The Annals of Statistics
1991, Vol. 19, No. 2, 579–602

MINIMUM HELLINGER DISTANCE ESTIMATION OF PARAMETER IN THE RANDOM CENSORSHIP MODEL

BY SONG YANG

Texas Tech University

This paper discusses the minimum Hellinger distance estimation (MHDE) of the parameter that gives the "best fit" of a parametric family to a density when the data are randomly censored. In studying the MHDE, the tail behavior of the product-limit (P-L) process is investigated, and the weak convergence of the process on the real line is established. An upper bound on the mean square increment of the normalized P-L process is also obtained. With these results, the asymptotic behavior of the MHDE is established and it is shown that, when the parametric model is correct, the MHD estimators are asymptotically efficient among the class of regular estimators. This estimation procedure is also minimax robust in small Hellinger neighborhoods of the given parametric family. The work extends the results of Beran for the complete i.i.d. data case to the censored data case. Some of the proofs employ the martingale techniques by Gill.

1. Introduction. Let $X_1, \dots, X_n$ be i.i.d. random variables with life-time c.d.f. F on $[0, \infty)$, and $Y_1, \dots, Y_n$ be independent of X_i's and i.i.d. with censoring c.d.f. G on $[0, \infty]$ (i.e., G may assign positive mass to ∞). In the random censorship model, the pairs $\{\min(X_i, Y_i), [X_i \le Y_i]\}$, $1 \le i \le n$, are observed, where $[A]$ denotes the indicator function of the event A. Suppose that F has a density f with respect to the 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.

There are many results in the literature for the case where G is degenerate at ∞, that is, when we are able to observe the complete data $X_1, \dots, X_n$. Millar (1983) illustrates 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 estimators usually have 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 is no contamination, as discussed in Beran (1977b, 1981). It is heuristically illustrated in Beran (1977b) that the minimum Hellinger distance estimator

Received July 1988; revised March 1990.
AMS 1980 *subject classifications*. Primary 62F35, 62E20; secondary 60F05, 62P10.
Key words and phrases. Product-limit estimator, censored data, martingale, stochastic integral, weak convergence, minimum Hellinger distance estimate, asymptotically efficient estimate, robust statistics, minimax robust.

considered there is closely related to the maximum likelihood estimator and therefore asymptotic efficiency seems plausible.

In this paper, 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 (1977b), 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 Section 2, some preliminary results are introduced. The tail behavior of the product-limit (P-L) process is investigated and the weak convergence of the process on the entire support set is 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 P-L process is developed. In Section 3, the differentiability of the minimum Hellinger distance functional is studied. In Section 4, the asymptotic behavior of the MHDE is investigated and it is shown that this procedure is asymptotically efficient if there is no contamination. In Section 5, a minimax robustness property of the MHDE is briefly discussed and some numerical simulation results are reported in a simple exponential example.

NOTATIONAL REMARKS. Throughout this paper, $X_1, \dots, X_n, Y_1, \dots, Y_n$ are independent r.v.'s. Unless mentioned otherwise, for $i = 1, \dots, n$, X_i, Y_i have distributions F, G, respectively. $\delta_i = [X_i \le Y_i]$ 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, $D^{-1}(t) = \inf\{u\colon D(u) \ge t\}$, $\tau_D = D^{-1}(1) \le \infty$, and $\overline{D} = 1 - D$, $\Delta D = D - D_-$. Note that $\overline{H} = \overline{F}\overline{G}$ and $\tau_H = \min(\tau_F, \tau_G)$. Abbreviate τ_H to τ. Let R denote the real line and $O_p(1)$ denote any sequence of r.v.'s bounded in probability. $\int_s^t = \int_{(s,t]}$ for $s > 0$ and $\int_0^t = \int_{[0,t]}$.

2. Preliminaries. For inference with censored data, the P-L estimator [Kaplan and Meier (1958)] has many optimal properties and lends itself readily to an analysis using counting processes and stochastic integrals. We will use the P-L estimator in constructing our estimators in this paper. We start by investigating the asymptotic behavior of the P-L process. Since we need to estimate the life-time distribution and the censoring distribution simultaneously, there will actually be two P-L processes.

Assuming the life-time distribution function to be continuous, Gill (1983) considers the stopped Kaplan–Meier processes and obtains their convergence in Skorohod topology on the whole real line. In Chapter 7 of their book, Shorack and Wellner (1986) use the uniform topology and hence allow the life-time distribution function to be possibly discontinuous. To allow a possibly discontinuous censoring distribution, we will use the uniform topology. Let $D^2[0,\tau]$ be the twofold product of the usual space of *cadlag* functions on $[0,\tau]$, with the sup norm $\|z\| = \|x\| + \|y\|$ for $z = (x, y)$ and the σ-field $\mathscr{P}$ generated by open balls. Define a random element of $D^2[0,\tau]$ to be a $\mathscr{P}$-measurable mapping from some probability space. We say random elements $\{W_n\}_{n=1}^{\infty}$

converge weakly to W, or $W_n \to W$ in $\mathscr{U}^2[0,\tau]$, if $Ef(W_n) \to Ef(W)$ for any bounded, continuous and $\mathscr{P}/\mathscr{B}^2$ measurable function f, where $\mathscr{B}^2$ denotes the Borel σ-field on R^2. We will also use similar notation for one-dimensional processes in $D[0,\tau]$.

Now formally define the product-limit estimators $\hat{F}_n, \hat{G}_n$ by

$$1 - \hat{F}_n(t) = \prod_{s \le t} \left[1 - \frac{\Delta H_n^1(s)}{\overline{H}_{n-}(s)}\right],$$

$$1 - \hat{G}_n(t) = \prod_{s \le t} \left[1 - \frac{\Delta H_n^0(s)}{\overline{H}_{n-}(s)}\right],$$

where $H_n = H_n^0 + H_n^1$ and H_n^0, H_n^1 are the basic empirical processes in $D^2[0,\tau]$ for the censored data:

$$H_n^j(t) = n^{-1} \sum_i \left[\tilde{X}_i \le t, \delta_i = j\right], \qquad j = 0, 1.$$

Upon inspecting Shorack and Wellner's (1986) discussion on $\hat{F}_n$, one finds that $\hat{G}_n$ only estimates G^* satisfying $G^*(t) = \int_0^t \overline{G}_-^* / \overline{H}_- \, dH^0$, which is identical with G on $[0,\tau]$ if and only $\int_0^\infty (\Delta F)\, dG = 0$. For the simultaneous estimation of F and G, define the process $M_n = (M_0^1, M_n^0)$ in $D^2[0,\tau]$ by $M_n^j(t) = n^{1/2}[H_n^j(t) - \int_0^t \overline{H}_{n-}\, d\Lambda^j]$, $j = 0, 1$, where $\Lambda^1(t) = \int_0^t (1/\overline{F}_-)\, dF$ and $\Lambda^0(t) = \int_0^t (1/\overline{G}_-)\, dG$. Then under the assumption $\int_0^\infty (\Delta F)\, dG = 0$, for the complete σ-field σ_t^n generated by $\{[\tilde{X}_i \le s]\delta_i, [\tilde{X}_i \le s]: 1 \le i \le n, 0 \le s \le t\}$, one can check that $\{M_n(t), \sigma_t^n: 0 \le t < \tau\}$ is a two-dimensional square integrable martingale with mean 0, predictable variation processes $\langle M_n^j \rangle(t) = \int_0^t \overline{H}_{n-}(1 - \Delta\Lambda^j)\, d\Lambda^j$, $j = 0, 1$, and predictable covariation process $\langle M_n^0, M_n^1 \rangle = 0$.

Now define the P-L processes

$$P_n^1 = n^{1/2}(\hat{F}_n - F), \qquad P_n^0 = n^{1/2}(\hat{G}_n - G)$$

and let $P^1 = \overline{F}B^1(C^1)$, $P^0 = \overline{G}B^0(C^0)$, where B^1, B^0 are two independent standard Brownian motions and

$$C^1(t) = \int_0^t (\overline{F}\overline{G}_-)^{-1}\, d\Lambda^1, \qquad C^0(t) = \int_0^t (\overline{G}\overline{F}_-)^{-1}\, d\Lambda^0. \tag{2.1}$$

Define a stopping time $T = \max \tilde{X}_i$. Let $R^T(t) = [t \le T]R(t) + [t > T]R(T)$ for any process R. The following theorem establishes the convergence of $\{P_n^j\}_{n=1}^\infty$, $j = 0, 1$. Notice that the convergence of $\{P_n^1\}_{n=1}^\infty$ is on $[0,\tau]$ and free of weight. One naturally hopes the same holds for $\{P_n^0\}_{n=1}^\infty$, but that would require some contradicting conditions. At τ the values of the limiting processes are interpreted to be their limits as $t \uparrow \tau$ [cf. Remark 2.2 in Gill (1983)].

