Characterizations of GIG laws: A survey

Angelo Efoévi Koudou[*]

Université de Lorraine and CNRS, Institut Elie Cartan de Lorraine, UMR 7502,
Vandoeuvre-lès-Nancy, F-54506, France
e-mail: Efoevi.Koudou@univ-lorraine.fr

and

Christophe Ley

ECARES and Département de Mathématique
Université libre de Bruxelles
Boulevard du Triomphe, CP210, B-1050 Brussels, Belgium
e-mail: chrisley@ulb.ac.be

Abstract: Several characterizations of the Generalized Inverse Gaussian (GIG) distribution on the positive real line have been proposed in the literature, especially over the past two decades. These characterization theorems are surveyed, and two new characterizations are established, one based on maximum likelihood estimation and the other is a Stein characterization.

1. Introduction

The *Generalized Inverse Gaussian* (hereafter GIG) distribution on the positive real line has been proposed by Good [14] in his study of population frequencies, yet its first appearance can be traced back to Etienne Halphen in the forties [16], whence the GIG is sometimes called Halphen Type A distribution. This distribution with parameters $p \in \mathbb{R}$, $a > 0$, $b > 0$ has density

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

with K_p the modified Bessel function of the third kind. The parameters a and b regulate both the concentration and scaling of the densities, the former via $\sqrt{ab}$ and the latter via $\sqrt{b/a}$. This is the reason why some authors (e.g., Jørgensen [19]) introduce the parameters $\theta = \sqrt{ab}$ and $\eta = \sqrt{b/a}$, leading to a GIG density of the form

$$f_{p,\theta,\eta}(x) := \frac{1}{2\eta K_p(\theta)} \left(\frac{x}{\eta}\right)^{p-1} e^{-\theta(x/\eta + \eta/x)/2}, \quad x > 0. \tag{1}$$

[*]Corresponding author.

The parameter p bears no concrete statistical meaning, but some particular values of p lead to well-known sub-models of the GIG such as the Inverse Gaussian ($p = -\frac{1}{2}$), the Reciprocal Inverse Gaussian ($p = \frac{1}{2}$), the hyperbolic ($p = 0$, in which case one also speaks of the Harmonic law), the positive hyperbolic ($p = 1$), the Gamma ($p > 0$ and $b = 0$) or the Reciprocal Gamma distribution ($p < 0$ and $a = 0$).

Several papers have investigated the probabilistic properties of the GIG distributions. For instance, Barndorff-Nielsen and Halgreen [1] have studied their convolution properties and established their infinite divisibility as well as interesting properties such as the equivalence $X \sim \mathrm{GIG}(p, a, b) \iff \frac{1}{X} \sim \mathrm{GIG}(-p, b, a)$. Vallois [59] has shown that, depending on the sign of p, the GIG laws can be viewed as the distributions of either first or last exit times for certain diffusion processes, generalizing the well-known fact that the Inverse Gaussian and Reciprocal Inverse Gaussian distributions are respectively the distributions of the first and the last hitting time for a Brownian motion. Madan, Roynette and Yor [36] have shown that the Black-Scholes formula in finance can be expressed in terms of the distribution function of GIG variables. GIG distributions belong to the family of generalized Γ-convolutions, as shown in Halgreen [15] (see also Eberlein and Hammerstein [11] for a detailed proof). Sichel ([54], [55]) used this distribution to construct mixtures of Poisson distributions.

The interesting properties of the GIG have led to the definition of matrix GIG distributions, containing the special case of Wishart matrices for $b = 0$. In this paper, we only deal with the one-dimensional GIG distribution. Since the matrix case will sometimes be referred to in the course of this survey, let us recall the definition of this distribution for matrix variates (see for instance Letac and Wesołowski [31]). Let $r \geq 1$ be an integer. Denote by $\mathcal{M}_r$ the Euclidean space of $r \times r$ real symmetric matrices. Consider the inner product $\langle a, b \rangle = \mathrm{tr}(ab)$ on $\mathcal{M}_r$. Let $\mathcal{M}_r^+$ be the cone of positive definite matrices in $\mathcal{M}_r$. For $p \in \mathbb{R}$, $a, b \in \mathcal{M}_r^+$, the GIG distribution with parameters p, a, b is the probability measure on $\mathcal{M}_r^+$ whose density with respect to the Lebesgue measure on $\mathcal{M}_r^+$ is proportional to

$$(\det x)^{p - \frac{r+1}{2}} \exp\left(-\frac{1}{2}(\langle a, x \rangle + \langle b, x^{-1} \rangle)\right), \quad x \in \mathcal{M}_r^+.$$

For an overview of results on matrix GIG laws, we refer the reader to Massam and Wesołowski ([38], [39]). Throughout the paper, unless explicitly stated, the name *GIG distribution* refers to the one-dimensional case $r = 1$.

The GIG distributions have been used in the modelization of diverse real phenomena such as, for instance, waiting time (Jørgensen [19]), neural activity (Iyengar and Liao [18]), or, most importantly, hydrologic extreme events (see Chebana *et al.* [6] and references therein). Despite this popularity of GIG models, statistical aspects of the GIG distributions have received much less attention in the literature than their probabilistic counterparts. The cornerstone reference in this respect is Jørgensen [19], complemented by the stream of literature on general Halphen distributions (of which the GIG or Halphen A is one of three representatives), see e.g. Perreault *et al.* ([44], [45]) or Chebana *et al.* [6]. Quite

recently, Koudou and Ley [24] have applied the Le Cam methodology (see [27]) to GIG distributions in order to construct optimal (in the maximin sense) testing procedures within the GIG family of distributions.

