Les densités de probabilité « distinguées » et l'équation d'Alexis Clairaut: regards croisés de Maurice Fréchet et de Jean-Louis Koszul

FREDERIC BARBARESCO[1]

[1] THALES AIR SYSTEMS, Advanced Radar Concepts Department
Voie Pierre-Gilles de Gennes, 91470 Limours, France

[1]frederic.barbaresco@thalesgroup.com

Résumé – En 1943, Maurice Fréchet a écrit un article introduisant ce qui fut appelée ensuite la borne de Cramer-Rao. Cet article peu lu, contient en fait beaucoup plus que cette importante découverte. En particulier, Maurice Fréchet y introduit des notions plus générales relatives aux « densités distinguées », pour lesquelles l'estimateur atteint cette borne, définies comme dépendantes d'une fonction solution de l'équation d'Alexis Clairaut. Les solutions « enveloppe de l'équation de Clairaut » sont équivalentes à la transformée de Legendre standard sans condition de convexité, mais uniquement sous contrainte de dérivabilité. Cette analyse de Fréchet peut être revisitée sur la base des travaux de Jean-Louis Koszul comme prémisse aux fondements de la « Géométrie de l'Information ».

Abstract - In 1943, Maurice Fréchet wrote a seminal paper introducing what was then called the Cramer-Rao bound. This paper contains in fact more than this important discovery. In particular, Maurice Fréchet introduces more general notions relative to "distinguished functions", densities with estimator reaching the bound, defined with a function, solution of Clairaut's equation. The solutions "envelope of the Clairaut's equation" are equivalent to standard Legendre transform without convexity constraints but only smoothness assumption. This Fréchet's analysis can be revisited on the basis of Jean-Louis Koszul works as seminal foundation of "Information Geometry".

1 Préambule

En 1943, deux ans avant Rao, Maurice Fréchet publie un article [4], dans lequel il introduit la borne qui portera abusivement le nom de borne de Cramer-Rao (la publication de Rao datant de 1945). D'autant plus que comme Fréchet le précise en annotation de son article, ce contenu est déjà présent dans son cours de l'hiver 1939 de l'Institut Henri Poincaré : « *Le contenu de ce mémoire a formé une partie de notre cours de statistique mathématique à l'Institut Henri Poincaré pendant l'hiver 1939-1940* ». L'article traite le cas monovarié, mais son collègue Georges Darmois publie la version multivariée en 1945. Il serait donc plus exacte d'appeler cette borne, la borne de Fréchet-Darmois.

Au-delà de la découverte de cette borne, Fréchet étudie les densités de probabilité dont l'estimateur atteint cette borne. Ils les appellent « ***densités distinguées*** ». Il montre que ces fonctions dépendent d'une fonction solution de l'équation d'Alexis Clairaut [5], et dont on peut établir les liens avec la transformée de Legendre. Cette transformation de Clairaut-Legendre constitue un élément des fondements de la géométrie de l'information dont Fréchet entrouvre la porte, en montrant que l'inverse de la matrice de Fisher est le hessien d'une fonction, définie elle-même comme la transformée de Legendre d'une seconde fonction, que nous identifierons avec le logarithme de la fonction caractéristique [3]. Nous replacerons alors le modèle de Fréchet dans un cadre théorique plus récent introduit par le mathématicien Jean-Louis Koszul [1,2], qui a établi les équations générales liant ce hessien à la métrique du cône convexe associé. Cette métrique est à la base de la théorie de la « Géométrie de l'Information » [3].

2 Les densités distinguées de Fréchet

Posons les notations de Maurice Fréchet, où il considère l'estimateur: $T = H(X_1,...,X_n)$ (1)

et la variable aléatoire $A(X)=\dfrac{\partial \log p_\theta(X)}{\partial \theta}$ (2)

à laquelle il associe : $U=\sum_i A(X_i)$ (3)

La contrainte $\int_{-\infty}^{+\infty} p_\theta(x)dx=1$ impose :

$\int_{-\infty}^{+\infty}...\int_{-\infty}^{+\infty}\prod_i p_\theta(x_i)dx_i=1$ dont la dérivée par rapport à θ

