Asymptotically Efficient Minimum Distance Estimators for Location

by

Dennis D. Boos

Let F_n be the empirical distribution function from a sample having distribution $F_0(x-\theta)$. A weighted Cramer-von Mises distance between $F_n(x)$ and $F_0(x-\theta)$ is minimized to produce estimators $\hat{\theta}_n$ which are asymptotically normal. If the weight function is taken proportional to $(-\ell n\ f_0(x-\theta))''/f_0(x)$, then the asymptotic variance of $n^{\frac{1}{2}}(\hat{\theta}_n-\theta)$ is given by $1/I(f_0)$, where $I(f_0)$ is Fisher's information. In addition, the *minimized* distance, i.e., the distance between $F_n(x)$ and $F_0(x-\hat{\theta}_n)$, is investigated as a goodness-of-fit statistic.

KEY WORDS: Minimum distance; Asymptotic efficiency; Goodness-of-fit.

1. INTRODUCTION

Let $X_1,\ldots,X_n$ be a random sample with distribution function $F(x) = F_0(x-\theta)$ and density $f(x) = f_0(x-\theta)$, $\theta\varepsilon(-\infty,\infty)$. Let F_n be the usual empirical distribution function, and let $w(x)$ be a non-negative weight function defined on $(-\infty,\infty)$. We shall consider estimates of θ obtained by minimizing

$$d_{F_n}(\theta) = \int_{-\infty}^{\infty}\left[F_n(x+\theta) - F_0(x)\right]^2 w(x)dx \ . \tag{1.1}$$

Our main contribution is to show that the weight function

$$w_0(x) = \frac{d^2}{dx^2}(-\ell n\ f_0(x))/f_0(x) \tag{1.2}$$

yields estimates which are asymptotically efficient. However, if $F \not\in F_0 = \{F\colon\ F(x) = F_0(x-\theta),\ \theta\varepsilon(-\infty,\infty)\}$, then the estimator $\hat{\theta}_n$ may still be estimating a reasonable quantity such as the center of symmetry of F. In either case we provide conditions for asymptotic normality of $\hat{\theta}_n$ and a goodness-of-fit statistic to check the assumption $F(x) = F_0(x-\theta)$. Although minimum distance estimators have been extensively studied, e.g., Blackman (1955), Wolfowitz (1957), Knusel (1969), Rao, Schuster, and Littel (1975), Bolthausen (1977), and Parr (1978), only Beran (1977, 1978) exhibits asymptotically efficient estimators. Moreover, Beran's work is essentially different in that he uses the Hellinger metric which is defined on *densities* rather than on distribution functions.

In Section 2 we define the estimator and give a consistency theorem. In Section 3 we provide a general asymptotic normality theorem and obtain as a corollary the above-mentioned asymptotic efficiency result based on (1.2). Section 4 investigates use of the *minimized* distance as a measure of goodness-of-fit. The limiting distribution of this statistic is given in the context of location-scale families for both null, $F(x) = F_0((x-\theta)/\sigma)$, and alternative situations.

2. CONSISTENCY

For distribution functions G, let the functional $\theta(G)$ satisfy

$$\min_{-\infty<\theta<\infty} d_G(\theta) = d_G(\theta(G)) \ , \tag{2.1}$$

where $d_G(\cdot)$ is defined by (1.1) with G in place of F_n. The following lemma shows that (2.1) has at least one solution $\theta(G)$, though perhaps more. A similar development is given in Beran (1978).

LEMMA 2.1. *Let the weight function* w *satisfy*

$$0 \leq w(x) \leq a|x|^r \tag{2.2}$$

for some constant $a > 0$ *and for some integer* $r \geq 0$. *Let* G *and* F_0 *have finite* $(r+1)$th *moments. Then*

(i) *the set of* $\theta(G)$ *which satisfy* (2.1) *is nonempty and compact;*

(ii) *if* $G(x) = F_0(x-\theta')$, *then* $\theta(G) = \theta'$ *is the unique solution of* (2.1).

PROOF. (i) We first show that $d_G(\theta)$ is continuous for all $\theta\varepsilon(-\infty,\infty)$. Let $|\theta_n-\theta|<\delta$. Then

$$\begin{aligned} |d_G(\theta_n)-d_G(\theta)| &\leq \int\left|[G(x+\theta_n)-F_0(x)]^2 - [G(x+\theta)-F_0(x)]^2\right|w(x)dx \\ &\leq 2\int\left|G(x+\theta_n-\theta) - G(x)\right|w(x-\theta)dx . \end{aligned} \tag{2.3}$$

The integrand is bounded by $|G(x+\delta)-G(x-\delta)|a|x-\theta|^r$, which can be shown to be integrable using integration by parts and the finite $(r+1)$th moment assumption. Then, as $\theta_n \to \theta$ we have $G(x+\theta_n-\theta) \to G(x)$ a.e. Lebesgue, and thus the right-hand side of (2.3) converges to 0 by dominated convergence.

Since $d_G(\theta)$ is continuous, it achieves a minimum over every compact interval. Also, as $\theta \to -\infty$ $d_G(\theta) \uparrow \int F_0^2(x)w(x)dx$, and as $\theta \to +\infty$ $d_G(\theta) \uparrow \int[1-F_0(x)]^2w(x)dx$. Thus $d_G(\theta)$ achieves a minimum over all $\theta\varepsilon(-\infty,\infty)$, and the set of all such values $\theta(G)$ lie in a compact interval. By continuity of $d_G(\theta)$, these values form a closed set and hence a compact set. (ii) If $G(x) = F_0(x-\theta')$, then $d_G(\theta') = 0$ and $d_G(\theta) > 0$ for $\theta \neq \theta'$. □

For the case that G is symmetric about θ' and F_0 is symmetric about 0 and $w(x) = w(-x)$, it appears that θ' will be the unique solution of (2.1). Thus far, however, we have been unable to prove this assertion. For the special weight function $w(x) = 1$ Knusel (1969) showed that $d_G(\theta)$ is strictly convex if F_0 has a positive density on $(-\infty,\infty)$.

