Journal of the Korean
Data & Information Science Society
2006, Vol. 17, No. 3, pp. 963 ~972

Likelihood Based Confidence Intervals for the Common Scale Parameter in the Inverse Gaussian Distributions

Woo Dong Lee[1)] · Kil Ho Cho[2)]
Young Joon Cha[3)] · Jung Hwan Ko[4)]

Abstract

This paper focuses on the likelihood based confidence intervals for two inverse gaussian distributions when the parameter of interest is common scale parameter. Confidence intervals based on signed loglikelihood ratio statistic and modified signed loglikelihood ratio statistics will be compared in small sample through an illustrative simulation study.

Keywords : Confidence intervals, Modified signed log likelihood ratio, Signed log likelihood ratio

1. Introduction

The inverse gaussian distribution has potentially useful applications in a wide variety of fields such as biology, economics, reliability theory, life testing and social science as discussed in Chhikara and Folks (1978,1989) and Seshdri (1999). Tweed (1957a, 1957b) studied many important statistical properties of the inverse gaussian distribution and discussed the similarity between statistical methods based on the inverse gaussian distribution and those based on the normal theory.

The scale parameter of the inverse gaussian distribution plays important role in the analysis of reciprocals (Tweed, 1957a; Fries and Battacharyya, 1983) and regression models (Whitmore 1979).

---

1) Department of Asset Management, Daegu Haany University, Kyungsan, 712-240, Korea. E-mail : wdlee@dhu.ac.kr
2) Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.
3) Department of Information Statistics, Andong National University, Andong, 760-749, Korea.
4) Department of Information Statistics, Andong National University, Andong, 760-749, Korea.

In this paper, our purpose is to make statistical inference on the common scale parameter of two independent inverse gaussian distributions. Specially, we want to construct an approximate confidence interval for common scale parameter based on likelihood-based methods. We propose two likelihood-based methods which are the signed loglikelihood ratio statistic and the modified signed loglikelihood ratio statistic (Barndorff-Nielsen and Cox 1994). It is well known that the modified signed loglikelihood ratio statistic is highly accurate even in small sample.

This article is organized as follows. In Section 2, we introduce the likelihood-based inference methods, that is, the signed loglikelihood ratio statistic and the modified signed loglikelihood ratio statistic which will employ to construct confidence intervals for the common scale parameter. In Section3, we derive the statistics given in Section 2 in details. In Section 4, through an illustrative simulation study, the performance of the confidence interval from two statistics will be compared with respect to estimated coverage probabilities and expected lengths when the sample size is small.

2. Likelihood-based Methods

Let X be an inverse gaussian distribution with parameters λ and μ_1, and its probability density function is given by

$$f(x \mid \mu_1,\lambda)=\sqrt{\frac{\lambda}{2\pi}}\,x^{-\frac{3}{2}}\exp\left\{-\frac{\lambda(x-\mu_1)^2}{2\mu_1^2 x}\right\}, x>0, \lambda>0, \mu_1>0,$$

and let Y be also an inverse gaussian distribution with parameters λ and μ_2, and its probability density function is given by

$$g(y \mid \mu_2,\lambda)=\sqrt{\frac{\lambda}{2\pi}}\,y^{-\frac{3}{2}}\exp\left\{-\frac{\lambda(y-\mu_2)^2}{2\mu_2^2 y}\right\}, x>0, \lambda>0, \mu_2>0,$$

and assume X and Y are independently distributed.

