Ann. Inst. Statist. Math.
Vol. 48, No. 2, 257–265 (1996)

A MEASURE OF DISCRIMINATION BETWEEN TWO RESIDUAL LIFE-TIME DISTRIBUTIONS AND ITS APPLICATIONS

NADER EBRAHIMI[1] AND S. N. U. A. KIRMANI[2*]

[1] *Division of Statistics and Statistical Consulting Laboratory, Northern Illinois University, DeKalb, IL 60115-2854, U.S.A.*
[2] *Department of Mathematics, University of Northern Iowa, Cedar Falls, IA 50614, U.S.A.*

(Received March 20, 1995)

Abstract. A measure of discrepancy between two residual-life distributions is proposed on the basis of Kullback-Leibler discrimination information. Properties of this measure are studied and the minimum discrimination principle is applied to obtain the proportional hazards model.

Key words and phrases: Hazard function, cumulative hazard function, mean residual lifetime function, NBU, IFR, Kullback-Leibler discrimination information, MDI principle.

1. Introduction

Let X and Y be two non-negative random variables representing times to failure of two systems. For example, they might be the times to failure of two bio-systems say the left and right kidneys, or the times to failure of two engineering systems. Let $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$ be the failure distributions of X and Y with survival functions $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(y) = 1 - G(y)$ respectively, with $\bar{F}(0) = \bar{G}(0) = 1$. We assume that F and G are differentiable and that $f(t) = F'(t)$ and $g(t) = G'(t)$ denote the probability density functions of X and Y respectively. We denote the hazard rate functions of X and Y by $\lambda_F(x) = f(x)/\bar{F}(x)$ and $\lambda_G(x) = g(x)/\bar{G}(x)$ and their mean residual lifetime functions by $\delta_F(t) = E(X - t \mid X > t)$ and $\delta_G(t) = E(Y - t \mid Y > t)$ respectively.

The Kullback-Leibler discrimination information $I(X, Y)$ or $I(F, G)$ is defined as

$$(1.1) \qquad I(X, Y) = I(F, G) = \int_0^\infty f(x) \log \frac{f(x)}{g(x)} dx.$$

It is sometimes known as the relative entropy, and also as the directed information distance between F and G. $I(F, G)$ is defined for $f(x) \neq 0$ whenever $g(x) \neq 0$.

* Currently, at Northern Illinois University on a professional development leave.

$I(F,G)$ was introduced in statistics as early as (1951) by Kullback and Leibler and its use in hypothesis testing and model evaluation was propagated strongly by Kullback (1959). Since then it has been widely used in statistics, see Akaike (1973) and Ebrahimi *et al.* (1992).

Frequently, in survival analysis and in life testing one has information about the current age of the systems under consideration. In such cases, the age must be taken into account when comparing or discriminating between two systems. Obviously, the measure $I(F,G)$ is unsuitable in such situations and must be modified to take age into account. This can be achieved by replacing F and G by distributions of the corresponding residual lifetimes. Given that both systems have survived up to time t, we therefore define Kullback-Leibler discrimination information at t by

$$\begin{aligned}(1.2)\quad I(X,Y;t)=I(F,G;t)&=\int_t^\infty \frac{f(x)}{\bar F(t)}\log\frac{f(x)/\bar F(t)}{g(x)/\bar G(t)}dx\\ &=\log\bar G(t)+H(F;t)-\int_t^\infty \frac{f(x)}{\bar F(t)}\log g(x)dx,\end{aligned}$$

where

$$\begin{aligned}(1.3)\qquad H(F;t)&=\int_t^\infty \frac{f(x)}{\bar F(t)}\log\frac{f(x)}{\bar F(t)}dx\\ &=\frac{1}{\bar F(t)}\int_t^\infty f(x)\log f(x)dx-\log\bar F(t)\\ &=\frac{1}{\bar F(t)}\int_t^\infty (\log\lambda_F(x))f(x)dx-1.\end{aligned}$$

The third version in (1.3) can be obtained by integration by parts. Note that (1.1) and (1.2) coincide when $t=0$. It is clear that for each fixed $t>0$, $I(F,G;t)$ will have all the properties of the Kullback-Leibler discrimination information $I(F,G)$. In particular, $I(F,G;t)\ge 0$ with equality if and only if the probability density functions of residual lifetimes are equal almost everywhere.

