Ann. Inst. Statist. Math.
Vol. 42, No. 1, 37–49 (1990)

INFORMATION AMOUNT AND HIGHER-ORDER EFFICIENCY IN ESTIMATION

YUZO HOSOYA

Faculty of Economics, Tohoku University, Kawauchi, Sendai 980, Japan

(Received June 3, 1988; revised April 14, 1989)

Abstract. By means of second-order asymptotic approximation, the paper clarifies the relationship between the Fisher information of first-order asymptotically efficient estimators and their decision-theoretic performance. It shows that if the estimators are modified so that they have the same asymptotic bias, the information amount can be connected with the risk based on convex loss functions in such a way that the greater information loss of an estimator implies its greater risk. The information loss of the maximum likelihood estimator is shown to be minimal in a general set-up. A multinomial model is used for illustration.

Key words and phrases: Fisher information amount, decision theory, risk function, efficiency, maximum likelihood estimator, asymptotic theory, estimation, information loss.

1. Introduction

The concept of information amount is related to the summary aspect of statistics and measures the amount of information loss through the data reduction process. Fisher (1925) emphasized the relevance of the concept as a criterion for the comparison of estimators as opposed to such decision-theoretic criterion as mean square error. He indicated the maximum likelihood estimator suffers minimal loss of information amount among first-order asymptotically efficient estimators, and made this the main reason for its use. However, it is by no means clear how the summary aspect of an estimator is related to efficiency of estimation. As far as first-order asymptotic approximation is concerned, the sufficiency of a consistent estimator implies its efficiency. But the relationship seems not so straightforward in higher-order approximation, and the maximum likelihood estimator in itself does not exhibit any superiority in a general decision-theoretic estimation set-up; a specific correction needs to be applied in order to be optimal (see Ghosh and Subramanyam (1974), Pfanzagl and Wefelmeyer (1978), Ghosh *et al.* (1980), Akahira and Takeuchi (1981) and Akahira *et al.* (1988)).

Of all the above results about the higher-order asymptotic optimality of the maximum likelihood estimator, Pfanzagl and Wefelmeyer (1978), Takeuchi (1982) and Bickel *et al.* (1985) showed under general set-ups that the maximum likelihood estimator is third-order optimal for regular loss functions if the comparison is conducted in such a way that compared estimators are modified so that they have some asymptotic bias.

In establishing a higher-order decision-theoretic optimality of the maximum likelihood estimator if bias-corrected, Ghosh and Subramanyam (1974) noted that the relationship between the Fisher information criterion and the decision-theoretic optimality in estimation needed to be clarified. However, they do not seem to have gone further. In view of those studies on higher-order asymptotic properties of the maximum likelihood estimation, there still remains the problem of how the information amount of an estimator and the bias adjustment for higher-order optimality in estimation is related. Ghosh and Subramanyam (1974) do not seem to have clarified this point, in spite of their original intention.

This paper aims at directly relating the information amount of first-order efficient estimators satisfying certain conditions (which will be called BAN estimators below) and their decision-theoretic performance in view of a specific loss function in second-order asymptotic approximation (namely, up to and including the order $O(n^{-1})$), giving a general formula connecting the two and thus clarifying the relationship. The paper shows that a correction such that a class of estimators has the same asymptotic bias is essential for their risk to be related to their information amount, and the latter is invariant under bias-correction. More specifically, thanks to the formula of the second-order information amount given by Hosoya (1988), Theorem 2.1 in Section 2 shows the invariance property of the information amount and Theorem 2.2 establishes a relationship between the risk of BAN estimators and their information amount. The theorem implies that if estimators are modified in such a way that each of them has the same asymptotic bias, then their decision-theoretic performance becomes comparable and shows that the greater is the loss of information of an estimator, the greater is its risk. It should be noted that this relationship holds for any pair of BAN estimators, and that the comparison between the maximum likelihood estimator and another BAN estimator is only a special case. In other words, a decision-theoretic comparison of any two BAN estimators in terms of their information amount gets meaning in view of the theorem, whereas the studies so far seem to have been focused only on the decision-theoretic completeness of the maximum likelihood estimator (see Pfanzagl and Wefelmeyer (1978) and Akahira *et al.* (1988)). Section 3 shows that the maximum likelihood estimator suffers the least loss of information among BAN estimators under a general set-up

(Theorem 3.1). The property is known so far for the multinomial case (Rao (1962)) and for the stationary Gaussian linear process (Hosoya (1979)). Section 4 illustrates the results in Sections 2 and 3 by comparing the maximum likelihood and minimum χ^2 estimators for a multinomial model, evaluating explicitly their second-order information amount, and the Kullback (1959) and other risks.

Throughout, the asymptotic cumulants (which are denoted by such notation as $\underline{E}$, $\underline{\text{Var}}$, $\underline{\text{Third Cumulant}}$) are evaluated by means of formal term-by-term integration of statistical stochastic expansion. The rigorous conditions for the validity of such an operation are not pursued, but it is implicitly assumed. Also, the Einstein convention is used in this paper, meaning that if an index appears twice in a term—once as a superscript and once as a subscript—summation over the index is indicated.

