Ann. Inst. Statist. Math.
Vol. 51, No. 1, 31–49 (1999)

STRONG CONSISTENCY OF THE MAXIMUM PRODUCT OF SPACINGS ESTIMATES WITH APPLICATIONS IN NONPARAMETRICS AND IN ESTIMATION OF UNIMODAL DENSITIES

YONGZHAO SHAO[1]* AND MARJORIE G. HAHN[2]**

[1]*Department of Statistics, Columbia University, New York, NY 10027, U.S.A.*
[2]*Department of Mathematics, Tufts University, Medford, MA 02155, U.S.A.*

(Received November 14, 1996; revised June 16, 1997)

Abstract. Distributions with unimodal densities are among the most commonly used in practice. However, for many unimodal distribution families the likelihood functions may be unbounded, thereby leading to inconsistent estimates. The maximum product of spacings (MPS) method, introduced by Cheng and Amin and independently by Ranneby, has been known to give consistent and asymptotically normal estimators in many parametric situations where the maximum likelihood method fails. In this paper, strong consistency theorems for the MPS method are obtained under general conditions which are comparable to the conditions of Bahadur and Wang for the maximum likelihood method. The consistency theorems obtained here apply to both parametric models and some nonparametric models. In particular, in any unimodal distribution family the asymptotic MPS estimator of the underlying unimodal density is shown to be universally L^1 consistent without any further conditions (in parametric or nonparametric settings).

Key words and phrases: Asymptotic MPS estimator, L^1 consistency, monotone density, nonparametric, spacing, unimodal density.

1. Introduction

The maximum likelihood (ML) method is the most widely used statistical estimation technique. Under very general conditions, the maximum likelihood estimates (MLE) are consistent and asymptotically efficient. However, the likelihood function can be unbounded, as in the widely used three-parameter lognormal and Weibull models (Cox and Hinkley (1974), Smith (1985)) and in some mixture models (e.g. Lindsay (1995)). This can lead to inconsistent estimates. Various alternative methods might be considered to handle each specific case. Of particular

* Supported partially by NSF grants DMS-92-04333 and DMS-96-26658.
** Supported partially by NSF grant DMS-92-04333.

interest is the following spacing-based method, introduced by Cheng and Amin (1983) and independently by Ranneby (1984). This method does not suffer from unbounded likelihood difficulties and it provides a natural way to preserve the essential asymptotic optimalities that the MLE usually possesses.

To fix the idea, let $X_1, \ldots, X_n$ be independent observations from a continuous univariate distribution F_{θ_0} belonging to $\{F_\theta : \theta \in \Theta\}$. Consider the problem of estimating the unknown true θ_0 (or F_{θ_0}). Applying the *probability integral transform* $F_\theta(\cdot)$ to the order statistics $X_{1,n} \le \cdots \le X_{n,n}$ yields: $0 \equiv F_\theta(X_{0,n}) \le F_\theta(X_{1,n}) \le \cdots \le F_\theta(X_{n,n}) \le F_\theta(X_{n+1,n}) \equiv 1$. The *maximum product of spacings* (MPS) *method* chooses $\hat{\theta}_n$ (or $F_{\hat{\theta}_n}$) as the MPS *estimator* of θ_0 (or F_{θ_0}), which maximizes the product of spacings, i.e.,

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

Note that the spacings sum to 1. The definition given in (1.1) is motivated by trying to set $F_{\hat{\theta}_n}$ equal to the *true* but unknown distribution for which the spacings are identically distributed.

To observe the similarities and differences between the ML and MPS formulations, note that when F_θ has a density f_θ, the log likelihood can be written as $\Sigma \log f_\theta(X_{i,n})$; while the log product of spacings $\Sigma \log[F_\theta(X_{i,n}) - F_\theta(X_{i-1,n})]$ may be approximated by $\Sigma \log[f_\theta(X_{i,n})(X_{i,n} - X_{i-1,n})]$. Since $\sum \log[X_{i,n} - X_{i-1,n}]$ is a constant, we have

$$\sup_\theta \sum \log[f_\theta(X_{i,n})(X_{i,n} - X_{i-1,n})] \propto \sup_\theta \sum \log f_\theta(X_{i,n}). \tag{1.2}$$

Roughly, (1.2) says that maximizing $\Sigma \log[F_\theta(X_{i,n}) - F_\theta(X_{i-1,n})]$ is *approximately* equivalent to maximizing $\Sigma \log f_\theta(X_{i,n})$. This suggests that the two methods should lead to similar estimates when the ML method works. On the other hand, the product of spacings is always bounded and hence is more stable than the likelihood function. Therefore, the MPS method might give good estimates when the ML method is unstable.

The most widely used distributions in practice are those with unimodal densities. In particular, the monotone densities are a special case of unimodal densities. The nonparametric MLE is consistent for unimodal densities with known mode or with unknown mode but whose modal interval I_f has length bounded away from zero, i.e. $\inf_f |I_f| > \delta > 0$ (Wegman (1970)). However, consistency of the MLE fails when the location of the mode is unknown and $\inf_f |I_f| = 0$, due to unboundedness of the likelihood function. It is thus of interest to investigate the behavior of the MPS estimator for unimodal densities. It is also of interest to have consistency theorems for the MPS estimator under conditions at least as general as the classical general conditions (e.g. Bahadur (1971), Wang (1985)) for consistency of the MLE. Moreover, in reliability theory, survival analysis and other applied fields (Robertson *et al.* (1988)), it is important to estimate the unknown cumulative distribution function (cdf) which is only known to have a monotone density (or monotone failure rate) or more generally, a unimodal density. It is desirable to have theorems covering these nonparametric problems together with the parametric ones.