$\int_{-\infty}^{+\infty}...\int_{-\infty}^{+\infty}\left[\sum_i A(x_i)\right]\prod_i p_\theta(x_i)dx_i=0$ donne : $E_\theta[U]=0$ (4)

De même, en supposant que $E_\theta[T]=\theta$, c'est-à-dire que $\int_{-\infty}^{+\infty}...\int_{-\infty}^{+\infty}H(x_1,...,x_n)\prod_i p_\theta(x_i)dx_i=\theta$, on obtient en dérivant par rapport à θ : $E[(T-\theta)U]=1$ (5)

Or, puisque $E[T]=\theta$ et $E[U]=0$, on a :

$E[(T-E[T])(U-E[U])]=1$ (6)

Or d'après l'inégalité de Schwarz : $[E(ZT)]^2 \le E[Z^2]E[T^2]$

$1 \le E[(T-E[T])^2]E[(U-E[U])^2]=(\sigma_T\sigma_U)^2$ (7)

U étant la somme de variables indépendantes, l'égalité de Bienaymé donne :

$$(\sigma_U)^2 = \sum_i \left[\sigma_{A(X_i)}\right]^2 = n(\sigma_A)^2 \quad (8)$$

Ou encore la borne recherchée : $(\sigma_T)^2 \geq \dfrac{1}{n(\sigma_A)^2}$ (9)

Fréchet remarque que c'est une inégalité remarquable où le second membre est indépendant du choix de la fonction H définissant la « valeur empirique » T et où dans le premier membre, on peut prendre pour T toute valeur empirique $T = H(X_1,...,X_n)$ assujettie à la seule condition $E_\theta[T] = \theta$ quel que soit θ.

La condition classique pour que l'inégalité de Schwarz devienne une égalité permet de déterminer dans quel cas σ_T atteint sa borne inférieure $\dfrac{1}{\sqrt{n}\sigma_n}$.

L'inégalité précédente ne devient une égalité que si il existe deux nombres α et β (non aléatoires et non tous deux nuls) tels que $\alpha(H' - \theta) + \beta U = 0$, avec H' fonction particulière parmi les H admissibles telle qu'on a l'égalité. Cette égalité se réécrit : $H' = \theta + \lambda' U$ avec λ' un nombre certain. Or si on utilise la relation précédente :

$$E[(T - E[T])(U - E[U])] = 1 \Rightarrow E[(H' - \theta)U] = \lambda' E_\theta\left[U^2\right] = 1 \quad (10)$$

On obtient : $U = \sum_i A(X_i) \Rightarrow \lambda' n E_\theta\left[A^2\right] = 1$ (11)

dont on déduit λ' et la forme de l'estimateur H' associé:

$$\lambda' = \frac{1}{nE\left[A^2\right]} \Rightarrow H' = \theta + \frac{1}{nE\left[A^2\right]} \sum_i \frac{\partial \log p_\theta(X_i)}{\partial \theta} \quad (12)$$

On en déduit donc que l'estimateur qui atteint la borne est de la forme :

$$H' = \theta + \frac{\sum_i \dfrac{\partial \log p_\theta(X_i)}{\partial \theta}}{n \int_{-\infty}^{+\infty} \left[\dfrac{\partial p_\theta(x)}{\partial \theta}\right]^2 \dfrac{dx}{p_\theta(x)}} \quad (13)$$

avec $E[H'] = \theta + \lambda' E[U] = \theta$.

H' **serait donc bien une des fonctions admissibles si** H' **était indépendant de** θ. En effet, si on considère $E_{\theta_0}[H'] = \theta_0$, $E\left[(H' - \theta_0)^2\right] \leq E_{\theta_0}\left[(H - \theta_0)^2\right]$ $\forall H$ tq $E_{\theta_0}[H] = \theta_0$. Or $H = \theta_0$ vérifie l'équation et l'inégalité montre qu'elle est presque certainement égale à θ_0. Donc pour chercher θ_0, il faudrait connaitre au préalable θ_0.

