La méthode du maximum de vraisemblance dans le cas des lois de Meijer

Jean-Marie Nicolas

2018D005

décembre 2018

Département Image, Données, Signal
Groupe IMAGES : *Image, Modélisation, Analyse, GEométrie, Synthèse*

1 Introduction

1.1 La méthode du maximum de vraisemblance

Dans de nombreux problèmes rencontrés en traitement du signal ou des images, on dispose souvent de R échantillons (x_1, ..., x_R) suivant une certaine densité de probabilité $p(x)$ paramétrée par les grandeurs $\alpha_1, ..., \alpha_N$. Ce jeu d'échantillons peut permettre d'estimer ces grandeurs $\alpha_1, ..., \alpha_N$ et il existe différentes méthodes pour en effectuer l'estimation. Une des plus célèbres est la méthode du maximum de vraisemblance. Comme cette méthode suppose que les R échantillons sont indépendants, la probabilité $p(x_1, ..., x_R)$ d'avoir cet ensemble d'échantillons est tout simplement le produit des probabilités d'avoir chaque échantillon x_r :

$$p(x_1, ..., x_R) = \prod_{r=1}^{R} p(x_r)$$

En en prenant le logarithme, on obtient alors une expression additive, la log-vraisemblance :

$$\mathcal{LV} \;=\; \log\left(p(x_1, ..., x_R)\right) = \sum_{r=1}^{R} \log\left(p(x_r)\right). \tag{1}$$

Pour tout paramètre α_i décrivant la densité de probabilité sous jacente, l'estimée $\hat{\alpha}_i$ de ce paramètre au sens du maximum de vraisemblance doit maximiser cette log-vraisemblance, ce qui donne l'expression suivante :

$$\frac{\mathrm{d}\log\left(p(x_1, ..., x_N)\right)}{\mathrm{d}\,\alpha_i} \;=\; \sum_{r=1}^{R} \left.\frac{\mathrm{d}\log p}{\mathrm{d}\,\alpha_i}\right|_{x=x_r} \;=\; 0 \tag{2}$$

La difficulté de la méthode réside dans la dérivation de la densité de probabilité en fonction de ses paramètres : l'expérience montre qu'à coté de cas simples, d'autres nécessitent la connaissance de fonctions spéciales de plus en plus compliquées, et que pour finir il existe un très grand nombre de cas où on ne connaît pas la forme anaytique de cette dérivée.

Ce rapport va se spécialiser sur une famille de lois : celles définies sur $\mathbb{R}^+$, et plus particulièrement celles pour lesquelles un des paramètres peut être vu comme un facteur d'échelle (c'est à dire quasiment la majorité). Dans ce cas, en utilisant un formalisme peu usité, celui des fonctions de Meijer [1], nous verrons que la recherche de ce paramètre par la méthode du maximum de vraisemblance conduit à une étonnante expression implicite dont la forme structurelle est finalement très simple et aisée à mettre en œuvre dans tous les cas possibles.

1.2 Les lois de l'imagerie cohérente

L'imagerie cohérente (radar, sonar, échographie médicale, laser) présente toujours un certain nombre de difficultés liées à la nature multiplicative du bruit sous jacent qu'est le chatoiement (*speckle*). Actuellement, il existe de nombreux travaux permettant d'aborder ce type de données dans le cadre classique des statistiques traditionnelles (moments, cumulants, première et seconde fonctions caractéristiques), par exemple l'ouvrage de Goodman [2] qui offre un panorama assez complet de la situation.

Or le bruit multiplicatif, qui nuit à la lisibilité de l'amplitude des images cohérentes, s'aborde beaucoup plus simplement dans le cadre des log-statistiques. Sous cet angle, il est facile de constater que la quasi totalité des lois statistiques usitées pour analyser l'amplitude des images en imagerie cohérente peuvent se décrire avec des paramètres que l'on peut classer en deux catégories :

1. Les fonctions de Meijer existent dans des logiciels comme Maple et Python.

– une catégorie représentée par un premier paramètre qui sera noté μ et qui joue le rôle de paramètre d'échelle
– une catégorie représentée par un certain nombre N de paramètres de forme, que l'on pourra noter $(\theta_j, j \in [1, N])$.

Remarquons que la paramétrisation usuelle de certaines lois peut ne pas faire apparaître de paramètre d'échelle ou de paramètre de formes stricto sensu (nous en verrons un exemple dans le cas de la loi Gamma).

Cette constation conduit les lois statistiques usitées en imagerie cohérente à avoir une forme bien spécifique et, une fois défini le paramètre d'échelle μ, il est facile de montrer que, dans quasiment tous les cas rencontrés en imagerie cohérente, une loi à $N+1$ paramètres s'écrit :

$$p\left[\mu, \theta_1, \ldots, \theta_N\right](x) \;=\; A \frac{k}{\mu}\, f\left[\theta_1, \ldots, \theta_N\right]\left(\frac{k\,x}{\mu}\right) \tag{3}$$

la fonction f étant une fonction mathématique plus ou moins simple, telle que les N paramètres θ_j modifient la forme de la fonction sans modifier sérieusement l'échelle (par exemple le mode dans le cas d'une loi ayant un mode non nul). Deux nouveaux facteurs A et k apparaissent dans ce formalisme :

– k permet de relier le paramètre μ de la loi à une grandeur statistique spécifique à cette loi, comme le mode ou le moment d'ordre 1. Bien évidemment, k dépend des facteurs de formes $\theta_1, \ldots, \theta_N$
– A est un paramètre permettant la relation fondamentale que suit toute loi de probabilité :

$$\int_0^\infty p\left[\mu, \theta_1, \ldots, \theta_N\right](x)\, \mathrm{d}x \;=\; 1$$

Pour illustrer les problèmes liés à la paramétrisation d'une loi, nous allons prendre l'exemple de la loi Gamma.