In this paper we investigate these general consistency problems for the MPS estimator, exemplified by the MPS estimates for unimodal distributions. Of course, consistency is just the first step towards asymptotic analysis. One optimality of the ML method is that under standard regularity conditions (e.g. Lehmann (1991), Theorem 6.2.3), the MLE is consistent and asymptotically normal with variance equal to the inverse of the expected Fisher information. Roughly speaking, when consistency of the estimator is established, one can focus on a small neighborhood to prove the asymptotic normality. As argued in (1.2), the MPS method and the ML method are asymptotically equivalent. Thus, when consistency of the MPS estimate is established, one might also expect the MPS estimate to have the same asymptotic normality behavior as the MLE. In fact, Shao and Hahn (1994) established a CLT for the MPS estimator under the same regularity conditions as appeared in Lehmann ((1991), Theorem 6.2.3) for the asymptotic normality of the MLE. More general results in this direction are currently under investigation. Hence this paper is focused solely on consistency.

The paper is organized as follows. Section 2 discusses previous work related to the MPS method. General consistency of the maximum product of spacings method is discussed in Section 3. Consistency theorems for the MPS method are obtained under very general conditions which are comparable to the conditions of Bahadur (1971) and Wang (1985) for the maximum likelihood method. The consistency theorems obtained here apply to both exact and asymptotic MPS estimators, also to both parametric models and some nonparametric models as well. In particular, in Section 4, the AMPS estimates are shown to be universally consistent for any unimodal distribution without additional conditions.

2. Remarks and lemmas

The MPS method was first introduced for estimating parameters in univariate continuous distribution families by Cheng and Amin (1983) and independently by Ranneby (1984). Cheng and Amin (1983) proved that the MPS estimates in the three-parameter lognormal model, the Weibull model and the Gamma model are consistent and asymptotically efficient, while the ML method is known to break down since the likelihood functions can be unbounded in these cases. Ranneby (1984) observed that a good inference method ought to minimize some suitable distance between the distribution and the model. He proposed the MPS method to minimize the Kullback-Leibler divergence. Ranneby also discussed some examples with inconsistent MLE (e.g. the mixture of normal distributions) and obtained consistent MPS estimates. Titterington (1985) remarked that the MPS method can be regarded as a maximum likelihood method for grouped data. Lind (1994) discussed the connection between the MPS method and information theory.

One advantage of the MPS method over the ML approach is that the same statistic

$$S_n(F) = (n+1)^{-1}\sum_{i=1}^{n+1} \log\{(n+1)[F(X_{i,n}) - F(X_{i-1,n})]\} \tag{2.0}$$

which is maximized to get the MPS estimator, can also be used to form a goodness-of-fit test to check the validity of a given model, i.e. test $H_0 : F = F_0$ vs. $H_1 :$

$F \neq F_0$ based on $S_n(F_0)$. In fact, the goodness-of-fit test based on $S_n(F_0)$ has been considered in Darling (1953), Pyke (1965), Gebert and Kale (1969), Shorack (1972), Cressie (1976), Cheng and Stephens (1989), Roeder (1990, 1992), and Shao and Hahn (1996).

Discussion of the major drawbacks of the MPS method (e.g. lack of a multivariate counterpart) can be found in Cheng and Traylor (1995). As argued intuitively in Section 1, the asymptotic optimality of the ML method in regular cases (in the sense of Cramér) may be well achieved by the MPS method in more general cases. This paper focuses on the problem of general consistency of the MPS method and pays special attention to estimation of distributions with unimodal densities.

The maximum of the product of spacings $\mathcal{P}_n(F_\theta, X) := \prod[F_\theta(X_{j,n}) - F_\theta(X_{j-1,n})]$ may not be achievable or may not be easy to find. Thus, we define the *asymptotic maximum product of spacings* (AMPS) estimate of θ_0 to be $\hat{\theta}_n$ such that

$$\varliminf_{n\to\infty} \frac{\mathcal{P}_n(F_{\hat{\theta}_n}, X)}{\mathcal{P}_n(F_{\theta_0}, X)} \geq C \quad \text{for some} \quad 0 < C \leq 1.$$

Without loss of generality, we will choose $C = 1$. The next proposition gives a useful criterion for determining whether or not a sequence of estimates are AMPS estimates.

LEMMA 2.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. with a continuous distribution function F_{θ_0}. Then $\{\hat{\theta}_n : n \geq 1\}$ is a sequence of AMPS estimates of θ_0 iff*

$$\varliminf_{n\to\infty} S_n(F_{\hat{\theta}_n}) \geq -\gamma \quad \textit{almost surely,} \tag{2.1}$$

where γ is Euler's Constant $\gamma = 0.577\cdots$.