A cette étape, Fréchet cherche les « ***densités distinguées*** », toute densité de probabilité $p_\theta(x)$ telle que la fonction :

$$h(x) = \theta + \frac{\dfrac{\partial \log p_\theta(x)}{\partial \theta}}{\int_{-\infty}^{+\infty} \left[\dfrac{\partial p_\theta(x)}{\partial \theta}\right]^2 \dfrac{dx}{p_\theta(x)}} \quad (14)$$

soit indépendante de θ. L'objectif de Fréchet est alors de déterminer la fonction minimisante $T = H'(X_1,...,X_n)$ qui atteint la borne. L'identité (14) peut se réécrire :

$$\lambda(\theta) \frac{\partial \log p_\theta(x)}{\partial \theta} = h(x) - \theta \quad (15)$$

Or comme $\lambda(\theta) > 0$, ***on peut considérer*** $\dfrac{1}{\lambda(\theta)}$ ***comme la dérivée seconde d'une fonction*** $\Phi(\theta)$ telle que :

$$\frac{\partial \log p_\theta(x)}{\partial \theta} = \frac{\partial^2 \Phi(\theta)}{\partial \theta^2}[h(x) - \theta] \quad (16)$$

dont on déduit que :

$$\ell(x) = \log p_\theta(x) - \frac{\partial \Phi(\theta)}{\partial \theta}[h(x) - \theta] - \Phi(\theta) \quad (17)$$

est une quantité indépendante de θ. ***Une densité distinguée*** sera donc de la forme :

$$p_\theta(x) = e^{\frac{\partial \Phi(\theta)}{\partial \theta}[h(x) - \theta] + \Phi(\theta) + \ell(x)} \quad (18)$$

avec la contrainte $\int_{-\infty}^{+\infty} p_\theta(x)dx = 1$.

Ces 2 conditions sont suffisantes. En effet, soient réciproquement 3 fonctions $\Phi(\theta)$, $h(x)$ et $\ell(x)$ telles qu'on ait quel que soit θ :

$$\int_{-\infty}^{+\infty} e^{\frac{\partial \Phi(\theta)}{\partial \theta}[h(x) - \theta] + \Phi(\theta) + \ell(x)} dx = 1 \quad (19)$$

alors la fonction est distinguée :

$$\theta + \frac{\dfrac{\partial \log p_\theta(x)}{\partial \theta}}{\int_{-\infty}^{+\infty} \left[\dfrac{\partial p_\theta(x)}{\partial \theta}\right]^2 \dfrac{dx}{p_\theta(x)}} = \theta + \lambda(x) \frac{\partial^2 \Phi(\theta)}{\partial \theta^2}[h(x) - \theta] \quad (20)$$

si $\lambda(x) \dfrac{\partial^2 \Phi(\theta)}{\partial \theta^2} = 1$ quand :

$$\frac{1}{\lambda(x)} = \int_{-\infty}^{+\infty} \left[\frac{\partial \log p_\theta(x)}{\partial \theta}\right]^2 p_\theta(x)dx = (\sigma_A)^2 \quad (21)$$

La fonction se réduit à $h(x)$ et donc ne dépend pas de θ. On a alors la relation suivante :

$$\frac{1}{\lambda(x)} = \int_{-\infty}^{+\infty} \left(\frac{\partial^2 \Phi(\theta)}{\partial \theta^2}\right)^2 [h(x) - \theta]^2 e^{\frac{\partial \Phi(\theta)}{\partial \theta}(h(x) - \theta) + \Phi(\theta) + \ell(x)} dx \quad (22)$$

La relation étant vérifiée quel que soit θ, on peut dériver l'expression précédente par rapport à θ :

$$\int_{-\infty}^{+\infty} e^{\frac{\partial \Phi(\theta)}{\partial \theta}(h(x) - \theta) + \Phi(\theta) + \ell(x)} \left(\frac{\partial^2 \Phi(\theta)}{\partial \theta^2}\right)[h(x) - \theta]dx = 0 \quad (23)$$

On peut diviser par $\dfrac{\partial^2 \Phi(\theta)}{\partial \theta^2}$ qui ne dépend pas de x.

