MINIMUM HELLINGER DISTANCE ESTIMATION OF PARAMETERS IN THE RANDOM CENSORSHIP MODEL

BY

Song Yang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1988

ABSTRACT

MINIMUM HELLINGER DISTANCE ESTIMATION OF PARAMETERS IN THE RANDOM CENSORSHIP MODEL

By

Song Yang

Let $X_1, \cdots, X_n$ be i.i.d. with c.d.f. F, and $Y_1, \cdots, Y_n$ be independent of X_i's and i.i.d. with an unknown censoring c.d.f. G. In the random censorship model, the pairs $(\min(X_i, Y_i), [X_i \le Y_i])$, $i = 1, \cdots, n$, are observed, where [A] denotes the indicator of the set A. Let F have a density f and $\{f_\theta : \theta \in \Theta\}$ be a parametric family of densities, where Θ is a subset of the p-dimensional Euclidean space. This thesis discusses the minimum Hellinger distance estimation (MHDE) of the parameter that gives the "best fit" of the parametric family to the data.

In studying the MHDE, the tail behavior of the product-limit processes is investigated and the weak convergence of these processes on the real line is established. In addition, an upper bound on the mean square increment of the normalized product-limit process is obtained. Based on the global behavior of the product-limit processes, kernel density estimators are constructed and shown to be consistent under Hellinger metric. Using these results, it is shown that, when f belongs to the parametric model, the MHD estimators are asymptotically efficient among the class of "regular" estimators; they are also minimax robust in small Hellinger neighborhoods of the given parametric family.

The work extends the results of Beran (1977; Minimum Hellinger distance estimates for parametric models. Ann. Statist. 5, 445-463) for the complete i.i.d. data case to the censored data case. Some of the proofs employ the martingale techniques developed by Gill(1980; Censoring and Stochastic Integrals. Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam).

To my parents and my brother Tao

ACKNOWLEDGEMENTS

I would like to thank my thesis adviser Hira L. Koul for his support and encouragement during the preparation of this dissertation. His guidance in my endeavor in statistics is greatly appreciated. I would also like to thank Professor V. Mandrekar, Professor James Hannan and Professor Clifford Weil for serving on my guidance committee and reviewing this manuscript. My special thanks go to Professor Mandrekar and Professor Hannan for their many helpful suggestions and stimulating discussions.

Thanks are also due to Cathy Sparks and especially Loretta Ferguson for their help in typing this manuscript.

Finally I express my deep appreciation to my parents and my brother Tao for their support and encouragement, which have strengthened me a great deal since my earlier education.

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY

Let $X_1, \cdots, X_n$ be independent and identically distributed (i.i.d.) random variables with cumulative distribution function (c.d.f.) F on $[0,\infty)$, and $Y_1, \cdots, Y_n$ be independent of X_i's and i.i.d. with (sub-)c.d.f. G on $[0,\infty]$ (i.e., G may assign positive mass to ∞). In the random censorship model, the pairs $\{\min(X_i, Y_i), [X_i \leq Y_i]\}$, $1 \leq i \leq n$, are observed, where [A] denotes the indicator function of the event A. Suppose that F has a density f with respect to Lebesgue measure, and some physical theory suggests that f belongs to a parametric family $\{f_\theta: \theta \in \Theta\}$, where Θ is a subset of p-dimensional Euclidean space. At the same time we recognize that, due to a variety of data contamination, f may possibly differ from any of the f_θ's. The problem is to estimate the parameter that gives the "best fit" of the parametric model to the data.

When G is degenerate at ∞, i.e. when we are able to observe the complete data $X_1, \cdots, X_n$, there have been many results in the literature. Millar(1983) illustrated that in many cases when the "best fit" is given via a minimum distance recipe, there usually exists a minimax structure, and the minimum distance estimator usually has the local asymptotic minimaxity property, which is defined to be robustness there. While there is quite a bit of freedom in choosing the distance, one distance — Hellinger distance — has the merit that the estimation procedure is asymptotically efficient if there were no contamination, as discussed in Beran(b,1977; 1981). It is heuristically illustrated in Beran(b,1977) that the minimum Hellinger distance estimator considered there is closely related to the maximum likelihood estimator and therefore asymptotic efficiency seems plausible.

In this thesis, the minimum Hellinger distance estimation (MHDE) in the random censorship model is considered. It turns out that, as in the i.i.d. complete data case discussed in Beran(b,1977), when there is no contamination, this procedure is asymptotically efficient among the class of "regular" estimators; it is also robust in a minimax sense in small Hellinger neighborhoods of the parametric model.

The material is organized as follows. In Chapter 2, some preliminary results are introduced. The tail behaviors of the product—limit processes are investigated and the weak convergences of these processes on the entire support set are established. The convergence of the kernel density estimators in the Hellinger metric is obtained. In addition, an upper bound on the mean square increment of the normalized product—limit process is developed. In Chapter 3, the differentiability of the minimum Hellinger distance functionals is studied. In Chapter 4, the asymptotic behavior of the MHDE is investigated and it is shown that this procedure is asymptotically efficient if there were no contamination. In Chapter 5, a minimax robustness property of the MHDE is established.

Notational Remarks. Throughout this thesis, $X_1, \cdots, X_n$, $Y_1, \cdots, Y_n$ are independent r.v.'s. Unless mentioned otherwise, for $i = 1, \cdots, n$, X_i, Y_i have distributions F, G respectively. $\delta_i = [X_i \leq Y_i]$, where [A] denotes the indicator function of the set A, and $\tilde{X}_i = \min(X_i, Y_i)$ with c.d.f. H. For any function ξ, $\xi_-(x) = \xi(x-)$, $\xi_+(x) = \xi(x+)$. For any (sub-)c.d.f. D, $D^{-1}(t) = \inf\{u: D(u) \geq t\}$, $\tau_D = D^{-1}(1) \leq \infty$, and $\bar{D} = 1-D$, $\Delta D = D - D_-$. Note that $\bar{H} = \bar{F}\,\bar{G}$ and $\tau_H = \min(\tau_F, \tau_G)$. Abbreviate τ_H to τ. Let $T = \max\{\tilde{X}_1, \cdots, \tilde{X}_n\}$, $J_n(t) = [t \leq T]$. Let R^p denote p-dimensional Euclidean space with $R = R^1$. Any v in R^p is a px1 matrix. $o_p(1)$ ($O_p(1)$) denotes any sequence of r.v.'s converging to zero in probability(bounded in probability). $\int_s^t = \int_{(s,t]}$ for $s > 0$ and $\int_0^t = \int_{[0,t]}$. The symbol "$\equiv$" means "is defined by".

2. PRELIMINARIES

In this chapter first we cite some frequently used results from analysis. Then we investigate the behaviors of some basic processes involved in the random censorship model.

Lemma 2.1. Let f_n, g_n, h_n, f, g and h be measurable functions on a measurable space Ω with measure μ.

(i) Suppose that f_n, g_n and h_n converge in μ measure to f, g and h respectively and they are all integrable. Also, $\int_\Omega g_n \, d\mu \longrightarrow \int_\Omega g \, d\mu$, $\int_\Omega h_n \, d\mu \longrightarrow \int_\Omega h \, d\mu$ and $g_n \leq f_n \leq h_n$ a.e.[μ]. Then $\int_\Omega f_n d\mu \longrightarrow \int_\Omega f \, d\mu$.

(ii) Suppose that $f_n \longrightarrow f$ in μ measure and $\|f_n\|_{L_p(\mu)} \longrightarrow \|f\|_{L_p(\mu)}$ where $p > 0$. Then $f_n \longrightarrow f$ in $L_p(\mu)$. □

For a version of the above results in which the convergence in measure is replaced by almost sure convergence, see Fabian and Hannan(1985, p.32) and Rudin(1974, p.76). From these references and a subsequence argument we can get our results here.

The following integration by parts formula has a proof in Hewitt and Stromberg(1965, p.419).

Lemma 2.2. Let U, V be of bounded variation on the real line, then for $a < b$,

$$U_+(b)V_+(b) - U_-(a)V_-(a) = \int_{[a,b]} U_- dV + \int_{[a,b]} V_+ dU. \quad \square$$

In this thesis we always assume that F and G do not have common points of discontinuity:

(2.1) $$\int_0^\infty (\Delta F)\, dG = 0 .$$

Let

$$H^1(t) = P[\delta_1 = 1, \tilde{X}_1 \le t] = \int_0^t \bar{G}_-\, dF,$$

$$H^0(t) = P[\delta_1 = 0, \tilde{X}_1 \le t] = \int_0^t \bar{F}\, dG$$

$$= \int_0^t \bar{F}_- dG , \quad \text{by (2.1)},$$

$$H_n(t) = n^{-1} \Sigma [\tilde{X}_i \le t],$$

$$H_n^1(t) = n^{-1} \Sigma [\tilde{X}_i \le t]\delta_i,$$

$$H_n^0(t) = n^{-1} \Sigma [\tilde{X}_i \le t](1 - \delta_i),$$

$$\Lambda^1(t) = \int_0^t (1/\bar{F}_-)\, dF = \int_0^t (1/\bar{H}_-)\, dH^1, \quad \text{for } t < \tau,$$

$$\Lambda^0(t) = \int_0^t (1/\bar{G}_-)\, dG = \int_0^t (1/\bar{H}_-)\, dH^0, \quad \text{for } t < \tau,$$

(2.2) $$Q_n^1(t) = n^{1/2}[H_n^1(t) - H^1(t)],$$

$$Q_n^0(t) = n^{1/2}[H_n^0(t) - H^0(t)],$$

$$Q_n = \begin{pmatrix} Q_n^1 \\ Q_n^0 \end{pmatrix},$$

for $i = 1,2,\ t < \tau$,

$$M_n^i(t) = n^{1/2}[H_n^i(t) - \int_0^t (1 - H_n)\, d\Lambda^i]$$

$$= Q_n^i(t) + \int_0^t (Q_{n-}^0 + Q_{n-}^1)/\bar{H}_-\, dH^i$$ (notice that M_n^1 is Gill(1980)'s M(t) as in his definition (3.2.4)),

$$M_n(t) = \begin{pmatrix} M_n^1(t) \\ M_n^0(t) \end{pmatrix}, \text{for } t < \tau.$$

The Kaplan - Meier (1958) product-limit estimators $\hat{F}_n$ and $\hat{G}_n$ of F and G are given as follows:

(2.3) $$1 - \hat{F}_n(t) \equiv \prod_{s \le t} [1 - \Delta H_n^1(s)/ \bar{H}_{n-}(s)],$$

$$1 - \hat{G}_n(t) \equiv \prod_{s \le t} [1 - \Delta H_n^0(s)/ \bar{H}_{n-}(s)].$$

The corresponding processes are

$$(2.4) \qquad P_n^1 \equiv n^{1/2}(\hat{F}_n - F),$$
$$P_n^0 \equiv n^{1/2}(\hat{G}_n - G), \quad \text{on } [0,\infty).$$

Shorack and Wellner (1986) (S-W) have a discussion on $\hat{F}_n$ in their Chapter 7. Notice that since $\int_0^\infty (\Delta F)\, dG = 0$, the roles of F, G are interchangeable, hence we have parallel results for $\hat{G}_n$. In fact, for the σ-field $\sigma_t^n = N \cup \sigma\{[Z_i \le s]\delta_i,\ [Z_i \le s]:\ 1 \le i \le n,\ 0 \le s \le t\}$, where N denotes the collection of all null sets and their commplements in the original probability space, one can check via Theorem 3.11 of Gill(1980) that $\{M_n(t),\ \sigma_t^n : 0 \le t < \tau\}$ is a 2-dimensional square integrable martingale with mean zero and the predictable variation processes

$$\langle M_n^i\rangle(t) = \int_0^t \bar{H}_{n-}(1 - \Delta\Lambda^i)\, d\Lambda^i \ ,\ i = 0,\ 1,$$

and that the predictable variation process for the martingale $M_n^0 + M_n^1$ is

$$\langle M_n^0 + M_n^1\rangle(t) = \int_0^t \bar{H}_{n-}(1 - \Delta\Lambda^0 - \Delta\Lambda^1)\, d(\Lambda^0 + \Lambda^1)$$
$$= \langle M_n^0\rangle(t) + \langle M_n^1\rangle(t),$$

since $\int_0^\infty (\Delta F)\, dG = 0$. So the predictable covariation process

$$\langle M_n^0,\ M_n^1\rangle(t) = 0.$$

Thus an argument similar to Theorem 4.2.1 of Gill(1980) shows that M_n converges in $D^2[0,\tau]$ to a limit process

$$(2.5) \qquad M = \begin{pmatrix} M^1 \\ M^0 \end{pmatrix},$$

where M^1, M^0 are independent zero mean Gaussian processes, each with independent increments and for i = 0,1, $\langle M^i\rangle(t) = \int_0^t \bar{H}_-(1 - \Delta\Lambda^i)\, d\Lambda^i$. Later on we need to consider the integration operation and therefore it is more convenient to use the uniform topology. We follow Pollard(1984)'s approach. Define the sup metric $\|\binom{X}{Y}\| = \|X\| + \|Y\|$ in $D^2[0,\tau]$. Equip $D^2[0,\tau]$ with the σ-field generated by open balls, or equivalently, by the projection maps $\{\pi_t : \binom{X}{Y} \longrightarrow \binom{X(t)}{Y(t)} \mid t \in R\}$. Denote this space by $(D^2[0,\tau], \|\cdot\|)$. Then each random element in this space is a measurable mapping from some probability space. We say W_n converges to W in $(D^2[0,\tau], \|\cdot\|)$ if $Ef(W_n) \longrightarrow Ef(W)$ for any bounded, continuous and measurable function f. Here the measurability refers to the open ball σ-field on $D^2[0,\tau]$ and the Borel σ-field on R. We have the following.

Theorem 2.1. Suppose that F and G do not have common points of discontinuity: $\int_0^\infty (\Delta F)\, dG = 0$. Then on a common probability space there exists a special construction of triangular array $\{(\tilde{X}_{ni}, \delta_{ni}),\ 1 \le i \le n,\ n = 1, 2, \cdots\}$, each row consisting of i.i.d. pairs with the same distribution as $(\tilde{X}_1, \delta_1)$, and a 2-dimensional Gaussian process

$$Q = \begin{pmatrix} V^1(H^1) \\ V^0(H^0) \end{pmatrix}$$

where V^1 and V^0 are Brownian bridges, with covariance $Cov(V^1(H^1), V^0(H^0)) = -H^0H^1$, such that

$$\|Q_n - Q\| \xrightarrow[a.s.]{} 0.$$

Remark. For the case when there is no censoring, S-W(1986) have the special construction of X_{ni}'s for the convergence of the ordinary empirical processes in their theorem 3.1.1 and Section 3.2. For the convergence of Q_n process, without the condition $\int_0^\infty (\Delta F)\, dG = 0$, they claim to have the special construction of X_{ni}'s and Y_{ni}'s by a minor variation of their ordinary theory of empirical processes, i.e., by "rebuilding" the random variables from the process. There is some difficulty in defining δ_{ni}'s if $\int_0^\infty (\Delta F)\, dG \neq 0$ and the construction of X_{ni}'s and Y_{ni}'s from Q_n does not seem obvious.

To prove Theorem 2.1, we need the following two lemmas. In Lemma 2.3 we use Billingsley(1968)'s techniques for fluctuation of partial sums to get an upper bound on the tail probability of the empirical process Q_n. In Lemma 2.4, we show that $(\tilde{X}_i, \delta_i)$'s can be "reconstructed from Q_n".

Lemma 2.3. For any a > b, any $\epsilon > 0$, there exists a constant K_ϵ, depending only on ϵ, such that for i = 0,1, and n exceeding some $n_0 = n_0(a,b,\epsilon) > 0$,

$$P[\sup_{t\in[a,b)} |Q_n^i(t) - Q_n^i(a)| > \epsilon] \leq K_\epsilon (H_-^i(b) - H^i(a))^2.$$

Lemma 2.4. Suppose that F and G do not have common points of discontinuity and that $\hat{Q}_n$ has the same distribution as Q_n. Then with probability 1 random variables $\{\tilde{X}_i^n, \delta_i^n, i = 1,\cdots, n\}$ can be constructed from $\hat{Q}_n$, with the same joint distribution as the ordered $\tilde{X}_{(i)}$'s and their corresponding δ_i's.