The present paper is concerned with characterization results of GIG laws. In probability and statistics, a characterization theorem occurs when a given distribution is the only one which satisfies a certain property. Besides their evident mathematical interest *per se*, characterization theorems also deepen our understanding of the distributions under investigation and sometimes open unexpected paths to innovations which might have been uncovered otherwise. Over the years, an important number of such characterization theorems for GIG laws have been presented in the literature, especially over the past two decades. We therefore propose here a survey of existing characterizations (without pretending to be exhaustive), amended by two novel characterizations, one based on the maximum likelihood estimator (MLE) of the scale parameter $\eta = \sqrt{b/a}$ of GIG laws and the other is a so-called Stein characterization.

The paper is organized as follows. A review of known characterizations of GIG distributions is presented in Section 2, while the short Section 3 is devoted to our two new characterizations.

2. Known characterizations of GIG laws

2.1. A characterization by continued fractions

The following theorem characterizes the GIG distribution as the law of a continued fraction with independent Gamma entries. We adopt the notation $\Gamma(p, a/2) = \mathrm{GIG}(p, a, 0)$ for the Gamma distribution with parameters $p, a > 0$ and $=_d$ for equality in distribution.

Theorem 2.1 (Letac and Seshadri, [29]).

- *Let X and Y be two independent random variables such that $X > 0$ and $Y \sim \Gamma(p, a/2)$ for $p, a > 0$. Then $X =_d \frac{1}{Y+X}$ if and only if $X \sim \mathrm{GIG}(-p, a, a)$.*
- *Let X, Y_1 and Y_2 be three independent random variables such that $X > 0$, $Y_1 \sim \Gamma(p, b/2)$ and $Y_2 \sim \Gamma(p, a/2)$ for $p, a, b > 0$. Then $X =_d \frac{1}{Y_1 + \frac{1}{Y_2 + X}}$ if and only if $X \sim \mathrm{GIG}(-p, a, b)$.*
- *If $(Y_i)_{i \geq 1}$ is a sequence of independent random variables such that*

$$\mathcal{L}(Y_{2i-1}) = \mathcal{L}(Y_1) = \Gamma(p, b/2) \ \ and \ \ \mathcal{L}(Y_{2i}) = \mathcal{L}(Y_2) = \Gamma(p, a/2); \ \ i \geq 1,$$

then the distribution of the random variable

$$\cfrac{1}{Y_1 + \cfrac{1}{Y_2 + \cfrac{1}{Y_3 + \ddots}}} \tag{2}$$

is $GIG(-p, a, b)$.

The proof relies on properties of continued fractions, which are exploited to show that the GIG distribution is the unique stationary distribution of the Markov chain $(X_m)_m$ defined by

$$X_{m+1} = \frac{1}{Y_{m+1} + X_m},$$

in the case $a = b$, or by

$$X_{m+1} = \frac{1}{Y_{2m+1} + \frac{1}{Y_{2m+2} + X_m}}$$

in the general case.

This characterization by continued fractions is one key ingredient in the proof given by Letac and Wesołowski [31] of the characterization of GIG laws by the Matsumoto-Yor property, which we now recall.

2.2. The Matsumoto-Yor property

Consider two independent, positive random variables X and Y such that

$$X \sim \text{GIG}(-p, a, b), \quad Y \sim \Gamma(p, a/2)$$

for some $p, a, b > 0$. The Matsumoto-Yor property is the fact that the random variables

$$U = \frac{1}{X + Y} \quad \text{and} \quad V = \frac{1}{X} - \frac{1}{X + Y} \tag{3}$$

are independent. This property has been discovered in Matsumoto and Yor [40] for the case $a = b$ while studying certain exponential functionals of Brownian motion. Letac and Wesołowski have noticed afterwards that this property remains true if $a \neq b$, although their paper has finally been published earlier than the Matsumoto-Yor paper from 2001, namely in 2000. Letac and Wesołowski [31] have further proved that this property is in fact a characterization of GIG laws (more exactly, of the product of a GIG and a Gamma law with suitable parameters).

Theorem 2.2 (Letac and Wesołowski, [31])**.** *Consider two non-Dirac, positive and independent random variables X and Y. Then the random variables $U = (X + Y)^{-1}$ and $V = X^{-1} - (X + Y)^{-1}$ are independent if and only if there exist $p > 0$, $a > 0$ and $b > 0$ such that $X \sim \text{GIG}(-p, a, b)$ while $Y \sim \Gamma(p, a/2)$.*

Let us make a few remarks on the theorem:

- As shown in the same paper, this result holds true also for matrix variates (characterization of the product of matrix GIG and Wishart variables) under a smoothness assumption not needed in the scalar case. To prove this, the authors use the extension to the matrix case, established by Bernadac [3], of the above mentioned continued-fraction characterization of GIG distributions.

- An interpretation of this property in terms of Brownian motion has been pointed out by Matsumoto and Yor [41]. Massam and Wesołowski [38] have provided a tree-version of the Matsumoto-Yor property, and an interpretation of that tree-version by means of a family of Brownian motions is given in Wesołowski and Witkowski [62].
- For $p = -1/2$, the Matsumoto-Yor property can be seen as a consequence of an independence property established by Barndorff-Nielsen and Koudou [2] on a tree-network of Inverse Gaussian resistances (see Koudou [23]).
- Koudou and Vallois ([25], [26]) investigated a generalization of the Matsumoto-Yor property, by looking for smooth decreasing functions f from $(0, \infty)$ onto $(0, \infty)$ with the following property: there exist independent, positive random variables X and Y such that the variables $U = f(X+Y)$ and $V = f(X) - f(X+Y)$ are independent. This led to other independence properties of the Matsumoto-Yor type. In particular, an independence property characterizing Kummer distributions has been established.

