Test d'adéquation pour la loi gaussienne inverse basé sur la propriété de Matsumoto-Yor

Efoevi Angelo Koudou, Severien Nkurunziza

A GOODNESS-OF-FIT TEST FOR THE INVERSE GAUSSIAN DISTRIBUTION BASED ON THE MATSUMOTO-YOR PROPERTY

Angelo Efoévi Koudou & Sévérien Nkurunziza

Institut Elie Cartan, Laboratoire de Mathématiques, B.P. 239, F-54506 Vandoeuvre-lès-Nancy cedex,
University of Windsor, 401 Sunset Avenue, Windsor, Ontario, N9B 3P4.

Abstract

Consider X and Y, two positive and independent random variables. Define the random variables $U = (X + Y)^{-1}$ and $V = X^{-1} - (X + Y)^{-1}$. The Matsumoto-Yor property is the following: the random variables U and V are independent if and only if X follows a generalized inverse Gaussian distribution while Y is gamma-distributed with suitable parameters. We use this property to propose a goodness-of-fit test for the inverse Gaussian distribution.

Résumé :
Soient X et Y des variables aléatoires positives indépendantes. D'après la propriété de Matsumoto-Yor, les variables $U = 1/(X+Y)$ et $V = 1/X - 1/(X+Y)$ sont indépendantes si et seulement si X suit une loi gaussienne inverse généralisée et Y une loi gamma. Nous utilisons cette propriété pour proposer un test d'adéquation pour la loi gaussienne inverse.

Keywords: Gamma distribution; Generalized inverse Gaussian distribution; Matsumoto-Yor property; Kendall's tau; Spearman's rho.

1 Introduction

The generalized inverse Gaussian distribution has been revealed to be more convenient in modeling and statistical data analysis where the observations are highly right-skewed. Indeed, the literature gives several applied fields such as cardiology, demography, finance, surviving bacteria, hydrology, pharmacokinetics, where the inverse Gaussian is the most appropriate in modeling and analyzing the data. To give some references which give illustrative applications of inverse Gaussian, we quote Chhikara and Folks (1989), Seshadri (1993), and Seshadri (1999). However, much development are still to be made in testing procedure that should be useful in deciding whether or not a given data set is from an inverse Gaussian distribution.

In this paper, we suggest a goodness-of-fit test for the inverse Gaussian distribution based on the Matsumoto-Yor property (2001), and a sequence of Kendall tests.

Section 2 gives some properties of generalized inverse Gaussian distributions. In Section 3, we present the suggested goodness-of-fit test. Section 4 is devoted to illustrative examples.

2 The generalized inverse Gaussian distributions

In this section, we recall the definition and some well known properties of the generalized inverse Gaussian (GIG) laws. For more details about these properties, the reader is referred to Barndorff-Nielsen (1994), Matsumoto and Yor (2001).

2.1 Definition

Let $\mu \in \mathbb{R}$, $a > 0$ and $b > 0$. Also, let $\mathbf{1}_A$ denote the indicator function of the event A. The GIG distribution with parameters μ, a, b is the probability measure :

$$GIG(\mu; a, b)(dx) = \left(\frac{b}{a}\right)^{\mu} \frac{x^{\mu-1}}{2K_{\mu}(ab)} e^{-\frac{1}{2}(a^2x^{-1}+b^2x)} \mathbf{1}_{(0,\infty)}(x)dx$$

where K_μ is the classical McDonald special function. One can have $a = 0$ if $\mu > 0$ or $b = 0$ if $\mu < 0$.

