TOKYO J. MATH.
VOL. 25, NO. 1, 2002

Differentiability of Densities

Yoichi MIYATA

Waseda University Honjo Senior High School

(Communicated by K. Komatsu)

Abstract. Suppose that p_θ is a probability density of sample X, T is a mapping, $g_\theta(t)$ is an induced probability density by T and $k_\theta(x)$ is a conditional density given $T = t$. Then, the following results are proved under some conditions. (a) L^2-differentiability of the family $(\sqrt{p_\theta})$ is equivalent to that of $(\sqrt{g_\theta})$ and $(\sqrt{k_\theta})$. (b) Regularity of the family (p_θ) is equivalent to that of (g_θ) and (k_θ).

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 $\mathcal{L}$. ν_0 is the measure induced in $\mathcal{L}$ on the σ-algebra $\mathcal{A}$; i.e. $\mathcal{A}$ is the σ-algebra of sets A in $\mathcal{L}$ such that $T^{-1}A \in \mathcal{F}$, and $\nu_0(A) = \mu(T^{-1}A)$.

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 $\mathcal{L}$ from μ_1 is finite and dominates ν_0. Let Θ be an open parameter set in $\mathbb{R}^k$, θ an element of Θ and p_θ a density function on $\mathcal{X}$. Put

$$Q_\theta(A) = \int_{T^{-1}A} p_\theta(x)d\mu\,, \quad A \in \mathcal{A}\,,$$
$$\nu(A) = 0 \Rightarrow \mu(T^{-1}A) = 0 \Rightarrow Q_\theta(A) = 0\,.$$

Hence $\nu \gg Q_\theta$, and so by the Radon-Nikodym theorem there exists a function g_θ on $\mathcal{L}$ such that

$$\int_{T^{-1}A} p_\theta d\mu = Q_\theta(A) = \int_A g_\theta d\nu\,, \quad A \in \mathcal{A}\,.$$

We shall write $g_\theta(t) = E[p_\theta | T = t]$ which is the conditional expectation given $T = t$.

We denote an inner product by $(\cdot,\cdot)$, the transpose of a row vector $\mathbf{a}$ by $\mathbf{a}'$ and $|\mathbf{a}| = \sqrt{(\mathbf{a},\mathbf{a})}$. $f \in L^p(\mu)$ (resp. $f \in L^p(Q_\theta)$) means $\int |f|^p d\mu < \infty$ (resp. $\int |f|^p g_\theta(t)d\nu < \infty$). We define

$$k_\theta(x) = \begin{cases} p_\theta(x)/g_\theta(T(x)) & \text{on } \{x : g_\theta(T(x)) \neq 0\} \\ 0 & \text{otherwise} \end{cases}.$$

Received July 28, 2001; revised October 22, 2001

This is a conditional density given T. It is well known that the family $(\sqrt{g_\theta})$ induced by T is L^2-differentiable with the derivative $E[\dot{p}_\theta(x)|T]/(2\sqrt{g_\theta(t)})$ if the family $(\sqrt{p_\theta})$ is L^2-differentiable. Recent references are Bickel, et al. [1] and Ibragimov and Hasminskii [4]. Further the family $(\sqrt{k_\theta})$ is *smooth* in the sense of conditional densities if the family $(\sqrt{p_\theta})$ is *smooth*. This result is proved by Kuboki [2], using Lebesgue Convergence Theorem and a concept of loosely convergence. The subject of this paper is to give some properties of L^2-differentiability of the families $(\sqrt{p_\theta})$ and $(\sqrt{k_\theta})$. In Section 3, we shall prove L^2-differentiability of $(\sqrt{k_\theta})$ under that of $(\sqrt{p_\theta})$ by a direct calculation (Theorem 3.1). Conversely, when both the families $(\sqrt{g_\theta})$ and $(\sqrt{k_\theta})$ are L^2-differentiable respectively, is the family $(\sqrt{p_\theta})$ L^2-differentiable? This is true. We shall prove this in Theorem 3.2 and also refer the factorization of Fisher information matrix. Furthermore, it is proved that under some conditions, regularity of (p_θ) is equivalent to that of (g_θ) and (k_θ) (Theorems 3.5 and 3.6). Section 4 deals with the proofs of these theorems.

2. Definition and properties of the conditional expectation.

To simply notations, we shall write $g_\theta(t)$ or $g_\theta T(x)$ for $g_\theta(T(x))$ and denote the square root of densities by $s_\theta := \sqrt{p_\theta}$, $q_\theta := \sqrt{g_\theta}$ and $r_\theta := \sqrt{k_\theta}$.

