Test of goodness of fit for the inverse-gaussian distribution

VICTOR MERGEL*

Abstract. *Entropy based test of goodness of fit is proposed for inverse-gaussian distribution.*

Key words: *entropy, statistical estimate, inverse-gaussian distribution, test of goodness of fit*

AMS subject classifications: 60G07, 60G17

Received January 18, 1999 Accepted May 6, 1999

1. Introduction

Denote by $\mathbb{R}^1$ a real line with the distance $\rho(x,y) = |x-y|$ between $x \in \mathbb{R}^1$ and $y \in \mathbb{R}^1$. Then let $v(y,r) = \{x \in \mathbb{R}^1 : \rho(x,y) < r\}$.

Consider now the random variable ξ with unknown density function $f(x)$, $x \in \mathbb{R}^1$. Denote by $supp(f) = \{x \in \mathbb{R}^1 : f(x) > 0\}$. The problem is to estimate the entropy

$$H = -\int_{\mathbb{R}^1} f(x) \ln f(x) dx < \infty \tag{1}$$

based on the independent identically distributed sample $X_1, \ldots, X_N$, $N \geq 2$, of random variable ξ.

For a fixed observation X_i, $i \in \{1,...,N\}$, and fixed $k \in \{1,...,N-1\}$, we define random variables $\rho_{i,k}$ as follows:

$\rho_{i,1} := \min\{\rho(X_i, X_j), j \in \{1,...,N\} \setminus \{i\}\} = \rho(X_i, X_{j_1})$,

$\rho_{i,2} := \min\{\rho(X_i, X_j), j \in \{1,...,N\} \setminus \{i, j_1\}\} = \rho(X_i, X_{j_2})$,

$\vdots$

$\rho_{i,k} := \min\{\rho(X_i, X_j), j \in \{1,...,N\} \setminus \{i, j_1, \ldots, j_{k-1}\}\} = \rho(X_i, X_{j_k})$,

$\vdots$

$\rho_{i,N-1} := \max\{\rho(X_i, X_j), j \in \{1,...,N\} \setminus \{i\}\} = \rho\left(X_i, X_{j_{N-1}}\right)$, where $X_{j_{k-1}}$ is a vector such that with probability 1:

$$\rho\left(X_i, X_{j_{k-2}}\right) < \rho\left(X_i, X_{j_{k-1}}\right) < \rho\left(X_i, X_{j_k}\right)$$

*Department of Mathematics, Taras Shevchenko University of Kyiv, Vladimirskaya 64, 252 601 Kyiv, Ukraine, e-mail: `VIK@g-auto.pul.kiev.ua`

Consequently with probability 1, we have

$$\rho\left(X_i, X_{j_1}\right) < \rho\left(X_i, X_{j_2}\right) < \ldots < \rho\left(X_i, X_{j_{N-1}}\right).$$

For a fixed $k \in \{1, \ldots, N-1\}$, we let

$$\bar{\rho}_k = \{\prod_{i=1}^{N} \rho_{i,k}\}^{1/N} \tag{2}$$

denote the geometric mean of random variables $\rho_{1,k}, \ldots, \rho_{N,k}$. We define the statistical estimate of (1)

$$H_{k,N} = \ln \bar{\rho}_k + \ln 2\gamma(N-1),$$

where $k \in \{1, 2, \ldots, N-1\}$ and $\gamma = \exp\{-\int_0^\infty e^{-t} \ln t dt\}$ is an Euler constant. The following Theorems are proved in [5]

Theorem 1. *Let $k \in \{1, \ldots, N-1\}$ be fixed and suppose that there exists an $\varepsilon > 0$ such that*

$$\int_{\mathbb{R}^1} |\ln f(x)|^{1+\varepsilon} f(x) dx < \infty. \tag{3}$$

Then

$$\lim_{N\to\infty} EH_{k,N} = H,$$

where entropy H is defined in (2)*.*

Theorem 2. *Let $k, \varepsilon > 0$ as in* Theorem 1 *and*

$$\int_{\mathbb{R}^1} |\ln f(x)|^{2+\varepsilon} f(x) dx < \infty. \tag{4}$$

Then the entropy estimate $H_{k,N}$ given in (2), is a consistent estimate of H, as $N \to \infty$.