Let $X_1, X_2, \cdots, X_n$ be a random sample from X and $Y_1, Y_2, \cdots, Y_m$ be a random sample from Y.
Then the likelihood function of λ, μ_1 and μ_2 is given by

$$L(\lambda,\mu_1,\mu_2) = \left(\frac{\lambda}{2\pi}\right)^{\frac{n+m}{2}}\left(\prod_{i=1}^{n} x_i^{-3/2}\right)\left(\prod_{j=1}^{m} y_j^{-3/2}\right) \times\exp\left\{-\frac{\lambda}{2\mu_1^2}\sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{x_i}-\frac{\lambda}{2\mu_2^2}\sum_{j=1}^{m}\frac{(y_j-\mu_2)^2}{y_j}\right\}.$$

Let $\Theta=(\lambda,\mu_1,\mu_2)=(\lambda,\omega)$, where $\omega=(\mu_1,\mu_2)$. Here λ is the scalar parameter of interest and ω is a nuisance parameter vector. Then the loglikelihood function of Θ is given by

$$l(\Theta) \propto \frac{n+m}{2}\log(\lambda)-\frac{3}{2}\left(\sum_{i=1}^{n}\log(x_i)+\sum_{j=1}^{m}\log(y_j)\right)-\frac{\lambda}{2\mu_1^2}\sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{x_i} -\frac{\lambda}{2\mu_2^2}\sum_{j=1}^{m}\frac{(y_j-\mu_2)^2}{y_j}$$

The above loglikelihood function can be re-expressed as

$$l(\Theta) \propto \frac{n+m}{2}\log(\lambda)-\frac{3}{2}\left(\sum_{i=1}^{n}\log(x_i)+\sum_{j=1}^{m}\log(y_j)\right)+\lambda\left(\frac{n}{\mu_1}+\frac{m}{\mu_2}\right) -\frac{\lambda}{2}\left(\sum_{i=1}^{n}x_i^{-1}+\sum_{j=1}^{m}y_j^{-1}\right)-\frac{\lambda}{2\mu_1^2}\sum_{i=1}^{n}x_i-\frac{\lambda}{2\mu_2^2}\sum_{j=1}^{m}y_j.$$

From this loglikelihood, the maximum likelihood estimates of Θ are given by

$$\widehat{\mu_1}=\sum_{i=1}^{n}\frac{x_i}{n},\quad \widehat{\mu_2}=\sum_{j=1}^{m}\frac{y_j}{m}\quad\text{and}\quad \widehat{\lambda}=\frac{n+m}{\sum_{i=1}^{n}(x_i^{-1}-\overline{x^{-1}})+\sum_{j=1}^{m}(y_j^{-1}-\overline{y^{-1}})}.$$

For fixed λ, the constrained maximum likelihood estimates of ω are

$$\widetilde{\mu_1}(\lambda)=\overline{x}\ \text{and}\ \widetilde{\mu_2}(\lambda)=\overline{y}.$$

Denote it as $\widetilde{\omega}(\lambda)=(\widetilde{\mu_1}(\lambda),\widetilde{\mu_2}(\lambda))$.

One can make statistical inference about λ based on the signed loglikelihood ratio statistics given by

$$r\equiv r(\lambda)=sgn(\widehat{\lambda}-\lambda)\{l(\widehat{\lambda},\widehat{\omega})-l(\lambda,\widetilde{\omega}(\lambda))\}^{1/2}. \tag{1}$$

It is well known that this statistic r distributed as standard normal distribution with asymptotic error $O(n^{-1/2})$. That is r is the first-order approximation. For testing $H_0:\lambda=\lambda_0$, a two sided p-value can be obtained from

$$p\text{-value}=2P\{r>|r_0|\}\approx 2[1-\Phi(|r_0|)], \tag{2}$$

where $\Phi(\cdot)$ is the distribution function of standard normal distribution and $r_0=r(\lambda_0)$ is the observed values of r under the null hypothesis. Furthermore, the approximate $100(1-\alpha)\%$ confidence interval for λ can be obtained from

$$\{\lambda : |r(\lambda)| \le z_{\alpha/2}\}, \tag{3}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution.