DEFINITION 2.1. The family $(\sqrt{p_\theta})$ is L^2-differentiable if there exists $\dot{s}_\theta \in L^2(\mu)$ such that for every $\theta \in \Theta$,

$$\int |\sqrt{p_{\theta+h}(x)} - \sqrt{p_\theta(x)} - (\dot{s}_\theta(x), h)|^2 d\mu = o(|h|^2)\,. \tag{2.1}$$

DEFINITION 2.2. The family $(\sqrt{k_\theta})$ is L^2-differentiable in the sense of conditional densities if there exists $\dot{r}_\theta \in L^2(Q_\theta)$ such that for every $\theta \in \Theta$,

$$\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)} - (\dot{r}_\theta(x), h)|^2 g_\theta(T(x)) d\mu = o(|h|^2)\,. \tag{2.2}$$

We shall say that the family $(\sqrt{k_\theta})$ is *conditional L^2-differentiable given T* if the family $(\sqrt{k_\theta})$ is L^2-differentiable in the sense of (2.2). We shall prove in Theorem 3.1 that under L^2-differentiability of $(\sqrt{p_\theta})$, the family (k_θ) satisfies (2.2) with the derivative

$$\dot{r}_\theta(x) = \begin{cases} \dfrac{\dot{s}_\theta(x)}{q_\theta T(x)} - \dfrac{s_\theta(x)\dot{q}_\theta T(x)}{q_\theta T(x)^2} & \text{on } \{x : g_\theta T(x) \neq 0\} \\ 0 & \text{otherwise} \end{cases}\,,$$

where $\dot{q}_\theta$ is L^2-derivative of $\sqrt{g_\theta}$.

DEFINITION 2.3. The family $(\sqrt{p_\theta})$ is continuous L^2-differentiable if it satisfies (2.1) and the L^2-derivative $\dot{s}_\theta$ is L^2-continuous, i.e., for every $\theta \in \Theta$,

$$\int |\dot{s}_{\theta+h}(x) - \dot{s}_\theta(x)|^2 d\mu \to 0 \quad \text{as } |h| \to 0\,. \tag{2.3}$$

DEFINITION 2.4. The family $(\sqrt{k_\theta})$ is continuous L^2-differentiable if it satisfies (2.2) and the L^2-derivative $\dot{r}_\theta$ is L^2-continuous in the following sense.

For every $\theta \in \Theta$,

$$\int |\dot{r}_{\theta+h}(x) - \dot{r}_\theta(x)|^2 g_\theta T(x) d\mu \to 0 \quad \text{as } |h| \to 0 . \tag{2.4}$$

We sometimes denote that (p_θ) (resp. (k_θ)) is *regular* if the family $(\sqrt{p_\theta})$ (resp. $(\sqrt{k_\theta})$) is continuous L^2-differentiable.

REMARK 2.5. By the property of the conditional expectation (see Nabeya [5]), $E[\frac{p_\theta(x)}{g_\theta(t)} 1_{[g_\theta>0]}|T] = \frac{1_{[g_\theta>0]}}{g_\theta(t)} E[p_\theta(x)|T] = 1_{[g_\theta>0]}$. Hence, it follows that $E[k_\theta(x)|T] = 1_{[g_\theta(t)>0]} \leq 1$ a.e.v.

The score functions of X and a mapping $T = t$ are defined respectively by

$$l_\theta(x) = \frac{\dot{p}_\theta(x)}{p_\theta(x)} 1_{[x:p_\theta>0]} , \quad l_\theta(t) = \frac{\dot{g}_\theta(t)}{g_\theta(t)} 1_{[t:g_\theta>0]} ,$$

where $\dot{p}_\theta$ (resp. $\dot{g}_\theta$) is L^1-derivative of p_θ (resp. g_θ). It is well-known that $l_\theta(x) = 2\dot{s}_\theta(x)/s_\theta(x) 1_{[s_\theta>0]}$ and $l_\theta(t) = 2\dot{q}_\theta(t)/q_\theta(t) 1_{[q_\theta>0]}$ if the family $(\sqrt{p_\theta})$ is L^2-differentiable. See Bickel, et al. [1].

3. Main results.

First, we consider the following conditions.

CONDITION A. $\int |l_\theta(t)|^2 k_{\theta+h}(x) g_\theta T(x) d\mu$ exists for small h.

CONDITION B. $\int |l_{\theta+h}(x)|^2 k_{\theta+h}(x) g_\theta T(x) d\mu$ and $\int |l_{\theta+h}(t)|^2 g_\theta T(x) d\mu$ exist for small h, respectively.

The above conditions are needed as the integrability when we shall prove Theorems 3.1, 3.2 and so on. Condition A is a weak assumption. Actually, condition A is satisfied automatically from $\int k_\theta(x, y) d\nu \leq 1$ in the case of $k_\theta(x, y) = p_\theta(x, y)/p_\theta(x)$. The detail is referred in Example 3.3.

THEOREM 3.1. *Suppose that* $(\sqrt{p_\theta})$ *is* L^2*-differentiable and condition A is satisfied. Then the family* $(\sqrt{k_\theta})$ *is* L^2*-differentiable in the sense of* (2.2) *with the derivative*

$$\dot{r}_\theta(x) = \frac{\dot{s}_\theta(x) q_\theta T(x) - s_\theta(x) \dot{q}_\theta T(x)}{q_\theta T(x)^2} 1_{[x:g_\theta T(x)>0]} .$$

THEOREM 3.2. *Suppose that the family* $(\sqrt{k_\theta})$ *is* L^2*-differentiable in the sense of* (2.2) *with the derivative* $\dot{r}_\theta(x)$, $(\sqrt{g_\theta})$ *is* L^2*-differentiable with the derivative* $\dot{q}_\theta(t)$ *and that condition A is satisfied. Then the family* $(\sqrt{p_\theta})$ *is* L^2*-differentiable with the derivative*

$$\dot{s}_\theta(x) = \dot{r}_\theta(x)\sqrt{g_\theta(T(x))} + \sqrt{k_\theta(x)}\dot{q}_\theta(T(x)) .$$

L^2-differentiability of induced probability densities is common knowledge (see Bickel, et al. [1]). Hence, by combining Theorems 3.1 and 3.2, we can see that the family $(\sqrt{p_\theta(x)})$ is

L^2-differentiable iff $(\sqrt{g_\theta(t)})$ is L^2-differentiable and $(\sqrt{k_\theta(x)})$ is conditional L^2-differentiable given T.