PROOF. $\{\hat{\theta}_n : n \geq 1\}$ is a sequence of asymptotic MPS estimates iff

$$\varliminf_{n\to\infty} S_n(F_{\hat{\theta}_n}) \geq \varlimsup_{n\to\infty} S_n(F_{\theta_0}).$$

Since $\{F_{\theta_0}(X_{j,n}), 1 \leq j \leq n\}$ is an ascending sequence of uniform order statistics, the proof is completed upon application of the following strong law of large numbers for the logarithm of uniform spacings (Theorem 1.2 from Shao and Hahn (1995)):

$$\lim_{n\to\infty} S_n(F_{\theta_0}) = -\gamma \quad \text{a.s.} \qquad \square$$

Similarly, the maximum of the likelihood function may not be achievable, so one defines the *approximate maximum likelihood estimate* (AMLE) as in Wald (1949). Results on general consistency of the AMLE, can be found in Wald (1949), Le Cam (1953), Kiefer and Wolfowitz (1956), Huber (1967), Bahadur (1971), Perlman (1972) and Wang (1985). But as pointed out by Le Cam ((1986), p. 621), the only available general ones are variants of a result of Wald (1949).

To prove general consistency for the AMPS estimator we will use some techniques appearing in Bahadur (1971) and Wang (1985). Bahadur (1971) extends the results in Wald (1949) for a finite dimensional parameter space to a general parameter space. Wang (1985) obtains general consistency theorems for the AMLE which are applicable to parametric models as well as some nonparametric models. Our consistency theorems also apply to both exact and asymptotic MPS estimates and to parametric models as well as some nonparametric models. Just as the strong law of large numbers plays a significant role in Wald's proof of consistency of the AMLE, the following lemma is fundamental to the proof of consistency of the AMPS estimates in Section 3.

$F(x)$ is called a *pseudo-distribution* if and only if $F(x)/F(\infty)$ is a probability distribution function (cdf).

LEMMA 2.2. (Information-type inequality) *Suppose $X_1, X_2, \ldots, X_n$ is an i.i.d. sample with a continuous univariate cdf F_{θ_0}. For any pseudo-distribution F_θ, define $F_\theta(X_{0,n}) \equiv F_\theta(-\infty)$ and $F_\theta(X_{n+1,n}) \equiv F_\theta(+\infty)$. Denote the derivative of $F_\theta(F_{\theta_0}^{-1}(x))$ by $g_{\theta,\theta_0}(x)$. Then, almost surely,*

$$\overline{\lim_{n\to\infty}} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_\theta(X_{j,n}) - F_\theta(X_{j-1,n})}{F_{\theta_0}(X_{j,n}) - F_{\theta_0}(X_{j-1,n})} \le \int_{R} \log g_{\theta,\theta_0}(x) dF_{\theta_0}(x). \tag{2.2}$$

If F_{θ_0} and F_θ have densities f_{θ_0} and f_θ respectively, with $F_\theta \neq F_{\theta_0}$, then

$$\overline{\lim_{n\to\infty}} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_\theta(X_{j,n}) - F_\theta(X_{j-1,n})}{F_{\theta_0}(X_{j,n}) - F_{\theta_0}(X_{j-1,n})} \le \int_{R} \log \frac{f_\theta(x)}{f_{\theta_0}(x)} dF_{\theta_0}(x) < 0. \qquad a.s.$$

PROOF. See Theorem 4.1 in Shao and Hahn (1995). □

3. Consistency theorems

As pointed out by Bahadur (1971), the maximum that can be expected from a statistical estimation procedure is successful estimation of the entire underlying distribution. Convergence of the estimates to particular parameters is a relatively subsidiary issue determined by how well the true distribution can be estimated if the model is well parameterized. When the parameter is a well-defined continuous functional of the distributions, convergence of the estimated distributions to the true distribution certainly yields convergence of the parameter estimates to the true parameter. However, a non-continuous parametrization can easily fail any general statistical procedure, such as maximum likelihood, as clearly seen from the simple example of Basu (1955). Under general assumptions for consistency of the AMLE found in Wald (1949) or Le Cam (1953), Landers and Rogge (1972) show that the usual convergence for parameters is equivalent to the convergence of the corresponding probability measures in the total variation distance. To deal with the general consistency problem of the AMPS estimator, it is natural to take the probability measures as parameters just as Bahadur (1971) did for the AMLE.

Let $\mathcal{P}$ be a family of probability measures on $\boldsymbol{R}$ dominated by Lebesgue measure λ and let $f_P(x) = \frac{dP}{d\lambda}(x)$ for all P in $\mathcal{P}$. Suppose $X_1, X_2, \ldots, X_n$ is an i.i.d. sample from $P_0 \in \mathcal{P}$. We want to estimate the unknown P_0 and give conditions that can ensure that the AMPS sequence $\{\hat{P}_n\}$ converges to P_0.