Generally, the first-order approximation is not accurate especially when the sample size is small. There are several ways to improve the accuracy of the approximation by adjusting the signed loglikelihood ratio statistics r.
Among the others, the modified signed loglikelihood ratio statistic, r^*, developed by Barndorff-Nielsen (1986, 1991), is quite accurate approximation. In fact, this is the third-order approximation and is given by

$$r^* \equiv r^*(\lambda) = r(\lambda) + r(\lambda)^{-1}\log\left\{\frac{u(\lambda)}{r(\lambda)}\right\}, \tag{4}$$

where $u(\lambda)$ is a statistic given by

$$u(\lambda) = \frac{|\, l_{;\hat{\theta}}(\hat{\lambda},\hat{\omega}) - l_{;\hat{\theta}}(\lambda,\tilde{\omega}(\lambda)) \;\; l_{\omega;\hat{\theta}}(\lambda,\tilde{\omega}(\lambda))\,|}{\{|j_{\theta\theta}(\hat{\lambda},\hat{\omega})|\;|j_{\omega\omega}(\lambda,\tilde{\omega}(\lambda))|\}^{1/2}}, \tag{5}$$

where the sample-space derivatives are defined by

$$l_{;\hat{\theta}}(\lambda,\omega) = \frac{\partial}{\partial\hat{\theta}}\, l(\lambda,\omega;\hat{\theta}), \tag{6}$$

the mixed derivatives as

$$l_{\omega;\hat{\theta}}(\lambda,\omega) = \frac{\partial}{\partial\omega}\, l_{;\hat{\theta}}(\lambda,\omega), \tag{7}$$

and $j_{\theta\theta}(\hat{\lambda},\hat{\omega})$ is the observed information matrix and $j_{\omega\omega}(\lambda,\tilde{\omega}(\lambda))$ is the observed nuisance information matrix (Barndorff-Nielsen, 1991). This statistic r^* also distributed as standard normal distribution with asymptotic error $O(n^{-3/2})$. One can also obtain the p-value of two sided test for testing $H_0 : \lambda = \lambda_0$ based on r^* as

$$p\text{-value} = 2P\{r^* > |r_0^*|\} \approx 2[1 - \Phi(|r_0^*|)], \tag{8}$$

which will be more accurate than (2). And the approximate $100(1-\alpha)\%$ confidence interval can be obtained from

$$\{\lambda : |r^*(\lambda)| \le z_{\alpha/2}\}. \tag{9}$$

3. Derivation of the Signed Log Likelihood Ratio Statistic and the Modified Signed Log Likelihood Ratio Statistic

Now, the signed log likelihood ratio statistic given in (1) can be easily obtained from the log likelihood function. But the modified signed log likelihood ratio is rather difficult.

Since the two sample inverse gaussian model is a full rank exponential model, the log likelihood function based on the sample data is only related to a minimum sufficient statistic

$$t = (t_1, t_2, t_3),$$

where

$$t_1 = \sum_{i=1}^{n} x_i^{-1} + \sum_{j=1}^{m} y_j^{-1}, \quad t_2 = \sum_{i=1}^{n} x_i \text{ and } \quad t_3 = \sum_{u=1}^{m} y_j.$$

The loglikelihood function can be rewritten as

$$l(\lambda, \omega; t) \propto \frac{n+m}{2}\log(\lambda) - \frac{3}{2}\left(\sum_{i=1}^{n}\log(x_i) + \sum_{j=1}^{m}\log(y_j)\right) + \lambda\left(\frac{n}{\mu_1} + \frac{m}{\mu_2}\right) - \frac{\lambda}{2}t_1 - \frac{\lambda}{2\mu_1^2}t_2 - \frac{\lambda}{2\mu_2^2}t_3.$$