EXAMPLE 3.3. Suppose that $\{\mathcal{X}, \mathcal{A}, \mu\}$ and $\{\mathcal{Y}, \mathcal{B}, \nu\}$ are two measure spaces. Let $\mathcal{A}$ and $\mathcal{B}$ be σ-algebras of subsets of $\mathcal{X}$ and $\mathcal{Y}$, and μ and ν be σ-finite measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively. $\{\mathcal{X}\times\mathcal{Y}, \mathcal{A}\times\mathcal{B}, \mu\times\nu\}$ is the cartesian product space of $\{\mathcal{X}, \mathcal{A}, \mu\}$ and $\{\mathcal{Y}, \mathcal{B}, \nu\}$. Let (P_θ^{XY}) be probability distributions on $\mathcal{X}\times\mathcal{Y}$ with densities $p_\theta(x, y)$ relative to a σ-finite measure $\mu\times\nu$. Then the induced probability density by a mapping $T : \mathcal{X}\times\mathcal{Y} \to \mathcal{X}$ is marginal $p_\theta(x)$. Thus the family $(\sqrt{p_\theta(x, y)})$ is L^2-differentiable iff both the families $(\sqrt{p_\theta(x)})$ and $(\sqrt{p_\theta(y|x)})$ are L^2-differentiable. Here, L^2-differentiability of $(\sqrt{p_\theta(y|x)})$ is correspond to (2.2), i.e., there exists $\dot{s}_\theta(y|x) \in L^2(P_\theta^X \times \nu)$ such that

$$\iint |\sqrt{p_{\theta+h}(y|x)} - \sqrt{p_\theta(y|x)} - (h, \dot{s}_\theta(y|x))|^2 p_\theta(x) d\mu d\nu = o(|h|^2)\,.$$

The concept of L^2-differentiability is available for a discrete probability function with an open parameter set since counting measure μ is σ-finite.

EXAMPLE 3.4. Suppose that the conditional density of x given k is

$$f_p(x|k) = \binom{n}{x-[k]} p^{x-[k]}(1-p)^{n-x+[k]} \quad x = [k], [k]+1, \cdots, [k]+n$$

and a random variable k is according to $N(\mu, \sigma^2)$, where $[\cdot]$ is Gaussian integer and $0 < p < 1$, μ and σ are unknown parameters, that is $\theta = (p, \mu, \sigma)$. Since $p^{x-[k]}(1-p)^{n-x+[k]}$ is ordinary differentiable with respect to p, it is obvious that

$$\sum_{x=[k]}^{n+[k]} \left(\frac{\sqrt{f_{p+h}(x|k)} - \sqrt{f_p(x|k)}}{h} - \dot{s}_p(x|k) \right)^2 \to 0 \quad \text{as } |h| \to 0\,, \tag{3.1}$$

where $\dot{s}_p(x|k) = \sqrt{\binom{n}{x-[k]}} \left(\frac{x-[k]-np}{2}\right) p^{\frac{x-[k]-2}{2}}(1-p)^{\frac{n-x+[k]-2}{2}}$. Furthermore, putting $w = x - [k]$, (3.1) does not depend on k. Thus it is verified that $(\sqrt{f_p(x|k)})$ is conditional L^2-differentiable given k. On the other hand, it is well known that the normal density of k is L^2-differentiable. Hence so is the joint density of (X, K).

Subsequently, we shall state the factorization of information matrix. We define the information matrix of p_θ, g_θ as

$$I_X(\theta) = 4\int \dot{s}_\theta(x)\dot{s}_\theta(x)' d\mu\,, \quad I_T(\theta) = 4\int \dot{r}_\theta(t)\dot{r}_\theta(t)' d\nu\,,$$

respectively and the conditional information matrix of $k_\theta(x)$ given T as $I_{X|T}(\theta) = E[(\dot{k}_\theta\dot{k}_\theta')/k_\theta|T]$. It is known that $I_X(\theta) = E_\theta^T\{I_{X|T}(\theta)\} + I_T(\theta)$ under L^2-differentiability of the family $(\sqrt{p_\theta})$. See Kuboki [2].