2. The second-order Kullback risk and the second-order Fisher information

Suppose that the observable random n-vector X_n has a probability measure in a real n-space R^n and has a density $f(x_n|\theta)$ with respect to a σ-finite measure μ, where $\theta = \{\theta^1, \ldots, \theta^p\}$ is a parameter p-vector. The density is assumed to be sufficiently smooth with respect to θ in an open set $\Theta \subset R^p$. The log likelihood function is denoted as $l_n(x_n|\theta)$; namely, $l_n(x_n|\theta) = \log f(x_n|\theta)$. Set the derivatives as $l_{n,i}(x_n|\theta) = \partial l_n(x_n|\theta)/\partial\theta^i$, $l_{n,ij}(x_n|\theta) = \partial^2 l_n(x_n|\theta)/\partial\theta^i\partial\theta^j$ and so on, and assume that for $\theta \in \Theta$

(i) $E_\theta\{l_{n,i}(X_n|\theta)\} = 0$,

(ii) $E_\theta\{l_{n,i}(X_n|\theta)l_{n,j}(X_n|\theta)\} = -E\{l_{n,ij}(X_n|\theta)\}$,

(iii) $E_\theta\{(1/n)l_{n,i}(X_n|\theta)l_{n,j}(X_n|\theta)\} = I_{ij}(\theta) + O(n^{-1})$ and $I(\theta)$ is positive-definite where I denotes the matrix $\{I_{ij}\}$,

(iv) the joint distribution of $n^{-1/2}l_{n,i}(X_n|\theta)$ tends weakly to a multivariate normal distribution with mean 0 and covariance matrix $I(\theta)$ as $n \to \infty$ if θ is the true value.

The class of estimators of θ which is investigated in the following is the one termed the best asymptotically normal (BAN) estimators. An estimator $\tilde{\theta}(X_n)$ is BAN if it has the asymptotic expansion when θ is true

$$
\begin{aligned}
(2.1) \qquad \tilde{\theta}^i_n &= \sqrt{n}\,(\tilde{\theta}^i - \theta^i) \\
&= n^{-1/2} l_{n,i}(X_n|\theta) + n^{-1/2} m_i(X_n, \theta) + o_p(n^{-1/2}),
\end{aligned}
$$

where $m_i(X_n, \theta)$ is measurable with respect to X_n and smooth with respect to θ such that $E_\theta(|m_i(X_n, \theta)|)$ exists and is bounded as $n \to \infty$ where I^{ij} is the (i,j)-th element of the inverse of I. Also the BAN estimator $\tilde{\theta}$ is assumed to have the asymptotic cumulants which have the following expression:

$$\underline{\mathrm{E}}_\theta(\tilde{\theta}_n^i) = \frac{1}{\sqrt{n}} K_1^i(\theta, \tilde{\theta}) + \frac{1}{n^{3/2}} K_2^i(\theta, \tilde{\theta}) + o(n^{-3/2}) ,$$

(2.2) $$\underline{\mathrm{Cov}}_\theta(\tilde{\theta}_n^i, \tilde{\theta}_n^j) = I^{ij}(\theta) + \frac{1}{n} K^{ij}(\theta, \tilde{\theta}) + o\left(\frac{1}{n}\right) ,$$

$$\underline{\text{Third Cum}}(\tilde{\theta}_n^i, \tilde{\theta}_n^j, \tilde{\theta}_n^k) = \frac{1}{\sqrt{n}} K_1^{ijk}(\theta) + \frac{1}{n} K_2^{ijk}(\theta, \tilde{\theta}) + o\left(\frac{1}{n}\right) ,$$

where note that the argument $\tilde{\theta}$ is omitted in $K_1^{ijk}(\theta)$ because for the class of estimators (2.1) K_1^{ijk} is seen to be common (see Akahira and Takeuchi (1981), p. 161). Then assume a maximum likelihood estimator $\hat{\theta}$ is BAN ($\sqrt{n}(\hat{\theta} - \theta)$ is also denoted as $\hat{\theta}_n$).

The second-order Fisher information amount is the asymptotic approximation, up to and including the order $O(1)$ of the Fisher information amount, and the general formula for asymptotically normal statistics was given by Hosoya (1988). Its application to $\tilde{\theta}_n$ leads to the following expression of the second-order information amount matrix I_{ij}^* of $\tilde{\theta}_n$:

(2.3) $$\begin{aligned} I_{ij}^*(\theta, \tilde{\theta}_n) = {} & nI_{ij}(\theta) + \frac{1}{2} \operatorname{tr} \{I(\theta)^{-1} I_{,i}(\theta) I(\theta)^{-1} I_{,j}(\theta)\} \\ & + \frac{1}{2} \{K_1^{lmn}(\theta) I_{lm,[i}(\theta) I_{j]n}(\theta) \\ & + K_1^{lmn}(\theta) K_1^{opq}(\theta) I_{lm}(\theta) I_{ni}(\theta) I_{oj}(\theta) I_{pq}(\theta)\} \\ & - I_{li}(\theta) I_{mj}(\theta) \{K^{lm}(\theta, \tilde{\theta}) - I^{k[l}(\theta) K_{1,k}^{m]}(\theta, \tilde{\theta})\} , \end{aligned}$$

where the brackets [ ] denote the symmetrization with respect to its arguments, and the subscripts preceded by commas denote the differentiation from the corresponding elements of θ; $I_{lm,i} = \partial I_{lm}(\theta)/\partial \theta^i$ for example.