The discrimination function in (1.2) is a measure of disparity between $F_t(x)=P(X_t\le x)=\frac{F(t+x)}{\bar F(t)}$, and $G_t(x)=P(Y_t\le x)=\frac{G(t+x)}{\bar G(x)}$, where $X_t\overset{d}{=}X-t\mid X>t$ and $Y_t\overset{d}{=}Y-t\mid Y>t$ denote the remaining lifetimes, and d stands for distribution. If we have a system with true survival function $\bar F$ then $I(F,G;t)$ can also be interpreted as a measure of distance between G_t and the true distribution F_t. G is generally referred to as the reference distribution. The following example demonstrates computation and usefulness of $I(F,G;t)$.

Example 1.1. Consider a parallel system of two independent components, each one with lifetime uniformly distributed over $[0,1]$. More specifically, suppose Y_1 and Y_2 are the lifetimes of components 1 and 2, respectively, with common probability density function g. If X is the lifetime of the system, then $X=\max(Y_1,Y_2)$ and the true probability density function of X is

$$f(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{elsewhere}\end{cases}.$$

One can easily verify that $I(F, G; t) = \log 2 - \frac{1}{2} - \log(1+t) - \frac{t^2}{1-t^2}\log t$ which is non-increasing in t. Furthermore $\lim_{t\to 1} I(F, G; t) = 0$. It simply means that as the system gets older and older, the distribution of remaining lifetime of the system gets closer and closer to the distribution of the remaining lifetime of each one of the components. Intuitively speaking, after a certain age the redundancy has negligible effect on the performance of the system.

As a dynamic measure of discriminatory information $I(F, G; t)$ turns out to be rather useful. Ebrahimi and Kirmani (1995) have proved the fact that $I(F, G; t)$ is constant in t if and only if F and G satisfy a proportional hazards model $\bar{G}(x) = (\bar{F}(x))^\beta$, for $x \geq 0$, where $\beta > 0$. The importance and utility of proportional hazards model are well-known, see e.g. Efron (1981) and Cox (1959).

This paper is organized as follows. In Section 2, we study some properties of this measure. In Section 3, it is shown that the proportional hazards model can be obtained from the minimum discrimination information (MDI) principle of Kullback (1954).

2. Properties of $I(F, G; t)$

In this section we study the properties of $I(F, G; t)$ defined by (1.2) and examine the implications of these properties. Our first result shows how $I(X, Y; t)$ is affected by a common increasing transformation of X and Y.

THEOREM 2.1. *$I(X, Y; \phi^{-1}(t)) = I(\phi(X), \phi(Y); t)$ for all increasing functions ϕ.*

PROOF.

$$\begin{aligned}
I(\phi(X), \phi(Y); t) &= \int_t^\infty \frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))\bar{F}(\phi^{-1}(t))} \log \frac{f(\phi^{-1}(x))/\bar{F}(\phi^{-1}(t))}{g(\phi^{-1}(x))/\bar{G}(\phi^{-1}(t))} \\
&= \int_{\phi^{-1}(t)}^\infty \frac{f(y)}{\bar{F}(\phi^{-1}(t))} \log \frac{f(y)/\bar{F}(\phi^{-1}(t))}{g(y)/\bar{G}(\phi^{-1}(t))} dy \\
&= I(X, Y; \phi^{-1}(t)).
\end{aligned}$$

This completes the proof.

Theorem 2.1 has a straightforward but interesting interpretation for the accelerated-life model. Suppose that two systems have lifetimes X_i^0, $i = 1, 2$, when operated under field conditions. Suppose also that failure testing can be carried out more rapidly in the laboratory in which case $X_i^* = X_i^0/\theta$, $i = 1, 2$, is the time to failure of system i. The acceleration factor θ is a constant possibly depending on variables, such as temperature and pressure, that accelerate failure. It follows from Theorem 2.1 that the discriminatory information between systems of laboratory age t equals the information between systems of field age θt.