2.3. Characterizations by constant regression

In a subsequent work, Wesołowski [61] relaxed the independence condition between U and V in Theorem 2.2 by assuming that V and $1/V$ have constant regression on U. Following Lukacs [35], we say that a random variable Z with finite expectation has a constant regression on a random variable T if $\mathbb{E}(Z|T) = \mathbb{E}(Z)$ almost surely. As can be seen in the following theorems, the price to pay for the relaxation at the level of independence is to assume the existence of moments of X and/or $1/X$.

Theorem 2.3 (Wesołowski [61])**.** *Consider two non-Dirac, positive and independent random variables X and Y such that $\mathbb{E}(1/X) < \infty$. Let U and V be defined by (3), and assume that $\mathbb{E}(V|U) = c$ and $\mathbb{E}(1/V|U) = d$ for some real constants c, d.*

Then $cd > 1$ and there exists $a > 0$ such that $X \sim \mathrm{GIG}(-p, 2a, 2b)$ and $Y \sim \Gamma(p, a)$ where $p = cd/(cd - 1)$ and $b = d/(cd - 1)$.

One can prove that X is GIG distributed, assuming that Y is Gamma distributed, and vice-versa:

Theorem 2.4 (Seshadri and Wesołowski [51])**.** *Consider two non-Dirac, positive and independent random variables X and Y such that $\mathbb{E}(1/X)$ and $\mathbb{E}(X)$ are finite. Define U and V as in (3). Suppose $Y \sim \Gamma(p, a/2)$ for some $p > 0, a > 0$. If the regression of V on U is constant, i.e. $\mathbb{E}(V|U) = c$, then $X \sim \mathrm{GIG}(-p, a, b)$, with $b = 2p/c$.*

Theorem 2.5 (Seshadri and Wesołowski [51])**.** *Consider two non-Dirac, positive and independent random variables X and Y such that $\mathbb{E}(Y) < \infty$. Assume that $X \sim \mathrm{GIG}(-p, a, b)$ for some positive p, a, b. If $\mathbb{E}(V|U) = 2p/b$, then $Y \sim \Gamma(p, a/2)$.*

Theorems [2.3], [2.4] and [2.5] are established by proving that Laplace transforms of measures linked with the laws of X and Y are probabilistic solutions of second-order differential equations. Wesołowski [61] has pointed out that the condition $\mathbb{E}(V|U) = c$ can be expressed in terms of linearity of the regression of Y/X on $X + Y$, under the form

$$\mathbb{E}\left(\left.\frac{Y}{X}\,\right|\, X + Y\right) = c(X + Y).$$

This can easily be seen from the definition of U and V in [(3)].

The results of Wesołowski [61] and Seshadri and Wesołowski [51] have been slightly extended in Chou and Huang [9].

Theorem 2.6 (Chou and Huang [9])**.** *Consider two non-Dirac, positive and independent random variables X and Y and define U and V as in ([3]). Assume that, for some fixed integer r, $\mathbb{E}(X^{-r-2})$, $\mathbb{E}(X^{-r})$, $\mathbb{E}(Y^r)$ and $\mathbb{E}(Y^{r+2})$ are finite.*

Assume that, for some constants c_r and c_{r+1},

$$\mathbb{E}(V^{r+1}|U) = c_r\mathbb{E}(V^r|U)$$

and

$$\mathbb{E}(V^{r+2}|U) = c_{r+1}\mathbb{E}(V^{r+1}|U).$$

Then, $c_{r+1} > c_r > 0$ and there exists $a > 0$ such that $X \sim \mathrm{GIG}(-p, 2a, 2b)$ and $Y \sim \Gamma(p, a)$ with $p = c_r/(c_{r+1} - c_r) - r > 0$ and $b = 1/(c_{r+1} - c_r) > 0$.

Lukacs [35] (Theorem 6.1) characterized the common normal distribution of i.i.d. random variables $X_1, X_2, \ldots, X_n$ by the constancy of regression of a quadratic function of $X_1, X_2, \ldots, X_n$ on the sum $\Lambda = \sum_{i=1}^n X_i$. The following theorem characterizing the GIG distribution belongs also to the family of characterizations of distributions by constancy of regression of a statistic S on the sum Λ of observations.

Theorem 2.7 (Pusz [47])**.** *Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed copies of a random variable $X > 0$ with $\mathbb{E}(1/X^2)$, $\mathbb{E}(1/X)$ and $\mathbb{E}(X)$ finite. Consider $q > 0$ and $p \in \mathbb{R}$ such that $p\mathbb{E}(1/X) + q\mathbb{E}(1/X^2) > 0$ and define $\Lambda = \sum_{i=1}^n X_i$ and*

$$S = \Lambda\left(p\sum_{i=1}^n \frac{1}{X_i} + q\sum_{i=1}^n \frac{1}{X_i^2}\right) - nq\sum_{i=1}^n \frac{1}{X_i}.$$

Then $\mathbb{E}(S|\Lambda) = c$ for some constant c if and only if there exist $\mu \in \mathbb{R}$, $a, b, \delta > 0$ such that

$$p = \delta(\mu - 1),\ q = \delta b/2,\ c = \delta n(n\mu - 1)$$

and $X \sim \mathrm{GIG}(p, a, b)$.

