Statistica Sinica **10**(2000), 937-951

THE INFORMATION LOSS IN THE INVERSE GAUSSIAN MODEL

Hui-Nien Hung

National Chiao Tung University

Abstract: In the inverse Gaussian model, the sample mean and sample reciprocal mean are minimum sufficient, and the distribution of sample reciprocal mean depends only on the dispersion parameter. Traditional inference about the dispersion parameter considers only the sample reciprocal mean instead of the whole sufficient statistic. This causes information loss, especially when the sample size is small. The purpose of this paper is to utilize the information of the dispersion parameter contained in the sample mean, and to improve the estimation of the dispersion parameter.

Key words and phrases: Average Likelihood, information loss, inverse Gaussian distribution, modified profile likelihood.

1. Introduction

The inverse Gaussian distribution was introduced by Schröedinger (1915). Since this distribution can be regarded as the first passage time in a Brownian motion, it has applications in fields such as economics, biology, medicine, and reliability testing (see Chhikara and Folks (1988)). The inverse Gaussian distribution has been studied extensively by many authors, for example, Tweedie (1945, 1946, 1957), Wald (1947), Wasan (1968) and Hsieh and Korwar (1990).

The probability density of an inverse Gaussian distribution is of the form

$$f(x;\mu,\lambda) = (\lambda/2\pi x^3)^{1/2}\exp\{-\lambda(x-\mu)^2/2\mu^2 x\};\ x>0,\ \mu>0,\ \lambda>0,$$

where μ is the mean parameter and λ is called the dispersion parameter. We denote this distribution by $\mathrm{IG}(\lambda,\mu)$. We are interested in the dispersion parameter λ and will treat μ as a nuisance parameter.

Let $X=(X_1,\ldots,X_n)$ be a random sample of size n from $\mathrm{IG}(\lambda,\mu)$. The likelihood function is

$$L(\lambda,\mu)=\prod_{i=1}^{n} f(X_i|\mu,\lambda)=(\frac{\lambda}{2\pi})^{\frac{n}{2}}(\prod_{i=1}^{n} X_i^{-\frac{3}{2}})\exp\{-\frac{n\lambda}{2}V-\frac{n\lambda(\bar{X}-\mu)^2}{2\mu^2\bar{X}}\};\ \lambda,\mu>0, \tag{1.1}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $V = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{X_i} - \frac{1}{\bar{X}})$, and $(\bar{X}, V)$ is a complete sufficient statistic for (λ, μ). Tweedie (1957) proved that the sample mean $\bar{X}$ has distribution $\mathrm{IG}(n\lambda, \mu)$, the statistic V has distribution $(\frac{1}{n\lambda})\chi^2_{n-1}$, and these two statistics are independent. The dispersion parameter λ is traditionally estimated by the MLE ($\hat{\lambda}_p = V^{-1}$), the UMVUE ($\tilde{\lambda} = \frac{n-3}{n}V^{-1}$) or $\hat{\lambda}_{mp} = \frac{n-1}{n}V^{-1}$, which maximizes the modified profile likelihood function given by Barndorff-Nielsen (1983).

We notice that all these estimates depend only on V, and not on $\bar{X}$. The purpose of the paper is to find the information of λ contained in $\bar{X}$ to improve the estimation of λ.

The organization of the paper is as follows. In Section 2, we discuss the intuitive information of λ contained in $\bar{X}$. In Section 3 we review the method of average likelihood (Hung and Wong (1996)), a method of eliminating nuisance parameters. In Section 4, we apply this method to the inverse Gaussian model. In the last section, we compare the information of λ contained in $\bar{X}$, as discussed in Section 2, and the information from the average likelihood. Also, we consider the estimation of any other monotone transformation of λ.

2. Information of λ in $\bar{X}$

To understand the information of λ contained in $\bar{X}$, consider how the inverse Gaussian distribution can be interpreted as the first passage time of a Brownian motion.

Suppose that $\{B(t), t \geq 0\}$ is a one-dimensional Brownian motion process with positive drift v and diffusion β, i.e., $B(0) = 0$, $\{B(t), t \geq 0\}$ has stationary and independent increments, and for every $t > 0$, $B(t)$ is normally distributed with mean vt and variance βt. Let T be the first time the process hits 1 (see Figure 1).

Schröedinger (1915) showed that T has the $\mathrm{IG}(\lambda, \mu)$ distribution with $\mu = 1/v$ and $\lambda = 1/\beta$. Since $\bar{X}$ is the average first passage time and λ is a reciprocal measure of the diffusion parameter in $B(t)$, $\bar{X}$ does contain information about λ. Intuitively, when λ is close to 0, we can see that the Brownian motion $B(t)$ will hit 1 very fast no matter what μ is, hence $\bar{X}$ should be small with high probability. In other words, when $\bar{X}$ is large, we can conclude that λ should not be small no matter what the value of μ.

