Ann. Inst. Statist. Math.
36 (1984), Part A, 181-197

A GENERALIZATION OF THE RELATIVE CONDITIONAL EXPECTATION—FURTHER PROPERTIES OF PITMAN'S T^* AND THEIR APPLICATIONS TO STATISTICS

HISATAKA KUBOKI

(Received July 4, 1983; revised Jan. 19, 1984)

Summary

This paper is concerned with the mapping T^* which is a generalization of the relative conditional expectation. It has been introduced by E. J. G. Pitman (1979, *Some Basic Theory for Statistical Inference*, Chapman and Hall).

First we extend the definition of the mapping T^* and describe its fundamental properties. Moreover, we establish inequalities for convex functions with respect to T^*.

The mapping T^* is very useful in analysing quantities associated with the distribution of a statistic T. The application of the mapping T^* to statistics is another interest of this paper.

1. Introduction

Let μ be a σ-finite measure on a σ-algebra $\mathcal{F}$ of sets in a space $\mathcal{X}$. T is a mapping from $\mathcal{X}$ into a space $\mathcal{T}$. ν_0 is the measure induced in $\mathcal{T}$ on the σ-algebra $\mathcal{A}$; i.e. $\mathcal{A}$ is the σ-algebra of sets A in $\mathcal{T}$ such that $T^{-1}A \in \mathcal{F}$, and $\nu_0(A)=\mu(T^{-1}A)$. We shall assume that the single point sets of $\mathcal{T}$ are $\mathcal{A}$ measurable.

Notice that ν_0 is not necessarily σ-finite. But there always exists a σ-finite measure ν on $\mathcal{A}$ which dominates ν_0. Indeed μ is dominated by some finite measure μ_1 and the measure induced in $\mathcal{T}$ from μ_1 is finite and dominates ν_0.

Let f be a real-valued measurable function on $\mathcal{X}$ which is integrable. Put

$$Q(A)=\int_{T^{-1}A} f d\mu , \qquad A \in \mathcal{A} .$$

$$\nu(A)=0 \Longrightarrow \mu(T^{-1}A)=0 \Longrightarrow Q(A)=0 .$$

Hence $\nu \gg Q$, and so by the Radon-Nikodym theorem there exists a function g on $\mathcal{T}$, determined up to ν equivalence, such that

$$\int_{T^{-1}A} f d\mu = Q(A) = \int_A g d\nu\ , \qquad A \in \mathcal{A}\ .$$

We shall write $g = T^*f$.

The above notation T^*, which is a mapping of integrable functions on $\mathcal{X}$ into integrable functions on $\mathcal{T}$, was introduced by Pitman [5]. This is a generalization of the concept of the *relative conditional expectation*, for which we refer to Loève [4]. Of course, if μ is a probability measure and ν is the induced probability measure in $\mathcal{T}$, then

$$T^*f = \mathrm{E}\{f|T\}\ .$$

Pitman gave some fundamental properties of T^* in [5], pp. 100–102.

The definition of T^* can be slightly extended as follows. Let μ be an arbitrary measure on $\mathcal{F}$. We assume that the induced measure ν_0 is dominated by some σ-finite measure ν. Let $\mathcal{E}(\mathcal{X}, \mu)$ (resp. $\mathcal{E}(\mathcal{T}, \nu)$) be a family of all extended real-valued measurable functions on $\mathcal{X}$ (resp. $\mathcal{T}$) whose integrals exist, that is,

$$\int f^+ d\mu < \infty \quad \text{or} \quad \int f^- d\mu < \infty\ , \qquad \text{if } f \in \mathcal{E}(\mathcal{X}, \mu)$$

$$\left(\text{resp. } \int g^+ d\nu < \infty \quad \text{or} \quad \int g^- d\nu < \infty\ , \qquad \text{if } g \in \mathcal{E}(\mathcal{T}, \nu)\right)\ .$$

DEFINITION 1.1. The mapping T^* of $\mathcal{E}(\mathcal{X}, \mu)$ into $\mathcal{E}(\mathcal{T}, \nu)$ is defined by the following formula: for each $f \in \mathcal{E}(\mathcal{X}, \mu)$,

$$\int_A (T^*f) d\nu = \int_{T^{-1}A} f d\mu \qquad \text{for every } A \in \mathcal{A}\ ,$$

where T^*f is determined up to ν equivalence.

The definition is justified by the extended Radon-Nikodym theorem (e.g. Loève [3]).

This paper is concerned with various properties of the mapping T^* and their applications to statistics. We can easily see that T^* has many properties similar to those of the conditional expectation. But in general, $T^*1 \neq 1$. For this reason the following Jensen's inequality is not necessarily justified:

$$T^*\Psi(f) \geqq \Psi(T^*f)\ ,$$

where Ψ is a convex function on $(-\infty, \infty)$ and f is an integrable function. In the following section we shall establish other inequalities for convex functions. In Section 3, we demonstrate some applications of the mapping T^*. In particular, let f be a probability density function and let T be a statistic. Then the induced probability measure Q is the distribution of T and the image of f, $g=T^*f$ is a density of Q relative to a σ-finite measure ν. In this case, the mapping T^* is useful in analysing various quantities associated with the distribution of T.

2. Properties of the mapping T^*

First we describe some fundamental properties of T^*. To avoid constant repitations, it will be assumed that the functions which figure under the T^* sign belong to $\mathcal{E}(\mathcal{X}, \mu)$.

(i) If $c_1\int f_1 d\mu + c_2\int f_2 d\mu$ exists for constants c_1, c_2, then

$$T^*(c_1f_1+c_2f_2)=c_1T^*f_1+c_2T^*f_2\,, \qquad \text{a.e. } \nu\,.$$

(ii) $f\geqq 0$, a.e. $\mu \Longrightarrow T^*f\geqq 0$, a.e. ν;
$f\geqq 0$, a.e. μ and $T^*f=0$, a.e. $\nu \Longrightarrow f=0$, a.e. μ.