The proof is based on the fact that the GIG distribution is characterized by its moments, on a lemma by Kagan *et al.* [20] giving a necessary and sufficient condition for the constancy of $\mathbb{E}(S|\Lambda)$, and on a careful manipulation of a differential equation satisfied by the function $f(t) = \mathbb{E}(1/X^2)\exp(itX)$.

2.4. Entropy characterization

Characterizing a distribution by the maximum entropy principle dates back to Shannon [53] who showed that Gaussian random variables maximize entropy among all real-valued random variables with given mean and variance. Since then, many examples of such characterizations have appeared in the literature. For instance, Kagan *et al.* [20] characterized several well-known distributions such as the exponential, Gamma or Beta distributions in terms of maximum entropy given various constraints, e.g.: *(i)* the exponential distribution with parameter λ maximizes the entropy among all the distributions of X supported on the positive real line under the constraint $\mathbb{E}(X) = 1/\lambda$; *(ii)* the Gamma distribution $\Gamma(p, a)$ maximizes the entropy under the constraints $\mathbb{E}(X) = p/a$ and $\mathbb{E}(\log X) = \Psi(p) - \log a$, where Ψ is the Digamma function; *(iii)* the Beta distribution with parameters $a, b > 0$ maximizes the entropy among all the distributions of X supported on $[0, 1]$ under the constraints $\mathbb{E}(\log X) = \Psi(a) - \Psi(a + b)$ and $\mathbb{E}(\log(1 - X)) = \Psi(b) - \Psi(a + b)$. This type of characterization exists as well for GIG distributions, as established in the following theorem due to Kawamura and Kōsei [22], for which we recall that the Bessel function of the third kind is defined by

$$K_p(z) = 2^{-p-1} z^p \int_0^\infty x^{-p-1} e^{-x - \frac{z^2}{4x}} dx, \quad \mathrm{Re}(z) > 0.$$

Theorem 2.8. *The distribution of X with density f on $(0, \infty)$ which maximizes the entropy*

$$H(X) = -\int_0^\infty f(x) \log f(x)\, dx$$

under the constraints

$$\mathbb{E}\left(\log \frac{X}{\sqrt{b/a}}\right) = \partial_p \log K_p(\sqrt{ab})$$

and

$$\mathbb{E}\left(\frac{X}{\sqrt{b/a}} + \frac{\sqrt{b/a}}{X}\right) = \frac{K_{p+1}(\sqrt{ab}) + K_{p-1}(\sqrt{ab})}{K_p(\sqrt{ab})}$$

is the distribution $GIG(p, a, b)$.

One application of entropy characterizations are goodness-of-fit tests. An instance is Vasicek [60] who designed an entropy test of normality, using Shannon's characterization. In our GIG setting, Mudholkar and Tian [42] established a goodness-of-fit test for the special case of an Inverse Gaussian (IG) distribution based on an entropy characterization that we recall below. Since the IG is a major particular case of GIG distributions —the latter were actually introduced in order to meet the needs of adding a third parameter to the IG distribution—we devote a brief subsection to characterizations of the IG distribution.

2.5. *Some characterizations of the IG distribution*

Of particular interest for the present survey is the Inverse Gaussian distribution with density

$$\frac{\sqrt{b}}{\sqrt{2\pi}}e^{\sqrt{ab}}x^{-\frac{3}{2}}e^{-(ax+b/x)/2}, \quad x > 0.$$

The IG distribution is useful in many fields for data modeling and analysis when the observations are right-skewed. Numerous examples of applications can be found in Chhikara and Folks [8], Seshadri [49] and Seshadri [50]. The IG distribution possesses many similarities in terms of statistical properties with the Gaussian distribution, as pointed out by e.g. Mudholkar and Tian [42]. We give here a few results characterizing the IG distribution.

2.5.1. *Entropy characterizations*

Let us first observe that, as an immediate corollary of Theorem 2.8, we obtain the following (new) entropy characterization of the IG distribution.

Theorem 2.9. *Denote by* $c_{a,b}$ *the value of* $\partial_p \log K_p(\sqrt{ab})$ *at* $p = -1/2$. *The distribution of* X *with density* f *on* $(0,\infty)$ *having maximum entropy under the constraints*

$$\mathbb{E}\left(\log \frac{X}{\sqrt{b/a}}\right) = c_{a,b} \quad and \quad \mathbb{E}\left(\frac{X}{\sqrt{b/a}} + \frac{\sqrt{b/a}}{X}\right) = 2 + \frac{1}{\sqrt{ab}}$$

is the distribution $\mathrm{GIG}(-1/2, a, b) = \mathrm{IG}(a, b)$.

Mudholkar and Tian [42] have obtained another entropy characterization of the IG distribution.

Theorem 2.10 (Mudholkar and Tian [42]). *A random variable* X *has the* $IG(\sqrt{b/a}, b)$ *distribution if and only if* $1/\sqrt{X}$ *attains maximum entropy among all absolutely continuous random variables* $Y \geq 0$ *such that*

$$\mathbb{E}(Y^{-2}) = \sqrt{b/a} \quad and \quad \mathbb{E}(Y^2) = \sqrt{a/b} + 1/b.$$

Inspired by the Vasicek [60] methodology for the entropy test of normality, Mudholkar and Tian [42], as stated earlier, used the above theorem to construct a goodness-of-fit test for the IG distribution. The test statistic is based on a non-parametric estimation of the entropy of $1/\sqrt{X}$ and rejection occurs when this estimation is not close enough to the theoretical value under the null hypothesis of X being IG-distributed. The rejection criterion is assessed at given significance level by critical values obtained via Monte Carlo simulations.