Next, we give a theoretical interpretation of the above information. Consider two Brownian motion processes, $B_1(\omega, t)$ and $B_2(\omega, t)$ with positive drifts $1/\mu_1$ and $1/\mu_2$, respectively, and common diffusion $1/\lambda$. It is not difficult to see that $B_1(\omega, t)$ has the same distribution as $B_2(\omega, t) + (\frac{1}{\mu_1} - \frac{1}{\mu_2})t$. Let Y and Z be the first passage times of $B_1(\omega, t)$ and $B_2(\omega, t)$ through 1, respectively. Then, when

$\mu_2 > \mu_1$, we have $Y \sim \mathrm{IG}(\lambda, \mu_1)$, $Z \sim \mathrm{IG}(\lambda, \mu_2)$, and Z is stochastically greater than Y. Therefore, for fixed $a > 0$, we have

$$\begin{aligned}\inf_{\mu} P_{\lambda,\mu}\{\ \bar{X} \le a\ \} &= \lim_{\mu\to\infty} P_{\lambda,\mu}\{\ \bar{X} \le a\ \}\\ &= \lim_{\mu\to\infty}\Big[\Phi\{\sqrt{\frac{n\lambda}{a}}(\frac{a}{\mu}-1)\} + e^{2n\lambda/\mu}\Phi\{-\sqrt{\frac{n\lambda}{a}}(\frac{a}{\mu}+1)\}\Big]\\ &= \Phi\Big\{-\sqrt{\frac{n\lambda}{a}}\Big\} + \Phi\Big\{-\sqrt{\frac{n\lambda}{a}}\Big\}\\ &= 2\Phi\Big\{-\sqrt{\frac{n\lambda}{a}}\Big\}, \end{aligned} \tag{2.1}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Examining the right hand side of (2.1), we find that $2\Phi\{-\sqrt{\frac{n\lambda}{a}}\}$ is a decreasing function of λ and $2\Phi\{-\sqrt{\frac{n\lambda}{a}}\} \to 1$ as $\lambda \to 0$. Hence (2.1) demonstrates that if λ is small, $\bar{X}$ should be small with high probability no matter what μ is. Thus, when a large $\bar{X}$ is observed, we infer that λ is large.

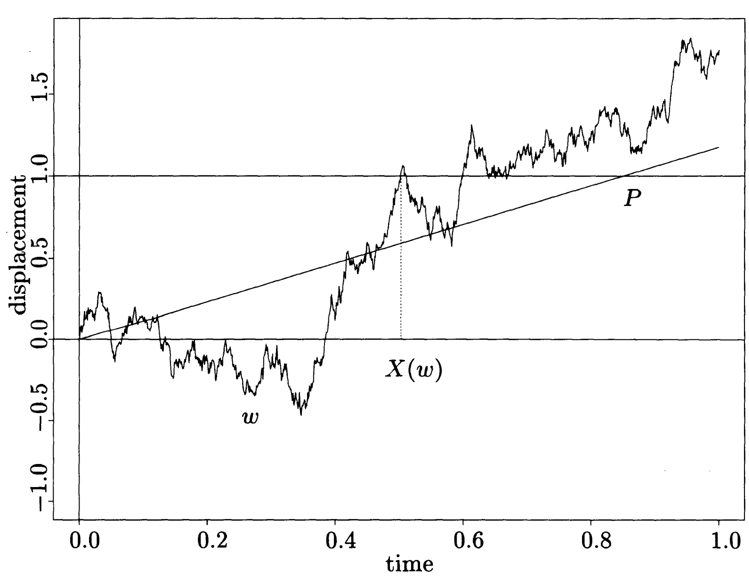

Figure 1. Brownian Motion and first passaging time where $P = (\mu, 1)$.

3. Average Likelihood

The method we use to recover the information of λ contained in $\bar{X}$ is the average likelihood method proposed by Hung and Wong (1996). For eliminating the nuisance parameter the profile likelihood method, which maximizes the likelihood function for each fixed parameter of interest, is the simplest method. Unfortunately, it does not take into account the uncertainty due to lack of knowledge of the nuisance parameters, and can be misleading in both precision (degrees of freedom) and location (bias). Therefore, for each fixed parameter of interest, instead of maximizing the likelihood function, the average likelihood method considers the whole likelihood function and averages it.

The construction of the average likelihood function is as follows. Let a statistical model be parameterized by (θ, φ), where θ is the parameter of interest and φ is the nuisance parameter. Here we assume that θ is orthogonal to φ in the sense that condition (M) in Hung and Wong (1996) is satisfied (in some regular statistical models, this is equivalent to $E(\frac{\partial^2 \log f(X;\theta,\varphi)}{\partial\theta\partial\varphi}) = 0$). Otherwise we need to reparameterize the nuisance parameters. After reparameterization, Hung and Wong (1996) define the average likelihood to be

$$L_{ave}(\theta) = \lim_{n\to\infty} \frac{\int_{\Phi_n} L(\theta,\varphi)\pi(\varphi|\theta)\ d\varphi}{\int_{\Phi_n} \pi(\varphi|\theta)\ d\varphi} \Big/ \frac{\int_{\Phi_n} L(\theta_0,\varphi)\pi(\varphi|\theta_0)\ d\varphi}{\int_{\Phi_n} \pi(\varphi|\theta_0)\ d\varphi},$$

where Φ is the range of φ, $(\Phi_n)_{n=1}^{\infty}$ is a sequence of compact subsets of Φ such that Φ_n goes to Φ as n goes to infinity, and θ_0 is some fixed point. For the choice of weighting function $\pi(\varphi|\theta)$ in the above formula, the idea is to put equal mass on each small interval of equal Hellinger length. Under some regularity conditions, Hung and Wong (1996) proved that this leads to the choice $\pi(\varphi|\theta) = (E(-\frac{\partial^2}{\partial\varphi^2} \log\ f(X;\theta,\varphi)))^{\frac{1}{2}}$, which is the Jeffrey's prior for φ when θ is fixed. It is well known that Jeffrey's prior has some good properties in the Bayesian approach when the dimension of the parameter is one. However, Jeffrey's prior only exists when the model is regular. Therefore, the weighting function based on Hellinger distance can be regarded as an extension of the Jeffrey's prior to non-smooth models, or to cases where the nuisance parameter is discrete. Details can be found in Hung and Wong (1996). We note that in smooth models this approach is related to the reference prior approach of Berger and Bernardo (1992). However the issue of parameterization of the nuisance parameters, which is essential for good results, was not discussed in the reference prior approach.