(iii) $f_1\geqq f_2$, a.e. $\mu \Longrightarrow T^*f_1\geqq T^*f_2$, a.e. ν.

(iv) If h is a measurable function on $\mathcal{T}$, then

$$T^*[h(T)\cdot f]=h\cdot T^*f\,, \qquad \text{a.e. } \nu\,.$$

(v) $T^*1=j$, a.e. ν, where $j=d\nu_0/d\nu$; hence from (iv)

$$T^*[h(T)]=h\cdot j\,, \qquad \text{a.e. } \nu\,.$$

These properties follows at once from the definition of T^* and properties of integrals.

Let $\Phi(u)$ and $\Psi(u)$ denote arbitrary convex functions defined on $(0, +\infty)$ and $(-\infty, +\infty)$, respectively. For Φ, we assume the following notational conventions which are due to Csiszár [1]:

$$
\begin{aligned}
&\Phi(0)=\lim_{u\to+0}\Phi(u) \qquad (=\Phi_0, \text{ say});\\
&0\cdot\Phi(0/0)=0\,; \qquad\qquad (2.1)\\
&0\cdot\Phi(a/0)=\lim_{\varepsilon\to+0}\varepsilon\cdot\Phi(a/\varepsilon)\\
&\qquad\qquad =a\lim_{u\to+\infty}\Phi(u)/u \qquad (=a\cdot\Phi_\infty, \text{ say}),\ 0<a<+\infty\,.
\end{aligned}
$$

Then we establish the following inequalities for convex functions.

THEOREM 2.1. *Let f and g be nonnegative and integrable functions on $\mathcal{X}$. Then*

$$T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\}\geqq T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right)\,, \qquad a.e.\ \nu\,. \tag{2.2}$$

If Φ is strictly convex, then the equality holds if and only if

$$f\cdot T^*g(T)=g\cdot T^*f(T)\ ,\qquad a.e.\ \mu\ .$$

THEOREM 2.2. *Let f be an integrable function on $\mathscr{X}$, and let g be a positive integrable function on $\mathscr{X}$. Then*

$$T^*\left\{g\cdot\Psi\left(\frac{f}{g}\right)\right\}\geqq T^*g\cdot\Psi\left(\frac{T^*f}{T^*g}\right)\ ,\qquad a.e.\ \nu\ . \tag{2.3}$$

If Ψ is strictly convex, then the equality holds if and only if

$$f\cdot T^*g(T)=g\cdot T^*f(T)\ ,\qquad a.e.\ \mu\ .$$

Remark 2.1. Notice that $T^*\{g\cdot\Phi(f/g)\}$ is always well-defined. Indeed, using conventions (2.1) and the convexity of Φ,

$$g\cdot\Phi(f/g)\geqq c_1f+c_2g\ ,$$

for some constants c_1 and c_2, and which implies that

$$\int g\cdot\Phi(f/g)d\mu>-\infty\ ,$$

since f and g are integrable. Similarly, $T^*\{g\cdot\Psi(f/g)\}$ is also well-defined. Moreover notice that inequalities (2.2) and (2.3) are not reduced to Jensen's inequalities even if we put $g=1$. Because, in general, $g=1$ is not integrable and $T^*1=j\neq 1$.

Now we prove Theorem 2.1 only, since we can similarly prove Theorem 2.2.

PROOF OF THEOREM 2.1. We hereafter denote by $\chi_A=\chi_A(\cdot)$ and $\bar{A}$ the indicator function and the complement of a set A, respectively. Put

$$F=\{t;\ T^*f(t)=0\}\ ,\qquad G=\{t;\ T^*g(t)=0\}$$

and

$$N=\{t;\ T^*f(t)=\infty\}\cup\{t;\ T^*g(t)=\infty\}\ .$$

Notice that $\nu(N)=0$ since f and g are integrable. From (ii),

$$f=0\ ,\qquad \text{a.e. } \mu \text{ on } T^{-1}F$$

and

$$g=0\ ,\qquad \text{a.e. } \mu \text{ on } T^{-1}G\ .$$

Thus

$$g\cdot\Phi\left(\frac{f}{g}\right)=\Phi_0\cdot g\,, \qquad \text{a.e. } \mu \text{ on } T^{-1}F$$

and

$$g\cdot\Phi\left(\frac{f}{g}\right)=\Phi_\infty\cdot g\,, \qquad \text{a.e. } \mu \text{ on } T^{-1}G\,.$$

From (iv)

$$T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\}=\Phi_0\cdot T^*g=T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right)\,, \qquad \text{a.e. } \nu \text{ on } F$$

and

$$T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\}=\Phi_\infty\cdot T^*f=T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right)\,, \qquad \text{a.e. } \nu \text{ on } G\,.$$

Thus combining these two relations, we have

$$\chi_{F\cup G}\cdot T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\}=\chi_{F\cup G}\cdot T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right)\,, \qquad \text{a.e. } \nu\,. \tag{2.4}$$

Now denote the right-hand derivative of Φ at u by $\Phi'_+(u)$, which is nondecreasing in u and thus is measurable. Then for $0<u_0<+\infty$,

$$\Phi(u)\geqq\Phi(u_0)+\Phi'_+(u_0)(u-u_0)\,, \qquad u\geqq 0$$

and

$$\Phi_\infty\geqq\Phi'_+(u_0)\,.$$

From these relations and conventions (2.1), we have

$$g\cdot\Phi\left(\frac{f}{g}\right)\geqq g\cdot\Phi\left(\frac{T^*f(T)}{T^*g(T)}\right)+\Phi'_+\left(\frac{T^*f(T)}{T^*g(T)}\right)\left(f-g\frac{T^*f(T)}{T^*g(T)}\right)\,, \quad \text{on } T^{-1}(\overline{F\cup G\cup N})\,, \tag{2.5}$$

by putting $u=f/g$ and $u_0=T^*f(T)/T^*g(T)$. Define

$$A_n=\left\{t\,;\ \frac{1}{n}\leqq\frac{T^*f(t)}{T^*g(t)}\leqq n\right\}\,, \qquad n=1, 2,\cdots\,.$$