2.5.2. *A martingale characterization*

Seshadri and Wesołowski [52] established the following martingale characterization of the IG distribution.

Theorem 2.11 (Seshadri and Wesołowski [52])**.** *Let* $(X_n)_{n\geq 1}$ *be a sequence of positive, non degenerate i.i.d. random variables. For* $n \geq 1$*, define* $S_n = X_1 + X_2 + \cdots + X_n$ *and consider the* σ*-algebra* $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \ldots)$*. Let* $b > 0$*. The sequence*

$$\left(\frac{n}{S_n} - \frac{1}{2bn}, \, \mathcal{F}_n \right)_{n \geq 1}$$

is a backward martingale if and only if $X_1 \sim \mathrm{IG}(a,b)$ *for some* $a > 0$*.*

Remark. In fact, the result established in Seshadri and Wesołowski [52] is more general than stated in Theorem 2.11. They proved that $(\frac{\alpha_n}{S_n} - \beta_n, \, \mathcal{F}_n)_{n\geq 1}$ is a backward martingale for some sequences $(\alpha_n)_n$ and $(\beta_n)_n$ if and only if X_1 follows one of four distributions, among which the IG distribution. The proof uses the Letac and Mora [28] characterization of natural exponential families with cubic variance function.

We conclude this section by writing down the well-known Khatri characterization of the IG distribution (see Letac and Seshadri [30] and also Seshadri [49] which contains a review of characterization problems for the IG distribution).

Theorem 2.12 (Khatri [21])**.** *Let* $X_1, X_2, \ldots, X_n$ *be i.i.d. copies of a random variable* X *such that* $\mathbb{E}(1/X)$*,* $\mathbb{E}(X)$*,* $\mathbb{E}(X^2)$ *and* $\mathbb{E}(1/\sum_{i=1}^n X_i)$ *exist and do not vanish.*

Let

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad and \quad \bar{X}_{-1} = \frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}.$$

The random variables $\bar{X}$ *and* $\bar{X}_{-1} - 1/\bar{X}$ *are independent if and only if there exist* $a, b > 0$ *such that* $X \sim \mathrm{IG}(a,b)$*.*

If one considers the parameters $\mu = \sqrt{b/a}$ and $\lambda = b$, then $\bar{X}$ and $1/(\bar{X}_{-1} - 1/\bar{X})$ are respectively the maximum likelihood estimators of μ and λ. The independence between these estimators characterizes the IG distribution, as the independence between the empirical mean and standard deviation as maximum likelihood estimators of location and scale characterizes the normal distribution. This is one of the similarities between the IG and normal distributions, as observed by Mudholkar and Tian [42].

3. New characterizations of GIG laws

3.1. An MLE characterization

A famous characterization theorem in statistics due to Carl Friedrich Gauss [12] says the following: the sample mean $\bar{X}$ is for any i.i.d. sample $X_1, \ldots, X_n$ of any sample size n the maximum likelihood estimator (MLE) of the location parameter in a location family $\{f(x - \mu), \mu \in \mathbb{R}\}$ if and only if the samples are drawn from a Gaussian population (with variance not fixed). This very first *MLE characterization theorem* has important implications, as it clearly

indicates that, the further one is away from the Gaussian situation, the less efficient becomes the sample mean as estimator. Several other characterization theorems have emerged in the past, linking particular forms of MLEs to a specific distribution such as the one-parameter exponential family (Poincaré [46]), the (negative) exponential distribution (Teicher [58]), the Laplace distribution (Kagan *et al.* [20]), the Gamma distribution (Marshall and Olkin [37]) or the Harmonic Law (Hürlimann [17]). For a recent overview and a unified theory on these results, we refer to Duerinckx *et al.* [10].

In this section, our aim is to apply the general result of Duerinckx *et al.* [10] to the GIG distribution in order to construct an MLE characterization theorem for GIG laws. To this end, we shall rather have recourse to the re-formulation (1) of the density, as in that parameterization the family $\{f_{p,\theta,\eta}(x),\ \eta > 0\}$ is a scale family. Let us first observe, by a simple calculation, that the MLE of η for fixed p and θ is

$$\hat{\eta} = \frac{\sqrt{p^2 + \theta^2 \bar{X}\bar{X}_{-1}} - p}{\theta \bar{X}_{-1}},$$

where we recall that $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{X}_{-1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}$ if $X_1,\ldots X_n$ are i.i.d. with common distribution $\mathrm{GIG}(p, a, b)$. This enables us to formulate the following.

Theorem 3.1. *Consider an integer $n \geq 3$. For $p \in \mathbb{R}$ and $\theta > 0$, $\frac{\sqrt{p^2 + \theta^2 \bar{X}\bar{X}_{-1}} - p}{\theta \bar{X}_{-1}}$ is the MLE of the scale parameter η in a scale family $\left\{\frac{1}{\eta}f(x/\eta), \eta > 0\right\}$ on $(0, \infty)$ for all samples $X_1,\ldots,X_n$ of fixed sample size n if and only if there exists $d > 0$ such that the family is $\{f_{pd,\theta d,\eta}(x),\ \eta > 0\}$, i.e. the family of densities of the distributions $\mathrm{GIG}(pd, \theta d/\eta, \theta \eta d),\ \eta > 0$.*

