CONSISTENCY OF THE MAXIMUM PRODUCT OF SPACINGS METHOD AND ESTIMATION OF A UNIMODAL DISTRIBUTION

Yongzhao Shao

Columbia University

Abstract: The first part of this paper gives some general consistency theorems for the maximum product of spacings (MPS) method, an estimation method related to maximum likelihood. The second part deals with nonparametric estimation of a concave (convex) distribution and more generally a unimodal distribution, without smoothness assumptions on the densities. The MPS estimator for a distribution function with a monotone density is shown to have a simple explicit representation analogous to the Grenander estimator, and is asymptotically minimax with respect to Kolmogorov-Smirnov type loss. A simple consistent MPS estimator for a unimodal distribution is also discussed.

Key words and phrases: Grenander estimator, nonparametric MPS estimator, spacings, total variation distance, unimodal density.

1. Introduction

The maximum product of spacings (MPS) method for estimating parameters in continuous univariate distributions was proposed by Cheng and Amin (1983) and independently by Ranneby (1984). A simple description of the MPS method goes as follows. Let $X_1, \cdots, X_n$ be i.i.d. observations from some unknown cdf F_{θ_0} in a collection of continuous univariate distribution functions $\{F_\theta : \theta \in \Theta\}$. To estimate θ_0, applying the probability integral transform $F_\theta(\cdot)$ to the order statistics $X_{1:n} \leq \cdots \leq X_{n:n}$ yields $0 \equiv F_\theta(X_{0:n}) \leq F_\theta(X_{1:n}) \leq \cdots \leq F_\theta(X_{n:n}) \leq F(X_{n+1:n}) \equiv 1$. The MPS estimator $\hat{\theta}_n$ of θ_0 maximizes the product of spacings, i.e.,

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \prod_{j=1}^{n+1} [F_\theta(X_{j:n}) - F_\theta(X_{j-1:n})]. \tag{1.1}$$

The MPS method is related to maximum likelihood (ML). When $F_\theta(x)$ has a density $f_\theta(x)$, the logarithm of $\prod_{j=1}^{n+1}[F_\theta(X_{j:n}) - F_\theta(X_{j-1:n})]$ can be approximated by

$$\sum \log f_\theta(X_{j:n})[X_{j:n} - X_{j-1:n}] = \sum \log f_\theta(X_{j:n}) + \sum \log[X_{j:n} - X_{j-1:n}]. \tag{1.2}$$

Since $\sum \log[X_{j:n} - X_{j-1:n}]$ is a constant not depending on θ, maximizing the product of spacings in (1.1) is asymptotically equivalent to maximizing $\sum \log f_\theta(X_{j:n})$, i.e., the log likelihood. Since the ML method generates asymptotically optimal estimates in many situations, the MPS estimator can be expected to have the same asymptotic optimalities. However the likelihood functions can be unbounded, as in the three parameter log-normal families or many mixture models (Le Cam (1990)), thereby leading to inconsistent estimates. Since the product of spacings in (1.1) is always bounded, the MPS method can generate asymptotically optimal estimates even when ML breaks down due to unbounded likelihood functions; see e.g., Cheng and Traylor (1995).

The consistency problem for the MPS method has been investigated previously by Cheng and Amin (1979), Ranneby (1984) and Shao and Hahn (1999). The last reference contains consistency theorems with the weakest regularity conditions among the existing results. Ekström (1997) showed that the consistency results of Shao and Hahn (1999) hold also for higher order spacings and for some variations of the MPS method. The regularity conditions for the previous consistency results are very general in the sense that they cover most of the known counterexamples for the ML method. However, these conditions are not *strictly* weaker than the classical regularity conditions for the consistency of the ML method (Wald (1949)). In particular, a local dominance-type condition (see Perlman (1972) for exact definitions) is known to be necessary for consistency of all the approximate MLE as defined in Wald (1949). This dominance condition is replaced by similar conditions, e.g., Condition 3.1 of Theorem 3.1 in Shao and Hahn (1999), for the MPS method. We show in Section 2 of this paper that the local dominance-type condition of Perlman (1972) and its replacement in Shao and Hahn (1999) are not necessary for the new consistency theorems for the MPS method. Of course, there are many other general consistency theorems for the MLE, but as pointed out by Le Cam (1986, p.621), the only available general ones are variants of Wald (1949). We point out that Wang (1985) introduced some interesting techniques to get consistency results applicable to parametric families as well as several nonparametric families including concave distribution functions. Namely, for any $\theta \in \Theta$ and any neighborhood $V_r(\theta_0)$ of the true parameter θ_0, Wang (1985) suggested working with the log-likelihood ratio of the type $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$, where $\theta_r(\theta)$ is chosen from $V_r(\theta_0)$ so that for any θ^* not in $V_r(\theta_0)$ there exists a neighborhood $B(\theta^*)$ of θ^* on which $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is dominated. For consistency of the ML method, these local dominance conditions cannot be removed without compensation, as can be seen from examples in Ferguson (1982) and Le Cam (1990). Theorem 2.2 provides consistency results for the MPS method assuming the existence of a density, but does not require assumptions 2 and 6 of Wald (1949).