1.3 L'exemple de la loi Gamma

La loi Gamma est la loi emblématique de l'imagerie cohérente puisqu'elle modélise les effets du chatoiement sur des données "en intensité". C'est une loi à deux paramètres et nous allons voir que, selon sa paramétrisation, sa mise en œuvre peut s'avérer plus ou moins limpide.

– Sous le formalisme de la relation 3, elle s'écrit (avec $L = \theta_1$) dans une paramétrisation bien connue en imagerie radar :

$$\mathcal{G}\left[\mu, L\right](x) \;=\; \frac{1}{\Gamma(L)} \frac{L}{\mu} \left(\frac{Lx}{\mu}\right)^{L-1} \mathrm{e}^{-\frac{Lx}{\mu}} \quad L > 0 \tag{4}$$

Sa moyenne, c'est à dire son moment d'ordre 1, s'écrit :

$$m_1 \;=\; \mu$$

et son mode, pour $L > 1$ s'écrit :

$$m_{\text{mode}} \;=\; \frac{L-1}{L}\, \mu$$

On observe donc que μ est bien directement lié à la moyenne et au mode, qui sont deux grandeurs couramment utilisées pour décrire l'allure d'une loi statistique..

Connaissant son écart type σ :

$$\sigma \;=\; \frac{\mu}{\sqrt{L}}$$

on en déduit le coefficient de variation γ :

$$\gamma \;=\; \frac{1}{\sqrt{L}}$$

qui ne dépend que du paramètre L : on peut donc affirmer que L est un facteur de forme. On peut aussi noter qu'avec cette paramétrisation la matrice de Fisher de la loi Gamma est une matrice diagonale (voir par exemple [3]).

– En revanche, si on prend la paramétrisation de Wikipedia, la loi Gamma s'écrit :

$$f(x;k,\theta) \;=\; \frac{x^{k-1}\,\mathrm{e}^{-\frac{x}{\theta}}}{\Gamma(k)\,\theta^k}$$

et on trouve pour sa moyenne, c'est à dire son moment d'ordre 1, l'expression :

$$m_1 \;=\; k\theta$$

et pour son mode (pour $k > 1$), l'expression :

$$m_{\text{mode}} \;=\; (k-1)\,\theta$$

On voit alors que moyenne et mode dépendent linéairement à la fois du paramètre k et du paramètre θ.

Connaissant son écart type σ :

$$\sigma \;=\; \sqrt{k}\theta$$

on en déduit le coefficient de variation γ :

$$\gamma \;=\; \frac{1}{\sqrt{k}}$$

Si k semble être un facteur de forme, θ ne peut être un simple facteur d'échelle puisque le premier moment dépend à la fois de θ et de k.

On peut aussi noter qu'avec cette paramétrisation la matrice de Fisher de la loi Gamma n'est pas une matrice diagonale.

1.4 Une propriété essentielle

La modélisation d'une loi de probabilité selon la forme de la relation 3 :

$$p\,[\mu,\theta_1,\ldots,\theta_N]\,(x) \;=\; \frac{Ak}{\mu}\,f\,[\theta_1,\ldots,\theta_N]\left(\frac{k\,x}{\mu}\right)$$

va conduire à une expression très particulière de sa dérivée selon le paramètre μ puisque l'on a :

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mu}\left(p\,[\mu,\theta_1,\ldots,\theta_N]\,(x)\right) &= \frac{\mathrm{d}}{\mathrm{d}\mu}\left(A\frac{k}{\mu}\,f\,[\theta_1,\ldots,\theta_N]\left(\frac{k\,x}{\mu}\right)\right)\\
&= -\frac{Ak}{\mu^2}\,f\,[\theta_1,\ldots,\theta_N]\left(\frac{k\,x}{\mu}\right) - \frac{Ak^2x}{\mu^3}\,\frac{\mathrm{d}}{\mathrm{d}y}\left(f\,[\theta_1,\ldots,\theta_N]\,(y)\right)\Big|_{y=\frac{k\,x}{\mu}}\\
&= -\frac{1}{\mu}\frac{Ak}{\mu}f\,[\theta_1,\ldots,\theta_N]\left(\frac{k\,x}{\mu}\right) - \frac{1}{\mu}\frac{Ak}{\mu}\frac{kx}{\mu}\,\frac{\mathrm{d}}{\mathrm{d}y}\left(f\,[\theta_1,\ldots,\theta_N]\,(y)\right)\Big|_{y=\frac{k\,x}{\mu}} \qquad (5)
\end{aligned}$$

expression que l'on peut aussi réécrire sous la forme suivante :

