Ann. Inst. Statist. Math.
Vol. 41, No. 2, 279-287 (1989)

CONDITIONAL INFORMATION FOR AN INVERSE GAUSSIAN DISTRIBUTION WITH KNOWN COEFFICIENT OF VARIATION*

KATUOMI HIRANO[1] AND KŌSEI IWASE[2]

[1] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan*
[2] *Department of Applied Mathematics, Faculty of Engineering, Hiroshima University, Higashi-Hiroshima 724, Japan*

(Received June 30, 1988; revised October 4, 1988)

Abstract. Conditional inference about a mean of an inverse Gaussian distribution with known coefficient of variation is discussed. For a random sample from the distribution, sufficient statistics with respect to the mean parameter include an ancillary statistic. The effects of conditioning on the ancillary statistic are investigated. It is shown that the model provides a good illustration of R. A. Fisher's recommendation concerning use of the observed second derivative of the log likelihood function in normal approximations.

Key words and phrases: Ancillary statistic, coefficient of variation, conditional inference, inverse Gaussian distribution, sample information.

1. Introduction

Let $X_1, X_2, \dots, X_n$ be independently distributed with common probability density function $f_\theta(x)$. Here θ is an unknown parameter. If $\hat{\theta}$ denotes the maximum likelihood estimator of θ, then standard theory shows that $\hat{\theta}$ is approximately $N(\theta, i_n^{-1})$ where i_n is the expected Fisher information. We denote by $l_n''(\theta)$ the second derivative of the log likelihood function $l_n(\theta)$ where $l_n(\theta) = \log \Pi f_\theta(x_i)$. Fisher (1925) suggested that information not contained in the maximum likelihood estimator $\hat{\theta}$ could be partially recovered by $l_n''(\hat{\theta})$, and essentially recommended that $l_n''(\hat{\theta})$ be treated as an ancillary statistic for which

$$\text{var}\,\{\hat{\theta} \mid l_n''(\hat{\theta})\} \doteqdot -\{l_n''(\hat{\theta})\}^{-1} . \tag{1.1}$$

*This work was started while Kōsei Iwase was visiting the Institute of Statistical Mathematics in Spring, 1987, and was partly supported by the ISM Cooperative Research Program (88-ISM·CRP-7), and by Scientific Research Fund No. 62540173 from the Ministry of Education, Science and Culture of Japan.

Efron and Hinkley (1978) proved that in large samples $\hat{\theta}$ has an approximate normal distribution conditional on an ancillary statistic, with normal variance equal to the reciprocal of the observed Fisher information. In other words, they showed that in translation families (1.1) holds. Note that the observed Fisher information is, in general, an approximate ancillary statistic. Hinkley (1977) demonstrated that the $N(\mu, c^2\mu^2)$ model with $\theta = \log\mu$ ($\mu > 0$) provides an excellent illustration of Fisher's recommendation, the numerical agreement being remarkable in small samples. Under the assumption $\mu > 0$, he gave both the unconditional and conditional Fisher information about $\theta = \log\mu$ and numerical illustration of the ratio (conditional information)/(unconditional information) when $n = 10$ and $c = 1$. Here the ancillarity of $l_n''(\hat{\theta})$ depends on the choice of parameterization and $\mu > 0$. It would seem that the assumption $\mu > 0$ is unnatural. In this paper we introduce another model where Fisher's result can be verified exactly. The purpose is to justify (1.1) for the model. In the next section, we reexamine the case $N(\mu, c^2\mu^2)$ with known c and $\mu > 0$. In Section 3 we investigate the case that X_i obeys the inverse Gaussian distribution with unknown mean μ and known coefficient of variation $c > 0$, denoted by $I(\mu, c^{-2}\mu)$. It is shown that the $I(\mu, c^{-2}\mu)$ model with $\theta = \log\mu$ also provides a good illustration of Fisher's recommendation (1.1).

2. Normal distribution $N(\mu, c^2\mu^2)$ with known c

Let $X_1, X_2, \ldots, X_n$ be independent random variables from $N(\mu, c^2\mu^2)$, where c is a known positive constant and $0 < \mu < \infty$. The minimal sufficient statistic is (T_1, T_2) where $T_1 = \sum_{i=1}^{n}(X_i^2/n)$, $T_2 = T_1/\bar{X}_A^2$ and $\bar{X}_A = (1/n)\sum_{i=1}^{n} X_i$. The distribution of T_2 does not depend on μ, then T_2 is an ancillary statistic. Therefore the next proposition holds.

