Ann. Inst. Statist. Math.
35 (1983), Part A, 151-165

THE DECOMPOSITION OF THE FISHER INFORMATION

NOBUO INAGAKI

(Received July 19, 1982; revised Oct. 25, 1982)

1. Introduction and summary

Let $(R^n, \mathfrak{B}^n)$ be the Borel measurable space of the n-dimensional Euclidean space and let a parameter space Θ be an open set of the k-dimensional Euclidean space R^k with $k<n$. We consider a family of probability measures on $\mathfrak{B}^n$, $\Pi=\{P_\theta;\ \theta\in\Theta\}$, which are dominated by a σ-finite measure μ on $\mathfrak{B}^n$. That is, $P_\theta \ll \mu$ for all $\theta\in\Theta$. We denote the Radon-Nikodym derivative of P_θ with respect to μ by

$$dP_\theta/d\mu = f(x;\theta)\ , \qquad \theta\in\Theta\ . \tag{1.1}$$

For an observation X having the distribution P_θ, let $T=T(X)$ be an estimator of θ which is a measurable function from R^n to Θ $(\subset R^k)$, and consider the factorization of the likelihood function

$$f(X;\theta)=g(T;\theta)h(X;\theta\,|\,T) \tag{1.2}$$

which is proposed in Section 3. The main result of this paper is that this factorization induces a decomposition of the Fisher information as follows:

$$I_f(\theta)=I_g(\theta)+I_h(\theta)\ . \tag{1.3}$$

Such a decomposition plays an important role in the theory of statistical estimation; see, for instance, Edwards [2], Barndorff-Nielsen [1], and Shimizu [7].

In Section 2, we introduce some notions of differentiability of the square root of the likelihood function and the likelihood ratio function in the line of LeCam [5]. In Section 3, we propose a factorization of the likelihood function of the observation into the marginal and conditional likelihood functions, and state several properties related to the factorization. Section 4 is devoted to a study of differentiations of the marginal and conditional likelihood functions. Our main aim is to show that their differentiability is inherited from that of the likelihood function of the observation. We utilize the results related to conditioning devised in Loève [6], especially the concept of the relative conditional expectation.

2. Differentiations of the square root of the likelihood function and the likelihood ratio function

We consider the square root of the likelihood function and the likelihood ratio function defined by

$$\phi(\theta)=f^{1/2}(X;\theta) \tag{2.1}$$

and

$$X_\theta(\tau)=\phi(\theta+\tau)/\phi(\theta)-1=f^{1/2}(X;\theta+\tau)/f^{1/2}(X;\theta)-1\,, \quad \text{for } \tau\in R^k\,, \tag{2.2}$$

respectively. The L_2-norm of ϕ is equal to 1:

$$\|\phi(\theta)\|=\left\{\int\phi^2(\theta)\mu(dx)\right\}^{1/2}=1\,. \tag{2.3}$$

Hájek and Sidák [3] point out that the function $X_\theta(\tau)$ has additional advantages than the log likelihood ratio function, $\log\{f(X;\theta+\tau)/f(X;\theta)\}$, since $X_\theta(\tau)$ has always finite variance and is not troubled with the circumstance of whether probability density functions have the common support or not. But it is necessary to take into consideration the following quantity which evaluates the difference of their supports:

$$\begin{aligned}\beta_\theta^f(\tau)&=\int_{\{x;\,f(x;\theta)=0\}}f(x;\theta+\tau)\mu(dx)\,, \qquad \text{say}\,,\\ &=\int\{1-\chi_\theta^f(x)\}f(x;\theta+\tau)\mu(dx)\,,\end{aligned} \tag{2.4}$$

where χ_θ^f is the indicator function of the support of $f(x;\theta)$, $S_f(\theta)=\{x;\ f(x;\theta)>0\}$ (say).

We define $f(x;\theta)$ to be differentiable at θ in mean with respect to μ if there exists an absolutely integrable vector-valued function $\dot f(x;\theta)$ called the derivative of $f(x;\theta)$ such that

$$\lim_{|\tau|\to 0}\frac{1}{|\tau|}\int|f(x;\theta+\tau)-f(x;\theta)-\dot f(x;\theta)\cdot\tau|\,\mu(dx)=0\,. \tag{2.5}$$

We define that $\phi(\theta)$ is differentiable at θ in quadratic mean with respect to μ if there exists a square integrable vector-valued function $\dot\phi(\theta)$ called the derivative of ϕ such that

$$\lim_{|\tau|\to 0}\|\phi(\theta+\tau)-\phi(\theta)-\dot\phi(\theta)\cdot\tau\|/|\tau| \tag{2.6}$$

$$=\lim_{|\tau|\to 0}\frac{1}{|\tau|}\left\{\int|\phi(\theta+\tau)-\phi(\theta)-\dot\phi(\theta)\cdot\tau|^2\mu(dx)\right\}^{1/2}=0\ .$$

Further, we say that X_θ is differentiable at θ in mean or in quadratic mean with respect to P_θ if there exists a random vector $\dot X(\theta)$ called the derivative of X_θ such that $\dot X(\theta)$ is absolutely integrable and

$$\lim_{|\tau|\to 0}\frac{1}{|\tau|}E_\theta|X_\theta(\tau)-\dot X(\theta)\cdot\tau|=0 \tag{2.7}$$

or such that $\dot X(\theta)$ is square integrable and

$$\lim_{|\tau|\to 0}\frac{1}{|\tau|}\{E_\theta|X_\theta(\tau)-\dot X(\theta)\cdot\tau|^2\}^{1/2}=0\ , \tag{2.8}$$

respectively. Of course, X_θ is differentiable in mean with the common derivative if it is so in quadratic mean.