First note that, by the information-type inequality, if $\mathcal{P}$ is a finite set then the MPS estimators $\{\hat{P}_n\}$ equal P_0 after some finite n_0. The distributions do not need to be absolutely continuous. If $\mathcal{P}$ is not finite, the hope is that the MPS estimates might behave similarly as $\mathcal{P}$ being a finite set, e.g. a compact set. Although $\mathcal{P}$ is not generally compact, it can be embedded in a compact set $\bar{\mathcal{P}}$. Different compactifications may be preferred in different situations. However, it is always possible to compactify $\mathcal{P}$ in the topology of vague convergence of subprobability measures as follows: Consider the space $\mathcal{S}$ of all subprobability measures on $\boldsymbol{R}$ endowed with the topology of *vague convergence* (Chung (1974)). $\mathcal{S}$ is a metrizable, compact, topological space. Let d be any distance defined on $\mathcal{S} \times \mathcal{S}$ such that $\lim_{k\to\infty} d(Q_k, Q_0) = 0$ if and only if $\{Q_k, k \in \boldsymbol{N}\} \subset \mathcal{S}$ converges vaguely to Q_0. In particular, if $\{Q_k, k \in \boldsymbol{N}\}$ is a sequence of probability measures and Q_0 is a probability measure, then $\lim_{k\to\infty} d(Q_k, Q_0) = 0$ if and only if $\{Q_k, k \in \boldsymbol{N}\}$ converges weakly to Q_0. Notice that $(\mathcal{P}, d)$ is a subspace of the metric space $(\mathcal{S}, d)$. Let $\bar{\mathcal{P}}$ be the closure of $\mathcal{P}$ in $\mathcal{S}$. Then $(\bar{\mathcal{P}}, d)$ is a compact metric space.

Given any compactification $\bar{\mathcal{P}}$ of $\mathcal{P}$, if $P \in \bar{\mathcal{P}} \setminus \mathcal{P}$, then P is a subprobability measure which may not be dominated by the measure λ. However, according to the Lebesgue Decomposition Theorem, there is a unique representation $P = P_{ac} + P_s$, where $P_{ac} \ll \lambda$ and $P_s \perp \lambda$. In this situation, define *the subdensity of* P as $f_P(x) = \frac{dP_{ac}}{d\lambda}(x)$. Then, associated to any family $\mathcal{P}$ of probability measures dominated by λ, there is a compactification $\bar{\mathcal{P}}$ of $\mathcal{P}$ with subdensities $\{f_P(x) : P \in \bar{\mathcal{P}}\}$ as defined above.

For any $P \in \bar{\mathcal{P}}$, let $S_P(x,r) \equiv \sup\{f_Q(x) : d(Q,P) < r, Q \in \mathcal{P}\}$. The following are three of the basic conditions for consistency of AMLE in Bahadur (1971):

CONDITION A. $\bar{\mathcal{P}}$ is a "suitable compactification" of $\mathcal{P}$, i.e. $S_P(x,r)$ is measurable for each r and $\gamma_P(x) \equiv \lim_{r\to 0} S_P(x,r)$ is a subdensity.

CONDITION B. For each $P \in \mathcal{P}$, $E_P \log \frac{\mathcal{L}(x)}{f_P(x)} < \infty$ where $\mathcal{L}(x) \equiv \sup\{f_P(x) : P \in \mathcal{P}\}$.

CONDITION C. $\{x : \gamma_P(x) \neq f_{P_0}(x)\}$ has positive measure if $P \in \bar{\mathcal{P}} \setminus \mathcal{P}$.

Bahadur's Condition A is defined in the weak topology of $\mathcal{P}$. From now on, a compactification $\bar{\mathcal{P}}$ of $\mathcal{P}$ will be called a *Bahadur suitable compactification* if Condition A is satisfied for some topology on $\mathcal{P}$, not necessarily the weak topology on $\mathcal{P}$. It is readily seen that Condition A implies the measurablity of $\gamma_P(x)$ and consequently the measurablity of $\mathcal{L}(x) \equiv \lim_{r\to\infty} S_P(x,r)$. Furthermore, since $\gamma_P(x) \geq f_P(x)$ where $\gamma_P(x)$ is a subdensity and $f_P(x)$ is a density for P in $\mathcal{P}$, it follows that $\gamma_P(x) = f_P(x)$ almost everywhere, i.e.

$$\lim_{r\to 0} S_P(x,r) = f_P(x) \quad \text{a.e.} \quad \text{for all} \quad P \text{ in } \mathcal{P}. \tag{3.0}$$

(3.0) gives upper semi-continuity of the model. As was mentioned earlier, some continuity condition is necessary for general consistency results. Thus, Bahadur's Condition A seems reasonable for general consistency of the AMPS estimators as well. Condition C is an identification condition which is rather weak. Bahadur's Condition B is a local dominance condition whose removal without compensation can lead to inconsistency of AMLE (see Kiefer and Wolfowitz (1956) and Bahadur (1971)).

Bahadur's Condition B is a typical regularity condition for general consistency of the AMLE in parametric models. Wang (1985) pointed out that the classical regularity conditions for consistency of the AMLE, (e.g. Wald (1949), Le Cam (1953), Kiefer and Wolfowitz (1956), Huber (1967), Bahadur (1971), and Perlman (1972)) share the common assumption that the log likelihood ratio of a distribution to the true one is locally dominated (or semidominated) by zero. More specifically, let $f_{P_0}(x)$ be the true density and let P^* be any probability measure in the parameter space. Local dominance (see Perlman (1972) for more precise definitions) requires the existence of a neighborhood $\mathcal{V}$ of P^* such that $\log[f_P(x)/f_{P_0}(x)]$ is dominated for P in $\mathcal{V}$. Additionally, Wang (1985) noted that in many nonparametric families the MLE is consistent but the local dominance assumption is violated. One example is the family of distributions with decreasing density on $[0,\infty)$. Using an alternative approach to the local dominance conditions, Wang (1985) (see also Pfanzagl (1988)) obtained theorems about consistency, which are shown to be applicable to several nonparametric families including the concave distribution functions.