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

(i) *If* $\Delta G(\tau) = 0$, *then for any* $\alpha \in (0, 1/2)$,

$$(2.2)\qquad \left(\bar{F}^{1-\alpha}P_n^0\right)^T \to \bar{F}^{1-\alpha}P^0 \quad in\ \mathscr{U}[0,\tau].$$

(ii) *If* $\Delta F(\tau) = 0$ *and* $A(\tau) < \infty$, *where*

$$(2.3)\qquad A(t) = \int_0^t (1/\bar{G}_-)\,dF,$$

then

$$(2.4)\qquad P_n^1 \to P^1 \quad in\ \mathscr{U}[0,\tau].$$

When all three assumptions hold, the joint convergence is valid.

PROOF. Let $H^j(t) = P[\tilde{X}_1 \le t,\ \delta_1 = j]$ for $j = 0,1$. Define the process $Q_n = (Q_n^1, Q_n^0)$ by $Q_n^j = n^{1/2}(H_n^j - H^j)$, $j = 0,1$. As in Shorack and Wellner (1986), the convergence of the P-L processes can be derived from the convergence of $\{Q_n\}_{n=1}^\infty$. Let $Q = (V^1(H^1), V^0(H^0))$, where V^1 and V^0 are Brownian bridges, with convariance $\mathrm{Cov}(V^1(H^1(s)), V^0(H^0(t))) = -H^1(s)H^0(t)$. Shorack and Wellner (1986) state that the result $Q_n \to Q$ in $\mathscr{U}^2[0,\tau]$ can be proved by a minor variation of their theory of ordinary empirical processes. Alternatively, its proof can be based on Theorem 5.5 of Pollard (1984), with a proper modification. The finite-dimensional convergence part is straightforward. For the small oscillation condition [cf. Pollard (1984), 5.4], we can show, using an argument similar to the proof of Theorem 13.1 in Billingsley (1968), that for any $a > b$, and $\varepsilon > 0$, there exists a constant K_ε depending only on ε, such that for $i = 0, 1$,

$$\limsup_{n\to\infty} P\left[\sup_{t\in[a,b)} \left|Q_n^i(t) - Q_n^i(a)\right| > \varepsilon\right] \le K_\varepsilon\left(H_-^i(b) - H^i(a)\right)^2.$$

From this the small oscillation property, and hence the convergence of $\{Q_n\}_{n=1}^\infty$, follow. Now the result (2.2) is immediate from (9) of Theorem 7.7.1 in Shorack and Wellner (1986) by taking their q-function to be $q(t) = t^\alpha$ on (0, 1/2] and using (1.2) of Gill (1983). Notice that the θ in Theorem 7.4.2 of Shorack and Wellner (1986) should be confined to be a continuity point of H. To prove (2.4), replace the role of $(1 - K_D)/q(K_D)$ in Shorack and Wellner (1986) by $\bar{F}$ and use a similar argument and (2.3) to obtain $(P_n^1)^T \to P^1$ in $\mathscr{U}[0,\tau]$. Then observe that $\sup|(P_n^1)^T - P_n^1| \le n^{1/2}(F(\tau) - F(T))$, hence by the assumption $\Delta F(\tau) = 0$ it only remains to show $n^{1/2}(F(\tau-) - F(T)) \to_p 0$. Note that for the function A as defined in (2.3), for all $t < \tau$,

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

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

$$n^{1/2}(F(\tau -) - F(T)) \le \left\{ n\overline{H}(T) \int_{(T,\tau)} dA \right\}^{1/2}.$$

Since we can show that $n\overline{H}(T)$ is $O_p(1)$ using $HH^{-1}(x) \le x + \Delta H(H^{-1}(x))$, by (2.3) we have (2.4). The joint convergence follows from (2.2), (2.4) and the fact $\langle M_n^1, M_n^0 \rangle = 0$. □

In their Theorem 7.1.1, Shorack and Wellner (1986) use the convergence of $\{Q_n\}_{n=1}^{\infty}$, with an almost sure representation for a special construction of X_i's and Y_i's. They state that such a special construction can be obtained by a minor variation of their theory for ordinary empirical processes, i.e., by "rebuilding" the random variables from the process. While the special construction of X_i's and Y_i's from Q_n may not seem obvious, a special construction of Z_i's and δ_i's is sufficient. These Z_i, δ_i can be rebuilt from Q_n by the measurability of the projections π_j: $(x_1, x_2) \to x_j$ from $D^2[0,\tau]$ to $D[0,\tau]$, $j = 1, 2$. Alternatively, the convergence of P_n^0, P_n^1 can be proved using the differentiability of those functions defining P_n^0, P_n^1 on the basis of Q_n [cf. Gill (1989)]. This avoids the special construction but needs some nontrivial measurability and differentiability arguments.

As consequences of Lemma 2.1, we make the following observations for later use:

$$\sup_{[0,\tau]} |P^1| = O_p(1), \qquad \sup_n \sup_{[0,\tau]} |P_n^1| = O_p(1) \tag{2.5}$$

and for any $\alpha \in (0, 1/2)$,

$$\sup_{[0,\tau]} |P^0 \overline{F}^{1-\alpha}| = O_p(1), \qquad \sup_n \sup_{[0,T]} |P_n^0 \overline{F}^{1-\alpha}| = O_p(1), \qquad P_n^0 \overline{F}^{1-\alpha}(T) \to_p 0.$$

Now let us use $\hat{F}_n$ to construct the kernel density estimator

$$f_n(x) = a_n^{-1} \int_R K((x-y)/a_n)\, d\hat{F}_n(y), \tag{2.6}$$

where K is some kernel function and a_n is some positive constant. We will need the following theorem on the convergence of f_n to f in the Hellinger metric. Note that if f is truncated at τ, then the result is also true without the assumption $\tau = \tau_F$. But in later use the following form is what we need.

THEOREM 2.1. *Suppose*

(i) $\int_0^\tau (1/\overline{G}_-)\, dF < \infty$ *and* $\tau = \tau_F$,

(ii) F *has a continuous density* f,

(iii) K *is nonnegative, continuous and of bounded variation on* R, $\int_R K(s)\, ds = 1$, $K(s) \to 0$ *as* $s \to -\infty$,

(iv) $a_n \to 0$ *and* $n^{1/2} a_n \to \infty$. *Then*

$$\| f_n^{1/2} - f^{1/2} \|_2 \to_p 0,$$

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

PROOF. Integration by parts [cf. Hewitt and Stromberg (1965), page 419] gives us

$$(2.7)\qquad \begin{aligned} f_n(x) - f(x) &= \left(n^{1/2}a_n\right)^{-1}\int_R P_n^1(x - a_n t)\, dK(t) \\ &\quad + \int_R [f(x - a_n t) - f(x)]K(t)\, dt. \end{aligned}$$

The random term is uniformly bounded by $(n^{1/2}a_n)^{-1}\sup_{[0,\tau]}|P_n^1|\,|\int_R dK|$, which goes to 0 in probability by (2.5). Since $\int_R\int_R f(x - a_n)K(t)\,dx\,dt = \int_R\int_R f(t)\,dx\,dt = 1$ and f is continuous, the nonrandom term converges to 0 in Lebesgue measure. Also $\|f_n^{1/2}\|_2^2 = \hat{F}_n(T) = o_p(1) + F(T) \to_p 1 = \|f^{1/2}\|_2^2$. From these facts and the subsequence characterization of convergences in probability and in measure the desired result follows. □

Later in Section 4, we will use a proper Taylor expansion in studying our estimators. To deal with higher-order terms there, we need to control the increments of the process P_n^1. When there is no censoring, direct calculation gives the mean square increment of the empirical process. In our case we have the following lemma, which gives an inequality for the mean square increment of the process $Z_n = n^{1/2}(\hat{F}_n - F)/\overline{F}$ in terms of the function A defined in (2.3). Let $J_n(t) = [t \le T]$.