The following lemma is called "L_r-convergence theorem" (see Loève [6], p. 165).

LEMMA 2.1. *Let $\{X_n\}_{n=1}^\infty$ and X be in an L_r-space. Then,*
(i) $\|X_n-X\|_r\to 0$ *as* $n\to\infty$,
if and only if
(ii) $X_n\to X$ *in probability as* $n\to\infty$ *and*
$\|X_n\|_r\to\|X\|_r$ *as* $n\to\infty$.

The following theorem and its corollary are rearrangements of several results due to LeCam [5], the proofs of which we give for the present paper to be self-contained.

THEOREM 2.1. (i) *X_θ is differentiable in quadratic mean at θ if $\phi(\theta)$ is differentiable in quadratic mean at θ.*
(ii) *The converse is also true if the following condition is assumed*:

$$\lim_{|\tau|\to 0}\frac{1}{|\tau|^2}\beta_\theta^f(\tau)=0\ . \tag{2.9}$$

PROOF. For $\lambda>0$ and $\tau\in R^k$ with $|\tau|=1$, we have

$$\begin{aligned}\|\{\phi(\theta+\lambda\tau)-\phi(\theta)\}/\lambda-\dot\phi(\theta)\cdot\tau\|^2 \\ =E_\theta|X_\theta(\lambda\tau)/\lambda-\dot X(\theta)\cdot\tau|^2+\|(1-\chi_\theta^f)\{\phi(\theta+\lambda\tau)/\lambda-\dot\phi(\theta)\cdot\tau\}\|^2\ ,\end{aligned} \tag{2.10}$$

where we set

$$\dot X(\theta)=\dot\phi(\theta)/\phi(\theta)\qquad\text{or}\qquad\dot\phi(\theta)=\dot X(\theta)\phi(\theta)\ . \tag{2.11}$$

(i) is an immediate consequence of (2.10) with $\dot X(\theta)=\dot\phi(\theta)/\phi(\theta)$. Further, we have from (2.10) that

$$\lim_{\lambda\to 0}\frac{1}{\lambda^2}\beta_\theta^f(\lambda\tau)=\|(1-\chi_\theta^f)\dot{\phi}(\theta)\cdot\tau\|^2\,, \tag{2.12}$$

and hence that

$$\lim_{\lambda\to 0}\frac{1}{\lambda}\beta_\theta^f(\lambda\tau)=0\;. \tag{2.13}$$

Now, we shall show (ii). Letting $\dot{\phi}(\theta)=\dot{X}(\theta)\phi(\theta)$, we have $(1-\chi_\theta^f)\cdot\dot{\phi}(\theta)=0$. Hence (2.10) becomes

$$\begin{aligned}\|\{\phi(\theta+\lambda\tau)-\phi(\theta)\}/\lambda-\dot{\phi}(\theta)\cdot\tau\|^2\\ =E_\theta|X_\theta(\lambda\tau)/\lambda-\dot{X}(\theta)\cdot\tau|^2+\beta_\theta^f(\lambda\tau)/\lambda^2\,.\end{aligned} \tag{2.14}$$

This and (2.9) lead to the conclusion of (ii).

COROLLARY 2.1. *If X_θ is differentiable in quadratic mean at θ and the condition* (2.13) *holds, then*

$$E_\theta\dot{X}(\theta)=0\;. \tag{2.15}$$

PROOF. Let $\lambda>0$ and $\tau\in R^k$ with $|\tau|=1$. By Lemma 2.1, it follows from (2.1)-(2.3), (2.8) and (2.13) that

$$\begin{aligned}-2E_\theta\{\dot{X}(\theta)\cdot\tau\}&=\lim_{\lambda\to 0}-2E_\theta\{X_\theta(\lambda\tau)/\lambda\}\\ &=\lim_{\lambda\to 0}\|\phi(\theta+\lambda\tau)-\phi(\theta)\|^2/\lambda\\ &=\lim_{\lambda\to 0}\{E_\theta|X_\theta(\lambda\tau)|^2/\lambda+\beta_\theta^f(\lambda\tau)/\lambda\}=0\;.\end{aligned} \tag{2.16}$$

THEOREM 2.2. *If X_θ is differentiable in quadratic mean at θ and the condition* (2.13) *holds, then $f(x;\theta)$ is differentiable in mean at θ with the derivative*

$$\dot{f}(x;\theta)=2\dot{X}(\theta)f(x;\theta)\;. \tag{2.17}$$

PROOF. It is obvious that $\dot{f}(x;\theta)$ is absolutely integrable since $\dot{X}(\theta)$ is so with respect to P_θ. For $\lambda>0$ and τ with $|\tau|=1$, it follows from (2.2) and (2.17) that

$$\begin{aligned}&\int|\{f(x;\theta+\lambda\tau)-f(x;\theta)\}/\lambda-2\dot{X}(\theta)\cdot\tau f(x;\theta)|\mu(dx)\\ &\quad=E_\theta|\{X_\theta(\lambda\tau)/\lambda\}^2\lambda+2X_\theta(\lambda\tau)/\lambda-2\dot{X}(\theta)\cdot\tau|+\beta_\theta^f(\lambda\tau)/\lambda\\ &\quad\leqq E_\theta|X_\theta(\lambda\tau)/\lambda|^2\lambda+2E_\theta|X_\theta(\lambda\tau)/\lambda-\dot{X}(\theta)\cdot\tau|^2+\beta_\theta^f(\lambda\tau)/\lambda\;.\end{aligned}$$

Therefore, by Lemma 2.1 we see from (2.8) and (2.13) that the last three terms tend to zero as $\lambda\to 0$. Hence, the proof is complete.