Proof of Theorem 3.1. We want to apply Theorem 5.1 of Duerinckx *et al.* [10]. To this end, we first need to calculate the scale score function of the standardized GIG density $f_{p,\theta}(x) := \frac{1}{2K_p(\theta)}x^{p-1}e^{-\frac{1}{2}\theta(x+1/x)}$ defined on $(0, \infty)$. This score, defined as $\psi_{f_{p,\theta}}(x) = 1 + x\frac{f'_{p,\theta}(x)}{f_{p,\theta}(x)}$, corresponds to

$$\psi_{f_{p,\theta}}(x) = p - \frac{\theta}{2}x + \frac{\theta}{2x}.$$

One easily sees that $\psi_{f_{p,\theta}}$ is invertible over $(0, \infty)$ with image $\mathbb{R}$. Thus Theorem 5.1 of Duerinckx *et al.* [10] applies and the so-called MNSS (Minimal Necessary Sample Size, the smallest sample size for which an MLE characterization holds) equals 3. This theorem then yields that $\frac{\sqrt{p^2 + \theta^2 \bar{X}\bar{X}_{-1}} - p}{\theta \bar{X}_{-1}}$ is the MLE of the scale parameter η in a scale family $\left\{\frac{1}{\eta}f(x/\eta), \eta > 0\right\}$ on $(0, \infty)$ for all samples $X_1,\ldots,X_n$ of fixed sample size $n \geq 3$ if and only if the densities are proportional to $x^{d-1}(f_{p,\theta}(x))^d$ for $d > 0$, which is nothing else but $f_{pd,\theta d}(x)$, the (standardized) GIG density with parameters pd and θd. This yields the announced MLE characterization for GIG laws. $\square$

Clearly, this MLE characterization theorem for GIG laws contains, *inter alia*, MLE characterizations for the Inverse Gaussian, the Reciprocal Inverse Gaussian and the hyperbolic or Harmonic Law, hence the characterization of Hürlimann [17] is a special case of our Theorem 3.1.

3.2. A Stein characterization

The celebrated Stein's method of normal approximation, introduced in Stein [56], has over the years been adapted to several other probability distributions, including the Poisson (Chen [7]), the exponential (Chatterjee *et al.* [5]), the Gamma (Luk [34]), the multinomial (Loh [33]), the geometric (Peköz [43]), the negative binomial (Brown and Phillips [4]) or the Beta (Goldstein and Reinert [13]). A first step in this method consists in finding a suitable Stein operator, whose properties determine the quality of the approximation; see Ross [48] for a recent overview on the intricacies of Stein's method. This operator satisfies a *Stein characterization theorem* which clearly links it to the targeted distribution.

To the best of the authors' knowledge, so far no such Stein characterization has been proposed in the literature for the GIG distribution, whence the reason of the following result.

Theorem 3.2. *A positive random variable X follows the $GIG(p, a, b)$ distribution if and only if for any differentiable function h satisfying*

$$\lim_{x \to \infty} f_{p,a,b}(x) h(x) = \lim_{x \to 0} f_{p,a,b}(x) h(x) = 0$$

we have

$$\mathbb{E}\left[h'(X) + \left(\frac{p-1}{X} + \frac{b}{2X^2} - \frac{a}{2} \right) h(X) \right] = 0. \tag{4}$$

The functional $h \mapsto \mathcal{T}_{f_{p,a,b}}(h)(x) := h'(x) + \left(\frac{p-1}{x} + \frac{b}{2x^2} - \frac{a}{2} \right) h(x)$ is the GIG Stein operator.

Proof of Theorem 3.2. The sufficient condition is readily checked by noting that $h'(x) + \left(\frac{p-1}{x} + \frac{b}{2x^2} - \frac{a}{2} \right) h(x) = \frac{(h(x) f_{p,a,b}(x))'}{f_{p,a,b}(x)}$, since then

$$\mathbb{E}\left[h'(X) + \left(\frac{p-1}{X} + \frac{b}{2X^2} - \frac{a}{2} \right) h(X) \right] = \int_0^\infty (h(x) f_{p,a,b}(x))' dx = 0$$

by the conditions on h. To see the necessity, write f for the density of X and define for $z > 0$ the function

$$l_z(u) := \mathbb{I}_{(0,z]}(u) - F_{p,a,b}(z), \quad u > 0,$$

with $F_{p,a,b}$ the GIG cumulative distribution function and $\mathbb{I}$ an indicator function. Then the function

$$h_z(x) := \frac{1}{f_{p,a,b}(x)} \int_0^x l_z(u) f_{p,a,b}(u) du \left(= -\frac{1}{f_{p,a,b}(x)} \int_x^\infty l_z(u) f_{p,a,b}(u) du \right)$$

is differentiable and satisfies $\lim_{x\to\infty} f_{p,a,b}(x)h_z(x) = \lim_{x\to 0} f_{p,a,b}(x)h_z(x) = 0$ for all z (since $\int_0^\infty l_z(u)f_{p,a,b}(u) = 0$), hence is a candidate function for the functions h that verify (4). Using hence this h_z leads to

$$
\begin{aligned}
0 &= \mathbb{E}\left[h_z'(X) + \left(\frac{p-1}{X} + \frac{b}{2X^2} - \frac{a}{2}\right)h_z(X)\right] \\
&= \mathbb{E}\left[\frac{(h_z(X)f_{p,a,b}(X))'}{f_{p,a,b}(X)}\right] \\
&= \int_0^\infty l_z(x)f(x)dx \\
&= F(z) - F_{p,a,b}(z)
\end{aligned}
$$

for every $z > 0$, from which we can conclude that $F = F_{p,a,b}$. $\square$

This result is a particular instance of the *density approach* to Stein characterizations initiated in Stein *et al.* [57] and further developed in Ley and Swan [32]. We attract the reader's attention to the fact that we could replace the functions $h(x)$ with $h(x)x^2$, turning the GIG Stein operator $\mathcal{T}_{f_{p,a,b}}(h)$ into the perhaps more tractable form

$$
\mathcal{T}_{f_{p,a,b}}(h)(x) = x^2 h'(x) + \left(-x^2\frac{a}{2} + (p+1)x + \frac{b}{2}\right)h(x).
$$

This new Stein characterization and the associated Stein operator(s) could be exploited to derive rates of convergence in some asymptotics related to GIG distributions, e.g. the rate of convergence of the law of the continued fraction in (2) to the GIG distribution. This kind of application will be investigated in future work.