LEMMA 2.2. *Suppose F is continuous. Then for $0 \le s < t < \tau_F$,*

$$E\left[Z_n^T(t) - Z_n^T(s)\right]^2 \le 4\overline{F}^{-2}(t)[A(t) - A(s)] + 4A(s)\left[\overline{F}^{-2}(t) - \overline{F}^{-2}(s)\right].$$

PROOF. By Lemma 2.5 in Gill (1983), for continuous F and any $\rho < \tau_F$, Z_n^T is a square integrable martingale on $[0, \rho]$ with predictable variation process $\langle Z_n^T\rangle(x) = \int_0^x[(1 - \hat{F}_{n-})/\overline{F}]^2 J_n/\overline{H}_{n-}\, d\Lambda^1$. Let $L(x) = E\langle Z_n^T\rangle(x)$. Then, since $n\overline{H}_{n-} \ge 1$, we have

$$\begin{aligned} L(x) &= E\int_0^x \left[1 - n^{-1/2}Z_{n-}^T\right]^2 J_n/\overline{H}_{n-}\, d\Lambda^1 \\ &\le 2\int_0^x E\left(J_n/\overline{H}_{n-}\right) d\Lambda^1 + 2\int_0^x E\left[Z_{n-}^T\right]^2 J_n\, d\Lambda^1. \end{aligned}$$

Note that $\overline{H}_{n-}(x)$ is binomial, so

$$\begin{aligned} E\left(J_n/\overline{H}_{n-}\right)(x) &= \sum_{k=1}^{n} \frac{n}{k}\binom{n}{k}\overline{H}_-^k H_-^{n-k}(x) \\ &\le 2\sum_1^n \frac{n+1}{k+1}\binom{n}{k}\overline{H}_-^k H_-^{n-k}(x) \le 2\left(\overline{H}_-(x)\right)^{-1}. \end{aligned}$$

Also by Fatou's lemma, for a sequence of real $x_k \uparrow x$,

$$E\big[Z_{n-}^T(x)\big]^2 J_n(x) \le E\big[Z_{n-}^T(x)\big]^2 \le \liminf_{k\to\infty} E\big[Z_n^T(x_k)\big]^2 = \liminf_{k\to\infty} E\langle Z_n^T\rangle(x_k) \le L(x).$$

Therefore, we obtain

$$L(x) \le \alpha(x) + \int_0^x L\,d\beta, \tag{2.8}$$

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

Now we use the argument as in Gronwall's lemma: Iterating m times in (2.8) and letting $m \to \infty$ yields

$$\begin{aligned} L(x) &\le \alpha(x) + \int_0^x e^{\beta(x)-\beta(y)}\alpha(y)\,d\beta(y) \\ &= 4\bar{F}^{-2}(x)A(x), \end{aligned} \tag{2.9}$$

where the last equation follows from Fubini's theorem. Now

$$\begin{aligned} &E\big[Z_n^T(t) - Z_n^T(s)\big]^2 \\ &\quad= E\big[\langle Z_n^T\rangle(t) - \langle Z_n^T\rangle(s)\big] \\ &\quad\le 2\int_s^t E(J/\bar{H}_{n-})\,d\Lambda^1 + 2\int_s^t E\big[Z_{n-}^T\big]^2\,d\Lambda^1 \\ &\quad\le 4\int_s^t (1/\bar{H}_-)\,d\Lambda^1 + 2\int_s^t L\,d\Lambda^1 \\ &\quad\le 4\bar{F}^{-2}(t)\big[A(t) - A(s)\big] + 4A(s)\big[\bar{F}^{-2} - \bar{F}^{-2}(s)\big], \end{aligned}$$

where in the last inequality we have used (2.9) and Fubini's theorem. □

From Lemma 2.2 and using $F(t) - F(s) < \bar{F}(s)$ to simplify, we have for $0 \le s < t < \tau_F$,

$$\begin{aligned} &E\big[P_n^1(t) - P_n^1(s)\big]^2[t < T] \\ &\quad\le 2\bar{F}^2(t)E\big[Z_n^T(t) - Z_n^T(s)\big]^2 + 2\big[\bar{F}(t) - \bar{F}(s)\big]^2 E\big[Z_n^T(s)\big]^2 \\ &\quad\le 8\big[A(t) - A(s)\big] + 24A(\tau)\big[F(t) - F(s)\big]\bar{F}^{-1}(s). \end{aligned} \tag{2.10}$$

In Section 4, we will use this more convenient form.

3. The minimum Hellinger distance functional. Our estimator of θ will be defined as the value of some functional at the kernel density estimator and the P-L estimator. Having established some convergence results for those estimators in the previous section, we define and study the differentiability of the functional here. The approach will be similar to that in Beran (1977b).

While one naive attempt might be to use exactly the same MHD functional as in Beran's paper and evaluate the functional at f_n, defined in (2.6), to get the estimator, Beran's heuristic argument about the relationship between the MHDE and the maximum likelihood estimator leads us to consider the likelihood function in the censored data case. Suppose F has a density f with respect to the Lebesgue measure λ. We then have

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

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

Thus $(\tilde{X}_1, \delta_1)$ has a density $\overline{F}^{1-y}(x) f^y(x)$ with respect to the measure μ_G on $R \times \{0, 1\}$, where μ_G is defined by the relation

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

for any nonnegative measurable function m on $R \times \{0, 1\}$. In view of this and the fact that $\hat{F}_n, \hat{G}_n$ may be defective, we define, for any subdensity d on R w.r.t. λ, a subdensity $L(d)$ on $R \times \{0, 1\}$ w.r.t. μ_G by

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

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

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

where $\| \cdot \|_G$ denotes the L^2-norm in $L^2(\mu_G)$. Due to the fact the $G(t)$ can only be estimated for t up to τ, we need to consider a more complicated variation of the above definition: For $0 \le \gamma \le \infty$, define $\Psi(\cdot; G; \gamma)$ similarly restricting all integration to $x \in (-\infty, \gamma]$. Later we will use $\|\{\cdot\}(-\infty, \gamma]\|_G$ to denote the norm under the restricted integration. Note that $\Psi(\cdot; \cdot; \infty) = \Psi(\cdot; \cdot)$. For subdensities f, f_n on R w.r.t. λ and (sub-) c.d.f. G_n, G and γ^n such that

$$\sup_{(-\infty, \gamma]} |G_n - G| \to 0 \quad \text{and} \quad \gamma^n \uparrow \gamma, \tag{3.1}$$

we will use the notation $\theta_0 = \Psi(f; G; \gamma)$, $\theta_n = \Psi(f; G_n; \gamma)$, $\theta_{nn} = \Psi(f; G_n; \gamma^n)$ and $\mu = \mu_G$, $\mu_n = \mu_{G_n}$. Also, we will use $\int_{-\infty}^{\gamma} d\mu$ to denote the integral on $(x, y) \in (-\infty, \gamma] \times \{0, 1\}$ and F_θ the c.d.f. of f_θ. To simplify the notation, we only look at the case when the parameter is one dimensional. For the multidimensional case we have parallel results.

LEMMA 3.1. *Suppose Θ is a compact subset of R, $\theta \neq \theta'$ implies $f_\theta \neq f_{\theta'}$ 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, subdensity function f and $0 \le \gamma \le \infty$, $\Psi(f; G; \gamma)$ exists,*

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

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

(iii) *for G_n, G and γ^n satisfying (3.1) and $\Delta G(\gamma) = 0$, $\|f_n^{1/2} - f^{1/2}\|_2 \to 0$ implies $\Psi(f_n; G_n; \gamma^n) \to \Psi(f; G; \gamma)$ if $\Psi(f; G; \gamma)$ is unique.*

PROOF. By the assumption $F_\theta(x)$ is continuous in θ for fixed x, thus (i) can be proved as in Theorem 1 of Beran (1977b). (ii) is obvious. To prove (iii), first note that

$$\left\|[L(f_n)]^{1/2} - [L(f)]^{1/2}\right\|_G \to 0 \quad \text{as } \|f_n^{1/2} - f^{1/2}\|_2 \to 0.$$

This is because

$$\sup_R \left[\bar{F}_n^{1/2} - \bar{F}^{1/2}\right]^2 \le \sup_R |F_n - F| \le \int_R |f_n - f| \le 2\left\{\int_R \left[f_n^{1/2} - f^{1/2}\right]^2\right\}^{1/2}.$$

Now define N by

$$N(\theta, f) = \left\|\left\{[L(f_\theta)]^{1/2} - [L(f)]^{1/2}\right\}(-\infty, \gamma]\right\|_G$$

and define $N_n(\theta, f)$ similarly using γ^n, G_n. By the triangle inequality, we have