The next theorem gives a relation between $I(F, G)$ and $I(F, G; t)$. We first recall that F is said to be new better (worse) than used if $\bar{F}(x+y) \leq (\geq) \bar{F}(x)\bar{F}(y)$ for all $x, y \geq 0$.

THEOREM 2.2. *Suppose*
(a) $\frac{\lambda_F(x)}{\lambda_G(x)}$ *is increasing (decreasing) in* x,
(b) *both* F *and* G *are new better (worse) than used.*
Then $I(F,G;t) \geq (\leq) I(F,G)$.

PROOF. We shall prove the result for the case when $\lambda_F(x)/\lambda_G(x)$ is increasing and both F and G are new better than used. If $f_t(x)$ and $g_t(x)$ are probability density functions of X_t and Y_t respectively and $H^{-1}(u) = \operatorname{Inf}\{u : H(x) \geq u\}$, then from (1.2) we have

$$I(F,G,t) = \int_0^\infty f_t(x) \log \frac{f_t(x)}{g_t(x)} dx = \int_0^\infty \left(\log \frac{f_t(x)}{g_t(x)}\right) dF_t(x) = \int_0^1 \log \frac{f_t(F_t^{-1}(y))}{g_t(F_t^{-1}(y))} dy. \tag{2.1}$$

Further,

$$\begin{aligned} f_t(F_t^{-1}(y)) &= \lambda_{F_t}(F_t^{-1}(y))\bar{F}_t(F_t^{-1}(y)) \\ &= \lambda_F(t + F_t^{-1}(y))\bar{F}_t(F_t^{-1}(y)) \\ &= \lambda_F(F^{-1}(1-(1-y)\bar{F}(t))) \cdot (1-y), \end{aligned} \tag{2.2}$$

where we used the fact that $F_t^{-1}(y) = F^{-1}(1-(1-y)\bar{F}(t)) - t$, $0 < y < 1$. Also,

$$\begin{aligned} g_t(F_t^{-1}(y)) &= \lambda_{G_t}(F_t^{-1}(y)) \cdot \bar{G}_t(F_t^{-1}(y)) \\ &= \lambda_G(t + F_t^{-1}(y))\bar{G}_t(F_t^{-1}(y)) \\ &= \lambda_G(F^{-1}(1-(1-y)\bar{F}(t)))\bar{G}_t(F_t^{-1}(y)). \end{aligned} \tag{2.3}$$

Combining (2.2) and (2.3),

$$\frac{f_t(F_t^{-1}(y))}{g_t(F_t^{-1}(y))} = \frac{\lambda_F(F^{-1}(1-(1-y)\bar{F}(t)))}{\lambda_G(F^{-1}(1-(1-y)\bar{G}(t)))} \cdot \frac{1-y}{\bar{G}_t(F_t^{-1}(y))}. \tag{2.4}$$

Under the assumption made on $\lambda_F(x)/\lambda_G(x)$,

$$\frac{\lambda_F(F^{-1}(1-(1-y)\bar{F}(t)))}{\lambda_G(F^{-1}(1-(1-y)\bar{F}(t)))} \geq \frac{\lambda_F(F^{-1}(y))}{\lambda_G(F^{-1}(y))}. \tag{2.5}$$

Now, when both F and G are new better than used (NBU), we have

$$\bar{G}(F_t^{-1}(u)) \leq \bar{G}(F^{-1}(u)). \tag{2.6}$$

Using (2.2)–(2.6),

$$\begin{aligned} \frac{f_t(F_t^{-1}(y))}{g_t(F_t^{-1}(y))} &\geq \frac{\lambda_F(F^{-1}(y))}{\lambda_G(F^{-1}(y))} \cdot \frac{1-y}{\bar{G}_t(F_t^{-1}(y))} \geq \frac{\lambda_F(F^{-1}(y))}{\lambda_G(F^{-1}(y))} \cdot \frac{1-y}{\bar{G}(F^{-1}(y))} \\ &= \frac{f(F^{-1}(y))}{g(F^{-1}(y))}. \end{aligned} \tag{2.7}$$

If follows from (2.1) that

$$I(F,G,t) \geq \int_0^\infty \log \frac{f(F^{-1}(y))}{g(F^{-1}(y))} dy = I(F,G).$$

This completes the proof.