Notice that

$$\overline{F\cup G\cup N}=\bigcup_{n=1}^{\infty} A_n\,. \tag{2.6}$$

Since each term of the right-hand side of (2.5) is integrable on $T^{-1}A_n$, from (i), (iii) and (iv) we have

$$T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\} \geqq T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right)+\Phi'_+\left(\frac{T^*f}{T^*g}\right)\left(T^*f-T^*g\frac{T^*f}{T^*g}\right)$$
$$=T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right), \qquad \text{a.e. } \nu \text{ on } A_n,$$

that is, for each $n=1, 2, \cdots$,

$$\chi_{A_n}\cdot T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\} \geqq \chi_{A_n}\cdot T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right), \qquad \text{a.e. } \nu .$$

From (2.6) we have

$$\chi_{\overline{F\cup G\cup N}}\cdot T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\} \geqq \chi_{\overline{F\cup G\cup N}}\cdot T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right), \qquad \text{a.e. } \nu .$$

Hence combining this with (2.4) and noting that $\nu(N)=0$, we obtain

$$T^*\left\{g\cdot\Phi\left(\frac{f}{g}\right)\right\} \geqq T^*g\cdot\Phi\left(\frac{T^*f}{T^*g}\right), \qquad \text{a.e. } \nu .$$

From (2.5) and (ii), the equality holds if and only if for each $n=1, 2, \cdots$,

$$g\cdot\Phi\left(\frac{f}{g}\right)=g\cdot\Phi\left(\frac{T^*f(T)}{T^*g(T)}\right)+\Phi'_+\left(\frac{T^*f(T)}{T^*g(T)}\right)\left(f-g\frac{T^*f(T)}{T^*g(T)}\right),$$
$$\text{a.e. } \mu \text{ on } A_n .$$

Since Φ is strictly convex, this implies that for each $n=1, 2, \cdots$,

$$\chi_{A_n}(T)\left(f-g\frac{T^*f(T)}{T^*g(T)}\right)=0, \qquad \text{a.e. } \mu .$$

Thus

$$\chi_{\overline{F\cup G\cup N}}(T)\left(f-g\frac{T^*f(T)}{T^*g(T)}\right)=0, \qquad \text{a.e. } \mu .$$

Hence the equality holds if and only if

$$f\cdot T^*g(T)=g\cdot T^*f(T), \qquad \text{a.e. } \mu .$$

Thus we complete the proof.

As stated in Remark 2.1, inequalities (2.2) and (2.3) are not generally reduced to Jensen's inequalities. But we have the following result.

COROLLARY 2.1 (Jensen's inequality). *Under the same notations as in Theorem* 2.1 *and Theorem* 2.2, *suppose that either* 1° *or* 2° *is satisfied*:
1° μ *is a finite measure, and then take* $\nu=\nu_0$.

2° ν_0 *is a σ-finite measure, and then take* $\nu=\nu_0$. *Moreover,* $\Phi(f)\in \mathcal{E}(\mathcal{X},\mu)$ *and* $\Psi(f)\in\mathcal{E}(\mathcal{X},\mu)$, *respectively.*

Then

$$T^*\Phi(f)\geqq\Phi(T^*f)\,, \qquad a.e.\ \nu$$

and

$$T^*\Psi(f)\geqq\Psi(T^*f)\,, \qquad a.e.\ \nu\,,$$

respectively.

PROOF. Notice that since $\nu=\nu_0$, $T^*1=1$, a.e. ν. Thus for every $h\in\mathcal{E}(\mathcal{T},\nu)$, $T^*h(T)=h$, a.e. ν. Then the above inequalities are proved by the same argument as in the proof of Theorem 2.1.

Using the above results, we obtain some useful consequences.

Example 2.1. Let $f\in L^k(\mathcal{X})$ and $g\in L^{k'}(\mathcal{X})$, where $k>1$ and $1/k+1/k'=1$. Applying Theorem 2.1 to a function $\Phi(u)=-u^{1/k}$, $u>0$, we have

$$T^*\left\{|g|^{k'}\left(\frac{|f|^k}{|g|^{k'}}\right)^{1/k}\right\}\leqq T^*|g|^{k'}\left(\frac{T^*|f|^k}{T^*|g|^{k'}}\right)^{1/k}, \qquad \text{a.e. } \nu\,.$$

According to conventions (2.1), this yields to the Hölder inequality with respect to T^*, that is,

$$T^*|f\cdot g|\leqq(T^*|f|^k)^{1/k}(T^*|g|^{k'})^{1/k'}\,, \qquad \text{a.e. } \nu\,,$$

with equality if and only if

$$|f|^k\cdot T^*|g|^{k'}=|g|^{k'}\cdot T^*|f|^k\,, \qquad \text{a.e. } \mu\,.$$

Using this result, we can easily prove the Minkowski inequality with respect to T^*: for $f,g\in L^k(\mathcal{X})$, $k\geqq1$,

$$(T^*|f+g|^k)^{1/k}\leqq(T^*|f|^k)^{1/k}+(T^*|g|^k)^{1/k}\,, \qquad \text{a.e. } \nu\,.$$