Section 2 deals with the consistency result. Section 3 deals with the problem of nonparametric estimation of distributions with unimodal densities but without

smoothness conditions on the densities. An early study of such problems goes back to Grenander (1956), who obtained the nonparametric maximum likelihood estimate for a concave cdf on $[0, \infty)$ as the least concave majorant (LCM) of the empirical distribution function $F_n(x)$. The Grenander estimator is a very good spatially adaptive estimator from a nonasymptotic point of view, as discussed by Birgé (1989, 1997). It would be desirable to have such a spatially adaptive estimator for unimodal densities with unknown mode. The construction of the Grenander estimator can be easily extended to unimodal densities with a known mode (e.g., Lo (1986)). However, when the mode is unknown the likelihood function can be unbounded. Estimating a unimodal density with unknown mode has attracted the attention of many researchers. For example, Wegman (1969, 1970a, b) considered the constrained MLE by assuming a modal interval of length greater than some positive ϵ. Other estimates have been proposed by Reiss (1973) and Prakasa Rao (1983). More references can be found in Barlow, Bartholomew, Brenner and Brunk (1972), Robertson, Wright and Dykstra (1988), and Birgé (1997).

The MPS estimator for a cdf with a monotone density is shown, in Section 3 of this paper, to have a simple explicit representation analogous to the Grenander estimator. A simple consistent nonparametric MPS estimator for a general unimodal distribution with unknown mode is also discussed.

2. General Consistency Theorems for the MPS Method

If the maximum of the product of spacings $\sup_\theta \prod_{j=1}^{n+1} [F_\theta(X_{j:n}) - F_\theta(X_{j-1:n})]$ is not attained or only numerical solutions are feasible, we call $\{\hat{\theta}_n, n > 1\}$ an *asymptotic* MPS estimate of θ_0 if

$$\liminf_{n\to\infty} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})} \geq 0 \qquad \text{a.s.} \qquad (2.1)$$

Let $d_{TV}(P, Q) = \sup_A |P(A) - Q(A)|$ be the total variation distance between two probability measures P and Q. When densities f_P and f_Q exist for P and Q, we have $d_{TV}(P, Q) = \frac{1}{2} \int |f_P(x) - f_Q(x)| dx$; thus d_{TV} and L^1 distances generate equivalent topologies and give the same consistency results.

Theorem 2.1. *Let $X_1, \cdots, X_n$ be an i.i.d. sample from P_{θ_0} with $\theta_0 \in \Theta$, where (Θ, d) is a compact metric space with the property that if $\theta' \neq \theta \in \Theta$, $P_{\theta'} \neq P_\theta$. Suppose $\lim_{d(\theta', \theta) \to 0} d_{TV}(P_\theta', P_\theta) = 0$ and the cdf $F_{\theta_0}(x)$ is continuous in x. Then almost surely $\lim_{n\to\infty} d(\hat{\theta}_n, \theta_0) = 0$ for any asymptotic MPS estimator $\{\hat{\theta}_n\}$.*

Since we can identify θ with its probability measure P_θ, it suffices to prove the following version of Theorem 2.1.

Theorem 2.1*. *Let $X_1, \cdots, X_n$ be an i.i.d. sample from $P_0 \in \mathcal{P}$ where $\mathcal{P}$ is a family of probability measures on $\mathbf{R}$ such that $(\mathcal{P}, d_{TV})$ is a compact metric space for the total variation distance d_{TV}. Suppose P_0 has a continuous cdf F_{P_0}. Then any asymptotic MPS sequence $\{\hat{P}_n\}$ is consistent, i.e., $\lim_{n \to \infty} d_{TV}(\hat{P}_n, P_0) = 0$.*

The proof of Theorem 2.1* is in the Appendix.

Remark 2.1. When densities $f_\theta(x)$ exist for all θ in Θ, the conditions in Theorem 2.1 are easy to check. In particular if $f_\theta(x)$ is continuous for θ, then Scheffé's lemma implies that $\lim_{d_{(\theta', \theta)} \to 0} d_{TV}(P'_\theta, P_\theta) = 0$.

Remark 2.2. Ferguson (1982) gave parametric examples in which the MLE exists but is not consistent. However, it is easy to check that the conditions of Theorem 2.1 hold and thus all the asymptotic MPS estimates are consistent for Ferguson's examples. Theorem 2.1 works also for nonparametric families. For example, consider the problem of estimating some unimodal density on $[0, 1]$ with unknown mode. It is obvious that the likelihood function can be unbounded; thus the likelihood method cannot be directly applied. Note that the unimodal densities on $[0, 1]$ is a compact set equipped with the L^1 distance (Reiss (1973)) so Theorem 2.1 implies that the approximate MPS estimates are consistent.

Remark 2.3. Theorem 2.1 assumes that convergence of parameters implies the convergence of the corresponding probability measures in total variation distance. This condition is implied by the general assumptions for the consistency of the approximate MLE as in Wald (1949) and Le Cam (1953), according to the result of Landers and Rogge (1972). Moreover, when the parameter space is homeomorphic to a closed subset of the cube $[0, 1]^{N_0}$ for some integer N_0 and when densities exist, Corollary 1 in Le Cam and Schwartz (1960), together with its accompanying remarks, imply that the sufficient condition in Theorem 2.1 is also necessary for the existence of consistent estimates.

Next we present consistency theorem without the compactness assumption.

Theorem 2.2. *Let $\{X_1, \cdots, X_n\}$ be an i.i.d. sample from a density f_{θ_0} with $\theta_0 \in \Theta$, where Θ is a closed set of $\mathbf{R}^d$. Assume that $\theta \neq \theta'$ implies that $\{x : f'_\theta(x) \neq f_\theta(x)\}$ has positive measure. Suppose $\lim_{\theta' \to \theta} \int |f_{\theta'}(x) - f_\theta(x)| dx = 0$ and $\lim_{|\theta| \to \infty} f_\theta(x) = 0$ for almost all x. Then $\lim_{n \to \infty} \hat{\theta}_n = \theta_0$ a.s. for any asymptotic MPS estimator $\{\hat{\theta}_n\}$.*

The proof of Theorem 2.2 is in the Appendix.