Theorem 2.2 simply says that under the assumptions made, the disparity between two systems of age t is never smaller (larger) than the disparity when the two systems were new.

The following example gives a simple application of Theorem 2.2 in the set-up of competing risks.

Example 2.1. Let $X = \min\{Y_1, \ldots, Y_n\}$ where Y_i, $i = 1, 2, \ldots, n$, are independently and identically distributed with common distribution function G. Suppose also that G has a density g and that X has distribution function F. Then $\lambda_F(x)/\lambda_G(x) = n$ for all x. Further, it is easy to verify, if G is NBU (NWU) so is F. Hence, if G is NBU (NWU), $I(F, G; t) \geq (\leq) I(F, G)$ for all t.

The conclusion of Theorem 2.2 can be considerably strengthened if the distributions F and G are both IFR (DFR). F is said to be an increasing (decreasing) failure rate (IFR (DFR)) if $\lambda_F(t)$ is increasing (decreasing) in t. The following theorem provides useful sufficient conditions for monotonicity of $I(F, G; t)$ in t.

THEOREM 2.3. *If* (a) $\lambda_F(x)/\lambda_G(x)$ *is increasing (decreasing) in* x, *and* (b) *both* F *and* G *are* IFR (DFR) *then* $I(F, G; t)$ *is increasing (decreasing) in* t.

PROOF. When both F and G are IFR (DFR)

$$\frac{1}{\bar{G}_{t_1}(F_{t_1}^{-1}(y))} \geq (\leq) \frac{1}{\bar{G}_{t_2}(F_{t_2}^{-1}(y))}$$

for all $0 \leq t_2 < t_1$ and all $0 < y < 1$. Further, under assumption (a),

$$\frac{\lambda_F(F^{-1}(1-(1-y)\bar{F}(t)))}{\lambda_G(F^{-1}(1-(1-y)\bar{F}(t)))}$$

is increasing (decreasing) in $t \geq 0$ for all $0 < y < 1$. Hence, proceeding as in the previous theorem,

$$I(F,G,t) = \int_0^1 \log \frac{f_t(F_t^{-1}(y))}{g_t(F_t^{-1}(y))} dy$$

is increasing (decreasing) in $t \geq 0$.

The following example provides an application of Theorem 2.3 to the important case of Kullback-Leibler information for Weibull distributions.

Example 2.2. Let X and Y have Weibull survival functions $\bar{F}(x) = \exp(-(\lambda_1 x)^\alpha)$ and $\bar{G}(x) = \exp(-(\lambda_2 x)^\beta)$, $x > 0$, $\lambda_1, \lambda_2, \alpha, \beta > 0$. Then $\lambda_F(x)/\lambda_G(x)$ is increasing (decreasing) according as $\alpha \geq (\leq)\beta$. It follows that if $\alpha \geq \beta \geq 1$ $(0 < \alpha \leq \beta \leq 1)$ then $I(F, G, t)$ is increasing (decreasing) in $t \geq 0$.

Our next theorem provides easily verifiable sufficient conditions for $I(F_1, G; t)$ to be no more than $I(F_2, G; t)$.

THEOREM 2.4. *Consider three non-negative random variables X_1, X_2, and Y with probability density functions f_1, f_2 and g respectively. If* (a) $\frac{f_1(x)}{g(x)}$ *is increasing in x, and* (b) $\lambda_{F_2}(x) \leq \lambda_{F_1}(x)$, *then*

$$I(F_1, G; t) \leq I(F_2, G; t).$$

PROOF. From (1.2)