Now let $X = (X_1, \ldots, X_n)$ be a random sample of size n from $\mathrm{IG}(\lambda, \mu)$. We have

$$E\Big(\frac{\partial^2 \log f(X;\lambda,\mu)}{\partial\lambda\partial\mu}\Big) = \frac{n}{\mu^3}E(\bar{X}) - \frac{n}{\mu^2} = \frac{n}{\mu^3}\mu - \frac{n}{\mu^2} = 0,$$

i.e., λ is orthogonal to μ and

$$E\Big(-\ \frac{\partial^2 \log L(\lambda,\mu)}{\partial\mu^2}\Big) = \frac{n\lambda}{\mu^3}.$$

Hence, we choose $\pi(\mu|\lambda) = \mu^{-\frac{3}{2}}$, and the average likelihood function of λ is

$$
\begin{aligned}
L_{ave}(\lambda) &\propto \int_0^\infty L(\lambda,\mu)\cdot\mu^{-\frac{3}{2}}\, d\mu \\
&\propto \int_0^\infty \big(\frac{\lambda}{2\pi}\big)^{\frac{n}{2}}\big(\prod_{i=1}^n X_i^{-\frac{3}{2}}\big)\exp\{-\frac{n\lambda}{2}V - \frac{n\lambda(\bar{X}-\mu)^2}{2\mu^2\bar{X}}\}\cdot\mu^{-\frac{3}{2}}\, d\mu \\
&\propto \lambda^{\frac{n}{2}}\exp\Big\{-\frac{n\lambda}{2}V\Big\}\int_0^\infty \mu^{-\frac{3}{2}}\exp\Big\{-\frac{n\lambda}{2\bar{X}}\Big(\frac{\bar{X}}{\mu}-1\Big)^2\Big\}\, d\mu \\
&\propto \exp\Big\{-\frac{n\lambda}{2}V\Big\}\,\lambda^{\frac{n-1}{2}}\cdot\int_0^\infty \exp\Big\{-\frac{n\lambda}{2\bar{X}}(t-1)^2\Big\}\cdot t^{-\frac{1}{2}}\, dt. \qquad (3.1)
\end{aligned}
$$

4. Results

In this section, we compare the estimator of λ derived from the average likelihood of λ (which contains the information of λ in $\bar{X}$) and the three estimators of λ mentioned in Section 1 ($\hat{\lambda}_p, \hat{\lambda}_{mp}$ and $\tilde{\lambda}$) which depend only on V. To compare these estimators, we consider the Pitman (1937) criterion: if T_1 and T_2 are two estimators of λ, we say that T_1 is **closer** to λ than T_2 if $P\{\ |T_1-\lambda| < |T_2-\lambda|\ \} > 0.5$.

Before finding the information of λ contained in $\bar{X}$, let us review a property of the inverse Gaussian distribution. If $Y \sim \mathrm{IG}(\lambda,\mu)$, then, for $a>0$, $aY \sim \mathrm{IG}(a\lambda, a\mu)$. For invariance, if T is an estimator of λ and $X=(X_1,\ldots,X_n)$ is a random sample, T should satisfy

$$T(aX) = aT(X), \quad \text{for} \quad a>0. \qquad (4.1)$$

It is easy to check that $\hat{\lambda}_p$, $\hat{\lambda}_{mp}$ and $\tilde{\lambda}$ satisfy (4.1). Suppose that $T_1(X)$ and $T_2(X)$ are two invariant estimators of λ. For $a>0$, let $Y=(Y_1,\ldots,Y_n)=(aX_1,\ldots,aX_n)$. Then $Y_1,\ldots,Y_n$ are i.i.d. $\mathrm{IG}(a\lambda,a\mu)$ and $T_1(Y)$ and $T_2(Y)$ are estimators of $a\lambda$. Since

$$
\begin{aligned}
P_{\lambda,\mu}\{\ |T_1(X)-\lambda| < |T_2(X)-\lambda|\ \} &= P_{\lambda,\mu}\{\ |aT_1(X)-a\lambda| < |aT_2(X)-a\lambda|\ \} \\
&= P_{\lambda,\mu}\{\ |T_1(aX)-a\lambda| < |T_2(aX)-a\lambda|\ \} \\
&= P_{a\lambda,a\mu}\{\ |T_1(Y)-a\lambda| < |T_2(Y)-a\lambda|\ \},
\end{aligned}
$$

we obtain the following.

Lemma 4.1. *For the inverse Gaussian model,* $\mathrm{IG}(\lambda,\mu)$*, if* T_1 *and* T_2 *are two invariant estimators of* λ*, then* $P_{\lambda,\mu}\{\ |T_1-\lambda| < |T_2-\lambda|\ \}$ *depends on* (λ,μ) *only through* $\frac{\lambda}{\mu}$.

Going further, if a likelihood function of λ, $L(\lambda;X)$, satisfies $L(\lambda;X) = L(a\lambda;aX)$ up to a factor depending only on X and a, then we say the likelihood function is an invariant likelihood function of λ. For an invariant likelihood function, it is not difficult to establish the following result.