The next lemma gives conditions for $\theta(\cdot)$ to be continuous with respect to the L_1 norm on $(-\infty,\infty)$. Let $||h||_\infty = \sup_{-\infty<x<\infty}|h(x)|$ be the usual sup-norm on $(-\infty,\infty)$.

LEMMA 2.2. *Let* G_n *be a sequence of distribution functions such that*

$$||G_n-F||_1 = \int_{-\infty}^{\infty} |G_n(x)-F(x)|dx \to 0 \ , \ n\to\infty. \tag{2.4}$$

Let F_0, F, *and* G_n, *each* n, *have finite first moments. Let* w *be nonnegative and bounded. If* $\theta(F)$ *is unique, then* $\theta(\cdot)$ *is continuous with respect to* $||\cdot||_1$, *i.e.,*

$$\theta(G_n) \to \theta(F) \ , \ n\to\infty. \tag{2.5}$$

PROOF. We begin with the bound

$$\begin{aligned} |d_{G_n}(\theta)-d_F(\theta)| &\le \int |G_n(x+\theta) - F(x+\theta)|\cdot|G_n(x+\theta) + F(x+\theta) - 2F_0(x)|w(x)dx \\ &\le 2||w||_\infty||G_n-F||_1 \ . \end{aligned} \tag{2.6}$$

Since the right-hand side of (2.6) is not a function of θ

$$\sup_{-\infty<\theta<\infty} |d_{G_n}(\theta) - d_F(\theta)| \to 0 \ , \ n\to\infty. \tag{2.7}$$

Then using the definition of $\theta(\cdot)$, we have

$$|d_F(\theta(G_n)) - d_F(\theta(F))| \le |d_F(\theta(G_n)) - d_{G_n}(\theta(G_n))| + |d_{G_n}(\theta(F_n)) - d_F(\theta(F))|$$

$$\le 2 \sup_{-\infty<\theta<\infty} |d_{G_n}(\theta) - d_F(\theta)| \to 0 \ , \ n\to\infty \ .$$

Since $d_F(\theta) \uparrow \int F_0^2(x)w(x)dx$ or $\int [1-F_0(x)]^2 w(x)dx$ according as $\theta \to -\infty$ or $\theta \to +\infty$, we have

$$\lim_{n\to\infty} d_F(\theta(G_n)) = d_F(\theta(F)) < \min\left\{\int F_0^2(x)w(x)dx, \int [1-F_0(x)]^2 w(x)dx\right\} . \tag{2.8}$$

If (2.5) is false, then there exists a subsequence $\{\theta_n\} \subset \{\theta(G_n)\}$ and an $\varepsilon > 0$ such that for n sufficiently large $|\theta_n - \theta(F)| > \varepsilon$. Nevertheless, $\{\theta_n\}$ does lie in some compact interval $[\theta_1, \theta_2]$. Otherwise, there would exist a subsequence $\{\theta_\ell\} \subset \{\theta_n\}$ such that $\theta_\ell \to -\infty$ or $+\infty$, which would entail by continuity of $d_F(\theta)$ that $d_F(\theta_\ell) \to d_F(-\infty)$ or $d_F(+\infty)$, contradicting (2.8). Since $\{\theta_n\}$ lies in $[\theta_1, \theta_2]$, there exists another subsequence $\{\theta_m\} \subset \{\theta_n\}$ and a point θ_0 such that $\theta_m \to \theta_0 \ne \theta(F)$. Thus

$$\lim_{m\to\infty} d_F(\theta_m) = d_F(\theta_0) > d_F(\theta(F)) \ ,$$

contradicting (2.8). □

THEOREM 2.1. *Let* $X_1, \ldots, X_n$ *be independent r.v.'s with distribution* F. *Let* F *and* F_0 *have finite first moments. Let* w *be nonnegative and bounded. If* $\theta(F)$ *is unique, then*

$$\theta(F_n) \xrightarrow{p} \theta(F) \ , \ n\to\infty \ .$$

PROOF. We need only verify that $||F_n-F||_1 \xrightarrow{p} 0$. Since $E|F_n(x)-F(x)| \leq E|I(X_1 \leq x)-F(x)| = 2F(x)(1-F(x))$, Fubini's theorem and the first moment condition allow the interchange $E||F_n-F||_1 = \int E|F_n(x)-F(x)|dx$. Using an expression for the mean deviation of a binomial random variable found in Frame (1945), we have

$$E|F_n(x)-F(x)| = 2F(x)(1-F(x)) \ P(B(n-1,F(x)) = [nF(x)]) \ ,$$

where $B(n-1,F(x))$ stands for a binomial random variable with parameters $n-1$ and $F(x)$ and $[\ ]$ is the greatest integer function. Then, since $F(x)(1-F(x))$ is integrable and $\lim_{n\to\infty} P(B(n-1,F(x)) = [nF(x)]) = 0$, each x, $\int E|F_n(x)-F(x)|dx \to 0$ follows by dominated convergence and thus $||F_n-F||_1 \xrightarrow{p} 0$ follows by the Markov inequality. □

REMARKS. (i) If $w(x)$ is Lebesgue integrable, $\theta(\cdot)$ may be shown to be continuous with respect to the sup-norm $||\cdot||_\infty$. The Gliventio-Cantelli theorem then yields $\theta(F_n) \xrightarrow{wp1} \theta(F)$.
(ii) The condition that w be nonnegative and bounded is exactly (2.2) with $r = 0$. Thus we have shown consistency for only the most restrictive of our general conditions. If (2.7) can be extended, then Theorem 2.1 can be easily extended.

