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 representa-
tion 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, spac-
ings, 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)
goes as follows. Let \(X_1, \ldots, X_n\) be i.i.d. observations from some unknown cdf
\(F_{\theta}\) in a collection of continuous univariate distribution functions \(\{F_{\theta} : \theta \in \Theta\}\).
To estimate \(\theta_0\), applying the probability integral transform \(F_{\theta}()\) 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 \(\theta_0\) maximizes the product of spacings, i.e.,
\[
\hat{\theta}_n = \arg \max_{\theta \in \Theta} \prod_{j=1}^{n+1} [F_{\theta}(X_{j:n}) - F_{\theta}(X_{j-1:n})].
\] (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
\[
\sum \log f_{\theta}(X_{j:n})[X_{j:n} - X_{j-1:n}] = \sum \log f_{\theta}(X_{j:n}) + \sum \log[X_{j:n} - X_{j-1:n}].
\] (1.2)
Since $\sum \log[X_{j:n} - X_{j-1:n}]$ is a constant not depending on $\theta$, 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 optimality. 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 $\theta_0$, Wang (1985) suggested working with the log-likelihood ratio of the type $\log[f(\theta)/f_{\theta_r}(\theta)]$, where $\theta_r(\theta)$ is chosen from $V_r(\theta_0)$ so that for any $\theta^* \not \in V_r(\theta_0)$ there exists a neighborhood $B(\theta^*)$ of $\theta^*$ on which $\log[f(\theta)/f_{\theta_r}(\theta)]$ 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 $\epsilon$. 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 $\theta_0$ if

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

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 $X_1, \ldots, X_n$ be an i.i.d. sample from $P_{\theta_0}$ with $\theta_0 \in \Theta$, where $(\Theta, d)$ 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 $F_{\theta}(x)$ 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 $\theta$ with its probability measure $P_\theta$, 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 $\mathbb{R}$ 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 $\theta$ in $\Theta$, the conditions in Theorem 2.1 are easy to check. In particular if $f_{\theta}(x)$ is continuous for $\theta$, then Scheffé’s lemma implies that $\lim_{\theta' \to \theta} d_{TV}(P_{\theta'}, P_{\theta}) = 0$.

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 $f_{\theta_0}$ with $\theta_0 \in \Theta$, where $\Theta$ is a closed set of $\mathbb{R}^d$. Assume that $\theta \neq \theta'$ implies that $\{x : f_{\theta}(x) \neq f_{\theta'}(x)\}$ has positive measure. Suppose $\lim_{\theta \to 0} \int |f_{\theta'}(x) - f_{\theta}(x)| \, dx = 0$ and $\lim_{|\theta| \to \infty} f_{\theta}(x) = 0$ for almost all $x$. Then $\lim_{n \to \infty} \theta_n = \theta_0$ 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 $F_n$ is not continuous, it may not be proper
as an estimator of \( F_0 \). Pyke’s modified empirical \( \pi_n \) 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 \( \pi_n \) is a MPS estimator of \( F_0 \) and

\[
\sup_{-\infty < x < +\infty} |\pi_n(x) - F_n(x)| = \frac{1}{n+1}.
\]

Both \( \pi_n \) and \( F_n \) are asymptotically minimax with respect to the Kolmogorov-Smirnov type risk as defined in (3.1) (Dvoretzky, 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 \( F_n \). This estimator is also asymptotically optimal in the minimax sense. More specifically, Kiefer and Wolfowitz (1976) proved that the empirical distribution \( F_n \) is asymptotically minimax for estimating a concave (or convex) cdf for the following Kolmogorov-Smirnov type risk function

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

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 \( \pi_n \).
(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 $F_n$ 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 $F_n(x)$ 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.