A bias correction of a BAN estimator $\tilde{\theta}$ means in this paper the modification such that

(2.4) $$\tilde{\theta}^{*i} = \tilde{\theta}^i + n^{-1} \phi_1^i(\tilde{\theta}) + n^{-2} \phi_2^i(\tilde{\theta}) ,$$

where ϕ_1^i and ϕ_2^i are assumed to be first-order continuously differentiable. If $\phi_1(\cdot) = -K_1^i(\cdot, \tilde{\theta})$, (2.4) implies that $\tilde{\theta}^*$ is asymptotically (up to n^{-1}) unbiased; the bias correction below is not restricted to this unbiasedness correction. By the correction (2.4), the cumulants $\tilde{\theta}_n^*$ become:

(2.5)
$$\underline{E}_\theta(\tilde{\theta}_n^{*i}) = \frac{1}{\sqrt{n}}\{K_1^i(\theta,\tilde{\theta}) + \phi_1^i(\theta)\} + \frac{1}{n^{3/2}}\{K_2^i(\theta,\tilde{\theta}) + \phi_2^i(\theta)\}$$
$$+ \frac{1}{n^{1/2}}\underline{E}\{\phi_1^i(\tilde{\theta}_1) - \phi_1^i(\theta)\} + o(n^{-3/2}),$$
$$\underline{\mathrm{Cov}}_\theta(\tilde{\theta}_n^{*i}, \tilde{\theta}_n^{*j}) = I^{ij}(\theta) + \frac{1}{n}\{K^{ij}(\theta,\tilde{\theta}) + I^{k[i}\phi_{1,k}^{j]}(\theta)\} + o\left(\frac{1}{n}\right),$$
$$\underline{\text{Third Cum}}_\theta(\tilde{\theta}_n^{*i}, \tilde{\theta}_n^{*j}, \tilde{\theta}_n^{*k}) = \frac{1}{\sqrt{n}} K_1^{ijk}(\theta) + \frac{1}{n} K_2^{ijk}(\theta,\tilde{\theta}^*) + o\left(\frac{1}{n}\right).$$

Then the substitution of these values in the formula (2.3) gives this result in a straightforward way.

THEOREM 2.1. *The second-order Fisher information amount is invariant for the correction* (2.3); *namely, it holds that*

$$I_{ij}^*(\theta,\tilde{\theta}_n) = I_{ij}^*(\theta,\tilde{\theta}_n^*).$$

In order to see how this invariance property is related to the evaluation of the performance of estimators in a specific estimation situation, suppose that a convex loss function $h(\theta,\tilde{\theta})$ is given and that it satisfies: (i) $h(\theta_1,\theta_2) \geq 0$; $h(\theta_1,\theta_2) = 0$ only if $\theta_1 = \theta_2$, (ii) $h(\theta_1,\theta_2)$ is fourth-order continuously differentiable with respect to the second arguments and (iii) $h_{,ij}(\theta_1,\theta_1)$ is positive definite, where $h_{,ij}(\theta_1,\theta_2) = \partial^2 h(\theta_1,\theta_2)/\partial\theta_2^i\partial\theta_2^j$. The risk of using $\tilde{\theta}$ is then given as $E_\theta h(\theta,\tilde{\theta})$ and second-order approximation of $nE_\theta h(\theta,\tilde{\theta})$ is said to be the second-order risk of $\tilde{\theta}$ based on the loss function h and denoted as $R(\theta,\tilde{\theta})$. It follows from the stochastic expansion

$$\begin{aligned} nh(\theta,\tilde{\theta}) = &\sqrt{n}\,h_{,i}(\theta,\theta)\tilde{\theta}_n^i + h_{,ij}(\theta,\theta)\tilde{\theta}_n^i\tilde{\theta}_n^j/2 \\ &+ n^{-1/2}h_{,ijk}(\theta,\theta)\tilde{\theta}_n^i\tilde{\theta}_n^j\tilde{\theta}_n^k/6 \\ &+ n^{-1}h_{,ijkl}\tilde{\theta}_n^i\tilde{\theta}_n^j\tilde{\theta}_n^k\tilde{\theta}_n^l/24 + o_p(n^{-1}) \end{aligned}$$

that

(2.6)
$$\begin{aligned} R(\theta,\tilde{\theta}) = &h_{,i}\{K_1^i(\theta,\tilde{\theta}) + n^{-1}K_2^i(\theta,\tilde{\theta})\} \\ &+ h_{,ij}\{I^{ij}(\theta) + n^{-1}K^{ij}(\theta,\tilde{\theta}) \\ &+ n^{-1}K_1^i(\theta,\tilde{\theta})K_1^j(\theta,\tilde{\theta})\}/2 \\ &+ n^{-1}h_{,ijk}\{K_1^{ijk}(\theta) + 3I^{ij}K_1^k(\theta,\tilde{\theta})\}/6 \\ &+ n^{-1}h_{,ijkl}I^{ij}I^{kl}/8. \end{aligned}$$