3. Nonparametric MPS Estimates for a cdf with a Unimodal Density

Consider the problem of estimating a cdf F_0 knowing only that it is continuous. Since the empirical cdf $\mathbf{F}_n$ is not continuous, it may not be proper

as an estimator of F_0. *Pyke's modified empirical* π_n is a continuous cdf which puts equal mass $\frac{1}{n+1}$ on every spacing and is uniformly distributed on each finite spacing. It is easy to see that π_n is a MPS estimator of F_0 and

$$\sup_{-\infty < x < +\infty} |\pi_n(x) - \mathbf{F}_n(x)| = \frac{1}{n+1}. \tag{3.0}$$

Both π_n and $\mathbf{F}_n$ are asymptotically minimax with respect to the Kolmogorov-Smirnov type risk as defined in (3.1) (Dvoreetzky, Kiefer and Wolfowitz (1956)).

Suppose it is also known that the underlying cdf F_0 has a non-increasing density on $[0,1]$, Grenander (1956) applied the ML method very elegantly and obtained the nonparametric MLE of F_0: the least concave majorant (LCM) of $\mathbf{F}_n$. This estimator is also asymptotically optimal in the minimax sense. More specifically, Kiefer and Wolfowitz (1976) proved that the empirical distribution $\mathbf{F}_n$ is asymptotically minimax for estimating a concave (or convex) cdf for the following Kolmogorov-Smirnov type risk function

$$r(F_0, F_n^*) = E_F W(\sqrt{n} \sup_t |F(t) - F_n^*(t)|), \tag{3.1}$$

where $W(\cdot)$ is any nonnegative nondecreasing function on $[0, \infty)$ and $\int_0^\infty W(r) re^{-2r^2} dr < \infty$, $W(r) \not\equiv 0$. Moreover, Lemma B in Marshall (1970) implies that the Grenander estimator is also asymptotically minimax for estimating a concave cdf with respect to any risk function in (3.1).

A monotone density is a special case of a unimodal density. The likelihood approach is not directly applicable to the estimation of a unimodal density with unknown mode since the likelihood functions are unbounded. We derive the MPS estimates for unimodal distributions and show that they are spatially adaptive and asymptotically optimal in some minimax sense.

3.1. The MPS estimate for a concave distribution function

Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics of an i.i.d. sample of a concave distribution function F. Define $\sigma_0(F) = \sup\{x : F(x) = 0\}$ and $\sigma_1(F) = \inf\{x : F(x) = 1\}$, with the convention that $\sigma_1(F) = \infty$ if no x satisfies $F(x) = 1$.

Theorem 3.1. *The nonparametric MPS estimator for a concave cdf is asymptotically minimax with respect to any risk function in* (3.1) *and has the following simple form:*

(1) *The MPS estimator for a concave cdf with known finite support* $[\sigma_0(F), \sigma_1(F)]$ *is the least concave majorant (LCM) of Pyke's modified empirical distribution* π_n.

(2) *An MPS estimator of a concave cdf $F(x)$ with known support $[\sigma_0(F), \infty)$ is any concave distribution function which is $n/(n+1)$ times the least concave majorant (LCM) of $\mathbf{F}_n$ for x in $[\sigma_0(F), X_{n:n}]$ and which puts mass $1/(n+1)$ on $[X_{n:n}, \infty)$.*

(3) *An MPS estimator of any concave cdf with unknown support is a concave distribution function which is $n/n+1$ times the least concave majorant of the empirical distribution $\mathbf{F}_n(x)$ for x in $[X_{1:n}, X_{n:n}]$ and which puts mass $1/n+1$ on each of the intervals $(-\infty, X_{1:n}]$ and $[X_{n:n}, \infty)$.*

The proof of Theorem 3.1 is contained in the Appendix.

Remark 3.1. Note that the MPS estimator in cases (2) and (3) is not unique, since there are different ways to put the mass $1/(n+1)$ on $[X_{n:n}, \infty)$ to ensure concavity. For example, one can put mass $1/(n+1)$ uniformly on $[X_{n:n}, B]$ for some large B, or mass $1/(n+1)$ on $[X_{n:n}, \infty)$ using any density tail of the gamma type.

We give a numerical example to illustrate the nonparametric MPS estimate of a concave cdf. Nonparametric MPS estimates can be plotted easily. In the following, based on 20 uniform random points from $[0,1]$, we first plot Pyke's modified empirical distribution and its LCM (i.e., the MPS estimator) in Figure 2.1, then for comparison we plot the empirical cdf and its LCM (i.e., Greander's estimator) in Figure 2.2.
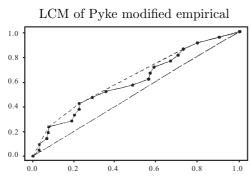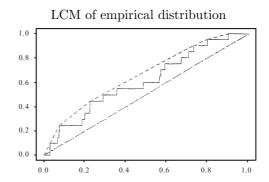

Figure 2.1. MPS Estimate

Figure 2.2. Greander Estimate

3.2. The nonparametric MPS estimate for a cdf with a unimodal density

If the mode of the underlying density is known, it is clear that the MPS method yields asymptotically minimax estimators. When the mode is unknown, the likelihood function can be unbounded and many difficulties arise for efficient estimation of the cdf. In fact, finding an asymptotically minimax estimator of a

cdf having a unimodal density with unknown mode has been an open problem since it was proposed by Kiefer and Wolfowitz (1976). Choosing a consistent estimator which is asymptotically efficient (or at least promising) is quite challenging. Of course, the empirical cdf $\mathbf{F}_n(x)$ (or $\pi_n(x)$) is consistent, in fact asymptotically minimax (Lo (1986)) for the risk in (3.1), but it is not unimodal so it may not be suitable as an estimator.

