) pour $p_\theta(x) = p(x|\theta)$, qui vaut $J_\theta = nJ_{\theta,1}$ dans le cas d'observations i.i.d. (conditionnellement à θ)[4]. Le livre de référence [25] contient de nombreux exemples d'application de la BCRB.

[3]Avec des hypothèses de régularité et de décroissance convenables sur l'*a priori* $p(\theta)$.

[4]On voit en particulier que l'influence de l'*a priori* finit par disparaître pour un très grand nombre d'observations car $J = \tilde{J} + n\,\mathbb{E}_\theta J_{\theta,1} \sim \mathbb{E}_\theta\, J_\theta$ quand $n \to \infty$, qui se réduit à la borne classique si J_θ ne dépend pas de θ.

La démonstration de van Trees est directement inspirée de l'inégalité de Cauchy-Schwarz (4) utilisée dans le cas classique, mais appliquée à la loi conjointe $p(x, \theta)$:

$$\mathbb{E}_{X,\theta}\big\{(\hat{\theta}(X) - \theta)^2\big\}\, \mathbb{E}_{X,\theta}\Big\{\Big(\frac{\partial \log p(X, \theta)}{\partial \theta}\Big)^2\Big\} \geqslant \Big[\mathbb{E}_{X,\theta}\Big\{(\hat{\theta}(X) - \theta)\frac{\partial \log p(X, \theta)}{\partial \theta}\Big\}\Big]^2 \quad (11)$$

Le terme de covariance au second membre se réécrit ici[5] $\iint (\hat{\theta}(x) - \theta)\frac{\partial p(x,\theta)}{\partial \theta} = \iint p(x, \theta) + \int \frac{\partial}{\partial \theta}\int (\hat{\theta}(x) - \theta) p(x, \theta) = 1 + \int \frac{\partial}{\partial \theta}\{p(\theta)B_\theta(\theta)\} = 1$ sous l'hypothèse suivante sur le biais :

$$\lim_{|\theta| \to \infty} p(\theta) B_{\hat{\theta}}(\theta) = 0, \quad (12)$$

hypothèse cruciale sous laquelle van Trees démontre l'inégalité (7). De plus, la condition d'égalité dans l'inégalité de Cauchy-Schwarz (11) permet à van Trees de montrer que la BCRB est atteinte lorsque l'*a posteriori* $p(\theta|x)$ est gaussienne.

Tous ces résultats obtenus par van Trees sont devenus classiques aujourd'hui, et sont repris tels quels dans la plupart des manuels sur le sujet (par exemple [15]). Mais peu de chercheurs savent qu'il a été précédé de plus d'une décennie par Marcel-Paul Schützenberger.

2 Marcel-Paul Schützenberger

Ce n'est qu'en 2007, lors de l'édition d'un recueil d'articles sur les bornes bayésiennes [25], que van Trees mentionne que sa borne avait été trouvée indépendamment par « Shutzenberger » (sic) et commente : « *Cette dernière dérivation est un modèle d'économie (1/3 de page) mais ne semble pas avoir été remarquée par les communautés d'ingénieurs ou de statisticiens.* » La figure 1 montre le tiers de page en question, un petit paragraphe d'un quinzaine de lignes. C'est en fait une simple annonce dans le bulletin de l'AMS qui fera un peu plus tard l'objet d'un article plus détaillé en français [20].

321*t*. M. P. Schützenberger: ***A generalization of the Frĕchet-Cramĕr inequality to the case of Bayes estimation.***

Let $f(x)$ be the a priori density function of x; $g(y|x)$ the conditional density function of y. For fixed x, the set of n independent y-variates is represented by z. The density function of z is $f'(z)$ and $g'(x|z)$ is the a posteriori density function of x, for given z. The a posteriori variance of the Bayes estimate is $v_z^2 = \int (x - \bar{x})^2 g'(x|z)dx$ and $v^2 = E_z v_z^2 = \int v_z^2 f'(z)dz$ is its average over z. $F = \int (\partial f(x)/\partial x)^2\, (f(x))^{-1}dx$; $G = E_x G_x$ with $G_x = \int ((\partial/\partial x)g(y|x))^2 (g(y|x))^{-1}dy$; $G' = E_z G'_z$ with $G'_z = \int ((\partial/\partial x)y'(x|z))^2 (g(x|z))^{-1}dx$. The usual assumptions on f and g, which insure that F, G_x, G'_z are finite are made. Since $O = F' = \int ((\partial/\partial x)f'(z))^2 (f'(z))^{-1}dz$, it is easily seen that $F + nG = G'$ (Third London Symposium on Information Theory, 1955, p. 18). Furthermore, it is a classical result that $v_z^2 G'_z \geqq 1$. Thus $v^2 = E_z v_z^2 \geqq (E_z 1/v_z^2)^{-1} \geqq (E_z G'_z)^{-1} = (F + nG)^{-1}$, which is the desired inequality that tends to the usual form when n goes to infinity. It reduces to an equality if and only if $v^2 = v_z^2 = (G'_z)^{-1}$ for all z, that is, if and only if $g'(x|z)$ is gaussian with variance independent of z. If, furthermore, $y - x = t$ has a distribution $h(t)$ independent of x, this implies that $f(x)$ and $h(t)$ are also gaussian. (This work was supported in part by the Army (Signal Corps), the Air Force (Office of Scientific Research, Air Research and Development Command), and the Navy (Office of Naval Research).) (Received November 5, 1956.)

