The Annals of Statistics
1982, Vol. 10, No. 2, 621–624

ASYMPTOTIC PROPERTIES OF WEIGHTED L^2 QUANTILE DISTANCE ESTIMATORS

BY VINCENT N. LARICCIA

University of Nebraska-Lincoln

The asymptotic properties of a family of minimum quantile function distance estimators are considered. These procedures take as the parameter estimates that vector which minimizes a weighted L^2 distance between the empirical quantile function and an assumed parametric family of quantile functions. Regularity conditions needed for these estimators to be consistent and asymptotically normal are presented. For single parameter families of distributions, the optimal form of the weight function is presented.

1. Definitions. A family of minimum quantile distance estimators are proposed, and their properties investigated. The parameter estimates are taken as that vector which minimizes a weighted L^2 distance between the empirical quantile function and the assumed parametric family of quantile functions. Let $X_{(1)}, \cdots, X_{(n)}$ be the order statistics of a random sample of size n from the distribution function $F(x; \boldsymbol{\theta}^0)$, where $\boldsymbol{\theta}^0 \in \Theta \subseteq R^s$. For each $\boldsymbol{\theta} \in \Theta$, define the quantile function by

$$Q(u; \boldsymbol{\theta}) = \inf\{x: F(x; \boldsymbol{\theta}) \geq u\}, \quad 0 < u < 1,$$

and for continuous random variables the density quantile function by

$$fQ(u; \boldsymbol{\theta}) = f(Q(u; \boldsymbol{\theta}); \boldsymbol{\theta}), \quad 0 < u < 1$$

where $f(x; \boldsymbol{\theta}) = F'(x; \boldsymbol{\theta})$. Let $\Omega = \{Q(u; \boldsymbol{\theta}): \boldsymbol{\theta} \in \Theta \subseteq R^s$, and $Q(\cdot; \cdot)$ is a known function of $(u, \boldsymbol{\theta})\}$. Further, define the empirical quantile function (eqf) by

$$Q_n(u) = X_{(m)}, \quad (m-1)/(n+1) < u \leq m/(n+1), \quad m = 1, \cdots, n.$$

The members of the proposed family of estimators are now defined.

DEFINITION 1.1. Let $H(u; \boldsymbol{\theta})$ map $(0, 1) \times \Theta \to [0, \infty)$. The estimator is taken to be any vector $\boldsymbol{\theta} \in \Theta$ which minimizes the function

$$R(Q_n, \boldsymbol{\theta}) = \int_0^1 \{Q_n(u) - Q(u; \boldsymbol{\theta})\}^2 H(u; \boldsymbol{\theta})\, du. \tag{1}$$

The estimators are denoted by $\hat{\boldsymbol{\theta}}(Q_n, H)$.

Specific large sample properties of these procedures are investigated in Section 2. However, certain general points should be noted. Minimum distance estimators based upon the c.d.f., the p.d.f., and the characteristic function have been proposed and their properties investigated. These estimators have been shown to be consistent, asymptotically normal, to have excellent robust properties (both local and global), and to have good small sample properties. A general review and bibliography of minimum distance estimators is given in Parr (1981). In Section 2, regularity conditions are presented under which many of the above asymptotic properties are valid for $\hat{\boldsymbol{\theta}}(Q_n, H)$. Further, for many common distributions, $\hat{\boldsymbol{\theta}}(Q_n, H)$ will have computational advantages over other weighted L^2

Received December 1980; revised September 1981.
AMS 1970 *subject classification.* Primary 62F12; secondary 62G30.
Key words and phrases. Asymptotic normality, linear combinations of order statistics, minimum distance estimators.

estimators. For example, for location/scale families of distributions, $Q(u; \boldsymbol{\theta}) = \mu + \sigma Q_0(u)$ where $Q_0(u)$ is a known function. Hence if the weight function does not depend upon $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}(Q_n, H)$ is of closed form.

Besides estimates, these procedures also provide information on the appropriateness of the assumed parametric family of distributions. For example, $R(Q_n, \hat{\boldsymbol{\theta}})$ is a measure of the distance between the true quantile function of the data and Ω. Further, they are directly related to $Q - Q$ plotting techniques. Thus the estimators are easily incorporated into a statistical package which not only estimates the parameters but also provides both checks for the appropriateness of the assumed parametric family, and graphs which can be employed to propose a more appropriate family of distributions.