Example 2.2. Let $f,f_1,f_2,\cdots\in L^1(\mathcal{X})$. From (ii) and (iii),

$$T^*|f_n-f|\geqq|T^*f_n-T^*f|\,, \qquad \text{a.e. } \nu\,.$$

Thus

$$f_n\to f \text{ in mean}\Longrightarrow T^*f_n\to T^*f \text{ in mean}\,.$$

When $f,f_1,f_2,\cdots\in L^k(\mathcal{X})$, $k\geqq2$, the corresponding result is not generally justified. But if ν_0 is σ-finite and we take $\nu=\nu_0$, then

$$f_n\to f \text{ in the } k\text{th mean}\Longrightarrow T^*f_n\to T^*f \text{ in the } k\text{th mean}\,.$$

In fact, from Corollary 2.1

$$T^*|f_n - f|^k \geqq |T^*f_n - T^*f|^k , \qquad \text{a.e. } \nu .$$

3. Applications of the mapping T^* to statistics

Consider a set $\{P_\theta ; \theta \in \Theta\}$ of probability measures on $\mathcal{X}$ with densities $f_\theta(x)$ relative to a σ-finite measure μ. Let T be a mapping from $\mathcal{X}$ into $\mathcal{T}$, and let $\{Q_\theta ; \theta \in \Theta\}$ be the set of distributions of T, that is, induced probability measures in $\mathcal{T}$. Then each Q_θ has the density $g_\theta(t) = T^*f_\theta(t)$ relative to a σ-finite measure ν.

3.1. *Properties of the divergence measure of two distributions*

Let $\Phi(u)$ be a convex function defined on $(0, +\infty)$. Consider the following quantities:

$$I_\Phi(P_\tau, P_\theta)|E = \int_E f_\theta \cdot \Phi\left(\frac{f_\tau}{f_\theta}\right) d\mu , \qquad E \in \mathcal{F}$$

and

$$I_\Phi(Q_\tau, Q_\theta)|A = \int_A g_\theta \cdot \Phi\left(\frac{g_\tau}{g_\theta}\right) d\nu , \qquad A \in \mathcal{A} ,$$

under conventions (2.1). $I_\Phi(P_\tau, P_\theta)|\mathcal{X}$ is called the *Φ-divergence* of the distributions P_τ and P_θ (Csiszár [1]). We now prove some important properties of the divergence measure by our method.

THEOREM 3.1 (Csiszár [1]). 1° *For every $E \in \mathcal{F}$,*

$$I_\Phi(P_\tau, P_\theta)|E \geqq P_\theta(E) \cdot \Phi\left(\frac{P_\tau(E)}{P_\theta(E)}\right) .$$

If $P_\theta(E) > 0$ and $\Phi(u)$ is strictly convex at $u_0 = P_\tau(E)/P_\theta(E)$, then the strict inequality holds except for the case

$$f_\tau(x) = u_0 \cdot f_\theta(x) , \qquad \text{a.e. } \mu \text{ on } E .$$

2° $$I_\Phi(P_\tau, P_\theta)|\mathcal{X} \geqq I_\Phi(Q_\tau, Q_\theta)|\mathcal{T} .$$

If $\Phi(u)$ is strictly convex and $I_\Phi(Q_\tau, Q_\theta)|\mathcal{T}$ is finite, then the equality holds if and only if T is sufficient.

PROOF. Consider a mapping $T(x) = \chi_E(x)$, for $E \in \mathcal{F}$. Let ν be a measure on $\mathcal{T} = \{0, 1\}$ such that $0 < \nu(0) < \infty$ and $0 < \nu(1) < \infty$. It is easily seen that

$$T^*f_\theta(t) = \frac{1}{\nu(t)} \int_{T^{-1}\{t\}} f_\theta d\mu = \frac{1}{\nu(t)} P_\theta(T^{-1}\{t\}) ,$$

for each $t=0, 1$. Thus from Theorem 2.1, we have

$$I_{\Phi}(P_{\tau}, P_{\theta})|E=\nu(1)\cdot T^{*}\left\{f_{\theta}\cdot\Phi\left(\frac{f_{\tau}}{f_{\theta}}\right)\right\}(1)\geqq\nu(1)\cdot T^{*}f_{\theta}(1)\cdot\Phi\left(\frac{T^{*}f_{\tau}(1)}{T^{*}f_{\theta}(1)}\right)$$
$$=P_{\theta}(E)\cdot\Phi\left(\frac{P_{\tau}(E)}{P_{\theta}(E)}\right).$$

This proves the first part of 1°. The last part of 1° is obvious. 2° is the direct consequence of Theorem 2.1.

Remark 3.1. Let μ be an arbitrary measure on $\mathcal{X}$, and let α and β be two nonnegative measurable function on $\mathcal{X}$ and integrable on E. Then by the same argument as in the above proof, we have

$$\int_{E}\beta\cdot\Phi\left(\frac{\alpha}{\beta}\right)d\mu\geqq\int_{E}\beta d\mu\cdot\Phi\left(\frac{\int_{E}\alpha d\mu}{\int_{E}\beta d\mu}\right).$$

3.2 *Smoothness of a family of probability distributions*

For simplicity, we here assume that Θ is an open subset of $\boldsymbol{R}^{1}$. Moreover to simplify notation we shall write f for f_{θ}, and f_{0} for f_{θ_0} where convenient. The following definition of smoothness of the family $\{f_{\theta};\theta\in\Theta\}$ is due to Pitman [5]. In what follows, we refer to pp. 11–28 of [5] by Pitman [5].