FIGURE 1 : L'intégralité de l'article [19] écrit en 1956.

L'auteur de cet article est en réalité Marcel-Paul Schützenberger, médecin, généticien, épidémiologiste et mathématicien français, personnage haut en couleurs [16] qui avait soutenu sa thèse en 1953 sous la direction de Georges Darmois, après avoir fait sa thèse en médecine intitulée *Contribution à l'étude*

[5]Les intégrales sont prises par rapport aux mesures dominantes adéquates par rapport auxquelles on définit respectivement les densités du vecteur d'observation X et du paramètre θ.

du sexe à la naissance. Déjà dans sa thèse de 1953, il établit des connexions profondes entre statistique (information de Fisher notamment) et théorie de l'information (information mutuelle de Shannon). Par une démonstration originale, il trouve la fameuse *inégalité de Pinsker* avec des constantes optimales au deuxième ordre, 7 ans avant Pinsker lui-même et 17 ans avant la republication précise de son inégalité par Kullback (voir la référence [16] pour plus de détails).

Schützenberger n'ignorait pas le travail de Maurice Fréchet, son président du jury de thèse, sur la borne dite de Cramér-Rao qu'il appelait d'ailleurs *inégalité de Fréchet-Cramér*. C'est invité par Claude Shannon au MIT pendant l'année universitaire 1956-57 qu'il écrit son article sur cette inégalité dans le contexte bayésien. Il est probable que son inspiration provient notamment de discussions avec David Slepian, qui travaillait aux laboratoires Bell à l'époque sur des problèmes d'estimation [21] et connaissait bien la borne de Cramér-Rao via le livre de Cramér. Slepian est en effet mentionné dans une note de bas de page par Schützenberger [20] comme ayant « obtenu indépendamment » la borne bayésienne, bien que son travail n'ait apparemment pas été publié.

Une seconde personne est également mentionnée par Schützenberger dans [20], comme ayant obtenu indépendamment la BCRB : « Mr J. Dard (article à paraitre dans *Annals of Mathematical Statistics*) ». Il semble qu'il s'agisse en fait de John J. Gart, dont l'article est effectivement publié dans cette revue en 1959 [11]. Cet article est également intégré à la fin du recueil de van Trees [25] sans trop de commentaires. La preuve de Gart réécrit simplement l'inégalité du cas classique (4) en moyennant aussi sur l'*a priori* $p(\theta)$, pour obtenir une version bayésienne de (5) où apparaît la dérivée du biais $B_{\hat{\theta}}(\theta)$ et où n'apparaît pas l'information *a priori* $\tilde{J}$:

$$\text{EQM} \geqslant \frac{(1 + \mathbb{E}_\theta\, B'_{\hat{\theta}}(\theta))^2}{\mathbb{E}_\theta\, J_\theta}. \tag{13}$$

Il traite aussi du cas vectoriel $\theta \in \mathbb{R}^d$, mais se contente des termes diagonaux de la matrice de Fisher sans tenir compte les corrélations inter-paramètres.

Comme on va le voir, le travail de Schützenberger lui-même sur la BCRB est particulièrement intéressant : c'est l'un de ses tous derniers dans le domaine des statistiques, sur lequel il n'est jamais revenu et qu'il n'a guère évoqué par la suite, se focalisant dès son passage au MIT sur la théorie des codes, des langages formels, des automates, de la combinatoire des mots, etc. autant de domaines de l'informatique théorique et de la combinatoire pour lesquels il est aujourd'hui le plus connu.