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

The second term is dominated by a constant multiple of $\sup_{(-\infty,\gamma]}|\bar{G}_n - \bar{G}|$ and so is the third term from the integration by parts formula. Thus for G_n, G and γ^n satisfying (3.1), $N_n(\theta, f_n) - N_n(\theta, f) \to 0$ uniformly in θ as $\|f_n^{1/2} - f^{1/2}\|_2 \to 0$. Similarly the triangle inequality and the additional assumption give us, for G_n, G and γ^n satisfying (3.1) and $\Delta G(\gamma) = 0$, $N_n^2(\theta, f) - N^2(\theta, f) \to 0$ uniformly in θ as $\|f_n^{1/2} - f^{1/2}\|_2 \to 0$. From these two uniform convergences it follows that, for G_n, G and γ^n satisfying (3.1) and $\Delta G(\gamma) = 0$, $N_n(\theta, f_n) - N(\theta, f) \to 0$ uniformly in θ as $\|f_n^{1/2} - f^{1/2}\|_2 \to 0$. As in Beran (1977b), from this, compactness of Θ, continuity of $N(\theta, f)$ in θ and uniqueness of $\Psi(f; G; \gamma)$, one has $\theta_{nn} \to \theta_0$, that is, $\Psi(f_n; G_n; \gamma^n) \to \Psi(f; G; \gamma)$. □

To study the asymptotic behavior of the minimum Hellinger distance functional, we need to establish the following expansion for $s_\theta \equiv [L(f_\theta)]^{1/2}$.

When the first-order partial derivative of f_θ w.r.t. θ exists, we will denote it by $\dot{f}_\theta$; when the second-order partial derivative of $\dot{f}_\theta$ w.r.t. θ exists, it will be denoted by $\ddot{f}_\theta$.

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 $\ddot{f}_\theta(x)$ is continuous in θ for $x \notin N$, where N is a λ-null set,*

(ii) $U_1(\theta) \equiv \int [\dot{f}_\theta]^4/f_\theta^3 \, d\lambda$, $U_2(\theta) \equiv \int [\ddot{f}_\theta]^2/f_\theta \, d\lambda$ *are continuous on V,*

(iii) *for some $\varepsilon, \delta > 0$,*

$$V_1(\theta) \equiv \int \frac{|\ddot{f}_\theta|^{2+\varepsilon}}{f_\theta^{1+\varepsilon}} \, d\lambda, \qquad V_2(\theta) \equiv \int \frac{|\dot{f}_\theta|^{4+\delta}}{f_\theta^{3+\delta}} \, d\lambda,$$

$$W_1(\theta) \equiv \int -\frac{|\ddot{f}_\theta|}{f_\theta} \, d\bar{F}_\theta^{1/2} \quad \text{and} \quad W_2(\theta) \equiv \int -\frac{|\dot{f}_\theta|}{f_\theta} \, d\bar{F}_\theta^{1/4}$$

are bounded in a neighborhood of ρ. Then, for G_n, G and γ^n in (3.1), *ρ_n in a neighborhood of ρ and $(x, y) \notin N_0 \times \{1\}$, where N_0 is a λ-null set,*

$$s_{\rho_n} = s_\rho + (\dot{s}_\rho + r_n)(\rho_n - \rho), \tag{3.2}$$

$$\dot{s}_{\rho_n} = \dot{s}_\rho + (\ddot{s}_\rho + R_n)(\rho_n - \rho), \tag{3.3}$$

where both $\|r_n(-\infty, \gamma^n]\|_{G_n}$ and $\|R_n(-\infty, \gamma^n]\|_{G_n}$ tend to 0 as $\rho_n \to \rho$.

PROOF. Here we just prove (3.2). The argument for (3.3) is similar and more involved. Note that the assumptions imply that $U(\theta) \equiv \int [\dot{f}_\theta]^2/f_\theta \, d\lambda$ is continuous on V, $\int |\dot{f}_\theta|^{2+\varepsilon}/f_\theta^{1+\varepsilon} \, d\lambda$ and $-\int |\dot{f}_\theta|/f_\theta \, d\bar{F}_\theta^{1/2}$ are bounded in a neighborhood of ρ. These three conditions are sufficient for (3.2). Write

$$\int |\dot{f}_\theta(s)| \, ds = \int \{ |\dot{f}_\theta(s)| f_\theta^{-1/2}(s) \} f_\theta^{1/2}(s) \, ds.$$

Then by the Cauchy–Schwarz inequality and the generalized dominated convergence theorem [cf. Fabian and Hannan (1985), page 32, for the GDCT], $\int |\dot{f}_\theta(s)| \, ds$ is finite and continuous. Integrating out $f_\theta(s)$ shows that, 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. Now from

$$\frac{\left[\dot{\bar{F}}_\theta(x)\right]^2}{\bar{F}_\theta(x)} \le \int_x^\infty \frac{\dot{f}_\theta^2}{f_\theta}(s) \, ds, \tag{3.4}$$

$$\int_R \int_x^\infty \frac{[\dot{f}_\theta]^2}{f_\theta} \, ds \, dG(x) = \int G \frac{[\dot{f}_\theta]^2}{f_\theta} \, d\lambda \tag{3.5}$$

and the proof of Lemma A.2 in Hájek (1972), we obtain that there is a neighborhood of ρ, on which s_θ is absolutely continuous in θ for $(x, y) \notin N_0 \times \{1\}$, where N_0 is a λ-null set. For those (x, y) and ρ_n in a neighborhood of ρ

we have

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

where

$$\begin{aligned}\|r_n(-\infty,\gamma]\|_G^2 &= \left\| \left\{ \frac{1}{\rho_n - \rho} \int_\rho^{\rho_n} (\dot{s}_t - \dot{s}_\rho)\, dt \right\} (-\infty,\gamma] \right\|_G^2 \\ &\le \left| \frac{1}{\rho_n - \rho} \int_\rho^{\rho_n} \|\{\dot{s}_t - \dot{s}_\rho\}(-\infty,\gamma]\|_G^2\, dt \right| \\ &= \|\{\dot{s}_{\xi_n} - \dot{s}_\rho\}(-\infty,\gamma]\|_G^2\end{aligned}$$

for some ξ_n between ρ_n and ρ. Thus to prove (3.2) it suffices to prove that $Z(G_n) \equiv \|\{\dot{s}_{\rho_n} - \dot{s}_\rho\}(-\infty,\gamma^n]\|_{G_n} \to 0$ as $\rho_n \to \rho$. By (3.4), (3.5) and the GDCT, $\|\{\dot{s}_\theta - \dot{s}_\rho\}(-\infty,\gamma]\|_G$ is continuous in θ. Thus $Z(G)$ converges to 0. Now consider

$$\begin{aligned}Z(G_n) - Z(G) &= \tfrac{1}{4}\int_{-\infty}^{\gamma^n} \left[\frac{\dot{f}_{\rho_n}}{f_{\rho_n}^{1/2}} - \frac{\dot{f}_\rho}{f_\rho^{1/2}} \right]^2 (\overline{G}_n - \overline{G})\, d\lambda \\ &\quad + \tfrac{1}{4}\int_{-\infty}^{\gamma^n} \left[\frac{\dot{\overline{F}}_{\rho_n}}{\overline{F}_{\rho_n}^{1/2}} - \frac{\dot{\overline{F}}_\rho}{\overline{F}_\rho^{1/2}} \right]^2 d(G_n - G).\end{aligned}$$

By (3.4) and repeated use of the L_p convergence theorem [cf. Rudin (1974), page 76] the first term converges to 0. From the integration by parts formula and (3.4) we can bound the second term by the product of

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

and

$$\left\{ \int_{-\infty}^{\gamma^n} \left| \frac{d}{ds} \left[\frac{\dot{\overline{F}}_{\rho_n}}{\overline{F}_{\rho_n}^{1/2}} - \frac{\dot{\overline{F}}_\rho}{\overline{F}_\rho^{1/2}} \right] \right| ds + 1 \right\}.$$

Hence from $\sup_{(-\infty,\gamma]}|G_n - G| \to 0$ it suffices to prove that

$$I_n = \int_{-\infty}^{\gamma^n} \left| \frac{d}{ds} \left[\frac{\dot{\overline{F}}_{\rho_n}}{\overline{F}_{\rho_n}^{1/2}} - \frac{\dot{\overline{F}}_\rho}{\overline{F}_\rho^{1/2}} \right] \right| ds$$

remains bounded. Letting $p = 2 + \varepsilon$, q be the conjugate of p and $a = 1 - 1/p$, Hölder's inequality gives us

$$\left|\dot{\bar{F}}_{\rho_n}(x)\right| \le \bar{F}_{\rho_n}^{1/q}\left[\int_x^\infty \frac{\left|\dot{f}_{\rho_n}\right|^{2+\varepsilon}(s)}{f_{\rho_n}^{1+\varepsilon}(s)}\,ds\right]^{1/p}.$$

Notice that $1/q - 1/2 > 0$. Thus the conditions mentioned at the beginning of our proof imply that I_n remains bounded. So we have $\|\{\dot{s}_{\rho_n} - \dot{s}_\rho\}(-\infty, \gamma^n]\|_{G_n} \to 0$ as $\rho_n \to \rho$ and (3.2) follows. □

Now we are ready for the main result of this section: the differentiability of the minimum Hellinger distance functional.

