



**HAL**  
open science

# Propriété de normalité asymptotique locale uniforme pour les modèles de loi gaussienne inverse généralisée

Angelo Efoévi Koudou, Christophe Ley

► **To cite this version:**

Angelo Efoévi Koudou, Christophe Ley. Propriété de normalité asymptotique locale uniforme pour les modèles de loi gaussienne inverse généralisée. 45èmes Journées de Statistique, May 2013, Toulouse, France. hal-01288619

**HAL Id: hal-01288619**

**<https://hal.science/hal-01288619v1>**

Submitted on 15 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PROPRIÉTÉ DE NORMALITÉ ASYMPTOTIQUE LOCALE UNIFORME POUR LES MODÈLES DE LOI GAUSSIENNE INVERSE GÉNÉRALISÉE

Angelo Efoévi Koudou <sup>1</sup> & Christophe Ley <sup>2</sup>

<sup>1</sup> *Université de Lorraine, Institut Elie Cartan de Lorraine, UMR 7502,  
Vandoeuvre-lès-Nancy, F-54506, France. Efoevi.Koudou@univ-lorraine.fr*

<sup>2</sup> *ECaRES and Département de Mathématique, Université Libre de Bruxelles, Boulevard  
du Triomphe, CP210, 1050 Brussels, Belgium. chrisley@ulb.ac.be*

**Résumé.** Nous démontrons la propriété de normalité asymptotique locale uniforme (propriété ULAN) pour le modèle paramétrique constitué par les lois d'échantillons iid de variables gaussiennes inverses généralisées. Cela ouvre la voie, dans des travaux futurs, à l'utilisation de la méthodologie de Le Cam en vue de construire des test optimaux pour les paramètres.

**Mots-clés.** Loi gaussienne inverse généralisée, propriété ULAN, théorie de Le Cam.

**Abstract.** We establish the uniform local asymptotic normality (ULAN) property for generalized inverse Gaussian models. This paves the way, in future work, for constructing optimal test procedures for the parameters, using the Le cam methodology.

**Keywords.** Generalized inverse Gaussian distribution, ULAN property, Le Cam theory.

## 1 Introduction.

La loi *gaussienne inverse généralisée* (GIG) de paramètres  $p \in \mathbb{R}$ ,  $a > 0$ ,  $b > 0$  a pour densité

$$f_{p,a,b}(x) := c(p, a, b)x^{p-1}e^{-(ax+b/x)/2}, \quad x > 0, \quad (1.1)$$

où la constante de normalisation est  $c(p, a, b) := \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})}$ ,  $K_p$  étant la fonction modifiée de Bessel de troisième espèce.

au sein de la famille des lois GIG, aussi appelée "famille de Halphen", on note deux modèles importants : la loi *gaussienne inverse* (IG) pour  $p = -\frac{1}{2}$  et la loi *gaussienne inverse réciproque* (RIG) pour  $p = \frac{1}{2}$ .

La loi gaussienne inverse est utilisée pour modéliser des données dans de nombreux domaines, notamment en cardiologie, en démographie, en finance, en biologie, en hydrologie, etc. On peut trouver des exemples d'applications dans Chhikara et Folks (1989), Seshadri (1999).

Un problème naturel est celui de tester les modèles IG ou RIG parmi les lois GIG, c'est-à-dire de considérer les deux hypothèses nulles  $\mathcal{H}_0^\pm : p = \pm\frac{1}{2}$  contre des alternatives de la forme  $\mathcal{H}_1^\pm : p \neq \pm\frac{1}{2}$ . La manière standard d'aborder cette question est d'utiliser un test de rapport de vraisemblance. Le numérateur du rapport de vraisemblance prend en compte uniquement les estimateurs du maximum de vraisemblance des paramètres  $a$  et  $b$  sous  $p = \pm\frac{1}{2}$  respectivement, dont les expressions sont les suivantes : pour  $p = -\frac{1}{2}$ ,

$$\hat{a} = \frac{1}{\bar{X}(\bar{X}\bar{X}_{-1} - 1)}, \quad \hat{b} = \frac{\bar{X}}{\bar{X}\bar{X}_{-1} - 1},$$

tandis que, pour  $p = \frac{1}{2}$ ,

$$\hat{a} = \frac{\bar{X}_{-1}}{\bar{X}\bar{X}_{-1} - 1}, \quad \hat{b} = \frac{1}{\bar{X}_{-1}(\bar{X}\bar{X}_{-1} - 1)},$$

où  $\bar{X} = (\sum X_i)/n$  and  $\bar{X}_{-1} = (\sum 1/X_i)/n$  (voir Seshadri 1999, page 7). Cependant, on a une difficulté pour le dénominateur, car le triplet  $(p, a, b)$  doit être estimé par la méthode du maximum de vraisemblance pour les densités (1.1). Il n'y a pas d'expression explicite d'un tel estimateur, et des méthodes numériques sont nécessaires. ainsi, l'efficacité des statistiques de test du rapport de vraisemblance obtenus sont réduites par la complexité numérique et la non exactitude des expressions.