Si on dérive à nouveau par rapport à θ, on aura :

$$\int_{-\infty}^{+\infty} e^{\frac{\partial \Phi(\theta)}{\partial \theta}(h(x) - \theta) + \Phi(\theta) + \ell(x)} \left(\frac{\partial^2 \Phi(\theta)}{\partial \theta^2}\right)[h(x) - \theta]^2 dx = \int_{-\infty}^{+\infty} e^{\frac{\partial \Phi(\theta)}{\partial \theta}(h(x) - \theta) + \Phi(\theta) + \ell(x)} dx = 1$$

En combinant cette expression avec celle de $\dfrac{1}{\lambda(x)}$, on obtient : $\lambda(x).\dfrac{\partial^2 \Phi(\theta)}{\partial \theta^2} = 1$ et comme $\lambda(x) > 0$ alors $\dfrac{\partial^2 \Phi(\theta)}{\partial \theta^2} > 0$. Fréchet souligne à cette étape une autre façon d'aborder le problème. On peut prendre arbitrairement $h(x)$ et $l(x)$ et alors $\Phi(\theta)$ est déterminée par :

$$\int_{-\infty}^{+\infty} e^{\frac{\partial \Phi(\theta)}{\partial \theta}[h(x) - \theta] + \Phi(\theta) + \ell(x)} dx = 1 \quad (24)$$

que l'on peut réécrire : $e^{\theta . \frac{\partial \Phi(\theta)}{\partial \theta} - \Phi(\theta)} = \int_{-\infty}^{+\infty} e^{\frac{\partial \Phi(\theta)}{\partial \theta} h(x) + \ell(x)} dx$ (25)

Si on fixe alors arbitrairement $h(x)$ et $l(x)$ et soit s une variable arbitraire, la fonction suivante sera une fonction positive connue représentée par $e^{\Psi(s)}$:

$$\int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}dx = e^{\Psi(s)} \tag{26}$$

On obtient alors la fonction $\Phi(\theta)$ par l'équation :

$$\Phi(\theta) = \theta.\frac{\partial\Phi(\theta)}{\partial\theta} - \Psi\left(\frac{\partial\Phi(\theta)}{\partial\theta}\right) \tag{27}$$

Fréchet remarque qu'il s'agit de l'équation d'Alexis Clairaut. Le cas $\frac{\partial\Phi(\theta)}{\partial\theta} = cste$ réduirait la densité à une fonction indépendante de θ, donc $\Phi(\theta)$ est donné par la solution singulière de cette équation de Clairaut, qui est unique et s'obtient en éliminant *s* entre :

$$\Phi = \theta.s - \Psi(s) \text{ et } \theta = \frac{\partial\Psi(s)}{\partial s} \tag{28}$$

ou encore entre

$$e^{\theta.s-\Phi(\theta)} = \int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}dx \text{ et } \int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}[h(x)-\theta]dx = 0 \tag{29}$$

$\Phi(\theta) = -\log\int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}dx + \theta.s$ où s est donné de façon implicite par $\int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}[h(x)-\theta]dx = 0$ ·

Quelle est alors quand on connait la densité distinguée, celle H' des fonctions $H(X_1,...,X_n)$ vérifiant $E_\theta[H] = \theta$ et telle que σ_H atteigne pour chaque valeur de θ, un minimum absolu, égale à $\frac{1}{\sqrt{n}\sigma_A}$?

A partir de (14), on peut écrire l'estimateur sous la forme :

$$H'(X_1,...,X_n) = \frac{1}{n}[h(X_1)+...+h(X_n)] \tag{30}$$

On peut calculer alors la valeur empirique associée :

$$t = H'(x_1,...,x_n) = \frac{1}{n}\sum_i h(x_i) = \theta + \lambda(\theta)\sum_i \frac{\partial \log p_\theta(x_i)}{\partial\theta}$$

et en prenant $\theta = t$, on a puisque $\lambda(\theta) > 0$:

$$\sum_i \frac{\partial \log p_t(x_i)}{\partial t} = 0 \tag{31}$$

Quand $p_\theta(x)$ est une densité distinguée, la valeur empirique t de θ correspondant à un échantillon $x_1,...,x_n$ est une racine de l'équation précédente en t. Cette équation à une racine et une seule quand *X* est une variable distinguée. En effet, d'après :