There seems to be no way to connect $R(\theta,\tilde{\theta})$ and $I^*(\theta,\tilde{\theta})$ in the form they are given; but suppose bias correction is applied so that corrected estimators have the same bias, up to and including $O(n^{-2})$. Namely, given two BAN estimators $\tilde{\theta}_1$, $\tilde{\theta}_2$, let ϕ_1 and ϕ_2 be correction functions such that $\tilde{\theta}_1^* = \tilde{\theta}_1 + n^{-1}\phi_1(\tilde{\theta}_1) + n^{-2}\phi_2(\tilde{\theta}_1)$ and $\tilde{\theta}_2^* = \tilde{\theta}_2 + n^{-1}\psi_1(\tilde{\theta}_2) + n^{-2}\psi_2(\tilde{\theta}_2)$ for which

$$
\begin{aligned}
(2.7)\qquad & K_1^i(\theta,\tilde{\theta}_1) + \phi_1^i(\theta) = K_1^i(\theta,\tilde{\theta}_2) + \psi_1^i(\theta), \quad \text{and} \\
& K_2^i(\theta,\tilde{\theta}_1) + n\underline{\mathrm{E}}\{\phi_1^i(\tilde{\theta}) - \phi_1^i(\theta)\} + \phi_2^i(\theta) \\
& \quad = K_2^i(\theta,\tilde{\theta}_2) + n\underline{\mathrm{E}}\{\psi_1^i(\tilde{\theta}_2) - \psi_1^i(\theta)\} + \psi_2^i(\theta) + o(1)\,.
\end{aligned}
$$

Then, the next theorem holds.

THEOREM 2.2.

$$
\begin{aligned}
(2.8)\qquad & R(\theta,\tilde{\theta}_1^*) - R(\theta,\tilde{\theta}_2^*) \\
& \quad = -\{h_{,ij}(\theta)I^{il}(\theta)I^{jk}(\theta)\}\{I_{lk}^*(\theta,\tilde{\theta}_1) - I_{lk}^*(\theta,\tilde{\theta}_2)\}/(2n)\,.
\end{aligned}
$$

PROOF. It follows from $K_1^i(\theta,\tilde{\theta}_1^*) = K_1^i(\theta,\tilde{\theta}_2^*) = \eta^i(\theta)$ that

$$R(\theta,\tilde{\theta}_1^*) - R(\theta,\tilde{\theta}_2^*) = h_{,ij}(\theta,\tilde{\theta}_1^*)\{K^{ij}(\theta,\tilde{\theta}_1^*) - K^{ij}(\theta,\tilde{\theta}_2^*)\}/(2n)\,.$$

On the other hand, thanks to (2.5)

$$
\begin{aligned}
K^{ij}(\theta,\tilde{\theta}_1^*) - K^{ij}(\theta,\tilde{\theta}_2^*) &= K^{ij}(\theta,\tilde{\theta}_1) + I^{k[i}\phi_{1,k}^{j]}(\theta) \\
&\quad - \{K^{ij}(\theta,\tilde{\theta}_2) + I^{k[i}\psi_{1,k}^{j]}(\theta)\}
\end{aligned}
$$

whereas in view of (2.7) the right-hand side is equal to

$$
\begin{aligned}
(2.9)\qquad & K^{ij}(\theta,\tilde{\theta}_1) - I^{k[i}K_{1,k}^{j]}(\theta,\tilde{\theta}_1) - \{K^{ij}(\theta,\tilde{\theta}_2) - I^{k[i}K_{1,k}^{j]}(\theta,\tilde{\theta}_2)\} \\
& \quad = -I^{il}I^{jk}\{I_{lk}^*(\theta,\tilde{\theta}_{1n}) - I_{lk}^*(\theta,\tilde{\theta}_{2n})\}\,,
\end{aligned}
$$

by the formula (2.3). Thus the result follows. □

COROLLARY 2.1. *If* $I_{ij}^*(\theta,\tilde{\theta}_1) - I_{ij}^*(\theta,\tilde{\theta}_2)$ *is positive-definite, then* $R(\theta,\tilde{\theta}_1^*) < R(\theta,\tilde{\theta}_2^*)$.

PROOF. Since $h_{,ij}$ is positive-definite, it has the spectral decomposition $h_{,ij} = \sum_m \lambda_m a_{mi} a_{mj}$ (λ_m is positive). Then the right-hand side of (2.8) is written as

$$- \Sigma \lambda_m (a_{mi} I^{il})(a_{mj} I^{jk}) \{ I_{lk}^*(\theta, \tilde{\theta}_1) - I_{lk}^*(\theta, \tilde{\theta}_2) \} / (2n) . \qquad \square$$

The consequence of Theorem 2.2 and Corollary 2.1 is evident; once the bias of BAN estimators is adjusted so that they have a common bias, not only their decision-theoretic performances become comparable, but they can be interpreted in terms of information amount. Unless such adjustment is made, there seems no general way of connecting the two concepts (see Efron (1982) for a discussion of the distinction between the data summary aspect of an estimator and the decision-theoretic nature of estimation).