Dans cette communication, nous démontrons que les modèles GIG ont la propriété ULAN (*uniform local asymptotic normality*). L'intérêt de démontrer cette propriété ULAN pour les GIG est qu'elle constitue un outil clé dans la célèbre méthodologie de Le CaM (voir par exemple Le Cam et Yang, 1990). Cela permettra, dans une recherche future, de construire un test optimal de Le Cam pour les deux hypothèses nulles  $\mathcal{H}_0^\pm : p = \pm\frac{1}{2}$  contre  $\mathcal{H}_1^\pm : p \neq \pm\frac{1}{2}$ , qui utilise uniquement l'expression explicite des estimateurs du maximum de vraisemblance des paramètres  $a$  et  $b$  sous  $p = \pm\frac{1}{2}$ , et soit aussi efficace que le test du rapport de vraisemblance. Cette procédure ne serait pas uniquement valable pour les hypothèses nulles ci-dessus, mais pour tout test concernant les paramètres  $p, a$  et  $b$ , donc pour tester d'autres sous-modèles au sein des modèles GIG. Dans la section 2, nous établissons la propriété de normalité asymptotique locale uniforme (ULAN) pour

des modèles GIG généraux. La preuve est donnée dans la section 3, et nous évoquons brièvement des perspectives de recherche dans la section 4.

## 2 La propriété ULAN pour les modèles $GIG(p, a, b)$

Soit  $X_1, \dots, X_n$  un échantillon i.i.d. suivant la loi  $GIG(p, a, b)$  de densité (1.1). Soit  $\mathbb{P}_{p,a,b}^{(n)}$  la loi conjointe de  $(X_1, \dots, X_n)$ . Considérons le modèle paramétrique GIG

$$\mathcal{P}_{GIG}^{(n)} := \left\{ \mathbb{P}_{p,a,b}^{(n)} : p \in \mathbb{R}, a \in \mathbb{R}_0^+, b \in \mathbb{R}_0^+ \right\}.$$

**Theorem 2.1** *Soit  $\boldsymbol{\vartheta} = (p, a, b)'$  pour tout  $p \in \mathbb{R}, a \in \mathbb{R}_0^+$  and  $b \in \mathbb{R}_0^+$ . alors, pour des perturbations d'ordre  $n^{-1/2}$ , la famille de lois de probabilité  $\mathcal{P}_{GIG}^{(n)}$  est ULAN en  $\boldsymbol{\vartheta}$ , avec comme suite centrale*

$$\begin{aligned} \Delta^{(n)}(\boldsymbol{\vartheta}) &:= \begin{pmatrix} \Delta_1^{(n)}(\boldsymbol{\vartheta}) \\ \Delta_2^{(n)}(\boldsymbol{\vartheta}) \\ \Delta_3^{(n)}(\boldsymbol{\vartheta}) \end{pmatrix} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \left( \frac{\partial_p c(p,a,b)}{c(p,a,b)} + \log(X_i) \right) \\ \left( \frac{\partial_a c(p,a,b)}{c(p,a,b)} - \frac{X_i}{2} \right) \\ \left( \frac{\partial_b c(p,a,b)}{c(p,a,b)} - \frac{1}{2X_i} \right) \end{pmatrix}, \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} \left( \frac{1}{2} \log(a/b) - \partial_p \log K_p(\sqrt{ab}) + \log(X_i) \right) \\ \left( \frac{p}{2a} - \partial_a \log K_p(\sqrt{ab}) - \frac{X_i}{2} \right) \\ \left( -\frac{p}{2b} - \partial_b \log K_p(\sqrt{ab}) - \frac{1}{2X_i} \right) \end{pmatrix}, \end{aligned}$$