Now we show that consistency of the AMPS estimator does not require local dominance conditions. In fact, boundedness of $S_P(x,r)$ on large sets suffices.

THEOREM 3.1. *Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from $P_0 \in \mathcal{P}$ where $\mathcal{P}$ is a family of probability measures on $\boldsymbol{R}$ dominated by λ. Suppose there is a version of the densities $\{f_P(x) : P \in \mathcal{P}\}$ such that $(\mathcal{P}, d)$ has a "Bahadur suitable compactification" $(\bar{\mathcal{P}}, d)$ which satisfies Bahadur's Condition C. Also assume for all P in $\bar{\mathcal{P}}$,*

$$\lim_{M\to\infty} \lim_{r\downarrow 0} P_0(\overline{A_{P,M,r}}) = 0 \tag{3.1}$$

where $\overline{A_{P,M,r}}$ denotes the closure of $A_{P,M,r} = \{x : S_P(x,r) \geq M\}$. Then any AMPS sequence is consistent, i.e. if $\{\hat{P}_n\}$ is an AMPS estimator, then $\lim_{n\to\infty} d(\hat{P}_n, P_0) = 0$.

PROOF. For any $P \in \mathcal{P}$ and $P \neq P_0$, let

$$R_n(F_P) \equiv \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F_P(X_{j,n}) - F_P(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})}.$$

Define $b(P,r) \equiv \{Q \in \bar{\mathcal{P}} : d(Q,P) < r\}$. We want to show that maximizing $R_n(F_P)$ over P in $\mathcal{P}$ yields an element of $b(P_0, \delta)$ for any small positive δ when n

is big enough. By Lemma 2.2, almost surely

$$\overline{\lim_{n\to\infty}} R_n(F_P) = \overline{\lim_{n\to\infty}} \frac{1}{n+1}\sum_{j=1}^{n+1} \log \frac{F_P(X_{j,n}) - F_P(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})}$$
$$\leq \int_R \log \frac{f_P(x)}{f_{P_0}(x)} dF_{P_0}(x) = \int_R \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

For Q in $\bar{\mathcal{P}}\backslash\mathcal{P}$, define $R_n(F_Q) \equiv \lim_{r\to 0}\sup_{P\in b(Q,r)\cap\mathcal{P}} R_n(F_P)$. Then maximizing $R_n(F_P)$ over P in $\mathcal{P}$ is equivalent to maximizing $R_n(F_Q)$ over Q in the compact set $\bar{\mathcal{P}}$.

Fix any $P \neq P_0$, $P \in \bar{\mathcal{P}}$. Without loss of generality one can assume that

$$-\infty < \int_R \log \frac{\gamma_P(x)}{f_{P_0}(x)}\, dF_{P_0}(x) < 0.$$

Then for each $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$, such that $P_0(A) < \delta(\varepsilon)$ implies

$$\left|\int_A \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x)\right| < \varepsilon.$$

Since P_0 is tight, there exists a compact interval K such that $P_0(K^c) < \frac{1}{2}\delta(\varepsilon)$. By assumption (3.1), for every $\delta(\varepsilon)$, $\exists M > 0$, $\exists r_M > 0$ such that $P_0(\overline{A_{P,M,r_M}}) < \frac{\delta(\varepsilon)}{2}$ where $A_{P,M,r_M} \equiv \{x : S_P(x, r_M) \geq M\}$. Let $K_M = \overline{A_{P,M,r_M}} \cap K$. K_M is a compact set and $P_0(K_M) < \frac{\delta(\varepsilon)}{2}$. Hence there are finitely many intervals $I_1, \ldots, I_m$ which cover K_M and $P_0(\cup_{j=1}^m I_j) < \frac{\delta(\varepsilon)}{2}$. Let $I_M = (\cup_{j=1}^m I_j) \cup K^c$. Then I_M is a union of finitely many intervals and $P_0(I_M) < \delta(\varepsilon)$. So

$$\left|\int_{I_M} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x)\right| < \varepsilon. \tag{3.2}$$

Define

$$f_{P,M}(x) \equiv \begin{cases} S_P(x, r_M), & x \in I_M^c \\ f_{P_0}(x), & x \in I_M \end{cases}.$$

Note that $f_{P,M}$ is integrable. Let

$$R_n(F_{P,M}) \equiv \frac{1}{n+1}\sum_{j=1}^{n+1} \log \frac{F_{P,M}(X_{j,n}) - F_{P,M}(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})},$$

where $F_{P,M}(x) \equiv \int_{-\infty}^x f_{P,M}(t)\lambda(dt)$ and $F_{P,M}(X_{n+1,n}) \equiv F_{P,M}(+\infty)$. By Lemma 2.2,

$$\overline{\lim_{n\to\infty}} R_n(F_{P,M}) \leq \int_R \log \frac{f_{P,M}(x)}{f_{P_0}(x)} dF_{P_0}(x) = \int_{I_M^c} \log \frac{S_P(x, r_M)}{f_{P_0}(x)} dF_{P_0}(x). \tag{3.3}$$