In general, one can construct a consistent unimodal MPS estimator of a unimodal distribution with unknown mode. It is also asymptotically minimax in all special cases studied. A sketch of the derivation of a unimodal MPS estimator goes as follows. Let $\mathcal{E} \equiv \{F(x) : F(x) \text{ is a cdf with a unimodal density } f(x)\}$. The MPS estimator of a unimodal cdf maximizes

$$\sup_{F \in \mathcal{E}} \prod_{i=1}^{n+1} [F(X_{i:n}) - F(X_{i-1:n})] = \max_{1 \leq j \leq n+1} \sup_{\{F \in \mathcal{F}_j\}} \prod_{i=1}^{n+1} [F(X_{i:n}) - F(X_{i-1:n})],$$

where $\mathcal{F}_j$ denotes the family of unimodal distributions with modes on the jth spacing. Without loss of generality, we may assume that the MPS estimator is in $\mathcal{F}_{j_0}$. For any $F \in \mathcal{F}_{j_0}$, F is convex on $(-\infty, X_{j_0-1:n}]$ and is concave on $[X_{j_0:n}, +\infty)$. Let $\alpha_F = F(X_{j_0-1:n})$ and $\beta_F = 1 - F(X_{j_0:n})$. Then $F(X_{j_0:n}) - F(X_{j_0-1:n}) = 1 - \alpha_F - \beta_F$ and

$$\prod_{i=1}^{n+1} [F(X_{i:n}) - F(X_{i-1:n})] = (1 - \alpha_F - \beta_F) \prod_{i \neq j_0} [F(X_{i:n}) - F(X_{i-1:n})]$$

$$= (1 - \alpha_F - \beta_F) \alpha_F^{j_0-1} \beta_F^{n-j_0+1} A_{j_0} B_{j_0},$$

where $A_{j_0} \equiv \alpha_F^{-1} \prod_{i=1}^{j_0-1} [F(X_{i:n}) - F(X_{i-1:n})]$, $B_{j_0} \equiv \beta_F^{-1} \prod_{i=j_0+1}^{n+1} [F(X_{i:n}) - F(X_{i-1:n})]$. Notice that B_{j_0} achieves its maximum value at the 'least concave majorant' of the function $(n - j_0 + 1)^{-1} \sum_{i=j_0+1}^{n+1} I_{(-\infty,x]}(X_{j:n})$. Consequently, $\sup_{F \in \mathcal{F}_{j_0}} B_{j_0}$ is a constant that does not depend on α_F and β_F (it only depends on the observations). A similar assertion can be made about A_{j_0}. Simple maximization of the right-hand side yields $\alpha_F = (j_0 - 1)/(n + 1)$ and $\beta_F = (n - j_0 + 1)/(n + 1)$, and a unimodal MPS estimate can thus be constructed. The consistency of such estimators in total variation distance is implied by Theorem 4.1 of Shao and Hahn (1999), however the justification of the asymptotic minimaxity of the MPS estimator with unknown mode is an interesting open problem.

Acknowledgements

would like to thank an associate editor for valuable comments. The author thanks Professor Jon A. Wellner for a conversation which motivated this study. This research is supported by NSF grant DMS-96-26658.

Appendix A. Proofs of Theorems 2.1*, 2.2 and 3.1

Proof of Theorem 2.1*. For ease of exposition first assume that, for each $P \in \mathcal{P}$, $F_P(x)$ has a density $f_P(x)$. If $P \neq P_0$, by Theorem 4.1 in Shao and Hahn (1995),

$$\limsup_{n \to \infty} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_P(X_{j:n}) - F_P(X_{j-1:n})}{F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n})} \leq \int_{\mathbf{R}} \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

Moreover, for some small positive number c,

$$\int_{\mathbf{R}} \log \frac{f_P(x) + c f_{P_0}(x)}{f_{P_0}(x)} dF_{P_0}(x) \leq \frac{1}{2} \int_{\mathbf{R}} \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

Thus without loss of generality, we assume that $f_P(x) \geq c f_{P_0}(x)$ for some $c > 0$. Next we show that for some small r_P and when n is large enough, we have

$$\sup_{Q \in b(P, r_P)} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{j:n}) - F_Q(X_{j-1:n})}{F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n})} \leq \frac{1}{4} \int_{\mathbf{R}} \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0,$$

(A.1)

where $b(P, r_P) \equiv \{Q \in \mathcal{P} : d_{TV}(Q, P) < r_P\}$. For $Q \in b(P, r_P)$, $\int |f_Q(x) - f_P(x)| dx < 2 r_P$. Denote $Y_{j,n}^P := F_P(X_{j+1:n}) - F_P(X_{j:n})$ for $1 \leq j \leq n+1$. Then for $Q \in b(P, r_P)$,

$$\frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{j:n}) - F_Q(X_{j-1:n})}{F_P(X_{j:n}) - F_P(X_{j-1:n})}$$

$$\leq \frac{1}{n+1} \sum_{j=1}^{n+1} \log \left(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{F_P(X_{j:n}) - F_P(X_{j-1:n})} \right)$$

$$\leq \frac{1}{n+1} \sum_{j=1}^{n+1} \log \left(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{c(F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n}))} \right). \qquad \text{(A.2)}$$

