# A COMMENT ON A STATISTICAL PROOF OF THE CONCAVITY ON THE EFFICIENT FISHER INFORMATION 

Etsuo Kumagai

Received December 24, 2012; revised January 11, 2013


#### Abstract

The efficient Fisher information is defined by the Schur complement in the Fisher information. We statistically prove the concavity of the matrix function on the efficient Fisher information more sophisticated way than the previous one [2].


1 Introduction Let $X$ be arbitrary random vector with the parameter $\theta$, which is distributed with a distribution $F(x)$ with respect to some probability measure on $\mathbf{R}^{s}$. Note that $s=p+q$ and the parameter is partitioned by $\theta=\left(\theta^{(1) T}, \theta^{(2) T}\right)^{T}$, where the notation ${ }^{T}$ stands for the transpose and the dimension of $\theta^{(1)}$ is $p$. Let $f(x ; \theta)$ be the probability density function with an expectation $\boldsymbol{\mu}=\boldsymbol{\mu}(\theta)$ and a covariance matrix $\boldsymbol{\Sigma}$ which is positive definite. Let the score function be partitioned as follows:

$$
\frac{\partial \log f(x ; \theta)}{\partial \theta}=\boldsymbol{J}(\theta)=\binom{\boldsymbol{J}_{1}(\theta)}{\boldsymbol{J}_{2}(\theta)}
$$

where $\boldsymbol{J}_{i}(\theta)$ are derived by $\theta^{(i)},(i=1,2)$, so that we have the Fisher information which is partitioned as follows:

$$
\boldsymbol{I}_{X}(\theta)=E_{\theta}\left(\boldsymbol{J}(\theta) \boldsymbol{J}(\theta)^{T}\right)=\left(\begin{array}{cc}
\boldsymbol{I}_{11, X}(\theta) & \boldsymbol{I}_{12, X}(\theta) \\
\boldsymbol{I}_{21, X}(\theta) & \boldsymbol{I}_{22, X}(\theta)
\end{array}\right) .
$$

Assume that the Fisher information is positive definite. For the Fisher information, the monotonicity and the additivity hold as follows: If $T=T(X)$ is a statistic with density function $g(t ; \theta)$, then the Fisher information $\boldsymbol{I}_{T}(\theta)$ on $T$ satisfies

$$
\begin{equation*}
\boldsymbol{I}_{X}(\theta) \geq \boldsymbol{I}_{T}(\theta) \tag{1}
\end{equation*}
$$

If $X, Y$ are independent random variables then $\boldsymbol{I}_{(X, Y)}(\theta)=\boldsymbol{I}_{X}(\theta)+\boldsymbol{I}_{Y}(\theta)$.
[1](page 28) introduced the efficient Fisher information matrix on $\theta^{(1)}$ in $X$ by

$$
\begin{equation*}
\widehat{\boldsymbol{I}}_{1, X}(\theta)=\boldsymbol{I}_{11, X}-\boldsymbol{I}_{12, X} \boldsymbol{I}_{22, X}^{-1} \boldsymbol{I}_{21, X} \tag{2}
\end{equation*}
$$

which is known as the Schur complement of $\boldsymbol{I}_{22, X}$ in $\boldsymbol{I}_{X}$. [2] showed that the monotonicity and the superadditivity on (2) hold as follows: If $T=T(X)$ is a statistic with $\boldsymbol{I}_{T}(\theta)$ positive definite, then $\widehat{\boldsymbol{I}}_{1, T}(\theta) \leq \widehat{\boldsymbol{I}}_{1, X}(\theta)$. If $X, Y$ are independent random variables then

$$
\begin{equation*}
\widehat{\boldsymbol{I}}_{1, X}(\theta)+\widehat{\boldsymbol{I}}_{1, Y}(\theta) \leq \widehat{\boldsymbol{I}}_{1,(X, Y)}(\theta) \tag{3}
\end{equation*}
$$

Although they tried to show the statistical proof of the concavity of the efficient Fisher information, the inequality (14) in page 347 was a little complicated or misleading to complete their proof. Here we prove it more sophisticated way than theirs, and our proof is easier to understand statistically.

