FOURTH AND FIFTH ORDER EFFICIENCY: FISHER INFORMATION

YUTAKA KANO *
Institute of Mathematics, University of Tsukuba
Tsukuba, Ibaraki 305, Japan
E-mail: kano@math.tsukuba.ac.jp

ABSTRACT

Loss of (Fisher) information, whose concept was defined by Fisher, is one of important measures of asymptotic efficiency. Based on Fisher information, this paper studies fourth and fifth order asymptotic efficiency of estimators in a curved exponential family of distributions with a structural parameter vector. In particular, we show that the bias-corrected MLE with a certain bias correction factor is fourth and fifth order efficient in a class of Fisher consistent estimators bias-corrected similarly.

1. Introduction

Efficiency is one of key concepts in the estimation problems in statistics. Asymptotic efficiency is often studied to discuss efficiency of estimators for parameters of models with a non-linear structure. There are mainly three measures or criteria to study asymptotic efficiency, namely loss of information, quadratic risk and concentration probability. Historically, taking loss of information, Fisher(1925) and Rao(1961) successfully proved the third order efficiency of the MLE in a class of the BAN (best asymptotic normal) estimators in a curved one-parameter multinomial distribution, that is, the MLE is the best of the class of first order efficient estimators. This result is often called the Fisher-Rao theorem. Amari (1982) and Eguchi (1983) took a geometric approach to extend the Fisher-Rao theorem to the general curved exponential family with a structural parameter vector. The third order efficiency has been shown to hold under other criteria, quadratic risk and concentration probability as well (e.g., Ghosh and Subramanyam 1974; Efron 1975; Pfanzagl and Wefelmeyer 1978; Akahira and Takeuchi 1981; Amari 1985). Notice that the term "third order " using in this paper means "second order " that is called by Fisher, Rao or Eguchi.

The basic results on third order efficiency have been extended to more general cases, see Takeuchi-Morimune (1985), Taniguchi (1991) and Taniguchi and Watanabe (1994) among others. Ghosh (1994) recently published a well-written review book on the topic.

There are many third order efficient estimators, however. For instance, Rao, Sinha and Subramanyam (1982) and Eguchi (1983) have formed some estimators asymp-

*This work is partially supported by a grant-in-aid from the Ministry of Education.

totically equivalent to the MLE up to order $^{-2}$; estimators obtained by iterating the scoring method twice are third order efficient as well (see e.g., Eguchi 1983). Thus, it is important to develop theory of *more* higher order efficiency to distinguish these third order efficient estimators.

There are a few works on asymptotic efficiency beyond the third order. Kano (1993a, 94, 95) have taken quadratic risk or concentration probability as a measure of efficiency to study fourth and fifth order asymptotic efficiency. Here, we shall study the same topic under *Fisher information* as a measure of efficiency. The loss of information is difference in the Fisher information between an estimator and the original sample, and it requires tedious calculation to get an explicit form. To avoid it, we evaluate the difference in the Fisher information between the MLE and a competitive estimator.

The result obtained here is an extension of the classical Fisher-Rao theorem to the case at *more* higher order. The reason that the proof of the fourth and fifth order efficiency of the MLE under the Fisher information comes after that under quadratic risk or concentration probability is that Fisher information is more intractable than the other two.

2. Curved Exponentials and Approximate Fisher Information

Let $\in \mathbf{R}$ be a parameter vector of interest and let an open set $\Omega(\subset \mathbf{R})$ be a parameter space of . Let $_0 \in \Omega$ be an arbitrarily fixed (inner) point of Ω. For every $\in \mathcal{N}(\ _0)$, a neighborhood of $_0$, assume that random -vectors $_1 \ldots$ are independent and identically distributed according to a (continuous) curved exponential family with density

$$\exp\{(\ )' - ((\ ))\} \qquad (\ )$$

where $(\cdot)$ is a carrier measure on $\mathbf{R}$ (see Amari 1985, Section 4.1) and $(\cdot)$ is an $\mathbf{R}$ -valued measurable function from $\Omega(\subset \mathbf{R})$ smooth in $\in \mathcal{N}(\ _0)$. The real-valued function, $()$, is defined as a normalizing constant $() = \log \int_{\mathrm{R}^p} \exp(' \ )(\ )$. Write $\Theta_1 = \frac{\mathrm{d}}{\mathrm{d}\ '}(\ )$, $\Psi_{10} = \frac{\mathrm{d}}{\mathrm{d}}$ and $\Psi_{11} = \frac{\mathrm{d}^2}{\mathrm{d}\mathrm{d}'}$. It is well-known that

$$[\ ] = \Psi_{10}((\ )) \ (= (\ ) \ \text{say}) \quad \text{and} \quad \mathrm{Cov}[\qquad] = \Psi_{11}((\ )) \ (= \Psi_{11} \ \text{say})$$

and that the Fisher information (matrix) of the original sample of each is expressed as

$$= \Theta_1' \Psi_{11} \Theta_1$$

Assume that Ψ_{11} and are positive definite.