Finally, the least squares estimation technique for the mixture of two normal or lognormal populations, presented in Fowlkes (1979), is closely related to $\hat{\boldsymbol{\theta}}(Q_n, H)$ with $H(u; \boldsymbol{\theta}) = 1$. Fowlkes showed that, for these distributions, the small sample properties compared favorably with the maximum likelihood estimator.

2. Asymptotic properties. The consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}(Q_n, H)$ are shown in Theorem 2.1. This theorem can be proven under many varying sets of restrictions on the pair Ω and H. The more restrictions that are placed upon the tail behavior of the weight function, the fewer restrictions that are needed on Ω. Further, for most distributions $\sup_{0<u<1} |Q_n(u) - Q(u; \boldsymbol{\theta}^0)| = \infty$ a.s. (Csörgö and Révész, 1981, page 144), and convergence results can only be proven for weighted sup norms. Hence the assumptions required in this case tend to be more cumbersome than those needed for minimum c.d.f. distance estimators (Millar, 1981). For the above reasons, conditions on the functional $R(Q_n, \boldsymbol{\theta})$, instead of assumptions on Ω and H, are presented. Some possible sets of restrictions on Ω and H can be found in LaRiccia (1981).

Let $\mathscr{G}$ denote the class of left continuous functions on (0, 1) that are of bounded variation on $(\gamma, 1 - \gamma)$ for all $0 < \gamma < \frac{1}{2}$. For any metric $d(\cdot, \cdot)$ on $\mathscr{G}$ and for all $h, g \in \mathscr{G}$ and $\boldsymbol{\theta}, \gamma \in \Theta$, define

$$\rho_d[(h, \boldsymbol{\theta}), (q, \gamma)] = \max\{d(q, h), |\theta_1 - \gamma_1|, \cdots, |\theta_s - \gamma_s|\}.$$

For brevity, the following notational conventions are employed. For any function $Z(u; \boldsymbol{\theta})$ let $Z^i(u; \boldsymbol{\theta}) = \partial Z(u; \boldsymbol{\theta})/\partial\theta_i$, $Z^{ij}(u; \boldsymbol{\theta}) = \partial^2 Z(u; \boldsymbol{\theta})/\partial\theta_i\partial\theta_j$, $Z'(u; \boldsymbol{\theta}) = \partial Z(u; \boldsymbol{\theta})/\partial u$, and $Q_{\boldsymbol{\theta}} = Q(\cdot; \boldsymbol{\theta})$.

Also let $\mathbf{W}(Q_n; \boldsymbol{\theta}) = \partial R(Q_n, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $G(Q_n, \boldsymbol{\theta}) = \partial^2 R(Q_n, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^2$.

For Theorem 2.1 it is required that Ω and $H(u; \boldsymbol{\theta})$ are such that the following assumptions are satisfied.

(i′) $\boldsymbol{\theta}^0$ is an interior point of Ω.

(ii′) There exists a metric d: $\mathscr{G} \times \mathscr{G} \to R$ such that $\sqrt{n}\, d(Q_n, Q_{\boldsymbol{\theta}^0}) = 0_p(1)$.

(iii′) There exists an open neighborhood, A, of $(Q_{\boldsymbol{\theta}^0}, \boldsymbol{\theta}^0)$ such that $\mathbf{W}(Q, \boldsymbol{\theta})$ are $G(Q, \boldsymbol{\theta})$ are continuous functions with respect to the metric $\rho_d(\cdot, \cdot)$.

(iv′) The $s \times s$ matrix $G(Q_{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is positive definite and of full rank for all $\boldsymbol{\theta}$ is an open neighborhood of $\boldsymbol{\theta}^0$.

(v′) There exists a linear functional $\mathbf{T}(Q_n)$, where $\mathbf{T}$ is an $s \times 1$ vector with $T_i(Q_n) = \int_0^1 \{Q_n(u) - Q(u; \boldsymbol{\theta}^0)\} J_i(u)\, du$ and $\sqrt{n}\, \mathbf{T}(Q_n) \to_d N(\mathbf{0}, \Sigma)$ with

$$\Sigma_{ij} = \int_0^1 \int_0^1 J_i(u) J_j(v) \{fQ(u; \boldsymbol{\theta}^0) fQ(v; \boldsymbol{\theta}^0)\}^{-1}\{\min(u, v) - u \cdot v\}\, du\, dv,$$

such that $\sqrt{n}\, \| W(Q_n, \boldsymbol{\theta}^0) - T(Q_n)\| = o_p(d(Q_n, Q_{\boldsymbol{\theta}^0}))$.