$$\begin{aligned}
(2.8)\qquad & I(F_1, G, t) - I(F_2, G; t) \\
&= \int_t^\infty \frac{f_1(x)}{\bar{F}_1(t)} \log \frac{f_1(x)\bar{G}(t)}{g(x)\bar{F}_1(x)} dx - \int_t^\infty \frac{f_2(x)}{\bar{F}_2(t)} \log \frac{f_1(x)\bar{G}(t)}{g(x)\bar{F}_1(t)} dx \\
&\quad - \int_t^\infty \frac{f_2(x)}{\bar{F}_2(t)} \log \frac{f_2(x)\bar{F}_1(t)}{f_1(x)\bar{F}_2(t)} \\
&\leq \int_t^\infty \frac{f_1(x)}{\bar{F}_1(t)} \log \frac{f_1(x)\bar{G}(t)}{g(x)\bar{F}_1(t)} dx - \int_t^\infty \frac{f_2(x)}{\bar{F}_2(t)} \log \frac{f_1(x)\bar{G}(t)}{g(x)\bar{F}_1(t)} dx \\
&= \int_t^\infty \frac{f_1(x)}{\bar{F}_1(t)} \log \frac{f_1(x)}{g(x)} - \int_t^\infty \frac{f_2(x)}{\bar{F}_2(t)} \log \frac{f_1(x)}{g(x)} dx,
\end{aligned}$$

where the inequality comes from the fact that $I(F_2, F_1; t) \geq 0$. Now, the assumption (b) is equivalent to saying that $X_t^{(2)}$ is stochastically larger than $X_t^{(1)}$ where $X_t^{(i)}$, $i = 1, 2$, is a random variable with density $f_i(x)/\bar{F}_i(t)$, $x > t$. It follows that, when $f_1(x)/g(x)$ is increasing, the expression (2.8) is nonpositive. This completes the proof.

The above result simply says that, under the assumptions of Theorem 2.4, at any point of time, the reference distribution G is closer to F_1 than to F_2.

We give below an application of Theorem 2.4 to nonhomogeneous Poisson processes.

Example 2.3. Consider a (possibly) nonhomogeneous Poisson process $\{N(t), t \geq 0\}$ with a differentiable mean value function $M(t) = E(N(t))$ such that $M(t) \to \infty$ as $t \to \infty$. Let R_n, $n = 1, 2, \ldots$, denote the occurrence time of the n-th event in such a process. Then R_n has distribution function F_n with density

$$f_n(x) = \frac{\{M(x)\}^{n-1}}{(n-1)!} f_1(x), \qquad x > 0, \quad n = 1, 2, \ldots$$

where $f_1(x) = -\frac{d}{dx}\exp(-M(x))$, $x > 0$. Clearly, $f_n(x)/f_1(x)$ is increasing in x. Moreover, $f_n(x)/f_{n-1}(x)$ is increasing in x so that $\lambda_{F_n}(x) \le \lambda_{F_{n-1}}(x)$. It follows from Theorem 2.4 that $I(F_n, F_1; t)$ is increasing in n for all $t \ge 0$.

Finally, we look at the behavior of $I(F, G; t)$ in the important special case when G is an exponential distribution.

THEOREM 2.5. *If* (a) $\lambda_F(t) \le \lambda$, (b) $\delta_F(t) = E(X - t \mid X > t)$ *is an increasing function of* t, *and* $G(x) = 1 - \exp(-\lambda x)$, $x > 0$, *then* $I(F, G; t)$ *is increasing in* t.

PROOF. It can be easily verified that

$$I(F, G; t) = H(F; t) - \log\lambda + \lambda\delta_F(t). \tag{2.9}$$

Differentiating (2.9) with respect to t, we get

$$\begin{aligned} I'(F, G, t) &= H'(F; t) + \lambda\delta'_F(t) \\ &\ge -\lambda_F(t)\delta'_F(t) + \lambda\delta'_F(t) \\ &= (\lambda - \lambda_F(t))\delta'_F(t) \ge 0. \end{aligned}$$

The above result simply says that the residual life time distribution of a system moves farther and farther away from exponential as the system is aging.

3. MDI principle and $I(F, G; t)$

Kullback (1954) introduced the minimum discrimination information (MDI) principle for statistical analysis. Since then it has been widely used; see Guiasu (1990). According to this principle, the best substitute F of the true distribution function G can be found by minimizing $I(F, G)$ subject to the constraints $\int f(x)dx = 1$ and $w_r = \int T_r(x)f(x)dx$, $r = 1, \dots, m$, where $T_r(x)$ are specified functions and w_r are specified constraints. The resulting distribution function F^* is called the MDI distribution and $I(F^*, G)$ the minimum discrimination information. In this section we show that MDI principle, when applied to modeling survival function, leads to the very useful proportional hazards model.