Lemma 4.2. *If T is an estimator of λ which maximizes an invariant likelihood function, T is an invariant estimator of λ. In particular, the average likelihood function of λ in* (3.1) *is invariant. Hence $\hat{\lambda}_{ave}$, the estimator of λ obtained by maximizing* (3.1)*, is an invariant estimator of λ.*

By the above results, $\hat{\lambda}_p$, $\tilde{\lambda}$, $\hat{\lambda}_{mp}$, and $\hat{\lambda}_{ave}$ are all invariant estimators of λ and to compare them, we need only consider different values of λ/μ in our study. The results are shown in Tables 1 to 3. These tables show that $\hat{\lambda}_{ave}$ is better than $\hat{\lambda}_p$, $\hat{\lambda}_{mp}$ and $\tilde{\lambda}$ in all situations. In particular when the sample size is small, $\hat{\lambda}_{ave}$ is much better than $\hat{\lambda}_p$, $\hat{\lambda}_{mp}$ and $\tilde{\lambda}$. In other words, the information on λ contained in $\bar{X}$ plays an important role and should not be ignored.

5. Discussion

1. Note that the average likelihood function of λ depends on both $\bar{X}$ and V, and that $\bar{X}$ does contain information about λ. We will see why, from the average likelihood point of view, the information of λ contained in $\bar{X}$ matches the intuitive and theoretical information of λ contained in $\bar{X}$, as discussed in Section 2.

Let

$$k(t) = \begin{cases} \frac{1}{\sqrt{t}} & \text{if } t > 0, \\ 0 & \text{if } t \le 0. \end{cases}$$

From (3.1), we have

$$\begin{aligned} L_{ave}(\lambda) &\propto \exp\Big\{-\frac{n\lambda}{2}V\Big\}\lambda^{\frac{n}{2}} \int_0^\infty \exp\Big\{-\frac{n\lambda}{2\bar{X}}(t-1)^2\Big\} t^{-\frac{1}{2}}\,dt \\ &\propto \exp\Big\{-\frac{n\lambda}{2}V\Big\}\lambda^{\frac{n-1}{2}} \cdot \sqrt{\frac{2\pi\bar{X}}{n}} \int_0^\infty t^{-1/2} \cdot \sqrt{\frac{n\lambda}{2\pi\bar{X}}} \exp\Big\{-\frac{n\lambda}{2\bar{X}}(t-1)^2\Big\}dt \\ &\propto L_{mp}(\lambda) \cdot E_\lambda[k(T)|\bar{X}], \end{aligned}$$

where $L_{mp}(\lambda)$ is the modified profile likelihood of λ (i.e., the marginal density of V) given by Barndorff-Nielsen (1983), and T given $\bar{X}$ is distributed as $N(1, \frac{\bar{X}}{n\lambda})$. Hence we can regard $L_{ave}(\lambda)$ as $L_{mp}(\lambda)$ multiplied by a weight function $E_\lambda[k(T)|\bar{X}]$ which depends on $\bar{X}$ but not on V. Since $E_{a\lambda}[k(T)|a\bar{X}] = E_\lambda[k(T)|\bar{X}]$ and $E_\lambda[k(T)|\bar{X}]$ goes to 0 and 1 as λ goes to 0 and ∞ respectively, we can conclude that, from the average likelihood function, λ should be large when large $\bar{X}$ is observed. Hence the average likelihood provides the right information about λ contained in $\bar{X}$ and suggests that $\hat{\lambda}_{ave}$ sholud be better than estimators using only V. The following figure is the average likelihood function of λ due to the observation $\bar{X}$, (i.e., $E_{\lambda}[k(T)|\bar{X}]$), and illustrates the results mentioned above.

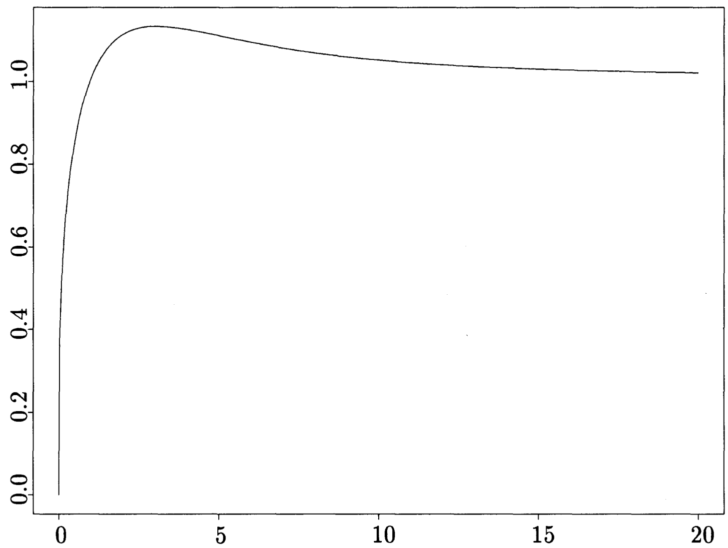

Figure 2. The average likelihood function of λ due to $\bar{X}$. The X-axis is the value of $\sqrt{\frac{n\lambda}{\bar{X}}}$, the Y-axis is the value of $E_{\lambda}[k(T)|\bar{X}]$.

2. Mean squared error (MSE) is the most common criterion for evaluating estimators. It is not a good criterion in this study however. Our interest is in the parameter λ in an $\mathrm{IG}(\lambda, \mu)$ model. Recall that the variance of the inverse Gaussian distribution is μ^3/λ. Hence λ is a reciprocal measure of variance and also a reciprocal measure of the diffusion parameter in the Brownian motion. From the mean squared error point of view, the difference between $\lambda = 40$ and $\lambda = 50$ is much larger than the difference between $\lambda = 0.1$ and $\lambda = 1$. But when we consider distributions, the difference between $\lambda = 40$ and $\lambda = 50$ is much smaller than the difference between $\lambda = 0.1$ and $\lambda = 1$. Hence the mean squared error is not really suitable and we choose Pitman closeness instead. It is also reasonable to compare two estimators of λ, T_1 and T_2, by considering $P\{|\log T_1 - \log\lambda| < |\log T_2 - \log\lambda|\}$. By a proof similar to that of Lemma 4.1, it is easy to show that, for the inverse Gaussian model, if T_1 and T_2 are two invariant estimators of λ, then $P_{\lambda,\mu}\{|\log T_1 - \log\lambda| < |\log T_2 - \log\lambda|\}$ depends on (λ, μ) only through λ/μ. The comparisons in terms of this new criterion are shown in Tables 4 to 6. These tables show that there is no obvious difference between the use of $P\{|T_1 - \lambda| < |T_2 - \lambda|\}$ or $P\{|\log T_1 - \log\lambda| < |\log T_2 - \log\lambda|\}$ as the criterion.