DEFINITION 3.1. The family $\{f_{\theta};\theta\in\Theta\}$ is called *smooth* at θ_{0} if it satisfies the following conditions:

(1) f is differentiable in mean at θ_{0}, i.e. there exists an integrable f_{0}' such that

$$\lim_{\theta\to\theta_0}\int\left|\frac{f-f_{0}}{\theta-\theta_{0}}-f_{0}'\right|d\mu=0;$$

(2) $$\lim_{\theta\to\theta_0}\int\left|\left(\frac{\sqrt{f}-\sqrt{f_{0}}}{\theta-\theta_{0}}\right)^{2}-\phi_{0}\right|d\mu=0,$$

where

$$\phi_{0}(x)=\begin{cases}\dfrac{f_{0}'(x)^{2}}{4f_{0}(x)}, & \text{on } \{x;f_{0}(x)>0\}\\ 0, & \text{on } \{x;f_{0}(x)=0\},\end{cases}$$

and

$$\int\phi_{0}d\mu<\infty.$$

As pointed out in [5], smoothness of the family at θ_0 is exactly equivalent to the differentiability in mean square of $\sqrt{f}$ at θ_0, i.e.

$$\lim_{\theta\to\theta_0}\int\left(\frac{\sqrt{f}-\sqrt{f_0}}{\theta-\theta_0}-\sqrt{\phi_0}\right)^2 d\mu=0 .$$

When the family $\{f_\theta;\theta\in\Theta\}$ is smooth at θ_0, so is the family of distributions of T. More precisely,

THEOREM 3.2 (Pitman [5]). *If the family* $\{f_\theta;\theta\in\Theta\}$ *is smooth at* θ_0, *so is the family* $\{g_\theta;\theta\in\Theta\}$, *i.e. it satisfies the following conditions:*

(1)
$$\lim_{\theta\to\theta_0}\int\left|\frac{g-g_0}{\theta-\theta_0}-g_0'\right|d\nu=0 ,$$

where

$$g_0'=T^*f_0' , \qquad a.e.\ \nu ;$$

(2)
$$\lim_{\theta\to\theta_0}\int\left|\left(\frac{\sqrt{g}-\sqrt{g_0}}{\theta-\theta_0}\right)^2-\psi_0\right|d\nu=0 ,$$

where

$$\psi_0(t)=\begin{cases}\dfrac{g_0'(t)^2}{4g_0(t)}, & on\ \{t;g_0(t)>0\}\\ 0, & on\ \{t;g_0(t)=0\} ,\end{cases}$$

and

$$\int\psi_0 d\nu\leqq\int\phi_0 d\mu .$$

For the proof of this theorem, we essentially use the Schwarz inequality and the convergence in mean with respect to T^* (see Examples 2.1 and 2.2).

Here we shall establish a further result about smoothness. Define

$$k_\theta(x)=\begin{cases}f_\theta(x)/g_\theta(Tx) , & \text{on } \{x;g_\theta(Tx)>0\}\\ h_\theta(x) , & \text{on } \{x;g_\theta(Tx)=0\} ,\end{cases}$$

where $h_\theta(x)$ is a properly chosen nonnegative measurable function. Using (ii), we can easily check that $f_\theta(x)$ is factorable in the form

$$f_\theta(x)=g_\theta(Tx)\cdot k_\theta(x) , \qquad \text{a.e. } \mu .$$

$k_\theta(x)$ is regarded as the density of the conditional distribution given $t=Tx$. Notice that $k_\theta\in\mathcal{E}(\mathcal{X},\mu)$ and thus T^*k_θ is well-defined. According to (iv),

$$T^*k_\theta(t)=1\ , \qquad \text{a.e. } \nu \text{ on } \{t;\, g_\theta(t)>0\}\ .$$

In the following discussion, we assume that it is possible to choose $h_\theta(x)$ so that

$$(3.1) \qquad T^*k_\theta(t)=1\ , \qquad \text{a.e. } \nu \text{ on } \{t;\, g_{\theta_0}(t)>0\}, \quad \text{for every } \theta \text{ in some neighbourhood of } \theta_0\,.$$

Of course, if the following condition is satisfied, then (3.1) holds:

$$(3.2) \qquad g_\theta(t)>0\ , \qquad \text{a.e. } \nu \text{ on } \{t;\, g_{\theta_0}(t)>0\}, \quad \text{for every } \theta \text{ in some neighbourhood of } \theta_0\,.$$

In particular when both ν_0 and ν are σ-finite, we can always determine k_θ so that it may satisfy (3.1). Since ν_0 is σ-finite, there is a function $T_0^*f_\theta$, determined up to ν_0 equivalence, such that

$$\int_{T^{-1}A} f_\theta d\mu=\int_A T_0^*f_\theta d\nu_0=\int_A j\cdot T_0^*f_\theta d\nu\ , \qquad A \in \mathcal{A}\,.$$

Thus

$$g_\theta=j\cdot T_0^*f_\theta\ , \qquad \text{a.e. } \nu\,.$$

Notice that $0\leqq j<\infty$, a.e. ν, since both ν_0 and ν are σ-finite. Hence

$$(3.3) \qquad 0<j(t)<\infty\ , \qquad \text{a.e. } \nu \text{ on } \{t;\, g_{\theta_0}(t)>0\}\ .$$

Define

$$k_\theta(x)=\begin{cases} f_\theta(x)/g_\theta(Tx)\ , & \text{on } \{x;\, g_\theta(Tx)>0\} \\ 1/j(Tx)\ , & \text{on } \{x;\, g_\theta(Tx)=0\}\ . \end{cases}$$

It follows from (iv), (v) and (3.3) that

$$T^*k_\theta(t)=1\ , \qquad \text{a.e. } \nu \text{ on } \{t;\, g_{\theta_0}(t)>0\}, \text{ for every } \theta \in \Theta\,.$$

THEOREM 3.3. *Suppose that the family* $\{f_\theta;\, \theta \in \Theta\}$ *is smooth at* θ_0. *If the family* $\{k_\theta;\, \theta \in \Theta\}$ *satisfies condition* (3.1), *then it is also smooth at* θ_0 *in the following sense*:

$$(1) \qquad \lim_{\theta\to\theta_0}\int g_0(T)\left|\frac{k-k_0}{\theta-\theta_0}-k_0'\right|d\mu=0\ ,$$

where

$$k_0'(x)=\begin{cases} \dfrac{f_0'(x)}{g_0(Tx)}-\dfrac{f_0(x)g_0'(Tx)}{g_0(Tx)^2}\ , & \textit{on } \{x;\, g_0(Tx)>0\} \\ 0\ ,\ \textit{say} & \textit{on } \{x;\, g_0(Tx)=0\}\,; \end{cases}$$