COROLLARY 2.2. *If $\phi(\theta)$ is differentiable in quadratic mean at θ,* *then* $f(x;\theta)$ *is differentiable in mean at* θ *with the derivative*

$$f(x;\theta)=2\dot{\phi}(\theta)\phi(\theta)\ . \tag{2.18}$$

Remarks. (a) By Lemma 2.1, it is easy to see that

$$\int \dot{f}(x;\theta)\mu(dx)=0 \tag{2.19}$$

if $f(x;\theta)$ is differentiable in mean at θ.
(b) According to LeCam [5], the condition (2.9) implies that $\{P_\theta^m\}$ and $\{P_{\theta+\tau/\sqrt{m}}^m\}$ are contiguous.

3. Factorization of likelihood function

We refer to Chapter VIII of Loève [6] for the properties of conditioning mentioned in this section. From the fact that $P_\theta \ll \mu$ with $dP_\theta/d\mu=f(X;\theta)$, we see that the same domination holds between the measures on $(R^k, \mathfrak{B}^k)$ induced by an estimator of θ, $T=T(X)$: $P_\theta^T \ll \mu^T$. We denote the Radom-Nikodym derivative of P_θ^T with respect to μ^T by

$$dP_\theta^T/d\mu^T=g(t;\theta)\ , \qquad \theta\in\Theta\ , \tag{3.1}$$

which we may regard as the marginal likelihood function of T. For the support of $g(T(x);\theta)$,

$$S_g(\theta)=\{x;\, g(T(x);\theta)>0\}\ , \qquad \text{say}\ ,$$

it holds that $S_g(\theta)\in T^{-1}(\mathfrak{B}^k)$ and

$$P_\theta\{S_g^c(\theta)\}=\int_{S_g^c(\theta)} f(x;\theta)\mu(dx)=\int_{S_g^c(\theta)} g(T(x);\theta)\mu(dx)=0\ . \tag{3.2}$$

The following theorem is an immediate consequence of (3.2).

THEOREM 3.1. *Set*

$$h(x;\theta\,|\,t)=\begin{cases} f(x;\theta)/g(t;\theta)\ , & \text{if } T(x)=t \text{ and } g(t;\theta)>0\ ,\\ 1\ , & \text{if } T(x)=t \text{ and } g(t;\theta)=0\ , \quad \text{and}\\ 0\ , & \text{otherwise}\ . \end{cases} \tag{3.3}$$

Then, it holds:

$$f(X;\theta)=g(T;\theta)h(X;\theta|T)\ , \qquad a.s.\ [P_\theta]\ . \tag{3.4}$$

We call this the factorization of the likelihood function of observation, $f(X;\theta)$, into the marginal likelihood function of an estimator

$T=T(X)$, $g(T;\theta)$, and the conditional likelihood function of X given T, $h(X;\theta|T)$.

On the other hand, since P_θ is a probability measure on $(R^n, \mathfrak{B}^n)$, it is known that there exists a regular conditional probability of P_θ given $T(X)=t$, $P_\theta(\cdot\,|t)$, satisfying

$$P_\theta(A\cap T^{-1}(B))=\int_B P_\theta(A|t)g(t;\theta)\mu^T(dt)$$
$$=\int_B\left\{\int_A P_\theta(dx|t)\right\}g(t;\theta)\mu^T(dt)\,,$$

for all $A\in\mathfrak{B}^n$ and all $B\in\mathfrak{B}^k$. More generally, it follows that for an integrable function $u(x)v(T(x))$

$$E_\theta\{u(X)v(T(X))\}=\int u(x)v(T(x))f(x;\theta)\mu(dx) \tag{3.5}$$
$$=\int\left\{\int u(x)\,P_\theta(dx|t)\right\}v(t)g(t;\theta)\mu^T(dt)\,.$$

We put the sense of (3.5) down symbolically as

$$f(x;\theta)\mu(dx)=g(t;\theta)\mu^T(dt)\,P_\theta(dx|t)\,. \tag{3.6}$$

Then, comparing (3.4) with (3.6), we could identify

$$h(x;\theta|t)\qquad\text{and}\qquad P_\theta(dx|t)/\mu(dx)\,,$$

in many regular cases. But we note that (3.6) means only (3.5) while (3.4) itself makes sense. It is well-known that, if $T(X)$ is a sufficient statistic for $\Pi=\{P_\theta;\ \theta\in\Theta\}$, there exists the conditional probability kernel $h(x|t)$, which is independent of parameter θ, so that the factorization theorem holds.

Now, similarly as in (2.1) and (2.2) we consider the square roots of the marginal and conditional likelihood functions and their likelihood ratio functions defined by

$$\psi(\theta)=g^{1/2}(T;\theta)\,,\qquad \zeta(\theta)=h^{1/2}(X;\theta|T)\,, \tag{3.7}$$

and

$$\begin{aligned}Y_\theta(\tau)&=\psi(\theta+\tau)/\psi(\theta)-1=g^{1/2}(T;\theta+\tau)/g^{1/2}(T;\theta)-1\,,\\ Z_\theta(\tau)&=\zeta(\theta+\tau)/\zeta(\theta)-1=h^{1/2}(X;\theta+\tau|T)/h^{1/2}(X;\theta|T)-1\,,\end{aligned} \tag{3.8}$$

for θ and $\theta+\tau\in\Theta$, respectively. Then, we have:

LEMMA 3.1.

$$E_\theta|X_\theta(\tau)-Y_\theta(\tau)-Z_\theta(\tau)-Y_\theta(\tau)Z_\theta(\tau)|=0\,. \tag{3.9}$$

PROOF. It follows from (3.3) and (3.4) that

$$E_\theta|(X_\theta(\tau)+1)-(Y_\theta(\tau)+1)(Z_\theta(\tau)+1)|$$
$$=\int_{\{g(T(x);\theta+\tau)=0\}}\phi(\theta+\tau)\phi(\theta)\mu(dx)=0 .$$

This is the conclusion of this lemma.