There are two ways to evaluate higher order Fisher information of estimators, one is to make use of the representation in conditional expectations (e.g., Fisher 1925; Rao 1961; Eguchi 1983) and the other is to utilize an Edgeworth expansion (Hosoya 1988). We shall take the latter here.

Let $(^{-})$ be an estimator for and let $(\ |\ )$ be a probability density function of the distribution of $\sqrt{^{-}}(\ \ (^{-})-\ \ )$ w.r.t. the Lebesgue measure on $(\mathbf{R}\ \ \mathbf{B}\ )$, where $\mathbf{B}$ denotes the -dimensional Borel -field. The density function of $\ \ (^{-})$ is then written as $(\sqrt{^{-}}(\ \ (^{-})-\ \ )|\ \ )^{2}$. Let

$$= \text{—} - \frac{1}{\sqrt{^{-}}} \text{—}$$

The Fisher information (matrix) of $\ \ (^{-})$ is expressible in the form:

$$\begin{aligned}
&(\ \ (^{-})) \\
&= \left[\left\{\text{—}\log(\sqrt{^{-}}(\ \ (^{-})-\ \ )|\ \ )^{2}\right\}\ \ \left\{\text{—}\log(\sqrt{^{-}}(\ \ (^{-})-\ \ )|\ \ )^{2}\right\}\ \ '\right] \\
&= \int \frac{\{(\ \ |\ \ )\}\{(\ \ |\ \ )\}'}{(\ \ |\ \ )}
\end{aligned}$$

It is too restrictive to assume that the distribution of $\ \ (^{-})$ permits a density function because usually $\ \ (^{-})$ is defined implicitly as a solution to nonlinear simultaneous equations and the differentiability is guaranteed only in a neighborhood of $(\ \ ) = [^{-}]$. On the other hand, under minimal regularity conditions (see e.g. Bhattacharya and Denker 1990), the distribution of $\sqrt{^{-}}(\ \ (^{-})-\ \ )(=(^{(1)}(^{-})\ldots{}^{()}(^{-}))'$ say) can be evaluated with the help of Edgeworth expansions: For a given positive integer ,

$$\begin{aligned}
&\Pr\{^{(1)}(^{-})\leq{}_{1}\ldots{}^{()}(^{-})\leq\ \ \} \\
&\quad = \sum_{=0}^{-2}\ \ \int_{-\infty}^{1}\cdots\int_{-\infty}\ \ (\ \ \ )(\ \ |\mathbf{0}\ \ ^{-1})_{1}\cdots\ \ +(^{-2}\ \ \ ) \qquad (1)
\end{aligned}$$

where $(\ \ \ )$ is a polynomial in with coefficients that are continuously differentiable in $\in\mathcal{N}(\ \ _{0})$. Let us term

$$(\ \ |\ \ ) = \sum_{=0}^{-2}\ \ (\ \ \ )(\ \ |\mathbf{0}\ \ ^{-1})$$

an approximate density function of the distribution of $\sqrt{^{-}}(\ \ (^{-})-\ \ )$ up to the order $^{-2}$, and the Fisher information (matrix) based on $(\ \ |\ \ )$ *approximate Fisher information (AFI)* of $\ \ (^{-})$. We further expand

$$\begin{aligned}
(\ \ (^{-})) \qquad &= \int \frac{\{\ (\ \ |\ \ )\}\{\ (\ \ |\ \ )\}'}{(\ \ |\ \ )} \\
&= \ {}_{1}+{}^{\frac{1}{2}}{}_{2}+\cdots+{}^{-\frac{}{2}+1}{}_{+1}\ \ +(^{-\frac{}{2}+1})
\end{aligned}$$

and then we shall define asymptotic efficiency on the basis of the expansion above (see Hosoya 1988).