PROPOSITION 2.1. (Hinkley (1977)) *The unconditional information about $\theta = \log\mu$ is $i_n = n(2 + c^{-2})$, and conditionally on $T_2^{1/2} = a$, the conditional information is*

$$i_n(a) = n\left\{2 + (ac)^{-1}n^{-1/2}\frac{I_n(a^{-1}c^{-1}\sqrt{n})}{I_{n-1}(a^{-1}c^{-1}\sqrt{n})}\right\},$$

where

$$I_m(b) = \int_0^\infty x^m \exp\left\{-\frac{(x-b)^2}{2}\right\}dx\,.$$

It is easy to see that the conditional information $i_n(a)$ can be rewritten as

$$i_n(a) = n\left\{2 + \frac{\sqrt{2}\Psi\left(\frac{n+1}{2}, \frac{1}{2}; \frac{n}{2a^2c^2}\right)}{\Psi\left(\frac{n+1}{2}, \frac{3}{2}; \frac{n}{2a^2c^2}\right)}\right\},$$

where $\Psi(\alpha, \gamma; z)$ is a confluent hypergeometric function, defined by the formula 6.5.(2) in Bateman (1953).

The behavior of $\Psi(\alpha, \gamma; z)$ as $\alpha \to \infty$ and $z \to \infty$, while γ is bounded, is mentioned by Bateman ((1953), pp. 280–282). Using the results, we obtain that, for $n \to \infty$,

$$i_n(a) \doteqdot n\left[2 + c^{-2}\left\{\left(\frac{1}{2a^4} + \frac{2c^2}{a^2}\right)^{1/2} - \left(\frac{1}{2a^4}\right)^{1/2}\right\}\right].$$

The maximum likelihood estimator is $\hat{\theta} = \log \hat{\mu}$, where

$$\hat{\mu} = \frac{1}{2} c^{-2}\bar{X}_A\{(1 + 4a^2c^2)^{1/2} - 1\},$$

then

$$-l_n''(\hat{\theta}) = n\left\{2 + \frac{(1 + 4a^2c^2)^{1/2} + 1}{2a^2c^2}\right\}.$$

Hence we have the next proposition.

PROPOSITION 2.2.
(1) *If* $a^2 = 2c^2$, $a \gg 1$ *and* $n \to \infty$, *then* $i_n \doteqdot i_n(a)$.
(2) *If* $a^2 = 1 + c^2$, *then* $i_n = -l_n''(\hat{\theta})$.
(3) *If* $a^2c^2 = 4 + 3\sqrt{2}$ *and* $n \to \infty$, *then* $i_n(a) \doteqdot -l_n''(\hat{\theta})$.

Hinkley (1977) assumes that $c = 1$ and $\mu > 0$, and then states that the event $\bar{X}_A \leq 0$ has probability $\Phi(-\sqrt{n}c^{-1})$ which one can ignore in discussing large-sample properties, but the probability cannot be ignored for large c. When the sample moments agree with the model property that the variance is equal to $c^2(\text{mean})^2$, i.e., when $a^2 = c^2 + 1$, then $i_n(a) \doteqdot i_n$. Also Propositions 2.2. (1) and (2) show that if $c = 1$, then $i_n(a) \doteqdot -l_n''(\hat{\theta})$ for large n. Hence Fisher's recommendation for this model depends on the assumption. In the next section we introduce a new model and discuss the recommendation.

3. Inverse Gaussian distribution $I(\mu, c^{-2}\mu)$ with known c

In Tweedie's (1957) notation the probability density function of a random variable X distributed as an inverse Gaussian with parameters μ and λ, denoted by $X \sim I(\mu, \lambda)$, is given by

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

where μ and λ are positive. The mean and variance of X are given by μ and $\mu^3\lambda^{-1}$, respectively. The coefficient of variation of X is $(\mu\lambda^{-1})^{1/2}$. If the coefficient of variation is a known number $c > 0$, then the probability density function is reduced to

$$f(x) = \left(\frac{\mu}{2\pi c^2 x^3} \right)^{1/2} \exp\left\{ -\frac{1}{2c^2}(\mu^{-1}x + \mu x^{-1} - 2) \right\}.$$

Let M be a family of inverse Gaussian distributions $I(\mu, c^{-2}\mu)$ with known constant c. Then the family M is incomplete. The family of inverse Gaussian distributions $I(\mu, \lambda)$ specified by the two parameters μ and λ forms an exponential family S. It is clear that M is a curved exponential family, imbedded in the space S. This property is analogous to that in the normal case. Note that M forms a hyperbola in S, but a parabola in the normal case.