la matrice d'information de Fisher correspondante étant

$$\Gamma(\boldsymbol{\vartheta}) := \begin{pmatrix} \Gamma_{p,p}(\boldsymbol{\vartheta}) & \Gamma_{p,a}(\boldsymbol{\vartheta}) & \Gamma_{p,b}(\boldsymbol{\vartheta}) \\ \Gamma_{p,a}(\boldsymbol{\vartheta}) & \Gamma_{a,a}(\boldsymbol{\vartheta}) & \Gamma_{a,b}(\boldsymbol{\vartheta}) \\ \Gamma_{p,b}(\boldsymbol{\vartheta}) & \Gamma_{a,b}(\boldsymbol{\vartheta}) & \Gamma_{b,b}(\boldsymbol{\vartheta}) \end{pmatrix},$$

où

$$\begin{aligned} \Gamma_{p,p}(\boldsymbol{\vartheta}) &:= \partial_{pp}^2 \log K_p(\sqrt{ab}), & \Gamma_{a,a}(\boldsymbol{\vartheta}) &:= \frac{p}{2a^2} + \partial_{aa}^2 \log K_p(\sqrt{ab}), \\ \Gamma_{b,b}(\boldsymbol{\vartheta}) &:= -\frac{p}{2b^2} + \partial_{bb}^2 \log K_p(\sqrt{ab}), & \Gamma_{p,a}(\boldsymbol{\vartheta}) &:= -\frac{1}{2a} + \partial_{ap}^2 \log K_p(\sqrt{ab}), \\ \Gamma_{p,b}(\boldsymbol{\vartheta}) &:= \frac{1}{2b} + \partial_{ba}^2 \log K_p(\sqrt{ab}), & \Gamma_{a,b}(\boldsymbol{\vartheta}) &:= \partial_{ab}^2 \log K_p(\sqrt{ab}). \end{aligned}$$

Plus précisément, pour tout  $\boldsymbol{\vartheta}^{(n)} = (p^{(n)}, a^{(n)}, b^{(n)})' = \boldsymbol{\vartheta} + O(n^{-1/2})$  et pour toute suite bornée  $\boldsymbol{\tau}^{(n)} = (\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)})' \in \mathbb{R}^3$  telle que  $a^{(n)} + n^{-1/2}\tau_2^{(n)} > 0$  and  $b^{(n)} + n^{-1/2}\tau_3^{(n)} > 0$ , on a

$$\begin{aligned} \Lambda_{\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\tau}^{(n)} / \boldsymbol{\vartheta}^{(n)}}^{(n)} &:= \log(dP_{\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\tau}^{(n)}}^{(n)} / dP_{\boldsymbol{\vartheta}^{(n)}}^{(n)}) \\ &= \boldsymbol{\tau}^{(n)'} \boldsymbol{\Delta}^{(n)}(\boldsymbol{\vartheta}^{(n)}) - \frac{1}{2} \boldsymbol{\tau}^{(n)'} \Gamma(\boldsymbol{\vartheta}) \boldsymbol{\tau}^{(n)} + o_P(1), \end{aligned} \quad (2.2)$$

et  $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\vartheta}^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Gamma(\boldsymbol{\vartheta}))$ , sous  $P_{\boldsymbol{\vartheta}^{(n)}}^{(n)}$  lorsque  $n \rightarrow \infty$ .

### 3 Preuve du théorème

La preuve utilise le Lemme 1 de Swensen (1985)—plus précisément, sur sa généralisation dans Garel and Hallin (1995). Les conditions suffisantes pour avoir la propriété ULAN dans ces travaux découlent d'arguments standard, si l'on prouve que  $(p, a, b)' \mapsto f_{p,a,b}^{1/2}(x)$  est différentiable au moyenne quadratique en tout point  $(p, a, b)'$ . C'est l'objet du lemme suivant.

**Lemme 3.1** *Considérons*

$$D_p f_{p,a,b}^{1/2}(x) := \frac{1}{2} x^{(p-1)/2} e^{-(ax+b/x)/4} \left( \frac{\partial_p c(p, a, b)}{(c(p, a, b))^{1/2}} + (c(p, a, b))^{1/2} \log(x) \right),$$

$$D_a f_{p,a,b}^{1/2}(x) := \frac{1}{2} x^{(p-1)/2} e^{-(ax+b/x)/4} \left( \frac{\partial_a c(p, a, b)}{(c(p, a, b))^{1/2}} - (c(p, a, b))^{1/2} \frac{x}{2} \right),$$

et

$$D_b f_{p,a,b}^{1/2}(x) := \frac{1}{2} x^{(p-1)/2} e^{-(ax+b/x)/4} \left( \frac{\partial_b c(p, a, b)}{(c(p, a, b))^{1/2}} - (c(p, a, b))^{1/2} \frac{1}{2x} \right).$$