We use the "relative conditional expectation" of an integrable function $u(x)$ given $T(x)=t$ with respect to a σ-finite measure μ, $E_\mu(u|t)$, such that $E_\mu(u|t)$ is $\mathfrak{B}^k$-measurable and satisfies

$$\int u(x)\mu(dx)=\int E_\mu(u|t)\mu^T(dt)=\int E_\mu(u|T(x))\mu(dx) . \tag{3.10}$$

We could also refer to Loève [6] above, or go back to Halmos and Savage [4], for the concept of the relative conditional expectation.

LEMMA 3.2. (i) $g(T;\theta)=E_\mu[f(X;\theta)|T]$, *a.s.* $[\mu^T]$.
(ii) *More generally, for an integrable function* $u(x)$,

$$E_\theta[u(X)|T]g(T;\theta)=E_\mu[u(X)f(X;\theta)|T] , \qquad a.s.\ [\mu^T] . \tag{3.11}$$

PROOF. It follows from the definitions of conditional and relative conditional expectations according to P_θ and μ, respectively, that

$$E_\theta\{u(X)\}=\int E_\theta[u(X)|t]g(t;\theta)\mu^T(dt)$$
$$=\int E_\mu[u(X)f(X;\theta)|T]\mu^T(dt) .$$

This means (3.11).

It is easy to see from (3.3) and Lemma 3.2 that $h(x;\theta|t)$ behaves as a conditional probability density function: that is $h(x;\theta|t)\geqq 0$ and

$$E_\mu[h(X;\theta|t)|t]=1 \qquad \text{a.s. } [\mu^T] . \tag{3.12}$$

Further, we have the following relations which play useful and essential roles in the present paper:

LEMMA 3.3.
(i) $E_\theta\{X_\theta(\tau)\}=-\|\phi(\theta+\tau)-\phi(\theta)\|^2/2=-\{E_\theta|X_\theta(\tau)|^2+\beta_\theta^f(\tau)\}/2.$
(ii) $E_\theta\{Y_\theta(\tau)\}=-\|\psi(\theta+\tau)-\psi(\theta)\|^2/2=-\{E_\theta|Y_\theta(\tau)|^2+\beta_\theta^g(\tau)\}/2,$
where $\beta_\theta^g(\tau)$ *is defined in the same way as in* (2.4) *taking* $g(t;\theta)$ *in place of* $f(x;\theta)$.
(iii) $E_\theta[Z_\theta(\tau)|T]=-E_\mu[|\zeta(\theta+\tau)-\zeta(\theta)|^2|T]/2$
$=-\{E_\mu[|Z_\theta(\tau)|^2|T]+E_\mu[(1-\chi_\theta^h(X))h(X;\theta+\tau|T)|T]\}/2,$
where χ_θ^h *is the indicator function of the support of* $h(x;\theta|T)$.

PROOF. (i) is immediately proved from (2.1)–(2.4) and (ii) is similarly done. It follows from (3.11) that

$$E_\theta[Z_\theta(\tau)|T]=E_\theta[h^{1/2}(X;\theta+\tau|T)h^{1/2}(X;\theta|T)-1|T]\ . \tag{3.13}$$

This and (3.12) lead to the conclusion of (iii).

Hereafter, we show that $g(T;\theta)$ and, consequently, $h(X;\theta|T)$ also inherit from $f(X;\theta)$ such properties as (2.9) and (2.16).

THEOREM 3.2. *Suppose*

$$\lim_{|\tau|\to 0} E_\theta\{X_\theta(\tau)\}/|\tau|=0\ , \tag{3.14}$$

then it holds that

$$\lim_{|\tau|\to 0} E_\theta\{Y_\theta(\tau)\}/|\tau|=0\ , \tag{3.15}$$

and

$$\lim_{|\tau|\to 0} E_\theta\{(Y_\theta(\tau)+1)\,E_\theta[Z_\theta(\tau)|T]\}/|\tau|=0\ . \tag{3.16}$$

PROOF. We see from (3.9) and by Lemma 3.3 that

$$E_\theta\{X_\theta(\tau)\}/|\tau|=E_\theta\{Y_\theta(\tau)\}/|\tau|+E_\theta\{(Y_\theta(\tau)+1)\,E_\theta[Z_\theta(\tau)|T]\}/|\tau|\ ,$$

each term of which is non-positive, and hence we conclude that the convergence to zero of the term on the left-hand side implies the same fact of each term on the right-hand side.