Let $\mathcal{F}$ be a class of estimators for under consideration. An estimator in $\mathcal{F}$ is said to be first order efficient (1OE) iff it maximizes $_1$ among all the estimators in $\mathcal{F}$; a 1OE estimator (in $\mathcal{F}$) is said to be second order efficient (2OE) in $\mathcal{F}$ iff it maximizes $_2$ among all the 1OE estimators (in $\mathcal{F}$); 3OE, 4OE and so fourth are defined consecutively in the same manner.

Let $(\cdot)$ be a measurable function from $\mathbf{R}$ to $\mathbf{R}$, four times continuously differentiable in a neighborhood of $(\ _0)$. Consider, as $\mathcal{F}$, a class of second order bias-corrected version $^*(^-)$ of Fisher consistent estimators:

$$\mathcal{F} = \left\{ \ ^*(^-) = \ (^-) - \frac{1}{} \ ^{(1)}(^-) - \frac{1}{2} \ ^{(2)}(^-) \right| \\ ((\ )) = \quad \text{for every} \quad \in \mathcal{N}(\ _0) \ [\ ^*(^-) - \ ] = (^{-3}) \Big\}$$

The measurable functions $^{(1)}(^-)$ and $^{(2)}(^-)$ are factors for bias correction, continuously differentiable in $^- \in \mathcal{N}((\ _0))$. The specific form of $^{(1)}(^-)$ will be given in (3).

It is well known that $^*(^-)$ is 1OE iff

$$_1 = \left. \frac{(^-)}{^-{}'} \right|_{^-=(\ )} = {}^{-1}\Theta_1'$$

and that the MLE meets the condition. For every estimator in $\mathcal{F}$, $_2 = 0$. Thus, the MLE is 1OE and 2OE. We have also known that the MLE is 3OE (Fisher-Rao theorem).

3. Fourth and Fifth Order Efficiency

Let $= [_1 \dots \ ]'$ and $(\ ) = [\ (\ )]$ be an $_1 \times _2$ matrix-valued continuously differentiable function. Put $\frac{(\ )}{} = \left[\frac{(\ )}{}\right]'$. Define $\frac{\mathrm{d}(\ )}{\mathrm{d}\ '} = \frac{\mathrm{d}}{\mathrm{d}\ '}(\ ) = \left[\frac{(\ )}{_1} \dots \frac{(\ )}{}\right]$ and $\frac{\mathrm{d}}{\mathrm{d}}(\ ) = \left(\frac{\mathrm{d}}{\mathrm{d}\ '}(\ )'\right)'$. The higher order (matrix) derivatives are defined inductively by $\left(\frac{\mathrm{d}}{\mathrm{d}\ '}\right)^{} = \frac{\mathrm{d}}{\mathrm{d}\ '}\left(\frac{\mathrm{d}}{\mathrm{d}\ '}\right)^{-1}$ with $\left(\frac{\mathrm{d}}{\mathrm{d}\ '}\right)^0 = 1$. For some properties of such matrix derivatives, see Kano (1993a,b) (cf. Bentler and Lee 1980; Magnus and Neudecker 1988). Let us simply write $= \overbrace{\otimes \cdots \otimes}^{\text{-fold}}$ -fold right Kronecker (tensor) product of the same matrix . Note $^0 = 1$. Let $^*(^-)$ be in $\mathcal{F}$, and let $\hat{}^*(^-)(\in \mathcal{F})$ be the bias-corrected version of the MLE. Under certain minimal regularity conditions on the population distribution, the distributions of $\hat{}^*(^-) \in \mathcal{F}$ and $^*(^-) \in \mathcal{F}$ permit Edgeworth expansions as in (1) with polynomials $^-(\ \ )$ and $(\ \ )$, respectively. Let $^-(\ |\ )$ and $(\ |\ )$ be the approximate

densities associated with $\hat{}^*(^-)$ and $^*(^-)$. We further write

$$^- (^-) = \left(\frac{\mathrm{d}}{\mathrm{d}^{-\prime}}\right) \quad \hat{}(^-) \quad ^- = ^- ((\ \ ))$$
$$(^-) = \left(\frac{\mathrm{d}}{\mathrm{d}^{-\prime}}\right) \quad (^-) \quad = ^- ((\ \ ))$$

Note that $(\cdot)$ and $^-(\cdot)$ are defined on a neighborhood of $(\ _0)$, and they are extended to be measurable functions on the entire $\mathbf{R}$.