Proof of Theorem 2.1. First let us prove the convergence of Q_n to Q in $(D^2[0,\tau], \|\cdot\|)$. We rely on Theorem 5.3 in Pollard(1984). The change from D[0,1] to $D^2[0,\tau]$ causes no problem.

Just as in $D[0,\tau]$, from the fact that $D^2[0,\tau]$ is equipped with the σ-field generated by open balls in the uniform metric, we obtain that every point in $D^2[0,\tau]$ is "completely regular" as in Definition 4.6 of Pollard(1984). Therefore if we can show that with probability 1 the limit process Q lies in a separable subset of $D^2[0,\tau]$, then the necessary and sufficient conditions for the weak convergence of the processes Q_n to Q are their finite dimensional convergences and the "small oscillation" condition(cf. Pollard 5.1.(4)). The finite dimensional convergence part being straightforward, we prove the separability and small oscillation property. Also, we only look at the first coordinate Q_n^1.

Let $S = \{s_0, s_1, \cdots\}$ be the countable set of jump points of H^1 and H^0, and A be the set of all functions in $D[0,\tau]$ whose jumps occur only possibly at points in S. For any $x \in A$, any $\delta > 0$, by Lemma 14.1 of Billingsley(1968), there exists a partition $0 = t_0 < t_1 < \cdots < t_m = \tau$, such that for $I_i = [t_{i-1}, t_i)$,

$$\sup_{I_i} |x(t) - x(t_{i-1})| < \delta, \ i = 1, \cdots, m.$$

Now we modify the partition as follows. If x is continuous at t_i, replace t_i by a rational point $\tilde{t}_i$ close enough to t_i so that the above inequalities still hold. If x is discontinuous at t_i, then t_i must be in S. Therefore, the points in the partition can be chosen from a countable set. Let us denote this set by $U = \{u_0, u_1, \cdots\}$.

Let $B_n = \{x \in D[0,\tau]:$ x takes constant rational values on each interval $[t_{i-1}, t_i)$, where $0 = t_0 < t_1 < \cdots < t_m = \tau$ is a partition of $[0, \tau]$ from points in U, and $x(\tau)$ is also rational$\}$. Then $B = \cup B_n$ is countable and certainly dense in A. Thus A is separable and clearly we have $P[Q \in A] = 1$.

As to the small oscillation condition, for each $\epsilon, \delta > 0$, take a partition $0 = t_0 < \cdots < t_m = \tau$ such that

$$\sup_{I_i} |H^1(t) - H^1(t_{i-1})| < \delta, \ i = 1, \cdots, m.$$

By Lemma 2.3, for n exceeding some $n_0(t_{i-1}, t_i, \epsilon)$,

$$P[\sup_{I_i} |Q_n^1(t) - Q_n^1(t_{i-1})| > \epsilon]$$
$$\leq K_\epsilon(H_-^1(t_i) - H^1(t_{i-1}))^2.$$

Since there are only finitely many partition points, we obtain

$$\limsup_{n \to \infty} P[\max_i \sup_{I_i} |Q_n^1(t_i) - Q_n^1(t_{i-1})| > \epsilon]$$
$$\leq \limsup_{n \to \infty} \sum_i K_\epsilon(H_-^1(t_i) - H^1(t_{i-1}))^2$$
$$\leq K_\epsilon \ \delta.$$

Thus the proof for weak convergence is completed.

Since with probability 1 the limit process Q sits in a separable set of completely regular points, the representation Theorem 4.13 in Pollard(1984) guarantees that on a common probability space there exist versions $\hat{Q}_n$, $\hat{Q}$ of Q_n, Q, respectively, such that $\|\hat{Q}_n - \hat{Q}\| \xrightarrow[a.s.]{} 0$. Now we can obtain the special construction $\{(\tilde{X}_{ni}, \delta_{ni}): 1 \le i \le n\}$ from $\{(\tilde{X}_i^n, \delta_i^n), i = 1, \cdots, n\}$ in Lemma 2.4 through a random permutation. □

Proof of Lemma 2.3. Let us just look at the case for $n^{1/2}(Q_n^1 - Q^1)$. The proof is similar to the one used in the proof of Theorem 13.1 in Billingsley(1968). In place of the condition (13.17) there, one can verify that

$$\begin{aligned}
&E\{|Q_n^1(s+p_1) - Q_n^1(s)|^2 |Q_n^1(s+p_1+p_2) - Q_n^1(s+p_1)|^2\} \\
&\le \text{Constant } (H^1(s+p_1) - H^1(s)) \\
&\quad \cdot (H^1(s+p_1+p_2) - H^1(s+p_1)), \\
&\text{for } s,\ s+p_1,\ s+p_1+p_2 \in [a, b).
\end{aligned}$$

Thus for $r < b - a$, by considering the random variables $Q_n^1(a+(b-a-r)i/m) - Q_n^1(a+(b-a-r)(i-1)/m)$, $i = 1, \cdots, m$, we have, in place of (13.22) in Billlingsley(1968),

$$P[\sup_{[a,b-r]} |Q_n^1(t) - Q_n^1(a)| > \epsilon]$$

$$\le B_\epsilon (H^1(b-r) - H^1(a))^2 + P[|Q_n^1(b-r) - Q_n^1(a)| \ge \epsilon/2], \tag{2.6}$$

where B_ϵ is some constant depending only on ϵ. By triangle inequality,

$$P[|Q_n^1(b-r) - Q_n^1(a)| \ge \epsilon/2]$$

$$\le P[|Q_n^1(b-) - Q_n^1(a)| \ge \epsilon/4] + P[|Q_n^1(b-) - Q_n^1(b-r)| \ge \epsilon/4]. \tag{2.7}$$

Since the fourth central moment of binomial(n, p) does not exceed $8np + n(n-1)p^2$, and $P[|X| > \epsilon] \le \epsilon^{-4}E\{X^4\}$, we obtain

$$P[|Q_n^1(b-) - Q_n^1(b-r)| \ge \epsilon/4]$$

$$\le (4/\epsilon)^4 E\{Q_n^1(b-) - Q_n^1(b-r)\}^4$$

$$\le (4/\epsilon)^4\{8n^{-1}(H^1(b-) - H^1(b-r)) + n^{-1}(n-1)(H^1(b-) - H^1(b-r))^2\}. \tag{2.8}$$

Substitute (2.7), (2.8) in (2.6) and then take the limit as $r \downarrow 0$, we have

$$P[\sup_{[a,b)} |Q_n^1(t) - Q_n^1(a)| > \epsilon]$$

$$\le B_\epsilon(H^1(b-) - H^1(a))^2 + P[|Q_n^1(b-) - Q_n^1(a)| \ge \epsilon/4].$$

Now for $\sigma = H^1(b-) - H^1(a)$, as $n \longrightarrow \infty$, the Central Limit Theorem gives us

$$P[|Q_n^1(b-) - Q_n^1(a)| \ge \epsilon/4]$$

$$\longrightarrow P[\{\sigma(1-\sigma)\}^{1/2}|N(0,1)| \ge \epsilon/4]$$

$$\le (4/\epsilon)^4\sigma^2 E\{N(0,1)\}^4 \le B_\epsilon(H^1(b-) - H^1(a))^2$$

if we take the constant B_ϵ large enough. Thus for n exceeding some $n_0 = n_0(a, b, \epsilon)$,

$$P[\sup_{[a,b)} |Q_n^1(t) - Q_n^1(a)| > \epsilon]$$

$$\le K_\epsilon(H^1(b-) - H^1(a))^2$$

for some constant K_ϵ depending only on ϵ. □

Proof of Lemma 2.4. First note that since $\int_0^\infty (\Delta F)dG = 0$, when there are ties among $\tilde{X}_i$'s, the corresponding δ_i's must be all 0's or all 1's: for $i \neq j$,

$$P[\tilde{X}_i = \tilde{X}_j, \delta_i = 0, \delta_j = 1] \le P[Y_i = X_j] = 0.$$

So for $W_n = \begin{pmatrix} W_n^1 \\ W_n^0 \end{pmatrix}$, where $W_n^i = Q_n^i - n^{1/2}H^i$ for $i = 0, 1$, with probability 1, we have

(2.9) $(\pi_t - \pi_s)W_n \in \{(in^{-1/2},\ jn^{-1/2}) : i,\ j \geq 0,\ i + j \leq n\},$

$$\Delta_t W_n = \lim_{m \to \infty} (\pi_t - \pi_{t-1/m})W_n$$

$$\in \{(0,\ kn^{-1/2}), (kn^{-1/2},\ 0) : 0 \leq k \leq n\},$$

$$\pi_1 W_n \in \{(in^{-1/2},\ jn^{-1/2}) : i,\ j \geq 0,\ i + j = n\}.$$

i.e., with probability 1, W_n^0, W_n^1 are increasing, taking constant values on consecutive intervals, and jumping at different points. also, the total number of jump points is n, if we count k jump points when $\Delta W^i = kn^{-1/2}$, $1 \leq k \leq n$, for $i = 0$ or 1. By our assumption the same is true for $\hat{W}_n = \hat{Q}_n - n^{1/2}\binom{H^1}{H^0}$. Denote the ordered jump points of $\hat{W}_n$ by $\tilde{X}_1^n, \cdots, \tilde{X}_n^n$ and define the corresponding δ_i^n's to be 1 or 0 according to whether $\hat{W}_n^1$ or $\hat{W}_n^0$ jumps at $\tilde{X}_i^n$. Then by (2.9) the joint distribution of $(\tilde{X}_i^n,\ \delta_i^n)$, $i = 1, \cdots, n$, can be determined by the projection π_t's acting on $\hat{W}_n$. In fact, it is the joint distribution of the ordered $\tilde{X}_i$'s and the corresponding δ_i's. □

By Theorem 2.1, we can adapt Theorem 7.1.1 in S-W to show

(2.10) $$\sup_{[0,\tau]} |M_n - M| \xrightarrow[\text{a.s.}]{} 0$$

for the special construction of $\tilde{X}_{ni}$, δ_{ni}, $i = 1, \cdots, n$, and a 2-dimensional Gaussian process as in (2.5). Now for $0 \leq t < \tau$, let

$$P^1(t) = \bar{F}(t) \int_0^t \bar{F}\_/(\bar{F}\bar{H}\_)\ dM^1,$$

$$P^0(t) = \bar{G}(t) \int_0^t \bar{G}\_/(\bar{G}\bar{H}\_)\ dM^0.$$

Then $P^1/\bar{F}$, $P^0/\bar{G}$ are martingales on $[0, \tau)$ with quardratic variation processes

$$C_1(t) = \int_0^t [\bar{F}_-/(\bar{F}\bar{H}_-)]^2 \bar{H}_-(1 - \Delta\Lambda^1)\, d\Lambda^1$$
$$= \int_0^t (\bar{F}\bar{G}_-)^{-1}\, d\Lambda^1,$$
$$C_0(t) = \int_0^t (\bar{G}\bar{F}_-)^{-1}\, d\Lambda^0,$$

respectively. Now we can obtain the following convergence results for P_n^1 and P_n^0.

Theorem 2.2. Suppose that F and G do not have common points of discontinuity.

(i) if G is continuous at τ, then for the special construction in Theorem 2.1 and any $\alpha \in (0,1/2)$,

(2.11) $$\sup_{t \in [0,\, T]} |[P_n^0 - P^0](t)\, \bar{F}^{1-\alpha}(t)| \xrightarrow[P]{} 0.$$

(ii) if F is continuous at τ and

(2.12) $$A(\tau) < \infty, \text{ where } A(t) \equiv \int_0^t (1/\bar{G}_-)\, dF,$$

then for the special construction we also have

(2.13) $$\sup_{t \in [0,\tau)} |[P_n^1 - P^1](t)| \xrightarrow[P]{} 0.$$

Proof. Since the roles of F and G are interchangeable, from (2.10) an argument similar to (9) of Theorem 7.7.1 in S-W shows that, if $\Delta G(\tau) = 0$, we have

(2.14) $$\sup_{[0,T]} |(P_n^0 - P^0)\, \bar{K}_0[\bar{G}\, q(K_0)]^{-1}| = o_p(1),$$

for $K_0(t) = \dfrac{C_0(t)}{1 + C_0(t)}$ and any function q such that, on $(0, 1/2]$ q is $\uparrow$ and $t^{-1/2}q(t)$ is $\downarrow$, q is symmetric about

$t = 1/2$, and $\int_0^1 q^{-2}(t)\,dt < \infty$. Notice that in their proof, S–W use their Theorem 7.4.2, which states the uniform convergence of P_n^1 to P^1 on each compact subinterval $[0,\rho]$ of $[0,\tau)$, $\rho < \tau$. One way to avoid some flaws in their argument is to restrict ρ to be a continuity point of H. Since H has only countably many discontinuity points, this restriction causes no problem in deriving their Theorem 7.7.1. Now Take $q = t^\alpha$ on $(0,1/2]$, where $\alpha \in (0,1/2)$. Notice that

$G(t)[\bar{G}(t)]^{-1} = \int_0^t (\bar{G}\bar{G}_-)^{-1}dG \le \int_0^t(\bar{G}\bar{G}_-\bar{F}_-)^{-1}\,dG = C_0(t)$

$\le [\bar{F}(t)]^{-1} \int_0^t (\bar{G}\bar{G}_-)^{-1}dG = G(t)[\bar{H}(t)]^{-1}$. So

$\bar{G} \ge \bar{K}_0 = (1 + D_0)^{-1} \ge \bar{H}$. Hence we have

$$\bar{K}_0[\bar{G}\ q(K_0)]^{-1} \ge (\bar{G})^{-1}\ \bar{K}_0^{\ 1-\alpha} \ge (\bar{G})^{-1}\ \bar{H}^{\ 1-\alpha} \ge \bar{F}^{\ 1-\alpha}.$$

Thus (2.11) follows from the above inequality and (2.14).