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left(p\,[\mu,\theta_1,\ldots,\theta_N]\,(x)\right) \;=\; -\frac{1}{\mu}p\,[\mu,\theta_1,\ldots,\theta_N]\,(x) - \frac{1}{\mu}\left(y\frac{\mathrm{d}}{\mathrm{d}y}\left(p\,[\mu,\theta_1,\ldots,\theta_N]\,(y)\right)\right)\Big|_{y=\frac{kx}{\mu}} \qquad (6)$$

1.5 Application à la méthode du maximum de vraisemblance

Comme nous l'avons rappelé au paragraphe 1.1, la méthode du maximum de vraisemblance cherche, à partir de R réalisations x_1, ..., x_R et d'une forme analytique d'une loi donnée $p(x)$, les paramètres tels que l'ensemble des réalisations soient les plus vraisemblables.

Pour une loi vérifiant le formalisme de la relation 3 :

$$p\left[\mu,\theta_1,\ldots,\theta_N\right](x) \;=\; A\frac{k}{\mu}\, f\left[\theta_1,\ldots,\theta_N\right]\left(\frac{k\,x}{\mu}\right)$$

et dans le cas du paramètre μ, cette expression s'écrit :

$$\frac{\mathrm{d}\log\left(p(x_1,...,x_N)\right)}{\mathrm{d}\,\mu} \;=\; \sum_{r=1}^{R}\left.\frac{\mathrm{d}\log p(x)}{\mathrm{d}\,\mu}\right|_{x=x_r} \;=\; 0 \qquad (7)$$

En utilisant la relation 6, on obtient alors :

$$\begin{aligned}
&\sum_{r=1}^{R}\left.\frac{\mathrm{d}\log p(x)}{\mathrm{d}\,\mu}\right|_{x=x_r}\\
&\quad= \sum_{r=1}^{R}\frac{1}{p\left[\hat{\mu},\theta_1,\ldots,\theta_N\right](x_r)}\left(-\frac{1}{\hat{\mu}}\,p\left[\hat{\mu},\theta_1,\ldots,\theta_N\right](x_r) \;-\; \frac{1}{\hat{\mu}}\left.\left(y\frac{\mathrm{d}}{\mathrm{d}y}\left(p\left[\hat{\mu},\theta_1,\ldots,\theta_N\right](y)\right)\right)\right|_{y=\frac{k\,x_r}{\hat{\mu}}}\right)\\
&\quad= 0
\end{aligned} \qquad (8)$$

ce qui donne au final une relation implicite fondamentale :

$$\frac{1}{R}\sum_{r=1}^{R}\left.\frac{y\frac{\mathrm{d}}{\mathrm{d}y}\left(p\left[\hat{\mu},\theta_1,\ldots,\theta_N\right](y)\right)}{p\left[\hat{\mu},\theta_1,\ldots,\theta_N\right](y)}\right|_{y=\frac{k\,x_r}{\hat{\mu}}} \;=\; -1 \qquad (9)$$

Elle nécessite bien entendu de connaître la forme analytique de la dérivée première de la fonction p. Il faut noter que, sous cette forme la plus générale et concise possible, l'expression est implicite puisque $\hat{\mu}$ n'apparaît que dans l'expression de la variable y.

Si on utilise la modélisation donnée par l'expression 3, on peut réécrire cette expression en une nouvelle expression ne faisant intervenir que la fonction $f\left[\theta_1,\ldots,\theta_N\right]$ (qui ne dépend pas directement de μ), ce qui donne :

$$\frac{1}{R}\sum_{r=1}^{R}\left.\frac{y\,\frac{\mathrm{d}}{\mathrm{d}y}\left(f\left[\theta_1,\ldots,\theta_N\right](y)\right)}{f\left[\theta_1,\ldots,\theta_N\right](y)}\right|_{y=\frac{k\,x_r}{\hat{\mu}}} \;=\; -1 \qquad (10)$$

que l'on peut réécrire :

$$\hat{\mu} \;=\; -\frac{1}{R}\sum_{r=1}^{R} kx_r\,\left.\frac{\frac{\mathrm{d}}{\mathrm{d}y}f\left[\theta_1,\ldots,\theta_N\right](y)}{f\left[\theta_1,\ldots,\theta_N\right](y)}\right|_{y=\frac{k\,x_r}{\hat{\mu}}} \qquad (11)$$

C'est aussi une relation implicite, puisque $\hat{\mu}$ apparaît dans le terme de gauche et aussi dans le terme de droite. Elle nécessite bien entendu de connaître la forme analytique de la dérivée première de la fonction f.