We might ask how $\hat{\lambda}_{ave}$ performs if we use other scales in λ instead. To answer this question, we consider the following.

Lemma 5.1. *Let $h(\lambda)$ be a monotone function of λ. Then*

$$P\{\ |h(T_1(X)) - h(\lambda)| < |h(T_2(X)) - h(\lambda)|\ \}$$
$$\geq P\{\ |T_1(X) - \lambda| < |T_2(X) - \lambda|\ \} - P\{\ (T_1(X) - \lambda)(T_2(X) - \lambda) < 0\ \}.$$

From this lemma and the entries in Tables 1, 2, 3 and 7, we can see that, for the most part, $P\{\ |h(\hat{\lambda}_{ave}(X)) - h(\lambda)| < |h(\hat{\lambda}_{mp}(X))) - h(\lambda)|\ \} > 0.5$. Thus, $\hat{\lambda}_{ave}$ is really better than $\hat{\lambda}_{mp}$ no matter what kind of scale about λ we use.

3. From Tables 1 through 6, $P(|\hat{\lambda}_{ave} - \lambda| < |\hat{\lambda}_{other} - \lambda|)$ appears to converge to 1/2. This suggests that all these estimators are asymptotically efficient. This result can be understood by the following argument. From Section 1, we have $\hat{\lambda}_p = V^{-1}$, $\tilde{\lambda} = \frac{n-3}{n}V^{-1}$ and $\hat{\lambda}_{mp} = \frac{n-1}{n}V^{-1}$. Hence, they agree to $O_p(n^{-\frac{1}{2}})$. From the asymptotic result of average likelihood, (Hung and Wong (1996)), the logarithm of the average likelihood is of order n and differs from the logarithm of profile likelihood in a term of $O_p(1)$ and, therefore, $\hat{\lambda}_{ave} - \hat{\lambda}_p = O_p(\frac{1}{n})$. Thus, the four estimators mentioned in this paper are equal to $O_p(n^{-\frac{1}{2}})$. Since $\hat{\lambda}_p$ is well-known to be asymptotically efficient, the above argument deduce has all four estimators asymptotically efficient.

Acknowledgement

I am grateful to Mr. Liu Meng-Fu for making the tables in this paper. Also, I would like to express my gratitude to a referee for his/her valuable comments which greatly improved the quality of the paper. This research work was partially supported by the National Science Council of the Republic of China, Grant No. NSC 86-2115-M-009-037-T.

Appendix

Proof of Lemma 5.1. Without loss of generality, we may assume that h is increasing. Since

$$\begin{aligned}
&P\{|T_1(X) - \lambda| < |T_2(X) - \lambda|\} \\
&\leq P\{\lambda < T_1(X) < T_2(X)\} + P\{T_2(X) < T_1(X) < \lambda\} + P\{(T_1(X) - \lambda)(T_2(X) - \lambda) < 0\} \\
&= P\{h(\lambda) < h(T_1(X)) < h(T_2(X))\} + P\{h(T_2(X)) < h(T_1(X)) < h(\lambda)\} \\
&\quad + P\{(T_1(X) - \lambda)(T_2(X) - \lambda) < 0\} \\
&\leq P\{|h(T_1(X)) - h(\lambda)| < |h(T_2(X)) - h(\lambda)|\} + P\{(T_1(X) - \lambda)(T_2(X) - \lambda) < 0\},
\end{aligned}$$

it follows that

$$P\{|h(T_1(X)) - h(\lambda)| < |h(T_2(X)) - h(\lambda)|\}$$
$$\geq P\{|T_1(X) - \lambda| < |T_2(X) - \lambda|\} - P\{(T_1(X) - \lambda)(T_2(X) - \lambda) < 0\}.$$

Table 1. $P\{|\hat{\lambda}_{ave} - \lambda| < |\hat{\lambda}_p - \lambda|\}$.

