The Korean Communications in Statistics
Vol. 13 No. 3, 2006, pp.657-667

Minimum Variance Unbiased Estimation for the Maximum Entropy of the Transformed Inverse Gaussian Random Variable by $Y=X^{-1/2}$

Byungjin Choi[1)]

Abstract

The concept of entropy, introduced in communication theory by Shannon (1948) as a measure of uncertainty, is of prime interest in information-theoretic statistics. This paper considers the minimum variance unbiased estimation for the maximum entropy of the transformed inverse Gaussian random variable by $Y=X^{-1/2}$. The properties of the derived UMVU estimator is investigated.

Keywords : Entropy; inverse Gaussian distribution; minimum variance unbiased estimator; asymptotic distribution.

1. Introduction

The concept of entropy, introduced in communication theory by Shannon (1948) as a measure of uncertainty, has been of prime interest in information-theoretic statistics. The Shannon entropy of a continuous random variable X with a density function $f(x)$ is given by

$$H(f)=-\int_{-\infty}^{\infty} f(x)\log f(x)\,dx. \tag{1.1}$$

Attempts have been made to extend and generalize the Shannon entropy and as a result, various types of entropy measures developed can be found in the literature; for instance, see Havrda and Charvat (1967), Burbea and Rao (1982), and Kapur and Kesavan (1992).

Jayens (1957) was the first to use the Shannon entropy in the domain of statistical inference. The maximum entropy(ME) principle formulated by Jayens (1957) is a method of deriving the probability distribution that maximizes the Shannon entropy subject to given constraints, and it can be stated as follows. Let a continuous random variable X take a density function $f(x)$ defined on the

1) Assistant Professor, Department of Applied Information Statistics, Kyonggi University, San 94-6, Iui-Dong, Yeongtong-Gu, Suwon, Gyeonggi-Do 443-760, Korea.
E-mail : bjchoi92@kyonggi.ac.kr

interval $[a, b]$, $-\infty \le a < b \le \infty$. Given m independent constraints

$$E_f\{T_i(X)\}= \int_a^b T_i(x) f(x)\, dx = \theta_i,\ i = 1, \cdots, m, \tag{1.2}$$

where $T_i(x)$'s are absolutely integrable functions with respect to $f(x)$, the probability distribution with maximum entropy, if it exists, takes the density function given by

$$f^*(x)= \exp\{-\lambda_0 - \lambda_1 T_1(x)- \lambda_2 T_2(x)- \cdots - \lambda_m T_m(x)\}, \tag{1.3}$$

where $\lambda_0, \lambda_1, \cdots, \lambda_m$ are Lagrange multipliers and can be determined from the constraints (1.2). The ME principle has quite broad range of applications. One example is that most of well-known distributions in the field of statistics are characterized as the ME distribution subject to certain constraints; see Kapur & Kesavan (1992) and Ahmed & Gokhale (1989) for a detailed list of specific constraints and the corresponding ME distributions.

The inverse Gaussian distribution, abbreviated $IG(\mu, \lambda)$, with the probability density function

$$f(x; \mu, \lambda)= \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0, \mu > 0, \lambda > 0 \tag{1.4}$$

does not fall into this category, as stated in Mudholkar and Tian (2002). The technical details, not provided in Mudholkar and Tian (2002), can be described as follows. Following the manner of Singh (1998), the expectation of $-\log f(x; \mu, \lambda)$ yields the constraints for (1.4) as

$$E_f(\log X)= \int_0^\infty \log x\, f(x; \mu, \lambda)\, dx,$$

$$E_f(X)= \int_0^\infty x\, f(x; \mu, \lambda)\, dx,$$

$$E_f(X^{-1})= \int_0^\infty x^{-1} f(x; \mu, \lambda)\, dx.$$

The density function with maximum entropy under the constraints is given by

$$f^*(x)= \exp(-\lambda_0 - \lambda_1 \log x - \lambda_2 x - \lambda_3 x^{-1}). \tag{1.5}$$