THEOREM 3.1. *Suppose*

(i) *the assumptions in Lemma* 3.1 *hold*,
(ii) $\theta_0 = \Psi(f; G; \gamma)$ *exists, is unique and lies in the interior of* Θ,
(iii) $\int_0^\gamma(\dot{s}_{\theta_0}^2 + \ddot{s}_{\theta_0}(s_{\theta_0} - [L(f)]^{1/2}))\,d\mu \neq 0$,
(iv) *the assumptions of Lemma* 3.2 *hold for* $\rho = \theta_0$.

Then, for f_n *in an Hellinger neighborhood of* f, G_n, G *and* γ^n *satisfying* (3.1) *and* $\Delta G(\gamma) = 0$,

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

$$= \left\{\int_{-\infty}^{\gamma}\left(\dot{s}_{\theta_0}^2 + \ddot{s}_{\theta_0}\left(s_{\theta_0} - [L(f)]^{1/2}\right)\right)d\mu + v_n\right\}^{-1} \tag{3.6}$$

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

where v_n *converges to* 0 *as* $\|f_n^{1/2} - f^{1/2}\|_2 \to 0$.

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

$$\int_{-\infty}^{\gamma^n}\dot{s}_{\theta_{nn}}\left(s_{\theta_{nn}} - [L(f)]^{1/2}\right)d\mu_n = 0. \tag{3.7}$$

Using (3.2) and (3.3) to expand $s_{\theta_{nn}}$, $\dot{s}_{\theta_{nn}}$ around θ_0, we can rewrite (3.7) as

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

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

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

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

An argument similar to the one used to prove Lemma 3.2 shows that for G_n, G and γ^n satisfying (3.1), as $\| f_n^{1/2} - f^{1/2}\|_2 \to 0$,

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

Thus from the above equation we have

$$0 = \int_{-\infty}^{\gamma^n}\dot{s}_{\theta_0}\Big(s_{\theta_0} - [L(f_n)]^{1/2}\Big)\,d\mu_n + \left\{\int_{-\infty}^{\gamma}\Big(\dot{s}_{\theta_0}^2 + \ddot{s}_{\theta_0}\Big(s_{\theta_0} - [L(f)]^{1/2}\Big)\Big)\,d\mu + v_n\right\}(\theta_{nn} - \theta_0),$$

where v_n converges to 0 as $\| f_n^{1/2} - f^{1/2}\|_2 \to 0$. Therefore the result follows. □

4. Asymptotic distributions.

The MHD estimator of $\Psi(f;G;\tau)$ is defined by

$$\hat{\theta}_n = \Psi(f_n;\hat{G}_n;T),$$

where $T = \max \tilde{X}_i$, $\hat{G}_n$ is the product-limit estimator and f_n is the kernel density estimator as defined in (2.6), with some kernel function K and constant $a_n > 0$. We now prove the consistency of $\hat{\theta}_n$.

THEOREM 4.1. *Suppose that*

(i) *the assumptions of Lemma* 3.1 *hold*,

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

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

(iv) $\int_0^\tau(1/\overline{G}_-)\,dF < \infty$ *and* $\Delta G(\tau) = 0$.

Then

$$\hat{\theta}_n \to_p \Psi(f;G;\tau) \quad \textit{if } \Psi(f;G;\tau) \textit{ is unique}.$$

PROOF. By Wang (1987) and $\Delta G(\tau) = 0$, $\sup_{[0,\tau]}|\hat{G}_n - G| \to_p 0$. Also (iv) implies that $\tau = \tau_F$. Theorem 2.1 and Lemma 3.1 give the result immediately. □

For the asymptotic distribution of $\hat{\theta}_n$, we establish the following lemma to break the long proof and to illustrate the essential steps. Let $\|\cdot\|_\infty$ denote the $L^\infty(R)$-norm.

LEMMA 4.1. *Suppose that*

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

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

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

(iv) $n^{1/2}a_n^2 \to 0$.

Then for the (sub-) c.d.f. F_n *of* f_n *and any bounded function* U *on* $[0, \tau]$,

$$\int_0^\tau n^{1/2}[F_n - F]U\,dG \to_p \int_0^\tau P^1 U\,dG, \tag{4.1}$$

and for any right-continuous function V *that is of bounded variation on* $[0, \tau]$,

$$\int_0^\tau n^{1/2}[f_n - f]V\,d\lambda \to_p - \int_0^\tau P_-^1\,dV. \tag{4.2}$$

If, in addition,

(v) $\inf\{f[f > 0]\} > 0$ *and for some* $\varepsilon > 0$, $n^{1/2}a_n^{1+\varepsilon} \to \infty$, *then*

$$\int_0^\tau n^{1/2}(f_n - f)^2\,d\lambda \to_p 0. \tag{4.3}$$

Proof. Let

$$\tilde{f}_n(x) = a_n^{-1}\int K((x-y)/a_n)\,dF(y) = \int f(x - a_n t)K(t)\,dt,$$

$$\tilde{F}_n(x) = \int_{-\infty}^x \tilde{f}_n\,d\lambda.$$

Then we have

$$\begin{aligned} n^{1/2}\big[F_n(x) - \tilde{F}_n(x)\big] &= \int P_n^1(x - a_n t)K(t)\,dt, \\ n^{1/2}\big[\tilde{F}_n(x) - F(x)\big] &= \int n^{1/2}[F(x - a_n t) - F(x)]K(t)\,dt. \end{aligned} \tag{4.4}$$

By Fubini's theorem we can write

$$\int_0^\tau \big[\tilde{F}_n - F\big]U\,dG = \int A(t)K(t)\,dt, \tag{4.5}$$

where $A(t) = \int_0^\tau [F(x - a_n t) - F(x)]U(x)\,dG(x)$. Thus Taylor's expansion for A at $t = 0$ and the symmetry of K give us $\int_0^\tau n^{1/2}[\tilde{F}_n(x) - F(x)]U\,dG \to 0$. Now (4.1) follows since by (2.5)

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

As for (4.2), we have

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

Hence it remains only to prove $\int_0^\tau n^{1/2}[\tilde{f}_n(x) - f(x)]V\,d\lambda \to 0$. But again by Fubini's theorem we have

$$\int_0^\tau [\tilde{f}_n(x) - f(x)]V\,d\lambda = \int B(t)K(t)\,dt, \tag{4.6}$$

where $B(t) = \int_0^\tau [f(x - a_n t) - f(x)]V\,d\lambda$. Thus similarly as for $\tilde{F}_n - F$, we obtain $\int_0^\tau n^{1/2}[\tilde{f}_n(x) - f(x)]V\,d\lambda \to 0$.