THEOREM 3.5. *Suppose that the family $(\sqrt{k_\theta})$ is conditional L^2-differentiable given T and $(g_\theta(t))$ is L^2-differentiable. Then,*

$$I_X(\theta) = E_\theta^T\{I_{X|T}(\theta)\} + I_T(\theta)\,.$$

PROOF. Applying Theorem 3.2, the family $(\sqrt{p_\theta})$ is L^2-differentiable from the assumption. By noting $\dot{g}_\theta(t) = E[\dot{p}_\theta(x)|T]$ and $\dot{r}_\theta(x) = \dot{s}_\theta(x)/q_\theta(t) - s_\theta(x)\dot{q}_\theta(t)/(q_\theta(t)^2)$, the assertion is proved. □

There are some papers which treat continuous L^2-differentiability instead of mere L^2-differentiability because things are easier. Here we shall state the following theorems concerned with regularity. Let a score function be $l_\theta(x) = (\dot{p}_\theta(x)/p_\theta(x))1_{[p_\theta>0]}$.

THEOREM 3.6. *Suppose that the family* (p_θ) *is regular and conditions A, B are satisfied. Then the family* (k_θ) *is regular in the sense of Definition* 2.4.

THEOREM 3.7. *Suppose that the family* (g_θ) *is regular, the family* (k_θ) *is regular in the sense of Definition* 2.4 *with the derivative* $\dot{r}_\theta$ *and conditions A, B are satisfied. Then the family* (p_θ) *is regular.*

When the family (p_θ) is regular, regularity of induced probability densities by T is proved by Bickel, et al. [1]. Hence, by combining Theorems 3.6 and 3.7, we can see that under conditions A, B, regularity of the family (p_θ) is equivalent to that of (g_θ) and (k_θ).

Finally, we shall introduce the useful theorem for checking L^2-differentiability of regression models. Let the conditional density of y given x be $p_\theta(y|x) = p_\theta(x,y)/p_\theta(x)$ on $\{x : p_\theta > 0\}$ and 0 on $\{x : p_\theta = 0\}$. In some cases, it is difficult to check L^2-differentiability of $(\sqrt{p_\theta(y|x)})$ in measure $P_\theta^X \times \nu$ even if $(\sqrt{p_\theta(y|x)})$ is L^2-differentiable in measure ν. The following Theorem 3.8 gives the sufficient condition of conditional L^2-differentiability given $T : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$, i.e.,

$$\iint |s_{\theta+h}(y|x) - s_\theta(y|x) - (\dot{s}_\theta(y|x), h)|^2 d\nu dP_\theta^X = o(|h|^2) \quad \text{for every } \theta\,.$$

This result is proved by Strasser [3], Theorem 3.4, p. 120. Suppose that for every $x \in \mathcal{X}$, the family $(\sqrt{p_\theta(y|x)})$ is L^2-differentiable in measure ν, i.e.,

$$\int |s_{\theta+h}(y|x) - s_\theta(y|x) - (\dot{s}_\theta(y|x), h)|^2 d\nu = o(|h|^2) \quad \text{for every } \theta\,.$$

Note that we shall use the same notation $\dot{s}_\theta(y|x)$ for the L^2-derivative in measure $P_\theta^X \times \nu$ as well as for the L^2-derivative in measure ν. For the following theorem, we consider continuous L^2-differentiability instead of mere L^2-differentiability. We denote Fisher's information matrix with respect to ν by $I_{Y|x}(\theta) := 4\int \dot{s}_\theta(y|x)\dot{s}_\theta(y|x)'d\nu$. Let $I_{Y|x,h}(\theta) := I_{Y|x}(\theta + h)$. $I_{Y|x,h}(\theta)$ is called to be *uniformly* P_θ^X*-integrable* if $\lim_{M\to\infty} \sup_h \int_{I_{Y|x,h}(\theta)>M} |I_{Y|x,h}(\theta)| dP_\theta^X = 0$ for every $\theta \in \Theta$. If $I_{Y|x,h}(\theta)$ is uniformly P_θ^X*-integrable*, it follows that

$$E_\theta^X\{I_{Y|x,h}(\theta)\} \to E_\theta^X\{I_{Y|x}(\theta)\} \quad \text{as } |h| \to 0\,. \tag{3.2}$$

Of course, if there exist P_θ^X-integrable functions $H_h(x : \theta)$ such that for any small h, $I_{Y|x,h}(\theta) \le H_h(x : \theta)$ a.e. P_θ^X and $\int H_h(x : \theta)dP_\theta^X \to \int H(x : \theta)dP_\theta^X$ as $|h| \to 0$, then (3.2) holds from Lebesgue Convergence Theorem. Let $\lambda = \nu \times P_\theta^X$. $\dot{s}_\theta(y|x)$ is called to be $\nu \times P_\theta^X$*-continuous* if for every $\theta \in \Theta$, $\lambda\{(x,y) : |\dot{s}_{\theta+h}(y|x) - \dot{s}_\theta(y|x)| \ge \varepsilon\} \to 0$ as $|h| \to 0$.

THEOREM 3.8. *Suppose that for every $x \in \mathcal{X}$, the family $(\sqrt{p_\theta(y|x)})$ is continuous L^2-differentiable in measure ν with the derivative $\dot{s}_\theta(y|x)$. Let Fisher's information matrix with respect to ν be $I_{Y|x}(\theta)$. If the family of functions $I_{Y|x,h}(\theta)$ satisfies* (3.2) *and $\dot{s}_\theta(y|x)$ is $\nu \times P_\theta^X$-continuous, then the family $(\sqrt{p_\theta(y|x)})$ is continuous L^2-differentiable in the sense of Definition* 2.4.