Example. Suppose the loss function is given by what Kullback (1959) calls the mean information for discrimination, which is represented as

$$h(\theta_1, \theta_2) = E_{\theta_1} \{ l_n(X_n, \theta_1) - l_n(X_n, \theta_2) \} ; \tag{2.10}$$

then the difference of the second-order risk of two bias corrected estimators $\tilde{\theta}_1^*$ and $\tilde{\theta}_2^*$ is expressed as

$$R(\theta, \tilde{\theta}_1^*) - R(\theta, \tilde{\theta}_2^*) = - I^{ij}(\theta) \{ I_{ij}^*(\theta, \tilde{\theta}_1) - I_{ij}^*(\theta, \tilde{\theta}_2) \} / (2n) .$$

Remark 1. Note that Theorem 2.2 does not assume symmetricity for the loss function $h(\theta_1, \theta_2)$. The result (2.8) is derived thanks to the particular way the first cumulant $K_1^i(\theta, \tilde{\theta})$ is involved in the last term of the second-order information amount formula (2.3) and thanks to the fact that the BAN estimators have a common third-order cumulant $K_1^{ijk}(\theta)$. See for the related optimality result of the maximum likelihood estimator Takeuchi (1982) and Bickel *et al.* (1985), where they showed that the bias-adjusted maximum likelihood estimator is third-order efficient for specific asymmetric loss functions.

Remark 2. In the derivation of (2.3), a little generalization is made in the application of Hosoya's (1988) result where $K_2(\theta, \tilde{\theta})$ is assumed to be equal to 0. But the assumption turns out to be unnecessary, since the contribution to I_{ij}^* by that term is equal to

$$n^{-1} \underline{\mathrm{E}} \left[\frac{\partial^2}{\partial \theta^i \partial \theta^j} \left\{ \frac{1}{6} I_{lp} I_{mq} I_{nr} \tilde{\theta}_n^p \tilde{\theta}_n^q \tilde{\theta}_n^r K_2^{pqr}(\theta, \tilde{\theta}) \right\} \right]$$

$$= \underline{\mathrm{E}} \left[\frac{1}{6} I_{lp} I_{mi} I_{nj} \tilde{\theta}_n^p K_2^{pqr}(\theta, \tilde{\theta}) \right] + o(1)$$

which is of order $o(1)$.

3. The second-order information amount of BAN estimators

Endow an order structure to the set of the second-order information amount of BAN estimators in such a way that $I^*(\theta, \tilde{\theta}_1) \leq I^*(\theta, \tilde{\theta}_2)$ if and only if $I^*(\theta, \tilde{\theta}_2) - I^*(\theta, \tilde{\theta}_1)$ is non-negative definite. The purpose of this section is to prove the next theorem, which shows that the information matrix of the maximum likelihood estimator $\hat{\theta}$ is maximal with respect to this order in the general set-up of Section 2. (Fisher (1925) noted the superiority of the maximum likelihood estimator in this sense and also regarded this as the reason for superiority as an estimator. See Rao (1962) for this property in the multinomial case, Hosoya (1979) in linear stationary time-series models and Hosoya (1988) for simultaneous equation models with nuisance parameters.) In addition to the assumptions of Section 2, assume:

(i)
$$\lim_{n\to\infty} n^{-1} \underline{\mathrm{E}}_\theta\{l_{n,i} l_{n,jk}\} = L_{i\cdot jk}(\theta) ,$$
$$\lim_{n\to\infty} n^{-1} \underline{\mathrm{E}}_\theta\{l_{n,ijk}\} = L_{ijk}(\theta) .$$

Those limits exist and there is a relationship that

$$I_{ij,k}(\theta) = -L_{ijk}(\theta) - L_{ij\cdot k}(\theta) ; \tag{3.1}$$

(ii) For any $\tilde{\theta}$, $\underline{\mathrm{E}}_\theta[n^{-1/2}(l_{n,ij} + nI_{ij})n(\hat{\theta}^j - \tilde{\theta}^j)]$ has a derivative with respect to θ which is bounded in n, and $\underline{\mathrm{E}}[n^{-1/2}(l_{n,ijk} - nL_{ijk}(\theta))n(\hat{\theta}^j - \tilde{\theta}^j)]$ and $\underline{\mathrm{E}}(n(\hat{\theta}^j - \tilde{\theta}^j))$ are bounded in n.