(2) $$\lim_{\theta\to\theta_0}\int g_0(T)\left|\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2-\kappa_0\right|d\mu=0\ ,$$

where

$$\kappa_0(x)=\begin{cases}\dfrac{k_0'(x)^2}{4k_0(x)}\ , & on\ \{x; f_0(x)>0\}\\ 0\ , & on\ \{x; f_0(x)=0\}\ .\end{cases}$$

Moreover

(3.4) $$\int\phi_0 d\mu=\int\psi_0 d\nu+\int\kappa_0\cdot g_0(T)d\mu\ ,$$

that is,

$$\mathrm{E}_{\theta_0}\left(\frac{f_0'}{f_0}\right)^2=\mathrm{E}_{\theta_0}\left(\frac{g_0'}{g_0}\right)^2+\mathrm{E}_{\theta_0}\left(\frac{k_0'}{k_0}\right)^2 .$$

Remark 3.2. Let E be an event of positive probability at θ_0, i.e.

$$P_0(E)=\int_E f_0 d\mu>0\ .$$

Consider a mapping $T(x)=\chi_E(x)$, and let ν be a measure on $\mathcal{T}=\{0, 1\}$ such that $\nu(0)=\nu(1)=1$. As stated in the proof of Theorem 3.1,

$$g_\theta(t)=T^*f_\theta(t)=P_\theta(T^{-1}\{t\})\ ,\qquad t=0, 1\ .$$

If the family $\{f_\theta; \theta\in\Theta\}$ is smooth at θ_0, then $P(E)\to P_0(E)$ as $\theta\to\theta_0$, since $f\to f_0$ in mean as $\theta\to\theta_0$. Thus $\{g_\theta; \theta\in\Theta\}$ satisfies condition (3.2). Consider

$$k_\theta(x)=\begin{cases}f_\theta(x)/P_\theta(E)\ , & x\in E \text{ and } P_\theta(E)>0\\ h_\theta(x)\ , & x\in E \text{ and } P_\theta(E)=0\ ,\end{cases}$$

which is the density of the conditional distribution given E at θ. From Theorem 3.3, we can easily verify that the family $\{k_\theta; \theta\in\Theta\}$ is smooth at θ_0 if the family $\{f_\theta; \theta\in\Theta\}$ is smooth at θ_0. This fact has been established by Pitman [5].

Now we shall prove Theorem 3.3. For this, we need the concept of *loose convergence* and the extended form of the dominated convergence theorem, for which we refer to Pitman [5], pp. 98–100.

DEFINITION 3.2. We shall say that g_n *converges loosely* to g, and write $g_n\xrightarrow{l}g$, if every subsequence of $\{g_n\}$ containes a subsequence which converges almost everywhere to g.

Notice that $g_n \to g$ in mean, or in measure, implies $g_n \xrightarrow{l} g$.

LEMMA 3.1. *The following extension of the dominated convergence theorem holds:*

1° $g_n \xrightarrow{l} g$, $|g_n| \leqq H_n$ *a.e.*, H_n *integrable and* $\xrightarrow{l} H$ *integrable*, $\int H_n \to \int H \Longrightarrow$ $g_n \to g$ *in mean.*

As consequences of 1°,

2° $g_n \xrightarrow{l} g$, $|g_n| \leqq |G_n|$ *a.e.*, G_n *integrable and* $\to G$ *in mean* $\Longrightarrow g_n \to g$ *in mean*;

3° $H_n \geqq 0$ *and integrable*, $H_n \xrightarrow{l} H$ *integrable*, $\int H_n \to \int H \Longrightarrow H_n \to H$ *in mean.*

PROOF OF THEOREM 3.3. Put

$$F_0 = \{x; f_0(x) > 0\} \qquad \text{and} \qquad G_0 = \{t; g_0(t) > 0\}.$$

Using (ii), we can easily verify that $F_0 \subset T^{-1}G_0 \cup N$, where N is a μ null set. Since the family $\{f_\theta; \theta \in \Theta\}$ is smooth at θ_0,

$$(3.5) \qquad \begin{aligned} &f \to f_0 \text{ in mean} &&\text{and} && \frac{f-f_0}{\theta-\theta_0} \to f_0' \text{ in mean},\\ &g \to g_0 \text{ in mean} &&\text{and} && \frac{g-g_0}{\theta-\theta_0} \to g_0' \text{ in mean}. \end{aligned}$$

These imply that

$$(3.6) \qquad \begin{aligned} &f \xrightarrow{l} f_0 &&\text{and} && \frac{f-f_0}{\theta-\theta_0} \xrightarrow{l} f_0',\\ &g(T) \xrightarrow{l} g_0(T) &&\text{and} && \frac{g(T)-g_0(T)}{\theta-\theta_0} \xrightarrow{l} g_0'(T), \end{aligned}$$

because $\nu_0 \ll \nu$. Thus

$$\begin{aligned} \frac{k-k_0}{\theta-\theta_0} &= \left(\frac{f-f_0}{\theta-\theta_0}\right)\frac{1}{g(T)} - \left(\frac{g(T)-g_0(T)}{\theta-\theta_0}\right)\frac{f_0}{g(T)g_0(T)}\\ &\xrightarrow{l} \frac{f_0'}{g_0(T)} - \frac{g_0'(T)f_0}{g_0(T)^2}, \qquad \text{on } T^{-1}G_0 \text{ as } \theta \to \theta_0. \end{aligned}$$