Let $X_1, X_2, \ldots, X_n$ be independent random variables from $I(\mu, c^{-2}\mu)$ with known c. The log likelihood function is, apart from a constant, given by

$$l_n(\mu) = \frac{n}{2}\log\mu - \frac{n\mu}{2c^2}\bar{X}_H^{-1} - \frac{n}{2c^2\mu}\bar{X}_A\,,$$

where

$$\bar{X}_H = \left\{ \frac{1}{n}\sum_{i=1}^{n} X_i^{-1} \right\}^{-1}.$$

The incomplete sufficient statistic for μ is $(\bar{X}_A, \bar{X}_H)$. Set $T_1 = \bar{X}_H^{-1}$ and $T_2 = \bar{X}_H^{-1}\bar{X}_A$, then the pair of T_1 and T_2 is also the sufficient statistic for μ. Tweedie (1957) proved that

$$\bar{X}_A \sim I(\mu, nc^{-2}\mu) \qquad \text{and} \qquad nc^{-2}\mu(\bar{X}_H^{-1} - \bar{X}_A^{-1}) \sim \chi^2(n-1)\,,$$

and furthermore that $\bar{X}_A$ and $\bar{X}_H^{-1} - \bar{X}_A^{-1}$ are stochastically independent. By using these results, the joint probability density function of (T_1, T_2) is

$$f_{T_1,T_2}(t_1,t_2) = \frac{(n\mu c^{-2})^{n/2}\exp(nc^{-2})}{2^{n/2}\pi^{1/2}\Gamma\left(\frac{1}{2}n-\frac{1}{2}\right)} t_2^{-3/2}(1-t_2^{-1})^{n/2-3/2}t_1^{n/2-1}$$

$$\cdot \exp\left\{-\frac{n\mu}{2c^2}t_1 - \frac{n}{2c^2\mu}t_2t_1^{-1}\right\},$$

for $t_1 \geq 0$ and $t_2 \geq 1$. Integrating the above formula with respect to t_1, we get the probability density function of T_2 as follows;

$$f_{T_2}(t_2) = \frac{(nc^{-2})^{n/2}\exp(nc^{-2})}{2^{n/2-1}\pi^{1/2}\Gamma\left(\frac{1}{2}n-\frac{1}{2}\right)} t_2^{n/4-3/2}(1-t_2^{-1})^{n/2-3/2}K_{n/2}(nc^{-2}t_2^{1/2}),$$

$$t_2 \geq 1,$$

where $K_\nu(z)$ is a modified Bessel function of the second kind (e.g., see Abramowitz and Stegun (1964)). It is clear that $f_{T_2}(t_2)$ does not depend on μ, then $T_2 = \bar{X}_H^{-1}\bar{X}_A$ is an ancillary statistic with respect to μ, and $T_1 = \bar{X}_H^{-1}$ is used as a sufficient statistic in inference for μ conditionally on $T_2 = a$. It is noted that T_2 is the ratio of an arithmetic mean to a harmonic mean, i.e., $T_2 = \bar{X}_A\bar{X}_H^{-1}$. With the aid of the relation

$$f_{T_1,T_2}(t_1,t_2) = f_{T_2}(t_2)\cdot f_{T_1|T_2}(t_1|t_2),$$

we obtain the conditional probability density function of T_1 given $T_2 = a$ as follows;

$$f_{T_1|T_2=a}(t_1|a) = \frac{t_1^{n/2-1}\exp\left\{-\frac{n\mu}{2c^2}t_1 - \frac{na}{2c^2\mu}t_1^{-1}\right\}}{2a^{n/4}\mu^{-n/2}K_{n/2}(nc^{-2}\sqrt{a})}.$$