From the fact that

$$\int_0^\infty f^*(x)\, dx = \int_0^\infty \exp(-\lambda_0 - \lambda_1 \log x - \lambda_2 x - \lambda_3 x^{-1})\, dx = 1, \tag{1.6}$$

it can be obtained that

$$e^{\lambda_0} = \int_0^\infty x^{-\lambda_1} \exp(-\lambda_2 x - \lambda_3 x^{-1})\, dx = 2(\lambda_3/\lambda_2)^{(1-\lambda_1)/2} K_{1-\lambda_1}(2\sqrt{\lambda_2 \lambda_3}), \tag{1.7}$$

where K_ν denotes the modified Bessel function of the third kind with index ν . Thus, λ_0 is obtained as

$$\lambda_0 = \log\left\{2(\lambda_3/\lambda_2)^{(1-\lambda_1)/2} K_{1-\lambda_1}\left(2\sqrt{\lambda_2\lambda_3}\right)\right\}. \quad (1.8)$$

From (1.6), λ_0 is also found as

$$\lambda_0 = \log\int_0^\infty x^{-\lambda_1}\exp\left(-\lambda_2 x - \lambda_3 x^{-1}\right)dx. \quad (1.9)$$

To determine λ_1, λ_2 and λ_3, each of (1.8) and (1.9) is differentiated with respect to λ_1, λ_2 and λ_3 and then the obtained equations are solved simultaneously.

However, it is a technically difficult task. For this reason, Mudholkar and Tian (2002) introduced an entropy characterization of the $IG(\mu,\lambda)$ distribution using a different approach. Mudholkar and Tian (2002) also presented a consistent estimator of the entropy, formed by replacing a parameter with its consistent estimator, and used it to construct a test of fit for the $IG(\mu,\lambda)$ distribution. However, the derivation of the UMVU estimator of the entropy was not considered.

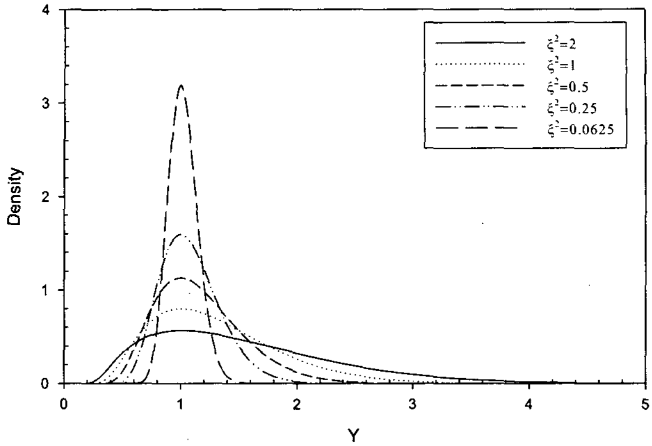
This paper deals with the minimum variance unbiased estimation for the maximum entropy of the transformed inverse Gaussian random variable by $Y = X^{-1/2}$. In section 2, the UMVU estimator of the entropy is derived and its variance is given. In section 3, the distributional behaviour of the derived UMVU estimator is investigated. The brief conclusions are made in section 4.

2. Minimum variance unbiased estimation

Let X be an inverse Gaussian random variable with the density function (1.4). Following Mudholkar and Tian (2002), the transformed inverse Gaussian random variable by $Y = X^{-1/2}$ takes the density function

$$g(y;\nu,\xi^2) = \left(\frac{2}{\pi\xi^2}\right)^{1/2}\exp\left\{-\frac{(y^2-\nu)^2}{2\xi^2 y^2}\right\}, \quad y \geq 0,\ \nu > 0,\ \xi^2 > 0, \quad (2.1)$$

where $\nu = 1/\mu$ and $\xi^2 = 1/\lambda$. Several density curves of Y are depicted in <Figure 1> when X is distributed as $IG(\mu,\lambda)$ with $\mu = 1$ for the values of $\lambda = 0.5, 1, 2, 4, 16$. The shape of the density $g(y;\nu,\xi^2)$ is highly skewed when λ is small(i.e. ξ^2 is large), whereas it is nearly symmetric when λ is large(i.e. ξ^2 is small).



<Figure 1> Probability density curves of Y with $\nu = 1$ for five values of ξ^2

Mudholkar and Tian (2002) introduced an entropy characterization of the inverse Gaussian distribution on the basis of the density function (2.1). The following is the characterization result.