COROLLARY 3.1. *Suppose the same assumption as in Theorem* 3.2 *and*

$$\overline{\lim_{|\tau|\to 0}}\, E_\theta|Y_\theta(\tau)|/|\tau|<\infty\ . \tag{3.17}$$

Then, we have

$$\lim_{|\tau|\to 0} E_\theta|Y_\theta(\tau)Z_\theta(\tau)|/|\tau|=0\ , \tag{3.18}$$

and hence,

$$\lim_{|\tau|\to 0} E_\theta\{Z_\theta(\tau)\}/|\tau|=0\ . \tag{3.19}$$

PROOF. Let $\lambda>0$ and $|\tau|=1$. By (iii) of Lemma 3.3, it holds that

$$\begin{aligned} E_\theta|Z_\theta(\lambda\tau)|^2/\lambda &\leqq -2\,E_\theta\{Z_\theta(\lambda\tau)\}/\lambda \\ &\leqq 2\,E_\theta|Y_\theta(\lambda\tau)|/\lambda-2\,E_\theta\{(Y_\theta(\lambda\tau)+1)\,E_\theta[Z_\theta(\lambda\tau)|T]\}/\lambda\ , \end{aligned}$$

and therefore that

$$\overline{\lim_{\lambda\to 0}}\, E_\theta|Z_\theta(\lambda\tau)|^2/\lambda\leqq 2\,\overline{\lim_{\lambda\to 0}}\, E_\theta|Y_\theta(\lambda\tau)|/\lambda<\infty\ . \tag{3.20}$$

Since

$$E_\theta|Y_\theta(\lambda\tau)Z_\theta(\lambda\tau)|/\lambda \leqq \{E_\theta|Y_\theta(\lambda\tau)|^2/\lambda\}^{1/2}\{E_\theta|Z_\theta(\lambda\tau)|^2/\lambda\}^{1/2},$$

(3.18) is proved by (ii) of Lemma 3.3, (3.15) and (3.20). This and (3.16) lead to (3.19). The proof is complete.

THEOREM 3.3. (i) *The condition* (2.9) *for* $f(X;\theta)$ *produces the same condition for* $g(T;\theta)$:

$$\lim_{|\tau|\to 0} \beta_\theta^g(\tau)/|\tau|^2 = 0 . \tag{3.21}$$

(ii) *Suppose*

$$\overline{\lim_{|\tau|\to 0}}\, E_\theta|X_\theta(\tau)|^2/|\tau|^2 < \infty , \tag{3.22}$$

then the condition (2.9) *produces the similar one for* $h(X;\theta|T)$:

$$\lim_{|\tau|\to 0} E_\theta\{E_\mu[(1-\chi_\theta^h(X))h(X;\theta+\tau|T)|T]\}/|\tau|^2 = 0 . \tag{3.23}$$

Thus, it holds in these situations that

$$\begin{aligned} &\lim_{|\tau|\to 0} \{2\,E_\theta(X_\theta(\tau)) + E_\theta|X_\theta(\tau)|^2\}/|\tau|^2 = 0 , \\ &\lim_{|\tau|\to 0} \{2\,E_\theta(Y_\theta(\tau)) + E_\theta|Y_\theta(\tau)|^2\}/|\tau|^2 = 0 , \qquad \textit{and} \\ &\lim_{|\tau|\to 0} \{2\,E_\theta(Z_\theta(\tau)) + E_\theta|Z_\theta(\tau)|^2\}/|\tau|^2 = 0 . \end{aligned} \tag{3.24}$$

PROOF. (i) Since $P_\theta\{g(T(X);\theta)=0\}=0$, it is sufficient for (3.21) to show that, for $A\in\mathfrak{B}^n$ with $P_\theta(A)=0$,

$$\lim_{|\tau|\to 0} P_{\theta+\tau}(A)/|\tau|^2 = 0 . \tag{3.25}$$

Recalling that $S_f(\theta)$ is the support of $f(x;\theta)$, we easily see

$$0 = P_\theta(A) = \int_{A\cap S_f(\theta)} f(x;\theta)\mu(dx)$$

and therefore, we have

$$\mu[A\cap S_f(\theta)] = 0 .$$

Thus, we have

$$P_{\theta+\tau}(A) = \int_{A\cap S_f(\theta)} f(x;\theta+\tau)\mu(dx) + \int_{A\cap S_f^c(\theta)} f(x;\theta+\tau)\mu(dx) \leqq \beta_\theta^f(\tau) .$$

Consequently, this and (2.9) imply (3.25) and hence (3.21).
(ii) Set

$$\beta_\theta^h(\tau) = E_\mu[(1-\chi_\theta^h(X))h(X;\theta+\tau|T)|T] . \tag{3.26}$$

It follows from (2.9), (3.3) and (3.11) that

$$(3.27)\qquad \begin{aligned} E_{\theta+\tau}\{\beta_\theta^h(\tau)\}/|\tau|^2 &= E_{\theta+\tau}\{1-\chi_\theta^h(X)\}/|\tau|^2 \\ &\leqq E_{\theta+\tau}\{1-\chi_\theta^f(X)\}/|\tau|^2 = \beta_\theta^f(\tau)/|\tau|^2 \to 0 \\ &\qquad \text{as } |\tau|\to 0 . \end{aligned}$$

Since $0\leqq \beta_\theta^h(\tau)\leqq 1$, it holds that

$$(3.28)\qquad \begin{aligned} &|E_{\theta+\tau}\{\beta_\theta^h(\tau)\}-E_\theta\{\beta_\theta^h(\tau)\}|/|\tau|^2 \\ &= \left|\int (\psi(\theta+\tau)-\psi(\theta))(\psi(\theta+\tau)+\psi(\theta))\beta_\theta^h(\tau)\mu^T(dt)\right| \Big/ |\tau|^2 \\ &\leqq \beta_\theta^g(\tau)/|\tau|^2+\{E_\theta|Y_\theta(\tau)|^2/|\tau|^2\}^{1/2}[E_{\theta+\tau}\{\beta_\theta^h(\tau)\}/|\tau|+E_\theta\{\beta_\theta^h(\tau)\}/|\tau|] . \end{aligned}$$