1.6 Exemple : la loi Gamma

Considérons la loi Gamma :

$$\mathcal{G}\left[\mu,L\right](x) \;=\; \frac{1}{\Gamma(L)}\frac{L}{\mu}\left(\frac{Lx}{\mu}\right)^{L-1}\mathrm{e}^{-\frac{Lx}{\mu}} \quad L>0$$

Elle peut s'écrire :

$$\mathcal{G}\left[\mu, L\right](x) \;=\; \frac{L}{\Gamma(L)}\,\frac{1}{\mu}\,f[L]\,(y)\Big|_{y=\frac{k\,x}{\mu}}$$

avec :

$$f[L](y) \;=\; y^{L-1}\,\mathrm{e}^{-y}\;,\;\; k \;=\; L \quad \text{et} \quad A \;=\; \frac{1}{\Gamma(L)}$$

La dérivée de $f[L](y)$ par rapport à y s'écrit (pour $L > 1$) [2] :

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}y} f[L](y) \;&=\; -f(y) \;+\; (L-1)\,y^{L-2}\,\mathrm{e}^{-y}\\ &=\; -f(y) \;+\; \frac{L-1}{y}\,y^{L-1}\,\mathrm{e}^{-y}\end{aligned}$$

ce qui donne :

$$\frac{\mathrm{d}}{\mathrm{d}y} f[L](y) \;=\; \left(\frac{L-1}{y} - 1\right)\,f[L](y)$$

d'où :

$$\frac{\frac{\mathrm{d}}{\mathrm{d}y} f[L](y)}{f[L](y)} \;=\; \left(\frac{L-1}{y} - 1\right)$$

En reportant dans l'expression 11, on obtient :

$$\begin{aligned}\hat{\mu} \;&=\; -\frac{1}{R}\sum_{r=1}^{R} Lx_r\,\frac{\frac{\mathrm{d}}{\mathrm{d}y} f\,[L]\,(y)}{f\,[L]\,(y)}\Bigg|_{y=\frac{L\,x_r}{\hat{\mu}}}\\ &=\; -\frac{1}{R}\sum_{r=1}^{R} Lx_r\left(\frac{(L-1)\hat{\mu}}{Lx_r} - 1\right)\\ &=\; -(L-1)\hat{\mu} \;+\; \frac{1}{R}\sum_{r=1}^{R} Lx_r\end{aligned}$$

d'où :

$$\hat{\mu} \;=\; \frac{1}{R}\sum_{r=1}^{R} x_r \tag{12}$$

On retrouve ainsi après un calcul plutôt compliqué un résultat bien connu : l'estimateur du paramètre μ d'une loi Gamma est tout simplement la moyenne (voir [3]).

2 Le cas des "Lois de Meijer Normalisées"

2.1 Choix du formalisme

Dans le rapport [4], nous avons défini les lois de Meijer Normalisées $\overline{LMN}\,[\mu](x)$ caractérisées par une expression analytique sous forme de fonction de Meijer dont les paramètres sont d'une part un certain nombre de facteurs de forme et d'autre part une valeur μ globalement associée à la moyenne de la loi.

$$\overline{LMN}\,[\mu](x) \;=\; A\,\frac{k}{\mu}\overline{G}_{p,q}^{m,n}\left(\frac{k\,x}{\mu}\;\middle|\;\begin{array}{lcl} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p\\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q\end{array}\right)$$

ce qui mène aux remarques suivantes :

2. Le lecteur analysera de lui même les autres cas.

– les coefficients a_i et b_i, au nombre de $p+q$, sont les facteurs de forme de la loi.
– le terme $\frac{x}{\mu}$ garantit des propriétés de similitude.
– le terme k assure que la moyenne soit proche du paramètre μ
– le terme A, qui ne dépend que des facteurs de formes, assure que cette relation décrit bien une loi de probabilité ($\int_0^\infty \overline{LMN}\,[\mu](x)\mathrm{d}x = 1$).

Il se trouve que quasiment toutes les lois usuelles en imagerie cohérente peuvent s'exprimer sous forme d'une loi de Meijer Normalisée [4]. Cette paramétrisation permet d'écrire :

$$\overline{LMN}\,[\mu](x) \;=\; A\,\frac{k}{\mu}\, f[[a_1,\ldots,a_n],[a_{n+1},\ldots,a_p],[b_1,\ldots,b_m],[b_{m+1},\ldots,b_q]](y)|_{y=\frac{k\,x}{\mu}}$$

avec

$$f[[a_1,\ldots,a_n],[a_{n+1},\ldots,a_p],[b_1,\ldots,b_m],[b_{m+1},\ldots,b_q]](y) = \overline{G}^{m,n}_{p,q}\left(y \,\middle|\, \begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q \end{array}\right)$$

On peut noter que la mise en forme des paramètres d'une fonction de Meijer utilisée ici dans l'expression formelle de la fonction f correspond à la forme des procédures de calcul des fonctions de Meijer dans les logiciels actuels (Python, Maple).