THEOREM 2.1. *If Assumptions* (i′) *through* (iv′) *are satisfied, then*

(i) *As $n \to \infty$ there exists with probability tending to one, a unique function $\hat{\boldsymbol{\theta}}(Q_n, H)$, which locally minimizes $R(Q_n, \boldsymbol{\theta})$.*

(ii) *The function $\hat{\boldsymbol{\theta}}(Q_n, H)$ is a consistent estimator of $\boldsymbol{\theta}^0$.*

If in addition Assumption (v′) *is also satisfied, then*

(iii) $\sqrt{n}\ \{\hat{\boldsymbol{\theta}}(Q_n, H) - \hat{\boldsymbol{\theta}}^0\}$ *converges in distribution to an s-variate normal random variable, with covariance matrix* $C = G(Q_{\theta^0}, \boldsymbol{\theta}^0)^{-1}\Sigma G(Q_{\theta^0}, \boldsymbol{\theta}^0)^{-1}$.

PROOF. Since $R(Q_{\theta^0}, \boldsymbol{\theta}^0) = 0$, (i) implies that $\mathbf{W}(Q_{\theta^0}, \boldsymbol{\theta}^0) = \mathbf{0}$. Also Conditions (iii′) and (iv′) imply that the conditions of the implicit function theorem are satisfied. Thus there exists an open neighborhood, B, of $(Q_{\theta^0}, \boldsymbol{\theta}^0)$ and a unique continuous map, $\hat{\boldsymbol{\theta}}(Q_n, H)$ of an open ball $B_\alpha \subseteq \mathscr{G}$ containing Q_{θ^0}, into R^s such that $W(Q_n, \hat{\boldsymbol{\theta}}(Q_n, H)) = 0$ and $\hat{\boldsymbol{\theta}}(Q_{\theta^0}, H) = \boldsymbol{\theta}^0$. Since (ii′) implies $d(Q_n, Q_{\theta^0}) \to_p 0$, with a probability tending to one $Q_n \in B_\alpha$ and there exists, with probability tending to one, a unique point $\hat{\boldsymbol{\theta}}(Q_n, H)$ such that $W(Q_n, \hat{\boldsymbol{\theta}}(Q_n, H)) = 0$. Since $G(Q, \boldsymbol{\theta})$ is a continuous function, $G(Q_n, \hat{\boldsymbol{\theta}}(Q_n, H)) \to_p G(Q_{\theta^0}, \boldsymbol{\theta}^0)$. Thus by (iv′), $R(Q_n, \boldsymbol{\theta})$ is a minimum at $\hat{\boldsymbol{\theta}}(Q_n, H)$ and (i) has been shown. Further, the continuity of $\hat{\boldsymbol{\theta}}(\cdot, H)$ implies that $\hat{\boldsymbol{\theta}}(Q_n, H)$ is consistent, and part (ii) has been shown.

Letting $\boldsymbol{\theta}^* = \boldsymbol{\theta}^0 + g(u)\{\hat{\boldsymbol{\theta}}(Q_n, H) - \boldsymbol{\theta}^0\}$ where for all u, $g(u)$ is a diagonal matrix with elements between zero and one, a first order Taylor series expansion of $W(Q_n, \boldsymbol{\theta}^0)$ about $\hat{\boldsymbol{\theta}}(Q_n, H)$ gives

$$\sqrt{n}\ \{\hat{\boldsymbol{\theta}}(Q_n, H) - \boldsymbol{\theta}^0\} = -\,G(Q_n, \boldsymbol{\theta}^*)^{-1}\sqrt{n}\mathbf{W}(Q_n, \boldsymbol{\theta}^0)$$
$$= -\,G(Q_n, \boldsymbol{\theta}^*)^{-1}\{\sqrt{n}\mathbf{T}(Q_n) + o_p(1)\}.$$

Thus by the continuity of $G(Q, \boldsymbol{\theta})$, (iii′), and a Slutsky type argument $\sqrt{n}\ (\hat{\boldsymbol{\theta}}(Q_n, H) - \boldsymbol{\theta}^0) \to_d Z \sim N_s(\mathbf{0}, C)$, which shows (iii).