EXAMPLE 3.9 (Normal Linear Model).

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, \cdots, n)$$

where $\varepsilon_1, \cdots, \varepsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$, σ is known and α and β are unknown.

Suppose that random variables $x_1, \cdots, x_n$ are i.i.d with common density f_θ which is continuous L^2-differentiable at θ with the expectation $\mu(\theta) = E_\theta X_i$, and each measure of Y_i, x_i $(i = 1, \cdots, n)$ is Lebesgue measure. Let $Y = (Y_1, \cdots, Y_n)'$ and $x = (x_1, \cdots, x_n)'$. Let $z(\cdot)$ be a density of standard normal distribution and the family $(p_{\alpha,\beta}(y|x))$ be the conditional densities of Y given x. It is verified that $p_{\alpha,\beta}(y|x) = \sigma^{-n} \prod_{i=1}^n z((y_i - \alpha - \beta x_i)/\sigma)$ is continuous L^2-differentiable in measure μ^Y. (See Bickel, et al. [1], Proposition 2).

Since $I_{Y|x,h}(\alpha, \beta, \theta) = \begin{pmatrix} n/\sigma^2 & \sum_{i=1}^n x_i/\sigma^2 & 0 \\ \sum_{i=1}^n x_i/\sigma^2 & \sum_{i=1}^n x_i^2/\sigma^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ does not depend on h and $\mu(\theta)$ is continuous from L^2-differentiability of f_θ, $E^X\{I_{Y|x}(\alpha, \beta, \theta)\} = \begin{pmatrix} n/\sigma^2 & n\mu(\theta)/\sigma^2 & 0 \\ n\mu(\theta)/\sigma^2 & n(\sigma^2 + \mu(\theta)^2)/\sigma^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ satisfies (3.2).

Thus, $(p_{\alpha,\beta}(y|x))$ satisfies (2.2) and (2.4) from Theorem 3.8. Applying Theorem 3.6, $p_{\theta,\alpha,\beta}(x, y) = \sigma^{-n} \prod_{i=1}^n z((y_i - \alpha - \beta x_i)/\sigma) f_\theta(x_i)$ is continuous L^2-differentiable.

4. Proofs.

In this section, we shall prove the theorems stated in Section 3. Let the partition of $\mathcal{L}$, $A_h := \{t : g_\theta(t) > 0 \text{ and } g_{\theta+h}(t) > 0\}$, $B_h := \{t : g_\theta(t) > 0 \text{ and } g_{\theta+h}(t) = 0\}$ and $C_h := \{t : g_\theta(t) = 0 \text{ and } g_{\theta+h}(t) > 0\}$.

LEMMA 4.1. *Suppose that the family $(\sqrt{p_\theta})$ is L^2-differentiable and $\xi(T(x))$ is a function such that $\int \xi(t(x)) k_{\theta+h}(x) g_\theta T(x) d\mu < \infty$. Then the following holds*

$$\text{(i)} \quad \int (r_{\theta+h}(x) - r_\theta(x))^2 g_\theta T(x) d\mu = O(|h|^2)$$

$$\text{(ii)} \quad \int \xi(t(x))(r_{\theta+h}(x) - r_\theta(x))^2 g_\theta T(x) d\mu = o(1)\,.$$

PROOF. First, we shall prove Lemma 4.1(i).

$$\int_{T^{-1}A_h}\left(\frac{s_{\theta+h}(x)}{q_{\theta+h}T(x)}-\frac{s_\theta(x)}{q_\theta T(x)}\right)^2 g_\theta T(x)d\mu$$
$$\leq \int_{T^{-1}A_h}(q_{\theta+h}T(x)-q_\theta T(x))^2\frac{p_{\theta+h}(x)}{g_{\theta+h}T(x)}d\mu+\int(s_{\theta+h}(x)-s_\theta(x))^2 d\mu$$
$$=\int_{A_h}(q_{\theta+h}(t)-q_\theta(t))^2 E\Big[\frac{p_{\theta+h}(x)}{g_{\theta+h}T(x)}\,\Big|\,T\Big]d\nu+O(|h|^2) \tag{4.1}$$

From L^2-differentiability of $(\sqrt{g_\theta})$ and Remark 2.5, we have (4.1) $=O(|h|^2)$. On the other hand, it is easily verified that $\int_{T^{-1}B_h}|r_{\theta+h}(x)-r_\theta(x)|^2 g_\theta T(x)d\mu = O(|h|^2)$ and $\int_{T^{-1}C_h}|r_{\theta+h}(x)-r_\theta(x)|^2 g_\theta T(x)d\mu=0$. Hence (i) is proved. Next, we shall prove Lemma 4.1(ii). By using the partition of integral domain, we have

$$\int|\xi(T(x))|\,|\sqrt{k_{\theta+h}(x)}-\sqrt{k_\theta(x)}|^2 g_\theta T(x)d\mu$$
$$\leq\int_{T^{-1}\left[t:|\xi(t)|>\frac{1}{|h|}\right]}|\xi(T(x))|\,|\sqrt{k_{\theta+h}(x)}-\sqrt{k_\theta(x)}|^2 g_\theta T(x)d\mu \tag{4.2}$$
$$+\frac{1}{|h|}\int|\sqrt{k_{\theta+h}(x)}-\sqrt{k_\theta(x)}|^2 g_\theta T(x)d\mu \tag{4.3}$$