For every 1OE estimator $^*(^-) \in \mathcal{F}$, Kano (1995 formula (4.11)) derived

$$\begin{aligned} ^- (\ |\ ) - (\ |\ ) &= \frac{1}{4}(\bar{}_3(\ \ ) - {}_3(\ \ ))\ (\ |\mathbf{0}\ ^{-1}) + \left(\frac{1}{-}\right) \\ &= \frac{1}{4}\mathrm{tr}[(_2 - \bar{}_2)\Psi_{11}^2\ (_2 - \bar{}_2)'(\ - \ \ ')] \\ &\quad \cdot (\ |\mathbf{0}\ ^{-1}) + \left(\frac{1}{-}\right) \end{aligned} \tag{2}$$

Based on this, we can show a necessary and sufficient condition that $^*(^-)$ be 3OE in $\mathcal{F}$ (in terms of Fisher information) is that ${}_3 = \bar{}_3$. The proof is exactly the same as that for 5OE which will follows. Thus, we reach a known result that the MLE is 3OE. The necessary and sufficient condition on the derivatives of estimators may not be known.

The first order bias is $\frac{1}{2}{}_2\Psi_{20}$. Here we take

$$^{(1)}(^-) = \frac{1}{2}{}_2(^-)\Psi_{20}((\ \ (^-))) \tag{3}$$

In the case, Kano (1994 formula (5.9)) derived, for every 3OE estimator $^*(^-) \in \mathcal{F}$

$$\begin{aligned} ^- (\ |\ ) - (\ |\ ) &= \frac{1}{12^2}(\bar{}_5(\ \ ) - {}_5(\ \ ))\ (\ |\mathbf{0}\ ^{-1}) + \left(\frac{1}{2}\right) \\ &= \frac{1}{12^2}\mathrm{tr}[(_3 - \bar{}_3)\Psi_{11}^3\ (_3 - \bar{}_3)'(\ - \ \ ')] \\ &\quad \cdot (\ |\mathbf{0}\ ^{-1}) + \left(\frac{1}{2}\right) \end{aligned} \tag{4}$$

Theorem. Let the MLE $\hat{}^*(^-) \in \mathcal{F}$ and $^*(^-) \in \mathcal{F}$ be bias-corrected versions with the factors defined in (3). Assume that $^*(^-)$ is 3OE. Then, $^*(^-)$ is always 4OE in $\mathcal{F}$; and $^*(^-)$ is 5OE in $\mathcal{F}$ if and only if ${}_3 = \bar{}_3$. In particular, the bias-corrected version of the MLE is 5OE in $\mathcal{F}$.

Proof. We simply denote, by $^-$ and , $^-(\ |\ )$ and $(\ |\ )$. Consider the difference in between $\hat{}^*(^-)$ and $^*(^-)$:

$$(\hat{}^*(^-)) - (\ ^*(^-))$$

$$
\begin{aligned}
&= \int \frac{(\bar{\ })(\bar{\ })'}{\bar{\ }} \quad - \int \frac{(\ )(\ )'}{} \\
&= \int \frac{(\ - \bar{\ })(\bar{\ })(\bar{\ })'}{\bar{\ }} \quad + \int \frac{(\bar{\ } - \ )(\ )'}{} \\
&\qquad + \int \frac{(\ )(\bar{\ } - \ )'}{} \qquad (5)
\end{aligned}
$$

We need to evaluate the terms up to $\left(\frac{1}{2}\right)$. Noting (4), we can replace $= — - \frac{1}{\sqrt{}} —$ with $—$ and regard, as $(\ |\mathbf{0}\ ^{-1})$, all the terms $\bar{\ }$ and but those in the difference $\bar{\ } - \ $. Let $= (_3 - \bar{_3}\ )\Psi^3_{11}\ (_3 - \bar{_3}\ )'$. Then, the first term in the LHS of (5) is equal to

$$
\begin{aligned}
&\frac{1}{12^2}\int \mathrm{tr}[\ (\quad ' - \ )](\quad)(\quad)'\ (\ |\mathbf{0}\ ^{-1}) \quad + \left(\frac{1}{2}\right) \\
&\qquad = \frac{1}{12^2}\int \mathrm{tr}[\ (\quad ' - \ )]\quad '\ (\ |\mathbf{0}\quad) \quad + \left(\frac{1}{2}\right) \\
&\qquad = \frac{1}{6^2}\qquad + \left(\frac{1}{2}\right)
\end{aligned}
$$