It is convenient to take the parameter to be $\theta = \log\mu$. From the log likelihood function,

$$l_n(\theta) = \frac{n}{2}\theta - \frac{n\bar{X}_H^{-1}}{2c^2}e^{\theta} - \frac{n\bar{X}_A}{2c^2}e^{-\theta},$$

$$l_n'(\theta) = \frac{n}{2} - \frac{n\bar{X}_H^{-1}}{2c^2}e^{\theta} + \frac{n\bar{X}_A}{2c^2}e^{-\theta},$$

and

$$l_n''(\theta) = -\frac{n\bar{X}_H^{-1}}{2c^2}e^{\theta} - \frac{n\bar{X}_A}{2c^2}e^{-\theta}.$$

Hence, the unconditional information is calculated to be

$$i_n = -E(l_n''(\theta)) = n\left(\frac{1}{2} + c^{-2}\right).$$

Conditionally on $\bar{X}_H^{-1}\bar{X}_A = a$, $l_n''(\theta)$ is rewritten as

$$l_n''(\theta) = -W - \frac{n^2 a}{4c^4}\frac{1}{W},$$

where $W = (n\mu/2c^2)\bar{X}_H^{-1}$, and the conditional information is

$$i_n(a) = E(W \mid T_2 = a) + \frac{n^2 a}{4c^4} E\left(\frac{1}{W} \,\middle|\, T_2 = a\right).$$

The conditional probability density function of W given $T_2 = a$ is presented to be

$$f_{W \mid T_2 = a}(w \mid a) = \frac{(2c^2 n^{-1})^{n/2} w^{n/2-1}}{2a^{n/4} K_{n/2}(nc^{-2}a^{1/2})} \exp\left\{-w - \frac{n^2 a}{4c^4} w^{-1}\right\}.$$

Therefore we obtain the next proposition.

PROPOSITION 3.1. *The unconditional and conditional information about* $\theta = \log\mu$ *are given by*

$$i_n = n\left(\frac{1}{2} + c^{-2}\right)$$

and

$$i_n(a) = n\left\{\frac{1}{2} + c^{-2}\sqrt{a}\,\frac{K_{n/2-1}(nc^{-2}\sqrt{a})}{K_{n/2}(nc^{-2}\sqrt{a})}\right\},$$

respectively.

Remark. The parameter μ of $N(\mu, c^2\mu^2)$ with $\mu > 0$ or $I(\mu, c^{-2}\mu)$ may be a scale one. $\hat{\theta} = \log\hat{\mu}$ is the maximum likelihood estimator of the location parameter $\theta = \log\mu$ $(\mu > 0)$. Thus the conditional density of $\hat{\theta}$ given a is of the translation form. In other words, the $N(\mu, c^2\mu^2)$ and the

$I(\mu, c^{-2}\mu)$ models are nonobvious examples of form (2.2) in Efron and Hinkley (1978). Hence we show that $\mathrm{var}(\hat{\theta}|a) \fallingdotseq -\{l_n''(\hat{\theta})\}^{-1}$ for the $I(\mu, c^{-2}\mu)$ model with $\theta = \log \mu$.

It is easy to see that the conditional information $i_n(a)$ can be represented as

$$i_n(a) = -nc^{-2}\sqrt{a}\,\frac{K'_{n/2}(nc^{-2}\sqrt{a})}{K_{n/2}(nc^{-2}\sqrt{a})}\,.$$

By using uniformly asymptotic expansions of $K_\nu(\nu z)$ and $K'_\nu(\nu z)$ with respect to z when $\nu \to \infty$, it holds that, for large n, the conditional information is

$$\frac{i_n(a)}{n} \fallingdotseq \frac{1}{2}(1 + 4c^{-4}a)^{1/2}\,.$$

For the conditional likelihood, the maximum likelihood estimator is $\hat{\theta} = \log \hat{\mu}$ where

$$\hat{\mu} = \frac{1}{2}c^2\bar{X}_H\{1 + (1 + 4c^{-4}a)^{1/2}\}\,.$$

Thus the following result holds.

PROPOSITION 3.2. *For any sample size n, it holds that*

$$-l_n''(\hat{\theta}) = \frac{n}{2}(1 + 4c^{-4}a)^{1/2}\,.$$

The proposition shows that $-l_n''(\hat{\theta})$ is exactly ancillary. As stated in the Introduction, we discuss the effects of conditioning on the ancillary statistic and Fisher's recommendation. After some calculations we obtain the next proposition.