Let $n_M \equiv \{j : [X_{j-1,n}, X_{j,n}] \cap I_M \neq \emptyset\}$. Then

$$\begin{aligned}\sup_{Q\in b(P,r_M)} R_n(F_Q) &\equiv \sup_{Q\in b(P,r_M)\cap\mathcal{P}} \frac{1}{n+1}\sum_{j=1}^{n+1} \log \frac{F_Q(X_{j,n}) - F_Q(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})} \\ &= \sup_{Q\in b(P,r_M)\cap\mathcal{P}} \frac{1}{n+1}\left(\sum_{j\in n_M} + \sum_{j\notin n_M}\right) \log \frac{F_Q(X_{j,n}) - F_Q(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})}.\end{aligned}$$

Since $\sum_{j\in n_M}(F_Q(X_{j,n}) - F_Q(X_{j-1,n})) \leq 1$, $\forall Q \in b(P, r_M) \cap \mathcal{P}$,

$$\begin{aligned}(3.4)\qquad &\sup_{Q\in b(P,r_M)\cap\mathcal{P}} \frac{1}{n+1}\sum_{j\in n_M} \log \frac{F_Q(X_{j,n}) - F_Q(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})} \\ &\leq \frac{1}{n+1}\sum_{j\in n_M} \log\left\{\frac{1}{|n_M|}(F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n}))^{-1}\right\} \\ &= \frac{1}{n+1}\sum_{j\in n_M} \log\frac{n+1}{|n_M|} \\ &\quad - \frac{|n_M|}{n+1}\frac{1}{|n_M|}\sum_{j\in n_M} \log\{(F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n}))(n+1)\} \\ &\to -P_0(I_M)\log P_0(I_M) + P_0(I_M)\gamma \quad \text{as} \quad n\to\infty, \quad \text{a.s.,} \\ &\qquad \text{(by SLLN and Theorem 1.2 of Shao and Hahn (1995))}\end{aligned}$$

where $\gamma = 0.577\cdots$ is Euler's constant. Moreover,

$$\begin{aligned}&\sup_{Q\in b(P,r_M)\cap\mathcal{P}} \frac{1}{n+1}\sum_{j\notin n_M} \log \frac{F_Q(X_{j,n}) - F_Q(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})} \\ &\leq \frac{1}{n+1}\sum_{j\notin n_M} \log\left\{\frac{\int_{X_{j-1,n}}^{X_{j,n}} S_P(x, r_M)\lambda(dx)}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})}\right\}.\end{aligned}$$

Consequently,

$$\begin{aligned}(3.5)\qquad &\overline{\lim_{n\to\infty}} \sup_{Q\in b(P,r_M)\cap\mathcal{P}} \frac{1}{n+1}\sum_{j\notin n_M} \log \frac{F_Q(X_{j,n}) - F_Q(X_{j-1,n})}{F_{P_0}(X_{j,n}) - F_{P_0}(X_{j-1,n})} \\ &\leq \overline{\lim_{n\to\infty}} R_n(F_{P,M}).\end{aligned}$$

By (3.3), (3.4), and (3.5),

$$\begin{aligned}\overline{\lim_{n\to\infty}} \sup_{Q\in b(P,r_M)} R_n(F_Q) \leq &\int_{I_M^c} \log\frac{S_P(x, r_M)}{f_{P_0}(x)} dF_{P_0}(x) \\ &- P_0(I_M)\log P_0(I_M) + P_0(I_M)\gamma.\end{aligned}$$

Now

$$\begin{aligned}\int_{I_M^c} \log \frac{S_P(x,r_M)}{f_{P_0}(x)} dF_{P_0}(x) &= \int_{\boldsymbol{R}} \left[I_{I_M^c}(x) \log \frac{S_P(x,r_M)}{f_{P_0}(x)} + I_{I_M}(x) \log \frac{\gamma_P(x)}{f_{P_0}(x)} \right] dF_{P_0}(x) \\ &\quad - \int_{\boldsymbol{R}} I_{I_M}(x) \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x).\end{aligned}$$

By (3.2),

$$\left| \int_{\boldsymbol{R}} I_{I_M}(x) \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x) \right| < \varepsilon.$$

Furthermore, since

$$I_{I_M^c}(x) \log \frac{S_P(x,r_M)}{f_{P_0}(x)} + I_{I_M}(x) \log \frac{\gamma_P(x)}{f_{P_0}(x)} \downarrow \log \frac{\gamma_P(x)}{f_{P_0}(x)} \qquad P_0\text{-a.e.} \quad \text{when } r_M \to 0,$$

the monotone convergence theorem implies that

$$\int_{\boldsymbol{R}} \left[I_{I_M^c}(x) \log \frac{S_P(x,r_M)}{f_{P_0}(x)} + I_{I_M}(x) \log \frac{\gamma_P(x)}{f_{P_0}(x)} \right] dF_0(x) \downarrow \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_0(x).$$

Hence, when M is large enough and r_M is small enough with $\varepsilon \leq -\frac{1}{4} \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_0(x)$,

$$\begin{aligned}\int_{I_M^c} \log \frac{S_P(x,r_M)}{f_{P_0}(x)} dF_0(x) &< \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_0(x) + \varepsilon \\ &\leq \frac{3}{4} \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_0(x)\end{aligned}$$

and

$$- P_0(I_M) \log P_0(I_M) + P_0(I_M)\gamma < \varepsilon \leq -\frac{1}{4} \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_0(x).$$