Hence

$$(3.7) \qquad g_0(T)\frac{k-k_0}{\theta-\theta_0} \xrightarrow{l} g_0(T)k_0', \qquad \text{on } T^{-1}G_0 \text{ as } \theta \to \theta_0.$$

On the other hand,

$$g_0(T)\left|\frac{k-k_0}{\theta-\theta_0}\right| \leqq g_0(T)\left|\frac{k-f/g_0(T)}{\theta-\theta_0}\right| + \left|\frac{f-f_0}{\theta-\theta_0}\right|$$

$$=k\left|\frac{g(T)-g_0(T)}{\theta-\theta_0}\right|+\left|\frac{f-f_0}{\theta-\theta_0}\right|, \qquad \text{on } T^{-1}G_0,$$

and from (iv), (3.1) and (3.5),

$$\int_{T^{-1}G_0} k\left|\frac{g(T)-g_0(T)}{\theta-\theta_0}\right|d\mu=\int_{G_0}\left|\frac{g-g_0}{\theta-\theta_0}\right|d\nu$$

$$\to\int_{G_0}|g_0'|d\nu=\int_{T^{-1}G_0} k_0|g_0'(T)|d\mu, \qquad \text{as } \theta\to\theta_0,$$

$$\int_{T^{-1}G_0}\left|\frac{f-f_0}{\theta-\theta_0}\right|d\mu\to\int_{T^{-1}G_0}|f_0'|d\mu, \qquad \text{as } \theta\to\theta_0.$$

Thus it follows from (3.6), (3.7) and Lemma 3.1, 1° that

$$g_0(T)\frac{k-k_0}{\theta-\theta_0}\to g_0(T)k_0' \text{ in mean on } T^{-1}G_0, \qquad \text{as } \theta\to\theta_0.$$

Therefore

$$\lim_{\theta\to\theta_0}\int g_0(T)\left|\frac{k-k_0}{\theta-\theta_0}-k_0'\right|d\mu=0.$$

Next from (3.6) and (3.7) we have

$$\text{(3.8)} \qquad g_0(T)\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2=g_0(T)\left(\frac{k-k_0}{\theta-\theta_0}\right)^2\frac{1}{(\sqrt{k}+\sqrt{k_0})^2}$$

$$\xrightarrow{l} g_0(T)\frac{k_0'^2}{4k_0}, \qquad \text{on } F_0 \text{ as } \theta\to\theta_0.$$

Moreover

$$\text{(3.9)} \qquad g_0(T)\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2$$

$$\leqq 2g_0(T)\left(\frac{\sqrt{k}-\sqrt{f/g_0(T)}}{\theta-\theta_0}\right)^2+2g_0(T)\left(\frac{\sqrt{f/g_0(T)}-\sqrt{f_0/g_0(T)}}{\theta-\theta_0}\right)^2$$

$$=2k\left(\frac{\sqrt{g(T)}-\sqrt{g_0(T)}}{\theta-\theta_0}\right)^2+2\left(\frac{\sqrt{f}-\sqrt{f_0}}{\theta-\theta_0}\right)^2, \qquad \text{on } T^{-1}G_0.$$

Since

$$\text{(3.10)} \qquad \left(\frac{\sqrt{f}-\sqrt{f_0}}{\theta-\theta_0}\right)^2\to\phi_0 \text{ in mean} \quad \text{and} \quad \left(\frac{\sqrt{g}-\sqrt{g_0}}{\theta-\theta_0}\right)^2\to\psi_0 \text{ in mean},$$

we can easily see that

$$\text{(3.11)} \qquad \left(\frac{\sqrt{f}-\sqrt{f_0}}{\theta-\theta_0}\right)^2\xrightarrow{l}\phi_0 \quad \text{and} \quad k\left(\frac{\sqrt{g(T)}-\sqrt{g_0(T)}}{\theta-\theta_0}\right)^2\xrightarrow{l}k_0\cdot\psi_0(T),$$

$$\text{on } T^{-1}G_0.$$

Recalling definitions of ϕ_0 and k_0, and $F_0 \subset T^{-1}G_0 \cup N$, it follows from (3.8), (3.9) and (3.11) that

$$(3.12) \qquad g_0(T)\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2 \xrightarrow{l} g_0(T)\kappa_0 , \qquad \text{on } T^{-1}G_0 .$$

Combining this with (3.9)-(3.11), and applying Lemma 3.1, 1°, we have

$$g_0(T)\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2 \to g_0(T)\kappa_0 \text{ in mean on } T^{-1}G_0 , \qquad \text{as } \theta\to\theta_0 ,$$

because using (iv) and (3.1)

$$\int_{T^{-1}G_0} k\left(\frac{\sqrt{g(T)}-\sqrt{g_0(T)}}{\theta-\theta_0}\right)^2 d\mu = \int_{G_0}\left(\frac{\sqrt{g}-\sqrt{g_0}}{\theta-\theta_0}\right)^2 d\nu$$

$$\to \int_{G_0} \psi_0 d\nu = \int_{T^{-1}G_0} k_0\cdot\psi_0(T) d\mu , \qquad \text{as } \theta\to\theta_0 ,$$

$$\int_{T^{-1}G_0}\left(\frac{\sqrt{f}-\sqrt{f_0}}{\theta-\theta_0}\right)^2 d\mu \to \int_{T^{-1}G_0} \phi_0 d\mu , \qquad \text{as } \theta\to\theta_0 .$$

Therefore

$$\lim_{\theta\to\theta_0} \int g_0(T)\left|\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2 - \kappa_0\right| d\mu = 0 .$$

The above argument proves the first part of the theorem.