There is one-to-one transformation between the maximum likelihood estimator $\widehat{\Theta} = (\widehat{\lambda}, \widehat{\mu_1}, \widehat{\mu_2})$ and $t = (t_1, t_2, t_3)$, and the transformation Jacobian matrix is $\partial\widehat{\Theta}/\partial t$. Hence the sample space derivatives with respect to $\widehat{\Theta}$ in the formula (5) for u can be derived based on the sample space derivatives with respect to t. And using the identity $j_{\Theta\Theta}(\widehat{\Theta}) = l_{\Theta;\widehat{\Theta}}(\widehat{\Theta})$ (Barndorff-Nielsen and Cox, 1994) and by canceling the determinant of the transformation Jacobian matrix, one can show that u reduces to the following form:

$$u(\lambda) = \frac{\left| l_{;t}(\lambda,\omega) - l_{;t}(\lambda, \widetilde{\omega}(\lambda)) \; l_{\omega;t}(\lambda, \widetilde{\omega}(\lambda)) \right|}{\left| l_{\Theta;t}(\widehat{\lambda}, \widehat{\omega}) \right|} \left\{ \frac{\left| j_{\Theta\Theta}(\widehat{\lambda}, \widehat{\omega}) \right|}{\left| j_{\omega\omega}(\lambda, \widetilde{\omega}(\lambda)) \right|} \right\}^{1/2},$$

where the sample space derivatives $l_{;t}(\Theta) = \partial l(\Theta; t)/\partial t$ and the mixed derivatives $l_{\omega;t}(\Theta) = \partial^2 l(\Theta; t)/\partial\omega\partial t$ are given by

$$l_{;t}(\Theta)=\left(-\frac{\lambda}{2},-\frac{\lambda}{2\mu_1^2},-\frac{\lambda}{2\mu_2^2}\right)^T \text{ and } l_{\omega;t}(\Theta)=\begin{pmatrix}0 & \frac{\lambda}{\mu_1^3} & 0\\ 0 & 0 & \frac{\lambda}{\mu_2^3}\end{pmatrix}$$

respectively. The observed information matrix $j_{\Theta\Theta}(\hat{\lambda},\hat{\omega})$ is given by

$$j_{\Theta\Theta}(\hat{\lambda},\hat{\omega})=\begin{pmatrix}\frac{n+m}{2\hat{\lambda}^2} & \frac{n}{\hat{\mu}_1^2}-\frac{t_2}{\hat{\mu}_1^3} & \frac{m}{\hat{\mu}_2^2}-\frac{t_3}{\hat{\mu}_2^3}\\ \frac{n}{\hat{\mu}_1^2}-\frac{t_2}{\hat{\mu}_1^3} & \frac{3\hat{\lambda}t_2}{\hat{\mu}_1^4}-\frac{2n\hat{\lambda}}{\hat{\mu}_1^3} & 0\\ \frac{m}{\hat{\mu}_2^2}-\frac{t_3}{\hat{\mu}_2^3} & 0 & \frac{3\hat{\lambda}t_3}{\hat{\mu}_2^4}-\frac{2m\hat{\lambda}}{\hat{\mu}_2^3}\end{pmatrix}.$$

And the mixed derivative matrix $l_{\Theta;t}(\hat{\Theta})$ is given by

$$l_{\Theta;t}(\hat{\Theta})=\begin{pmatrix}-\frac{1}{2} & 0 & 0\\ -\frac{1}{2\hat{\mu}_1^2} & \frac{\hat{\lambda}}{\hat{\mu}_1^3} & 0\\ -\frac{1}{2\hat{\mu}_2^2} & 0 & \frac{\hat{\lambda}}{\hat{\mu}_2^3}\end{pmatrix}.$$

Finally, the observed nuisance information matrix $j_{\omega\omega}(\lambda,\tilde{\omega}(\lambda))$ is given by

$$j_{\omega\omega}(\lambda,\tilde{\omega}(\lambda))=\begin{pmatrix}\frac{3\lambda t_2}{\tilde{\mu}_1(\lambda)^4}-\frac{2n\lambda}{\tilde{\mu}_1(\lambda)^3} & 0\\ 0 & \frac{3\lambda t_3}{\tilde{\mu}_2(\lambda)^4}-\frac{2m\lambda}{\tilde{\mu}_2(\lambda)^3}\end{pmatrix}.$$

Therefore, the statistic $u(\lambda)$ can be calculated using the above equations. Hence $r^*(\lambda)$ can be obtained from $r(\lambda)$ and $u(\lambda)$. Now, we are ready to obtain confidence intervals for λ using (3) and (9).