Theorem 2.1 [Mudholkar and Tian (2002)]. The random variable X following the inverse Gaussian distribution is characterized by the property that $Y = X^{-1/2}$ attains maximum entropy among all nonnegative, absolutely continuous random variables Y subject to constraints $E(Y^{-2}) = 1/\nu$ and $E(Y^2) = \nu + \xi^2$.

If X is distributed as $IG(\mu, \lambda)$, then the maximum entropy of Y is given by $H(g) = \log(\xi^2 \pi e/2)/2$, where $\xi^2 = E(Y^2) - 1/E(Y^{-2})$. The entropy $H(g)$ depends on only the parameter ξ^2 and is very similar to that of the normal distribution with variance σ^2.

The UMVU estimator of the entropy $H(g)$ is now derived. Let a random sample of size n, $X_1, X_2, \cdots, X_n$, be drawn from an inverse Gaussian distribution. Then a transformed sample $Y_i = X_i^{-1/2}$, $i = 1, \cdots, n$, has the density function (2.1). It is easily seen that two statistics, $\sum_{i=1}^{n} Y_i^2$ and $\sum_{i=1}^{n} Y_i^{-2}$, are jointly complete and sufficient because the density function (2.1) is a member of exponential family. To obtain the UMVU estimator based on these statistics, the following lemma is used.

Lemma 2.1 Let Z be a chi-squared random variable with m degrees of freedom. Then, the mean of $Z^* = \log Z$ is given by

$$E(Z^*) = \log 2 + \psi\left(\frac{m}{2}\right), \tag{2.2}$$

where ψ is the digamma function defined by $\psi(k) = d\log\Gamma(k)/dk$.

Proof. Since the moment generating function of Z^* is given by

$$m_{Z^*}(t) = E\left(e^{t\log Z}\right) = E\left(Z^t\right) = 2^t\Gamma\left(\frac{m}{2}+t\right)/\Gamma\left(\frac{m}{2}\right), \tag{2.3}$$

the result is immediately obtained by differentiating $m_{Z^*}(t)$ with respect to t and setting $t=0$. □

Theorem 2.2 The uniformly minimum variance unbiased estimator of the entropy $H(g)$ is given by

$$\widetilde{H}(g) = \frac{1}{2}\log V_W + \frac{1}{2}\log\frac{\pi e}{4} - \frac{1}{2}\psi\left(\frac{n-1}{2}\right), \tag{2.4}$$

where $W_i = Y_i^2$ $(i=1,\cdots,n)$, $V_W = \sum_{i=1}^{n}\left(W_i - \overline{W_H}\right)$, $\overline{W_H} = n/\sum_{i=1}^{n} W_i^{-1}$ and ψ is the digamma function.

Proof. Following Chhikara and Folks (1989), the distribution of λV_X is the chi-squared with $n-1$ degrees of freedom, where $V_X = \sum_{i=1}^{n}\left(1/X_i - 1/\overline{X}\right)$. Two statistics V_W and V_X are equivalent and thus, it follows that $V_W/\xi^2 \sim \chi^2_{n-1}$. Applying the lemma 2.1 gives the expected value of $\log V_W$ as

$$E(\log V_W) = \log\xi^2 + \log 2 + \psi\left(\frac{n-1}{2}\right). \tag{2.5}$$

Taking $\phi(V_W) = \{\log V_W + \log(\pi e/4) - \psi((n-1)/2)\}/2$ as an estimator of $H(g)$ and using the result (2.5) give

$$E\{\phi(V_W)\} = \frac{1}{2}\log\left(\frac{\xi^2\pi e}{2}\right). \tag{2.6}$$

Thus, $\phi(V_W)$ is an unbiased estimator of $H(g)$. Since V_W is complete and sufficient, $\phi(V_W)$ is the uniformly minimum variance unbiased estimator of $H(g)$ by the Lehmann-Scheffe theorem(for instance, see Hogg, McKean and Craig (2005), p. 387). This completes the proof. □

Lemma 2.2 The variance of $\widetilde{H}(g)$ is given by

$$Var\{\widetilde{H}(g)\} = \frac{1}{4}\psi'\left(\frac{n-1}{2}\right), \tag{2.7}$$

where ψ' is the trigamma function defined by $\psi'(k) = d^2 \log\Gamma(k)/dk^2$.