2.2 Les Lois de Meijer Normalisées et la méthode du Maximum de Vraisemblance

Le principe du maximum de vraisemblance requiert, dans l'expression 10, le calcul de la grandeur :

$$y\,\frac{\mathrm{d}}{\mathrm{d}y} f\left[\theta_1,\ldots,\theta_N\right](y)$$

c'est à dire :

$$y\,\frac{\mathrm{d}}{\mathrm{d}y} f\left[[a_1,\ldots,a_n],[a_{n+1},\ldots,a_p],[b_1,\ldots,b_m],[b_{m+1},\ldots,b_q]\right](y)$$

Or dans le cas d'une fonction de Meijer, on connaît deux résultats essentiels pour le calcul de cette expression :

– La dérivée d'une fonction de Meijer est une fonction de Meijer (voir [1, 4]) :

$$\frac{\mathrm{d}}{\mathrm{d}z}\overline{G}^{m,n}_{p,q}\left(z \,\middle|\, \begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q \end{array}\right) = -\overline{G}^{m+1,n}_{p+1,q+1}\left(z \,\middle|\, \begin{array}{ccc} a_1-1,\ldots,a_n-1 & ; & 1, a_{n+1}-1,\ldots,a_p-1 \\ 0, b_1-1,\ldots,b_m-1 & ; & b_{m+1}-1,\ldots,b_q-1 \end{array}\right)$$

les paramètres de forme ayant subis une décrémentation unitaire et deux nouveaux paramètres de forme étant apparu sur l'antidiagonale (les valeurs dans un cercle).

– le produit d'une fonction de Meijer de variable z par la variable z est une fonction de Meijer

$$z\,\overline{G}^{m,n}_{p,q}\left(z \,\middle|\, \begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q \end{array}\right) = \overline{G}^{m,n}_{p,q}\left(z \,\middle|\, \begin{array}{ccc} a_1+1,\ldots,a_n+1 & ; & a_{n+1}+1,\ldots,a_p+1 \\ b_1+1,\ldots,b_m+1 & ; & b_{m+1}+1,\ldots,b_q+1 \end{array}\right)$$

Ainsi, le rapport recherché s'écrit :

$$\frac{\overline{G}^{m+1,n}_{p+1,q+1}\left(\frac{k\,x}{\hat{\mu}} \,\middle|\, \begin{array}{ccc} a_1,\ldots,a_n & ; & 0, a_{n+1},\ldots,a_p \\ 1, b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q \end{array}\right)}{\overline{G}^{m,n}_{p,q}\left(\frac{k\,x}{\hat{\mu}} \,\middle|\, \begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q \end{array}\right)}$$

et donc, si on dispose de R réalisations de la loi, on a la relation implicite que doit vérifier $\hat{\mu}$ pour être l'estimateur du facteur d'échelle au sens du maximum de vraisemblance :

$$\boxed{\frac{1}{R}\sum_{r=1}^{R} \frac{\overline{G}_{p+1,q+1}^{m+1,n}\left(\frac{k\,x_r}{\hat{\mu}}\left|\begin{array}{ccc} a_1,\ldots,a_n & ; & 0,a_{n+1},\ldots,a_p \\ 1,b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q\end{array}\right.\right)}{\overline{G}_{p,q}^{m,n}\left(\frac{k\,x_r}{\hat{\mu}}\left|\begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q\end{array}\right.\right)} \;=\; 1} \tag{13}$$

Pour être exhaustif, notons que la dérivée d'une fonction de Meijer a une autre expression (les nouveaux paramètres de valeurs 1 et 0 étant placés sur la diagonale, ce qui donne :

$$\frac{1}{R}\sum_{r=1}^{R} \frac{-\overline{G}_{p+1,q+1}^{m,n+1}\left(\frac{k\,x}{\hat{\mu}}\left|\begin{array}{ccc} 0,a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & 1,b_{m+1},\ldots,b_q\end{array}\right.\right)}{\overline{G}_{p,q}^{m,n}\left(\frac{k\,x}{\hat{\mu}}\left|\begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q\end{array}\right.\right)} \;=\; 1$$

2.3 La loi Gamma, cas de loi de Meijer Normalisée

Puisque la loi Gamma a une expression sous forme de loi de Meijer normalisée :

$$\mathcal{G}\left[\mu,L\right](x) \;=\; \frac{L}{\mu}\frac{1}{\Gamma(L)}\overline{G}_{0,1}^{1,0}\left(\frac{Lx}{\mu}\left|\begin{array}{ccc} . & ; & . \\ L-1 & ; & .\end{array}\right.\right)$$

la relation que doit vérifier $\hat{\mu}$ pour être l'estimateur du facteur d'échelle s'écrit :

$$\frac{1}{R}\sum_{r=1}^{R} \frac{\overline{G}_{1,2}^{2,0}\left(\frac{Lx_r}{\hat{\mu}}\left|\begin{array}{ccc} . & ; & 0 \\ 1,L-1 & ; & .\end{array}\right.\right)}{\overline{G}_{0,1}^{1,0}\left(\frac{Lx_r}{\hat{\mu}}\left|\begin{array}{ccc} . & ; & . \\ L-1 & ; & .\end{array}\right.\right)} \;=\; 1$$

Cette relation, dans laquelle on peut éventuellement remplacer la fonction de Meijer par une fonction de Whittaker [3], est bien évidemment inutile en pratique puisque dans ce cas on connaît par des calculs traditionnels la valeur de $\hat{\mu}$ (relation 12) :

$$\hat{\mu} \;=\; \frac{1}{R}\sum_{r=1}^{R} x_r$$

relation non seulement explicite mais surtout très facile à calculer (une simple moyenne).