4. An Illustrative Simulation Study

In this section, we will give an example and some simulation results. The aim of our simulation is to compare the estimated coverage probability, coverage error, upper and lower error probability, and average lengths for r and r^*. The coverage error is defined as the absolute difference between the estimated coverage probability and nominal value (0.90). And upper and lower error probability is the percentage of the intervals falling above and below the true parameter.

Example. The following two groups of data are artificially generated from two independent inverse gaussian distribution with parameters $IG(5,2)$ and $IG(0.5,2)$, respectively. The data are as follows:

X	2.610	3.302	1.121	12.769	0.706			
Y	0.164	0.572	0.201	0.896	0.390	0.562	0.285	0.668

From the above data, the maximum likelihood estimates of θ were $\widehat{\mu_1}=4.102$, $\widehat{\mu_2}=0.467$, $\hat{\lambda}=1.6178$. The 95% confidence interval for λ based on r was (0.672, 3.199) and r^* was (0.482, 2.745). The length of the 95% confidence interval based on r and r^* were 2.528 and 2.263, respectively. The two intervals contain the true parameter and the length of the interval based on r^* is shorter than that based on r.

We perform simulation to estimate the coverage probability, coverage error, upper and lower error probability and average length. We assume that sample sizes are (n,m)=(5,5),(10,5),(5,10),(10,10),(15,10), (10,15) and (15,15). And (λ, μ_1, μ_2) =(3,4,3),(0.5,4,3) and (1,4,3). We have generated 10,000 independent random samples. The simulation results are given from Table 1 to Table 3.

From tables, we observe that the coverage probabilities are lower than the nominal level 0.90, or equivalently, the coverage errors are relatively large for the interval based on the singed log likelihood statistic r. Moreover, the biases of the upper error probabilities of r are large and they extremely asymmetric in small sample size.

In contrast, the confidence interval based on r^* gives nearly the exact coverage probabilities. Furthermore, its upper and lower error probabilities are close to the nominal error levels and they are much more symmetric than those of the interval based on r.

The average length of the interval based on r^* is shorter than that of the interval based on r.

Overall, based on the comparison criteria examined in the above, the confidence interval based on $r*$ performs much better than the confidence interval based on r.

We have tried several parameter values different from Table 1 to Table 3, but their results were almost same.

<Table 1> Coverage probabilities, coverage errors, error probabilities and average lengths of two sided 90% confidence intervals for $r(\lambda)$ and $r^*(\lambda)$ $(\lambda=3, \mu_1=4, \mu_2=3)$

Sample size	Methods	Coverage Probability	Coverage Error	Upper Error Probability	Lower Error Probability	Average Length
5,5	r r^*	0.819 0.901	0.081 0.001	0.167 0.052	0.014 0.047	7.413 6.392
10,5	r r^*	0.848 0.902	0.052 0.002	0.134 0.052	0.018 0.046	4.940 4.485
5,10	r r^*	0.841 0.898	0.059 0.002	0.141 0.055	0.018 0.047	4.993 4.533
10,10	r r^*	0.857 0.899	0.043 0.001	0.125 0.055	0.018 0.046	3.962 3.688
15,10	r r^*	0.865 0.896	0.035 0.004	0.111 0.052	0.024 0.052	3.348 3.163
10,15	r r^*	0.871 0.900	0.029 0.000	0.105 0.049	0.024 0.051	3.329 3.145
15,15	r r^*	0.879 0.907	0.021 0.007	0.098 0.044	0.023 0.049	2.951 2.814

<Table 2> Coverage probabilities, coverage errors, error probabilities and average lengths of two sided 90% confidence intervals for $r(\lambda)$ and $r^*(\lambda)$ $(\lambda=0.5, \mu_1=4, \mu_2=3)$