Hence, when M is large enough and r_M is small enough

$$\overline{\lim_{n\to\infty}} \sup_{Q \in b(P,r_M)} R_n(F_Q) < \frac{1}{2} \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

So far we have been considering a fixed P. Generally, for any $P \in \bar{\mathcal{P}} \setminus \{P_0\}$, there exist $r_{M,P} > 0$ and $N_P \in \boldsymbol{N}$ such that whenever $n \geq N_P$,

$$\sup_{Q \in b(P,r_{M,P})} R_n(F_Q) < \frac{1}{2} \int_{\boldsymbol{R}} \log \frac{\gamma_P(x)}{f_{P_0}(x)} dF_{P_0}(x) < 0.$$

For each $\delta > 0$, all the neighborhoods $\{b(P, r_{M,P}) : P \in \bar{\mathcal{P}} \setminus \{P_0\}\}$ cover $\bar{\mathcal{P}} \setminus b(P_0, \delta)$. Since $\bar{\mathcal{P}} \setminus b(P_0, \delta)$ is compact, there exist finitely many neighborhoods, say

$\{b(P_1, r_{M,P_1}), \ldots, b(P_k, r_{M,P_k})\}$ which cover $\bar{\mathcal{P}} \setminus b(P_0, \delta)$. Let $N_\delta = \max(N_{P_1}, \ldots, N_{P_k})$. Then whenever $n \geq N_\delta$,

$$\sup_{Q \in \bar{\mathcal{P}} \setminus b(P_0,\delta)} R_n(F_Q)$$
$$< \frac{1}{2} \max\left(\int_R \log \frac{\gamma_{P_1}(x)}{f_{P_0}(x)} dF_{P_0}(x), \ldots, \int_R \log \frac{\gamma_{P_k}(x)}{f_{P_0}(x)} dF_{P_0}(x) \right) < 0.$$

Hence, maximizing $R_n(F_Q)$ over $Q \in \mathcal{P}$ yields an element of $b(P_0, \delta)$ for each $\delta > 0$. Thus, consistency of the AMPS estimator is substantiated. □

Remark 3.1. It should be noted that the "Bahadur suitable compactification" is a condition not only on the set $\mathcal{P}$ but also on the particular version f_P of $\frac{dP}{d\lambda}$ which is in force for each P in $\mathcal{P}$. Consistency of the AMLE depends on which versions of the densities are chosen. The consistency result for the AMPS estimator in Theorem 3.1 does not depend on which versions of the densities are chosen.

Theorem 3.1 has been formulated for families dominated by Lebesgue measure λ. λ can be replaced by any σ-finite measure with no atoms. Basically, Theorem 3.1 asserts consistency of the AMPS estimator under Bahadur's Conditions A, C and (3.1). Condition (3.1) replaces Bahadur's Condition B. Assumption (3.1) is used here to provide a simple, concise proof in the spirit of Wald (1949). Furthermore, it is usually straightforward to check. For example, if the densities are "reasonably" nice, e.g. have versions which are upper semicontinuous, then (3.1) is implied by Bahadur's Conditions A and C. More specifically,

COROLLARY 3.1. *Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from $P_0 \in \mathcal{P}$. Suppose there are upper semicontinuous versions of the densities $\{f_P(x) : P \in \mathcal{P}\}$ such that $(\mathcal{P}, d)$ has a "Bahadur suitable compactification" $(\bar{\mathcal{P}}, d)$ for which Bahadur's Condition C is satisfied. Then any AMPS sequence is consistent, i.e. if $\{\hat{P}_n\}$ is an AMPS, then $\lim_{n\to\infty} d(\hat{P}_n, P_0) = 0$. In particular, if $\{f_P(x) : P \in \mathcal{P}\}$ is a family of density functions and $\mathcal{P}$ has a compactification $\bar{\mathcal{P}}$ such that the subdensities $\{f_P(x), P \in \bar{\mathcal{P}}\}$ have an upper semicontinuous version both for P and for x, then any AMPS sequence is consistent.*

PROOF. It is easy to see that $S_P(x, r)$ is an upper semicontinuous function of x since each of the functions in $\{f_P(x) : P \in \mathcal{P}\}$ is. So for each $M > 0$, $A_{P,M,r} = \{x : S_P(x, r) \geq M\}$ is a closed set. Hence $\overline{A_{P,M,r}} = A_{P,M,r}$. So for any positive M,

$$\lim_{r\to 0} P_0(\overline{A_{P,M,r}}) = \lim_{r\to 0} P_0(A_{P,M,r}) = P_0(\{x : \gamma_P(x) \geq M\}).$$

Thus

$$\lim_{M\to\infty} \lim_{r\to 0} P_0(\overline{A_{P,M,r}}) = \lim_{M\to\infty} P_0(\{x : \gamma_P(x) \geq M\}) = 0$$

and consequently assumption (3.1) of Theorem 3.1 holds. Moreover, since $f_P(x)$ is upper semicontinuous for P, i.e. (3.0) holds, and $\gamma_P(x) = f_P(x)$ a.e. for any P

in $\bar{\mathcal{P}}$, we have a Bahadur suitable compactification for which Bahadur's Condition C is satisfied. □