$\frac{\lambda}{\mu}$	Sample Size				
	2	3	4	5	6
0.001	0.8074	0.7366	0.6955	0.6756	0.6548
(St.E)	(0.0039)	(0.0044)	(0.0046)	(0.0047)	(0.0048)
0.01	0.8106	0.7360	0.6930	0.6735	0.6574
(St.E)	(0.0039)	(0.0044)	(0.0046)	(0.0047)	(0.0047)
0.1	0.8071	0.7401	0.7010	0.6712	0.6532
(St.E)	(0.0039)	(0.0044)	(0.0046)	(0.0047)	(0.0048)
0.5	0.7960	0.7157	0.6771	0.6681	0.6401
(St.E)	(0.0040)	(0.0045)	(0.0047)	(0.0047)	(0.0048)
1	0.7958	0.7244	0.6628	0.6432	0.6225
(St.E)	(0.0040)	(0.0045)	(0.0047)	(0.0048)	(0.0048)
2	0.7797	0.7011	0.6593	0.6393	0.6212
(St.E)	(0.0041)	(0.0046)	(0.0047)	(0.0048)	(0.0049)
10	0.7711	0.6943	0.6676	0.6436	0.6339
(St.E)	(0.0042)	(0.0046)	(0.0047)	(0.0048)	(0.0048)
100	0.7762	0.7066	0.6821	0.6458	0.6337
(St.E)	(0.0042)	(0.0046)	(0.0047)	(0.0048)	(0.0048)
1000	0.7750	0.7159	0.6750	0.6499	0.6340
(St.E)	(0.0042)	(0.0045)	(0.0047)	(0.0048)	(0.0048)
$\frac{\lambda}{\mu}$	Sample Size				
	10	15	20	25	30
0.001	0.6141	0.5852	0.5720	0.5629	0.5580
(St.E)	(0.0049)	(0.0049)	(0.0049)	(0.0050)	(0.0050)
0.01	0.6156	0.5818	0.5699	0.5637	0.5536
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)
0.1	0.6102	0.5751	0.5604	0.5543	0.5430
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)
0.5	0.5877	0.5628	0.5552	0.5518	0.5379
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1	0.5968	0.5679	0.5628	0.5541	0.5401
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
2	0.5940	0.5702	0.5542	0.5467	0.5407
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
10	0.6016	0.5712	0.5535	0.5547	0.5860
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0049)
100	0.6031	0.5814	0.5644	0.5633	0.5452
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)
1000	0.6053	0.5766	0.5611	0.5542	0.5470
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)

Table 2. P$\{\ |\hat{\lambda}_{ave} - \lambda|\ <\ |\tilde{\lambda} - \lambda|\ \}$.

$\frac{\lambda}{\mu}$	Sample Size				
	2	3	4	5	6
0.001 (St.E)	—	—	0.5471 (0.0050)	0.5400 (0.0050)	0.5370 (0.0050)
0.01 (St.E)	—	—	0.5529 (0.0050)	0.5435 (0.0050)	0.5391 (0.0050)
0.1 (St.E)	—	—	0.5465 (0.0050)	0.5493 (0.0050)	0.5376 (0.0050)
0.5 (St.E)	—	—	0.5811 (0.0049)	0.5577 (0.0050)	0.5626 (0.0050)
1 (St.E)	—	—	0.5904 (0.0049)	0.5821 (0.0049)	0.5789 (0.0049)
2 (St.E)	—	—	0.5929 (0.0049)	0.5794 (0.0049)	0.5704 (0.0050)
10 (St.E)	—	—	0.5838 (0.0049)	0.5747 (0.0049)	0.5648 (0.0050)
100 (St.E)	—	—	0.5693 (0.0050)	0.5696 (0.0050)	0.5609 (0.0050)
1000 (St.E)	—	—	0.5779 (0.0049)	0.5739 (0.0049)	0.5591 (0.0050)

$\frac{\lambda}{\mu}$	Sample Size				
	10	15	20	25	30
0.001 (St.E)	0.5355 (0.0050)	0.5290 (0.0050)	0.5264 (0.0050)	0.5268 (0.0050)	0.5223 (0.0050)
0.01 (St.E)	0.5357 (0.0050)	0.5329 (0.0050)	0.5328 (0.0050)	0.5309 (0.0050)	0.5270 (0.0050)
0.1 (St.E)	0.5310 (0.0050)	0.5362 (0.0050)	0.5355 (0.0050)	0.5375 (0.0050)	0.5337 (0.0050)
0.5 (St.E)	0.5535 (0.0050)	0.5527 (0.0050)	0.5418 (0.0050)	0.5412 (0.0050)	0.5458 (0.0050)
1 (St.E)	0.5517 (0.0050)	0.5456 (0.0050)	0.5375 (0.0050)	0.5339 (0.0050)	0.5421 (0.0050)
2 (St.E)	0.5505 (0.0050)	0.5452 (0.0050)	0.5478 (0.0050)	0.5435 (0.0050)	0.5417 (0.0050)
10 (St.E)	0.5417 (0.0050)	0.5454 (0.0050)	0.5448 (0.0050)	0.5371 (0.0050)	0.4952 (0.0050)
100 (St.E)	0.5470 (0.0050)	0.5335 (0.0050)	0.5308 (0.0050)	0.5317 (0.0050)	0.5317 (0.0050)
1000 (St.E)	0.5419 (0.0050)	0.5341 (0.0050)	0.5300 (0.0050)	0.5320 (0.0050)	0.5313 (0.0050)

Table 3. $P\{\ |\hat{\lambda}_{ave} - \lambda|\ <\ |\hat{\lambda}_{mp} - \lambda|\ \}$.

$\frac{\lambda}{\mu}$	Sample Size				
	2	3	4	5	6
0.001	0.5897	0.5414	0.5253	0.5141	0.5085
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.01	0.5935	0.5487	0.5310	0.5233	0.5120
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.1	0.6422	0.5992	0.5813	0.5852	0.5799
(St.E)	(0.0048)	(0.0049)	(0.0049)	(0.0049)	(0.0049)
0.5	0.7786	0.7854	0.7707	0.7766	0.7637
(St.E)	(0.0042)	(0.0041)	(0.0042)	(0.0042)	(0.0042)
1	0.8568	0.8374	0.7912	0.7411	0.6662
(St.E)	(0.0035)	(0.0037)	(0.0041)	(0.0044)	(0.0047)
2	0.8816	0.7579	0.6485	0.5919	0.5636
(St.E)	(0.0032)	(0.0043)	(0.0048)	(0.0049)	(0.0050)
10	0.6818	0.6119	0.5959	0.5784	0.5721
(St.E)	(0.0047)	(0.0049)	(0.0049)	(0.0049)	(0.0049)
100	0.6811	0.6226	0.6070	0.5837	0.5762
(St.E)	(0.0047)	(0.0048)	(0.0049)	(0.0049)	(0.0049)
1000	0.6716	0.6315	0.6049	0.5805	0.5760
(St.E)	(0.0047)	(0.0048)	(0.0049)	(0.0049)	(0.0049)