Note that the estimate $H_{1,N}$ was considered by Kozachenko and Leonenko (see [9]), but our conditions are weaker than those proposed in [9] even in the case $k = 1$.

2. Maximum entropy principle for the inverse - gaussian distribution

We present first the following:

Theorem 3. *Let $\mathcal{K}$ be a class of probability density functions $f(x)$, $x \in \mathbb{R}^1$ with supp$\{f\} = (0, \infty)$ which satisfy conditions* (3) *and* (4)*. Note that the density*

$$f^*(x) = \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left\{-\frac{\lambda}{2m^2}\frac{(x-m)^2}{x}\right\} I_{(0,\infty)}(x) \tag{5}$$

of an inverse-gaussian distribution $IG(m, \lambda)$ belongs to the class $\mathcal{K}$. Then among all densities f from class $\mathcal{K}$ such that

$$\int_0^\infty x f(x) dx = m \tag{6}$$

$$\int_0^\infty \log x f(x)dx = 2\frac{\lambda}{m}e^{\left(\frac{\lambda}{m}-\frac{1}{\lambda}\right)}\log m + 2e^{\frac{\lambda}{m}}\frac{\sqrt{\lambda}}{m\sqrt{2\pi}}\left(K_\delta\left(\frac{1}{\lambda}\right)\right)'_\delta\Big|_{\delta=-\frac{1}{2}} \tag{7}$$

$$\int_0^\infty \frac{f(x)}{x}dx = \frac{1}{m}+\frac{1}{\lambda} \tag{8}$$

the entropy is maximized by the density of the inverse-gaussian distribution ($K_\delta(\lambda)$ is a modified Bessel function).

Proof. To prove this theorem we use the Maximum-Entropy Principle (see [8]). The Lagrangian in this case will be:

$$\begin{aligned} L = & -\int_0^\infty f(x)\log f(x)dx - \alpha\left(\int_0^\infty f(x)dx - 1\right) - \mu_1\left(\int_0^\infty xf(x)dx - m\right) \\ & -\mu_2\left(\int_0^\infty \log x f(x)dx - \left(2\frac{\lambda}{m}e^{\left(\frac{\lambda}{m}-\frac{1}{\lambda}\right)}\log m + 2e^{\frac{\lambda}{m}}\frac{\sqrt{\lambda}}{m\sqrt{2\pi}}\left(K_\delta\left(\tfrac{1}{\lambda}\right)\right)'_\delta\Big|_{\delta=-\frac{1}{2}}\right)\right) \\ & -\mu_3\left(\int_0^\infty \frac{f(x)}{x}dx - \left(\frac{1}{m}+\frac{1}{\lambda}\right)\right). \end{aligned}$$

Then

$$\frac{\partial L}{\partial f} = -1 - \log f(x) - \alpha - \mu_1 x - \mu_2\log x - \frac{\mu_3}{x} = 0$$

or

$$f(x) = A\exp\left\{-\mu_1 x - \mu_2\log x - \frac{\mu_3}{x}\right\}.$$

To determie A, μ_1, μ_2, μ_3 we use $(6), (7), (8)$ and $\int_0^\infty f(x)dx = 1$. So

$$m = A\int_0^\infty \exp\left\{-\mu_1 x - (\mu_2 - 1)\log x - \frac{\mu_3}{x}\right\},$$