3. ASYMPTOTIC NORMALITY

If $d_G(\theta)$ is differentiable, it follows from (2.1) that

$$d_G'(\theta(G)) = \frac{d}{d\theta} d_G(\theta) \Big|_{\theta=\theta(G)} = 0 \ . \qquad (3.1)$$

This characterization of $\theta(G)$ as the solution of an equation is vital for proving asymptotic normality of $\theta(F_n)$. The following lemma gives sufficient conditions for $d_G(\theta)$ to be continuously differentiable on $(-\infty,\infty)$. Whenever we talk about the derivative w' of w, we mean that w is continuous everywhere and that w' exists everywhere except possibly at a finite number of points where at least a right-hand derivative exists.

LEMMA 3.1. *Let* w *and* w' *both satisfy* (2.2) *for some* $a > 0$ *and integer* $r \geq 0$. *Let* G *and* F_0 *have finite* $(r+1)$th *moments. Let* F_0 *have a bounded density* f_0. *Then* $d_G(\theta)$ *is continuously differentiable on* $(-\infty,\infty)$ *with derivative*

$$d_G'(\theta) = 2\int [G(x+\theta) - F_0(x)]f_0(x)w(x)dx - \int [G(x+\theta) - F_0(x)]^2 w'(x)dx .$$

Further, if G *has a bounded density* g , *then*

$$d_G''(\theta) = 2\int g(x+\theta)f_0(x)w(x)dx - 2\int [G(x+\theta) - F_0(x)]g(x+\theta)w'(x)dx .$$

PROOF. Let $\theta_n \to \theta$. Then

$$\frac{d_G(\theta_n)-d_G(\theta)}{\theta_n-\theta} = \int \left[\frac{F_0(x-\theta)-F_0(x-\theta_n)}{\theta_n-\theta}\right]\left[2G(x)-F_0(x-\theta_n)-F_0(x-\theta)\right]w(x-\theta_n)dx \tag{3.2}$$

$$+ \int \left[G(x)-F_0(x-\theta)\right]^2\left[\frac{w(x-\theta_n)-w(x-\theta)}{\theta_n-\theta}\right]dx .$$

The result will follow if we can justify interchange of $\lim_{\theta_n \to \theta}$ and $\int$ in both terms. Let $|\theta_n - \theta| < \delta$. For $c > 0$ large enough divide $(-\infty,\infty)$ into three regions $(-\infty,-c)$, $[-c,c]$, and (c,∞). Then on (c,∞) the first integrand is bounded in absolute value by

$$||f_0||_\infty a(x-\theta+\delta)^r [2|G(x) - F_0(x-\theta)| + |F_0(x+\delta-\theta) - F_0(x-\delta-\theta)|] ,$$

which can be shown to be integrable using integration by parts and the moment conditions on G and F_0. The region $(-\infty,-c)$ is handled similarly. Then bounded convergence on $[c,c]$ and dominated convergence on $(-\infty,-c)$ and (c,∞) allow the interchange. The second term of (3.2) is handled in the same way. Continuity of $d_G'(\theta)$ and existence of $d_G''(\theta)$ are likewise straightforward to verify. □

The next theorem is our main asymptotic normality result. The basic restrictions are similar to those of Lemma 3.1 with the additional assumption that $\theta(F_n)$ is weakly consistent for $\theta(F)$. Define

$$h_F(t) = [d_F'(t) - d_F'(\theta(F))]/2(t-\theta(F)) \qquad t \neq \theta(F)$$

$$= \tfrac{1}{2} d_F''(\theta(F)) \qquad t = \theta(F)$$

THEOREM 3.1. *Let* $X_1,\ldots,X_n$ *be independent r.v.'s with distribution* F. *Let* w *and* w' *both satisfy* (2.2) *for some* $a > 0$ *and integer* $r \geq 0$. *Let* w' *be continuous a.e. Lebesgue. Suppose* *that* F *and* F_0 *have finite* (r+1)th *moments and bounded densities* f *and* f_0. *Let* f_0 *be continuous a.e. Lebesgue. If* $\theta(F_n) \xrightarrow{p} \theta(F)$, $h_F(\theta(F)) \neq 0$, *and*

$$0 < A(F) = \mathrm{Var}_F\left\{\int [I(X_1 \leq x) - F(x)]\{[F(x) - F_0(x-\theta(F))]w'(x-\theta(F)) - f_0(x-\theta(F))w(x-\theta(F))\}dx/h_F(\theta(F))\right\} < \infty ,$$

then

$$n^{\frac{1}{2}}[\theta(F_n) - \theta(F)] \xrightarrow{d} N(0, A(F)) .$$

(N(0,A) denotes a normal r.v. with mean 0 and variance A.)