THEOREM 3.1. *For any BAN estimator* $\tilde{\theta}$,

$$I^*_{ij}(\theta, \hat{\theta}_n) - I^*_{ij}(\theta, \tilde{\theta}_n)$$
$$= I_{li}(\theta) I_{mj}(\theta) \underline{\mathrm{Cov}}\,\{n(\tilde{\theta}^l - \hat{\theta}^l), n(\tilde{\theta}^m - \hat{\theta}^m)\} + o(1) .$$

PROOF. In order to evaluate $K^{lm}(\theta, \tilde{\theta})$, consider the equation

$$\underline{\mathrm{Cov}}\,(\tilde{\theta}^l_n, \tilde{\theta}^m_n) = \underline{\mathrm{Cov}}\,(\hat{\theta}^l_n, \tilde{\theta}^m_n)$$
$$+ n^{-1/2}[\underline{\mathrm{Cov}}\,\{\hat{\theta}^l_n, n(\hat{\theta}^m - \tilde{\theta}^m)\} + \underline{\mathrm{Cov}}\,\{\hat{\theta}^m_n, n(\hat{\theta}^l - \tilde{\theta}^l)\}]$$
$$+ n^{-1}\,\underline{\mathrm{Cov}}\,\{n(\tilde{\theta}^l - \hat{\theta}^l), n(\tilde{\theta}^m - \hat{\theta}^m)\} ,$$

where if it can be shown that

(3.2)
$$\underline{\mathrm{Cov}}\,\{\hat\theta^l_n, n(\hat\theta^m - \tilde\theta^m)\} = n^{-1/2}\{I^{lk}K^m_{1,k}(\theta,\hat\theta) - I^{lk}K^m_{1,k}(\theta,\tilde\theta)\} + o(n^{-1/2}),$$

the result follows, since then

$$\begin{aligned} n^{-1}\{K^{lm}(\theta,\tilde\theta) - K^{lm}(\theta,\hat\theta)\} &= \underline{\mathrm{Cov}}\,(\tilde\theta^l_n, \tilde\theta^m_n) - \underline{\mathrm{Cov}}\,(\hat\theta^l_n, \hat\theta^m_n) + o(n^{-1}) \\ &= n^{-1}\{I^{k[l}K^{m]}_{1,k}(\theta,\tilde\theta) - I^{k[l}K^{m]}_{1,k}(\theta,\hat\theta)\} \\ &\quad + n^{-1}\,\underline{\mathrm{Cov}}\,\{n(\tilde\theta^l - \hat\theta^l), n(\tilde\theta^m - \hat\theta^m)\} + o(n^{-1}), \end{aligned}$$

whereas, in view of (2.3),

$$\begin{aligned} I^*_{ij}(\theta,\hat\theta_n) - I^*_{ij}(\theta,\tilde\theta_n) = I_{li}I_{mj}\cdot[K^{lm}(\theta,\tilde\theta) - K^{lm}(\theta,\hat\theta) \\ - \{I^{k[l}K^{m]}_{1,k}(\theta,\tilde\theta) - I^{k[l}K^{m]}_{1,k}(\theta,\hat\theta)\}]. \end{aligned}$$

Therefore what needs to be established is the relationship (3.2). It follows from the stochastic expansion of $\hat\theta_n$

$$\begin{aligned} \hat\theta^i_n = n^{-1/2}I^{ij}l_{n,j}(\theta) + n^{-1/2}I^{ij}\hat\theta^l_n n^{-1/2}\{l_{n,jl}(\theta) + nI_{jl}\} \\ + o_p(n^{-1/2}), \end{aligned}$$

that

$$\begin{aligned} &\underline{\mathrm{Cov}}\,\{\hat\theta^i_n, n(\hat\theta^j - \tilde\theta^j)\} \\ &\quad = I^{ik}\,\underline{\mathrm{Cov}}\,\{n^{-1/2}l_{n,k}, n(\hat\theta^j - \tilde\theta^j)\} \\ &\quad + n^{-1/2}I^{im}I^{lk}\,\underline{\mathrm{Cov}}\,\{n^{-1/2}l_{n,k}n^{-1/2}(l_{n,ml} + nI_{ml}), n(\hat\theta^j - \tilde\theta^j)\} \\ &\quad + o(n^{-1/2}) \end{aligned}$$

where the first covariance in the right-hand side is equal to

$$\begin{aligned} &\underline{\mathrm{Cov}}\,\{n^{-1/2}l_{n,k}, n(\hat\theta^j - \tilde\theta^j)\} \\ &\quad = n^{-1/2}\frac{\partial}{\partial\theta^i}\,\underline{\mathrm{E}}\,\{n(\hat\theta^j - \tilde\theta^j)\} = n^{-1/2}\{K^j_{1,i}(\theta,\hat\theta) - K^j_{1,i}(\theta,\tilde\theta)\} + o(n^{-1/2}). \end{aligned}$$

As for the second covariance, it holds that

$$\begin{aligned} &\underline{\mathrm{Cov}}\,\{n^{-1/2}l_{n,k}n^{-1/2}(l_{n,ml} + nI_{ml}), n(\hat\theta^j - \tilde\theta^j)\} \\ &\quad = \underline{\mathrm{E}}\,\{n^{-1/2}l_{n,k}n^{-1/2}(l_{n,ml} + nI_{ml})n(\hat\theta^j - \tilde\theta^j)\} \\ &\quad - L_{k\cdot lm}\,\underline{\mathrm{E}}\,\{n(\hat\theta^j - \tilde\theta^j)\} + o(n^{-1/2}) \end{aligned}$$