$$\begin{aligned} & 2\frac{\lambda}{m}e^{\left(\frac{\lambda}{m}-\frac{1}{\lambda}\right)}\log m + 2e^{\frac{\lambda}{m}}\frac{\sqrt{\lambda}}{m\sqrt{2\pi}}\left(K_\delta\left(\frac{1}{\lambda}\right)\right)'_\delta\Big|_{\delta=-\frac{1}{2}} \\ & \qquad = A\int_0^\infty \log x\exp\left\{-\mu_1 x - \mu_2\log x - \frac{\mu_3}{x}\right\}, \end{aligned}$$

$$\frac{1}{m}+\frac{1}{\lambda} = A\int_0^\infty \exp\left\{-\mu_1 x - (\mu_2+1)\log x - \frac{\mu_3}{x}\right\}$$

and

$$1 = A\int_0^\infty \exp\left\{-\mu_1 x - \mu_2\log x - \frac{\mu_3}{x}\right\}.$$

From here it is easy to get that: $\mu_1 = \frac{\lambda}{2m^2}$, $\mu_2 = \frac{3}{2}$, $\mu_3 = \frac{\lambda}{2}$, $A = \frac{\sqrt{\lambda}}{\sqrt{2\pi}} \exp\left\{\frac{\lambda}{m}\right\}$ and

$$\begin{aligned} f^*(x) = & \frac{\sqrt{\lambda}}{\sqrt{2\pi}} \exp\left\{\frac{\lambda}{m}\right\} \exp\left\{-\frac{3}{2}\log x - \frac{\lambda}{2m^2}x - \frac{\lambda}{2x}\right\} I_{(0,\infty)}(x) \\ = & \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left\{-\frac{\lambda}{2m^2}\frac{(x-m)^2}{x}\right\} I_{(0,\infty)}(x) \end{aligned}$$

that is

$$\begin{aligned} H(f) \le \; & H(f^*) = -\int_0^\infty f(x)\log f(x)dx \\ = \; & -\int_0^\infty \frac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left\{-\frac{\lambda}{2m^2}\frac{(x-m)^2}{x}\right\} \left(\frac{1}{2}\log\frac{\lambda}{2\pi x^3} - \frac{\lambda}{2m^2}\frac{(x-m)^2}{x}\right) dx \\ = \; & -\int_0^\infty \frac{1}{2} f^*(x) \log\frac{\lambda}{2\pi}dx + \frac{3}{2}\int_0^\infty f^*(x)\log x dx + \frac{\lambda}{2m^2}\int_0^\infty f^*(x) x dx \\ & -\frac{\lambda}{m}\int_0^\infty f^*(x)dx + \frac{\lambda}{2}\int_0^\infty \frac{f^*(x)}{x}dx = \frac{1}{2}\log\frac{2\pi}{\lambda} + \frac{1}{2} \\ & +3\frac{\lambda}{m}e^{\left(\frac{\lambda}{m}-\frac{1}{\lambda}\right)}\log m + 3e^{\frac{\lambda}{m}} \frac{\sqrt{\lambda}}{m\sqrt{2\pi}} \left(K_\delta\left(\frac{1}{\lambda}\right)\right)'_\delta\Big|_{\delta=-\frac{1}{2}} \end{aligned}$$

□

3. Tests of goodness of fit for the inverse-gaussian distribution

We propose the entropy based test of goodness of fit for the so-called inverse-gaussian distribution (see, for example [10]). This distribution is important for both turbulence theory and finance (see, for example [1] and [2]).

The main idea for the construction of tests of goodness of fit is based on the maximum entropy principle (see [8]). Consider a class of densities satisfying certain restrictions. Find a consistent estimator of entropy for the members of the class. Next, using the so-called maximun entropy principle (see, for example [8] or for more recent developments [4]),determine a member of the class maximizing entropy and find its parametric consistent estimator. Finally, take a function of the above estimators as a test statistics of goodness fit for the member maximizing the entropy.

Tests of goodness of fit based on the sample entropy was proposed by Vasicek (see [11]) for one-dimensional normal distribution and by Dudewicz and Van der Meuler (see [3]) for uniform distribution, among the others. The essential difference between the tests proposed by these authors and ours lies in the choice of entropy estimator. In our work we have broaded this list of entropy-based tests, proposed by Goria, Leonenko and Mergel (see [5]).