Proof. Since V_W/ξ^2 is distributed as the chi-squared with $n-1$ degrees of freedom, the moment generating function of $\log(V_W/\xi^2)$ is obtainable by replacing m with $n-1$ in (2.3). Differentiating $m_{Z^*}(t)$ twice with respect to t and setting $t=0$ give $Var\{\log(V_W/\xi^2)\} = \psi'\{(n-1)/2\}$. The result immediately follows from the fact that $Var\{\log(V_W/\xi^2)\} = Var(\log V_W)$ and $Var\{\widetilde{H}(g)\} = Var(\log V_W)/4$. □

Lemma 2.3 The variance of $\widetilde{H}(g)$ given in (2.7) converges to zero as n tends to infinity.

Proof. By the series expansion of the trigamma function(see, Abramowitz and Stegun (1970), p. 260), the variance of $\widetilde{H}(g)$ is expressible as $\sum_{k=0}^{\infty}(n+2k-1)^{-2}$.

Since the infinite series $\sum_{k=0}^{\infty}(n+2k-1)^{-2}$ is equal to $(n+1)^2/\{4n(n-1)^2\}$ and the term $(n+1)^2/\{4n(n-1)^2\}$ converges to zero as $n\to\infty$, the result follows. □

3. Distributional behaviour of the UMVU estimator

It is known that a logarithmic transformed chi-squared random variable is well approximated by the normal distribution with proper mean and variance(see Olshen (1937)). The estimator $\widetilde{H}(g)$ is a function of a logarithmic transformed chi-squared random variable and thus, it is expected to be also well approximated by the normal distribution with mean $H(g)$ and variance given by (2.7).

Denote $\log(V_W/\xi^2)$ by V_W^*. Using the distributional result of V_W/ξ^2 and applying the delta method, it can be shown that V_W^* is asymptotically normally distributed with mean $\mu^* = \log(n-1)$ and variance $\sigma^{2^*} = 2/(n-1)$. The estimator $\widetilde{H}(g)$ is expressed in terms of V_W^* as

$$\begin{aligned}\widetilde{H}(g) &= \frac{1}{2}\log V_W + \frac{1}{2}\log\frac{\pi e}{4} - \frac{1}{2}\psi\left(\frac{n-1}{2}\right)\\ &= \frac{1}{2}\log\frac{V_W}{\xi^2} + H(g) - \frac{1}{2}\psi\left(\frac{n-1}{2}\right) - \frac{1}{2}\log 2\end{aligned}$$

$$= \frac{1}{2} V_W^* + H(g) - C(n), \qquad (3.1)$$

where $C(n) = [\psi\{(n-1)\}/2 + \log 2]/2$. Using the distributional result of V_W^* and applying the delta method, it can be shown that $\widetilde{H}(g)$ is distributed as the normal with mean $\widetilde{\mu} = H(g) + \log(n-1)/2 - C(n)$ and variance $\widetilde{\sigma}^2 = 1/\{2(n-1)\}$ as n increases. However, $\widetilde{\mu}$ approximates to $H(g)$ because $\log(n-1)/2 - C(n)$ is close to zero for large n (see Gradshteyn and Ryzhik (2000), p. 894). Also using the infinite series representation of the trigamma function(see Abramowitz and Stegun (1970), p. 260) to obtain the approximate form of $\psi'\{(n-1)/2\}$ yields $\psi'\{(n-1)/2\} = 2/(n-1) + R_{n-1}$, where R_{n-1} is the remainder term. By observ-ing that R_{n-1} vanishes with order $1/(n-1)^2$, the variance of $\widetilde{H}(g)$, for large n, can be approximated by the first term as $\psi'\{(n-1)/2\}/4 \approx 1/\{2(n-1)\}$, which is equivalent to $\widetilde{\sigma}^2$. Based on these results, it can be seen that $\widetilde{H}(g)$ is approximated by the normal distribution with mean $H(g)$ and variance $\psi'\{(n-1)/2\}/4$.