REMARK 1. Assumption (iv′) is similar to the usual identifiability assumption for minimum distance estimators (Millar, 1981). Weighted sup. and integral metrics which satisfy (ii′) can be found in Csörgö and Révész (1978, 1981) and Shorack (1972). Finally, for many continuous random variables $J_i(u) = H(u; \boldsymbol{\theta})Q^i(u; \boldsymbol{\theta})$ and for most reasonable weight functions, Assumptions (iii′) through (v′) are satisfied.

REMARK 2. For single parameter families of distributions, optimal forms of the weight function can be determined. Let $\theta^0 \in R$,

$$L(x; \theta) = \partial \ell_n f(x; \theta)/\partial\theta, \quad H_1(u; \theta) = \{\partial L(x; \theta)/\partial x\,|_{x=Q(u;\theta)}\}/Q^\iota(u; \theta).$$

If Ω and H_1 satisfy the conditions of Theorem 2.1 and $L(x; \theta)F^\iota(x; \theta) \to 0$ as $x \to \pm\infty$, then it is easily shown that the asymptotic variance of $\sqrt{n}\ \{\hat{\theta}(Q_n, H_1) - \theta^0\}$ attains its Cramer-Rao lower bound. Note that for scale or location families of distributions $H_1(u; \theta)$ does not depend upon θ, and hence $\hat{\theta}(Q_n, H)$ can be determined in closed form and are the optimal estimators developed in Chernoff, Gastwirth and Johns (1967). Finally note that for single parameter families, by slightly altering $R(u; \theta)$, fully efficient minimum quantile distance estimators for multiple type II censored samples can also be developed.

REMARK 3. Since the estimators are asymptotically equivalent to a weighted sum of s linear functions of the order statistics, it is easily shown that the influence curve for $\hat{\boldsymbol{\theta}}(Q_n, H)$ is the $s \times 1$ vector with ith element

$$IC_i(x; F, T) = \textstyle\sum_{j=1}^{s} a_{ij}\left[\displaystyle\int_0^1 \{1/fQ(u; \boldsymbol{\theta}^0)\}u\cdot Q^i(u; \boldsymbol{\theta}^0)H(u; \boldsymbol{\theta}^0)\,du - \int_{F(x;\boldsymbol{\theta}^0)}^1 \{1/fQ(u; \boldsymbol{\theta}^0)\}\,Q^i(u; \boldsymbol{\theta}^0)H(u; \boldsymbol{\theta}^0)\,du\right],$$

where a_{ij} is the (i, j)th element of $G(Q_{\theta^0}, \boldsymbol{\theta}^0)$. It should be noted that global robustness properties of these estimators can also be determined, in a method similar to that used in Parr (1980) and Millar (1981), by considering the properties of the estimator when $Q_n(u) \to Q(u) \notin \Omega$.

REFERENCES

CHERNOFF, H., GASTWIRTH, J. and JOHNS, M. V. (1967). Asymptotic distribution of linear combinations of order statistics with applications to estimation. *Ann. Math. Statist.* **38** 52–72.

CSÖRGÖ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882–894.

CSÖRGÖ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York.

FOWLKES, E. (1979). Some methods for studying the mixture of two normal (lognormal) distributions. *J. Amer. Statist. Assoc.* **74** 561–575.

LARICCIA, V. (1981). L^2 quantile distance estimators. Technical Report, Department of Mathematics and Statistics, Univ. of Nebraska-Lincoln.

MILLAR, P. (1981). Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete.* **55** 73–89.

PARR, W. and SCHUCANY, W. (1980). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.* **75** 616–624.

PARR, W. (1981). Minimum Distance Estimation: A Bibliography. *Commun. Statist.-Theor. Meth.* **A10** 1205–1224.

SHORACK, G. (1972). Functions of order statistics. *Ann. Math. Statist.* **43** 412–427.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEBRASKA-LINCOLN
LINCOLN, NEBRASKA 68588