$$= n^{-1/2} \frac{\partial}{\partial \theta^k} \underline{\mathrm{E}} \{n^{-1/2}(l_{n,ml} + I_{ml}) n(\hat{\theta}^j - \tilde{\theta}^j)\}$$
$$- n^{-1/2} \underline{\mathrm{E}} \{n^{-1/2}(l_{n,mlk} + nI_{ml,k}) n(\hat{\theta}^j - \tilde{\theta}^j)\}$$
$$- L_{k \cdot lm} \underline{\mathrm{E}} \{n(\hat{\theta}^j - \tilde{\theta}^j)\} + o(n^{-1/2}) ,$$

where the first term in the right-hand side is of order $o(1)$ by assumption (ii), whereas the sum of the second and third terms is equal to $n^{-1/2} \cdot \underline{\mathrm{E}} \{n^{-1/2}(l_{n,mlk} - nL_{mlk}) n(\hat{\theta}^j - \tilde{\theta}^j)\}$ in view of (3.1), and since it is of order $o(1)$, the relationship (3.2) follows. □

4. A multinomial model

Let the random numbers $N^1,\dots, N^s$ ($N^1 + \cdots + N^s = n$) have a multinomial distribution

$$(4.1) \qquad p\{N^1 = n^1,\dots, N^s = n^s | \theta\} = \frac{n!}{n^1! \cdots n^s!} p^1(\theta)^{n^1} \cdots p^s(\theta)^{n^s}$$

and suppose that p^i's are smooth positive functions of a scalar parameter θ. Set $l_n(\theta) = \log p(N^1,\dots, N^s|\theta)$ and set $N_n^i = \sqrt{n}\{(N^i/n) - p^i(\theta)\}$. Let $\hat{\theta}$ be a maximum likelihood estimator of θ based on (4.1) and let $\tilde{\theta}$ be a minimum χ^2 estimator which is a value minimizing $\sum_{i=1}^{s} (N^i - np^i(\theta))^2/p^i(\theta)$; then $\tilde{\theta}$ satisfies the equation

$$(4.2) \qquad D_n(\tilde{\theta}) = \sum_{i=1}^{s} (N_i/n)^2 \{\dot{p}^i(\tilde{\theta})/p^i(\tilde{\theta})\}^2 = 0$$

where the dot denotes differentiation with respect to θ. In general, both $\hat{\theta}$ and $\tilde{\theta}$ are BAN estimators for regular situations and $\hat{\theta}_n$ has the following asymptotic cumulants (see Peers (1978)):

$$\underline{\mathrm{E}} (\hat{\theta}_n) = n^{-1/2} K_1 + o(n^{-1}) ,$$
$$\underline{\mathrm{Var}} (\hat{\theta}_n) = m_2^{-1} + n^{-1} K_2 + o(n^{-1}) ,$$
$$\underline{\text{Third Cumulant}} (\hat{\theta}_n) = n^{-1/2} K_3 + o(n^{-1}) ,$$

where

$$K_1 = \{m_{001}/2 + m_{11}\} m_2^{-2} ,$$
$$K_2 = \{m_{0001} + 3(m_{02} - m_{01}^2) + 2(m_{21} - 2m_2 m_{01}) + 3m_{101}\} m_2^{-3}$$
$$+ \{(5m_{001}^2/2 + 5m_{11}^2) + 8m_{11} m_{001}\} m_2^{-4} ,$$

$$K_3 = (2m_{001} + 3m_{11})m_2^{-3} ,$$

in which $m_{abc\cdots}$ is defined as

$$m_{abc\cdots}(\theta) = \sum_{i=1}^{s} (\log \dot{p}^i(\theta))^a (\log \ddot{p}^i(\theta))^b (\log \dddot{p}^i(\theta))^c \cdots p^i(\theta) .$$

In Section 2, the lower limit of second-order risk for BAN estimators is seen to be attained by the maximum likelihood estimator. Consider the second-order Kullback risk based on the loss function (2.10) and the unbiased correction of estimators. Namely, set

$$\hat{\theta}^* = \hat{\theta} - n^{-1}\{m_{001}(\hat{\theta})/2 + m_{11}(\hat{\theta})\}/m_2(\hat{\theta})^2 ;$$

then $\underline{\mathrm{E}}\,(\hat{\theta}^*) = o(n^{-1})$. In view of the relationship

$$\begin{aligned} \dot{K}_1 &= (m_{0001}/2 + 3m_{101}/2 + m_{02} + m_{21})m_2^{-2} \\ &\quad - (m_{001}^2 + 3m_{001}m_{11} + 2m_{11}^2)m_2^{-3} , \end{aligned}$$

and in view of (2.7), the Kullback risk of $\hat{\theta}^*$ is represented as

$$\begin{aligned} K^*(\theta; \hat{\theta}^*) = 1 - n^{-1}[\{(m_{0001}/2) - (3m_{101}/2) - 2m_{01} \\ - m_{21} + 3m_{01}^2 + 4m_2 m_{01}\}m_2^{-2} \\ - \{(m_{001}^2/2) - 2m_{11}m_{001} - 3m_{11}^2\}m_2^{-3}] . \end{aligned}$$