It holds from Lemma 4.1(i) that (4.3) $\to 0$ as $|h|\to 0$. Since we have $E[k_\theta(x)|T]\leq 1$ a.e. ν for every θ from Remark 2.5, (4.2) is bounded above by

$$\int_{\left[t:|\xi(t)|>\frac{1}{|h|}\right]}|\xi(t)|g_\theta(t)(E[k_{\theta+h}(x)|T]+E[k_\theta(x)|T])d\nu$$
$$\leq 2\int|\xi(t)|1_{\left[|\xi(t)|>\frac{1}{|h|}\right]}g_\theta(t)d\nu\,. \tag{4.4}$$

It follows from Lebesgue Convergence Theorem that (4.4) $\to 0$ as $|h|\to 0$. Hence (ii) is proved. □

PROOF OF THEOREM 3.1. From $r_\theta=s_\theta/q_\theta$ and $r_{\theta+h}=s_{\theta+h}/q_{\theta+h}$ on $T^{-1}A_h$, it follows that

$$\int_{T^{-1}A_h}|r_{\theta+h}(x)-r_\theta(x)-(h,\dot r_\theta(x))|^2 g_\theta T(x)d\mu$$
$$\leq 2\int_{T^{-1}A_h}|s_{\theta+h}(x)-s_\theta(x)-(h,\dot s_\theta(x))|^2 d\mu \tag{4.5}$$
$$+2\int_{T^{-1}A_h}|(q_{\theta+h}T(x)-q_\theta T(x))r_{\theta+h}(x)-(h,r_\theta(x)\dot q_\theta T(x))|^2 d\mu\,. \tag{4.6}$$

It is clear that $(4.5) = o(|h|^2)$. (4.6) is bounded above by

$$4\int |q_{\theta+h}T(x) - q_\theta T(x) - (h, \dot{q}_\theta T(x))|^2 k_{\theta+h}(x)d\mu \tag{4.7}$$

$$+ 4\int |r_{\theta+h}(x) - r_\theta(x)|^2 \left(h, \frac{\dot{q}_\theta T(x)}{q_\theta T(x)}\right)^2 g_\theta T(x)d\mu\,. \tag{4.8}$$

By Remark 2.5, it holds that $(4.7) = o(|h|^2)$. Setting $l_\theta(t) = 2(\dot{q}_\theta T(x)/q_\theta T(x))$, it follows from Lemma 4.1 (ii) that $(4.8) = o(|h|^2)$. Hence the assertion is proved. □

LEMMA 4.2. *Suppose that the family* $(\sqrt{k_\theta})$ *is conditional* L^2*-differentiable given T. Then*,

$$\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)}|^2 g_\theta T(x)d\mu = O(|h|^2)\,.$$

PROOF.

$$\begin{aligned}
&\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)}|^2 g_\theta T(x)d\mu \\
&\quad \le 2\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)} - (\dot{r}_\theta(x), h)|^2 g_\theta T(x)d\mu + 2\int |(h, \dot{r}_\theta(x))|^2 g_\theta T(x)d\mu \\
&\quad = O(|h|^2)\,.
\end{aligned}$$

Thus, the lemma is proved. □

LEMMA 4.3. *Suppose that* $(\sqrt{k_\theta})$ *is* L^2*-differentiable and* $\int \xi(t(x))k_{\theta+h}(x)g_\theta T(x)d\mu < \infty$ *for small h.*

$$\int |\xi(T(x))|\,|\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)}|^2 g_\theta T(x)d\mu = o(1)\,.$$

PROOF. This is proved by the same argument as in the proof of Lemma 4.1 (ii). □

PROOF OF THEOREM 3.2. Since $p_\theta(x) = k_\theta(x)g_\theta(t)$ on $\{x : g_\theta > 0\}$, we have

$$\int_{T^{-1}A_h} |\sqrt{p_{\theta+h}(x)} - \sqrt{p_\theta(x)} - (\dot{s}_\theta(x), h)|^2 d\mu$$

$$\le 3\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)} - (\dot{r}_\theta(x), h)|^2 g_\theta T(x)d\mu \tag{4.9}$$

$$+ 3\int |\sqrt{g_{\theta+h}T(x)} - \sqrt{g_\theta T(x)} - (\dot{q}_\theta T(x), h)|^2 k_{\theta+h}(x)d\mu \tag{4.10}$$

$$+ 3\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)}|^2 (\dot{q}_\theta T(x), h)|^2 d\mu\,. \tag{4.11}$$