Now to prove (4.3), it suffices to show $\int_0^T n^{1/2}(f_n - \tilde{f}_n)^2\,d\lambda \to_p 0$. Let $D_n(x,t) = P_n^1(x + a_n t) - P_n^1(x - a_n t)$. By symmetry of K and the Cauchy–Schwarz inequality, $\int_0^T n^{1/2}(f_n - \tilde{f}_n)^2\,d\lambda$ is bounded by a constant multiple of

$$\frac{1}{n^{1/2}a_n^2}\int_0^T\int_0^M D_n^2(x,t)|K'(t)|\,dt\,dx = \frac{1}{n^{1/2}a_n^2}\int_0^M\int_0^T D_n^2(x,t)\,dx|K'(t)|\,dt.$$

Hölder's inequality and (2.10) give us

$$\begin{aligned} &E\int_{a_n t}^{T - a_n t}\{D_n(x,t)\}^{2-2\varepsilon}\,dx \\ &\quad\le \int_{a_n t}^{\tau - a_n t}\left\{ED_n(x,t)^2[a_n t < x < T - a_n t]\right\}^{1-\varepsilon}dx \\ &\quad\le W\int_{a_n t}^{\tau - a_n t}[A(x + a_n t) - A(x - a_n t)]^{1-\varepsilon}\,dx \\ &\qquad + W\int_{a_n t}^{\tau - a_n t}\{[F(x + a_n t) - F(x - a_n t)]\bar{F}^{-1}(x - a_n t)\}^{1-\varepsilon}\,dx, \end{aligned} \tag{4.7}$$

for some constant W independent of t. The first term in (4.7) does not exceed $W\tau^{\varepsilon}\{2Ma_n A(\tau)\}^{1-\varepsilon}$ since

$$\int_{a_n t}^{\tau - a_n t}[A(x + a_n t) - A(x - a_n t)]\,dx \le \int_0^\tau\int_{u - a_n t}^{u + a_n t}dx\,dA(u) \le 2Ma_n A(\tau).$$

The second term in (4.7) does not exceed

$$W\{2Ma_n\|f\|_\infty\}^{1-\varepsilon}\inf\{f[f>0]\}\int_0^\tau F^{\varepsilon-1}\,dF.$$

Thus the sum in (4.7) can be written as $B_n a_n^{1-\varepsilon}$ for some bounded quantity B_n independent of t. It follows that

$$\frac{1}{n^{1/2}a_n^2}\int_0^M\int_{a_nt}^{T-a_nt}D_n(x,t)^{2-2\varepsilon}\,dx|K'(t)|\,dt \to_p 0.$$

So we have

$$\frac{1}{n^{1/2}a_n^2}\int_0^M\int_{a_nt}^{T-a_nt}D_n(x,t)^2\,dx|K'(t)|\,dt \to_p 0,$$

because $\|D_n\|$ is bounded in probability by (2.5). Since we also have

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

the result follows. □

The following theorem establishes the asymptotic distribution of the estimator $\hat{\theta}_n$. Recall that from the beginning of Section 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 under $L(f)$. We will use the differentiability of Ψ as in (3.6), specifying $\mu_n = \mu_{\hat{G}_n}$, $\gamma = \tau$. Thus $\theta_0 = \Psi(f;G;\tau)$. 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 to R by defining them to be 0 outside the support of f_{θ_0} or the support of f.

THEOREM 4.2. *Suppose the assumptions of Theorem* 3.1 *and Lemma* 4.1 *hold, and in addition,*

(i) $\|f_{\theta_0}\|_\infty < \infty$, $\|\dot{f}_{\theta_0}\|_\infty < \infty$ *and* $\inf\{f_{\theta_0}[f_{\theta_0} > 0]\} > 0$,
(ii) φ_1 *is of bounded variation on* $[0,\tau]$,
(iii) $\tau_{F_{\theta_0}} \le \tau$ *and* $\Delta G(\tau) = 0$.

Then, under $L(f)$, $n^{1/2}(\hat{\theta}_n - \Psi(f;G;\tau))$ *converges weakly to a normal distribution with mean* 0 *and finite variance.*

In particular, under $L(f_\theta)$, $n^{1/2}(\hat{\theta}_n - \theta)$ *converges weakly to* $N(0,1/I)$, *where* I *is the Fisher information*:

$$I = E\left[\frac{\partial}{\partial\theta}\big(\ln L(f_\theta)(\tilde{X},\delta)\big)\right]^2.$$

PROOF. Under our assumptions the expansion (3.6), with G_n, γ^n, γ replaced by $\hat{G}_n, T, \tau$, respectively, is valid, where v_n converges to 0 in probability. Since the coefficient of the integral on the right-hand side of (3.6) converges to a nonrandom limit, we only have to deal with the integral in (3.6). We will need to use the algebraic identity (for $a, b > 0$)

$$b^{1/2} - a^{1/2} = \frac{1}{2a^{1/2}}(b-a) - \frac{1}{2a^{1/2}}\frac{b^{1/2} - a^{1/2}}{b^{1/2} + a^{1/2}}(b-a). \tag{4.8}$$

Note that under the assumptions φ_0 and $\bar{F}_{\theta_0}/\bar{F}$ are also 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 (3.7), we have

$$\int_0^\tau \dot{s}_{\theta_0}\Big(\big[L(f_{\theta_0})\big]^{1/2} - \big[L(f)\big]^{1/2}\Big)\,d\mu = 0.$$

Thus

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

We can write $S_1 = \int_0^T BP_n^0\,d\lambda$, where $B = -\rho_1[f^{1/2} - f_{\theta_0}^{1/2}]$ is bounded. Since

$$\int_0^\tau \bar{F}^{\alpha-1}\,d\lambda \le \int_0^\tau \bar{F}^{\alpha-1}\,dF\Big(\inf_R f(x)[f(x) > 0]\Big)^{-1} < \infty$$

for $\alpha \in (0, 1/2)$ by Lemma 2.1 and the fact that $T \to \tau$ w.p.1, we have $S_1 \to_p \int_0^\tau BP^0\,d\lambda$.

Next, integration by parts gives

$$S_2 = -\int_0^T P_{n-}^0 A\,d\lambda + \Big\{\rho_0\big[\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}\big]P_n^0\Big\}(T), \tag{4.9}$$

where

$$A = \frac{d}{dx}\Big\{\rho_0\big[\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}\big]\Big\}. \tag{4.10}$$

When $\tau_{F_{\theta_0}} = \tau$, A is bounded; when $\tau_{F_{\theta_0}} < \tau$, on $[0, \tau_{F_{\theta_0}}]$, $A \le W\bar{F}_\theta^{1/2}$ and $\bar{F} > 0$. The latter, with Lemma 2.1, implies that $P_n^0 \to P^0$ in $\mathscr{U}[0, \tau_{F_{\theta_0}}]$. Therefore, in both cases we have $\int_0^T P_{n-}^0 A\, d\lambda \to_p \int_0^\tau P_-^0 A\, d\lambda$. The remainder term in (4.9)

$$\left\{\rho_0\left[\bar{F}^{1/2} - \bar{F}_{\theta_0}^{1/2}\right]P_n^0\right\}(T) \le W(T)\left(P_n^0\bar{F}\right)(T) = \to_p 0$$

by (2.5). Hence $S_2 \to_p -\int_0^\tau P_-^0 A\, d\lambda$.

By (4.8), $S_3 = S_{31} + S_{32}$, where

$$S_{31} = 2^{-1}\int_0^T \varphi_0 n^{1/2}\left[\bar{F}_n - \bar{F}\right] dG,$$

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

So $S_{31} \to_p -2^{-1}\int_0^\tau \varphi_0 P^1\, dG$ by Lemma 4.1 and the fact $T \to \tau$ w.p.1. The integrand of S_{32} is dominated by that of S_{31} in absolute value, and for $x \in [0, \tau)$, $(\bar{F}_n^{1/2} - \bar{F}^{1/2})(\bar{F}_n^{1/2} + \bar{F}^{1/2})^{-1} \to 0$. Also $\Delta G(\tau) = 0$. Hence the GDCT gives $S_{32} \to_p 0$. Thus $S_3 \to_p -2^{-1}\int \varphi_0 P^1\, dG$.

By (4.8) again, $S_4 = S_{41} + S_{42}$, where

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

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

From (2.7), $f_n - f$ is bounded. So

$$\int_T^\tau n^{1/2}(f_n - f)\varphi_1 \bar{G}\, d\lambda \le W n^{1/2}(\tau - T) \to_p 0.$$

Hence $S_{41} \to_p -2^{-1}\int P_-^1\, d(\varphi_1 G)$ by (4.2). Since $|S_{42}| \le W\int_0^T n^{1/2}(f_n - f)^2 \bar{G}\, d\lambda$, by (4.3) $S_{42} \to_p 0$. Thus we have $S_4 \to_p -2^{-1}\int P^1\, d(\varphi_1 \bar{G})$.