The second and third terms in (5) vanish because $(\bar{\ } - \ )$ and are basically the third order and the first order multivariate Hermite polynomials, respectively, and they are orthogonal to each other. As a consequence, we obtain

$$
(\hat{\ }^*(\bar{\ })) - (\ ^*(\bar{\ })) = \frac{1}{6^2}\ \ (_3 - \bar{_3}\ )\Psi^3_{11}\ \ (_3 - \bar{_3}\ )' \ + \left(\frac{1}{2}\right) \qquad (6)
$$

from which the necessary and sufficient condition on 5OE in the Theorem follows. The equation (6) also implies that the terms of $\left(\frac{1}{\sqrt{}}\right)$ coincide between $(\hat{\ }^*(\bar{\ }))$ and $(\ ^*(\bar{\ }))$ for any 3OE $^*(\bar{\ })$ in $\mathcal{F}$. This shows that 3OE implies 4OE.

Staring from (2), we can iterate exactly the same process to prove the 2OE and 3OE of 1OE estimators in $\mathcal{F}$. While the proof is alternative to the existing proofs of 3OE based on the representation in conditional expectation (e.g., Rao 1961; Eguchi 1983), it requires more calculations than the traditional ones because the consequence (2) needs computation rather much. A merit of our proofs for 2OE and 3OE is easy generalizability to those of 4OE and 5OE.

Besides efficiency of the MLE, the loss of Fisher information of 5OE estimators and its geometrical structure are also interesting topics. Much more tedious calculation is involved in assessment of the loss of information itself, as compared with the calculation of the difference in Fisher information. The result will be coming soon.

ACKNOWLEDGMENT

The author would like to express his thanks to Professor S. Amari of The University of Tokyo for his suggestion.

References

Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency,* Lecture Notes in Statistics, Vol.7. Springer-Verlag, New York.

Amari, S. (1982). Differential geometry of curved exponential families—Curvatures and information loss. *Ann. Statist.*, **10**, 357-385.

Amari, S. (1985). *Differential-Geometrical Methods in Statistics,* Lecture Notes in Statistics, Vol.28. Springer-Verlag, Berlin.

Bentler, P. M. and Lee, Sik-Yum (1978). Matrix derivatives with chain rule and rules for simple, Hadamard, and Kronecker products. *J. Math. Psychol.*, **17**, 255-262.

Bhattacharya, R. and Denker, M. (1990). *Asymptotic Statistics.* Birkhäuser, Basel.

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

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statist.*, **11**, 793-803.

Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.*, **122**, 700-725

Ghosh, J. K. (1994). *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol.4. IMS, Hayward, CA.

Ghosh, J. K. and Subramanyam, K. (1974). Second order efficiency of maximum likelihood estimates. *Sankhyā, Ser. A*, **36**, 325-358.

Hosoya, Y. (1988). The second-order Fisher information. *Biometrika*, **75**, 265-274.

Kano, Y. (1993a). Beyond third-order efficiency. (submitted)

Kano, Y. (1993b). Beyond third-order efficiency: Supplement. *Technical Report*, University of Osaka Prefecture, Osaka, Japan.

Kano, Y. (1994). *More* higher-order efficiency: Concentration probability. *Technical Reports of Statistical Group:* DATA95-04, Tokyo, Japan.

Kano, Y. (1995). Third-order efficiency implies fourth-order efficiency. *Technical Reports of Statistical Group:* DATA95-07, Tokyo, Japan.

Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley, Chichester.

Pfanzagl, J. and Wefelmeyer, W. (1978). A third order optimum property of the maximum likelihood estimator. *J. Multivariate Anal.*, **8**, 1-29.

Rao, C. R. (1961). Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symp.*, **1**, 531-545.

Rao, C. R., Sinha, B. K. and Subramanyam, K. (1982). Third order efficiency of the maximum likelihood estimator in the multinomial distribution. *Statist. Decisions*, **1**, 1-16.

Takeuchi, K. and Morimune, K. (1985). Third-order efficiency of the extended maximum likelihood estimators in a simultaneous equation system. *Econometrica*, **53**, 177-200.

Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics, Vol.68. Springer-Verlag, Berlin.

Taniguchi, M and Watanabe, Y. (1994). Statistical analysis of curved probability densities. *J. Multivariate Anal.*, **48**, 228-248.