$$p_\theta(x) = e^{\frac{\partial\Phi(\theta)}{\partial\theta}[h(x)-\theta]+\Phi(\theta)+\ell(x)} \tag{32}$$

$$\sum_i \frac{\partial \log p_t(x_i)}{\partial t} = \frac{\partial^2\Phi(t)}{\partial t^2}\left[\frac{\sum_i h(x_i)}{n} - t\right] \text{ avec } \frac{\partial^2\Phi(t)}{\partial t^2} > 0 \tag{33}$$

On retrouve la racine unique $t = \frac{\sum_i h(x_i)}{n}$.

Cette fonction $T \equiv H'(X_1,...,X_n) = \frac{1}{n}\sum_i h(X_i)$ ne peut avoir une forme arbitraire, c'est une somme de fonctions de chacune des quantités et c'est même la moyenne arithmétique des *n* valeurs observées d'une même variable aléatoire auxiliaire $Y = h(X)$.

La dispersion est donnée par :

$$(\sigma_{T_n})^2 = \frac{1}{n(\sigma_A)^2} = \frac{1}{n\int_{-\infty}^{+\infty}\left[\frac{\partial p_\theta(x)}{\partial\theta}\right]^2 \frac{dx}{p_\theta(x)}} = \frac{1}{n\frac{\partial^2\Phi(\theta)}{\partial\theta^2}} \tag{34}$$

T_n a une loi de probabilité :

$$p_\theta(t) = \sqrt{n}\frac{1}{\sigma_A\sqrt{2\pi}}e^{-\frac{n(t-\theta)^2}{2.\sigma_A^2}} \text{ avec } (\sigma_A)^2 = \frac{\partial^2\Phi(\theta)}{\partial\theta^2} \tag{35}$$

3 Equation de Clairaut et transformée de Legendre

Nous venons donc de voir que Fréchet montre que les densités distinguées dépendent d'une fonction $\Phi(\theta)$, solution de l'équation de Clairaut:

$$\Phi(\theta) = \theta.\frac{\partial\Phi(\theta)}{\partial\theta} - \Psi\left(\frac{\partial\Phi(\theta)}{\partial\theta}\right) \tag{36}$$

ou donnée par la transformée de Legendre :

$$\Phi = \theta.s - \Psi(s) \text{ et } \theta = \frac{\partial\Psi(s)}{\partial s} \tag{37}$$

Il observe également que la fonction $\Phi(\theta)$ peut s'écrire $\Phi(\theta) = -\log\int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}dx + \theta.s$ où s est donné de façon implicite par $\int_{-\infty}^{+\infty} e^{s.h(x)+\ell(x)}[h(x)-\theta]dx = 0$ ·

Cette transformée « de Legendre » avait été introduite par Adrien-Marie Legendre en 1787 pour résoudre un problème de surface minimale posé par Gaspard Monge en 1784. Utilisant un résultat de Jean-Baptiste Marie Meusnier de La Place, un étudiant de Monge, il résout le problème par un changement de variable qui correspond à la transformée qui porte maintenant son nom. Legendre écrit : « *J'y suis parvenu simplement par un changement de variables qui peut être utile dans d'autres occasions* ». A propos de cette transformation, Darboux donne dans son livre une interprétation de Chasles : « *Ce qui revient suivant une remarque de M. Chasles, à substituer à la surface sa polaire réciproque par rapport à un paraboloïde* ». L'équation de Clairaut a été introduite 40 ans plus tôt en 1734 par Alexis de Clairaut [5]. Les solutions « enveloppe de l'équation de Clairaut » sont équivalentes à la transformée de Legendre sans condition de convexité, mais uniquement sous contrainte de dérivabilité. En effet, pour une fonction non convexe, la transformation de Legendre n'est plus définie là où le hessien de la fonction s'annule, alors que l'équation de Clairaut ne fait que l'hypothèse de dérivabilité. Le passage de la fonction strictement convexe *g* dans l'équation de Clairaut *y = px - g(p)* à la fonction *f* donnant l'enveloppe des solutions par la formule *y = f(x)* est justement la transformation de Legendre [6,7]. Nous allons voir que l'approche de Fréchet peut être reconsidérée dans un contexte plus général sur la base des travaux de Jean-Louis Koszul.