Acknowledgements

References

[1] BARNDORFF-NIELSEN, O. E. and HALGREEN, C. (1977). Infinite divisibility of the Hyperbolic and generalized inverse Gaussian distribution. *Z. Wahrsch. Verw. Geb.* **38**, 309–312. MR0436260

[2] BARNDORFF-NIELSEN, O. E. and KOUDOU, A. E. (1998). Trees with random conductivities and the (reciprocal) inverse Gaussian distribution. *Adv. Appl. Probab.* **30**, 409–424. MR1642846

[3] BERNADAC, E. (1995). Random continued fractions and inverse Gaussian distribution on a symmetric cone. *J. Theor. Probab.* **8**, 221–260. MR1325851

[4] BROWN, T. and PHILLIPS, M. (1999). Negative binomial approximation with Stein's method. *Methodol. Comput. Appl. Probab.* **1**, 407–421. MR1770372

[5] CHATTERJEE, S., FULMAN, S. and RÖLLIN, A. (2011). Exponential approximation by Stein's method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat.* **8**, 197–223. MR2802856

[6] CHEBANA, F., EL ADLOUNI, S. and BOBÉE, B. (2010). Mixed estimation methods for Halphen distributions with applications in extreme hydrologic events. *Stoch. Environ. Res. Risk Assess.* **24**, 359–376.

[7] CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

[8] CHHIKARA, R. S. and FOLKS, J. L. (1989). *The Inverse Gaussian Distribution, Theory, Methodology and Applications.* Marcel Dekker Inc, New-York.

[9] CHOU, C.-W. and HUANG, J.-W. (2004). On characterizations of the gamma and the generalized inverse Gaussian distributions. *Stat. Probab. Lett.* **69**, 381–388. MR2091757

[10] DUERINCKX, M., LEY, C. and SWAN, Y. (2013). Maximum likelihood characterization of distributions. *Bernoulli,* to appear. MR3178518

[11] EBERLEIN, E. and HAMMERSTEIN, E. A. (2004). Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes. *Prog. Probab.* **58**, 221–264. MR2096291

[12] GAUSS, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.* Cambridge Library Collection. Cambridge University Press, Cambridge. Reprint of the 1809 original. MR2858122

[13] GOLDSTEIN, L. and REINERT, G. (2013). Stein's method for the Beta distribution and the Pòlya-Eggenberger urn. *Adv. Appl. Probab.,* to appear. MR3161381

[14] GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–260. MR0061330

[15] HALGREEN, C. (1979). Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Z. Wahrsch. verw. Geb.* **47**, 13–17. MR0521527

[16] HALPHEN, E. (1941). Sur un nouveau type de courbe de fréquence. *Comptes Rendus de l'Académie des Sciences* **213**, 633–635. Published under the name of "Dugué" due to war constraints. MR0009271

[17] HÜRLIMANN, W. (1998). On the characterization of maximum likelihood estimators for location-scale families. *Comm. Statist. Theory Methods* **27**, 495–508. MR1621418

[18] IYENGAR, S. and LIAO, Q. (1997). Modeling neural activity using the generalized inverse Gaussian distribution. *Biol. Cybern.* **77**, 289–295.

[19] Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution.* Springer-Verlag, Heidelberg. MR0648107

[20] Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics.* Wiley, New York. MR0346969

[21] Khatri, C. G. (1962). A characterisation of the inverse Gaussian distribution. *Ann. Math. Statist.* **33**, 800–803. MR0137197

[22] Kawamura, T. and Kōsei, I. (2003). Characterisations of the distributions of power inverse Gaussian and others based on the entropy maximisation principle. *J. Japan Statist. Soc.* **33**, 95–104. MR2044965

[23] Koudou, A. E. (2006). A link between the Matsumoto-Yor property and an independence property on trees. *Stat. Probab. Lett.* **76**, 1097–1101. MR2269279

[24] Koudou, A. E. and Ley, C. (2014). Efficiency combined with simplicity: new testing procedures for Generalized Inverse Gaussian models. *Test*, to appear. DOI: 10.1007/s11749-014-0378-2.

[25] Koudou, A. E. and Vallois, P. (2011). Which distributions have the Matsumoto-Yor property? *Electron. Commun. Prob.* **16**, 556–566. MR2836761

[26] Koudou, A. E. and Vallois, P. (2012). Independence properties of the Matsumoto-Yor type. *Bernoulli* **18**, 119–136. MR2888701

[27] Le Cam, L. and Yang, G. L. (2000). *Asymptotics in statistics. Some basic concepts.* 2nd ed. Springer-Verlag, New York. MR1784901

[28] Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Ann. Statist.* **18**, 1–37. MR1041384

[29] Letac, G. and Seshadri, V. (1983). A characterization of the generalized inverse Gaussian distribution by continued fractions. *Z. Wahrsch. Verw. Geb.* **62**, 485–489. MR0690573

[30] Letac, G. and Seshadri, V. (1985). On Khatri's characterization of the inverse Gaussian distribution. *Can. J. Stat.* **13**, 249–252. MR0818329