PROPOSITION 3.3. *The relations among i_n, $i_n(a)$ and $-l_n''(\hat{\theta})$ are*

(1) *If $a = 1 + c^2$, then $i_n \fallingdotseq i_n(a)$ for large n.*
(2) *If $a = 1 + c^2$, then $i_n = -l_n''(\hat{\theta})$ for any $n > 0$.*
(3) *$i_n(a) \fallingdotseq -l_n''(\hat{\theta})$ for large n, any $a \geq 1$ and $c > 0$.*

Efron and Hinkley (1978) showed that in translation families $\mathrm{var}\{\hat{\theta}|a\} \fallingdotseq i_n(a)^{-1}$. Therefore, from Proposition 3.3, if Fisher's claim (1.1) is true, then $-l_n''(\hat{\theta})$ and $i_n(a)$ should agree within $O(1)$ in large samples. For numerical illustration we have computed the ratios $i_n(a)/i_n$ and $-l_n''(\hat{\theta})/i_n$ when $n=10$, $c=0.05$, 0.1, 0.5, 1.0 with a varying over 99% or more of its distribution. Table 1 gives values of the ratios and corresponding values of $T_2=a$. Table 1 shows that even for $n=10$ and unusual values of c (e.g., $c=1.0$) the agreement is remarkable. Fisher's recommendation (1.1) is substantiated. The new $I(\mu, c^{-2}\mu)$ model with $\theta=\log\mu$ provides an excellent illustration of Fisher's recommendation.

Table 1. Conditional Fisher information for θ when $n=10$.

	a	$\frac{i_{10}(a)}{i_{10}}$	$-\frac{l_{10}''(\hat{\theta})}{i_{10}}$
$c=0.05$	1.0	0.999	0.999
	2.0	1.413	1.412
	4.0	1.998	1.998
	6.0	2.447	2.446
$c=0.1$	1.0	0.996	0.995
	2.0	1.408	1.407
	4.0	1.991	1.990
	6.0	2.438	2.437
$c=0.5$	1.0	0.907	0.896
	2.0	1.273	1.262
	4.0	1.792	1.781
	6.0	2.191	2.180
$c=1.0$	1.0	0.772	0.745
	2.0	1.029	1.000
	4.0	1.405	1.374
	6.0	1.698	1.667

Now it is easily shown that the moment generating function of $\hat{\theta}$ given a is

$$\phi(\xi) = (\sqrt{a}\,e^{-\theta})^{\xi}\,\frac{K_{\xi+n/2}(nc^{-2}\sqrt{a})}{K_{n/2}(nc^{-2}\sqrt{a})}\,.$$

Then the exact conditional variance of $\hat{\theta}$ given a is calculated from $\phi(\xi)$ to be

$$\begin{aligned}\operatorname{var}(\hat{\theta}|a) &= \frac{d^2}{dv^2}\,K_v(nc^{-2}\sqrt{a})\Big|_{v=n/2}\{K_{n/2}(nc^{-2}\sqrt{a})\}^{-1} \\ &\quad -\left[\frac{d}{dv}\,K_v(nc^{-2}\sqrt{a})\Big|_{v=n/2}\{K_{n/2}(nc^{-2}\sqrt{a})\}^{-1}\right]^2.\end{aligned}$$

But it is not easy to see the accuracy of the approximation $\operatorname{var}(\hat{\theta}|a) \doteqdot$

$-\{l_n''(\hat{\theta})\}^{-1}$.

Finally, we give the statistical curvature γ_μ, as defined by Efron (1975), of the $I(\mu, c^{-2}\mu)$ model. We calculate $\gamma_\mu^2 = 4c^4/(c^2+2)^3$, taking its maximum $(2/3)^3$ at $c = 2$. In the case of $N(\mu, c^2\mu^2)$, see Efron (1975).

Acknowledgement

The authors would like to thank the referee for his valuable suggestions.

REFERENCES

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions*, Dover Publications, New York.

Bateman, H. (1953). *Higher Transcendental Functions* I, (ed. A. Erdely), McGraw-Hill, New York.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency), *Ann. Statist.*, **3**, 1189–1242.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika*, **65**, 457–487.

Fisher, R. A. (1925). Theory of statistical estimation, *Proc. Camb. Phil. Soc.*, **122**, 700–725.

Hinkley, D. V. (1977). Conditional inference about a normal mean with known coefficient of variation, *Biometrika*, **64**, 105–108.

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