Let $\xi_1, \ldots, \xi_{n+1}$ be i.i.d. standard exponential random variables. Using the exponential representation of uniform spacings (see Shorack and Wellner (1986, p.335)) and the fact that for $a > 0$, $\log(1 + ab) \leq \log(1 + a) + |b - 1|$, we have

$$\frac{1}{n+1} \sum_{j=1}^{n+1} \log \left(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{c(F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n}))} \right)$$

$$\stackrel{\text{dist}}{=} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\xi_j} \cdot \frac{\xi_1 + \xi_2 + \cdots + \xi_{n+1}}{n+1} \Big)$$

$$\leq \frac{1}{n+1} \sum_{j=1}^{n+1} \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\xi_j} \Big) + \Big| \frac{\xi_1 + \xi_2 + \cdots + \xi_{n+1}}{n+1} - 1 \Big|.$$

For a small positive $\alpha < 1$ (its value will be selected later),

$$\frac{1}{n+1} \sum_{j=1}^{n+1} \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\xi_j} \Big)$$

$$= \frac{1}{n+1} \Big[\sum_{\xi_j \leq \alpha} + \sum_{\xi_j > \alpha} \Big] \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\xi_j} \Big). \tag{A.3}$$

Note that

$$\sum_{j=1}^{n+1} |Y_{j,n}^Q - Y_{j,n}^P| \leq \sum_{j=1}^{n+1} \int_{X_{j-1:n}}^{X_{j:n}} |f_Q(x) - f_P(x)| dx = 2d_{TV}(F_Q, F_P).$$

Thus

$$\frac{1}{n+1} \sum_{\xi_j > \alpha} \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\xi_j} \Big)$$

$$\leq \frac{1}{n+1} \sum_{\xi_j > \alpha} \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\alpha} \Big)$$

$$\leq \frac{1}{n+1} \sum_{\xi_j > \alpha} \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\alpha} \leq \frac{2d_{TV}(F_Q, F_P)}{c\alpha}. \tag{A.4}$$

Also notice that

$$\frac{1}{n+1} \sum_{\xi_j > \alpha} \log \Big(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c\xi_j} \Big)$$

$$= \frac{1}{n+1} \sum_{\xi_j \leq \alpha} \Big[\log \Big(\xi_j + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c} \Big) - \log \xi_j \Big].$$

By the SLLN

$$\frac{1}{n+1} \sum_{\xi_j \leq \alpha} \log \xi_j \to \int_0^\alpha e^{-x} \log x \, dx. \tag{A.5}$$

Moreover,

$$\frac{1}{n+1} \sum_{\xi_j \le \alpha} \log \left(\xi_j + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c} \right) \le \frac{1}{n+1} \sum_{\xi_j \le \alpha} \log \left(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c} \right)$$

$$\le \frac{1}{n+1} \sum_{\xi_j \le \alpha} \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c}$$

$$\le \frac{2d_{TV}(F_Q, F_P)}{c}. \qquad (A.6)$$

Thus

$$\frac{1}{n+1} \sum_{j=1}^{n+1} \log \left(1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{F_P(X_{j:n}) - F_P(X_{j-1:n})} \right)$$

$$\le \frac{2d_{TV}(F_Q, F_P)}{c} + \frac{2d_{TV}(F_Q, F_P)}{c\alpha} - \int_0^\alpha e^{-x} \log x \, dx.$$

Take c, α and $d_{TV}(F_Q, F_P) \le r_P := c\alpha^2$ very small such that the right-hand side of the above is less than $-(1/4) \int \log[f_P(x)/f_{P_0}(x)] dF_{P_0}(x)$, then (A.1) holds. Consequently, for any $P \ne P_0$ there exists a small $r_P > 0$ and $N_P \in \mathbf{N}$ such that whenever $n \ge N_P$,

$$\sup_{Q \in b(P, r_P)} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{j:n}) - F_Q(X_{j-1:n})}{F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n})} < \frac{1}{4} \int_{\mathbf{R}} \log \frac{g_P^c(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

For each $\delta > 0$, the neighborhoods $\{b(P, r_P) : P \in \mathcal{P} \backslash b(P_0, \delta)\}$ covers $\mathcal{P} \backslash b(P_0, \delta)$. Since $\mathcal{P} \backslash b(P_0, \delta)$ is compact, there exist finitely many neighborhoods, say $\{b(P_1, r_{P_1}), \cdots, b(P_k, r_{P_k})\}$ which cover $\mathcal{P} \backslash b(P_0, \delta)$. Let $N_\delta = \max(N_{P_1}, \cdots, N_{P_k})$. Then whenever $n \ge N_\delta$,

$$\sup_{Q \in \mathcal{P} \backslash b(P_0, \delta)} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{j:n}) - F_Q(X_{j-1:n})}{F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n})}$$

$$< \frac{1}{4} \max_{1 \le i \le k} \int_{\mathbf{R}} \log \frac{f_{P_i}(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

Hence, maximizing the product of spacings over $Q \in \mathcal{P}$ yields an element of $b(P_0, \delta)$ for each $\delta > 0$. Note that by applying the probability integral transform, it is not essential to assume the existence of densities in order for the above proof to be valid. Thus, consistency of the asymptotic MPS estimate is substantiated.