Tombée dans l'oubli, sa démonstration de la BCRB est pourtant étonnamment originale. En effet, un décryptage attentif de son article (Fig. 1) montre que l'accent est mis sur la loi *a posteriori* $p(\theta|x)$ plutôt qu'*a priori*, et sur une version de l'*information de Fisher a posteriori*

$$\tilde{J}(x) = \mathbb{E}_{\theta|x}\Big\{\Big(\frac{\partial \log p(\theta|x)}{\partial\theta}\Big)^2\Big\} = -\,\mathbb{E}_{\theta|x}\Big\{\frac{\partial^2 \log p(\theta|x)}{\partial\theta^2}\Big\} \tag{14}$$

qui n'est semble-t-il mentionné quasiment nulle part dans la littérature et vérifie la relation $J = \mathbb{E}_X\, \tilde{J}(X)$, comme on le voit en décomposant $p(x,\theta) = p(x)p(\theta|x)$: la loi inconditionnelle des données $p(x)$ disparaît dans la dérivation car elle ne dépend pas du paramètre θ.

L'introduction de l'*a posteriori* est en effet naturelle puisque l'estimateur optimal, qui minimise l'EQM, est précisément donné par la moyenne de cette loi $\hat{\theta}^*(x) = \mathbb{E}(\theta|x)$. Schützenberger commence donc par montrer qu'il suffit de prouver la BCRB (7) sur l'EQM minimale

$$\text{EQMM} = \mathbb{E}_{X,\theta}\{(\theta - \mathbb{E}(\theta|X))^2\} = \mathbb{E}_X\, \mathbb{V}(\theta|X) \tag{15}$$

où $\mathbb{V}(\theta|x)) = \mathbb{E}_{\theta|x}\{(\theta - \mathbb{E}(\theta|x))^2\}$ est la variance de l'*a posteriori*. L'étape cruciale de la démonstration de Schützenberger est l'inégalité qui ressemble à la BCRB, mais pour un vecteur d'observation donnée :

$$\mathbb{V}(\theta|x) \geqslant \frac{1}{\tilde{J}(x)}. \tag{16}$$

Schützenberger dit simplement que c'est un « résultat classique » (*classical result*, cf. Figure 1) dans [19]... Heureusement, il précise dans [20] que c'est l'« inégalité de Weyl ».

Bien que ce ne soit pas évident, on peut effectivement le voir comme l'inégalité de Weyl-Heisenberg qui constitute le célèbre *principe d'incertitude* en mécanique quantique ! En effet, Weyl démontre dans son livre de 1928 sur la mécanique quantique [27, App. 1] l'inégalité suivante, que nous écrivons ici pour une fonction $f(t)$ réelle vérifiant $\int f^2(t)\,\mathrm{d}t = 1$ et de transformée de Fourier $\hat{f}$ (formellement $\hat{f}(\nu) = \int f(t)e^{-2i\pi t\nu}\,\mathrm{d}t)$:

$$\sigma_t\sigma_\nu \geqslant \tfrac{1}{4\pi} \tag{17}$$

où σ_t désigne l'écart-type de la distribution $|f(t)|^2$, et où σ_ν désigne l'écart-type de la distribution $|\hat{f}(\nu)|^2$, qui est bien normalisée par la relation de Parseval : $\int |f(t)|^2\,\mathrm{d}t = \int |\hat{f}(\nu)|^2\,\mathrm{d}\nu = 1$. La démonstration de cette inégalité par Weyl (ainsi que le cas d'égalité) est classique : l'inégalité équivaut à l'inégalité de Cauchy-Schwarz suivante faisant intervenir la dérivée f' de f :

$$\begin{aligned}\big(\textstyle\int t^2 f^2(t)\,\mathrm{d}t\big)\big(\int f'^2(t)\,\mathrm{d}t\big) &\geqslant \big(\textstyle\int t f(t) f'(t)\,\mathrm{d}t\big)^2 \\ &= \big(-\tfrac{1}{2}\textstyle\int f^2(t)\,\mathrm{d}t\big)^2 = \tfrac{1}{4}\end{aligned} \tag{18}$$

où la dernière inégalité résulte d'une intégration par parties dont on montre que le terme tout intégré est nul dès lors que le premier membre existe et est fini. On voit facilement que le cas d'égalité correspond au cas où $f(t)$ est gaussienne en résolvant une équation différentielle du premier ordre.