Let $X_1, X_2, \dots X_N$, $N \ge 2$, be an independent sample from a member of $\mathcal{K}$. Let $\bar{X}_N$, S_N^2 denote the sample mean and variance, respectively, then

$$\hat{m}_N = \bar{X}_N$$

$$\left(\hat{\frac{1}{\lambda}}\right)_N = \frac{S_N^2}{\hat{m}_N^3}$$

are consistent estimators of parameters m, λ, respectively.

Under the null hypothesis $H_0 : X_1, X_2, \ldots X_N$ are a sample from the inverse-gaussian distribution $IG(m, \lambda)$, we have that for each fixed $k \in \{1, \ldots, N-1\}$

$$\begin{aligned} \zeta_N = \quad & \log\left(\sqrt{2\pi e}\frac{S_N}{\sqrt{\hat{m}_N^3}}\right) + \frac{3\hat{m}_N^2 \log \hat{m}_N}{S_N^2} \exp\left\{\frac{\hat{m}_N^2}{S_N^2} - \frac{S_N^2}{\hat{m}_N^3}\right\} \\ & + \frac{3\sqrt{\hat{m}_N}}{S_N\sqrt{2\pi}} \exp\left\{\frac{\hat{m}_N^2}{S_N^2}\right\} \left(K_\delta\left(\frac{S_N^2}{\hat{m}_N^3}\right)\right)_\delta' \Bigg|_{\delta=-\frac{1}{2}} - H_{k,N} \to 0 \end{aligned}$$

in probability $N \to \infty$. While under the alternative $H_1 : X_1, X_2, \ldots X_N$ are a sample from any other distribution from class $\mathcal{K}$ with finite second order moment ,we obtain that $\zeta_N \to \mathit{const} \neq 0$ as $N \to \infty$ in probability for every fixed $k \in \{1, 2, \ldots, N-1\}$. This means that this test is consistent for such alternatives.

References

[1] D. E. BARNDORFF-NIELSEN, *Hyperbolic distributions and distributions on hyperbolas*, Scan. J. Statist. **5**(1978), 151–157.

[2] D. E. BARNDORFF-NIELSEN, *Models for non-Gaussian variation: with applications to turbulence*, Proc. R. Soc. London **A368**, 501–520.

[3] E. I. DUDEWICZ, E. C. VAN DER MEULER, *Entropy based test of uniformity*, JASA **76**(1981), 974–987.

[4] M. N. FRONTINI, A. TAGLIANI, *Entropy-Convergence in Stieltjes and hamburger moment problem*, Appl. Math. Comput. **88**(1997), 39–51.

[5] M. N. GORIA, N. N. LEONENKO, V. V.MERGEL, *Statistical Estimators of Entropy of Random Vector and its Applications in Testing of Statistical Hypotheses*, Bernoulli, submitted.

[6] D. V. ILIESKU, V. G. VODA, *On the inverse-gaussian distribution*, Bull. Math. de la Soc. Sci. Math. de la R. S. de Roumanic **25**(**73**)(1981), 1–15.

[7] A. M. KAGAN, YU. V. LINNIK, C. B. RAO, *Characterization Problems in Mathematical Statistics*, John Wiley, New York, 1973.

[8] J. N. KAPUR, *Maximum Entropy Models in Science and Engineering*, John Wiley, New York, 1989.

[9] L. F. KOSACHENKO, N. N. LEONENKO, *On statistical estimation of entropy of random vector*, Prob. Inf. Trans. **23**(1987), 9–16 (in Russian).

[10] V. SESHARDI, *The Inverse-Gaussian Distribution*, Clarendon Press, Oxford, 1993.

[11] O. VASICEK, *A test of normality based on sample entropy*, J. Roy. Stat.Soc., Ser. B **38**(1976), 254–256.