Since (2.9) and (3.22) imply

$$\varlimsup_{|\tau|\to 0}[-E_\theta\{X_\theta(\tau)\}/|\tau|^2]=\varlimsup_{|\tau|\to 0} E_\theta|X_\theta(\tau)|^2/|\tau|^2<\infty ,$$

(3.14) holds. Further, by using (3.21) and similarly as the proof of Theorem 3.2 we see that

$$\begin{aligned} \varlimsup_{|\tau|\to 0} E_\theta|Y_\theta(\tau)|^2/|\tau|^2 &= \varlimsup_{|\tau|\to 0}[-E_\theta\{Y_\theta(\tau)\}/|\tau|^2] \\ &\leqq \varlimsup_{|\tau|\to 0}[-E_\theta\{X_\theta(\tau)\}/|\tau|^2]<\infty \end{aligned}$$

and hence, that (3.17) holds. Then, by (iii) of Lemma 3.3 and Corollary 3.1 we have

$$(3.29)\qquad \varlimsup_{|\tau|\to 0} E_\theta\{\beta_\theta^h(\tau)\}/|\tau|=0 .$$

Thus, we obtain (3.23) from (3.21), (3.22) and (3.27)-(3.29).

4. Differentiability of the marginal and conditional likelihood functions and decomposition of Fisher information

In this section we show that the marginal likelihood function of an estimator T, $g(T;\theta)$, and the conditional likelihood function of X given T, $h(X;\theta|T)$ inherit differentiabilities with respect to θ from the likelihood function of observation X, $f(X;\theta)$, while we show in the previous section that the formers inherit some other regularity conditions from the last.

THEOREM 4.1. *If $f(x;\theta)$ is differentiable in mean at θ (see (2.5)), then $g(t;\theta)$ is also differentiable in mean at θ:*

$$(4.1)\qquad \lim_{|\tau|\to 0}\int |g(t;\theta+\tau)-g(t;\theta)-\dot g(t;\theta)\cdot\tau|\mu^T(dt)/|\tau|=0$$

with the derivative

$$\dot{g}(T;\theta)=E_\mu[\dot{f}(X;\theta)|T] \ . \tag{4.2}$$

PROOF. Let $\lambda>0$, $|\tau|=1$ and $\dot{g}(T;\theta)=E_\mu[\dot{f}(X;\theta)|T]$. It follows from (2.5) and (i) of Lemma 3.2 that

$$\begin{aligned}
&\int |\{g(t;\theta+\lambda\tau)-g(t;\theta)\}/\lambda-\dot{g}(t;\theta)\cdot\tau|\mu^T(dt)\\
&\quad=\int |E_\mu[\{f(X;\theta+\lambda\tau)-f(X;\theta)\}/\lambda-\dot{f}(X;\theta)\cdot\tau|T=t]|\mu^T(dt)\\
&\quad\leqq\int |\{f(X;\theta+\lambda\tau)-f(X;\theta)\}/\lambda-\dot{f}(X;\theta)\cdot\tau|\mu(dx)\to 0 \ ,\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{as } \lambda\to 0 \ .
\end{aligned}$$

This is the conclusion of the theorem.

Similarly as in (2.9), it is obvious that

$$\int \dot{g}(t;\theta)\mu^T(dt)=0 \ .$$

THEOREM 4.2. *If X_θ is differentiable in quadratic mean at θ (see* (2.8)) *and if the condition* (2.13) *holds, then not only X_θ but also Y_θ and Z_θ are differentiable in mean at θ, letting derivatives be defined by*

$$\dot{Y}(\theta)=E_\theta[\dot{X}(\theta)|T] \ , \qquad \textit{say} \ , \tag{4.3}$$

and

$$\begin{aligned}\dot{Z}(\theta)&=\dot{X}(\theta)-E_\theta[\dot{X}(\theta)|T] \ , \qquad \textit{say} \ ,\\ &=\dot{X}(\theta)-\dot{Y}(\theta) \ ,\end{aligned} \tag{4.4}$$

respectively.

PROOF. 1°. Let $\lambda>0$ and $|\tau|=1$. Since differentiability in quadratic mean implies differentiability in mean with the same derivative, it is obvious that X_θ is differentiable in mean:

$$\lim_{\lambda\to 0} E_\theta|X_\theta(\lambda\tau)/\lambda-\dot{X}(\theta)\cdot\tau|=0 \ . \tag{4.5}$$

2°. Theorem 2.2 states that under the same condition as in the present theorem $f(x;\theta)$ is differentiable in mean with the derivative $\dot{f}(X;\theta)=2\dot{X}(\theta)f(X;\theta)$. Therefore, by Theorem 4.1 it follows from (3.11) and (4.3) that $g(T;\theta)$ is differentiable in mean with the derivative

$$\dot{g}(T;\theta)=2\dot{Y}(\theta)g(T;\theta) \ . \tag{4.6}$$

We see from (3.7), (3.8) and (4.6) that

$$E_\theta|Y_\theta(\lambda\tau)/\lambda-\dot{Y}(\theta)\cdot\tau| \tag{4.7}$$

$$=\int |\{\psi(\theta+\lambda\tau)\psi(\theta)-g(t;\theta)\}/\lambda-\dot{g}(t;\theta)\cdot\tau/2|\mu^T(dt)$$
$$\leqq\frac{1}{2}\int |\{g(t;\theta+\lambda\tau)-g(t;\theta)\}/\lambda-\dot{g}(t;\theta)\cdot\tau|\mu^T(dt)$$
$$+\frac{1}{2}\|\psi(\theta+\lambda\tau)-\psi(\theta)\|^2/\lambda .$$