Proof of Theorem 2.2. By Theorem 2.1, it suffices to prove that the probability of having a sequence of unbounded MPS estimates $\hat{\theta}_n$ is zero. Suppose that,

with positive probability, we have a sequence of MPS estimates $\hat{\theta}_n \to \infty$. Then $f_{\hat{\theta}_n}(x) \to 0$ for almost all x. Egoroff's Theorem implies that there is a closed set K such that $P_{\theta_0}(K)$ can be arbitrarily close to 1, and $\lim_{n\to\infty} \sup_{x\in K} f_{\hat{\theta}_n}(x) = 0$. In particular, $P_{\hat{\theta}_n}(K) \to 0$ and $P_{\hat{\theta}_n}(K^c) \to 1$. The mapping arctan (x) : $[-\infty, +\infty] \to [-\pi/2, \pi/2]$ induces a topology on $[-\infty, +\infty]$ and makes it a compact metric space. On a compact space, $\{P_{\hat{\theta}_n}\}$ is tight and thus has a subsequence, say itself, which converges weakly to a probability measure P_∞. Note that K^c is an open set; thus it is the union of at most countably many disjoint open intervals, say $K^c = \cup(a_j, b_j)$. So $1 = \lim_{n\to\infty} P_{\hat{\theta}_n}(K^c) = \lim_{n\to\infty} \sum P_{\hat{\theta}_n}\{(a_j, b_j)\}$.

For $\delta > 0$, choose a large J such that $\sum_{j=1}^{J} P_{\hat{\theta}_n}(a_j, b_j) = P_{\hat{\theta}_n}(\cup_{j=1}^{J}(a_j, b_j)) > 1 - \delta$ for all large enough n. Note that $P_{\theta_0}\{\cup_{j=1}^{J}(a_j, b_j)\}$ can be arbitrarily small. Since $P_{\hat{\theta}_n} \to P_\infty$, we have $P_\infty(\cup_{j=1}^{J}[a_j, b_j]) \geq \limsup_{n\to\infty} P_{\hat{\theta}_n}(\cup_{j=1}^{J}[a_j, b_j]) > 1 - \delta$. Thus, there exist finitely many open intervals whose union, say U, contains $\cup_{j=1}^{J}[a_j, b_j]$, $P_\infty(U) > 1 - \delta$, and $P_{\theta_0}(U)$ can be arbitrarily small. The set U^c is a union of finitely many closed intervals with $P_\infty(U^c) < \delta$. Thus $\limsup_{n\to\infty} P_{\hat{\theta}_n}(U^c) \leq P_\infty(U^c) < \delta$. Consequently there exists a large enough N such that whenever $n \geq N$, we have $P_{\hat{\theta}_n}(U^c) < \delta$. Moreover,

$$\sum_{j=1}^{n+1} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})} = \Big(\sum_{j\in n_U} + \sum_{j\notin n_U} \Big) \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})},$$

where n_U is the index set of spacings $[X_{j:n}, X_{j+1:n}]$ which intersect the open set U. Since $\sum_{j\notin n_U}[F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})] \leq P_{\hat{\theta}_n}(U^c) < \delta$,

$$\sup_{\hat{\theta}_n} \frac{1}{n+1} \sum_{j\notin n_U} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})}$$

$$\leq \frac{1}{n+1} \sum_{j\notin n_U} \log \Big(\frac{\delta}{n - |n_U|} [F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})]^{-1} \Big)$$

$$= -\frac{n - |n_U|}{n+1} \log \Big(\frac{n - |n_U|}{n+1} \Big) + \frac{n - |n_U|}{n+1} \log \delta$$

$$-\frac{n - |n_U|}{n+1} \frac{1}{n - |n_U|} \sum_{j\notin n_U} \log([n + 1][F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})]).$$

By the SLLN, $\frac{n-|n_U|}{n+1} \to P_{\theta_0}(U)$. By Theorem 2.1 of Shao and Hahn (1995),

$$\frac{1}{n - |n_U|} \sum_{j\notin n_U} \log([n + 1][F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})]) \to -\gamma.$$

Thus almost surely,

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{j\notin n_U} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})}$$
$$\leq -P_{\theta_0}(U^c) \log P_{\theta_0}(U^c) + P_{\theta_0}(U^c)(\log \delta + \gamma).$$

Moreover, since $\prod_{j\in n_U}[F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})] \leq (\frac{1}{|n_U|})^{|n_U|}$,

$$\frac{1}{n+1} \sum_{j\in n_U} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})}$$
$$\leq \frac{1}{n+1} \sum_{j\in n_U} \log \left(\frac{1}{|n_U|} [F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})]^{-1} \right)$$
$$= -\frac{|n_U|}{n+1} \log \frac{|n_U|}{n+1} - \frac{|n_U|}{n+1} \frac{1}{|n_U|} \sum_{j\in n_U} \log((n+1)[F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})])$$
$$\to -P_{\theta_0}(U) \log P_{\theta_0}(U) + P_{\theta_0}(U)\gamma, \qquad \text{as} \quad n \to +\infty.$$

Thus

$$0 \leq \limsup_{n\to\infty} \frac{1}{n+1} \sum_{j\in n+1} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})}$$
$$\leq -P_{\theta_0}(U^c) \log P_{\theta_0}(U^c) - P_{\theta_0}(U) \log P_{\theta_0}(U) + \gamma + P_{\theta_0}(U^c) \log \delta.$$

This contradicts the fact that $P_{\theta_0}(U^c) \log \delta \to -\infty$ when $\delta \to 0$. Thus we cannot have a sequence of MPS estimates $\hat{\theta}_n \to \infty$ with positive probability.