Même s'il ne le détaille pas, le point clé de la démonstration de Schützenberger consiste à l'appliquer à la fonction

$$f(\theta) = \sqrt{p(\theta|x)} \tag{19}$$

pour x donné. L'inégalité de Weyl, après un changement de variable de translation sur $\theta \leftarrow (\theta - \mathbb{E}(\theta|x))$ qui ne modifie que la première intégrale, équivaut à

$$\int \big(\theta - \mathbb{E}(\theta|x)\big)^2 p(\theta|x)\,\mathrm{d}\theta \int \Big(\frac{\partial\sqrt{p(\theta|x)}}{\partial\theta}\Big)^2 \mathrm{d}\theta \geqslant \frac{1}{4} \tag{20}$$

Or on a $\Big(\frac{\partial\sqrt{p(\theta|x)}}{\partial\theta}\Big)^2 = \Big(\frac{1}{2}\frac{\frac{\partial p(\theta|x)}{\partial\theta}}{\sqrt{p(\theta|x)}}\Big)^2 = \frac{1}{4}\frac{\big(\frac{\partial p(\theta|x)}{\partial\theta}\big)^2}{p(\theta|x)} = \frac{1}{4}\big(\frac{\partial \log p(\theta|x)}{\partial\theta}\big)^2 p(\theta|x)$, de sorte que *l'inégalité de Weyl appliqué à* $\sqrt{p(\theta|x)}$ *équivaut bien à la BCRB* (16) *pour* x *donné*. Finalement, il suffit de prendre l'espérance sur la loi inconditionnelle $p(x)$ pour obtenir la BCRB :

$$\text{EQMM} = \mathbb{E}_X \mathbb{V}(\theta|X) \geqslant \mathbb{E}_X \frac{1}{\tilde{J}(X)} \geqslant \frac{1}{\mathbb{E}_X \tilde{J}(X)} = \frac{1}{J} \tag{21}$$

où la dernière inégalité provient de la convexité de la fonction $1/x$ pour $x > 0$. Schützenberger écrit (comme van Trees une décennie plus tard) cette borne sous la forme (7) avec $\frac{1}{J} = \frac{1}{\tilde{J} + n\,\mathbb{E}_\theta J_{\theta,1}}$ pour des observations i.i.d. [19,20] (cf. Figure 1).

3 Conclusion et perspectives

Il est dommage que le travail pionnier de Schützenberger soit tombé dans l'oubli si longtemps. Il faut dire que le résumé en anglais [19] n'est pas évident à décrypter et l'article [20], à peine un peu plus long, a certainement eu le défaut d'être écrit en français. Ce dernier article n'a réellement aujourd'hui que trois citations, deux du russe Alexander Veretennikov [1, 26] (qui l'a obtenu d'Hélène Schützenberger, fille de Marcel-Paul) et un présenté au GRETSI plus récemment [2], qui reprend les deux inégalités présentes dans (21).

La preuve de Schützenberger a pourtant plusieurs avantages sur la démarche habituelle, du fait qu'elle équivaut à une inégalité d'incertitude sur l'a posteriori : D'abord, contrairement à van Trees (ou Gart), elle ne suppose pas des conditions sur le biais, telles que (12). En effet, l'inégalité de Weyl est vérifiée automatiquement dès lors que les quantités en jeu dans la BCRB sont finies. Ensuite, la condition d'égalité devient évidente puisque c'est celle de l'inégalité de Weyl-Heisenberg, à savoir que l'*a posteriori* doit être gaussien. Enfin, on peut montrer qu'elle se généralise facilement (comme dans le cas de van Trees) au cas vectoriel avec des matrices de Fisher [17].

Il faut mentionner ici que l'idée d'une ressemblance ou d'une équivalence entre borne de Cramér-Rao et principe d'incertitude, présente implicitement chez Schützenberger, est réapparue plusieurs fois indépendamment par la suite. Dès 1972, on parle de la borne classique de Cramér-Rao en disant qu'elle « ressemble à celle de Heisenberg » [9, p. 198] (sans plus de précision). En 1991, Dembo *et al.* [7] démontrent une équivalence entre la borne de Cramér-Rao classique et le principe d'incertitude, étendu au cas vectoriel, mais uniquement pour un paramètre de localisation $p(x|\theta) = p(x-\theta)$. Il semble que ce genre d'équivalence n'est *pas* possible en général pour la borne de Cramér-Rao (non bayésienne) et qu'il faut se trouver dans le cas bayésien pour l'obtenir.

En conclusion, l'apport de Marcel-Paul Schützenberger, plus d'une décennie avant van Trees, met l'accent sur le fait que le BCRB n'est rien d'autre qu'un *principe d'incertitude sur la loi a posteriori*, plus précisément appliquée à $\sqrt{p(\theta|x)}$. Cela ouvre ainsi de nouvelles perspectives sur de nouvelles bornes basées sur des améliorations ou des variantes des inégalités d'incertitude.