It is clear from the assumption that $(4.9) = o(|h|^2)$. Applying Remark 2.5, $(4.10) = o(|h|^2)$. (4.11) does not exceed

$$\frac{3}{4}|h|^2 \int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)}|^2 |l_\theta(t)|^2 g_\theta T(x)d\mu\,,$$

where $l_\theta(t) = 2\dot{q}_\theta(t)/q_\theta(t)$. It follows from Lemma 4.3 that (4.9) $= o(|h|^2)$. On the other hand, we have $p_{\theta+h} = 0$ a.e. μ on $T^{-1}B_h$ since $0 = \int_{B_h} g_{\theta+h} d\nu = \int_{T^{-1}B_h} p_{\theta+h} d\mu$. Hence

$$\begin{aligned}\int_{T^{-1}B_h} &|\sqrt{p_{\theta+h}(x)} - \sqrt{p_\theta(x)} - (\dot{s}_\theta(x), h)|^2 d\mu \\ &\le \int_{T^{-1}B_h} p_\theta(x) d\mu + \int_{T^{-1}B_h} |(\dot{s}_\theta(x), h)|^2 d\mu \\ &\le \int_{B_h} g_\theta(t) d\nu + |h|^2 \int_{T^{-1}B_h} |\dot{s}_\theta(x)|^2 d\mu \\ &= o(|h|^2)\,.\end{aligned}$$

It is easily verified that $\int_{T^{-1}C_h} |\sqrt{p_{\theta+h}(x)} - \sqrt{p_\theta(x)} - (\dot{s}_\theta(x), h)|^2 d\mu = o(|h|^2)$. Thus, the assertion is proved. □

LEMMA 4.4. *Suppose that the family* $(\sqrt{g_\theta})$ *is regular and* $\int |l_{\theta+h}(t)|^2 g_\theta T(x) d\mu < \infty$. *Then the following holds*

$$\int |l_{\theta+h}(t) - l_\theta(t)^2 g_\theta T(x) d\mu = o(1)\,.$$

PROOF.

$$\begin{aligned}4 \int_{T^{-1}A_h} &\left| \frac{\dot{q}_{\theta+h}T(x)}{q_{\theta+h}T(x)} - \frac{\dot{q}_\theta T(x)}{q_\theta T(x)} \right|^2 g_\theta T(x) d\mu \\ &= 4 \int_{T^{-1}A_h} \left| \dot{q}_{\theta+h}T(x) \frac{q_\theta T(x)}{q_{\theta+h}T(x)} - \dot{q}_\theta T(x) \right|^2 d\mu \\ &\le 2 \int_{T^{-1}A_h} |q_{\theta+h}T(x) - q_\theta T(x)|^2 |l_{\theta+h}(t)|^2 d\mu + 8 \int |\dot{q}_{\theta+h}T(x) - \dot{q}_\theta T(x)|^2 d\nu \\ &= 4 \int_{A_h} \left(\frac{q_{\theta+h}(t)}{q_\theta(t)} - 1 \right)^2 |l_{\theta+h}(t)|^2 g_\theta(t) d\nu + o(1)\,. \end{aligned} \tag{4.12}$$

By the same argument as Lemma 4.1 (ii), (4.12) tends to 0 as $|h| \to 0$. It is verified that $\int_{T^{-1}B_h} |l_{\theta+h}(t) - l_\theta(t)|^2 g_\theta T(x) d\mu = o(1)$ and $\int_{T^{-1}C_h} |l_{\theta+h}(t) - l_\theta(t)|^2 g_\theta T(x) d\mu = 0$. Hence the assertion is proved. □

LEMMA 4.5. *Suppose that the family* $(\sqrt{p_\theta})$ *is regular and* $\int |l_{\theta+h}(x)|^2 k_{\theta+h}(x) g_\theta T(x) d\mu < \infty$. *Then, it follows that*

$$\int |l_{\theta+h}(x) r_{\theta+h}(x) - l_\theta(x) r_\theta(x)|^2 g_\theta T(x) d\mu = o(1)\,.$$

PROOF. Since $l_\theta(x)r_\theta(x) = \dot{s}_\theta(x)/q_\theta T(x)$, we have

$$\begin{aligned} 4\int_{T^{-1}A_h} &\left| \frac{\dot{s}_{\theta+h}(x)}{q_{\theta+h}T(x)} - \frac{\dot{s}_\theta(x)}{q_\theta T(x)} \right|^2 g_\theta T(x) d\mu \\ &\le 8\int_{T^{-1}A_h} \left| \frac{\dot{s}_{\theta+h}(x)}{q_{\theta+h}T(x)} q_\theta T(x) - \dot{s}_{\theta+h}(x) \right|^2 d\mu \qquad (4.13) \\ &\quad + 8\int |\dot{s}_{\theta+h}(x) - \dot{s}_\theta(x)|^2 d\mu\,. \end{aligned}$$

Putting $\xi_{\theta+h}(t) := E[|l_{\theta+h}(x)r_{\theta+h}(x)|^2 \mid T]$, the following inequality holds in (4.13) by the same argument as Lemma 4.1 (ii).

$$\begin{aligned} (4.13) &\le 8\int_{T^{-1}A_h} |q_{\theta+h}T(x) - q_\theta T(x)|^2 |l_{\theta+h}(x)r_{\theta+h}(x)|^2 d\mu \\ &= 8\int_{A_h} |q_{\theta+h}(t) - q_\theta(t)|^2 E[|l_{\theta+h}(x)r_{\theta+h}(x)|^2 \mid T] d\nu \\ &\le \frac{8}{|h|}\int |q_{\theta+h}(t) - q_\theta(t)|^2 d\nu + 8\int_{|\xi_{\theta+h}(t)| > \frac{1}{\sqrt{|h|}}} |\xi_{\theta+h}(t)|(g_{\theta+h}(t) + g_\theta(t) d\nu\,. \end{aligned}$$