3 Le cas des "Lois de Meijer Normalisées Généralisées"

3.1 Choix du formalisme

Dans le rapport [4], nous avons défini les lois de Meijer Normalisées Généralisées $\mathcal{LMG}[\mu](x)$ caractérisées par une expression analytique sous forme de loi de Meijer normalisée, mais dont la variable est élevé à la puissance η :

$$\mathcal{LMG}[\mu](x) \;=\; A\,\frac{k}{\mu}\overline{G}_{p,q}^{m,n}\left(\left(\frac{k\,x}{\mu}\right)^{\eta}\left|\begin{array}{ccc} a_1,\ldots,a_n & ; & a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m & ; & b_{m+1},\ldots,b_q\end{array}\right.\right)$$

3. jonglerie mathématique a priori sans importance mais utile pour des plateformes logicielles ne disposant pas des fonctions de Meijer, comme Matlab.

η est un facteur de forme : il joue un rôle dans les paramètres $a_1, \ldots, a_p$ et $b_1, \ldots, b_q$ de la fonction de Meijer.

On a bien :

$$\mathcal{LMG}[\mu](x) \;=\; A\,\frac{k}{\mu}\,f[[a_1,\ldots,a_n],[a_{n+1},\ldots,a_p],[b_1,\ldots,b_m],[b_{m+1},\ldots,b_q]](y^\eta)|_{y=\frac{k\,x}{\mu}}$$

3.2 Maximum de vraisemblance

Comme dans le cas précédent, nous nous limitons à l'estimation du paramètre d'échelle μ au sens du maximum de vraisemblance, ce qui requiert, dans l'expression 10, le calcul de la grandeur :

$$y\,\frac{\mathrm{d}}{\mathrm{d}y} f\left[\theta_1,\ldots,\theta_N\right](y)$$

c'est à dire :

$$y\,\frac{\mathrm{d}}{\mathrm{d}y} f\left[[a_1,\ldots,a_n],[a_{n+1},\ldots,a_p],[b_1,\ldots,b_m],[b_{m+1},\ldots,b_q]\right](y)$$

Comme précédemment, on sait que la dérivée d'une fonction de Meijer est une fonction de Meijer

$$\begin{aligned}
&\tfrac{\mathrm{d}}{\mathrm{d}z}\overline{G}^{m,n}_{p,q}\left(z\,\middle|\,\begin{array}{l} a_1,\ldots,a_n\;;\;a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right) \\
&\qquad = -\overline{G}^{m+1,n}_{p+1,q+1}\left(z\,\middle|\,\begin{array}{c} a_1-1,\ldots,a_n-1\;;\;-1,a_{n+1}-1,\ldots,a_p-1 \\ 0,b_1-1,\ldots,b_m-1\;;\;b_{m+1}-1,\ldots,b_q-1\end{array}\right)
\end{aligned}$$

Comme nous avons ici

$$z \;=\; y^\eta$$

le théorème des fonctions composées permet d'écrire :

$$\begin{aligned}
&\tfrac{\mathrm{d}}{\mathrm{d}y}\overline{G}^{m,n}_{p,q}\left(y^\eta\,\middle|\,\begin{array}{l} a_1,\ldots,a_n\;;\;a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right) \\
&\qquad = -\eta\,y^{\eta-1}\overline{G}^{m+1,n}_{p+1,q+1}\left(y^\eta\,\middle|\,\begin{array}{c} a_1-1,\ldots,a_n-1\;;\;-1,a_{n+1}-1,\ldots,a_p-1 \\ 0,b_1-1,\ldots,b_m-1\;;\;b_{m+1}-1,\ldots,b_q-1\end{array}\right)
\end{aligned}$$

et, connaissant la propriété liée à la multiplication d'une fonction de Meijer par la variable, on obtient :

$$\frac{\mathrm{d}}{\mathrm{d}y}\overline{G}^{m,n}_{p,q}\left(y^\eta\,\middle|\,\begin{array}{l} a_1,\ldots,a_n\;;\;a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right) = -\eta\frac{1}{y}\overline{G}^{m+1,n}_{p+1,q+1}\left(y^\eta\,\middle|\,\begin{array}{c} a_1,\ldots,a_n\;;\;0,a_{n+1},\ldots,a_p \\ 1,b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right)$$

d'où le terme recherché :

$$-\eta\;\overline{G}^{m+1,n}_{p+1,q+1}\left(y^\eta\,\middle|\,\begin{array}{c} a_1,\ldots,a_n\;;\;0,a_{n+1},\ldots,a_p \\ 1,b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right)$$

Par un raisonnement analogue au cas précédent, et dans le cas où on dispose de R réalisations x_1, ..., x_R de la loi, on obtient la relation implicite que doit vérifier $\hat{\mu}$:

$$\boxed{\frac{1}{R}\sum_{r=1}^{R}\frac{\overline{G}^{m+1,n}_{p+1,q+1}\left(\left(\frac{k\,x_r}{\hat{\mu}}\right)^\eta\,\middle|\,\begin{array}{c} a_1,\ldots,a_n\;;\;0,a_{n+1},\ldots,a_p \\ 1,b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right)}{\overline{G}^{m,n}_{p,q}\left(\left(\frac{k\,x_r}{\hat{\mu}}\right)^\eta\,\middle|\,\begin{array}{l} a_1,\ldots,a_n\;;\;a_{n+1},\ldots,a_p \\ b_1,\ldots,b_m\;;\;b_{m+1},\ldots,b_q\end{array}\right)} \;=\; \frac{1}{\eta}} \qquad (14)$$

Il est vraiment étonnant de retrouver –au terme en $1/\eta$ près– l'expression obtenue dans le cadre des lois standards (relation 13), le cas $\eta = 1$ redonnant bien évidemment ce cas précédent.