To prove (2.13), first note that if $\Delta F(\tau) = 0$, we can use the same method as in the proof of (9) in Th.7.7.1 of S-W, the role of $\bar{K}_1/q(K_1)$ being replaced by $\bar{F}$ and the integrability condition on q being replaced by (2.12), to obtain

$$\sup_{[0,T]} |P_n^1 - P^1| = o_p(1). \tag{2.15}$$

Now observe that by triangle inequality

$$\begin{aligned} &\sup_{[T,\ \tau)} |P_n^1 - P^1| \\ &\le |P_n^1\ (T) - P^1(T)| + n^{1/2}(F(\tau-) - F(T)) \\ &\quad + \sup_{[T,\ \tau)} |P^1(T) - P^1(t)|. \end{aligned} \tag{2.16}$$

The first term is $o_p(1)$ by (2.15). As for the last term, by the continuity of F (hence C^1) at τ and the fact $T \longrightarrow \tau$, it is $o_p(1)$ when $C^1(\tau) < \infty$; it is also $o_p(1)$ when $C^1(\tau) = \infty$, since by Remark 2.2 in Gill(1980) (the results (2.4),(2.5) in that remark and the inequalities following them), $\sup_{[\rho,\tau)} |P^1(t)| \cdot = o_p(1)$ as $\rho \uparrow \tau$. Hence it only remains to show $n^{1/2}(F(\tau-) - F(T)) = o_p(1)$. Note that for the function A as defined in (2.12), for all $t < \tau$,

$$(2.17) \qquad F(\tau-) - F(t) = \int_{(t,\tau)} \bar{G}_- \, dA \le \bar{G}(t)\int_{(t,\tau)} dA.$$

Now substitute t = T, multiply $n(F(\tau-) - F(T))$ through to get

$$(2.18) \qquad n^{1/2}(F(\tau-) - F(T)) \le \{n\bar{H}(T)\int_{(T,\tau)} dA\}^{1/2}.$$

So it suffices to show that $n\bar{H}(T)$ is $O_p(1)$, since A is bounded and $T \longrightarrow \tau$ w.p.1. Using $H\,H^{-1}(x) \le x + \Delta H(H^{-1}(x))$, for $x \in [0,1]$, we have

$$\begin{aligned} &P[n\bar{H}(T) > t] \\ &= P[H(T) < 1 - t/n] = [H_-H^{-1}(1 - t/n)]^n \\ &= [H\,H^{-1}(1 - t/n) - \Delta H(H^{-1}(1 - t/n))]^n \\ &\le [1 - t/n]^n \longrightarrow e^{-t} \,. \end{aligned}$$

which implies $n\bar{H}(T) = O_p(1)$. Thus $n^{1/2}(F(\tau-) - F(T)) = o_p(1)$. So finally we have (2.13). □

Remark. For later use we make the following observations. For t close to τ, $\bar{F}$, $\bar{F}^{1-\alpha}\,\bar{G}$ are nonnegative, right continuous and nonincreasing. Also,

$$\int \bar{F}^2 \, dD_1 \le \int_0^\tau (1/\bar{G}_-) \, dF < \infty,$$

$$\int (\bar{F}^{1-\alpha}\,\bar{G})^2 \, dD_0 \le \int_0^\tau \bar{F}^{1-2\alpha} \, dG \le 1, \text{ for } \alpha \in (0,1/2).$$

Hence, when $C^1(\tau) = \infty$, as illustrated in Remark 2.2 of Gill(1983),

$$P^1(t) \longrightarrow 0 \quad \text{w.p.1} \quad \text{as } t \longrightarrow \tau,$$
$$P^0(t)\bar{F}^{1-\alpha}(t) \longrightarrow 0 \quad \text{w.p.1} \quad \text{as } t \longrightarrow \tau.$$

So from Theorem 2.2 we have

(2.19) $$\sup_{[0,\tau)} |P^1| < \infty \quad \text{w.p.1},$$
$$\sup_n \sup_{[0,\tau)} |P_n^1| = O_p(1)$$
and for any $\alpha \in (0,1/2)$,
$$\sup_{[0,\tau)} |P^0\, \bar{F}^{1-\alpha}| < \infty, \text{ w.p.1},$$
$$\sup_n \sup_{[0,T)} |P_n^0\, \bar{F}^{1-\alpha}| = O_p(1),$$
$$P_n^0\, \bar{F}^{1-\alpha}(T) = o_p(1).$$

To obtain the main results of Chapter 4(Theorem 4.2, 4.3), we need to control the increment of the process
$$Z_n = n^{1/2}(\hat{F}_n - F)/\bar{F}.$$
For any process X, let $X^T(t) = [t \le T]X(t) + [t > T]X(T)$. Also recall the function A defined in (2.12). The following lemma gives an inequality for the mean square increment of the Z_n^T process in terms of the function A.

Lemma 2.5. Suppose F is continuous. Then for $0 \le s < t < \tau_F$,
$$E[Z_n^T(t) - Z_n^T(s)]^2 \le 4\,[A(t)/\bar{F}^2(t) - A(s)/\bar{F}^2(s)]$$
$$\le 4\,\bar{F}^{-2}(t)[A(t) - A(s)] + 4\,A(s)[\bar{F}^{-2}(t) - \bar{F}^{-2}(s)].$$

Proof. Lemma 2.5 in Gill (1983) states that for continuous F, Z_n^T is a square integrable martingale on $[0, \rho]$

for any $\rho < \tau_F$, with predictable variation process

(2.20) $\quad \langle Z_n^T \rangle(x) = \int_0^x [(1 - \hat{F}_{n-})/\bar{F}]^2 \, J_n/\bar{H}_{n-} \, d\Lambda^1$

for $x \in [0, \rho]$. Let $L(x) = E\langle Z_n^T \rangle(x)$. Then

$$L(x) = E\int_0^x [1 - n^{-1/2} Z_{n-}^T]^2 \, J_n/\bar{H}_{n-} \, d\Lambda^1$$

$$\le 2 \int_0^x E(J_n/\bar{H}_{n-}) \, d\Lambda^1 + 2 \int_0^x E[Z_{n-}^T]^2 J_n d\Lambda^1 .$$

Since $\bar{H}_{n-}(x)$ is binomial and for $1 \le k \le n$, $n/k \le (n+1)/(k+1)$, we have

$$E(J_n/\bar{H}_{n-})(x) = \sum_{k=1}^{n} n/k \binom{n}{k} \bar{H}_-^k \, H_-^{n-k}(x)$$

$$\le 2 \sum_{1}^{n} (n+1)/(k+1) \binom{n}{k} \bar{H}_-^k \, H_-^{n-k}(x) \le 2 \, (\bar{H}_-(x))^{-1} .$$

Also by Fatou's Lemma,

$$E[Z_{n-}^T(x)]^2 \, J_n(x) \le E[Z_{n-}^T(x)]^2 = E \lim_{x_k \uparrow x} [Z_n^T(x_k)]^2$$

$$\le \underline{\lim}_{x_k \uparrow x} E[Z_n^T(x_k)]^2 = \underline{\lim}_{x_k \uparrow x} E\langle Z_n^T \rangle(x_k) \le L(x),$$

since the integrand in (2.20) is nonnegative. Therefore, we obtain

(2.21) $\quad L(x) \le \alpha(x) + \int_0^x L \, d\beta$,

where $\alpha(x) = 4 \int_0^x (1/\bar{H}_-) \, d\Lambda^1$, $\beta = 2 \Lambda^1$.

Now we use the argument as in Gronwall's Lemma: iterating m times in (2.21) yields

$$L(x) \le \alpha(x) + \int_0^x \sum_{i=0}^{m-1} 1/i! \, [\beta(x) - \beta(y)]^i \, \alpha(y) \, d\beta(y)$$

$$+ \int_0^x 1/m! \, [\beta(x) - \beta(y)]^m \, L(y) \, d\beta(y)$$

$$\le \alpha(x) + \int_0^x e^{\beta(x)-\beta(y)} \, \alpha(y) \, d\beta(y)$$

$$+ 1/(m+1)! \, \beta^{m+1}(x) \, L(x).$$

Let $m \longrightarrow \infty$. Then

$$L(x) \le \alpha(x) + \int_0^x e^{\beta(x) - \beta(y)} \, \alpha(y) \, d\beta(y)$$

$$= \alpha(x) + 2 \int_0^x e^{2\ln(\bar{F}(y)/\bar{F}(x))} \alpha(y)\, d\Lambda^1(y)$$

$$= \alpha(x)$$

$$+ 8 \int_0^x [\bar{F}(y)/\bar{F}(x)]^2 \int_0^y (1/\bar{H}\_(u))\, d\Lambda^1(u)\, d\Lambda^1(y)$$

$$= 4\, \bar{F}^{-2}(x)\, A(x) \tag{2.22}$$

by Fubini's theorem.

Now $E[Z_n^T(t) - Z_n^T(s)]^2$

$$= E[\langle Z_n^T\rangle(t) - \langle Z_n^T\rangle(s)]$$

$$\le 2 \int_s^t E(J/\bar{H}_{n-})\, d\Lambda^1 + 2 \int_s^t E[Z_{n-}^T]^2\, d\Lambda^1$$

$$\le 4 \int_s^t (1/\bar{H}\_)\, d\Lambda^1 + 2 \int_s^t L\, d\Lambda^1$$

$$\le 4\, [A(t)/\bar{F}^2(t) - A(s)/\bar{F}^2(s)]$$

by (2.22) and the Fubini's theorem. □

Let us now define the kernel density estimators and smoothed version of the product-limit estimators as follows:

$$f_n(x) = a_n^{-1} \int_R K((x-y)/a_n)\, d\hat{F}_n(y), \tag{2.23}$$

$$F_n(x) = \int_{-\infty}^{x} f_n(s)\, ds,$$

where K is some kernel function and a_n is some positive constant. The following theorem establishes the convergence of f_n to f in the Hellinger metric.

Theorem 2.3. Suppose $\int_0^\tau (1/\bar{G}\_)\, dF < \infty$ and G is continuous at τ. Also suppose that F has a continuous density f, K is nonnegative, continuous and of bounded variation on R, $\int_R K(s)\, ds = 1$, $K(s) \longrightarrow 0$ as $s \longrightarrow -\infty$, $a_n \longrightarrow 0$ and $n^{1/2} a_n \longrightarrow \infty$. Then

$$\| f_n^{1/2} - f^{1/2} \|_2 \xrightarrow[P]{} 0. \tag{2.24}$$

where $\|\cdot\|_2$ denotes the L^2- norm with respect to the Lebesgue measure.

Proof. Integration by parts gives us

$$(2.25)\qquad f_n(x) - f(x) = (n^{1/2}a_n)^{-1} \int_R P_n^1(x - a_n t)\, dK(t) + \int_R [f(x - a_n t) - f(x)]\, K(t)\, dt = R_3 + R_4\ , \quad \text{say.}$$

By (2.19) we have

$$(2.26)\qquad \sup_{x \in R} |R_3| \le (n^{1/2}a_n)^{-1} \sup_{[0,\tau)} |P_n^1|\ |\int_R dK| = o_p(1).$$

Since $\int_R \int_R f(x - a_n t)\, K(t)\, dx\, dt = \int_R \int_R f(x)\, K(t)\, dx\, dt = 1$, by Lemma 2.1 (ii) the nonrandom term R_4 converges to zero in Lebesgue measure. Also

$$\|f_n^{1/2}\|_2^2 = \hat{F}_n(T) = o_p(1) + F(T) \xrightarrow[P]{} 1 = \|f^{1/2}\|_2^2.$$

Now to obtain (2.24), we use a subsequence argument. For any subsequence $\{n'\} \subset \{n\}$, there exists a further subsequence $\{n''\} \subset \{n'\}$ alone which

(a) $R_4 \longrightarrow 0$ for a.e.[λ]-x,

(b) $\sup_{x \in R} |R_3| \longrightarrow 0$ w.p.1,

(c) $\|f_{n''}^{1/2}\|_2 \longrightarrow \|f^{1/2}\|_2$ w.p.1.

By (a) and (b), with probability 1, $f_{n''}(x) \longrightarrow f(x)$ for a.e.[λ]-x. By Lemma 2.1 (ii), this and (c) imply that with probability 1, $\|f_{n''}^{1/2} - f^{1/2}\|_2 \longrightarrow 0$. Hence (2.24) follows. □

3. MINIMUM HELLINGER DISTANCE FUNCTIONALS

In the censored data case, the pair $(\tilde{X}, \delta)$ is observed. Assuming F has a density f with respect to the Lebesgue measure λ, we have

$$P[\delta_1 = 0, \tilde{X}_1 \leq t] = \int_0^t \bar{F}\, dG,$$

$$P[\delta_1 = 1, \tilde{X}_1 \leq t] = \int_0^t \bar{G}_- dF = \int_0^t f\, \bar{G}\, d\lambda.$$

Thus $(\tilde{X}, \delta)$ has a (sub-)density $\bar{F}^{1-y}(x) f^y(x)$ with respect to the measure μ_G on $R \times \{0,1\}$, where μ_G is defined by the relation

$$\int m\, d\mu_G = \int m(x, 0)\, dG(x) + \int m(x, 1)\, \bar{G}\, dx,$$

for any nonnegative measurable function m on $R \times \{0,1\}$.

For any (sub-)density d on R w.r.t. λ define a (sub-)density L(d) on $R \times \{0,1\}$ w.r.t. μ_G by

$$L(d)(x,y) = \bar{D}^{1-y}(x)\, d^y(x),$$

where D is the (sub-)c.d.f. of d. Recall the parametric family $\{f_\theta\colon \theta \in \Theta\}$ as mentioned in the introduction. For (sub-)c.d.f. G and (sub-)density function d, the minimum Hellinger distance functional $\Psi(d; G)$ is defined as a point in Θ, if exists, that minimizes the Hellinger distance between $L(f_\theta)$ and $L(d)$:

$$\text{(3.1)} \qquad \| [L(f_{\Psi(d;G)})]^{1/2} - [L(d)]^{1/2} \|_G = \inf_{\theta \in \Theta} \| [L(f_\theta)]^{1/2} - [L(d)]^{1/2} \|_G ,$$

where $\|\cdot\|_G$ denotes the L^2– norm in $L^2(\mu_G)$. In the case when there are more than one minimizer, a Borel measurable selection is possible(c.f. Brown and Purves(1973)). For

$0 \leq \tau \leq \infty$, define $\Psi(\cdot;\ G;\ \tau)$ similarly by restricting all integration to $x \in (-\infty, \tau]$. Later we will use $\|\{\cdot\}(-\infty,\ \tau]\|_G$ to denote the norm under the restricted integration. Note that $\Psi(\cdot;\ G;\ \infty) = \Psi(\cdot;\ G)$. For (sub-)densities f, f_n on R w.r.t. λ and c.d.f. G_n, τ^n such that

(3.2) $$\sup_{(-\infty,\tau]} |G_n - G| \longrightarrow 0 \quad \text{and} \quad \tau^n \uparrow \tau,$$

we will use notations $\theta_0 = \Psi(f;\ G;\ \tau)$, $\theta_{00} = \Psi(f;\ G)$, $\theta_n = \Psi(f;\ G_n;\ \tau)$, $\theta_{nn} = \Psi(f;\ G_n;\ \tau^n)$ and $\mu = \mu_G$, $\mu_n = \mu_{G_n}$. Also, we will use $\int_{-\infty}^{\tau} \cdot d\mu$ to denote the integral on $(x,y) \in (-\infty,\tau] \times \{0,1\}$ and F_θ the c.d.f. corresponding to f_θ. We have the following.

Lemma 3.1. Suppose

(a) Θ is a compact subset of R^p,

(b) $\bar{\theta} \neq \theta$ implies $f_{\bar{\theta}} \neq f_\theta$ on a set of positive Lebesgue measure, and for almost every x, $f_\theta(x)$ is continuous in θ.

Then,

(i) for any (sub-)c.d.f. G, (sub-)density function f and $0 \leq \tau \leq \infty$, $\Psi(f;\ G;\ \tau)$ exists,

(ii) $\Psi(f_\theta;\ G;\ \tau) = \theta$ uniquely if both τ_G and $\tau \geq \tau_{F_\theta}$.

(iii) for any (sub-)c.d.f. G, any (sub-)density functions f, f_n on R, $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow 0$ implies $\|[L(f_n)]^{1/2} - [L(f)]^{1/2}\|_G \longrightarrow 0$.

(iv) $\|[L(f_n)]^{1/2} - [L(f)]^{1/2}\|_G \longrightarrow 0$ implies $\Psi(f_n;\ G) \longrightarrow \Psi(f;\ G)$ if $\Psi(f;\ G)$ is unique.

If, in addition,

(c) the family $\{F_\theta(x): \theta \in \Theta\}$ is equicontinuous, then

(v) for G_n, τ^n, G and τ satisfying (3.2) and $\Delta G(\tau) = 0$, $\|[L(f_n)]^{1/2} - [L(f)]^{1/2}\|_G \longrightarrow 0$ implies $\Psi(f_n; G; \tau^n) \longrightarrow \Psi(f; G; \tau)$ if $\Psi(f; G; \tau)$ is unique.