Proof of Theorem 3.1. The MPS method does not depend on the version of the density, thus there is no loss of generality in assuming that each non-increasing density is left-continuous. First consider Case (1), i.e., both $\sigma_0(F)$ and $\sigma_1(F)$ are known finite constants. Without loss of generality, let $\sigma_0(F) = 0$ and $\sigma_1(F) = 1$. The MPS estimate of the unknown underlying distribution function is

$$\hat{F}_n = \arg \sup_{F\in\mathcal{F}} \prod_{i=1}^{n+1} [F(X_{j:n}) - F(X_{i-1:n})], \tag{A.7}$$

where $\mathcal{F} = \{F : F \text{ has a non-increasing left-continuous density on } [0,1]\}$. To derive the explicit form of the MPS estimate, recall the derivation of the MLE for a non-increasing density on $[0,\infty)$. Let f_F be the density of $F \in \mathcal{F}$. Then the MLE is defined as

$$\hat{f}_n = \arg \sup_{\{f_F : F\in\mathcal{F}\}} \prod_{i=1}^{n} f_F(X_{i:n}). \tag{A.8}$$

Given points $X_{1:n} < \cdots < X_{n:n}$, Grenander (1956) proved the following two results.

(I) To achieve the maximum in (A.8) over all non-increasing densities on $[0, \infty)$, it suffices to consider only those non-increasing densities which are step functions with possible jumps at the given points $X_{1:n} < \cdots < X_{n:n}$.

(II) The density f_F which achieves the maximum in (A.8) is the slope of the least concave majorant (LCM) of the function $(1/n) \sum_{j=1}^{n} I_{[0,x]}(X_{j:n})$.

Thus the MLE of a concave distribution function on $[0, \infty)$, (or on $[0, b)$ with b unknown) is the LCM of the empirical distribution function.

Next we prove that the MPS estimate of a concave distribution function is the LCM of Pyke's modified empirical π_n. Let $g_i(F) = [F(X_{i:n}) - F(X_{i-1:n})]/[X_{i:n} - X_{i-1:n}]$ for $1 \le i \le n+1$. Then by the concavity of F, $g_1(F) \ge g_2(F) \ge \cdots \ge g_n(F) \ge g_{n+1}(F)$. Define a step function $g_F(x)$ on $[0, 1)$ such that $g_F(x) \equiv g_i(F)$ for $x \in [X_{i-1:n}, X_{i:n})$. It is easy to see that $g_F(x)$ is a non-increasing density function on $[0, 1]$ with possible jumps only at the observation points. Furthermore, F maximizes the product of spacings in (A.7) if and only if the g_F so defined maximizes the following "likelihood"

$$\prod_{i=1}^{n+1} g_F(X_{i:n}).\tag{A.9}$$

One can regard $X_{1:n} \le \cdots \le X_{n:n} \le X_{n+1:n} \equiv 1$ as the $n+1$ ordered "observation points" from a non-increasing density on $[0, +\infty)$. By (I), maximizing (A.9) is equivalent to determining the MLE for non-increasing densities on $[0, +\infty)$ when the ordered "observations" $\{X_{1:n}, \cdots, X_{n:n}, X_{n+1:n}\}$ are given. Then (I) and (II) yield the explicit representation for "the MLE" which maximizes (A.9), i.e., the MPS estimate which maximizes (A.7). Thus, the MPS estimate of the underlying distribution function is the LCM of the following "empirical distribution function": $\mathbf{H}_{n+1}(x) = (n+1)^{-1} \sum_{i=1}^{n+1} I_{[0,x]}(X_{i:n})$. Equivalently, the MPS estimator of a concave distribution on $[0, 1]$ is the LCM of Pyke's modified empirical distribution π_n.

Next we study Case (2). Estimating a cdf with a non-increasing density on $[\sigma_0(F), \infty)$ is equivalent to estimating a cdf which has a nonincreasing density on $[\sigma_0(F), \sigma_1(F)]$ with $\sigma_1(F)$ unknown. For any non-increasing density function $g(x)$ on $[\sigma_0(F), +\infty)$, define

$$\alpha_g = \int_{X_{n:n}}^{\infty} g(t)dt, \qquad G_{\alpha_g}(x) = \int_{\sigma_0(F)}^{x} \frac{g(t)}{1 - \alpha_g} dt.$$

Then $\{G_{\alpha_g}(x) : x \in [\sigma_0(F), X_{n:n}], \ 1 > \alpha_g \ge 0\}$ is the family of all distribution functions with non-increasing densities on $[\sigma_0(F), X_{n:n}]$. Given the observations, the maximum (over α_g) of

$$\prod_{j=1}^{n} [G_{\alpha_g}(X_{j:n}) - G_{\alpha_g}(X_{j-1:n})]\tag{A.10}$$

does not depend on g. The maximum in (A.10) is achieved by spacings of the LCM of the empirical distribution function. Now consider the following decomposition of the product of spacings:

$$\prod_{j=1}^{n+1}[F(X_{j:n}) - F(X_{j-1:n})] = \Big(\int_{X_{n:n}}^{\infty} f(t)dt\Big)\prod_{j=1}^{n}[F(X_{j:n}) - F(X_{j-1:n})]$$

$$= \alpha_f(1-\alpha_f)^n\prod_{j=1}^{n}[F_{\alpha_f}(X_{j:n}) - F_{\alpha_f}(X_{j-1:n})] \leq \alpha_f(1-\alpha_f)^n M_n. \qquad (A.11)$$

Note that in (A.11), M_n does not depend on α_f. So maximizing $\prod_{j=1}^{n+1}[F(X_{j:n}) - F(X_{j-1:n})]$ over F is equivalent to maximizing $S(\alpha_f) \equiv \alpha_f(1-\alpha_f)^n M_n$ over α_f. Differentiating $S(\alpha_f)$ and setting the derivative equal to zero yields a maximum at $\alpha_f = 1/(n+1)$. Thus, the MPS estimator of a concave distribution function $F(x)$ on $[\sigma_0(F), \infty)$ is a concave distribution function which is $n/(n+1)$ times the least concave majorant of $\mathbf{F}_n$ for x in $[\sigma_0(F), X_{n:n}]$ and which puts mass $1/(n+1)$ on $[X_{n:n}, \infty)$. Such an estimator is easy to construct and is uniquely determined on $[\sigma_0(F), X_{n:n}]$.