In the case $\mu = -\frac{1}{2}$, the law $GIG(\mu; a, b)$ is the classical inverse Gaussian distribution with density

$$IG(a, b)(dx) = \frac{a}{\sqrt{2\pi}} e^{ab} x^{-\frac{3}{2}} e^{-\frac{1}{2}(a^2x^{-1}+b^2x)} \mathbf{1}_{(0,\infty)}(x)dx,$$

while in the case $\mu = \frac{1}{2}$ we have the reciprocal inverse Gaussian distribution with density

$$RIG(a, b)(dx) = \frac{b}{\sqrt{2\pi}} e^{ab} x^{-\frac{1}{2}} e^{-\frac{1}{2}(a^2x^{-1}+b^2x)} \mathbf{1}_{(0,\infty)}(x)dx.$$

Note that $RIG(0, b)$ is the gamma distribution with density

$$\frac{b}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{b^2}{2}x} \mathbf{1}_{(0,\infty)}(x)dx$$

i.e. the gamma distribution with shape parameter 1/2 and scale parameter $2/b^2$.

2.2 Laplace transforms and diffusion interpretation

Laplace transforms are respectively

$$L_{IG(a,b)}(\theta) = e^{ab-a\sqrt{b^2-2\theta}} \quad \text{(with } \theta \leq \tfrac{b^2}{2}\text{)}, \quad \text{and} \quad L_{RIG(a,b)}(\theta) = \frac{b}{\sqrt{b^2-2\theta}} L_{IG(a,b)}(\theta),$$

which can be used to derive the following formulae for the expectations and the variances. If $X \sim IG(a, b)$, then

$$\mathbf{E}(X) = a/b; \quad \mathbf{V}(X) = a/b^3. \tag{2.1}$$

If $X \sim RIG(\delta, \gamma)$, then $\mathbf{E}(X) = (1/b^2) + a/b; \quad \mathbf{V}(X) = (2/b^4) + a/b^3$.

There is a well-known diffusion interpretation for the inverse Gaussian distribution: the distribution of the first hitting time of level $\xi > 0$ for a Brownian motion with drift $\mu \geq 0$ and diffusion coefficient σ^2 is $IG(\delta, \gamma)$ for $\delta = \xi/\sigma$ and $\gamma = \mu/\sigma$. For a proof of this interpretation, see for instance Bhattacharya and Waymire (1990). Further there is a related interpretation of the reciprocal inverse Gaussian distribution that is given in Vallois (1991; p. 302). Specifically, for a Brownian motion as above, with $\mu > 0$, the *last hitting time* of level $\xi \geq 0$ is distributed as $RIG(\delta, \gamma)$ (where, again, $\delta = \xi/\sigma$ and $\gamma = \mu/\sigma$). This may be shown in a simple manner from the first hitting time result combined with the invariance in law of driftless Brownian motion b_t under the transformation $b_t \to t b_{1/t}$.

2.3 The Matsumoto-Yor property

Proposition 2.1 The Matsumoto-Yor property. *Consider two independent non-dirac and positive 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 $\mu < 0$, $a > 0$ and $b > 0$ such that $X \sim$ GIG(μ, a, b) and $Y \sim$ GIG$(-\mu, 0, b)$. Furthermore, $U \sim$ GIG(μ, b, a) and $V \sim$ GIG$(-\mu, 0, a)$.*

The sufficient part proposition was proved by Matsumoto and Yor in the case $a = b$. Letac and Wesolowski (2000) noticed that it is true also if $a \neq b$ and proved that it is in fact a characterization of GIG laws.

3 Independence and goodness-of-fit tests

w We present an algorithm which is used for the proposed goodness-of-fit test.

3.1 Algorithm

Consider a sample $(x_1, \ldots, x_n)$. Suppose one wishes to know if this is a sample from an inverse Gaussian distribution IG(a, b). We propose the following framework :

- By using the moment method, estimate the parameters a and b using the identities (2.1). Thus, the moment estimates of a and b are respectively

$$\widehat{a} = \sqrt{\overline{x}^3/\widehat{\sigma}^2} \quad \text{and} \quad \widehat{b} = \sqrt{\overline{x}/\widehat{\sigma}^2}$$

  with $\overline{x} = (1/n)\sum_{i=1}^n x_i$ and $\widehat{\sigma}^2 = (1/n)\sum_{i=1}^n (x_i - \overline{x})^2$.