COROLLARY. *If the conditions of Theorem* 3.1 *are satisfied for* $F(x) = F_0(x-\theta)$ *and* $w = w_0$ *given by* (1.2), *then*

$$n^{\frac{1}{2}}[\theta(F_n) - \theta] \xrightarrow{d} N\left(0, \frac{1}{I(f_0)}\right) ,$$

where

$$I(f_0) = \int \left(\frac{f_0'(x)}{f_0(x)}\right)^2 f_0(x)dx .$$

PROOF OF THEOREM 3.1. For notational simplicity let $\theta_n = \theta(F_n)$ and $\Delta_n(x) = F_n(x) - F(x)$. The method of proof is to show

$$\theta_n - \theta_F = T(F;\Delta_n) + o_p(n^{-\frac{1}{2}}) , \qquad (3.3)$$

where

$$T(F;\Delta_n) = \int \Delta_n(x)\{[F(x)-F_0(x-\theta(F))]w'(x-\theta(F))$$

$$-f_0(x-\theta(F))w(x-\theta(F))\}dx/h_F(\theta(F)) \ .$$

The result then follows since $T(F;\Delta_n)$ is an average of i.i.d. r.v.'s having mean 0 and variance $A(F)$. We justify (3.3) by showing

$$h_F(\theta_n)|\cdot|\theta_n-\theta(F) - \frac{h_F(\theta(F))}{h_F(\theta_n)} T(F;\Delta_n)| = o_p(n^{-\frac{1}{2}}) \ . \qquad (3.4)$$

Then, since $h_F(\theta_n) \xrightarrow{p} h_F(\theta(F)) \neq 0$ and $T(F;\Delta_n) = 0_p(n^{-\frac{1}{2}})$, (3.3) follows directly from (3.4). We start with the representation

$$\theta_n-\theta(F) = \left[d_F'(\theta_n)-d_F'(\theta(F))\right]/2h_F(\theta_n)$$

$$= \left[d_F'(\theta_n)-d_{F_n}'(\theta_n)\right]/2h_F(\theta_n) \ .$$

This last step uses the fact that $d_F'(\theta(F)) = d_{F_n}'(\theta_n) = 0$. Then the left-hand side of (3.4) can be bounded by

$$\left|\int \Delta_n(x)\left[f_0(x-\theta(F))w(x-\theta(F))-f_0(x-\theta_n)w(x-\theta_n)\right]dx\right|$$

$$+ \left|\int \Delta_n(x)\left\{\left[\frac{F(x)+F_n(x)-2F_0(x-\theta_n)}{2}\right]w'(x-\theta_n)-\left[F(x)-F_0(x-\theta(F))\right]w'(x-\theta(F))\right\}dx\right|$$

$$\leq ||\Delta_n||_\infty \int |f_0(x-\theta(F))w(x-\theta(F))-f_0(x-\theta_n)w(x-\theta_n)|dx$$

$$+ \tfrac{1}{2}\int \Delta_n^2(x)|w'(x-\theta_n)|dx + ||\Delta_n||_\infty \int |F_0(x-\theta(F))-F_0(x-\theta_n)||w'(x-\theta_n)|dx$$

$$+ ||\Delta_n||_\infty \int |F(x)-F_0(x-\theta(F))||w'(x-\theta_n)-w'(x-\theta(F))|dx \; . \qquad (3.5)$$

It is well-known that the Kolgomorov-Smirnov statistic $||\Delta_n||_\infty$ is $O_p(n^{-\frac{1}{2}})$ (e.g., see Dvoretsky, Kiefer, Wolfowitz (1956)). Thus for the first, third, and fourth terms of (3.5) we need only show that the integrals are $o_p(1)$. The first is handled by Sheffe's convergence theorem for densities since for each θ

$$f_0(x-\theta)w(x-\theta)/\int f_0(x)w(x)dx$$

is a density. For the third and fourth integrals we break up $(-\infty,\infty)$ into three regions and proceed as in the proof of Lemma 3.1. For the second integral of (3.5) we also look at the same three regions of integration, but this time for the purpose of taking expectations. For example, on (c,∞) with $|\theta_n-\theta(F)|<\delta$,

$$E\int_c^\infty \Delta_n^2(x)\,|w'(x-\theta_n|dx \leq E\int_c^\infty \Delta_n^2(x)a(x-\theta(F)+\delta)^r dx$$

$$\leq a\int_c^\infty F(x)(1-F(x))(x-\theta(F)+\delta)^r dx/n \; .$$

Since the moment assumption on F guarantees that the last integral is finite and $E\,\Delta_n^2(x) = F(x)(1-F(x))/n$, this last step is justified by Fubini's theorem. The other two regions are handled similarly, and using the Markov inequality we find that the second term of (3.5) is $O_p(n^{-1})$. □

EXAMPLE 3.1. Logistic, $F_0 = (1+e^{-x})^{-1}$. Using (1.2) we find $w_0(x) = 2$, and by taking derivatives $\theta(F_n)$ is seen to be the solution of

$$\sum_{i=1}^{n} [F_0(X_i-\theta)-\tfrac{1}{2}] = 0 ,$$

i.e., an M-estimator with ψ function $F_0(x)-\frac{1}{2}$. For any F symmetric about θ with bounded density f and finite first moment, Theorem 3.1 yields

$$n^{\frac{1}{2}}[\theta(F_n)-\theta] \xrightarrow{d} N(0,A(F)) ,$$

where

$$A(F) = \frac{\int [F_0(x-\theta)-\frac{1}{2}]^2 dF(x)}{\left(\int f_0(x-\theta)\,dF(x)\right)^2} \qquad F(x) \neq F_0(x-\theta) ,$$

$$= \frac{1}{12\left(\int f_0^2(x)\,dx\right)^2} \qquad F(x) = F_0(x-\theta) .$$

Actually, by theorem 3.1 the above convergence holds for any F_0 symmetric about 0 with bounded density f_0 and finite first moment. Results of Huber (1964) show that these conditions are stronger than necessary. If F_0 is logistic, then the last expression displayed is of course $1/I(f_0)$.

EXAMPLE 3.2. Double-exponential, $f_0(x) = \frac{1}{2}\exp(-|x|)$. Since w_0 is not defined in this case, we could replace the measure $w(x)dx$ in (1.1) by $(1/f_0(x))d\ \mathrm{sgn}(x)$. Then

$$d_{F_n}(\theta) = 2[F_n(\theta)-\tfrac{1}{2}]^2 ,$$

and $\theta(F_n)$ is a version of the sample median. Our theorems of course do not apply here.