It follows from Corollary 2.1 and (2.16) that the condition (3.14) in Theorem 3.2 holds and hence, from (ii) of Lemma 3.3 and (3.15) that the last term in (4.7) converges to zero as $\lambda\to 0$. This and the differentiability of $g(T;\theta)$ conclude that of Y_θ in mean:

$$\lim_{\lambda\to 0} E_\theta|Y_\theta(\lambda\tau)/\lambda-\dot{Y}(\theta)\cdot\tau|=0 . \tag{4.8}$$

3°. By Lemma 2.1 and (4.8), it holds that

$$\lim_{\lambda\to 0} E_\theta|Y_\theta(\lambda\tau)|/\lambda=E_\theta|\dot{Y}(\theta)\cdot\tau|<\infty . \tag{4.9}$$

Therefore, by Corollary 3.1 we have that (3.18) holds. It follows from (3.9), (4.3) and (4.4) that

$$\begin{aligned} E_\theta|\{X_\theta(\lambda\tau)/\lambda-\dot{X}(\theta)\cdot\tau\}-\{Y_\theta(\lambda\tau)/\lambda-\dot{Y}(\theta)\cdot\tau\}& \\ -\{Z_\theta(\lambda\tau)/\lambda-\dot{Z}(\theta)\cdot\tau\}+Y_\theta(\lambda\tau)Z_\theta(\lambda\tau)/\lambda|&=0 . \end{aligned} \tag{4.10}$$

Thus, we have that

$$\begin{aligned} &E_\theta|Z_\theta(\lambda\tau)/\lambda-\dot{Z}(\theta)\cdot\tau| \\ &\quad\leqq E_\theta|X_\theta(\lambda\tau)/\lambda-\dot{X}(\theta)\cdot\tau|+E_\theta|Y_\theta(\lambda\tau)/\lambda-\dot{Y}(\theta)\cdot\tau|+E_\theta|Y_\theta(\lambda\tau)Z_\theta(\lambda\tau)|/\lambda\to 0 \\ &\qquad\qquad \text{as } \lambda\to 0 , \end{aligned}$$

considering (3.18), (4.5) and (4.8). This concludes the differentiability of Z_θ in mean. Therefore, the proof of the present theorem is complete.

THEOREM 4.3. *If X_θ is differentiable in quadratic mean at θ and if the condition* (2.9) *holds, then Y_θ and Z_θ are also differentiable in quadratic mean at θ with the same derivative as in Theorem* 4.2, *respectively.*

PROOF. 1°. Let $\lambda>0$ and $|\tau|=1$. It follows from (2.8) and by Lemma 2.1 that

$$\lim_{\lambda\to 0} E_\theta|X_\theta(\lambda\tau)|^2/\lambda^2=E_\theta|\dot{X}(\theta)\cdot\tau|^2<\infty , \tag{4.11}$$

and hence that (3.23) and (3.24) hold. The condition (2.9) implies (2.13) and hence the conclusion of Theorem 4.2 holds.
2°. It follows from (4.3), (4.4) and (4.11) that

$$E_\theta|\dot{X}(\theta)\cdot\tau|^2=E_\theta|\dot{Y}(\theta)\cdot\tau|^2+E_\theta|\dot{Z}(\theta)\cdot\tau|^2 . \tag{4.12}$$

By Lemma 2.1 and Theorem 4.2, we have that

$$X_\theta(\lambda\tau)/\lambda \to \dot{X}(\theta)\cdot\tau\ ,$$
$$Y_\theta(\lambda\tau)/\lambda \to \dot{Y}(\theta)\cdot\tau\ , \qquad \text{and} \tag{4.13}$$
$$Z_\theta(\lambda\tau)/\lambda \to \dot{Z}(\theta)\cdot\tau$$

in probability as $\lambda\to 0$, respectively. Thus, by Fatou's Lemma we have that

$$\begin{aligned}&\varliminf_{\lambda\to 0} E_\theta|Y_\theta(\lambda\tau)|^2/\lambda^2 \geqq E_\theta|\dot{Y}(\theta)\cdot\tau|^2\ , \qquad \text{and}\\ &\varliminf_{\lambda\to 0} E_\theta|Z_\theta(\lambda\tau)|^2/\lambda^2 \geqq E_\theta|\dot{Z}(\theta)\cdot\tau|^2\ .\end{aligned} \tag{4.14}$$

In the same way, we have from (iii) of Lemma 3.3 that

$$\begin{aligned}&2\varliminf_{\lambda\to 0} E_\theta\{(Y_\theta(\lambda\tau)+1)(-Z_\theta(\lambda\tau))\}/\lambda^2\\ &\qquad \geqq E_\theta\{\varliminf_{\lambda\to 0}(Y_\theta(\lambda\tau)+1)\, E_\theta[|Z_\theta(\lambda\tau)|^2|T]/\lambda^2\} \geqq E_\theta|\dot{Z}(\theta)\cdot\tau|^2\ .\end{aligned} \tag{4.15}$$

Now, it follows from (3.9) similarly as in (4.10) that

$$E_\theta\{-X_\theta(\lambda\tau)\}/\lambda^2 = E_\theta\{-Y_\theta(\lambda\tau)\}/\lambda^2 + E_\theta\{(Y_\theta(\lambda\tau)+1)(-Z_\theta(\lambda\tau))\}/\lambda^2\ ,$$

which together with (3.24), (4.11), (4.12) and (4.15) implies that

$$\begin{aligned}\varlimsup_{\lambda\to 0} E_\theta|Y_\theta(\lambda\tau)|^2/\lambda^2 &= 2\varlimsup_{\lambda\to 0} E_\theta\{-Y_\theta(\lambda\tau)\}/\lambda^2\\ &= 2\lim_{\lambda\to 0} E_\theta\{-X_\theta(\lambda\tau)\}/\lambda^2 - 2\varliminf_{\lambda\to 0} E_\theta\{(Y_\theta(\lambda\tau)+1)(-Z_\theta(\lambda\tau))\}/\lambda^2\\ &\leqq E_\theta|\dot{Y}(\theta)\cdot\tau|^2\ .\end{aligned} \tag{4.16}$$