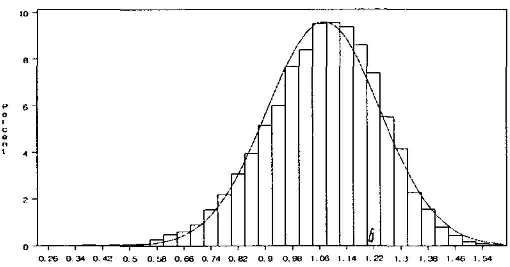
To confirm the normality of $\widetilde{H}(g)$, 10000 inverse Gaussian samples with $\mu = 1$ for four values of λ were generated by the algorithm of Michael et al. (1976). For sample size, three cases were considered: $n = 20$ (small size), $n = 50$ (moderate size) and $n = 100$ (large size). The estimator $\widetilde{H}(g)$ was calculated from each sample and then the calculated values were fitted to the normal distribution with theoretically specified mean and variance.

<Table 1> provides the results on three well-known EDF tests for nor-mality(the Kolmogorov-Smirnov D, the Carmer-von Mises W^2 and the Anderson-Darling A^2 tests), based on 10000 simulated values of $\widetilde{H}(g)$. When sample size is small($n = 20$), all of the tests, as might be expected, show significance at the level 5%. As sample size grows($n = 50$), goodness-of-fit for normality shows a tendency to be much improved. The p-values of the A^2 test is observed to be less than 5% for all cases of λ, whereas D and W^2 tests are appeared not to be significant at 5% for $\lambda = 1,2,4$, However, all the tests show no significance at 1% across the values of λ except for the A^2 test when $\lambda = 0.5$. Thus, it can be seen that the normality of $\widetilde{H}(g)$ is preserved. When sample size is large($n = 100$), an evidence that $\widetilde{H}(g)$ follows the normal distribution is clearly obtained, on the ground that all the tests are not significant at the level 5% except for the A^2 test when $\lambda = 0.5$. In addition to the testing results, the empirical distribution of $\widetilde{H}(g)$ with the solid line representing a fitted normal curve was depicted in <figures 2-4>, which show $\widetilde{H}(g)$ to be well

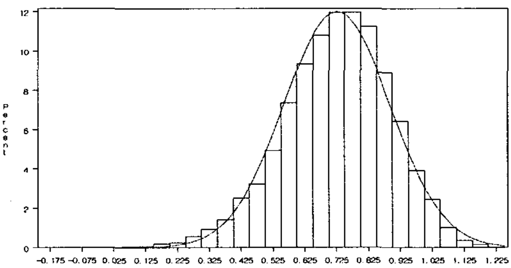
approximated by the normal distribution provided that sample size is not small.

<Table 1> Testing results on normality based on 10000 simulated values of $\widetilde{H}(g)$

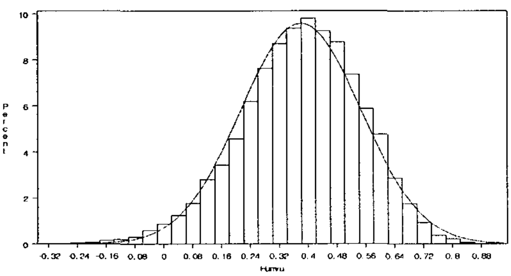
n	λ	D	$p-$value	W^2	$p-$value	A^2	$p-$value
20	0.5	0.0190	0.2%	1.3883	< 0.1%	11.3089	< 0.1%
	1.0	0.0248	< 0.1%	2.0605	< 0.1%	13.5023	< 0.1%
	2.0	0.0247	< 0.1%	2.2652	< 0.1%	13.8492	< 0.1%
	4.0	0.0293	< 0.1%	2.4958	< 0.1%	14.3270	< 0.1%
50	0.5	0.0156	1.7%	0.6149	2.2%	4.5751	0.5%
	1.0	0.0101	> 25%	0.3334	11.1%	2.8271	3.6%
	2.0	0.0113	15.8%	0.3010	13.6%	2.8766	3.3%
	4.0	0.0111	17%	0.3579	9.5%	3.3143	2%
100	0.5	0.0119	11.9%	0.3042	13.4%	3.4203	1.9%
	1.0	0.0089	> 25%	0.0994	> 25%	1.8377	13.5%
	2.0	0.0075	> 25%	0.1655	> 25%	1.6418	14.5%
	4.0	0.0075	> 25%	0.1459	> 25%	1.4819	18.5%