EXAMPLE 3.3. Normal, $f_0(x) = (2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$. Then $w_0(x) = 1/f_0(x)$ and the estimator $\theta(F_n)$ is a solution of

$$d'_{F_n}(\theta) = 2(\theta-\bar{X}) - \int \left\{F_n(x+\theta)-F_0(x)\right\}^2 \left\{x/f_0(x)\right\} dx = 0 .$$

Unfortunately w_0 does not satisfy (2.2) for any r, and thus theorem 3.1 does not apply. A direct approach through the above equation should produce the appropriate asymptotic normality result. One might prefer to use $w(x) = 1$ as in Example 3.1 with F_0 the standard normal. The resulting (M-) estimator $\theta(F_n)$ is robust and 95.5 per cent efficient at the normal.

EXAMPLE 3.4. The classical Cramer-von Mises goodness-of-fit statistic employs the weight function $w(x) = f_0(x)$. For what F_0 is this weight function optimal? Solving the differential equation

$$\frac{d^2}{dx^2}[-\ln f_0(x)] = af_0^{\,2}(x)$$

yields the *hyperbolic secant* distribution (Johnson and Kotz (1970), p. 15)

$$f_0(x) = \frac{2}{\pi}\cdot\frac{1}{e^x+e^{-x}} = \frac{1}{\pi}\operatorname{sech}(x) .$$

The results of this section and of the previous one assume that the scale of the observations is known. In practice one will almost always want to simultaneously estimate location and scale by replacing $F_n(x+\theta)$ in (1.1) by $F_n(\sigma x+\theta)$ and then minimize over both θ and σ. In the next section we consider this situation. An alternative procedure is to replace σ by some consistent estimate $\hat{\sigma}_n$ found by other methods and then minimize with respect to only θ. In either case the asymptotic efficiency result will be preserved for *symmetric* F_0 but not necessarily otherwise. A similar difficulty with respect to scale arises in M-estimation.

4. GOODNESS-OF-FIT

Minimum distance provides not only a method of estimating parameters but also a natural statistic for judging model validity. Consider location-scale families of the form $\mathcal{F}_0 = \{F\colon\ F(x) = F_0((x-\theta)/\sigma),\ -\infty<\theta<\infty,\ 0<\sigma<\infty\}$. For a sample $X_1,\ldots,X_n$ from F we can estimate location and scale parameters $\theta(F)$ and $\sigma(F)$ by minimizing

$$d_{F_n}(\theta,\sigma) = \int_{-\infty}^{\infty}[F_n(\sigma x+\theta)-F_0(x)]^2 w(x)dx\ . \tag{4.1}$$

The natural test statistic for $H_0\colon\ F\varepsilon\mathcal{F}_0$ is then the *minimized* distance

$$\min_{\theta,\sigma} d_{F_n}(\theta,\sigma)\ . \tag{4.2}$$

In this section we have two theorems which give the asymptotic distribution of (4.2). The first theorem considers the non-null situation $F\notin\mathcal{F}_0$ and shows that (4.2) is asymptotically normal if the parameter estimates $\theta(F_n)$ and $\sigma(F_n)$ are suitably consistent. The second theorem gives the limiting distribution of (4.2) for the null case $F\varepsilon\mathcal{F}_0$. Here the basic assumption is the existence of expansions for $\theta(F_n)$ and $\theta(\sigma_n)$ of the form (3.3). The limiting distribution is expressed as a functional of the Browman Bridge on C[0,1], which facilitates comparison with other Cramer-von Mises statistics. Of interest also are the normalizing constants required for convergence in the two situations, $n^{\frac{1}{2}}$ for the non-null and n for the null, which suggests that generally good power may be reached for moderate n.

Let the pair of functionals $(\theta(G),\sigma(G))$ satisfy

$$\min_{\theta,\sigma} d_G(\theta,\sigma) = d_G(\theta(G),\sigma(G)) \ , \tag{4.3}$$

where $d_G(\theta,\sigma)$ is given by (4.1) with G in place of F_n. An extension of Lemma 2.1 can be given to make this definition more concrete. However, we proceed directly to

THEOREM 4.1. *Let* $X_1,\ldots,X_n$ *be independent r.v.'s with distribution* $F\neq F_0$. *Let* w *and* w' *both satisfy* (2.2) *for some* $a > 0$ *and integer* $r \geq 0$. *Suppose that* F *and* F_0 *have finite* (r+2)th *moments and that* F_0 *has a bounded density* f_0. *If* $(\theta_n,\sigma_n) = (\theta(F_n),\sigma(F_n))$ *satisfy*

$$\theta_n-\theta(F) = 0_p(n^{-\frac{1}{2}})$$

and (4.4)

$$\sigma_n-\sigma(F) = 0_p(n^{-\frac{1}{2}}) \ ,$$

then

$$n^{\frac{1}{2}}[d_{F_n}(\theta_n,\sigma_n)-d_F(\theta(F),\sigma(F))] \xrightarrow{d} N(0,A(F)),$$

where

$$A(F) = 4\int_{-\infty}^{\infty}[F^*(\min(x,y))-F^*(x)F^*(y)][F^*(x)-F_0(x)][F^*(y)-F_0(y)]w(x)w(y)dxdy \ ,$$

with $F^*(x) = F(\sigma(F)x+\theta(F))$.