Therefore, it follows from (4.14) and (4.16) that

$$\lim_{\lambda\to 0} E_\theta|Y_\theta(\lambda\tau)|^2/\lambda^2 = E_\theta|\dot{Y}(\theta)\cdot\tau|^2 \tag{4.17}$$

and consequently, that

$$2\lim_{\lambda\to 0} E_\theta\{(Y_\theta(\lambda\tau)+1)(-Z_\theta(\lambda\tau))\}/\lambda^2 = E_\theta|\dot{Z}(\theta)\cdot\tau|^2\ . \tag{4.18}$$

We conclude from (4.13), (4.17) and Lemma 2.1 that Y_θ is differentiable in quadratic mean:

$$\lim_{\lambda\to 0} E_\theta|Y_\theta(\lambda\tau)/\lambda - \dot{Y}(\theta)\cdot\tau|^2 = 0\ . \tag{4.19}$$

3°. Since

$$\begin{aligned}&2\, E_\theta\{-X_\theta(\lambda\tau)\}/\lambda^2\\ &\qquad = 2\, E_\theta\{-Y_\theta(\lambda\tau)\}/\lambda^2 + 2\, E_\theta\{-Z_\theta(\lambda\tau)\}/\lambda^2 - 2\, E_\theta\{Y_\theta(\lambda\tau)Z_\theta(\lambda\tau)\}/\lambda^2\ ,\end{aligned}$$

we have from (3.24) and (4.11) that

$$\lim_{\lambda \to 0} E_\theta |Y_\theta(\lambda\tau) - Z_\theta(\lambda\tau)|^2/\lambda^2 = E_\theta |\dot{X}(\theta)\cdot\tau|^2 .$$

Similarly as (4.12), it is easy to see that

$$E_\theta |\dot{Y}(\theta)\cdot\tau - \dot{Z}(\theta)\cdot\tau|^2 = E_\theta |\dot{X}(\theta)\cdot\tau|^2$$

and hence that

$$\lim_{\lambda \to 0} E_\theta |Y_\theta(\lambda\tau) - Z_\theta(\lambda\tau)|^2/\lambda^2 = E_\theta |\dot{Y}(\theta)\cdot\tau - \dot{Z}(\theta)\cdot\tau|^2 . \tag{4.20}$$

From (4.13), we see

$$|Y_\theta(\lambda\tau) - Z_\theta(\lambda\tau)|/\lambda \to |\dot{Y}(\theta)\cdot\tau - \dot{Z}(\theta)\cdot\tau| . \tag{4.21}$$

Thus, by Lemma 2.1 it holds from (4.20) and (4.21) that

$$\lim_{\lambda \to 0} E_\theta |(Y_\theta(\lambda\tau) - Z_\theta(\lambda\tau))/\lambda - (\dot{Y}(\theta)\cdot\tau - \dot{Z}(\theta)\cdot\tau)|^2 = 0 . \tag{4.22}$$

(4.19) and (4.22) lead to the differentiability in quadratic mean of Z_θ:

$$\lim_{\lambda \to 0} E_\theta |Z_\theta(\lambda\tau)/\lambda - \dot{Z}(\theta)\cdot\tau|^2 = 0 .$$

The proof of the theorem is complete.

In the same situation as Theorem 4.3, considering (4.13), we define the Fisher information metrices of $f(X;\theta)$, $g(T;\theta)$ and $h(X;\theta|T)$ by

$$\begin{aligned} I_f(\theta) &= 4E_\theta\{\dot{X}(\theta)\dot{X}(\theta)'\} , \\ I_g(\theta) &= 4E_\theta\{\dot{Y}(\theta)\dot{Y}(\theta)'\} , \qquad \text{and} \\ I_h(\theta) &= 4E_\theta\{\dot{Z}(\theta)\dot{Z}(\theta)'\} , \end{aligned} \tag{4.23}$$

respectively, which are confirmed by the facts of (2.17) and (4.6). Then, from (4.12) we obtain the main result (1.3):

THEOREM 4.4. *The Fisher information matrix of the likelihood function of observation is decomposed into those of the marginal and conditional likelihood functions*:

$$I_f(\theta) = I_g(\theta) + I_h(\theta) .$$

Acknowledgement

The author is grateful to Professor Okamoto, M. for helpful suggestions which improved the presentation of the manuscript. The author also wishes to express his gratitude to the referee for his useful comments and kind suggestions.

OSAKA UNIVERSITY

REFERENCES

[1] Barndorff-Nielsen, O. (1978). *Information and Exponential Families*, John Wiley & Sons, Chichester.

[2] Edwards, A. W. F. (1972). *Likelihood*, Cambridge University Press, Cambridge.

[3] Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

[4] Halmos, P. R. and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistic, *Ann. Math. Statist.*, **20**, 225-241.

[5] LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates, *Ann. Math. Statist.*, **41**, 802-828.

[6] Loève, M. (1978). *Probability Theory* I, II, 4th ed., Springer-Verlag, New York.

[7] Shimizu, R. (1976). *Central Limit Theorems* (in Japanese), Kyoiku Shuppan, Tokyo.