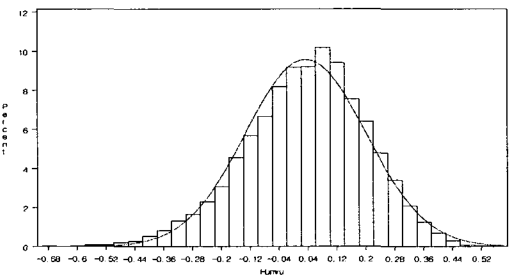
(a) $\mu=1$, $\lambda=0.5$



(b) $\mu=1$, $\lambda=1$

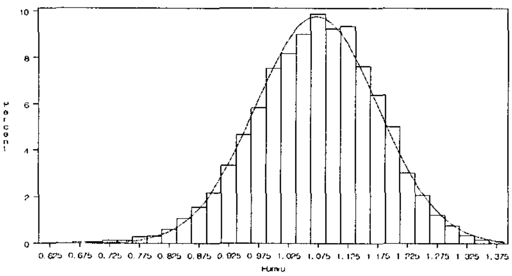


(c) $\mu=1$, $\lambda=2$

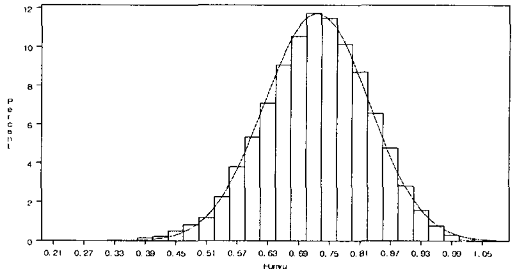


(d) $\mu=1$, $\lambda=4$

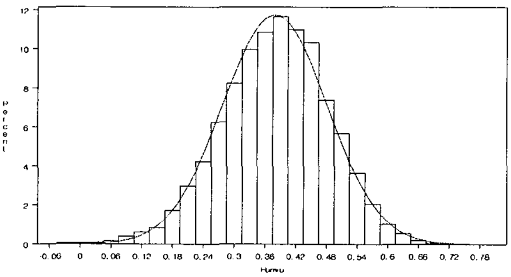
<Figure 2> Empirical distribution of $\widetilde{H}(g)$ based on 10000 simulated inverse Gaussian samples($n=20$)