4 Densités distinguées de Koszul

Jean-Louis Koszul [1,2,3] obtient des équations similaires à Fréchet pour les densités distinguées mais dans un cadre plus général, sur les cônes convexes saillants Ω (cône ne contenant pas de droites) dans un espace vectoriel E, via la fonction caractéristique ψ.

Définition de la fonction caractéristique de Koszul: *Soit* $d\xi$ *la mesure de Lebesque sur* E^*, *l'intégrale suivante:* $\psi_\Omega(x)=\int_{\Omega^*} e^{-\langle \xi,x\rangle}d\xi \quad \forall x\in\Omega$ (38)

avec Ω^* *le cône dual, est une fonction analytique sur* Ω, *avec* $\psi_\Omega(x)\in\,]0,+\infty[$, *appelée* ***fonction caractéristique*** *du cône* Ω.

Cette fonction possède les propriétés suivantes:

- Si $g\in Aut(\Omega)$ alors $\psi_\Omega(gx)=|\det g|^{-1}\psi_\Omega(x)$ (39)
- ψ_Ω est logarithmiquement strictement convexe ($\phi_\Omega(x)=\log(\psi_\Omega(x))$ est strictement convexe)

Un difféomorphisme définit les coordonnées duales:

$$x^*=-\alpha_x=-d\log\psi_\Omega(x) \quad (40)$$

$$x^*=\int_{\Omega^*}\xi.e^{-\langle \xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi$$

$$\langle -x^*,h\rangle=d_h\log\psi_\Omega(x)=-\int_{\Omega^*}\langle \xi,h\rangle e^{-\langle \xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi \quad (41)$$

Si on calcule la transformée de Legendre de $\Phi(x)$:

$\Phi^*(x^*)=\langle x,x^*\rangle-\Phi(x)$ avec $x^*=D_x\Phi$ et $x=D_{x^*}\Phi^*$ (42)

qui s'écrit sous la forme d'une équation de Clairaut :

$$\Phi^*(x^*)=\left\langle (D_x\Phi)^{-1}(x^*),x^*\right\rangle-\Phi\left[(D_x\Phi)^{-1}(x^*)\right] \quad (43)$$

En utilisant, $-\langle x^*,x\rangle=\int_{\Omega^*}\log e^{-\langle \xi,x\rangle}.e^{-\langle \xi,x\rangle}d\xi \Big/ \int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi$ et $-\langle \xi,x\rangle=\log e^{-\langle \xi,x\rangle}$, on obtient l'entropie de Shannon:

$$\Phi^*(x^*)=\left[\left(\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi\right).\log\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi-\int_{\Omega^*}\log e^{-\langle \xi,x\rangle}.e^{-\langle \xi,x\rangle}d\xi\right]\Big/\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi$$

$$\Phi^*(x^*)=\left[-\int_{\Omega^*}\frac{e^{-\langle \xi,x\rangle}}{\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi}.\log\left(\frac{e^{-\langle \xi,x\rangle}}{\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi}\right)d\xi\right] \quad (44)$$

Si on note $p_x(\xi)=e^{-\langle \xi,x\rangle}\Big/\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi$ comme une densité de probabilité, $\Phi^*()$ a la forme d'une Entropie de Shannon : $\Phi^*=-\int_{\Omega^*}p_x(\xi)\log p_x(\xi)d\xi$ (45)

On remarque que $x^*=\int_{\Omega^*}\xi.p_x(\xi)d\xi$ (46)

et $\log p_x(\xi)=-\langle x,\xi\rangle-\log\int_{\Omega^*}e^{-\langle \xi,x\rangle}d\xi=-\langle x,\xi\rangle+\Phi(x)$ (47)

Ce qui nous permet d'écrire la relation:

$$\int_{\Omega^*}\Phi^*(\xi)p_x(\xi)d\xi=\Phi^*\left(\int_{\Omega^*}\xi.p_x(\xi)d\xi\right) \quad (48)$$