3.3 Cas de la loi de Rayleigh-Nakagami

Pour illustrer un exemple de loi généralisée, prenons la loi Gamma Généralisée $\mathcal{GG}\left[\mu, L, \eta\right]$:

$$\mathcal{GG}\left[\mu, L, \eta\right](x) \;=\; \frac{|\eta|}{\mu}\frac{L^{\frac{1}{\eta}}}{\Gamma(L)}\left(\frac{L^{\frac{1}{\eta}}x}{\mu}\right)^{\eta L-1} \mathrm{e}^{-\left(\frac{L^{\frac{1}{\eta}}x}{\mu}\right)^{\eta}}$$

et plus précisément le cas de la loi de Rayleigh-Nakagami $\mathcal{RN}\left[\mu, L\right]$, qui est une loi Gamma Généralisée avec $\eta = 2$. Plus précisément, on a :

$$\mathcal{RN}\left[\mu, L\right](x) \;=\; \frac{2}{\mu}\frac{\sqrt{L}}{\Gamma(L)}\;\overline{G}_{0,1}^{1,0}\left(\left(\frac{\sqrt{L}x}{\mu}\right)^2 \middle|\; \begin{matrix} . & ; & . \\ L-\frac{1}{2} & ; & . \end{matrix}\right)$$

Dans ce cadre, la relation 14 s'écrit :

$$\frac{1}{R}\sum_{r=1}^{R}\frac{\overline{G}_{1,2}^{2,0}\left(\left(\frac{\sqrt{L}x_r}{\hat{\mu}}\right)^2 \middle|\; \begin{matrix} . & ; & 0 \\ 1, L-\frac{1}{2} & ; & . \end{matrix}\right)}{\overline{G}_{0,1}^{1,0}\left(\left(\frac{\sqrt{L}x_r}{\hat{\mu}}\right)^2 \middle|\; \begin{matrix} . & ; & . \\ L-\frac{1}{2} & ; & . \end{matrix}\right)} \;=\; \frac{1}{2} \tag{15}$$

Notons que dans le cas de la loi de Rayleigh Nakagami, on a une relation bien connue donnant explicitement la valeur $\hat{\mu}$ pour la méthode du maximum de vraisemblance :

$$\hat{\mu} \;=\; \sqrt{\sum_{r=1}^{R} x_r^2} \tag{16}$$

puisque pour une loi Gamma Généralisée on a :

$$\hat{\mu} \;=\; \left(\sum_{r=1}^{R} x_r^{\eta}\right)^{\frac{1}{\eta}}$$

La nouvelle approche semble proposer dans ce cas une méthode certes exacte mais inadaptée puisqu'il existe une autre méthode beaucoup plus simple et facile à mettre en œuvre.

4 Résultats de simulation

4.1 Mise en œuvre des simulations

Nous allons maintenant vérifier la validité de cette nouvelle approche sur des simulations de diverses lois de l'imagerie cohérente. Ces lois s'exprimant sous forme de fonctions de Meijer, leurs fonctions de répartition s'en déduisent aisément (puisque la primitive d'une fonction de Meijer est une fonction de Meijer), ce qui permet d'effectuer des tirages aléatoires par la méthode de l'inversion de la fonction de répartition.

Ensuite, pour résoudre le système implicite de l'équation 14, l'expérience montre que numérateur et dénominateur sont des fonctions relativement douces autour de la valeur $x = \mu$: une méthode de type essai-erreur pemet alors d'avoir très facilement d'avoir des estimés corrects pour la valeur $\hat{\mu}$.

Enfin, on va comparer les résultats obtenus avec ceux obtenus par d'autres méthodes : la méthode des moments et celles des log-cumulants, à ceci près que l'on se place dans un cadre plutôt restrictif où l'on connaît les paramètres de forme et que la recherche ne porte que sur le paramètres d'échelle μ. En effectuant un certain nombre de tirage de lois, on pourra évaluer la moyenne de l'estimateur obtenu et son écart-type. Il sera alors possible de vérifier que la méthode donnant la variance minimale est celle du maximum de vraisemblance (puisque dans le cas des lois de type exponentielle on sait que la variance atteint la borne de Cramèr Rao), et de comparer avec les autre méthodes.

Notons cependant que dans des cas simples où l'estimateur est une simple moyenne, méthode des moments et méthode du maximum de vraisemblance donnent le même résultat. Dans le cas contraire, la méthode des moments donne très souvent des résultats entachés d'une forte variance. Aussi il est intéressant de noter que dans le même temps, la méthode des log-cumulants donne toujours une estimation non biaisée (puisque l'on connait les paramètres de forme) et dont la variance est assez proche de la borne de Cramèr Rao (supposée atteinte par la méthode du maximum de vraisemblance pour les lois de l'imagerie cohérente).