For the purpose of comparison of second-order risk of $\hat{\theta}^*$ and $\tilde{\theta}^*$, the asymptotic cumulants of $\tilde{\theta}$ is required, but it is derived by means of the relationship

$$n(\tilde{\theta} - \hat{\theta}) = (1/2m_2)\, \Sigma\, (N^i/n)^2 \dot{p}_i(\theta)/p_i(\theta)^2 + o_p(1) \tag{4.3}$$

which is obtained by the Taylor expansion of $D(\tilde{\theta})$ around $\hat{\theta}$. It follows from (4.3) that

$$\underline{\mathrm{E}}\,\{n(\tilde{\theta} - \hat{\theta})\} = o(1) ,$$

$$\begin{aligned} &\underline{\mathrm{Var}}\,\{n(\tilde{\theta} - \hat{\theta})\} \\ &\quad = \sum_i \sum_j \underline{\mathrm{Cov}}\,\{(N^i/n)^2, (N^j/n)^2\}\dot{p}^i(\theta)\dot{p}^j(\theta)/\{p^i(\theta)p^j(\theta)\}^2 + o(1) , \end{aligned}$$

whence the asymptotic cumulants of $\tilde{\theta}$ are

$$\underline{E}\,(\tilde{\theta}_n) = n^{-1/2}K_1 + o(n^{-1}) ,$$

$$\underline{\mathrm{Var}}\,(\tilde{\theta}_n) = m_2^{-1} + [K_2 + \underline{\mathrm{Var}}\; n(\tilde{\theta} - \hat{\theta})]/n + o(n^{-1}) ,$$

$$\underline{\text{Third Cumulant}}\,(\tilde{\theta}_n) = n^{-1/2}K_3 + o(n^{-1/2}) .$$

Now, in view of (2.6) and (2.8), the difference in the risk is expressed as

$$K^*(\theta; \tilde{\theta}^*) - K^*(\theta; \hat{\theta}^*) = m_2\{\underline{\mathrm{Var}}\; n(\tilde{\theta} - \hat{\theta})\}/(2n) + o(n^{-1}) , \tag{4.4}$$

where note in this case that (2.5) holds also for uncorrected $\tilde{\theta}$, $\hat{\theta}$, since the bias of $\tilde{\theta}$ and $\hat{\theta}$ is the same up to $O(n^{-1})$.

Finally, take as a risk the one which would be most favourable to the minimum χ^2 estimator; namely

$$D_1(\theta, \tilde{\theta}) = nE\left[\sum_i (p^i(\theta) - p^i(\tilde{\theta}))^2/p^i(\theta)\right]$$

where the second-order approximations for $\hat{\theta}$ and $\tilde{\theta}$, respectively, are given as

$$D_1^*(\theta, \hat{\theta}) = 2 + n^{-1}[m_2(K_1^2 + K_2) + 2(2m_{11} + 3m_3)(K_3 + 3m_2^{-1}K_1)$$
$$+ 6(2m_{101} + 2m_{05} + 15m_{21} + 7m_4)m_2^{-2}] + o(n^{-1}) ,$$
$$D_1^*(\theta, \tilde{\theta}) = D_1(\theta, \hat{\theta}) + n^{-1}m_2\,\underline{\mathrm{Var}}\,\{n(\tilde{\theta} - \hat{\theta})\}/2 + o(n^{-1}) ,$$

whence it is concluded that the maximum likelihood estimator is better for this risk, too, no matter what bias correction is applied.

REFERENCES

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

Akahira, M., Hirakawa, F. and Takeuchi, K. (1988). *Second and Third Order Asymptotic Completeness of the Class of Estimators, Probability Theory and Mathematical Statistics, Lecture Notes in Mathematics* 1299, 11–27, Springer, New York.

Bickel, P. J., Götze, F. and Van Zwet, W. R. (1985). A simple analysis of third-order efficiency of estimates, *Proc. Berkeley Conference in Honor of J. Neyman and J. Kiefer*, Vol. II, (eds. L. M. Le Cam and R. A. Olshen), 749–768, Wadsworth, Monterey.

Efron, B. (1982). Maximum likelihood and decision theory, *Ann. Statist.*, **10**, 340–356.

Fisher, R. A. (1925). Theory of statistical estimation, *Proc. Camb. Phil. Soc.*, **22**, 700–725.

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

Ghosh, J. K., Sinha, B. K. and Wieand, H. S. (1980). Second-order efficiency of the MLE with respect to any bowl-shaped loss function, *Ann. Statist.*, **8**, 506–521.

Hosoya, Y. (1979). High-order efficiency in the estimation of linear processes, *Ann. Statist.*, **7**, 516–530.

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

Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.

Peers, H. W. (1978). Second-order sufficiency and statistical invariants, *Biometrika*, **65**, 489–496.

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. (1962). Efficient estimates and optimum inference procedures in large sample, *J. Roy. Statist. Soc. Ser. B*, **24**, 26–72.

Takeuchi, K. (1982). Higher order asymptotic efficiency of estimators in decision procedures, *Proc. III Purdue Symp. Decision Theory and Related Topics*, Vol. II, 351–361, Academic Press.