PROOF. The assumptions imply that $d_{F_n}(\theta,\sigma)$ is differentiable everywhere on $(-\infty,\infty) \times (0,\infty)$, and thus by the mean value theorem

$$d_{F_n}(\theta_n,\sigma_n)-d_{F_n}(\theta(F),\sigma(F)) = \qquad (4.5)$$
$$(\theta_n-\theta(F))\ \frac{\partial}{\partial\theta}\, d_{F_n}(\theta_n',\sigma_n')+(\sigma_n-\sigma(F))\ \frac{\partial}{\partial\sigma}\, d_{F_n}(\theta_n',\sigma_n') \ ,$$

where $(\theta_n',\sigma_n') \xrightarrow{p} (\theta(F),\sigma(F))$ along with (θ_n,σ_n). The above partial derivatives converge in probability to the partial derivatives of $d_F(\theta,\sigma)$ evaluated at $(\theta(F),\sigma(F))$, which are necessarily 0 by (4.3). Then by (4.4) the right-hand side of (4.5) is $o_p(n^{-\frac{1}{2}})$. Let $\Delta_n^*(x) = \Delta_n(\sigma(F)x-\theta(F))$. Thus

$$\begin{aligned} d_{F_n}(\theta_n,\sigma_n)-d_F(\theta(F),\sigma(F)) &= d_{F_n}(\theta(F),\sigma(F))-d_F(\theta(F),\sigma(F))+o_p(n^{-\frac{1}{2}}) \\ &= 2\int\Delta_n^*(x)[F^*(x)-F_0(x)]w(x)dx \\ &\quad + \int(\Delta_n^*(x))^2w(x)dx + o_p(n^{-\frac{1}{2}}) \ . \end{aligned}$$

But $\int(\Delta_n^*(x))^2w(x)dx$ is $O_p(n^{-1})$ as is shown in the proof of Theorem 3.1 for a similar quantity. Thus we have reduced the problem to consideration of

$$2\int \Delta_n^*(x)[F^*(x)-F_0(x)]w(x)dx$$

$$= \frac{1}{n}\sum_{i=1}^{n} 2\int [I(X_i \le \sigma(F)x+\theta(F))-F_0(x)][F^*(x)-F_0(x)]w(x)dx ,$$

an average of i.i.d. r.v.'s with mean 0 and variance $A(F)$. $A(F)$ is easily seen to be finite by the second moment condition on F, and the CLT finishes the proof. □

For the next theorem we will need to assume existence of expansions of the form (3.3). If $F(x) = F_0((x-\theta)/\sigma)$ and $\Delta^*(x) = \Delta(\sigma x+\theta)$, define

$$\theta(F;\Delta) = \frac{\sigma}{AD-B^2}\left\{D\int -\Delta^*(x)f_0(x)w(x)dx + B\int \Delta^*(x)f_0(x)xw(x)dx\right\}$$

and (4.6)

$$\sigma(F;\Delta) = \frac{\sigma}{AD-B^2}\left\{A\int -\Delta^*(x)f_0(x)xw(x)dx + B\int \Delta^*(x)f_0(x)w(x)dx\right\} ,$$

where

$$A = \int f_0^2(x)w(x)dx \ , \ B = \int f_0^2(x)xw(x)dx \ , \ D = \int f_0^2(x)x^2w(x)dx \ .$$

THEOREM 4.2. *Let* $X_1,\ldots,X_n$ *be independent r.v.'s with distribution* $F(x) = F_0((x-\theta)/\sigma)$. *Let* w *and* w' *both satisfy* (2.2) *for some* $a > 0$ *and* $r \ge 0$ *and suppose that* w' *is continuous a.e. Lebesgue. Let* F_0 *have a finite* $(r+3)$th *moment and a bounded and continuous density* f_0. *If* (θ_n,σ_n) *satisfy*

$$n^{\frac{1}{2}}[\theta_n-\theta-\theta(F;F_n-F)] \xrightarrow{p} 0$$

and (4.7)

$$n^{\frac{1}{2}}[\sigma_n-\sigma-\sigma(F;F_n-F)] \xrightarrow{p} 0 \ ,$$

then

$$nd_{F_n}(\theta_n,\sigma_n) \xrightarrow{d} \int_{-\infty}^{\infty}\left\{f_0(x)[\theta(F_0;U\circ F_0)+x\sigma(F_0;U\circ F_0)]+U\circ F_0(x)\right\}^2 w(x)dx$$

$$= \int_0^1 U^2(t)w\circ F_0^{-1}(t)dF_0^{-1}(t) \qquad (4.8)$$

$$- \left(\frac{D}{AD-B^2}\right)\left(\int_0^1 U(t)w\circ F_0^{-1}(t)dt\right)^2 - \left(\frac{A}{AD-B^2}\right)\left(\int_0^1 U(t)F_0^{-1}(t)w\circ F_0^{-1}(t)dt\right)^2$$

$$+ 2\left(\frac{B}{AD-B^2}\right)\int_0^1 U(t)w\circ F_0^{-1}(t)dt\int_0^1 U(t)F_0^{-1}(t)w\circ F_0^{-1}(t) \ ,$$

where $U(t)$ *is the Brownian Bridge on* $C[0,1]$.

REMARKS. (i) The last 3 terms reflect estimation of the parameters (θ,σ). (ii) If f_0 and w are symmetric about 0, then $B = 0$ and the limit distribution is

$$\int_0^1 U^2(t)w\circ F_0^{-1}(t)dF_0^{-1}(t) - \frac{\left\{\int_0^1 U(t)w\circ F_0^{-1}(t)dt\right\}^2}{\int f_0^2(x)w(x)dx} - \frac{\left\{\int_0^1 U(t)F_0^{-1}(t)w\circ F_0^{-1}(t)dt\right\}^2}{\int f_0^2(x)x^2w(x)dx}$$

(iii) If $w(x) = f_0(x)$, then the first term is $\int U^2(t)dt$, the limiting form of the Cramer-von Mises statistic. (iv) The proof of Theorem 4.2 relies mainly on (4.7). Since other methods of estimation, e.g., maximum likelihood, lead to expansions similar to (4.7), the conclusion of the theorem is more generally applicable (except for the final form (4.8)).