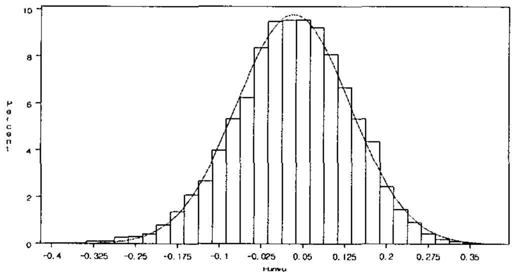
(a) $\mu=1, \ \lambda=0.5$



(b) $\mu=1, \ \lambda=1$

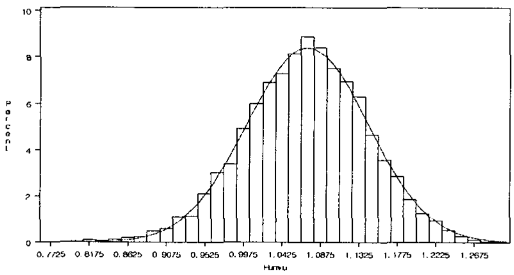


(c) $\mu=1, \ \lambda=2$

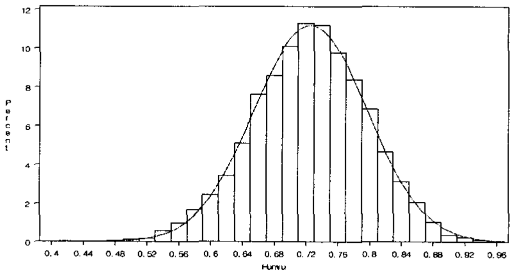


(d) $\mu=1, \ \lambda=4$

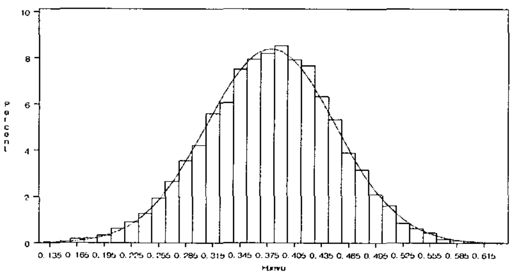
<Figure 3> Empirical distribution of $\widetilde{H}(g)$ based on 10000 simulated inverse Gaussian samples($n=50$)



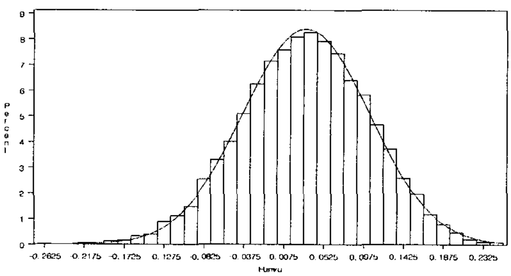
(a) $\mu=1, \ \lambda=0.5$



(b) $\mu=1, \ \lambda=1$



(c) $\mu=1, \ \lambda=2$



(d) $\mu=1, \ \lambda=4$

<Figure 4> Empirical distribution of $\widetilde{H}(g)$ based on 10000 simulated inverse Gaussian samples($n=100$)

4. Conclusion

In this paper, we have discussed the minimum variance unbiased estimation for the maximum entropy of the transformed inverse Gaussian random variable by $Y=X^{-1/2}$. The UMVU estimator of the entropy was derived and its variance was given. The distributional behaviour of the UMVU estimator was investigated. Monte Carlo simulation results reported the UMVU estimator to be well approximated by the normal distribution provided that sample size is not small.

The use of the UMVU estimator will be of interest in testing fit of goodness for the inverse Gaussian distribution. This work is currently in progress.

References

[1] Abramowitz, M. and Stegun, I.A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications, Inc, New York.

[2] Ahmed, N.A. and Gokhale, D.V. (1989). Entropy expressions and their esti -mators for multivariate distributions. *IEEE Transactions on Infor -mation Theory,* Vol. 35, 688-692.

[3] Burbea, J. and Rao, C.R. (1982). Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis,* Vol. 12, 576-579.

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

[5] Gradshteyn, I.S. and Ryzhik, I.M. (2000). *Tables of Integrals, Series, and Products*(6th Edition). Academic Press, San Diego.

[6] Havrda, J. and Charvat, F. (1967). Quantification method in classification processes: concept of structural α-entropy. *Kybernetika,* Vol. 3, 30-35.

[7] Hogg, R.V., McKean, J.W. and Craig, A.T. (2005). *Introduction to Math -ematical Statistics.* Pearson Education, Inc. Upper Saddle River.

[8] Jayens, E.T. (1957). Information theory and statistical mechanics. *Physical Review,* Vol. 106, 620-630.

[9] Kapur, J.N. and Kesavan, H.K. (1992). *Entropy Optimization Principles with Applications.* Academic Press, San Diego.

[10] Michael, J.R., Schucany, W.R. and Hass, R.W. (1976). Generating random variables using transformation with multiple roots. *The American Statistician,* Vol. 30, 88-90.

[11] Mudholkar, G.S. and Tian L. (2002). An entropy characterization of the

inverse Gaussian distribution and related goodness-of-fit test. *Journal of Statistical Planning and Inference*, Vol. 102, 211-221.

[12] Olshen, A.C. (1937). Transformations of the Pearson type III distributions. *The Annals of Mathematical Statistics*, Vol. 8, 176-200.

[13] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, Vol. 27, 379-423, 623-656.

[14] Singh, V.P. (1998). *Entropy-Based Parameter Estimation in Hydrology.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

[Received April 2006, Accepted September 2006]