4.2 Loi Gamma

La simulation traitée (2000 tirages) est celle de la loi Gamma $\mathcal{G}\left[\mu = 11.7, L = 1.7\right](x)$. Connaissant L, on obtient les estimés de μ suivant :

Méthode	$\hat{\mu}$	écart type
MMV	11.695	0.189
log-stat	11.676	0.227
MM	11.695	0.189

Valeurs moyennes et écart types ont été estimés sur 100 cas.

Comme on pouvait s'y attendre, la nouvelle méthode (notée MMV) et la méthode des moments donnent exactement les mêmes résultats, ce qui valide le code de recherche du paramètre $\hat{\mu}$ dans la relation implicite 14.

On note qu'effectivement la méthode des log-cumulants semble non biaisée et présente une variance d'estimateur supérieure à celle obtenue par la méthodes du maximum de vraisemblance.

4.3 Loi à 3 paramètres : loi de Fisher

La simulation traitée (2000 tirages) est celle de la loi de Fisher $\mathcal{F}\left[\mu = 11.7, L = 1.7, M = 2.33\right](x)$. Connaissant L et M, on obtient les estimés de μ suivant :

Méthode	$\hat{\mu}$	écart type
MMV	11.717	0.264
log-stat	11.724	0.273
MM	11.635	0.457

Valeurs moyennes et écart types ont été estimés sur 100 cas.

Comme dans le cas de la loi de Fisher $\mathcal{F}\left[\mu, L, M\right](x)$, le moment d'ordre 1 s'écrit (pour $M > 1$) :

$$m_1 = \frac{M}{M-1}\mu$$

la méthode du maximum de vraisemblance ne peut se réduire à une simple moyenne : on en voit la conséquence quand on compare méthode des moments avec méthode du maximum de vraisemblance et on note que la variance de la méthode des moments est presque le double de la variance de la méthode du maximum de vraisemblance. Dans le même temps, la méthode des log-cumulants donne une variance d'estimateur certes plus élevée mais finalement assez proche de celle de la méthode du maximum de vraisemblance.

4.4 Loi généralisée : loi de Nakagami

La simulation traitée (2000 tirages) est celle de la loi de Nakagami $\mathcal{RN}\left[\mu = 11.7, L = 1.7\right](x)$, qui est un exemple de loi Gamma Généralisée (avec η=2). Connaissant L, on obtient les estimés de μ suivant :

Méthode	$\hat{\mu}$	écart type
MMV	11.689	0.106
log-stat	11.695	0.120
MM	11.693	0.109

Valeurs moyennes et écart types ont été estimés sur 100 cas.

La méthode des moments donne un résultat très légèrement différent de celle du maximum de vraisemblance puisque l'expression 16 est une moyenne quadratique et non une simple moyenne.

5 Conclusion

La méthode du maximum de vraisemblance appliquée aux lois de Meijer permet d'obtenir une relation implicite étonnante pour estimer le paramètre d'échelle μ. En effet, d'une part le formalisme obtenu ne pose aucune difficulté pour être appliquée à n'importe quelle loi de Meijer : il suffit de rajouter deux paramètres (valeurs 1 et 0) sur l'antidiagonale dans la représentation très caractéristique des fonctions de Meijer utilisées pour construire la relation implicite donnant l'estimée de μ. D'autre part, si ce formalisme permet d'obtenir dans tous les cas une expression finalement assez simple vérifiée par l'estimée de μ, il est totalement inadapté pour traiter les lois dont un simple calcul classique permet d'obtenir une relation explicite simple (cas de la loi Gamma ou de la loi de Nakagami par exemple). Notons qu'il est parfaitement adapté aux lois généralisées.

Néanmoins, même si ce résultat peut se classer parmi les curiosités mathématiques des lois de l'imagerie cohérente, il n'en demeure pas moins qu'il est très utile pour montrer que, concernant le paramètre de forme, les estimateurs des log-statistiques ont des variances très proches des minima atteints par l'approche du maximum de vraisemblance. Ce résultat était resté jusqu'à présent à l'état de conjecture : désormais, il est possible dans tous les cas de figure de vérifier cette concordance. Etant donné la simplicité de mise en œuvre des techniques issues des log-statistiques, il était important d'en connaître effectivement les qualités intrinsèques, c'est à dire leur faible variance, et d'en garantir ainsi leur utilisation dans tous les cas concrets possibles.

Références

[1] H. Bateman *Higher transcendental functions* McGraw-Hill, 1953

[2] J.W. Goodman *Speckle phenomena in optics : Theory and applications*, Roberts & Company, 2007.

[3] J.M. Nicolas *Application de la transformée de Mellin : étude des lois statistiques de l'imagerie cohérente* Rapport Télécom ParisTech 2006D010, version revue et corrigée 2018D004

[4] J.M. Nicolas *Les distributions de Meijer et leurs propriétés en statistiques de Mellin* Rapport Télécom ParisTech 2011D002

Table des matières

Dépôt légal : 2018 – 4e trimestre
Imprimé à Télécom ParisTech – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Télécom ParisTech

Institut Mines-Télécom - membre de l'Université Paris Saclay

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.telecom-paristech.fr

Département IDS