$\frac{\lambda}{\mu}$	Sample Size				
	10	15	20	25	30
0.001	0.5082	0.5087	0.5119	0.5099	0.5107
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.01	0.5036	0.5172	0.5199	0.5087	0.5099
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.1	0.5927	0.6183	0.6331	0.6472	0.6374
(St.E)	(0.0049)	(0.0049)	(0.0048)	(0.0048)	(0.0048)
0.5	0.6624	0.5447	0.5250	0.5205	0.5105
(St.E)	(0.0047)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1	0.5544	0.5274	0.5306	0.5256	0.5114
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
2	0.5493	0.5327	0.5210	0.5158	0.5145
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
10	0.5564	0.5352	0.5208	0.5229	0.5622
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
100	0.5551	0.5464	0.5317	0.5331	0.5195
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1000	0.5611	0.5402	0.5318	0.5277	0.5206
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)

Table 4. $P\{\ |\log\hat{\lambda}_{ave} - \log\lambda|\ <\ |\log\hat{\lambda}_p - \log\lambda|\ \}$.

$\frac{\lambda}{\mu}$	Sample Size				
	2	3	4	5	6
0.001	0.8031	0.7339	0.6932	0.6750	0.6536
(St.E)	(0.0040)	(0.0044)	(0.0046)	(0.0047)	(0.0048)
0.01	0.8061	0.7341	0.6911	0.6725	0.6569
(St.E)	(0.0040)	(0.0044)	(0.0046)	(0.0047)	(0.0047)
0.1	0.8025	0.7381	0.6981	0.6693	0.6518
(St.E)	(0.0040)	(0.0044)	(0.0046)	(0.0047)	(0.0048)
0.5	0.7890	0.7123	0.6736	0.6647	0.6375
(St.E)	(0.0041)	(0.0045)	(0.0047)	(0.0047)	(0.0048)
1	0.7844	0.7163	0.6558	0.6386	0.6194
(St.E)	(0.0041)	(0.0045)	(0.0048)	(0.0048)	(0.0049)
2	0.7632	0.6905	0.6525	0.6338	0.6177
(St.E)	(0.0043)	(0.0046)	(0.0048)	(0.0048)	(0.0049)
10	0.7527	0.6851	0.6617	0.6408	0.6313
(St.E)	(0.0043)	(0.0046)	(0.0047)	(0.0048)	(0.0048)
100	0.7633	0.6987	0.6777	0.6428	0.6317
(St.E)	(0.0043)	(0.0046)	(0.0047)	(0.0048)	(0.0048)
1000	0.7616	0.7083	0.6700	0.6460	0.6318
(St.E)	(0.0043)	(0.0045)	(0.0047)	(0.0048)	(0.0048)

$\frac{\lambda}{\mu}$	Sample Size				
	10	15	20	25	30
0.001	0.6135	0.5851	0.5718	0.5629	0.5579
(St.E)	(0.0049)	(0.0049)	(0.0049)	(0.0050)	(0.0050)
0.01	0.6151	0.5817	0.5698	0.5636	0.5536
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)
0.1	0.6094	0.5742	0.5601	0.5542	0.5428
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)
0.5	0.5863	0.5619	0.5549	0.5514	0.5378
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1	0.5954	0.5671	0.5623	0.5539	0.5400
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
2	0.5923	0.5695	0.5540	0.5465	0.5404
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
10	0.6009	0.5703	0.5526	0.5545	0.5856
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0049)
100	0.6012	0.5808	0.5639	0.5629	0.5452
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)
1000	0.6038	0.5762	0.5608	0.5541	0.5469
(St.E)	(0.0049)	(0.0049)	(0.0050)	(0.0050)	(0.0050)

Table 5. $P\{\ |\log\hat{\lambda}_{ave} - \log\lambda|\ <\ |\log\tilde{\lambda} - \log\lambda|\ \}$.

$\frac{\lambda}{\mu}$	Sample Size				
	2	3	4	5	6
0.001	—	—	0.6175	0.5773	0.5627
(St.E)			(0.0049)	(0.0049)	(0.0050)
0.01	—	—	0.6255	0.5792	0.5657
(St.E)			(0.0048)	(0.0049)	(0.0050)
0.1	—	—	0.6137	0.5856	0.5620
(St.E)			(0.0049)	(0.0049)	(0.0050)
0.5	—	—	0.6376	0.5879	0.5834
(St.E)			(0.0048)	(0.0049)	(0.0049)
1	—	—	0.6432	0.6070	0.5941
(St.E)			(0.0048)	(0.0049)	(0.0049)
2	—	—	0.6433	0.6059	0.5876
(St.E)			(0.0048)	(0.0049)	(0.0049)
10	—	—	0.6398	0.6045	0.5806
(St.E)			(0.0048)	(0.0049)	(0.0049)
100	—	—	0.6274	0.5978	0.5793
(St.E)			(0.0048)	(0.0049)	(0.0049)
1000	—	—	0.6308	0.6016	0.5787
(St.E)			(0.0048)	(0.0049)	(0.0049)

$\frac{\lambda}{\mu}$	Sample Size				
	10	15	20	25	30
0.001	0.5439	0.5339	0.5292	0.5286	0.5240
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.01	0.5465	0.5370	0.5359	0.5325	0.5288
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.1	0.5407	0.5409	0.5374	0.5386	0.5345
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.5	0.5598	0.5555	0.5430	0.5426	0.5469
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1	0.5565	0.5477	0.5402	0.5357	0.5437
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
2	0.5564	0.5480	0.5493	0.5451	0.5426
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
10	0.5484	0.5485	0.5469	0.5393	0.4968
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
100	0.5528	0.5367	0.5330	0.5332	0.5324
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1000	0.5488	0.5373	0.5320	0.5329	0.5324
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)

Table 6. $P\{\ |\log\hat{\lambda}_{ave} - \log\lambda|\ <\ |\log\hat{\lambda}_{mp} - \log\lambda|\ \}$.