EXAMPLE 4.1. For the logistic $F_0 = (1+e^{-x})^{-1}$ with $w(x) = 1$ the right-hand side of (4.8) becomes

$$\int_0^1 U^2(t)[t(1-1)]^{-1}dt-6\left\{\int_0^1 U(t)dt\right\}^2-\left\{\frac{18}{\pi^2-6}\right\}\left\{\int_0^1 U(t)\ln\left\{\frac{t}{1-t}\right\}dt\right\}^2 .$$

$$= A^2-B^2-C^2$$

The first term is the limiting form of the Anderson-Darling statistic. As in Anderson and Darling (1952), let

$$U(t) = [t(1-r)]^{\frac{1}{2}} \sum_{j=1}^{\infty} \frac{Z_j}{[j(j+1)]^{\frac{1}{2}}} f_j(t) ,$$

where $Z_1, \ldots$ are i.i.d. standard normal r.v.'s and the f_j are the Ferrer associated Legendre polynomials. Then

$$A^2 = \sum_{j=1}^{\infty} \frac{Z_j^2}{[j(j+1)]}$$

and $B^2 = Z_1^2/2$, the first *component* of A^2. Unfortunately, C^2 is *not* the second component: C turns out to be a linear combination of $Z_2, Z_4, \ldots$.

PROOF OF THEOREM 4.2. Call the right-hand side of (4.8) $T(U)$. We will show that

$$nd_{F_n}(\theta_n, \sigma_n) = T(U_n) + o_p(1) \ , \qquad (4.9)$$

where $U_n(t) = n^{\frac{1}{2}}[F_n(\sigma F_0^{-1}(t) + \theta) - t]$ is the empirical process. Then we use Pyke and Shorack (1968, Theorem 2.1) to construct U_n^* having the same distribution as U_n and such that

$$d(U_n^*, U) = \sup_{0<t<1} \left| \frac{U_n^*(t) - U(t)}{[t(1-t)]^{\frac{1}{2}-\delta}} \right| \xrightarrow{p} 0 \ .$$

It is easy to show $T(U_n^*) - T(U) \xrightarrow{p} 0$. For example, consider the difference of first terms,

$$\left|\int_{\infty}^{\infty}[U_n{*}\circ F_0(x)]^2 w(x)dx - \int_{-\infty}^{\infty}[U\circ F_0(x)]^2 w(x)dx\right|$$

$$\leq d(U_n{*},U)\left(d(U_n{*},U) + 2\sup_{0<t<1}\left|\frac{U(t)}{[t(1-t)]^{\frac{1}{2}-\delta}}\right|\right)\int_{-\infty}^{\infty}[F_0(x)(1-F_0(x))]^{1-2\delta}w(x)dx \ .$$

For δ sufficiently small the moment condition on F_θ implies that this last integral is finite. The other terms of $|T(U_n{*}) - T(U)|$ are handled similarly and thus $T(U_n{*})$ and $T(U_n)$ have limiting distribution $T(U)$.

To show (4.9) we first expand $d_F(\theta_n,\sigma_n)$ in a Taylor series about (θ,σ).

$$d_F(\theta_n,\sigma_n) = d_F(\theta,\sigma) + (\theta_n-\theta)\frac{\partial}{\partial\theta}d_F(\theta,\sigma) + (\sigma_n-\sigma)\frac{\partial}{\partial\sigma}d_F(\theta,\sigma)$$

$$+ \ (\theta_n-\theta)^2\tfrac{1}{2}\frac{\partial^2}{\partial\theta^2}d_F(\theta_n{*},\sigma_n{*}) \ + \ (\sigma_n-\sigma)^2\tfrac{1}{2}\frac{\partial^2}{\partial\sigma^2}d_F(\theta_n{*},\sigma_n{*})$$

$$+ \ (\theta_n-\theta)(\sigma_n-\sigma)\frac{\partial^2}{\partial\theta\partial\sigma}d_F(\theta_n{*},\sigma_n{*}) \ ,$$

where $(\theta_n{*},\sigma_n{*})$ lies "between" (θ_n,σ_n) and (θ,σ). (The conditions on F_0 and w are sufficient for $d_F(\theta,\sigma)$ to be twice continuously differentiable.) Note that the first three terms above are 0. Then

$$d_{F_n}(\theta_n,\sigma_n) = d_F(\theta_n,\sigma_n) + [d_{F_n}(\theta_n,\sigma_n) - d_F(\theta_n,\sigma_n)]$$

$$= (\theta_n-\theta)^2 \tfrac{1}{2} \frac{\partial^2}{\partial\theta^2} d_F(\theta_n^*,\sigma_n^*) + (\sigma_n-\sigma)^2 \tfrac{1}{2} \frac{\partial^2}{\partial\sigma^2} d_F(\theta_n^*,\sigma_n^*)$$

$$+ (\theta_n-\theta)(\sigma_n-\sigma) \frac{\partial^2}{\partial\theta\partial\sigma} d_F(\theta_n^*,\sigma_n^*) + \frac{1}{\sigma_n}\int \Delta_n^2(x) w\left(\frac{x-\theta_n}{\sigma_n}\right)dx \qquad (4.10)$$

$$+ \frac{2}{\sigma_n} \int \Delta_n(x)\left\{F(x) - F_0\left(\frac{x-\theta_n}{\sigma_n}\right)\right\} w\left(\frac{x-\theta_n}{\sigma_n}\right)dx \ ,$$