Once the current age t of a system is known, attention shifts from $I(F, G)$ to $I(F, G; t)$. Thus, the interest now is in finding the model closest to the true survival function $\bar{G}$ given $Y > t$. We, therefore consider the problem of minimizing $I(F, G; t)$ subject to constraints

$$\theta(t) = \int_t^\infty \left(-\log\left(\frac{\bar{G}(x)}{\bar{G}(t)}\right)\right) f(x)dx, \tag{3.1}$$

and

$$\int_0^\infty f(x)dx = 1. \tag{3.2}$$

In survival analysis $-\log \bar{G}(x)$, $-\log \bar{G}(x) = \int_0^x \lambda(u)du$, is called the cumulative hazard function. Therefore, intuitively speaking, the (3.1) simply means that the average cumulative hazard function after time t is $\theta(t)$.

The following theorem shows how the proportional hazards is obtained as the solution to the stated problem.

THEOREM 3.1. *Under the constraints* (3.1) *and* (3.2), F^* *minimizes* $I(F, G; t)$, *where*

$$\frac{\bar{F}^*(x)}{\bar{F}^*(t)} = \left(\frac{\bar{G}(x)}{\bar{G}(t)}\right)^{1/\theta_t}, \quad \text{for} \quad x > t. \tag{3.3}$$

PROOF. Since

$$\begin{aligned} M(\tau) &= \int_t^\infty \frac{g(x)}{\bar{G}(x)} \left(\exp\left(-\tau \log \frac{\bar{G}(x)}{\bar{G}(t)}\right)\right) dx \\ &= \frac{1}{1-\tau} < \infty, \end{aligned}$$

for $\tau < 0$. Therefore applying Theorem (2.1) p. 39 of Kullback (1959), $I(F, G; t)$ is minimized if

$$\begin{aligned} \frac{f^*(x)}{\bar{F}^*(t)} &= \left(\exp\left(-\frac{\theta(t)-1}{\theta(t)} \log \frac{\bar{G}(x)}{\bar{G}(t)}\right)\right) \frac{g(x)}{\bar{G}(t)} (\theta(t))^{-1} \\ &= (\theta(t))^{-1} \left(\frac{g(x)}{\bar{G}(x)}\right) \left(\frac{\bar{G}(x)}{\bar{G}(t)}\right)^{1/(\theta(t)-1)}, \quad x > t. \end{aligned}$$

That is $\frac{\bar{F}^*(x)}{\bar{F}^*(t)} = (\frac{\bar{G}(x)}{\bar{G}(t)})^{1/\theta(t)}$, $x > t$. This completes the proof.

From Theorem 3.1, if $t = 0$, then we get $\bar{F}^*(x) = (\bar{G}(x))^{1/\theta(0)}$.

The following example illustrates an application of Theorem 3.1.

Example 3.1. Suppose $\bar{G}(x) = \exp(-x)$. Then, under the constraints $\int_0^\infty f(x)dx = 1$ and $\int_t^\infty (x-t)f(x)dx = \theta(t)$, $\frac{\bar{F}^*(x)}{\bar{F}^*(t)} = \exp(-\frac{1}{\theta(t)}(x-t))$. This means, the closest residual life time distribution is the exponential.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiádo, Budapest.

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure, *J. Roy. Statist. Soc. Ser. B*, **21**, 411–421.

Ebrahimi, N. and Kirmani, S. N. U. A. (1995). A characterization of the proportional hazards model through a measure of discrimination between two residual life distribution, *Biometrika* (to appear).

Ebrahimi, N., Habibullah, M. and Soofi, E. S. (1992). Testing exponentiality based on Kullback-Leibler information, *J. Roy. Statist. Soc. Ser. B*, **54**, 739–748.

Efron, B. (1981). Censored data and the bootstrap, *J. Amer. Statist. Assoc.*, **76**, 312–319.

Guiasu, S. (1990). A classification of the main probability distributions by minimizing weighted logarithmic measure of deviation, *Ann. Inst. Statist. Math.*, **42**, 269–279.

Kullback, S. (1954). Certain inequalities in information theory and the Cramér-Rao inequality, *Ann. Math. Statist.*, **25**, 745–751.

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

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Math. Statist.*, **22**, 79–86.