Proof. By (b) $F_\theta(x)$ is continuous in θ for fixed x, thus (i) can be proved as in Theorem 1 of Beran (b,1977). (ii) is obvious. For (iii), first note that for a, b $\geq$ 0, $|b - a| \leq b + a$, hence $(b - a)^2 \leq |b^2 - a^2|$. So we have

$$\|[L(f_n)]^{1/2} - [L(f)]^{1/2}\|_G^2$$
$$= \int_R [\bar{F}_n^{1/2} - \bar{F}^{1/2}]^2 \, dG + \int_R [f_n^{1/2} - f^{1/2}]^2 \bar{G} \, d\lambda$$
$$\leq \int_R |F_n - F| \, dG + \int_R [f_n^{1/2} - f^{1/2}]^2 \bar{G} \, d\lambda.$$

Since $\int_R |F_n - F| \, dG \leq \sup_R |F_n - F| \leq \int_R |f_n - f| \, d\lambda = \int_R |f_n^{1/2} - f^{1/2}|(f_n^{1/2} + f^{1/2}) \, d\lambda$, an application of Cauchy-Schwartz's inequality to the last integral gives us $\int_R |F_n - F| \, dG \leq 2\{\int_R [f_n^{1/2} - f^{1/2}]^2 \, d\lambda\}^{1/2}$. Hence we obtain

$$\|[L(f_n)]^{1/2} - [L(f)]^{1/2}\|_G^2$$
$$\leq 2\{\int_R [f_n^{1/2} - f^{1/2}]^2 \, d\lambda\}^{1/2}$$
$$+ \int_R [f_n^{1/2} - f^{1/2}]^2 \bar{G} \, d\lambda.$$

This proves (iii). Next we prove only (v) as the assertion (iv) follows similar to (v).

Define N, $N_n > 0$ by

$$N(\theta, f) = \|\{[L(f_\theta)]^{1/2} - [L(f)]^{1/2}\}(-\infty, \tau]\|_G ,$$
$$N_n(\theta, f) = \|\{[L(f_\theta)]^{1/2} - [L(f)]^{1/2}\}(-\infty, \tau^n]\|_{G_n} .$$

By the triangle inequality, we have

$$|N_n(\theta, f_n) - N_n(\theta, f)|^2$$
$$\leq \|\{[L(f_n)]^{1/2} - [L(f)]^{1/2}\}(-\infty, \gamma^n]\|^2_{G_n}$$
$$\leq \|[L(f_n)]^{1/2} - [L(f)]^{1/2}\|^2_G$$
$$+ \left|\int_{-\infty}^{\gamma^n} [f_n^{1/2} - f^{1/2}]^2 (\bar{G}_n - \bar{G})\, d\lambda\right|$$
$$+ \left|\int_{-\infty}^{\gamma^n} [\bar{F}_n^{1/2} - \bar{F}^{1/2}]^2\, d(G_n - G)\right|.$$

The first term converges to zero uniformly in θ as $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow 0$, and so does the second term for G_n, γ^n, G and γ satisfying (3.2), since it is dominated by

$\sup_{(-\infty,\gamma^n]} |\bar{G}_n - \bar{G}| \int_{-\infty}^{\gamma^n} [f_n^{1/2} - f^{1/2}]^2\, d\lambda \leq 2 \sup_{(-\infty,\gamma]} |\bar{G}_n - \bar{G}|$. Now using integration by parts formula, we can write the third term as $|-\int_{-\infty}^{\gamma^n} (G_{n-} - G_-)\, d(\bar{F}_n + \bar{F} - 2(\bar{F}_n\bar{F})^{1/2})$ $+ [\bar{F}_n^{1/2} - \bar{F}^{1/2}]^2(\gamma^n)(G_n - G)(\gamma^n)|$, which can be dominated by

$5 \sup_{(-\infty,\gamma]} |G_n - G| \longrightarrow 0$. Thus we have shown that, for G_n, γ^n, G and γ satisfying (3.2),

(3.3) $$N_n(\theta, f_n) - N_n(\theta, f) \longrightarrow 0$$

uniformly in θ as $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow 0$.

By the triangle inequality again we have

$$|N_n^2(\theta, f) - N^2(\theta, f)|$$
$$\leq \int_{-\infty}^{\gamma^n} [f_\theta^{1/2} - f^{1/2}]^2 |\bar{G}_n - \bar{G}|\, d\lambda$$
$$+ \left|\int_{-\infty}^{\gamma^n} [\bar{F}_\theta^{1/2} - \bar{F}^{1/2}]^2\, d(G_n - G)\right|$$
$$+ \|\{[L(f_\theta)]^{1/2} - [L(f)]^{1/2}\}(\gamma^n, \gamma]\|^2_G.$$

Similar to the previous argument, the first term is bounded by $2 \sup_{(-\infty,\gamma]} |G_n - G|$ and the second term by $5 \sup_{(-\infty,\gamma]} |G_n - G|$.

The third term is bounded by

$|(F_\theta + F + G)(\tau) - (F_\theta + F + G)(\tau^n)|$. Hence for G_n, τ^n, G and τ satisfying (3.2), (c) and the assumption $\Delta G(\tau) = 0$,

(3.4) $$N_n^2(\theta, f) - N^2(\theta, f) \longrightarrow 0$$

uniformly in θ as $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow 0$.

From (3.3), (3,4) and again the inequality $(b - a)^2 \leq |b^2 - a^2|$ for a, b ≥ 0, it follows that $N_n(\theta, f_n) - N(\theta, f) \longrightarrow 0$ uniformly in θ, which implies $N_n(\theta_{nn}, f_n) - N(\theta_0, f) = \min_\theta N_n(\theta, f_n) - \min_\theta N(\theta, f) \longrightarrow 0$, and $N_n(\theta_{nn}, f_n) - N(\theta_{nn}, f) \longrightarrow 0$. Hence

(3.5) $$N(\theta_{nn}, f\ ) - N(\theta_0, f) \longrightarrow 0.$$

As in Beran(b,1977), from (3.5), compactness of Θ, continuity of $N(\theta,f)$ in θ and uniqueness of $\Psi(f;\ G)$, one has $\theta_{nn} \longrightarrow \theta_0$. i.e. $\Psi(f_n;\ G;\ \tau^n) \longrightarrow \Psi(f;\ G;\ \tau)$. □

To study the asymptotic behavior of the minimum Hellinger distance functionals, we need to establish the following expansions for $s_\theta \equiv [L(f_\theta)]^{1/2}$. When the first order partial derivatives of f_θ w.r.t. θ exist, we will denote the column vector of the partials by $\dot{f}_\theta$ with i^{th} component $\dot{f}_\theta^{(i)}$; when the matrix of the second order partials exists, it will be denoted by $\ddot{f}_\theta$ with (i,j) entry $\ddot{f}_\theta^{(ij)}$. Also, A^t denotes the transpose of the matrix A.

Lemma 3.2. Let ρ be an interior point of Θ. Suppose that there exists a neighborhood V of ρ such that

(i) on V, f_θ is continuous in θ for every x and $\dot{f}_\theta(x)$

is contniuous in θ for $x \notin N$, where N is a λ-null set,

(ii) for $i = 1, \cdots, p$, $U^{(i)}(\theta) \equiv \int [\dot{f}_\theta^{(i)}]^2/f_\theta \, d\lambda$ is continuous on V.

Then for ρ_n in a neighborhood of ρ and $(x,y) \notin N_0 x\{1\}$, where N_0 is a λ-null set,

$$s_{\rho_n} = s_\rho + (\dot{s}_\rho^{\,t} + r_n^t)(\rho_n - \rho), \tag{3.6}$$

where

$$\| r_n^{(i)}(-\infty, \tau] \|_G \longrightarrow 0 \quad \text{as } \rho_n \longrightarrow \rho \tag{3.7}$$

for $i = 1, \cdots, p$, $\tau \in R$, and any (sub-)c.d.f. G.

If, in addition, we assume

(iii) for $i = 1, \cdots, p$, and some $\epsilon > 0$,

$V^{(i)}(\theta) \equiv \int |\dot{f}_\theta^{(i)}|^{2+\epsilon}/f_\theta^{1+\epsilon} \, d\lambda$ and

$W^{(i)}(\theta) \equiv \int - |\dot{f}_\theta^{(i)}|/f_\theta \, d\bar{F}_\theta^{1/2}$ are bounded in a neighborhood of ρ. Then

$$\| r_n^{(i)}(-\infty, \tau^n] \|_{G_n} \longrightarrow 0 \quad \text{as } \rho_n \longrightarrow \rho \tag{3.8}$$

for each $i = 1, \cdots, p$ and G_n, τ^n satisfying (3.2).

Lemma 3.3. Let ρ be an interior point of Θ. Suppose that there exists a neighborhood V of ρ such that

(i) on V, f_θ is continuous in θ for every x and $\ddot{f}_\theta(x)$ is contniuous in θ for $x \notin N$, where N is a λ-null set,

(ii) for $i, j = 1, \cdots, p$, $U_1^{(i)}(\theta) \equiv \int [\dot{f}_\theta^{(i)}]^4/f_\theta^3 \, d\lambda$, $U_2^{(ij)}(\theta) \equiv \int [\ddot{f}_\theta^{(ij)}]^2/f_\theta \, d\lambda$ are continuous on V.

Then for ρ_n in a neighborhood of ρ and $(x,y) \notin N_0 \times \{1\}$, where N_0 is a λ-null set, (3.9) $$\dot{s}_{\rho_n} = \dot{s}_\rho + (\ddot{s}_\rho + R_n)(\rho_n - \rho),$$

where

(3.10) $$\|R_n^{(i,j)}(-\infty,\tau]\|_G \longrightarrow 0 \quad \text{as } \rho_n \longrightarrow \rho$$

for i, j = 1,⋯, p, $\tau \in R$ and any (sub-)c.d.f. G.

If, in addition, we assume ,

(iii) for i, j = 1,⋯, p, and some ϵ, $\delta > 0$

$$V_1^{(ij)}(\theta) \equiv \int |\ddot{f}_\theta^{(ij)}|^{2+\epsilon}/f_\theta^{1+\epsilon}\, d\lambda,$$

$$V_2^{(i)}(\theta) \equiv \int |\dot{f}_\theta^{(i)}|^{4+\delta}/f_\theta^{3+\delta}\, d\lambda,$$

$$W_1^{(ij)}(\theta) \equiv \int - |\ddot{f}_\theta^{(ij)}|/f_\theta\, d\bar{F}_\theta^{1/2} \text{ and}$$

$$W_2^{(i)}(\theta) \equiv \int - |\dot{f}_\theta^{(i)}|/f_\theta\, d\bar{F}_\theta^{1/4}$$

are bounded in a neighborhood of ρ, then

(3.11) $$\|R_n^{(i,j)}(-\infty,\tau^n]\|_{G_n} \longrightarrow 0 \quad \text{as } \rho_n \longrightarrow \rho$$

for i, j = 1,⋯, p and G_n, τ^n satisfying (3.2).

Here we just prove Lemma 3.2. The argument for Lemma 3.3 is similar and more involved .

Proof of Lemma 3.2. First let us look at the case when the parameter is one dimensional.

On $[s: f_\theta(s) = 0]\backslash N$, $\dot{f}_\theta(s)$ must be zero since otherwise $f_t(s)$ would be negative for some t. Now we can write $\int_R |\dot{f}_\theta(s)|\, ds = \int_R \{|\dot{f}_\theta(s)|\, f_\theta^{-1/2}(s)\} f_\theta^{1/2}(s)\, ds$, hence by (i), the Cauchy-Schwartz inequality and Lemma 2.1(i), $\int_R |\dot{f}_\theta(s)|\, ds$ is finite and continuous. Now since for $s \notin N$

and $\theta \in V$ we have $f_\theta(s) = f_\rho(s) + \int_\rho^\theta \dot{f}_t(s)\,dt$, it follows by Fubini's theorem that $\bar{F}_\theta(x) = \bar{F}_\rho(x) + \int_\rho^\theta\{\int_x^\infty \dot{f}_t(s)\,ds\}\,dt$. So by Lemma 2.1 (i), for every x, $\dot{\bar{F}}_\theta(x)$ exists, is equal to $\int_x^\infty \dot{f}_\theta(s)\,ds$ and is continuous in θ. Next, note that for every $x < \tau_{F_\theta}$,

$$[\dot{\bar{F}}_\theta(x)]^2/\bar{F}_\theta(x)$$

$$= (1/\bar{F}_\theta(x))[\int_x^\infty (\dot{f}_\theta f_\theta^{-1/2})(s)\, f_\theta^{1/2}(s)\,ds]^2$$

(3.12) $$\le \int_x^\infty (\dot{f}_\theta^{\,2}/f_\theta)(s)\,ds$$

and $$\int_R \int_x^\infty [\dot{f}_\theta^{(i)}]^2/f_\theta\,ds\,dG(x)$$

(3.13) $$= \int G\,[\dot{f}_\theta^{(i)}]^2/f_\theta\,d\lambda.$$

Thus Lemma 2.1 gives us the continuity of $\|\{\dot{s}_\theta - \dot{s}_\rho\}(-\infty,\tau]\|_G$ in θ for any (sub-)d.f. G and $\tau \in R$. Now (3.12), (3.13), assumptions (i), (ii) and the proof of Lemma A.2 in Hajek(1972) show that, there is a neighborhood of ρ, in which $s_\theta(x,1)$ is absolutely continuous in θ for a.e. x $[\lambda]$ and $s_\theta(x,0)$ is absolutely continuous in θ for every x. Thus for $(x,y) \notin N_0 \times \{1\}$, where N_0 is a λ-null set, we have for ρ_n in a neighborhood of ρ,

$$s_{\rho_n} = s_\rho + \int_\rho^{\rho_n} \dot{s}_t\,dt$$

(3.14) $$= s_\rho + (\rho_n - \rho)\{\dot{s}_\rho + (\rho_n - \rho)^{-1}\int_\rho^{\rho_n} (\dot{s}_t - \dot{s}_\rho)\,dt\},$$

and

$$\|\{(\rho_n - \rho)^{-1}\int_\rho^{\rho_n} (\dot{s}_t - \dot{s}_\rho)\,dt\}(-\infty,\tau]\|_G^2$$

$$\le |(\rho_n - \rho)^{-1}\int_\rho^{\rho_n} \|\{\dot{s}_t - \dot{s}_\rho\}(-\infty,\tau]\|_G^2\,dt|$$

(3.15) $$= \|\{\dot{s}_{\xi_n} - \dot{s}_\rho\}(-\infty,\gamma]\|_G^2$$

for some ξ_n between ρ_n and ρ. Thus to prove (3.7) and (3.8), it suffices to prove that as $\rho_n \longrightarrow \rho$,

$$\|\{\dot{s}_{\rho_n} - \dot{s}_\rho\}(-\infty,\gamma]\|_G \longrightarrow 0, \quad \|\{\dot{s}_{\rho_n} - \dot{s}_\rho\}(-\infty,\gamma^n]\|_{G_n} \longrightarrow 0,$$

respectively. The first being guaranteed by the continuity of $\|\{\dot{s}_\theta - \dot{s}_\rho\}(-\infty,\gamma]\|_G$ in θ, we only have to prove $\|\{\dot{s}_{\rho_n} - \dot{s}_\rho\}(-\infty,\gamma^n]\|_{G_n} \longrightarrow 0$.

When θ is multidimensional, we can apply the above argument to each component of θ. Consequently, we have $\|\{\dot{s}^{(i)}_{\rho_n} - \dot{s}^{(i)}_\rho\}(-\infty,\gamma]\|_G \longrightarrow 0$ for $\rho_n \longrightarrow \rho$, and it only remains to prove $\|\{\dot{s}^{(i)}_{\rho_n} - \dot{s}^{(i)}_\rho\}(-\infty,\gamma^n]\|_{G_n} \longrightarrow 0$.