Now we consider the remainder terms R_1, R_2, R_3 and R_4. The results $R_1 \to_p 0$, $R_2 \to_p 0$ follow since

$$n^{1/2}(\tau - T) \le \left(\inf\{f[f > 0]\}\right)^{-1} n^{1/2}(F(\tau-) - F(T)) \to_p 0.$$

To deal with R_3, write it as

$$\int_0^t \left\{\rho_1(x)\left[f_n^{1/2}(x) + f^{1/2}(x)\right]^{-1}\right\}(f_n - f)(x) P_n^0(x)\, dx.$$

Note that the quantity in { } is bounded. Hence by (2.5), (2.7) and (4.6) we have $R_3 \to_p 0$. Now we write $R_4 = \int_0^t B_n\, d(\hat{G}_n - G)$ for some B_n. From $a^{1/2} - b^{1/2} = (a + b)/(a^{1/2} + b^{1/2})$, (4.4) and

$$\left|\rho_0\left(\bar{F}_n^{1/2} + \bar{F}^{1/2}\right)^{-1}(x)\right| \le \left|\rho_0\left[\bar{F}^{1/2}\right]^{-1}(x)\right| \le W,$$

the integrand B_n is uniformly bounded in probability. It also has the property that, for $x_n \to x \in (0, \tau)$, $B_n(x_n) \to_p 4^{-1}\dot{\bar{F}}_{\theta_0}(\bar{F}_{\theta_0}\bar{F})^{-1/2}(x)P^1(x)$. By the uniform strong consistency of $\hat{G}_n$ on $[0, t]$ for rational $t < \tau$ [cf. Shorack and Wellner (1986)], for w in a set of probability 1, $\hat{G}_n^{-1} \to G^{-1}$ at continuity points of G^{-1} in $(0, G(\tau))$. Note that for $t < 1$ we have $G^{-1}(t) < \tau$, hence $B_n \circ \hat{G}_n^{-1}(t) \to B \circ G^{-1}(t)$ for a.e. t. Now let $\rho_n = \hat{G}_n(T)$, $\eta_n = G(T)$ and $\rho = G(\tau)$. By Wang (1987), $\rho_n \to_p \rho$, and continuity of G at τ gives us $\eta_n \to \rho$. Hence

$$R_4 = \int_0^T B_n \, d(\hat{G}_n - G) = \int_0^{\rho_n} B_n \circ \hat{G}_n^{-1}(t) \, dt - \int_0^{\eta_n} B_n \circ G^{-1}(t) \, dt \to_p 0.$$

Therefore, we have proved that for A defined in (4.10),

$$\begin{aligned}
& n^{1/2} \int_0^T \dot{s}_{\theta_0} \Big([L(f_n)]^{1/2} - [(f_{\theta_0})]^{1/2} \Big) \, d\mu_n \\
\text{(4.11)} \qquad & \to_p - \int_0^T \rho_1 [f^{1/2} - f_{\theta_0}^{1/2}] P^0 \, d\lambda - \int_0^\tau P_-^0 A \, d\lambda \\
& \quad - \tfrac{1}{2} \int_0^\tau \varphi_0 P^1 \, dG - \tfrac{1}{2} \int_0^\tau P^1 \, d(\varphi_1 \overline{G}),
\end{aligned}$$

where the limit has a normal distribution with mean 0 and finite variance. Thus $n^{1/2}(\hat{\theta}_n - \Psi(f; G; \tau))$ also converges weakly to a normal distribution with mean 0 and finite variance. The variance can be computed using (2.1). In particular, when $f = f_\theta$ for some θ, then the limit becomes

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

where

$$\begin{aligned}
h_1(x) &= \tfrac{1}{2} \int_x^\tau \varphi_0 \overline{F}_\theta \, dG + \tfrac{1}{2} \int_x^\tau \overline{F}_\theta \, d(\varphi_1 \overline{G}) \\
&= \tfrac{1}{4} \left[\int_x^\tau \dot{\overline{F}}_\theta \, dG + \int_x^\tau \varphi_1 \overline{G} \, dF_\theta \right] - \tfrac{1}{2} \varphi_1 \overline{F}_\theta \overline{G}(x) \\
&= \tfrac{1}{4} \Big(\dot{\overline{F}}_\theta - 2 \varphi_1 \overline{F}_\theta \Big) \overline{G}.
\end{aligned}$$

From this, and the quadratic variation process of the martingale $P^1 / \overline{F}$ from Section 2, one can check that the variance of the limit of $n^{1/2}(\hat{\theta}_n - \Psi(f; G; \tau))$ is $1/I$. □

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

$$\text{(4.12)} \qquad \left\| n^{1/2}(d_n^{1/2} - d^{1/2}) - \alpha \right\|_2 \to 0.$$

Note that (4.12) implies $\alpha \perp d^{1/2}$ and

$$\text{(4.13)} \qquad \left\| n^{1/2} \Big([L(d_n)]^{1/2} - [L(d)]^{1/2} \Big) - \beta \right\|_G \to 0,$$

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 $\{\hat{\underline{\theta}}_n\}$ be a sequence of estimators of the functional $\Psi(d;G;\gamma)$ based on $(\tilde{X}_i,\delta_i)$, $i = 1,\dots,n$. We say that $\{\hat{\underline{\theta}}_n\}$ is regular at d if for $\{d_n\} \in K(d,G)$ and $X_1,\dots,X_n$ i.i.d. with density d_n, $n^{1/2}[\hat{\underline{\theta}}_n - \Psi(d_n;G;\gamma)]$ converges weakly to a distribution $\Gamma(d;\gamma;G)$ that does not depend on the particular sequence $\{d_n\}$. The following theorem extends Theorem 5 of Beran (1977a) to the censored data case.

THEOREM 4.3. *Suppose $\Psi(\cdot;G;\gamma)$ is differentiable at d with derivative ψ, in the sense that for d_n in a Hellinger neighborhood of d,*

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

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

PROOF. The proof is almost the same as in Beran (1977a). One only needs the following variation of (4.3) in his paper: For d_n, d in (4.12) and any $\varepsilon > 0$,

$$P_{L(d)}\left[\left|L_n - 2n^{-1/2}\sum_{i=1}^{n}\beta(\tilde{X}_i,\delta_i)[L(d)]^{-1/2}(\tilde{X}_i,\delta_i) + 2\int_{-\infty}^{\gamma}\beta^2\,d\mu_G\right| > \varepsilon\right] \to 0, \tag{4.14}$$

where $L_n = 2\prod_{i=1}^n[L(d_n)]^{1/2}(\tilde{X}_i,\delta_i)/[L(d)]^{1/2}(\tilde{X}_i,\delta_i)$. This can be easily deduced from Le Cam's second lemma and is similar to Lemma 1 of Wellner (1982). □

When the conclusion of Theorem 4.2 holds, the sequence of estimators $\{\hat{\theta}_n\}$ is regular at f_θ. In fact, under $L(f_\theta)$, the difference

$$n^{1/2}[\hat{\theta}_n - \theta] - \left\{\tfrac{1}{4}\int \dot{\overline{F}}_\theta \overline{F}_\theta^{-1} n^{1/2}[\hat{F}_n - F_\theta]\,dG + \tfrac{1}{4}\int \dot{f}_\theta f_\theta^{-1}\overline{G} n^{1/2}\,d[\hat{F}_n - F_\theta]\right\} \to_p 0,$$

as in the proof of Theorem 4.2. This is also true under $L(d_n)$, since (4.14) gives the contiguity of $\{L(d_n)\}$ to $\{L(f_\theta)\}$. Thus convergence in $D[0,\tau]$ of $P_{nn}^1 = n^{1/2}(\hat{F}_n - D_n)$ to P^1 under $L(d_n)$ and the differentiability of $\Psi(\cdot;G;\tau)$ will give the regularity of $\{\hat{\theta}_n\}$. By Theorem 5.3 in Pollard (1984) the necessary and sufficient conditions for the convergence of P_{nn}^1 to P^1 are the finite-dimensional convergence and small oscillation conditions. The former can be derived from the martingale representation of $P_{nn}^1/\overline{D}_n$ on $[0, \mu]$, for any $\mu < \tau_{F_\theta}$, and Theorem 8.13 of Pollard (1984). The small oscillation property is preserved under contiguity. Therefore $\hat{\theta}_n$ is a regular estimator of $\Psi(f; G; \tau)$ at f_θ, distinguished for having the smallest asymptotic variance when the parametric model is true.