[^0]2 Concavity on the efficient Fisher information The convexity of the matrix function $\phi(\boldsymbol{A})=\boldsymbol{A}^{-1}$ on the set of positive definite symmetric matrices $\left\{\boldsymbol{A}_{j}\right\}$ is

$$
\sum_{j=1}^{n} w_{j} \phi\left(\boldsymbol{A}_{j}\right) \geq \phi\left(\sum_{j=1}^{n} w_{j} \boldsymbol{A}_{j}\right)
$$

where the weights $w_{j}$ are nonnegative and $\sum_{j=1}^{n} w_{j}=1$. On the other hand, the concavity of the matrix function $\phi(\boldsymbol{A})=\left(\boldsymbol{A}^{11}\right)^{-1}$ is

$$
\sum_{j=1}^{n} w_{j} \phi\left(\boldsymbol{A}_{j}\right) \leq \phi\left(\sum_{j=1}^{n} w_{j} \boldsymbol{A}_{j}\right)
$$

For an analytic proof, see [3](page 678).
For both the above concavity and the convexity of the matrix functions, we shall prove them by the view of statistics simpler and wider than the previous way in [2]. Let $X_{1}, \ldots, X_{n}$ be independent $s$-variate normal random vectors with

$$
E_{\theta}\left(X_{j}\right)=v_{j} \theta, \quad V_{\theta}\left(X_{j}\right)=\boldsymbol{\Sigma}_{j}^{-1}, \quad(j=1, \ldots, n)
$$

where $v_{j}$ are known, $\sum_{j=1}^{n} v_{j}^{2}=1$, and $\boldsymbol{\Sigma}_{j}$ are positive definite. Setting $X=\left(X_{1}, \ldots, X_{n}\right)$, $v_{j}^{2}=w_{j}$ gives the Fisher information matrices as follows:

$$
\boldsymbol{I}_{X_{j}}=w_{j} \boldsymbol{\Sigma}_{j}, \quad \boldsymbol{I}_{X}=\sum_{j=1}^{n} \boldsymbol{I}_{X_{j}}=\sum_{j=1}^{n} w_{j} \boldsymbol{\Sigma}_{j}
$$

Since $S=\sum_{j=1}^{n} v_{j} X_{j}$ is a statistics of $X$ and is distributed with the normal distribution with

$$
E_{\theta}(S)=\sum_{j=1}^{n} v_{j}^{2} \theta=\theta, \quad V_{\theta}(S)=\sum_{j=1}^{n} w_{j} \boldsymbol{\Sigma}_{j}^{-1}
$$

so that the Fisher information on $S$ is

$$
\boldsymbol{I}_{S}=\left(\sum_{j=1}^{n} w_{j} \boldsymbol{\Sigma}_{j}^{-1}\right)^{-1}
$$

The monotonicity (1) on the Fisher information with respect to $S$ and $X$ gives

$$
\sum_{j=1}^{n} w_{j}\left(\boldsymbol{\Sigma}_{j}^{-1}\right)^{-1}=\boldsymbol{I}_{X} \quad \geq \quad \boldsymbol{I}_{S}=\left(\sum_{j=1}^{n} w_{j} \boldsymbol{\Sigma}_{j}^{-1}\right)^{-1}
$$

so that the convexity of the matrix function holds. On the other hand, since the matrices $\boldsymbol{\Sigma}_{j}$ and $\boldsymbol{\Sigma}_{j}^{-1}$ are partitioned by

$$
\boldsymbol{\Sigma}_{j}=\left(\begin{array}{cc}
\boldsymbol{\Sigma}_{11, j} & \boldsymbol{\Sigma}_{12, j} \\
\boldsymbol{\Sigma}_{21, j} & \boldsymbol{\Sigma}_{22, j}
\end{array}\right) \quad \text { and } \quad \boldsymbol{\Sigma}_{j}^{-1}=\left(\begin{array}{cc}
\boldsymbol{\Sigma}_{j}^{11} & \boldsymbol{\Sigma}_{j}^{12} \\
\boldsymbol{\Sigma}_{j}^{21} & \boldsymbol{\Sigma}_{j}^{22}
\end{array}\right)
$$

the superadditivity (3) on the efficient Fisher information gives

$$
\sum_{j=1}^{n} w_{j}\left(\boldsymbol{\Sigma}_{j}^{11}\right)^{-1}=\sum_{j=1}^{n} \widehat{\boldsymbol{I}}_{1, X_{j}} \leq \widehat{\boldsymbol{I}}_{1, X}=\left(\left(\sum_{j=1}^{n} w_{j} \boldsymbol{\Sigma}_{j}\right)^{11}\right)^{-1}
$$

so that the concavity of the matrix function holds.

## References

[1] P.J.Bickel, C.A.J.Klaassen, Y.Ritov, J.A.Wellner (1998), Efficient and Adaptive Estimation for Semiparametric Models, Springer.
[2] A.Kagan and C.R.Rao (2003), Some properties and applications of the efficient Fisher score, Journal of Statistical Planning and Inference, 116, 343-352.
[3] A.W. Marshall, I. Olkin, and B.C. Arnold (2010), Inequalities: Theory of Majorization and Its Applications 2nd Edition, Springer.

Communicated by Hisao Nagao

Graduate School of Engineering Science
Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531
Japan


[^0]:    2010 Mathematics Subject Classification. 62B10.
    Key words and phrases. The Schur complement, the efficient Fisher information, convexity, concavity.