We have

$$\|\{\dot{s}^{(i)}_{\rho_n} - \dot{s}^{(i)}_\rho\}(-\infty,\gamma^n]\|_{G_n}^2$$

$$= \|\{\dot{s}^{(i)}_{\rho_n} - \dot{s}^{(i)}_\rho\}(-\infty,\gamma^n]\|_G^2$$

$$+ 4^{-1}\int_{-\infty}^{\gamma^n} [\dot{f}^{(i)}_{\rho_n}/(f^{1/2}_{\rho_n}) - \dot{f}^{(i)}_\rho/(f^{1/2}_\rho)]^2(\bar{G}_n - \bar{G})\, d\lambda$$

(3.16) $$+ 4^{-1}\int_{-\infty}^{\gamma^n} [\dot{\bar{F}}^{(i)}_{\rho_n}/(\bar{F}^{1/2}_{\rho_n}) - \dot{\bar{F}}^{(i)}_\rho/(\bar{F}^{1/2}_\rho)]^2 d(G_n - G).$$

The first term converges to zero as mentioned above, and by (3.12) and repeated use of Lemma 2.1 (i), the second term also converges to zero. Applying the integration by parts formula and (3.12) to the third term we have

$$\left|\int_{-\infty}^{\gamma^n} [\dot{\bar{F}}^{(i)}_{\rho_n}/(\bar{F}^{1/2}_{\rho_n}) - \dot{\bar{F}}^{(i)}_\rho/(\bar{F}^{1/2}_\rho)]^2\, d(G_n - G)\right|$$

$$= \Big| -2\int_{-\infty}^{\gamma^n} (G_{n-} - G_-)[\dot{\bar F}^{(i)}_{\rho_n}/(\bar F^{1/2}_{\rho_n}) - \dot{\bar F}^{(i)}_{\rho}/(\bar F^{1/2}_{\rho})]$$

$$\cdot \frac{d}{dx}[\dot{\bar F}^{(i)}_{\rho_n}/(\bar F^{1/2}_{\rho_n}) - \dot{\bar F}^{(i)}_{\rho}/(\bar F^{1/2}_{\rho})]\ dx$$

$$+ (G_n - G)[\dot{\bar F}^{(i)}_{\rho_n}/(\bar F^{1/2}_{\rho_n}) - \dot{\bar F}^{(i)}_{\rho}/(\bar F^{1/2}_{\rho})]^2(\gamma^n)\Big|$$

$$(3.17)\quad \le 2 \sup_{(-\infty,\gamma]} |G_n - G| \{[U^{(i)}(\rho_n)]^{1/2} + [U^{(i)}(\rho)]^{1/2}\}$$

$$\cdot \int_{-\infty}^{\gamma^n} \Big|\frac{d}{ds}[\dot{\bar F}^{(i)}_{\rho_n}/(\bar F^{1/2}_{\rho_n}) - \dot{\bar F}^{(i)}_{\rho}/(\bar F^{1/2}_{\rho})]\Big|\ ds$$

$$+ \sup_{(-\infty,\gamma]} |G_n - G| \{[U^{(i)}(\rho_n)]^{1/2} + [U^{(i)}(\rho)]^{1/2}\}.$$

Since $\sup_{(-\infty,\gamma]} |G_n - G| \longrightarrow 0$, it suffices to prove that $\int_{-\infty}^{\gamma^n} |\frac{d}{ds}[\dot{\bar F}^{(i)}_{\rho_n}/(\bar F^{1/2}_{\rho_n}) - \dot{\bar F}^{(i)}_{\rho}/(\bar F^{1/2}_{\rho})]|\ ds$ remains bounded. We can bound it by $\int_R [|\dot f^{(i)}_{\rho_n}|/(\bar F^{1/2}_{\rho_n}) + \int |\dot f^{(i)}_{\rho}|/(\bar F^{1/2}_{\rho})]$ $+ 2^{-1}\int_R \dot{\bar F}^{(i)}_{\rho_n}/(\bar F^{3/2}_{\rho_n}) dF_{\rho_n} + 2^{-1}\int \dot{\bar F}^{(i)}_{\rho}/(\bar F^{3/2}_{\rho})\ dF_{\rho}$. The first two terms in the sum are $2W^{(i)}(\rho_n)$, $2W^{(i)}(\rho)$ respectively and therefore remain bounded. To deal with the last two terms in the sum, denote them by $B(\rho_n)$, $B(\rho)$ respectively. Let $p = 2 + \epsilon$, and q the conjugate of p: $p^{-1} + q^{-1} = 1$. Then $q^{-1} - 2^{-1} > 0$. Take $a = q^{-1}$, then $aq = 1$, $ap = p - 1$. Holder's inequality gives us

$$|\dot{\bar F}^{(i)}_{\rho_n}(x)|/\bar F^{3/2}_{\rho_n}(x)$$

$$\le \bar F_{\rho_n}^{-3/2}(x)[\int_x^\infty |\dot f^{(i)}_{\rho_n}/f^a_{\rho_n}|^p(s)\ ds]^{1/p}$$

$$\cdot [\int_x^\infty f^{aq}_{\rho_n}(s)\ ds]^{1/q}$$

$$= \bar F^{1/q - 3/2}_{\rho_n}(x)[\int_x^\infty |\dot f^{(i)}_{\rho_n}/f^a_{\rho_n}|^p(s)\ ds]^{1/p}$$

$$(3.18)\quad \le \bar F^{1/q - 3/2}_{\rho_n}(x)[V^{(i)}(\rho_n)]^{1/p}.$$

hence

(3.19) $$B(\rho_n) \le 2^{-1}[V^{(i)}(\rho_n)]^{1/p}(q^{-1} - 2^{-1})^{-1}.$$

Similarly,

(3.20) $$B(\rho) \le 2^{-1}[V^{(i)}(\rho)]^{1/p}(q^{-1} - 2^{-1})^{-1}.$$

Finally the result follows from (3.12) through (3.16). □

Now we are ready for the main results of this chapter: the differentiability of the minimum Hellinger distance functionals. We state the results separately for the case when G is known and the case when G is unknown.

Theorem 3.1. Suppose

(i) assumptions (a) and (b) in Lemma 3.1 hold,

(ii) $\theta_{00} = \Psi(f; G)$ exists, is unique and lies in the interior of Θ,

(iii) the matrix $\int \ddot{s}_{\theta_{00}}[L(f)]^{1/2}\, d\mu$ is nonsingular,

(iv) assumptions (i) and (ii) of Lemmas 3.2 and 3.3 hold for $\rho = \theta_{00}$. Then, for f_n in a Hellinger neighborhood of f,

(3.21) $$\begin{aligned}\Psi(f_n; G) - \Psi(f; G) &= [-\int \ddot{s}_{\theta_{00}}[L(f)]^{1/2}\, d\mu + u_n]^{-1} \\ &\quad \cdot \int \dot{s}_{\theta_{00}}([L(f_n)]^{1/2} - [L(f)]^{1/2})\, d\mu\end{aligned}$$

where all entries of the matrix u_n converge to zero as $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow 0$.

Theorem 3.2. Suppose

(i) assumptions (a), (b) and (c) in Lemma 3.1 hold,

(ii) $\theta_0 = \Psi(f;\ G;\ \gamma)$ exists, is unique and lies in the interior of Θ,

(iii) the matrix $\int_0^{\gamma} (\dot{s}_{\theta_0} \dot{s}_{\theta_0}^t + \ddot{s}_{\theta_0}(s_{\theta_0} - [L(f)]^{1/2}))d\mu$ is nonsingular,

(iv) assumptions (i), (ii) and (iii) of Lemmas 3.2 and 3.3 hold for $\rho = \theta_0$. Then, for f_n in a Hellinger neighborhood of f, G_n, γ^n satisfying (3.2) and $\Delta G(\gamma) = 0$,

$$\Psi(f_n;\ G_n;\ \gamma^n) - \Psi(f;\ G;\ \gamma)$$

$$= \{\int_{-\infty}^{\gamma} (\dot{s}_{\theta_0} \dot{s}_{\theta_0}^t + \ddot{s}_{\theta_0}(s_{\theta_0} - [L(f)]^{1/2}))\, d\mu + v_n\}^{-1} \cdot \int_{-\infty}^{\gamma^n} \dot{s}_{\theta_0}([L(f_n)]^{1/2} - [L(f_{\theta_0}]^{1/2})\, d\mu_n, \tag{3.22}$$

where all entries of the matrix v_n converge to zero as $\|f_n^{\ 1/2} - f^{\ 1/2}\|_2 \longrightarrow 0$.

We only give the proof for Theorem 3.2, since the proof for Theorem 3.1 is similar and simpler.

Proof of Theorem 3.2. First note that if assumptions (i), (ii) and (iii) hold for θ_0, then they hold for all points in a neighborhood of θ_0. Let n be sufficiently large so that θ_{nn} is in that neighborhood. Since θ_{nn} minimizes $\int_{-\infty}^{\gamma^n} (s_t^2 - 2\, s_t[L(f_n)]^{1/2})\, d\mu_n$, we have by Lemma 3.2, for suffficiently large n,

$$\int_{-\infty}^{\gamma^n} \dot{s}_{\theta_{nn}} (s_{\theta_{nn}} - [L(f)]^{1/2})\, d\mu_n = 0. \tag{3.23}$$

Expanding $s_{\theta_{nn}}$, $\dot{s}_{\theta_{nn}}$ around θ_0, we can rewrite (3.23) as

$$0 = \int_{-\infty}^{\gamma^n} [\dot{s}_{\theta_0} + (\ddot{s}_{\theta_0} + R_n)(\theta_{nn} - \theta_0)]$$

$$\cdot[s_{\theta_0} + (\dot{s}_{\theta_0} + r_n)^t (\theta_{nn} - \theta_0) - [L(f_n)]^{1/2}]\, d\mu_n$$

$$= \int_{-\infty}^{\gamma^n} \dot{s}_{\theta_0}(s_{\theta_0} - [L(f_n)]^{1/2})\, d\mu_n$$

$$+ \int_{-\infty}^{\gamma^n} \dot{s}_{\theta_0}(\dot{s}_{\theta_0} + r_n)^t\, d\mu_n\, (\theta_{nn} - \theta_0)$$

$$+ \int_{-\infty}^{\gamma^n} (\ddot{s}_{\theta_0} + R_n)(s_{\theta_0} - [L(f_n)]^{1/2})\, d\mu_n\, (\theta_{nn} - \theta_0)$$

$$+ \int_{-\infty}^{\gamma^n} (\ddot{s}_{\theta_0} + R_n)(\theta_{nn} - \theta_0)(s_{\theta_0} + r_n)^t\, d\mu_n\, (\theta_{nn} - \theta_0).$$

An argument similar to that used to prove Lemma 3.2 shows that for G_n, γ^n satisfying (3.2), as $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow \infty$, $\int_{-\infty}^{\gamma^n} (\dot{s}_{\theta_0}\dot{s}_{\theta_0}^t + \ddot{s}_{\theta_0}(s_{\theta_0} - [L(f_n)]^{1/2}))\, d\mu_n \longrightarrow$ $\int_{-\infty}^{\gamma} (\dot{s}_{\theta_0}\dot{s}_{\theta_0}^t + \ddot{s}_{\theta_0}(s_{\theta_0} - [L(f)]^{1/2}))\, d\mu$. Thus the above equation can be written as

$$0 = \int_{-\infty}^{\gamma^n} \dot{s}_{\theta_0}(s_{\theta_0} - [L(f_n)]^{1/2})\, d\mu_n$$

$$+ \{\int_{-\infty}^{\gamma} (\dot{s}_{\theta_0}\dot{s}_{\theta_0}^t + \ddot{s}_{\theta_0}(s_{\theta_0} - [L(f)]^{1/2}))\, d\mu$$

$$+ v_n\}(\theta_{nn} - \theta_0),$$

where all entries of the matrix v_n converge to zero as $\|f_n^{1/2} - f^{1/2}\|_2 \longrightarrow \infty$. Therefore the result follows. □

Notice that differentiating twice in the identity $\int s_\theta^2\, d\mu \equiv 1$ yields $\int (\dot{s}_\theta \dot{s}_\theta^t + \ddot{s}_\theta s_\theta)\, d\mu \equiv 0$, which results in the shorter expression (3.21).

4. ASYMPTOTIC DISTRIBUTIONS

When G is known, the MHD estimator of $\Psi(f;\ G)$ is defined as

(4.1) $$\hat{\theta}_{1n} = \Psi(f_n;\ G);$$

when G is unknown, the MHD estimator of $\Psi(f;\ G;\ \tau)$ is defined as

(4.2) $$\hat{\theta}_{2n} = \Psi(f_n;\ \hat{G}_n;\ T)$$

(recall $T = \max(\tilde{X}_1, \cdots, \tilde{X}_n)$), where $\hat{G}_n$ is the product-limit estimator as defined in (2.3), and f_n is the kernel density estimator as defined in (2.23):

$$f_n(x) = a_n^{-1} \int K((x-y)/a_n)\ d\hat{F}_n(y)$$

for some kernel function K and constant $a_n > 0$. We now prove the consistency of $\hat{\theta}_{1n}$, $\hat{\theta}_{2n}$.

Theorem 4.1. Suppose that

(i) assumptions (a) and (b) of Lemma 3.1 hold,

(ii) K is nonnegative, continuous and of bounded variation on R, $\int K\ d\lambda = 1$, $K(s) \longrightarrow 0$ as $s \longrightarrow -\infty$,

(iii) $a_n \longrightarrow 0$ and $n^{1/2}a_n \longrightarrow \infty$,

(iv) $\int_0^{\tau} (1/\bar{G}_-)dF < \infty$.

Then

$$\hat{\theta}_{1n} \xrightarrow[P]{} \Psi(f;\ G) \quad \textbf{if } \Psi(f;\ G) \textbf{ is unique.}$$

If, in addition, assumption (c) in Lemma 3.1 holds and G **is continuous at** τ, **then**

$$\hat{\theta}_{2n} \xrightarrow[P]{} \Psi(f;\ G;\ \tau) \textbf{ if } \Psi(f;\ G;\ \tau) \textbf{ is unique.}$$

Proof. By Wang(1987), $\sup_{[0,\tau]} |\hat{G}_n - G| = o_p(1)$. So Theorem 2.3 and Lemma 3.1 give the result immediately. □

To investigate the asymptotic distributions of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$, we need to establish some convergence results of the kernel density estimator f_n and the smoothed product-limit estimator F_n. Let $\|\cdot\|_\infty$ denote the $L^\infty(R)$- norm.

Lemma 4.1. Suppopse that

(i) $f' = \frac{d}{dx} f$ **exists and is absolutely continuous on** $[0, \tau]$, $\|f'\|_\infty < \infty$ **and** $\|f''\|_\infty < \infty$,

(ii) $\tau < \infty$ **and** $\int_0^\tau (1/\bar{G}_-)\, dF < \infty$,

(iii) K is nonnegative, symmetric and absolutely continuous, $\int K d\lambda = 1$, **support of** $K \subset [-M, M]$ **for some** $M < \infty$,

(iv) $a_n \longrightarrow 0$, $n^{1/2} a_n^2 \longrightarrow 0$ **and for some** $\epsilon > 0$, $n^{1/2} a_n^{1+\epsilon} \longrightarrow \infty$,

(v) U is bounded, V is right continuous and of bouded variation on $[0, \tau]$.

Then

$$\int_0^\tau n^{1/2}[F_n - F]\, U\, dG \xrightarrow[P]{} \int_0^\tau P^1\, U\, dG,$$
$$\int_0^\tau n^{1/2}[f_n - f]\, V\, d\lambda \xrightarrow[P]{} -\int_0^\tau P^1_-\, dV.$$

Proof. Let

$$\tilde{f}_n(x) = a_n^{-1} \int K((x-y)/a_n)\, dF(y), \tag{4.3}$$
$$\tilde{F}_n(x) = \int_{-\infty}^x \tilde{f}_n\, d\lambda.$$

Then we have

$$n^{1/2}[F_n(x) - \tilde{F}_n(x)] = \int P_n^1(x - a_n t)\, K(t)\, dt,$$
$$n^{1/2}[\tilde{F}_n(x) - F(x)] = \int n^{1/2}[F(x - a_n t) - F(x)]\, K(t)dt.$$

It follows from (2.13) that

$$\int_0^\tau n^{1/2}[F_n(x) - \tilde{F}_n(x)]\, U\, dG \xrightarrow[P]{} \int_0^\tau P^1 U\, dG.$$

By the integral form of the mean value theorem,

$$F(b) - F(a) = (b - a)f(a) + \int_a^b (b - u)f'(u)\, du$$

for a, b $\in$ R. Hence by the symmetry of K,

$$(4.4) \quad \sup_R n^{1/2}|\tilde{F}_n(x) - F(x)| \le M^2 \|f'\|_\infty\, n^{1/2} a_n^2 \longrightarrow 0$$

as $n \longrightarrow 0$. This gives us $\int_0^\tau n^{1/2}[\tilde{F}_n(x) - F(x)]\, U\, dG \longrightarrow 0$.