Estimating a concave cdf on $[\sigma_0(F), \sigma_1(F))$ with $\sigma_0(F)$ unknown is an interesting problem, since the likelihood function can be unbounded without further constraints.

Given n i.i.d. observations from a distribution $F(x)$ which has a non-increasing density $f(x)$ on $[\sigma_0(F), +\infty)$, $f(x)$ is non-increasing on $[X_{1:n}, +\infty)$. For any non-increasing density function $g(x)$ which is not identically zero on $(X_{1:n}, +\infty)$, define

$$\beta_g = \int_{-\infty}^{X_{1:n}} g(t)dt \quad and \quad G_{\beta_g}(x) = \int_{X_{1:n}}^{x} \frac{g(t)}{1-\beta_g}dt.$$

Then $\{G_{\beta_g}(x) : x \in [X_{1:n}, \infty),\ 1 > \beta_g \geq 0\}$ is the family of all distribution functions with non-increasing densities on $[X_{1:n}, \infty)$. An argument similar to the proof of Case (2) shows that the MPS estimate of $F(x)$ is a concave distribution function which is $n/(n+1)$ times the least concave majorant of the empirical distribution $\mathbf{F}_n(x)$ for x in $[X_{1:n}, X_{n:n}]$ and which puts mass $1/(n+1)$ on each of the intervals $(-\infty, X_{1:n}]$ and $[X_{n:n}, \infty)$.

References

Barlow, R., Bartholomew, D., Brenner, J. and Brunk, H. (1972). *Statistical Inference under Order Restrictions.* John Wiley, New York.

Birgé, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17**, 1532-1549.

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25**, 970-981.

Cheng, R. C. H. and Amin, N. A. K. (1979). Maximum product of spacings estimation with application to the lognormal distribution. Math. Report 79-1, University of Wales Institute of Science and Technology, Cardiff.

Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J. Roy. Statist. Soc. Ser. B* **45**, 394-403.

Cheng, R. C. H. and Traylor, L. (1995), Non-regular maximum likelihood problems (with discussions). *J. Roy. Statist. Soc. Ser. B* **57**, 3-44.

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.

Ekström, M. (1997). Maximum spacing methods and limit theorems for statistics based on spacings. Doctoral Dissertation, Umea University, Sweden.

Ferguson, T. S. (1982). An inconsistent maximum likelihood estimate. *J. Amer. Statist. Assoc.* **77**, 831-834.

Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skan. Aktuarietidskr* **39**, 125-153.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahr. verw. Geb.* **34**, 73-85.

Landers, D. and Rogge, L. (1972). Characterization of the topologies used in the theory of maximum likelihood estimation. *Z. Wahr. verw. Geb.* **21**, 197-200.

Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. of Calif. Public. in Statist.* **1**, 277-329.

Le Cam, L. and Schwartz, L. (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* **31**, 140-150.

Le Cam, L. (1990). Maximum likelihood: An introduction. *Internat. Statist. Rev.* **59**, 153-171.

Lo, S. H. (1986). Estimation of a unimodal distribution function. *Ann. Statist.* **14**, 1132-1138.

Marshall, A. W. (1970). *Nonparametric Techniques in Statistical Inference* (Edited by M. L. Puri), 175-176. Cambridge University Press, Cambridge.

Perlman, M. D. (1972). On the strong consistency of approximate maximum likelihood estimators. *Proc. of the Sixth Berkeley Symposium* **1**, 263-282.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhya Ser. A* **31**, 23-36.

Pyke, R. (1965). Spacings. *J. Roy. Statist. Soc. Sec. B* **27**, 395-449.

Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scand. J. Statist.* **11**, 93-112.

Reiss, R. D. (1973). On measurability and consistency of maximum likelihood estimates for unimodal densities. *Ann. Statist.* **1**, 888-901.

Roberson, T., Wright, F. and Dykstra, R. (1988). *Order restricted Inference*, Wiley, New York.

Shao, Y. and Hahn, M. G. (1995). Limit Theorems for the logarithm of sample spacings. *Statist. Probab. Lett.* **24**, 121-132.

Shao, Y. and Hahn, M. G. (1999). Strong consistency of the maximum product of spacings estimates with applications in nonparametrics and in estimation of unimodal densities. *Ann. Inst. Statist. Math.* **51**, 31-49.

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

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595-601.

Wang, J. L. (1985). Strong consistency of approximate maximum likelihood estimators with applications to nonparametrics. *Ann. Statist.* **13**, 932-946.

Wegman, E. J. (1969). A note on estimating a unimodal density. *Ann. Math. Statist.* **40**, 1661-1667.

Wegman, E. J. (1970a, 1970b). Maximum likelihood estimation of a unimodal density function. *Ann. Math. Statist.* **41**, 457-471, 2169-2174.

Department of Statistics, Columbia University, New York, NY 10027, U.S.A.

E-mail: yshao@stat.columbia.edu