alors, pour tout  $p \in \mathbb{R}$ ,  $a \in \mathbb{R}_0^+$  et  $b \in \mathbb{R}_0^+$ , on a, lorsque  $(\rho, \alpha, \beta) \rightarrow (0, 0, 0)$ ,

$$(i) \int_0^\infty \left( f_{p+\rho,a,b}^{1/2}(x) - f_{p,a,b}^{1/2}(x) - \rho D_p f_{p,a,b}^{1/2}(x) \right)^2 dx = o(\rho^2),$$

$$(ii) \int_0^\infty \left( f_{p+\rho,a+\alpha,b}^{1/2}(x) - f_{p+\rho,a,b}^{1/2}(x) - \alpha D_a f_{p+\rho,a,b}^{1/2}(x) \right)^2 dx = o(\alpha^2),$$

$$(iii) \int_0^\infty \left( f_{p+\rho,a+\alpha,b+\beta}^{1/2}(x) - f_{p+\rho,a+\alpha,b}^{1/2}(x) - \beta D_b f_{p+\rho,a+\alpha,b}^{1/2}(x) \right)^2 dx = o(\beta^2),$$

$$(iv) \int_0^\infty \left( D_a f_{p+\rho,a,b}^{1/2}(x) - D_a f_{p,a,b}^{1/2}(x) \right)^2 dx = o(1),$$

$$(v) \int_0^\infty \left( D_b f_{p+\rho, a+\alpha, b}^{1/2}(x) - D_b f_{p, a, b}^{1/2}(x) \right)^2 dx = o(1),$$

$$(vi) \int_0^\infty \left\{ f_{p+\rho, a+\alpha, b+\beta}^{1/2}(x) - f_{p, a, b}^{1/2}(x) - \begin{pmatrix} \rho \\ \alpha \\ \beta \end{pmatrix}' \begin{pmatrix} D_p f_{p, a, b}^{1/2}(x) \\ D_a f_{p, a, b}^{1/2}(x) \\ D_b f_{p, a, b}^{1/2}(x) \end{pmatrix} \right\}^2 dx = o(\|(\rho, \alpha, \beta)'\|^2).$$

PREUVE DU LEMME 3.1. (i) On commence par écrire l'intégrale dans le membre de gauche sous la forme

$$\int_0^\infty x^{p-1} e^{-(ax+b/x)/2} \left( (c(p+\rho, a, b))^{1/2} x^{\rho/2} - (c(p, a, b))^{1/2} - \frac{1}{2}\rho \left( \frac{\partial_p c(p, a, b)}{(c(p, a, b))^{1/2}} + (c(p, a, b))^{1/2} \log(x) \right) \right)^2 dx.$$

Cette intégrale peut être majorée par  $c_1 S_1 + c_2 S_2 + c_3 \rho^2 S_3$ , où  $c_i \in \mathbb{R}^+$  pour tout  $i = 1, 2, 3$  et

$$S_1 = \int_0^\infty \left( (c(p+\rho, a, b))^{1/2} x^{\rho/2} - (c(p, a, b))^{1/2} x^{\rho/2} - \frac{1}{2}\rho \frac{\partial_p c(p, a, b)}{(c(p, a, b))^{1/2}} x^{\rho/2} \right)^2 x^{p-1} e^{-(ax+b/x)/2} dx,$$

$$S_2 = \int_0^\infty \left( (c(p, a, b))^{1/2} x^{\rho/2} - (c(p, a, b))^{1/2} - \frac{1}{2}\rho (c(p, a, b))^{1/2} \log(x) \right)^2 x^{p-1} e^{-(ax+b/x)/2} dx,$$

et

$$S_3 = \int_0^\infty \left( \frac{\partial_p c(p, a, b)}{(c(p, a, b))^{1/2}} \right)^2 (x^{\rho/2} - 1)^2 x^{p-1} e^{-(ax+b/x)/2} dx.$$