Thus we have

$$\int_0^\tau n^{1/2}[F_n(x) - F(x)]\, U\, dG \xrightarrow[P]{} \int_0^\tau P^1 U\, dG.$$

As for the second assertion of our lemma, we have

$$\begin{aligned}
&\int_0^\tau n^{1/2}[f_n - \tilde{f}_n]\, V\, d\lambda \\
&= \int_R V(x)\, a_n^{-1} \int_R K((x-y)/a_n)\, dP_n^1(y)\, dx \\
&= \int_R \int_R V(y + a_n t)\, K(t)\, dt\, d\, P_n^1(y) \\
&= - \int P_{n-}^1(y - a_n t)\, dV(y)\, K(t)\, dt \\
&\xrightarrow[P]{} - \int P_-^1\, dV.
\end{aligned}$$

Hence it only remains to prove $\int_0^\tau n^{1/2}[\tilde{f}_n(x) - f(x)]\, V\, d\lambda \longrightarrow 0$. Again, by the integral form of the mean value theorem,

$$f(b) - f(a) = (b - a)f'(a) + \int_a^b (b - u)f''(u)\, du$$

for a, b $\in$ R. So

$$(4.5) \quad \sup_R n^{1/2}|\tilde{f}_n(x) - f(x)| \le M^2 \|f''\|_\infty\, n^{1/2} a_n^2 \longrightarrow 0,$$

which implies $\int_0^\tau n^{1/2}[\tilde{f}_n(x) - f(x)]\, V\, d\lambda \longrightarrow 0$. □

Lemma 4.2. Suppose

(i) $\{B_n\}$ is a family of uniformly bounded functions on

$[0, \tau)$, and for any $x \in [0,\tau)$, $x_n \longrightarrow x$ implies $B_n(x_n) \xrightarrow[P]{}$ $B(x)$ for a bounded function B.

(ii) G is continuous at τ.

Then $\int_0^T B_n \, d(\hat{G}_n - G) \xrightarrow[P]{} 0$.

Proof. S-W state in their Theorem 7.3.1 the uniform strong consistency of $\hat{F}_n$ on $[0, \tau)$. There is some difficulty in their argument about the uniform convergence on $[0, \tau)$. But their proof shows that for any $t \in [0, \tau)$, $\hat{F}_n(t)$ is strongly consistent for $F(t)$. Thus similarly we have the strong consistency of $\hat{G}_n(t)$ for any $t \in [0, \tau)$. This implies that for ω in a set of probability 1, $\hat{G}_n^{-1} \longrightarrow G^{-1}$ at continuity points of G^{-1} in $(0, G(\tau))$. Since $G^{-1} = \tau$ only at $t = 1$, for a.e.t we have $G^{-1}(t) < \tau$, hence for a.e.t, $B_n \circ \hat{G}^{-1}(t) \longrightarrow B \circ G^{-1}(t)$. Now let $\rho_n = \hat{G}_n(T)$, $\eta_n = G(T)$ and $\rho = G(\tau)$. By Wang(1987), $\rho_n \xrightarrow[P]{} \rho$. And continuity of G at τ gives us $\eta_n \longrightarrow \rho$. Therefore

$$\int_0^T B_n \, d(\hat{G}_n - G)$$
$$= \int_0^{\rho_n} B_n \circ \hat{G}_n^{-1}(t) \, dt - \int_0^{\eta_n} B_n \circ G^{-1}(t) \, dt \xrightarrow[P]{} 0. \quad \square$$

Lemma 4.3. Suppose that

(i) $f' = \frac{d}{dx} f$ exists and is absolutely continuous on $[0, \tau]$, $\|f'\|_\infty < \infty$, $\|f''\|_\infty < \infty$ and $\inf\{f[f > 0]\} > 0$,

(ii) $\tau < \infty$ and $\int_0^\tau (1/\bar{G}_-) \, dF < \infty$,

(iii) K is nonnegative, symmetric and absolutely continuous, $\int K d\lambda = 1$, support of $K \subset [-M, M]$ for some $M < \infty$,

(iv) $a_n \longrightarrow 0$, $n^{1/2} a_n^2 \longrightarrow 0$ and for some $\epsilon > 0$

$n^{1/2}a_n^{1+\epsilon} \longrightarrow \infty$. Then

$$\int_0^T n^{1/2}(f_n - f)^2 \, d\lambda \xrightarrow[P]{} 0.$$

Proof. By (4.5), it suffices to show $\int_0^T n^{1/2}(f_n - \tilde{f}_n)^2 \, d\lambda \xrightarrow[P]{} 0$. Let $D_n(x,t) = [P_n^1(x+a_nt) - P_n^1(x-a_nt)]$. By symmetry of K and the Cauchy-Schwartz inequality,

$$\begin{aligned} &\int_0^T n^{1/2}(f_n - \tilde{f}_n)^2 \, d\lambda \\ &= (n^{1/2}a_n^2)^{-1}\int_0^T [\int_0^M D_n(x,t)\, K'(t)dt]^2 \, dx \\ &\leq W(n^{1/2}a_n^2)^{-1} \int_0^T \int_0^M D_n^2(x,t)\ |K'(t)| \, dt \, dx \\ &= W(n^{1/2}a_n^2)^{-1} \int_0^M \int_0^T D_n^2(x,t)\, dx \, |K'(t)| \, dt. \end{aligned}$$

Writing $P_n(s) = Z_n^T(s)\bar{F}(s)$ for $s < T$ and using $(a+b)^2 \leq 2a^2 + 2b^2$, we have $E\, D_n(x,t)^2[a_nt < x < T - a_nt]$

$$\begin{aligned} &= E[Z_n^T(x+a_nt)\bar{F}(x+a_nt) - Z_n^T(x-a_nt)\bar{F}(x-a_nt)]^2[a_nt < x < T - a_nt] \\ &\leq 2\, \bar{F}^2(x+a_nt)\, E[Z_n^T(x+a_nt) - Z_n^T(x-a_nt)]^2 \\ &\quad + 2\, [\bar{F}(x+a_nt) - \bar{F}(x-a_nt)]^2\, E[Z_n^T\,(x-a_nt)]^2. \end{aligned}$$

By Lemma 2.5,

$$\begin{aligned} &\bar{F}^2(x+a_nt)\, E[Z_n^T(x+a_nt) - Z_n^T(x-a_nt)]^2 \\ &\leq 4[A(x+a_nt) - A(x-a_nt)] + 4\,\bar{F}^{\,2}(x+a_nt)\, A(x-a_nt) \\ &\qquad \cdot[\bar{F}^{\,-2}(x+a_nt) - \bar{F}^{\,-2}(x-a_nt)]. \end{aligned}$$

Since for $0 < a < b$, $a^2(a^{-2} - b^{-2}) = b^{-2}(a + b)(b - a)$ $< 2(b - a)/b$, the second term on the RHS of the last inequality does not exceed

$$8A(\tau)[F(x+a_nt) - F(x-a_nt)]\, \bar{F}^{\,-1}(x-a_nt).$$

Lemma 2.5 also gives us

$$\begin{aligned} &[\bar{F}(x+a_nt) - \bar{F}(x-a_nt)]^2\, E[Z_n^T\,(x-a_nt)]^2 \\ &\leq 4\, [\bar{F}(x+a_nt) - \bar{F}(x-a_nt)]^2\, \bar{F}^{\,-2}(x-a_nt)\, A(x-a_nt) \end{aligned}$$

$$\leq 4\ A(\tau)\ [F(x+a_n t) - F(x-a_n t)]\ \bar{F}^{-1}(x-a_n t).$$

Hence $\quad E\ D_n(x,t)^2[a_n t < x < T - a_n t]$

$$\leq 8[A(x+a_n t) - A(x-a_n t)]$$
$$+ 24\ A(\tau)[F(x+a_n t) - F(x-a_n t)]\ \bar{F}^{-1}(x-a_n t).$$

This and Holder's inequality gives us

$$E \int_{a_n t}^{T-a_n t} \{D_n(x,t)\}^{2-2\epsilon}\, dx$$
$$= \int_R E\ \{D_n(x,t)[a_n t < x < T - a_n t]\}^{2-2\epsilon}\, dx$$
$$\leq \int_{a_n t}^{\tau-a_n t} \{E\ D_n(x,t)^2[a_n t < x < T - a_n t]\}^{1-\epsilon}\, dx$$
$$\leq \int_{a_n t}^{\tau-a_n t} \{8[A(x+a_n t) - A(x-a_n t)] + 24\ A(\tau)$$
$$\cdot[F(x+a_n t) - F(x-a_n t)]\ \bar{F}^{-1}(x-a_n t)\}^{1-\epsilon}\, dx$$

(4.6) $$\leq W \int_{a_n t}^{\tau-a_n t} [A(x+a_n t) - A(x-a_n t)]\}^{1-\epsilon}\, dx$$
$$+ W \int_{a_n t}^{\tau-a_n t} \{[F(x+a_n t) - F(x-a_n t)]\ \bar{F}^{-1}(x-a_n t)\}^{1-\epsilon} dx,$$

where W is independent of t.

Since

$$\int_{a_n t}^{\tau-a_n t} [A(x+a_n t) - A(x-a_n t)]dx = \int_{a_n t}^{\tau-a_n t} \int_{x-a_n t}^{x+a_n t} dA(u)\, dx$$
$$\leq \int_0^\tau \int_{u-a_n t}^{u+a_n t} dx\, dA(u) = 2\, a_n t\ A(\tau) \leq 2M\ a_n A(\tau),$$

by Holder's inequality the first term in (4.6) does not exceed $W\ \tau^\epsilon\{2M\ a_n A(\tau)\}^{1-\epsilon}$. The second term in (4.6) does not exceed $W\{2M\ a_n \|f\|_\infty\}^{1-\epsilon}\ \inf\{f[f > o]\}\ \int_0^\tau F^{\epsilon-1}\, dF$. Thus the sum in (4.6) can be written as $B_n a_n^{1-\epsilon}$ for some bounded quantity B_n independent of t. It follows that

$$(n^{1/2}a_n^2)^{-1}\int_0^M \int_{a_n t}^{T-a_n t} D_n(x,t)^{2-2\epsilon}\, dx\ |K'(t)|\ dt \xrightarrow[P]{} 0.$$

Hence by (2.19) $(n^{1/2}a_n^2)^{-1}\int_0^M \int_{a_n t}^{T-a_n t} D_n(x,t)^2\, dx\, |K'(t)|\, dt$ $\xrightarrow[P]{} 0$. Since we also have

$$\int_0^{a_n t} D_n^2(x,t)\, dx + \int_{T-a_n t}^{T} D_n^2(x,t)\, dx$$
$$\leq 4\, a_n M \sup_n \sup_{[0,\tau)} |P_n^1|,$$

the result follows. □

The following theorems establish the asymptotic distributions of our estimators $\hat{\theta}_{1n}$, $\hat{\theta}_{2n}$. Recall that from the beginning of Chapter 3, when X has a density f w.r.t the Lebesgue measure and Y has distribution G, $(\tilde{X}, \delta)$ has a density L(f) w.r.t. μ_G. Since G remains unchanged throughout, we will simply refer to the weak convergence as under L(f). The theorems show that for a general density $f \neq f_\theta$, the asymptotic distributions are slightly different for the two cases when G is known and when G is unknown. At f_θ they coincide. We will use the differentiability of Ψ as in (3.21), (3.22), specifying $\mu_n = \mu_{\hat{G}_n}$, $\gamma = \tau$. Thus

$\theta_0 = \Psi(f;\, G;\, \tau)$, $\theta_{00} = \Psi(f;\, G)$. Denote $\rho_1 = 2^{-1}\dot{f}_{\theta_0} f_{\theta_0}^{-1/2}$, $\rho_0 = 2^{-1}\dot{\bar{F}}_{\theta_0} \bar{F}_{\theta_0}^{-1/2}$, $\varphi_1 = \rho_1 f^{-1/2}$, $\varphi_0 = \rho_0 \bar{F}^{-1/2}$ and extend them on R by defining them to be zero outside the support of f_{θ_0} or the support of f.

Theorem 4.2. Assume (i) through (iv) of Theorem 3.1 hold. In addition, Suppose

(i) $\|f_{\theta_{00}}\|_\infty < \infty$, $\|\dot{f}^{(i)}_{\theta_{00}}\|_\infty < \infty$ for $i = 1, \cdots, n$ and $\inf\{f_{\theta_{00}}[f_{\theta_{00}} > 0]\} > 0$.

(ii) φ_1 is of bounded variation on $[0, \tau]$.

(iii) f' exists and is absolutely continuous on $[0, \tau]$, $\|f'\|_\infty < \infty$, $\|f''\|_\infty < \infty$ and $\inf\{f[f > 0]\} > 0$.

(iv) $\tau < \infty$ and $\int_0^\tau (1/\bar{G}_-)\, dF < \infty$.

(v) K is nonnegative, symmetric and absolutely continuous, $\int K d\lambda = 1$, support of $K \subset [-M, M]$ for some $M < \infty$.

(vi) $a_n \longrightarrow 0$, $n^{1/2} a_n^2 \longrightarrow \infty$, and for some $\epsilon > 0$, $n^{1/2} a_n^{1+\epsilon} \longrightarrow \infty$.

Then, under L(f), $n^{1/2}(\hat{\theta}_{1n} - \Psi(f; G))$ converges weakly to a normal distribution with mean zero and finite variance.

In particular, under $L(f_\theta)$, $n^{1/2}(\hat{\theta}_{1n} - \Psi(f; G))$ converges weakly to $N(0, \Sigma^{-1})$, where Σ is the Fisher information matrix:

$$\Sigma = E \frac{\partial}{\partial\theta} (\ln L(f_\theta)(\tilde{X}, \delta)) \left[\frac{\partial}{\partial\theta} (\ln L(f_\theta)(\tilde{X}, \delta))\right]^t.$$

Theorem 4.3. Assume (i) through (iv) of Theorem 3.2 hold. In addition, Suppose

(i) $\|f_{\theta_0}\|_\infty < \infty$, $\|\dot{f}^{(i)}_{\theta_0}\|_\infty < \infty$ for $i = 1, \cdots, n$ and $\inf\{f_{\theta_0}[f_{\theta_0} > 0]\} > 0$.

(ii) φ_1 is of bounded variation on $[0, \tau]$.

(iii) f' exists and is absolutely continuous on $[0, \tau]$, $\|f'\|_\infty < \infty$, $\|f''\|_\infty < \infty$ and $\inf\{f[f > 0]\} > 0$.

(iv) $\tau_{F_{\theta_0}} \leq \tau < \infty$, $\int_0^\tau (1/\bar{G}_-)\, dF < \infty$ and G is continuous at τ.

(v) K is nonnegative, symmetric and absolutely continuous, $\int K d\lambda = 1$, support of $K \subset [-M, M]$ for some $M < \infty$.

(vi) $a_n \longrightarrow 0$, $n^{1/2} a_n^2 \longrightarrow 0$ and for some $\epsilon > 0$, $n^{1/2} a_n^{1+\epsilon} \longrightarrow \infty$.

Then, under L(f), $n^{1/2}(\hat{\theta}_{2n} - \Psi(f;\, G;\, \tau))$ converges weakly to a normal distribution with mean zero and finite variance.

In particular, under $L(f_\theta)$, $n^{1/2}(\hat{\theta}_{2n} - \Psi(f;\, G;\, \tau))$ converges weakly to $N(0, \Sigma^{-1})$.

We just prove Theorem 4.3. The proof for Theorem 4.2 is similar.