- For $j = 1, 2, \ldots, M$ where M is large, draw a sample $\left(y_1^{(j)}, \ldots, y_n^{(j)}\right)$ from the gamma distribution RIG$(0, \widehat{b})$.

- From the samples $(x_1, \ldots, x_n)$ and $\left(y_1^{(j)}, \ldots, y_n^{(j)}\right)$, for each $j = 1, 2, \ldots, M$ construct the samples
$\left(u_1^{(j)}, \ldots, u_n^{(j)}\right)$ and $\left(v_1^{(j)}, \ldots, v_n^{(j)}\right)$ where for each $j = 1, 2, \ldots, M$, for $i = 1, \ldots, n$

$$u_i^{(j)} = 1 \Big/ \left(x_i + y_i^{(j)}\right), \;\; v_i^{(j)} = (1/x_i) - 1 \Big/ \left(x_i + y_i^{(j)}\right).$$

- Compute appropriate test statistic in order to perform independence tests between the samples $\left(u_1^{(j)}, \ldots, u_n^{(j)}\right)$ and $\left(v_1^{(j)}, \ldots, v_n^{(j)}\right)$, for each $j = 1, 2, \ldots, M$. In this paper, the independence test is performed by using nonparametric methods. In particular, we use Kendall's test although in data analysis we present also the result of Spearman's test for comparison purposes. Namely, let $\{\tau_n^{(j)}\}$, $j = 1, 2, \ldots$ be a sequence of Kendall's tau corresponding to $\left(u_1^{(j)}, \ldots, u_n^{(j)}\right)$ and $\left(v_1^{(j)}, \ldots, v_n^{(j)}\right)$, i. e.

$$\tau_n^{(j)} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{l=1}^{n-1} \text{sign}\left(u_i^{(j)} - u_l^{(j)}\right) \text{sign}\left(v_i^{(j)} - v_l^{(j)}\right), \tag{3.1}$$

  where $\text{sign}(x) = 1$ if $x > 0, 0$ if $x = 0, -1$ if $x < 0$. It is noticed that, here $\text{sign}(x)$ takes values 1 and -1 almost surely since the probability of ties is 0. Also, let the test statistic

$$\tau_n(M) = \sum_{j=1}^{M} \tau_n^{(j)}/M. \tag{3.2}$$

  Then we reject the null hypothesis for a large value of $|\tau_n(M)|$, which is made more precise below.

  Finally, according to the Matsumoto-Yor property, the rejection of the null hypothesis of independence of the variables U and V is equivalent to the rejection of the hypothesis that the sample $(x_1, \ldots, x_n)$ is drawn from the distribution $\text{IG}(a, b)$.

As well-known, we have for each $j = 1, \ldots, M$, and $\tau_n^{(j)}$ defined by (3.1),

$$\mathbf{E}(\tau_n^{(j)}) = 0, \quad \text{and} \quad \mathbf{V}(\tau_n^{(j)}) = 2(2n+5)/\left(9n(n-1)\right),$$

so that, for the test statistic $\tau_n(M)$ defined by (3.2),

$$\mathbf{V}(\tau_n(M)) = 2(2n+5)/\left(9Mn(n-1)\right).$$

Thus, under the null hypothesis, as n and M tend to infinity, the asymptotic law of $\frac{3}{2}\sqrt{Mn}\tau_n(M)$ is the standard normal distribution. Then, at a significance level α, we reject the null hypothesis if

$$\frac{3}{2}\sqrt{Mn}|\tau_n(M)| > z_{\alpha/2},$$

where $z_{\alpha/2}$ is such that $P(Z > z_{\alpha/2}) = \alpha/2$ for a standard normal variable Z.

4 Illustrative examples

In this section, we illustrate the application of the suggested procedure through four real data sets. The four data sets have been used by Krishnamoorthy and Tian (2008) in order to illustrate their statistical procedure.