![LCM of Pyke modified empirical](image1)

![LCM of empirical distribution](image2)

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
The MPS estimator of a unimodal cdf maximizes \( \alpha \). Simple maximization of the right-hand side yields only depends on the observations. A similar assertion can be made about \( A \) frequently, sup \( A \) goes as follows. Let \( \mathcal{E} = \{ F(x) : F(x) \) is a cdf with a unimodal density \( f(x) \} \). The MPS estimator of a unimodal cdf maximizes

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

where \( F_j \) denotes the family of unimodal distributions with modes on the \( j^{th} \) spacing. Without loss of generality, we may assume that the MPS estimator is in \( F_{j_0} \). For any \( F \in F_{j_0} \), \( F \) 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

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

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

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\(_{F \in F_{j_0}} B_{j_0} \) is a constant that does not depend on \( \alpha_F \) and \( \beta_F \) (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**

The author is grateful to a referee who provided a long list of corrections and suggestions which greatly improved the presentation of this paper. He also
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),
\[
\limsup_{n \to \infty} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_P(X_{jn}) - F_P(X_{j-1:n})}{F_{P_0}(X_{jn}) - F_{P_0}(X_{j-1:n})} \leq \int_{\mathbb{R}} \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.
\]
Moreover, for some small positive number \( c \),
\[
\int_{\mathbb{R}} \log \frac{f_P(x) + cf_{P_0}(x)}{f_{P_0}(x)} dF_{P_0}(x) \leq \frac{1}{2} \int_{\mathbb{R}} \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.
\]
Thus without loss of generality, we assume that \( f_P(x) \geq cf_{P_0}(x) \) for some \( c > 0 \). Next we show that for some small \( r_P \) and when \( n \) is large enough, we have
\[
\sup_{Q \in b(P,r_P)} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{jn}) - F_Q(X_{j-1:n})}{F_{P_0}(X_{jn}) - F_{P_0}(X_{j-1:n})} \leq \frac{1}{4} \int_{\mathbb{R}} \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0,
\]
(A.1)
where \( b(P,r_P) = \{ Q \in \mathcal{P} : d_{TV}(Q,P) < r_P \} \). For \( Q \in b(P,r_P) \), \( f|Q(x) - f_P(x)|dx < 2r_P \). 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 \( Q \in b(P,r_P) \),
\[
\frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{jn}) - F_Q(X_{j-1:n})}{F_{P_0}(X_{jn}) - F_{P_0}(X_{j-1:n})}
\leq \frac{1}{n+1} \sum_{j=1}^{n+1} \log \left( 1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{F_P(X_{j:n}) - F_P(X_{j-1:n})} \right)
\leq \frac{1}{n+1} \sum_{j=1}^{n+1} \log \left( 1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{c(F_{P_0}(X_{jn}) - F_{P_0}(X_{j-1:n}))} \right). \tag{A.2}
\]
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
\[
\frac{1}{n+1} \sum_{j=1}^{n+1} \log \left( 1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|}{c(F_{P_0}(X_{jn}) - F_{P_0}(X_{j-1:n}))} \right)
\]
\[
\text{dist} \leq \frac{1}{n+1} \sum_{j=1}^{n+1} \log \left( 1 + \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c \xi_j} \right) \cdot \xi_1 + \xi_2 + \ldots + \xi_{n+1} \frac{1}{n+1}
\]

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

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

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

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

Also notice that

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

By the SLLN

\[
\frac{1}{n+1} \sum_{\xi_j \leq \alpha} \log \xi_j \rightarrow \int_{0}^{\alpha} e^{-x} \log x dx. \tag{A.5}
\]
Moreover,

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

\[
\leq \frac{1}{n+1} \sum_{j=1}^{n+1} \frac{|Y_{j,n}^Q - Y_{j,n}^P|(n+1)}{c} 
\]

\[
\leq \frac{2d_{TV}(F_Q, F_P)}{c}.
\]

Thus

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

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

Take \( c, \alpha \) and \( d_{TV}(F_Q, F_P) \leq r_P := c \alpha^2 \) 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 \( P \neq P_0 \) there exists a small \( r_P > 0 \) and \( N_P \in \mathbb{N} \) such that whenever \( n \geq N_P \),

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

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

\[
\sup_{Q \in \mathcal{P} \setminus b(P_0, \delta)} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_Q(X_{j:n}) - F_Q(X_{j-1:n})}{F_{P_0}(X_{j:n}) - F_{P_0}(X_{j-1:n})} < \frac{1}{4} \max_{1 \leq i \leq k} \int_{\mathbb{R}} \log \frac{f_{P_i}(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.
\]