Proof of Theorem 4.3. Throughout the proof, we adopt the special construction of $\tilde{X}_{ni}$'s and δ_{ni}'s as in Theorem 2.1. we will need to use the algebraic identity(for a, b > 0)

(4.7) $$b^{1/2} - a^{1/2}$$
$$= \frac{1}{2a^{1/2}}(b - a) - \frac{1}{2a^{1/2}} \frac{b^{1/2} - a^{1/2}}{b^{1/2} + a^{1/2}}(b - a)$$
$$= \frac{1}{2a^{1/2}}(b - a) - \frac{1}{2a^{1/2}} \frac{(b - a)^2}{[b^{1/2} + a^{1/2}]^2}.$$

Under our assumptions the expansion (3.22), with G_n, γ^n, γ replaced by $\hat{G}_n$, T, τ respectively, is valid, where all entries of the matrix v_n converge to zero in probability. Since the coefficient of the integral on the right hand side of(3.22) converges to a nonrandom limit, we only have to deal with the integral in (3.22).

Note that $|\dot{\bar{F}}_{\theta_o}^{(i)}(x)| = |\int_x^\infty \dot{f}_{\theta_0}^{(i)}(s)\, ds|$ $= \int_x^\infty |\dot{f}_{\theta_0}^{(i)}/f_{\theta_0}|\, f_{\theta_0} d\lambda \le \{\sup|\dot{f}_{\theta_0}^{(i)}(s)/f_{\theta_0}|\}\, \bar{F}_{\theta_0}(x)$, and since $\tau_{F_{\theta_0}} \le \tau$ the mean value theorem gives $\bar{F}_{\theta_0}(x) \le \sup f_{\theta_0} \inf\{f[f>0]\}\bar{F}$. Hence φ_0 and $\bar{F}_{\theta_0}/\bar{F}$ are bounded. For the sake of convenience we will use W to denote a bound for all bounded quantities in our argument. Notice that as in (3.23), we have $\int_0^\tau \dot{s}_{\theta_0}([L(f_{\theta_0})^{1/2} - [L(f)]^{1/2})d\mu = 0$. Thus

$$
\begin{aligned}
& n^{1/2}\int_0^T \dot{s}_{\theta_0}([L(f_n)]^{1/2} - [L(f_{\theta_0})]^{1/2})\, d\mu_n \\
&= -\int_0^T \rho_1[f^{1/2} - f_{\theta_0}^{1/2}]n^{1/2}(\hat{G}_n - G)\, d\lambda \\
&+ \int_0^T \rho_0\, [\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}]\, d[n^{1/2}(\hat{G}_n - G)] \\
&- n^{1/2} \int_T^\tau \rho_1[f^{1/2} - f_{\theta_0}^{1/2}]\, \bar{G}\, d\lambda \\
&- n^{1/2} \int_T^\tau \rho_0\, [\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}]\, dG \\
&- \int_0^T \rho_1\, [f_n^{1/2} - f^{1/2}]\, n^{1/2}(\hat{G}_n - G)\, d\lambda \\
&+ \int_0^T \rho_0\, n^{1/2}[\bar{F}_n^{1/2} - \bar{F}^{1/2}]\, d(\hat{G}_n - G) \\
&+ \int_0^T \rho_0\, n^{1/2}[\bar{F}_n^{1/2} - \bar{F}^{1/2}]\, dG \\
&+ \int_0^T \rho_1\, n^{1/2}[f_n^{1/2} - f^{1/2}]\, \bar{G}\, d\lambda \\
&= S_1 + S_2 + R_1 + R_2 + R_3 + R_4 + S_3 + S_4. \qquad (4.8)
\end{aligned}
$$

We can write

$$S_1 = \int_0^T B(x)\, P_n^0(x)\, dx$$

where $B(x) = -\rho_1[f^{1/2} - f_{\theta_0}^{1/2}]$ is bounded.

$$|\int_0^T B^{(i)}(x)[P_n^0(x) - P^0(x)]\,dx|$$
$$\leq \sup_R |B^{(i)}| \sup_{[0,T]} |[P_n^0 - P^0]\bar{F}^{1-\alpha}| \int_0^\tau \bar{F}^{\alpha-1}d\lambda,$$

for $0 < \alpha < 1/2$. Since $\int_0^\tau \bar{F}^{\alpha-1}d\lambda \leq \int_0^\tau \bar{F}^{\alpha-1}dF$ $\cdot(\inf_R f(x)[f(x) > 0])^{-1} < \infty$, by Theorem 2.2 and the fact that $T \longrightarrow \tau$ w.p.1, we have

$$S_1 \xrightarrow[P]{} \int_0^\tau B\,P^0 d\lambda. \tag{4.9}$$

Next, integration by parts gives

$$S_2 = -\int_0^T P_{n-}^0 A\,d\lambda + \{\rho_0[\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}]P_n^0\}(T)\ , \tag{4.10}$$

where

$$\begin{aligned} A &= \frac{d}{dx}\{\rho_0[\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}]\} \\ &= 2^{-1}[\dot{f}_{\theta_0} + 2^{-1}\dot{\bar{F}}_{\theta_0}\bar{F}_{\theta_0}^{-3/2}\bar{F}^{1/2} f_{\theta_0} \\ &\quad - \bar{F}^{1/2}\bar{F}_{\theta_0}^{-1/2} f_{\theta_0} - 2^{-1}\dot{\bar{F}}(\bar{F}_{\theta_0}\bar{F})^{-1/2} f]. \end{aligned} \tag{4.11}$$

When $\tau_{F_{\theta_0}} = \tau$, A is bounded; when $\tau_{F_{\theta_o}} < \tau$, on $[0, \tau_{F_{\theta_0}}]$ $A \leq W\,\bar{F}_\theta^{-1/2}$ and $\bar{F} > \bar{F}(\tau_{F_{\theta_0}}) > 0$. Hence by Theorem 2.2 P_n^0 converges to P^0 uniformly on $[0, \tau_{F_{\theta_0}}]$. Therefore in both cases we have $\int_0^T P_{n-}^0 A\,d\lambda \xrightarrow[P]{} \int_0^\tau P_-^0 A\,d\lambda$. The remainder term in (4.10) $\{\rho_0[\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}]P_n^0\}(T) \leq W(T)(P_n^0\,\bar{F})(T) = o_p(1)$ by (2.19). Hence

$$S_2 \xrightarrow[P]{} -\int_0^\tau P_-^0 A\,d\lambda. \tag{4.12}$$

By (4.7), $S_3 = S_{31} + S_{32}$, where

$$S_{31} = 2^{-1}\int_0^T \varphi_0 n^{1/2}[\bar{F}_n - \bar{F}]\,dG,$$
$$S_{32} = 2^{-1}\int_0^T \varphi_0 n^{1/2}[\bar{F}_n - \bar{F}] \cdot(\bar{F}_n^{1/2} - \bar{F}^{1/2})(\bar{F}_n^{1/2} + \bar{F}^{1/2})^{-1}dG.$$

So $S_{31} \xrightarrow[P]{} - 2^{-1}\int_0^{\tau} \varphi_0 P^1 dG$ by Lemma 4.1 and the fact $T \longrightarrow \tau$ w.p.1 . The integrand of $S_{32}^{(i)}$ is dominated by that of $S_{31}^{(i)}$ in absolute value, and for $x \in [0,\tau)$, $(\bar{F}_n^{1/2} - \bar{F}^{1/2})(\bar{F}_n^{1/2} + \bar{F}^{1/2})^{-1} \longrightarrow 0$. Also $\Delta G(\tau) = 0$. Hence Lemma 2.1 (i) gives $S_{32} \xrightarrow[P]{} 0$. Thus

(4.13) $$S_3 \longrightarrow - 2^{-1}\int \varphi_0 P^1 \, dG.$$

By (4.7) again $S_4 = S_{41} + S_{42}$, where

$$S_{41} = 2^{-1}\int_0^T \varphi_1 n^{1/2}[f_n - f] \, \bar{G} \, d\lambda,$$

$$S_{42} = - 2^{-1}\int_0^T \varphi_1 \bar{G} \, (f_n^{1/2} + f^{1/2})^{-2} \, n^{1/2}[f_n - f]^2 \, d\lambda.$$

From (2.25), $f_n - f$ is bounded. So $\int_T^{\tau} n^{1/2}(f_n - f) \, \varphi_1 \bar{G} \, d\lambda \leq W n^{1/2}(\tau - T) = o_p(1)$. Hence $S_{41} \xrightarrow[P]{} - 2^{-1}\int P_-^1 \, d(\varphi_1 G)$ by Lemma 4.1. Since $|S_{42}^{(i)}| \leq W\int_0^T n^{1/2}(f_n - f)^2 \, \bar{G} \, d\lambda$, by Lemma 4.3 $S_{42}^{(i)} \xrightarrow[P]{} 0$. Thus we have

(4.14) $$S_4 \xrightarrow[P]{} - 2^{-1}\int P^1 d(\varphi_1 \bar{G}).$$

As for the remainder terms R_1, R_2, R_3 and R_4. The results

(4.15) $$R_1 = o_p(1), \; R_2 = o_p(1)$$

follow since $n^{1/2}(\tau - T) \leq (\inf_R f(x)[f(x) > 0])^{-1} \cdot n^{1/2}(F(\tau-) - F(T)) = o_p(1)$. Write

$$R_3 = \int_0^T \{\rho_1(x)[(f_n^{1/2}(x) + f^{1/2}(x)]^{-1}\} \cdot (f_n - f)(x) P_n^0(x) \, dx,$$

then by (2.19), (2.25), (2.26) and the fact that the quantity in { } is bounded, we have

(4.16) $$R_3 \xrightarrow[P]{} 0.$$

Now look at

$$R_4 = \int_0^T \rho_0 (\bar{F}_n^{1/2} + \bar{F}^{1/2})^{-1}(x) \cdot n^{1/2}[F_n - F](x) \, d(\hat{G}_n - G)(x)$$

(4.17) $\qquad = \int_0^T B_n d(\hat{G}_n - G)$, say.

By (4.4) and the fact that $|\rho_0^{(1)}(\bar{F}_n^{1/2} + \bar{F}^{1/2})^{-1}(x)| \le |\rho_0^{(1)}[\bar{F}^{1/2}]^{-1}(x)| \le W$, the integrand is uniformly bounded in probability. Also, by the uniform convergence of P_n^1 to P^1 on each compact subinterval of $[0, \tau)$, and continuity of F, F_{θ_0} and $\dot{F}_{\theta_0}$, we have for $x_n \longrightarrow x \in (0,\tau)$, $B_n(x_n) \xrightarrow[P]{} 4^{-1} \dot{\bar{F}}_{\theta_0}(\bar{F}_{\theta_0}\bar{F})^{-1/2}(x)\, P^1(x)$. Hence by Lemma 4.2,

(4.18) $\qquad R_4 \xrightarrow[P]{} 0$.

Therefore, we have proved that for A defined in (4.11),

$$n^{1/2}\int_0^T \dot{s}_{\theta_0}([L(f_n)]^{1/2} - [L(f_\theta)]^{1/2})\, d\mu_n$$

(4.19) $$\xrightarrow[P]{} -\int_0^\tau \rho_1[f^{1/2} - f_{\theta_0}^{1/2}]\, P^0\, d\lambda - \int_0^\tau P_-^0 A\, d\lambda$$

$$- \int_0^\tau 2^{-1}\varphi_0\, P^1\, dG - \int_0^\tau 2^{-1}P^1\, d(\varphi_1\bar{G}),$$

where the limit has a normal distribution with mean zero and finite variance. Thus $n^{1/2}(\hat{\theta}_{2n} - \Psi(f;\, G;\, \tau))$ also converges weakly to a normal distribution with mean zero and finite variance. The variance can be computed using (2.5). In particular when $f = f_\theta$ for some θ, then the limit becomes

$$- \int_0^\tau 2^{-1}\, \varphi_0\, P^1\, dG - \int_0^\tau 2^{-1}\, P^1 d(\varphi_1\bar{G})$$
$$= \int_0^\tau h_1\, d(P^1/\bar{F}),$$

where

$$h_1(x) = 2^{-1}\int_x^\tau \varphi_0\, \bar{F}_\theta\, dG + 2^{-1}\int_x^\tau \bar{F}_\theta\, d(\varphi_1\bar{G})$$
$$= 2^{-1}[\int_x^\tau 2^{-1}\, \dot{\bar{F}}_\theta\, dG + \int_x^\tau \varphi_1\bar{G}\, dF_\theta] - 2^{-1}\varphi_1\bar{F}_\theta\bar{G}(x)$$
$$= 4^{-1}[\int_x^\tau \dot{\bar{F}}_\theta\, dG + \int_x^\tau \dot{f}_\theta\bar{G}] - 2^{-1}\varphi_1\bar{F}_\theta\bar{G}(x)$$
$$= 4^{-1}\, \frac{\partial}{\partial\theta}\, (\bar{F}_\theta\bar{G}) - 2^{-1}\varphi_1\, \bar{F}_\theta\bar{G}$$

$$= 4^{-1}(\dot{\bar{F}}_\theta - 2\,\varphi_1\bar{F}_\theta)\,\bar{G}\ .$$

Using the quardratic variation process of the martingale $P^1/\bar{F}$ from Section 2, we have

$$\mathrm{Cov}(\int_0^\tau h_1\ d(P^1/\ \bar{F}))$$

$$= 16^{-1}\int_0^\tau (\dot{\bar{F}}_\theta - 2\varphi_1\bar{F}_\theta)\,\bar{G}^{\,2}(\dot{\bar{F}}_\theta - 2\varphi_1\bar{F}_\theta)^t \cdot(\bar{F}_\theta{}^2\,\bar{G})^{-1}dF_\theta$$

$$= 4^{-1}\int_0^\tau(\varphi_1\varphi_1^t - \varphi_0\varphi_1^t - \varphi_1\varphi_0^t + \varphi_0\varphi_0^t)\,\bar{G}\ dF_\theta.$$

The (i,j) entry of $\int_0^\tau \varphi_0\varphi_0\bar{G}\ dF_\theta$ can be written as

$$4^{-1}\int_0^\tau \dot{\bar{F}}_\theta^{(i)}\ \dot{\bar{F}}_\theta^{(j)}\ \bar{F}_\theta{}^{-2}\ \bar{G}\ dF_\theta$$

$$= 4^{-1}\int_0^\tau \dot{\bar{F}}_\theta^{(i)}\ \dot{\bar{F}}_\theta^{(j)}\ \bar{G}\ d(\bar{F}_\theta{}^{-1})$$

$$= 4^{-1}[\int_0^\tau \bar{G}\ d(\dot{\bar{F}}_\theta^{(i)}\ \dot{\bar{F}}_\theta^{(j)}\ \bar{F}_\theta{}^{-1})$$

$$- \int_0^\tau \bar{G}\ \bar{F}_\theta{}^{-1}d(\dot{\bar{F}}_\theta^{(i)}\ \dot{\bar{F}}_\theta^{(j)})]$$

$$= 4^{-1}[\int_0^\tau \dot{\bar{F}}_\theta^{(i)}\ \dot{\bar{F}}_\theta^{(j)}\ \bar{F}_\theta{}^{-1}dG$$

$$+ \int_0^\tau \bar{G}\ \bar{F}_\theta{}^{-1}(\ \dot{\bar{F}}_\theta^{(i)}\ \dot{f}_\theta^{(j)} + \dot{\bar{F}}_\theta^{(j)}\ \dot{f}_\theta^{(i)})d\lambda]$$

$$= (i,j)\text{ entry of }\{\int_0^\tau \varphi_0\varphi_0^t\ \bar{F}_\theta dG + \int_0^\tau \varphi_0\ \varphi_1^t\ \bar{G}\ dF_\theta + \int_0^\tau \varphi_1\varphi_0^t\ \bar{G}\ dF_\theta\ \}.$$

Thus

$$\mathrm{Cov}\ (\int_0^\tau h_1\ d(P^1/\ \bar{F}))$$

$$= 4^{-1}[\int_0^\tau \varphi_1\varphi_1^t\ \bar{G}\ dF_\theta + \int_0^\tau \varphi_0\varphi_0^t\ \bar{F}_\theta\ dG]$$

$$= 16^{-1}\ \Sigma.$$

Consequently the covariance of the limit of $n^{1/2}(\hat{\theta}_{2n} - \Psi(f;\ G;\ \tau))$ is $[4^{-1}\Sigma]^{-1}(16^{-1}\Sigma)([4^{-1}\Sigma]^{-1})^t = \Sigma.$ □

When the X_i's are distributed according to the model f_θ, the asymptotic covariance matrix of $n^{1/2}[\hat{\theta}_{2n} - \Psi(f;\ G;\ \tau)]$ is the reciprocal of the Fisher information matrix. This fact reflects a certain optimality property of the estimator $\hat{\theta}_{2n}$. For $\alpha \in L^2(R)$, let $K(d, \alpha, G)$ denote the collection of all sequences of densities $\{d_n\}$ such that

(4.20) $$\|n^{1/2}(d_n^{1/2} - d^{1/2}) - \alpha\|_2 \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Note that (4.20) implies $\alpha \perp d^{1/2}$, as is easily shown. It also implies

(4.21) $$\|n^{1/2}([L(d_n)]^{1/2} - [L(d)]^{1/2}) - \beta\|_G \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

where $\beta(x,0) = [\int_x^\infty \alpha^2 \, d\lambda]^{1/2}$, $\beta(x,1) = \alpha(x)$, and $\beta \perp [L(d)]^{1/2}$. Let $K(d, G)$ denote the union of $K(d, \alpha, G)$ for all $\alpha \in L_2(R)$, and let $\{\hat{\theta}_n\}$ be a sequence of estimators of the functional $\Psi(d; G; \tau)$ based on $(\tilde{X}_i, \delta_i)$, $i = 1, \cdots, n$. We say that $\{\hat{\theta}_n\}$ is regular at d if for $\{d_n\} \in K(d, G)$ and $X_1, \cdots, X_n$ independently and identically distributed according to d_n, $n^{1/2}[\hat{\theta}_n - \Psi(d_n; G; \tau)]$ converges weakly to a distribution $\Gamma(d; \tau; G)$ that does not depend upon the particular sequence $\{d_n\}$. The following theorem extends Theorem 5 of Beran(a, 1977) to the censored data case.