Therefore the assertion is proved. □

PROOF OF THEOREM 3.5. Since L^2-differentiability of the family $(\sqrt{k_\theta})$ is proved in Theorem 3.1, it is sufficient to prove L^2-continuity of the derivative $\dot{r}_\theta(x)$. By using $\dot{r}_\theta(x) = (l_\theta(x) - l_\theta(t))r_\theta(x)$, it follows that

$$\begin{aligned} \int &|\dot{r}_{\theta+h}(x) - \dot{r}_\theta(x)|^2 g_\theta T(x) d\mu \\ &= \int |(l_{\theta+h}(x) - l_{\theta+h}(t))r_{\theta+h}(x) - (l_\theta(x) - l_\theta(t))r_\theta(x)|^2 g_\theta T(x) d\mu \\ &\le 2\int |l_{\theta+h}(x)r_{\theta+h}(x) - l_\theta(x)r_\theta(x)|^2 g_\theta T(x) d\mu \qquad (4.14) \\ &\quad + 2\int |l_{\theta+h}(t)r_{\theta+h}(x) - l_\theta(t)r_\theta(x)|^2 g_\theta T(x) d\mu\,. \qquad (4.15) \end{aligned}$$

It follows from Lemma 4.5 that (4.14) $\to 0$ as $|h| \to 0$. (4.15) does not exceed

$$4\int |l_{\theta+h}(t) - l_\theta(t)|^2 g_\theta(t) d\nu + 4\int |r_{\theta+h}(x) - r_\theta(x)|^2 |l_\theta(t)|^2 g_\theta(t) d\mu\,. \qquad (4.16)$$

Using Lemma 4.3 and 4.4, (4.16) $\to 0$ as $|h| \to 0$. Therefore (4.15) $\to 0$ as $|h| \to 0$. Hence the assertion is proved. □

PROOF OF THEOREM 3.6. It is sufficient to prove that the L^2-derivative $\dot{s}_\theta(x)$ is L^2-continuous since L^2-differentiability is proved in Theorem 3.2. For $\dot{s}_\theta = \dot{r}_\theta\sqrt{g_\theta} + \sqrt{k_\theta}\dot{q}_\theta$,

$$\begin{aligned}
&\int |\dot{s}_{\theta+h}(x) - \dot{s}_\theta(x)|^2 d\mu \\
&\quad = \int |\dot{r}_{\theta+h}(x)\sqrt{g_{\theta+h}(t)} + \sqrt{k_{\theta+h}(x)}\dot{q}_{\theta+h}(t) - \dot{r}_\theta(x)\sqrt{g_\theta(t)} - \sqrt{k_\theta(x)}\dot{q}_\theta(t)|^2 d\mu \\
&\quad \le 4\int |\dot{r}_{\theta+h}(x) - \dot{r}_\theta(x)|^2 g_\theta T(x) d\mu \qquad (4.17) \\
&\qquad + 4\int |\sqrt{g_{\theta+h}(t)} - \sqrt{g_\theta(t)}|^2 |\dot{r}_{\theta+h}(x)|^2 d\mu \qquad (4.18) \\
&\qquad + 4\int |\dot{q}_{\theta+h}(t) - \dot{q}_\theta(t)|^2 k_{\theta+h}(x) d\mu \qquad (4.19) \\
&\qquad + 4\int |\sqrt{k_{\theta+h}(x)} - \sqrt{k_\theta(x)}|^2 |\dot{q}_\theta(t)|^2 d\mu\,. \qquad (4.20)
\end{aligned}$$

We have (4.18) $\to 0$ as $|h| \to 0$ by the same argument in (4.12). Applying Remark 2.5, (4.19) $\to 0$ as $|h| \to 0$. By Lemma 4.3, (4.20) $\to 0$ as $|h| \to 0$. Hence the assertion is proved. □

ACKNOWLEDGEMENTS. The author thanks the editor and the referee for their helpful comments which have improved the presentation of the present paper.

References

[1] P. J. BICKEL, C. A. J. KLAASSEN, Y. RITOV and J. A. WELLNER, *Efficient and adaptive estimation for semiparametric models*, Johns Hopkins Univ. press (1993).

[2] H. KUBOKI, A generalization of the relative conditional expectation, Ann. Inst. Statist. Math. **36** (1984), 181–197.

[3] H. STRASSER, Differentiability of statistical experiment, Statistics & Decisions **16** (1998), 113–130.

[4] I. A. IBRAGIMOV and R. Z. HAS'MINSKII, *Statistical Estimation Asymptotic Theory*, Springer (1981).

[5] S. NABEYA, *Mathematical Statistics* (in Japanese), Kyōritu Press (1978).

Present Address:
WASEDA UNIVERSITY HONJO SENIOR HIGH SCHOOL,
OKUBOYAMA NISHITOMITA, HONJO-SHI, SAITAMA, 367–0045 JAPAN.
e-mail: ymiyata@mn.waseda.ac.jp