[31] Letac, G. and Wesołowski, J. (2000). An independence property for the product of GIG and gamma laws. *Ann. Probab.* **28**, 1371–1383. MR1797878

[32] Ley, C. and Swan, Y. (2013). Stein's density approach and information inequalities. *Electron. Comm. Probab.* **18**, 1–14. MR3019670

[33] Loh, W. L. (1992). Stein's method and multinomial approximation. *Ann. Appl. Probab.* **2**, 536–554. MR1177898

[34] Luk, H.-M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis. University of Southern California. Los Angeles, USA. MR2693204

[35] Lukacs, E. (1956). Characterization of populations by properties of suitable statistics. *Proc. Third Berkeley Symp. on Math. Statist. and Prob.*, Vol. 2 (Univ. of Calif. Press), 195–214. MR0084892

[36] Madan, D., Roynette, B. and Yor, M. (2008). Unifying Black-Scholes type formulae which involve Brownian last passage times up to a finite horizon. *Asia-Pacific Finan. Markets* **15**, 97–115.

[37] Marshall, A. W. and Olkin, I. (1993). Maximum likelihood characterizations of distributions. *Statist. Sinica* **3**, 157–171. MR1219297

[38] Massam, H. and Wesołowski, J. (2004). The Matsumoto-Yor property on trees. *Bernoulli* **10**, 685–700. MR2076069

[39] Massam, H. and Wesołowski, J. (2006). The Matsumoto-Yor property and the structure of the Wishart distribution. *J. Multivariate Anal.* **97**, 103–123. MR2208845

[40] Matsumoto, H. and Yor, M. (2001). An analogue of Pitman's $2M - X$ theorem for exponential Wiener functional, Part II: the role of the generalized inverse Gaussian laws. *Nagoya Math. J.* **162**, 65–86. MR1836133

[41] Matsumoto, H. and Yor, M. (2003). Interpretation via Brownian motion of some independence properties between GIG and Gamma variables. *Stat. Probab. Lett.* **61**, 253–259. MR1959132

[42] Mudholkar, S. M. and Tian, L. (2002). An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test. *J. Stat. Plan. Infer.* **102**, 211–221. MR1896483

[43] Peköz, E. (1996). Stein's method for geometric approximation. *J. Appl. Probab.* **33**, 707–713. MR1401468

[44] Perreault, L., Bobée, B. and Rasmussen, P. F. (1999a). Halphen distribution system. I: Mathematical and statistical properties. *J. Hydrol. Eng.* **4**, 189–199.

[45] Perreault, L., Bobée, B. and Rasmussen, P. F. (1999b). Halphen distribution system. II: Parameter and quantile estimation. *J. Hydrol. Eng.* **4**, 200–208.

[46] Poincaré, H. (1912). *Calcul des probabilités.* Carré-Naud, Paris.

[47] Pusz, J. (1997). Regressional characterization of the Generalized inverse Gaussian population. *Ann. Inst. Statist. Math.* **49**, 315–319. MR1463309

[48] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8**, 210–293. MR2861132

[49] Seshadri, V. (1993). *The Inverse Gaussian Distribution, a case study in exponential families.* Oxford Science Publications. MR1306281

[50] Seshadri, V. (1999). *The Inverse Gaussian Distribution.* Springer-Verlag, New York. MR1622488

[51] Seshadri, V. and Wesołowski, J. (2001). Mutual characterizations of the gamma and the generalized inverse Gaussian laws by constancy of regression. *Sankhya Ser. A* **63**, 107–112. MR1898552

[52] Seshadri, V. and Wesołowski, J. (2004). Martingales defined by reciprocals of sums and related characterizations. *Comm. Statist. Theory Methods* **33**, 2993–3007. MR2138669

[53] Shannon, C. E. (1949). *The Mathematical Theory of Communication.* Wiley, New York.

[54] Sichel, H. S. (1974). On a distribution representing sentence-length in written prose. *J. R. Stat. Soc. Ser. A* **137**, 25–34.

[55] Sichel, H. S. (1975). On a distribution law for word frequencies. *J. Am. Stat. Assoc.* **70**, 542–547.

[56] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of*

the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Vol 2, pp. 586–602). Berkeley: University of California Press. MR0402873

[57] STEIN, C., DIACONIS, P., HOLMES, S. and REINERT, G. (2004). Use of exchangeable pairs in the analysis of simulations. In *Persi Diaconis and Susan Holmes, editors, Stein's method: expository lectures and applications, volume 46 of IMS Lecture Notes Monogr. Ser, pages 1–26. Beachwood, Ohio, USA: Institute of Mathematical Statistics.* MR2118600

[58] TEICHER, H. (1961). Maximum likelihood characterization of distributions. *Ann. Math. Statist.* **32**, 1214–1222. MR0130726

[59] VALLOIS, P. (1991). La loi Gaussienne inverse généralisée comme premier ou dernier temps de passage de diffusion. *Bull. Sc. Math.*, 2e Série, **115**, 301–368. MR1117781

[60] VASICEK, O. (1976). A test for normality based on sample entropy. *J. R. Stat. Soc. Ser. B* **38**, 54–59. MR0420958

[61] WESOŁOWSKI, J. (2002). The Matsumoto-Yor independence property for GIG and Gamma laws, revisited. *Math. Proc. Camb. Philos. Soc.* **133**, 153–161. MR1888812

[62] WESOŁOWSKI, J. and WITKOWSKI, P. (2007). Hitting times of Brownian motion and the Matsumoto-Yor property on trees. *Stoch. Proc. Appl.* **117**, 1303–1315. MR2343941