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 \( \tilde{\theta}_n \) is zero. Suppose that,
for all large enough \( n \). In particular, \( P_{\theta_n}(K) \to 0 \) and \( P_{\theta_n}(K^c) \to 1 \). 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, \( \{P_{\theta_n}\} \) is tight and thus has a subsequence, say itself, which converges weakly to a probability measure \( P_\infty \). Note that \( K^c \) is an open set; thus it is the union of at most countably many disjoint open intervals, say \( K^c = \bigcup_{i=1}^\infty (a_i, b_i) \). Since \( K^c \) is open, we can partition \( K^c \) into at most countably many disjoint open intervals \( \delta \). Thus, there exist finitely many open intervals whose union, say \( U \), contains \( \bigcup_{j=1}^n (a_j, b_j) \), \( P_\infty(U) \geq 1 - \delta \), 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 \( \limsup_{n \to \infty} P_{\theta_n}(U^c) \leq P_\infty(U^c) < \delta \). Consequently, there exists a large enough \( N \) such that whenever \( n \geq N \), we have \( P_{\theta_n}(U^c) < \delta \). Moreover,

\[
\sum_{j=1}^{n+1} \log \frac{F_{\theta_n}(X_{j:n}) - F_{\theta_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})} = \left( \sum_{j \in n_U} + \sum_{j \notin n_U} \right) \log \frac{F_{\theta_n}(X_{j:n}) - F_{\theta_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})},
\]

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

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

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

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,

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

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

$$\frac{1}{n+1} \sum_{j \notin n_U} \log \frac{F_{\hat{\theta}_n}(X_{j:n}) - F_{\hat{\theta}_n}(X_{j-1:n})}{F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})} \leq \frac{1}{n+1} \sum_{j \notin n_U} \log \left( \frac{1}{|n_U|} [F_{\theta_0}(X_{j:n}) - F_{\theta_0}(X_{j-1:n})]^{-1} \right) $$

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

$$\to -P_{\theta_0}(U) \log P_{\theta_0}(U) + P_{\theta_0}(U) \gamma, \quad \text{as} \ n \to +\infty. \nonumber$$

Thus

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

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

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

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

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

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 \(\pi_n\). Let \(g_t(F) = [F(X_{i:n}) - F(X_{i-1:n})]/[X_{i:n} -
X_{i-1:n}]\) for \(1 \leq i \leq n + 1\). Then by the concavity of \(F\), \(g_1(F) \geq g_2(F) \geq \cdots \geq
g_n(F) \geq g_{n+1}(F)\). 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 \(X_{1:n} \leq \cdots \leq X_{n:n} \leq X_{n+1:n} \equiv 1\) 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 under-
lying distribution function is the LCM of the following “empirical distribution
function”: \(H_{n+1}(x) = (n + 1)^{-1} \sum_{i=1}^{n+1} I_{[0,x]} (X_{i:n})\). Equivalently, the MPS esti-
mator of a concave distribution on \([0, 1]\) is the LCM of Pyke’s modified empirical
distribution \(\pi_n\).

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

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

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

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

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

Note that in (A.11), \( M_n \) does not depend on \( \alpha_f \). 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 \( \alpha_f \). 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 \([\sigma(F)_0, \infty)\) is a concave distribution function which is \( n/(n + 1) \) times the least concave majorant of \( F_n \) for \( x \) in \([\sigma(F)_0, X_{n:n}]\) and which puts mass \( 1/(n + 1) \) on \([X_{n:n}, \infty)\). Such an estimator is easy to construct and is uniquely determined on \([\sigma(F)_0, X_{n:n}]\).

Estimating a concave cdf on \([\sigma(F)_0, \sigma(F)_1]\) with \( \sigma(F)_0 \) 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 \([\sigma(F)_0, +\infty), f(x) \) 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

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

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 \( F_n(x) \) 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


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

E-mail: yshao@stat.columbia.edu

(Received March 1998; accepted March 2001)