Theorem 4.4. Suppose $\Psi(\cdot; G; \tau)$ is differentiable at d with derivative ψ, in the sense that for d_n in a Hellinger neighborhood of d,

$$\Psi(d_n; G; \tau) - \Psi(d; G; \tau)$$
$$= \int_{-\infty}^{\tau} \psi\{[L(d_n)]^{1/2} - [L(d)]^{1/2}\} \, d\mu_G$$
$$+ \|[L(d_n)]^{1/2} - [L(d)]^{1/2}\|_G \, u_n,$$

where each component of $u_n \longrightarrow 0$ as $\|d_n^{1/2} - d^{1/2}\|_2 \longrightarrow 0$. Let $\{\hat{\theta}_n\}$ be a sequence of estimators of $\Psi(\cdot; G; \tau)$ which is regular at d. Then $\Gamma(d;\ \gamma;\ G)$ can be represented as the convolution of a $N(0,\ 4^{-1}\int_{-\infty}^{\gamma} \psi\ \psi^t\ d\mu_G)$ distribution with a distribution $\Gamma_1(d;\ \gamma;\ G)$.

Proof. Let

$$(4.22) \qquad L_n = 2 \prod_{i=1}^{n} \frac{[L(d_n)]^{1/2}(Z_i,\ \delta_i)}{[L(d)]^{1/2}(Z_i,\ \delta_i)} \ .$$

then we have for d_n, d in (4.20), as $n \longrightarrow \infty$,

$$(4.23) \quad P_{L(d)}[\,|L_n - 2\ n^{-1/2} \sum_{i=1}^{n} \beta(Z_i,\ \delta_i)\ [L(d)]^{-1/2}(Z_i,\delta_i) + 2 \int_{-\infty}^{\gamma} \beta^2\ d\mu_G|\ > \epsilon] \longrightarrow 0,$$

for any $\epsilon > 0$. This can be easily deduced from LeCam's second lemma and is similar to Lemma 1 of Wellner(1982). Now the rest is almost the same as in Theorem 6 in Beran(a, 1977). For any vector $v \in R^p$, the differentiability of $\Psi(\cdot;\ G;\ \gamma)$ at d and (4.20) give

$$(4.24) \qquad v^t[n^{1/2}(\Psi(d_n;\ G;\ \gamma) - \Psi(d;\ G;\ \gamma))] \longrightarrow \int_{-\infty}^{\gamma} (v^t\psi)\beta\ d\mu_G.$$

Thus we can proceed almost exactly as in Theorem 6 of Beran(1977,a): the choice $\beta = h\ v^t\psi$, $h \in R$ arbitrary, yields that along a subsequence, the random vectors

$$\{v^t[n^{1/2}(\hat{\theta}_{2n}) - \Psi(d;\ G;\ \gamma))],$$
$$n^{-1/2} \sum_{i=1}^{n} v^t\psi(Z_i,\delta_i)[L(d)]^{-1/2}(Z_i,\delta_i)\}$$

converge weakly under L(d) to $\{v^tS,\ v^tN\}$ where $N = N(0,\ \int_{-\infty}^{\gamma} \psi\psi^t\ d\mu_G)$ and S depend only on d, γ, G and not on $\{d_n\}$. Let φ denotes the characteristic function of the limit $\{v^tS,\ v^tN\}$. Then at the end we get

(4.25) $$\varphi(s,0) = \varphi(s,\ -2^{-1}s) \cdot \exp\ [-8^{-1}\{\textstyle\int_{-\infty}^{\tau} v^t\psi\psi^t v\ d\mu_G\}s^2].$$

The first factor is the characteristic function of $v^t(S - 2^{-1}N)$, the second factor is the characteristic function of $4^{-1}v^tN$. Thus the theorem follows. □

When the conclusions of of Theorem.4.2, 4.3 hold, the sequences of estimators $\{\hat{\theta}_{1n}\}$, $\{\hat{\theta}_{2n}\}$ are regular at f_θ. In fact, under $L(f_\theta)$,

(4.26) $$\begin{aligned} & n^{1/2}[\hat{\theta}_{2n} - \Psi(f_\theta;\ G;\ \tau)] \\ & \quad - 4^{-1}\int \dot{\bar{F}}_\theta\ \bar{F}_\theta^{-1}\ n^{1/2}[\hat{F}_n - F_\theta]\ dG \\ & \quad - 4^{-1}\int \dot{f}_\theta\ f_\theta^{-1}\ \bar{G}\ d\{n^{1/2}[\hat{F}_n - F_\theta]\} \\ & = o_p(1), \end{aligned}$$

as in the proof of Theorem.4.3. Since (4.24) gives contiguity of $\{L(d_n)\}$ to $\{L(f_\theta)\}$, (4.26) is also true under $L(d_n)$. Thus convergence in $D[0,\tau]$ of $P^1_{nn} = n^{1/2}(\hat{F}_n - D_n)$ to P^1 under $L(d_n)$ and the differentiability of $\Psi(\cdot;\ G;\ \tau)$ will give the regularity of $\{\hat{\theta}_{2n}\}$. Similarly we can obtain the regularity of $\{\hat{\theta}_{1n}\}$. Since with probability 1 P^1 sits in a separable subset of $D^2[0,\ \tau]$, by Theorem 5.3 in Pollard(1984) the necessary and sufficient condition for the convergence of P^1_{nn} to P^1 are the finite dimensional convergence and "small oscillation" condition. Recall the martingale representation of $P^1_n/\bar{F}$ under $L(f)$ as in Theorem 7.2.1 and Theorem 7.5.1 of S-W. We have similar representation for $P^1_{nn}/\bar{D}_n$ under $L(d_n)$. Thus convergence of P^1_{nn} on $[0,\mu]$ for any $\mu < \tau_{F_\theta}$ can be obtained by, say, Theorem 8.13 of Pollard(1984). This gives finite dimensional convergence of P^1_{nn}. Since small oscillation property is reserved under contiguity, we obtain the convergence of P^1_{nn}. Therefore $\hat{\theta}_{2n}$ is a distinguished regular estimator of $\Psi(f; G; \tau)$ for having the smallest asymptotic variance when the parametric model is true.

5. ROBUSTNESS PROPERTIES

Just as in the i.i.d. complete data case(i.e. G is degenerate at ∞) discussed in Beran (b,1977), the minimum Hellinger distance estimation procedure in the random censorship model posesses certain degree of robustness. In one way this is reflected in the continuity of $\Psi(\cdot;\ G)$; furthermore, $\Psi(f_n;\ G)$ proves to be optimally insensitive to perturbations of its argument in a minimax sense. Consider the class of functionals {U} such that for ρ a p-dimensional vector with components $\rho^{(i)}$ in $L_2(\mu)$.

(5.1) $$U(f_\theta) = \theta,$$

$$U(f) - \theta = \int \rho([L(f)]^{1/2} - [L(f_\theta)]^{1/2})\, d\mu$$
$$+\ o(\|[L(f)]^{1/2} - [L(f_\theta)]^{1/2}\|_G)u_n,$$

where each component of $u_n \longrightarrow 0$ as $[L(f)]^{1/2} \longrightarrow [L(f_\theta)]^{12}$ in $L_2(\mu)$. We can assume for each i, $\rho^{(i)} \perp s_\theta$ in $L_2(\mu)$, since otherwise we can replace ρ by

$$\tilde{\rho} = \rho - \{\int \rho\ s_\theta\ d\mu\}s_\theta,$$

with the difference caused by the replacement being absorbed into the remainder term in (5.1). Also for each unit vector e_i in R^p, scalar $\alpha \neq 0$, we have

$$e_i = a^{-1}[U(f_{\theta+\alpha e_i}) - U(f_\theta)] \longrightarrow [\int \rho \dot{s}_\theta^{(i)}\ d\mu_n] \quad \text{as } \alpha \longrightarrow 0.$$ So

$\int \rho\ \dot{s}_\theta\ d\mu = I$, the identity matrix. When Theorem 3.1 applies, $\{\Psi(\cdot;\ G)\}$ belongs to this class. One may be interested in seeing which functional in the class just described is least affected by infinitesimal perturbations of f_θ, at least asymptotically. To this end, let us examine the behavior of $c^t[U(f) - \theta]$, for every constant vector $c \in R^p$. By projection, $[L(f)]^{1/2}$ can be represented as

$$[L(f)]^{1/2} = \cos \gamma\ s_\theta + \sin \gamma\ \delta,$$

where $\gamma \in [0, 2^{-1}\pi]$, $\|\delta\|_G = 1$, $\int \delta\ s_\theta\ d\mu = 0$. Then

(5.2) $\quad U(f) - \theta = \gamma \int \rho\ \delta\ d\mu + o(\gamma) \quad$ as $\gamma \longrightarrow \infty$.

Thus for small γ, or equivatently, small $\|[L(f\ )]^{1/2} - s_\theta\|_G$, the behavior of $|c^t[U(f) - \theta]|$ is primarily determined by $|\int \rho\ \delta\ d\mu| = L_c(\rho,\ \delta)$. Thus the problem becomes: which ρ minimizes for each $c \in R^p$ the deviation L_c, against all possible direction δ? It turns out that $\Psi(\cdot;\ G)$ corresponds to the optimal choice of ρ, as the following result shows.

Theorem 5.1. **Suppose** s_θ **satisfies the conditions in Lemma** 3.2. **For each** i, $\rho^{(i)} \in L_2(\mu)$, $\int \rho^{(i)} s_\theta\ d\mu = 0$, $\int \rho\ \dot{s}_\theta^{\ t} d\ \mu = I$, $\|\delta\|_G = 1$, $\int \delta\ s_\theta\ d\mu = 0$.

Then for every $v \in R^p$,

(5.3)
$$\min_\rho \max_\delta L_c(\rho,\ \delta) = \max_\delta \min_\rho L_c(\rho,\ \delta) = L_c(\ \rho^0,\ \delta^0),$$

where

(5.4)
$$\rho^0 = [\int \dot{s}_\theta \dot{s}_\theta^{\ t}\ d\mu]^{-1} \dot{s}_\theta,$$
$$\delta^0 = \|c^t \rho\|_G^{-1}\ c^t\ \rho^0.$$

Proof. The proof is just as in Beran(b,1977). It suffices to show $\max_\delta \min_\rho L_c \geq \min_\rho \max_\delta L_c$, since the reverse inequality is trivial.

By the Cauchy - Schwartz inequality,

$$\max_\delta L_c (\rho, \delta) = \|c^t \rho\|_G , \tag{5.5}$$

at $\delta = \|c^t \rho\|_G^{-1} c^t \rho$. Since $\int \rho \dot{s}_\theta^t d\mu = I$ and $\int \rho^{(i)} \dot{s}_\theta d\mu = 0$ for each $i = 1, \cdots, p$, we can decompose $\rho = A \dot{s}_\theta + \sigma$, where $\int \sigma^t \dot{s}_\theta d\mu = 0$ and $\int \delta s_\theta d\mu = 0$. Then $I = \int \rho \dot{s}_\theta^t d\mu$ implies $A = [\int \dot{s}_\theta \dot{s}_\theta^t d\mu]^{-1}$. So

$$\rho = \rho^0 + \sigma, \quad \int \sigma^t \rho^0 d\mu = 0, \tag{5.6}$$

and therefore

$$\min_\rho \max_\delta L_c(\rho, \delta) = \min_\sigma \|c^t \rho + c^t \sigma\|_G = \|c^t \rho^0\|_G .$$

On the other hand,

$$\begin{aligned} \max_\delta \min_\rho L_c(\rho, \delta) &\geq \min_\rho L_c(\rho, \delta^0) \\ &= \|c^t \rho^0\|_G^{-1} \min_\rho |c^t \int \rho (\rho^0)^t d\mu \, c| \\ &= \|c^t \rho^0\|_G^{-1}, \text{ by (5.6)}. \end{aligned}$$

Hence the theorem follows. □

REFERENCES

1. Beran, R. J. (a,1977) Robust location estimates. Ann. Statist. 5 431-444

2. Beran, R. J. (b,1977) Minimum Hellinger distance estimates for parametric models. Ann. Statist. 5 445-463

3. Beran, R. J. (1981) Efficient robust estimation for parametric models. Z. Wahr. 55 91-108

4. Billingsley, P. (1968) Weak Convergence of Probability Measures. Wiley.

5. Brown, L. D. and Purves, F. (1973) Measurable selections of extrema. Ann. Statist. 3 902-912

6. Fabian, V. and Hannan, J. (1985) Introduction to Probability and Mathematical Statistics. Wiley.

7. Gill, R. (1980) Censoring and Stochastic Integrals. Mathematical Centre Tracts 124 Mathematisch Centrum, Amsterdam.

8. Gill, R. (1983) Large sample behavior of the product-limit estimator on the whole line. Ann. Statist. 11 49-58.

9. Hajek, J. (1972) Locally asymptotic minimax and admissibility in estimation. Sixth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1 175–194.

10. Hewitt, E. and Stromberg, K. (1965) Real and Abstract Analysis. Springer–Verlag, New York.

11. Kaplan, E. and Meier, P. (1958) Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53 457-481

12. Millar, P.W. (1983) The Minimax Principle in Asymptotic Statistical Theory. Lecture Notes in Mathematics. 976 75-265 Springer–Verlag, New York.

13. Pollard, D. (1984) Convergence of Stochastic Processes. Springer–Verlag, New York.

14. Rudin, W. (1974) Real and Complex Analysis. second edition. McGraw–Hill.

15. Shorack, G. R. and Wellner, J. A. (1986) Empirical

Processes with Applications to Statistics. Wiley.

16. Wang, J. G. (1987) A note on the uniform consistency of the Kaplan-Meier estimator. Ann. Statist. 15 1313-1316

17. Wellner, J. A. (1982) Asymptotic optimality of the product limit estimator. Ann. Statist. 10 595-602