Sample size	Methods	Coverage Probability	Coverage Error	Upper Error Probability	Lower Error Probability	Average Length
5,5	r r^*	0.816 0.897	0.084 0.003	0.169 0.053	0.015 0.050	1.234 1.064
10,5	r r^*	0.855 0.904	0.045 0.004	0.128 0.048	0.017 0.048	0.820 0.745
5,10	r r^*	0.845 0.901	0.055 0.001	0.138 0.053	0.017 0.046	0.830 0.754
10,10	r r^*	0.853 0.896	0.047 0.004	0.127 0.056	0.020 0.048	0.661 0.615
15,10	r r^*	0.867 0.900	0.033 0.000	0.110 0.050	0.023 0.050	0.558 0.527
10,15	r r^*	0.870 0.899	0.030 0.001	0.106 0.050	0.024 0.051	0.555 0.525
15,15	r r^*	0.882 0.908	0.018 0.008	0.096 0.045	0.022 0.047	0.491 0.469

<Table 3> Coverage probabilities, coverage errors, error probabilities and average lengths of two sided 90% confidence intervals for $r(\lambda)$ and $r^*(\lambda)$
($\lambda=1, \mu_1=4, \mu_2=3$)

Sample size	Methods	Coverage Probability	Coverage Error	Upper Error Probability	Lower Error Probability	Average Length
5,5	r	0.816	0.084	0.171	0.013	2.485
	r^*	0.897	0.003	0.055	0.048	2.143
10,5	r	0.853	0.047	0.130	0.017	1.640
	r^*	0.906	0.006	0.047	0.047	1.489
5,10	r	0.843	0.057	0.140	0.017	1.663
	r^*	0.901	0.001	0.054	0.045	1.510
10,10	r	0.853	0.047	0.128	0.019	1.321
	r^*	0.898	0.002	0.055	0.047	1.230
15,10	r	0.866	0.034	0.111	0.023	1.116
	r^*	0.899	0.001	0.051	0.050	1.055
10,15	r	0.869	0.031	0.107	0.024	1.110
	r^*	0.897	0.003	0.051	0.052	1.049
15,15	r	0.881	0.019	0.097	0.022	0.984
	r^*	0.904	0.004	0.047	0.049	0.938

References

1. Barndorff-Nielsen, O. E. (1986). Inference on Full and Partial Parameters, based on the Standardized Signed Log Likelihood Ratio, Biometrika, 73, 307-322.
2. Barndorff-Nielsen, O. E. (1991). Modified Signed Log Likelihood Ratio, Biometrika, 78, 557-563.
3. Barndorff-Nielsen, O. E. and Cox, D. R. (1994). Inference and Asymptotics. London : Chapman and Hall.
4. Chhikara, R. S. and Folks, L. (1978). The Inverse Gaussian Distribution and Its Statistical Application-A Review. Journal of Royal Statistical Society, B, 40, 263-289.
5. Chhikara, R. S. and Folks, L. (1989). The Inverse Gaussian Distribution; Theory, Methodology and Applications, Marcel Dekker, New York.
6. Fries, A. and Bhattacharyya, G. K. (1983). Analysis of Two-Factor Experiments under an Inverse Gaussian Model. Journal of the American Statistical Association, 78, 820-826.
7. Seshdri, V. (1999). The Inverse Gaussian Distribution: Statistical Theory and Applications, Springer, New York.
8. Tweed, M. C. K. (1957a). Statistical Properties of Inverse Gaussian

Distributions I. The Annals of Mathematical Statistics, 28, 362-377.
9. Tweed, M. C. K. (1957a). Statistical Properties of Inverse Gaussian Distributions II. The Annals of Mathematical Statistics, 28, 696-705.
10. Whitmore, G. A. (1979). An Inverse Gaussian Model for Labour Turnover, Journal of Royal Statistical Society, A, 142, 468-478.

[received date : Jun. 2006, accepted date : Jul. 2006]