In particular, the two first data sets represent the shelf life in days of two products as reported in Gacula and Kubaba (1975). Table 1 gives the observed values of the data set as presented in Krishnamoorthy and Tian (2008).

Table 1: The shelf life in days of two products

Product M	24	24	26	32	32	33	33	33	35	41	42	43	47	48	48	48
	50	52	54	55	57	57	57	61								
Product K	21	23	25	38	43	43	52	56	61	63	67	69	70	75	85	107

The two last data sets are reported in Shapiro *et al.* (1986) and these were obtained from a study of lymphocyte abnormalities in group of patients in remission from Hodgkin's desease as described in Shapiro *et al.* (1986). Namely, Hodgkin's disease group and Non-Hodgkin's disease group were considered and 20 observations were recorded in each group. The observations represent the number of T_4 cells per mm^2 in the patient's blood and these are given in Table 2.

Table 2: Number of T_4 cells per mm^3 in the patients' blood

Hodgkin's disease	396	567	1212	171	554	1104	257	435	295	397
	288	1004	431	795	1621	1378	902	958	1283	2415
Non-Hodgkin's disease	375	375	752	208	151	116	736	192	315	1252
	675	700	440	771	688	426	410	979	377	503

Table 3 presents inference results based on Kendall's rank correlation tau. In particular, Table 3 shows that each of the four data sets given in Table 1 (Products M and K data sets) and Table 2 (Hodgkin's disease and Non-Hodgkin's disease data sets) is generated by a generalized inverse Gaussian with a significance level of 5% (p-value>0.49). Also, Table 4 shows that Spearman's rank correlation rho leads to similar conclusion as Kendall's rank correlation tau. Namely, based on Table 4, we conclude that each of the four data sets has been collected from a generalized inverse Gaussian population, with significance level of 5 % (p-value>0.48).

Table 3: Kendall's rank correlation tau

	Product M	Product K	Hodgkin's disease	Non-Hodgkin's disease
Statistic	155.213	65.285	104.521	104.139
P-value	0.5080697	0.5097817	0.4973657	0.4953746
Kendall's τ	0.03475333	0.08808333	0.1002211	0.0962

Table 4: Spearman's rank correlation rho

	Product M	Product K	Hodgkin's disease	Non-Hodgkin's disease
Statistic	2463.286	598.836	1146.694	1150.246
P-value	0.4974664	0.4843884	0.4875552	0.4837759
Spearman's ρ	0.05258231	0.1193588	0.1378241	0.1351534

References

[1] Barndorff-Nielsen, O. E. (1994). A note on electrical networks and the inverse Gaussian distribution. *Advanced Applied Probability*, **26**, 63–67.

[2] Bhattacharya and Waymire (1990). *Stochastic Processes with Applications*. Wiley, New York.

[3] Chhikara, R. S., and Folks, J. L. (1989). *The Inverse Gaussian Distribution*. Marcel Dekker, New York.

[4] Gacula, J., M. C., and Kubaba, J. J. (1975). Statistical models for shelf life failures. *J. Food Sc.* **40**, 404–409.

[5] Krishnamoorthy, K. and Tian, L. (2008). Inferences on the difference and ratio of the means of two inverse Gaussian distributions. *Journal of Statistical Planning and Inference* **138**, 2082-2089.

[6] 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.

[7] Letac, G. and Wesolowski, J. (2000). An independence property for the product of GIG and gamma laws. *Annals of Probability* **28**, 1371-1383.

[8] Seshadri, V. (1993). *The Inverse Gaussian Distribution: A Case Study in Exponential Families*. Clarendon Press, Oxford.

[9] Seshadri, V. (1999). *The Inverse Gaussian Distribution: Statistical Theory and Applications*. Springer, New York.

[10] Shapiro, C. M., Beckmann, E., and Christiansen, N. (1986). Identifying men at high risk of heart attacks:strategy for use in general practice. *Amer. J. Medical Sci.*, **293**, 366-370.