Puisque, pour tout  $\rho \in \mathbb{R}$ ,  $x^{p+\rho-1} e^{-(ax+b/x)/2}$  est intégrable sur  $(0, \infty)$ ,  $S_1$  peut être majoré par  $d \left( (c(p+\rho, a, b))^{1/2} - (c(p, a, b))^{1/2} - \frac{1}{2}\rho \frac{\partial_p c(p, a, b)}{(c(p, a, b))^{1/2}} \right)$  pour une constante strictement positive  $d$ . Un argument standard de développement en série de Taylor appliqué à  $p \mapsto c(p, a, b)$  montre alors que  $S_1 = o(\rho^2)$ . De même, on peut montrer que  $S_3 = o(1)$ . Quant à  $S_2$ , le théorème de la moyenne appliqué à  $x^{\rho/2} - 1$  permet de réécrire  $S_2$  sous la forme

$$\frac{1}{4}\rho^2 \int_0^\infty (\log(x) x^{\rho^*/2} - \log(x))^2 c(p, a, b) x^{p-1} e^{-(ax+b/x)/2} dx,$$

avec  $\rho^* \in (0, \rho)$ . Puisque  $\log(x) x^{p+\tilde{\rho}-1} e^{-(ax+b/x)/2}$  est intégrable sur  $(0, \infty)$  pour tout  $\tilde{\rho} \in \mathbb{R}$ , cette dernière expression, et donc  $S_2$ , est un  $o(\rho^2)$  par convergence dominée de Lebesgue.

Les preuves de (ii)-(v) sont similaires.

(vi) Pour prouver (vi), qui représente la différentiabilité en moyenne quadratique de la fonction  $(p, a, b)' \mapsto f_{p,a,b}^{1/2}(x)$ , on majore l'intégrale de gauche par la somme  $e_1 S_1 + e_2 S_2 + e_3 S_3 + e_4 \alpha^2 S_4 + e_5 \beta^2 S_5$  avec  $e_i \in \mathbb{R}^+$  pour tout  $i = 1, 2, 3, 4, 5$  et

$$S_1 = \int_0^\infty \left( f_{p+\rho,a,b}^{1/2}(x) - f_{p,a,b}^{1/2}(x) - \rho D_p f_{p,a,b}^{1/2}(x) \right)^2 dx,$$

$$S_2 = \int_0^\infty \left( f_{p+\rho,a+\alpha,b}^{1/2}(x) - f_{p+\rho,a,b}^{1/2}(x) - \alpha D_a f_{p+\rho,a,b}^{1/2}(x) \right)^2 dx,$$

$$S_3 = \int_0^\infty \left( f_{p+\rho,a+\alpha,b+\beta}^{1/2}(x) - f_{p+\rho,a+\alpha,b}^{1/2}(x) - \beta D_b f_{p+\rho,a+\alpha,b}^{1/2}(x) \right)^2 dx,$$

$$S_4 = \int_0^\infty \left( D_a f_{p+\rho,a,b}^{1/2}(x) - D_a f_{p,a,b}^{1/2}(x) \right)^2 dx,$$

et

$$S_5 = \int_0^\infty \left( D_b f_{p+\rho,a+\alpha,b}^{1/2}(x) - D_b f_{p,a,b}^{1/2}(x) \right)^2 dx.$$

Le résultat découle alors de (i)-(v). □

## 4 Perspectives

L'idée centrale de la théorie de Le Cam est le concept de convergence de modèles statistiques, qui permet d'approcher une famille de mesures de probabilité par une autre famille plus simple à étudier. L'ingrédient majeur est la propriété ULAN, d'où nous pourrions déduire que le modèle paramétrique GIG,  $\mathcal{P}_{GIG}^{(n)}$ , est localement (dans un voisinage de  $(p, a, b)'$ ) et asymptotiquement (pour des échantillons de grande taille) équivalent à un simple modèle gaussien. Intuitivement, cela est dû au fait que le développement du rapport de vraisemblance (2.2), aux termes de reste près, ressemble au rapport de vraisemblance d'un modèle gaussien. Puisque les procédures optimales pour les modèles statistiques gaussiens sont bien connues, il est possible de les utiliser pour obtenir des procédures inférentielles asymptotiquement "optimales" pour le modèle paramétrique GIG.

## Bibliographie

[1] Chhikara, R. S. and Folks, J. L. (1989) *The Inverse Gaussian Distribution, Theory, Methodology and applications*. Marcel Dekker Inc, New-York.

- [2] Garel, B., and Hallin, M. (1995) Local asymptotic normality of multivariate aRMA processes with a linear trend. *ann. Inst. Statist. Math.* **47**, 551–579.
- [3] Jorgensen B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer-Verlag, Heidelberg.
- [4] Kreiss, J-P. (1987) On adaptive estimation in stationary aRMA processes. *ann. Statist.* **15**, 112– 133.
- [5] Le Cam, L. and Yang, L. Y. *asymptotics in Statistics, Some Basic Concepts*. Springer-Verlag.
- [6] Seshadri, V. (1999) *The Inverse Gaussian Distribution*. Springer-Verlag, New York.
- [7] Swensen, a.R. (1985) The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate anal.* **16**, 54–70.