5. Robustness properties and numerical simulation. The robustness discussion in Beran (1977b) can be carried over almost verbatim to our censored data case. The robustness of the MHDE in one way 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 $\rho \in L^2(\mu)$,

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

$$U(f) - \theta = \int \rho\left([L(f)]^{1/2} - [L(f_\theta)]^{1/2}\right) d\mu + u, \tag{5.1}$$

where $u \to 0$ as $[L(f)]^{1/2} \to [L(f_\theta)]^{1/2}$ in $L^2(\mu)$. We can assume $\rho \perp s_\theta$ 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 we have

$$1 = \frac{1}{\alpha}[U(f_{\theta+\alpha}) - U(f_\theta)] \to \left[\int \rho \dot{s}_\theta \, d\mu\right] \quad \text{as } \alpha \to 0.$$

So $\int \rho \dot{s}_\theta \, d\mu = 1$. When Theorem 3.1 applies, $\{\Psi(\cdot; G)\}$ belongs to this class. As in Beran (1977b) to see which functional in the class is asymptotically least affected by infinitesimal perturbations of f_θ, let us examine the behavior of $[U(f) - \theta]$. By projection, $[L(f)]^{1/2}$ can be represented as $[L(f)]^{1/2} = \cos \gamma s_\theta + \sin \gamma \delta$, where $\gamma \in [-, 2^{-1}\pi]$, $\|\delta\|_G = 1$, $\int \delta s_\theta \, d\mu = 0$. Then $U(f) - \theta = \gamma \int \rho \delta \, d\mu + o(\gamma)$ as $\gamma \to \infty$. Thus for small γ, or equivalently, small $\|[L(f)]^{1/2} - s_\theta\|_G$, the behavior of $|U(f) - \theta|$ is primarily determined by $|\int \rho \delta \, d\mu| = L(\rho, \delta)$. Thus the problem becomes: Which ρ minimizes the deviation L, against all possible directions δ? It turns out that $\Psi(\cdot; G)$ corresponds to the optimal choice of ρ, as the following result shows. The proof is almost the same as in Beran (1977b) and is omitted.

THEOREM 5.1. *Suppose s_θ satisfies the conditions in Lemma* 3.2. $\rho \in L_2(\mu)$, $\int \rho s_\theta \, d\mu = 0$, $\int \rho \dot{s}_\theta \, d\mu = 1$, $\|\delta\|_G = 1$, $\int \delta s_\theta \, d\mu = 0$.
Then

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

where

$$\rho^0 = \left[\int \dot{s}_\theta^2 \, d\mu\right]^{-1} \dot{s}_\theta, \qquad \delta^0 = \|\rho\|_G^{-1} \rho^0.$$

TABLE 1

$b_i = (i+1)/10$, $\theta = 2$, $\alpha = 0.1$ ($\theta_0 = 2.267$, *10% contamination*, $\frac{1}{3}$ *censoring*)

Cov($\tilde{\theta}$)

0.1565	1.658	0.1428	0.1384	0.1217	0.1076	0.09553
0.1658	0.2319	0.1988	0.1914	0.169	0.1503	0.1342
0.1428	0.1988	0.3286	0.3059	0.2694	0.239	0.2127
0.1384	0.1914	0.3059	0.3197	0.2843	0.2538	0.2268
0.1217	0.169	0.2694	0.2843	0.2551	0.2292	0.2059
0.1076	0.1503	0.239	0.2538	0.2292	0.2073	2.1871
0.09553	0.1342	0.2127	0.2268	0.2059	0.1871	0.1696

Mean

2.264	2.431	2.647	2.566	2.388	2.237	2.105

Mean square error from $\theta = 2$

0.228	0.418	0.747	0.640	0.406	0.263	0.181

In a more abstract setting, when the robustness of estimator of the functional Ψ is measured by the maximum risk in Hellinger neighborhoods of F_θ and G, Yang (1990) shows that the MHD estimators achieve the lower bound for the local asymptotic minimax risk in the Hájek–Le Cam sense. Notice that in that paper a local formulation is given in terms of projections to subspaces.

To demonstrate the finite-sample behaviors of the MHD estimators, a numerical simulation was performed for the exponential family $\overline{F}_\theta(x) = e^{-\theta x}[x > 0]$, the contaminated distribution $F(x) = (1-\alpha)F_\theta(x) + \alpha[x \geq 0.001]$ and exponential censoring: $\overline{G}(x) = e^{-x}[x > 0]$. Here the contamination was introduced by a point mass at a very small number 0.001; with a larger number the effect of contamination was more likely to be cancelled by censoring. Window size $a_n = b$ and the Epanecknikov kernel $K = \frac{3}{4}(1 - x^2)[|x| \leq 1]$ were chosen to evaluate the MHDE $\hat{\theta}(b) = \psi(f_n; \hat{G}_n; \infty)$. For comparison, the estimators $\hat{\theta}^1 = (n-1)/\Sigma X_i$ (to see how well one could do if the complete data $X_1, \dots, X_n$ were available), $\hat{\theta}^2 = \Sigma \delta_i / \Sigma \tilde{X}_i$, a sufficient statistic for θ, and $\hat{\theta}^3 = 1/\hat{F}_n^{-1}(1 - e^{-1})$, the percentile estimator based on $\hat{F}_n$, were also considered. The simulation was based on samples of size 50 and 600 repetitions. The mean, mean square error and covariance of the estimator vector $\tilde{\theta} = (\hat{\theta}^1, \hat{\theta}^2, \hat{\theta}^3, \hat{\theta}(b_1), \hat{\theta}(b_2), \hat{\theta}(b_3), \hat{\theta}(b_4))$ are given in Tables 1 and 2.

From Tables 1 and 2 we can see the influence of the window size a_n on the behavior of the MHD estimator. In practice, how should we choose a_n? For scale-invariant families one can first use numerical simulation with a known parameter value to determine a proper window size, then by invariance that window size works well for all parameters. In general, one may take the value of an initial estimator as the true parameter to choose a window size.

TABLE 2

$b_i = ((i + 1)/10)(2/9)$, $\theta = 9$, $\alpha = 0$ ($\theta_0 = \theta = 9$, *no contamination*, 10% *censoring*)

			Cov($\tilde{\theta}$)			
1.695	1.687	1.479	1.531	1.401	1.291	1.181
1.687	1.849	1.641	1.667	1.523	1.405	1.286
1.479	1.641	2.425	1.898	1.653	1.47	1.305
1.531	1.667	1.898	1.98	1.78	1.616	1.462
1.401	1.523	1.653	1.78	1.633	1.493	1.36
1.291	1.405	1.47	1.616	1.493	1.383	1.266
1.181	1.286	1.305	1.462	1.36	1.266	1.165
			Mean			
8.956	9.114	9.749	9.708	9.129	8.674	8.285
			Mean square error from $\theta = 9$			
1.697	1.862	2.986	2.481	1.649	1.489	1.676

Acknowledgments. This paper is part of my Ph.D. dissertation under the guidance of Professor Hira L. Koul at Michigan State University. I would like to thank Professor Koul for suggesting this problem and for his continuous encouragement and stimulating advice. I would also like to thank Professor James Stapleton for helping with the simulation, the Associate Editor and the referees for valuable comments and Professor Benjamin Duran for proofreading. All their efforts led to this improved version of the manuscript.

REFERENCES

BERAN, R. J. (1977a). Robust location estimates. *Ann. Statist.* **5** 431–444.

BERAN, R. J. (1977b). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BERAN, R. J. (1981). Efficient robust estimation for parametric models. *Z. Wahrsch. Verw. Gebiete* **55** 91–109.

BILLINGSLEY, P. (1968). *Weak Convergence of Probability Measures.* Wiley, New York.

FABIAN, V. and HANNAN, J. (1985). *Introduction to Probability and Mathematical Statistics.* Wiley, New York.

GILL, R. (1980). *Censoring and Stochastic Integrals.* Math. Centre Tract 124. Mathematisch Centrum, Amsterdam.

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

GILL, R. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. *Scand. J. Statist.* **16** 97–128.

HÁJEK, J. (1972). Locally asymptotic minimax and admissibility in estimation. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.

HEWITT, E. and STROMBERG, K. (1965). *Real and Abstract Analysis.* Springer, New York.

KAPLAN, E. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

MILLAR, P. W. (1983). *The Minimax Principle in Asymptotic Statistical Theory. Lecture Notes in Math.* **876** 75–265. Springer, New York.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

RUDIN, W. (1974). *Real and Complex Analysis*, 2nd ed. McGraw-Hill, New York.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

WANG, J. G. (1987). A note on the uniform consistency of the Kaplan–Meier estimator. *Ann. Statist.* **15** 1313–1316.

WELLNER, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10** 595–602.

YANG, S. (1990). Efficient robust estimation of parameter in the random censorship model. *Statist. Probab. Lett.* **10** 419–426.

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
BOX 4139
LUBBOCK, TEXAS 79409-1042