Références

[1] R. ABU-SHANAB et A. Yu. VERETENNIKOV : On asymptotic Borovkov–Sakhanenko inequality with unbounded parameter set. *Theor. Probability and Math. Statist.*, 90:1–12, 2015 (version en russe, 2014).

[2] Lucien BACHARACH, Carsten FRITSCHE, Umut ORGUNER et Éric CHAUMETTE : Une borne de Cramér-Rao bayésienne plus précise. *In Proc. GRETSI*, Lille, France, Aug. 2019.

[3] Y. BAR-SHALOM, R. W. OSBORNE, P. WILLETT et F. E. DAUM : Cramér-Rao-Leibniz lower bound — A new estimation bound for finite support measurement noise. *In 53rd IEEE Conf. Decision Control*, pages 2609–2614, L.A., CA, USA, 2014.

[4] Frédéric BARBARESCO : Les densités de probabilité « distinguées » et l'équation d'Alexis Clairaut : Regards croisés de Maurice Fréchet et de Jean-Louis Koszul. *In Proc. GRETSI*, pages 521–524, Juan-Les-Pins, France, 2017.

[5] Harald CRAMÉR : *Mathematical Methods of Statistics*, volume 9. Princeton Univ. Press, septembre 1946.

[6] Georges DARMOIS : Sur les limites de la dispersion de certaines estimations. *Revue Inst. Int. Stat.*, 13(1/4):9–15, 1945.

[7] Amir DEMBO, Thomas M. COVER et Joy A. THOMAS : Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, Nov. 1991.

[8] Daniel DUGUÉ : Application des propriétés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation. *Journal de l'Ecole Polytechnique*, 3(4):305–372, 1937.

[9] Claude FOURGEAUD et Aimé FUCHS : *Statistique*. Dunod, 1972.

[10] Maurice FRÉCHET : Sur l'extension de certaines évaluations statistiques au cas de petit échantillons. *Revue de l'Institut International de Statistique*, 11(3/4):182–205, 1943.

[11] John J. GART : An extension of the Cramér-Rao inequality. *Ann. Math. Statist.*, 30:367–380, juin 1959.

[12] J. D. GORMAN et A. O. HERO : Lower bounds for parametric estimation with constraints. *IEEE Transactions on Information Theory*, 36(6):1285–1301, novembre 1990.

[13] Calyampudi Radhakrishna RAO : Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[14] Olivier RIOUL : Yet another proof of the entropy power inequality. *IEEE Transactions on Information Theory*, 63(6):3595–3599, June 2017.

[15] Olivier RIOUL : *Estimation Paramétrique*. Spartacus-IDH, Sept. 2022.

[16] Olivier RIOUL : A historical perspective on Schützenberger-Pinsker inequalities (extended version). *Information Geometry*, 7:S737–S779, 2024.

[17] Olivier RIOUL : A historical perspective on the Schützenberger-van Trees inequality : A posterior uncertainty principle. *In Proc. 7th conference on Geometric Science of Information (GSI'25)*, Lecture Notes in Computer Science. Springer, Oct. 2025.

[18] T. ROUTTENBERG et J. TABRIKIAN : Non-bayesian periodic Cramér-Rao bound. *IEEE Transactions on Signal Processing*, 61(4):1019–1032, février 2013.

[19] Marcel-Paul SCHÜTZENBERGER : A generalization of the Fréchet-Cramér inequality to the case of Bayes estimation. *Bulletin of the American Mathematical Society*, 63:142, 1957.

[20] Marcel-Paul SCHÜTZENBERGER : A propos de l'inégalité de Fréchet-Cramer. *Publ. Inst. Statist. Univ. Paris*, 7:3–6, 1958.

[21] David S. SLEPIAN : Estimation of signal parameters in the presence of noise. *Trans. IRE Professional Group Inf. Theory*, 3(3):68–69, 1954.

[22] S. T. SMITH : Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans.Signal Proc.*, 53(5):1610–1630, mai 2005.

[23] M. VALLISNERI : Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Phys. Rev. D*, 77, juin 2008.

[24] Harry L. VAN TREES : *Detection, Estimation and Modulation Theory*, volume 1. John Wiley & Sons, 1968.

[25] Harry L. VAN TREES et Kristine L. BELL, éditeurs. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. IEEE Press, John Wiley & Sons, 2007.

[26] Alexander VERETENNIKOV : On asymptotic information integral inequalities. *Theory of Stochastic Processes*, 13(29)(1-2):294–307, 2007.

[27] Hermann WEYL : *Gruppentheorie und Quantenmechanik*. S.Hirzel Verlag, 1928.