where $\Delta_n = F_n-F$. In order to match terms of $d_{F_n}(\theta_n,\sigma_n)$ with those of $T(U_n)$, the latter can be expressed as

$$\frac{1}{n}T(U_n) = \left(\theta(F;\Delta_n)\right)^2 \tfrac{1}{2} \frac{\partial^2}{\partial\theta^2} d_F(\theta,\sigma) + \left(\sigma(F;\Delta_n)\right)^2 \tfrac{1}{2} \frac{\partial^2}{\partial\sigma^2} d_F(\theta,\sigma)$$

$$+ \theta(F;\Delta_n)\sigma(F;\Delta_n) \frac{\partial^2}{\partial\theta\partial\sigma} d_F(\theta,\sigma) + \frac{1}{\sigma}\int \Delta_n^2(x) w\left(\frac{x-\theta}{\sigma}\right)dx \qquad (4.11)$$

$$+ \theta(F;\Delta_n) \frac{1}{\sigma}\int \Delta_n(x) f(x) w\left(\frac{w-\theta}{\sigma}\right)dx + \sigma(F;\Delta_n) \frac{1}{\sigma}\int \Delta_n(x) f(x)\left(\frac{x-\theta}{\sigma}\right) w\left(\frac{x-\theta}{\sigma}\right)dx \ .$$

Using (4.7) and the continuity of the second partial derivatives, the difference between the first four terms of (4.10) and (4.11) is seen to be $o_p(n^{-1})$. The fifth term of (4.10) can be expanded in a Taylor series about (θ,σ) and is equal to the last two terms of (4.11) plus a remainder which is also $o_p(n^{-1})$. Thus (4.9) holds. □

5. COMPLEMENTS

(i) If the observations $X_1,\ldots,X_n$ come from a *symmetric* distribution F, an approach reminiscent of R-estimators is available for estimating the center of symmetry. Choose $\hat{\theta}_n$ to minimize

$$\int_{-\infty}^{\infty}[1-F_n(\theta+x)-F(\theta-x)]^2 w(x)dx \ .$$

Under suitable regularity conditions $\hat{\theta}_n$ is asymptotically normal, and in addition, $\hat{\theta}_n$ is asymptotically efficient if w is chosen according to (1.2). Knüsel (1969) noted the result for $w(x) = 1$, where $\hat{\theta}_n = \text{median}\{(X_i+X_j)/2,\ 1 \le i \le j \le n\}$ is a version of the R-estimator derived from the Wilcoxon signed rank test.

(ii) In the two-sample shift problem, $X_1,\ldots X_m$ from $F(X-\Delta)$ and $Y_1,\ldots,Y_n$ from $F(x)$, $\hat{\Delta}_n$ is chosen to minimize

$$\int_{-\infty}^{\infty}[F_m(x+\Delta)-G_n(x)]^2 w(x)dx \ ,$$

where F_m and G_n are the respective empirical distribution functions. Once again the choice of w according to (1.2) yields asymptotically efficient estimates. Fine (1966) noted the result for $w(x) = 1$, where $\hat{\Delta} = \text{median}\{Y_i-X_j\ ,\ 1 \le i \le n,\ 1 \le j \le m\}$ is the R-estimator derived from the Wilcoxon rank sum test.

(iii) In the pure scale situation, $X_1,\ldots,X_n$ from $F_0(x/\sigma)$, σ can be estimated by minimizing

$$\int_{-\infty}^{\infty} [F_n(\sigma x)-F_0(x)]^2 w(x)dx \ .$$

The w which yields asymptotic efficiency is given by

$$w_0(x) = \frac{\frac{d}{dx}\left(-1-x\,\frac{f_0'(x)}{f_0(x)}\right)}{x\,f_0(x)} \ .$$

REFERENCES

Anderson, T. W., and Darling, D. A. (1952), "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," *Ann. Math. Statist.*, 23, 193-212.

Beran, R. (1977), "Minimum Hellinger Distance Estimates for Parametric Models," *Ann. Statist.*, 5, 445-463.

_____ (1978), "An Efficient and Robust Adaptive Estimator of Location," *Ann. Statist.*, 6, 292-313.

Blackman, J. (1955), "On the Approximation of a Distribution Function by an Empirical Distribution," *Ann. Math. Statist.*, 26, 256-267.

Bolthausen, E. (1977), "Convergence in Distribution of Minimum-Distance Estimators," *Metrika*, 24, 215-227.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956), "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator," *Ann. Math. Statist.*, 27, 642-669.

Fine, T. (1966), "On the Hodges and Lehmann Shift Estimator in the Two Sample Problem, *Ann. Math. Statist.*, 37, 1814-1818.

Frame, J. S. (1945), "Mean Deviation of the Binomial Distribution," *American Mathematical Monthly*, 52, 377-379.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Ann. Math. Statist.*, 35, 73-101.

Johnson, N. L., and Kotz, S. (1970), *Continuous Univariate Distributions*-2," Boston: Houghton Mifflin Co.

Knüsel, L. F. (1969), "Über Minimum-Distanz-Schätzungen," Published Ph.D. Thesis, Swiss Federal Institute of Technology, Zürich.

Parr, W. (1978), "Minimum Distance and Robust Estimation," preprint.

Pyke, R., and Shorack, G. (1968), "Weak Convergence of a Two-Sample Empirical Process and a New Approach to Chernoff-Savage Theorems," *Ann. Math. Statist.*, 39, 755-771.

Rao, P. V., Schuster, E. F., and Littell, R. C. (1975), "Estimation of Shift and Center of Symmetry Based on Kolmogorov-Smirnov Statistics," *Ann. Statist.*, 3, 862-873.

Wolfowitz, J. (1957), "The Minimum Distance Method," *Ann. Math. Statist.*, 28, 75-88.