$\frac{\lambda}{\mu}$	Sample Size				
	2	3	4	5	6
0.001	0.5942	0.5433	0.5266	0.5150	0.5087
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.01	0.5973	0.5504	0.5321	0.5240	0.5125
(St.E)	(0.0049)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.1	0.6475	0.6004	0.5819	0.5860	0.5800
(St.E)	(0.0048)	(0.0049)	(0.0049)	(0.0049)	(0.0049)
0.5	0.7805	0.7861	0.7712	0.7766	0.7637
(St.E)	(0.0041)	(0.0041)	(0.0042)	(0.0042)	(0.0042)
1	0.8577	0.8373	0.7911	0.7410	0.6661
(St.E)	(0.0035)	(0.0037)	(0.0041)	(0.0044)	(0.0047)
2	0.8814	0.7576	0.6483	0.5918	0.5636
(St.E)	(0.0032)	(0.0043)	(0.0048)	(0.0049)	(0.0050)
10	0.6817	0.6119	0.5959	0.5784	0.5721
(St.E)	(0.0047)	(0.0049)	(0.0049)	(0.0049)	(0.0049)
100	0.6810	0.6226	0.6070	0.5837	0.5762
(St.E)	(0.0047)	(0.0048)	(0.0049)	(0.0049)	(0.0049)
1000	0.6716	0.6315	0.6049	0.5805	0.5760
(St.E)	(0.0047)	(0.0048)	(0.0049)	(0.0049)	(0.0049)

$\frac{\lambda}{\mu}$	Sample Size				
	10	15	20	25	30
0.001	0.5084	0.5088	0.5120	0.5100	0.5108
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.01	0.5039	0.5173	0.5199	0.5088	0.5100
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
0.1	0.5928	0.6185	0.6331	0.6472	0.6374
(St.E)	(0.0049)	(0.0049)	(0.0048)	(0.0048)	(0.0048)
0.5	0.6624	0.5447	0.5250	0.5205	0.5105
(St.E)	(0.0047)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1	0.5544	0.5274	0.5306	0.5255	0.5114
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
2	0.5493	0.5327	0.5210	0.5158	0.5145
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
10	0.5564	0.5352	0.5208	0.5229	0.5622
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
100	0.5551	0.5464	0.5317	0.5331	0.5195
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)
1000	0.5611	0.5402	0.5318	0.5277	0.5206
(St.E)	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0050)

Table 7. Number of $[\hat{\lambda}_{ave} - \lambda][\hat{\lambda}_{mp} - \lambda] < 0$ over 10000 simulations.

$\frac{\lambda}{\mu}$	Sample Size									
	2	3	4	5	6	10	15	20	25	30
0.001	834	690	584	568	511	423	296	278	239	215
0.01	804	687	595	574	527	392	281	264	247	183
0.1	796	594	545	464	396	272	179	110	87	54
0.5	593	369	292	200	155	116	97	75	59	39
1	370	241	177	174	181	137	66	70	25	14
2	276	218	193	170	136	53	15	11	12	16
10	137	63	35	19	20	8	2	1	2	1
100	8	3	1	0	3	1	1	0	1	0
1000	0	0	0	0	1	1	0	0	0	0

References

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-365.

Berger, J. O. and Bernardo, J. M. (1992). Ordered group reference priors with application to the multinomial problem. *Biometrika* **79**, 25-37.

Chhikara, R. S. and Folks, J. L. (1988). *The Inverse Gaussian Distribution.* Marcel Dekker, New York.

Folks, J. L. and Chhikara, R. S. (1978). The inverse Gaussian distribution and its statistical application–a review (with discussion). *J. Roy. Statist. Soc. Ser. B* **40**, 263-289.

Hsieh, H. K. and Korwar, R. (1990). Inadmissibility of the uniform minimum variance unbiased estimator of the inverse Gaussian variance. *Comm. Stat.* **19**, 2509-2516.

Hung, H. N. and Wong, W. H. (1996). Average likelihood. Technical Report, University of Chicago.

Pitman, E. J. G. (1937). The closest estimates of statistical parameters. *Proc. Camb. Philos. Soc.* **33**, 212-222.

Schrödinger, E. (1915). Zur theorie der fall-und steigversuche an teilchenn mit Brownsche bewegung. *Physikalische Zeitschrift* **16**, 289-295.

Tweedie, M. C. K. (1945). Inverse statistical variates. *Nature* **155**, 453.

Tweedie, M. C. K. (1946). Some statistical properties of inverse Gaussian distribution. *Virginia J. Sci.* **7**, 160-165.

Tweedie, M. C. K. (1957). Statistical properties of inverse Gaussian distribution, I. *Ann. Math. Statist.* **28**, 362-377.

Wald, A. (1947). *Sequential Analysis.* Wiley, New York.

Wasan, M. T. (1968). On an inverse Gaussian process. *Skand. Actuar.* **60**, 69-96.

Institute of Statistics, National Chiao Tung University, Hsinchu 300, Taiwan.

E-mail: hhung@stat.nctu.edu.tw

(Received November 1998; accepted October 1999)