Cette dernière relation peut s'écrire:

$$E\left[\Phi^*(\xi)\right]=\Phi^*(E[\xi])\ ,\ \xi\in\Omega^* \quad (49)$$

qui montre que le barycentre de l'Entropie est égale à l'Entropie du barycentre, obtenue pour $x^*=D_x\Phi$. Cette densité est la densité à Maximum d'Entropie:

$$\underset{p_x(.)}{Max}\left[-\int_{\Omega^*}p_x(\xi)\log p_x(\xi)d\xi\right] \text{ tel que } \int_{\Omega^*}\xi.p_x(\xi)d\xi=x^* \quad (50)$$

En posant $\bar{\xi}=\Theta(x)=\frac{d\Phi(x)}{dx}$ et son inverse $x=\Theta^{-1}(\bar{\xi})$, la densité est donnée par: $p_{\bar{\xi}}(\xi)=\dfrac{e^{-\langle \xi,\Theta^{-1}(\bar{\xi})\rangle}}{\int_{\Omega^*}e^{-\langle \xi,\Theta^{-1}(\bar{\xi})\rangle}d\xi}$ (51)

avec $\bar{\xi}=\int_{\Omega^*}\xi.p_{\bar{\xi}}(\xi)d\xi$ et $\Phi(x)=-\log\int_{\Omega^*}e^{-\langle x,\xi\rangle}d\xi$ (52)

Si nous nous ramenons à la définition classique de l'information de Fisher $I(x)$, le modèle de Koszul la fait apparaître comme le hessien d'une fonction réelle :

$$\log p_x(\xi)=-\langle x,\xi\rangle+\Phi(x)\Rightarrow\frac{\partial^2\log p_x(\xi)}{\partial x^2}=\frac{\partial^2\Phi(x)}{\partial x^2} \quad (53)$$

$$I(x)=-E_\xi\left[\frac{\partial^2\log p_x(\xi)}{\partial x^2}\right]=-\frac{\partial^2\Phi(x)}{\partial x^2}=\frac{\partial^2\log\psi_\Omega(x)}{\partial x^2}$$

$$I(x)=E_\xi\left[\xi^2\right]-E_\xi[\xi]^2 \quad (54)$$

5 Conclusion

Nous avons redécouvert le sens initial des travaux de Fréchet, en montrant qu'outre la découverte de la borne de Cramer-Rao, six ans avant Rao, Fréchet avait trouvé des relations qui sont les équations fondamentales de la « Géométrie de l'Information », liées à la transformation de Legendre-Clairaut. Il avait fait le lien aussi entre la matrice de Fisher et le hessien d'une fonction convexe, qui sera identifié par Koszul comme le logarithme de la fonction caractéristique [3].

6 Références

1. Koszul J.L., « Ouverts convexes homogènes des espaces affines », *Math. Z. 79, pp. 254-259.*, 1962
2. Koszul J.L., « Variétés localement plates et convexité », *Osaka J. Math. , n°2,p.285-290*, 1965
3. Barbaresco, F., « Koszul Information Geometry ». *MDPI Entropy, n°16, pp. 4521-4565*, 2014
4. Fréchet M., « Sur l'extension de certaines évaluations statistiques au cas de petits échantillons ». *Revue de l'Institut International de Statistique, vol. 11, n° 3/4, pp. 182–205*, 1943
5. Clairaut A., « Solution de plusieurs Problèmes où il s'agit de trouver des Courbes dont la propriété consiste dans une certaine relation entre leurs branches, exprimée par une Équation donnée », *Hist. Acad. royale des sciences, pp.196–215*, 1734
6. Arnold V.I., « Geometrical Methods in the Theory of Ordinary Differential Equations», *Springer,1988*
7. Duplij S., « Generalized Duality, Hamiltonian Formalism and New Brackets », *J. of Math. Phys., Anal., Geom., vol. 10, No. 2, pp. 189-220*, 2014
8. "Differential Geometrical Theory of Statistics", F. Barbaresco & F. Nielsen Ed., MDPI Book, 2017