From differentiability in mean at θ_0 of f, we can easily check that

$$f_0' = 0 , \qquad \text{a.e. } \mu \text{ on } \bar{F}_0 .$$

Consequently, noting that $F_0 \subset T^{-1}G_0 \cup N$, we have

$$\begin{aligned} 4\int \kappa_0\cdot g_0(T) d\mu &= \int_{F_0} \frac{k_0'^2}{k_0} g_0(T) d\mu \\ &= \int_{F_0} \frac{f_0'^2}{f_0} d\mu - 2\int_{F_0} \frac{f_0'\cdot g_0'(T)}{g_0(T)} d\mu + \int_{F_0} \frac{f_0\cdot g_0'(T)^2}{g_0(T)^2} d\mu \\ &= 4\int \phi_0 d\mu - 2\int_{T^{-1}G_0} \frac{f_0'\cdot g_0'(T)}{g_0(T)} d\mu + \int_{T^{-1}G_0} \frac{f_0\cdot g_0'(T)^2}{g_0(T)^2} d\mu \\ &= 4\int \phi_0 d\mu - 2\int_{G_0} \frac{g_0'^2}{g_0} d\nu + \int_{G_0} \frac{g_0'^2}{g_0} d\nu \\ &= 4\int \phi_0 d\mu - 4\int \psi_0 d\nu , \end{aligned}$$

since $T^* f_0 = g_0$ a.e. ν and $T^* f_0' = g_0'$ a.e. ν. Thus the proof is completed.

Assumption (3.1) is essential to prove the first part of Theorem 3.3. But notice that for the proof of (3.4), we does not use assumption

(3.1). Thus we have the following result.

COROLLARY 3.1. *The following decomposition always holds*:

$$\int(\sqrt{f}-\sqrt{f_0})^2 d\mu \tag{3.13}$$
$$=\int(\sqrt{g}-\sqrt{g_0})^2 d\nu+\int\sqrt{g(T)g_0(T)}(\sqrt{k}-\sqrt{k_0})^2 d\mu .$$

Consequently, if the family $\{f_\theta;\theta\in\Theta\}$ *is smooth at* θ_0, *then*

$$\lim_{\theta\to\theta_0}\int\left|\sqrt{g(T)g_0(T)}\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2-g_0(T)\kappa_0\right|d\mu=0 .$$

PROOF. To prove the first part, we have only to check that

$$T^*(\sqrt{f}-\sqrt{f_0})^2=(\sqrt{g}-\sqrt{g_0})^2+T^*\{\sqrt{g(T)g_0(T)}(\sqrt{k}-\sqrt{k_0})^2\} ,\quad \text{a.e. } \nu .$$

Since

$$1\geqq\int\sqrt{f\cdot f_0}\,d\mu=\int\sqrt{g(T)g_0(T)}\sqrt{k\cdot k_0}\,d\mu ,$$

the following sum of integrals exists:

$$\int\sqrt{g(T)g_0(T)}\,k d\mu+\int\sqrt{g(T)g_0(T)}\,k_0 d\mu-2\int\sqrt{g(T)g_0(T)}\sqrt{k\cdot k_0}\,d\mu .$$

Hence applying (i) and (iv), we have

$$\begin{aligned}&T^*\{\sqrt{g(T)g_0(T)}(\sqrt{k}-\sqrt{k_0})^2\}\\&=\sqrt{g\cdot g_0}\,T^*k+\sqrt{g\cdot g_0}\,T^*k_0-2\sqrt{g\cdot g_0}\,T^*\sqrt{k\cdot k_0}\\&=2\sqrt{g\cdot g_0}-2\sqrt{g\cdot g_0}\,T^*\sqrt{k\cdot k_0} ,\qquad \text{a.e. } \nu ,\end{aligned}$$

because

$$T^*k=1,\ \text{a.e. } \nu \quad\text{and}\quad T^*k_0=1,\ \text{a.e. } \nu ,$$
$$\text{on } \{t;g(t)>0\}\cap\{t;g_0(t)>0\} .$$

This proves the above relation. Next from (3.6) and (3.12),

$$\sqrt{g(T)g_0(T)}\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2\xrightarrow{l}g_0(T)\kappa_0 .$$

Since both $\{f_\theta;\theta\in\Theta\}$ and $\{g_\theta;\theta\in\Theta\}$ are smooth at θ_0, it follows from (3.4) and (3.13) that

$$\int\sqrt{g(T)g_0(T)}\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2 d\mu\to\int g_0(T)\kappa_0 d\mu .$$

Therefore using Lemma 3.1, 3° we see that

$$\sqrt{g(T)g_0(T)}\left(\frac{\sqrt{k}-\sqrt{k_0}}{\theta-\theta_0}\right)^2 \to g_0(T)\kappa_0 \text{ in mean}, \qquad \text{as } \theta\to\theta_0 .$$

Thus the corollary is established.

Recently, Inagaki [2] has proved the results which are almost equivalent to Theorem 3.2 and Theorem 3.3. His argument is based on random variables $\sqrt{f}/\sqrt{f_0}-1$, $\sqrt{g}/\sqrt{g_0}-1$ and $\sqrt{k}/\sqrt{k_0}-1$. He analysed his results in terms of the conditional expectation and the relative conditional expectation, under the implicit assumptions that ν_0 $(=\mu T^{-1})$ is σ-finite and $=\nu$, and that k is integrable. As referred to earlier, they are not always so. But using the mapping T^*, we can deal with our problem under the most general situation, as demonstrated in this section.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2**, 299-318.
[2] Inagaki, N. (1983). The decomposition of the Fisher information, *Ann. Inst. Statist. Math.*, **35**, 151-165.
[3] Loève, M. (1977). *Probability Theory* I, (4th edn.), Springer-Verlag, New York.
[4] Loève, M. (1978). *Probability Theory* II, (4th edn.), Springer-Verlag, New York.
[5] Pitman, E. J. G. (1979). *Some Basic Theory for Statistical Inference*, Chapman and Hall, London.