Example 3.1. Le Cam ((1990), Section 4) provides a modification of Bahadur's well-known counterexample for the MLE, in which the MLE exists but goes further and further away from the true parameter value. For that example, it is easy to see that $\bar{\mathcal{P}} = \mathcal{P} \cup \{P_\infty\}$ where $P_\infty(-\infty, x) = (1-c)\delta_0(x) + cI_{(0,1]}(x)$ for $x \in [0,1]$ and $\gamma_{P_\infty}(x) = cI_{[0,1]}(x)$. Moreover, for all $P \in \mathcal{P}$ and $\varepsilon > 0$, there exists $r_P > 0$ such that $\{Q : d(Q,P) < r_P\} = \{P\}$. Thus, when M is big enough, $A_{P,M,r_P} = \emptyset$ and the set $\{x : \sup\{f_Q(x) : d(Q, P_\infty) < r\} \geq M\}$ is an interval with length going to zero. Hence, in Theorem 3.1, $\lim_{r\to 0} P_0(\overline{A_{P,M,r}}) = 0$ is satisfied. Consequently, the MPS estimator is consistent. However, notice that $\sup_{Q\in\mathcal{P}}\{f_Q(x) : d(Q, P_\infty) < r\}$ is not integrable for any $r > 0$, so Bahadur's Condition B fails and the regularity conditions in Wang (1985) do not hold either.

Example 3.2. Ferguson (1982) constructs an example with the following properties:

(1) The parameter space Θ is the compact set $[0,1]$;

(2) The observations are i.i.d. according to a distribution $F_\theta(x)$ for some $\theta \in \Theta$;

(3) Densities $f_\theta(x)$ with respect to Lebesgue measure exist and are continuous in θ for all $x \in [-1,1]$;

(4) (Identifiability) If $\theta' \neq \theta$, then $F_{\theta'}(x)$ is not identical to $F_\theta(x)$.

More specifically, the densities are of the form

$$f_\theta(x) = \frac{\theta}{2} + \left[\frac{1-\theta}{\delta(\theta)}\right]\left[1 - \frac{|x-\theta|}{\delta(\theta)}\right]^+, \qquad x \in [-1,1],\ \theta \in [0,1]$$

where $\delta(\theta)$ is a continuous decreasing function of θ with $\delta(0) = 1$ and $0 < \delta(\theta) \leq 1-\theta$ for $0 < \theta < 1$. For $\theta = 1$, $f_1(x) = \frac{1}{2}I_{[-1,1]}(x)$ which is the density of the uniform.

The existence of the MLE is clear since a continuous function is maximized on a compact set. However, Ferguson proves that if $\delta(\theta)$ tends to zero rapidly enough as $\theta \to 1$, then the MLE converges almost surely to 1 regardless of the true parameter $\theta_0 \in [0,1]$. On the other hand, note that for fixed θ, $f_\theta(x)$ is upper semicontinuous with respect to x and continuous with respect to θ for fixed x. Consequently, Corollary 3.1 shows that any AMPS sequence is consistent. In particular, the MPS estimator always exists and is consistent for this example. □

Remark 3.2. Bahadur ((1971), Theorem 9.2.) gives an existence theorem for consistent MLE based on the following five conditions:

CONDITION A. $\bar{\mathcal{P}}$ is a "suitable compactification" of $\mathcal{P}$.

CONDITION B. For each $P \in \mathcal{P}$, $E_P \log \frac{\mathcal{L}(x)}{f_P(x)} < +\infty$, where $\mathcal{L}(x) \equiv \sup\{f_Q(x) : Q \in \mathcal{P}\}$.

CONDITION C. $\{x : \gamma_P(x) \neq f_{P_0}(x)\}$ has positive measure if $P \in \bar{\mathcal{P}} \setminus \mathcal{P}$.

CONDITION D. $\bar{\mathcal{P}} - \mathcal{P}$ is a closed set.
CONDITION E. $\gamma_P(x) = f_P(x)$ for each $P \in \mathcal{P}$ and for all x in X.

Note that Conditions A–C are required to ensure consistency of AMLE. The additional conditions D and E assert that the MLE exists. Condition E is stronger than "$\gamma_P(x) = f_P(x)$ almost everywhere," which is implied by Condition A. Since the MPS estimator does not depend on the version of the density, Condition E is not necessary for existence of the MPS estimator. Next we prove that adding Condition D to the above consistency theorems for the AMPS estimator ensures existence of the MPS estimator.

THEOREM 3.2. *If the conditions in Theorem* 3.1 *and Bahadur's Condition D are satisfied, then with probability one an MPS estimator exists for all sufficiently large n, and any MPS estimator is consistent.*

PROOF. Given any small neighborhood of the true distribution, Theorem 3.1 asserts that all the AMPS estimators will be in the neighborhood for sufficiently large n. By considering the $\frac{\epsilon}{2}$-neighborhood we can assume the closure $\overline{A_\epsilon}$ of the ϵ-neighborhood A_ϵ is in $\mathcal{P}$ because of Condition D. For a given sample with size n (sufficiently large), let $\mathcal{A}_n$ be the set of all the AMPS estimates corresponding to that sample. Then $\mathcal{A}_n$ is not empty and $\mathcal{A}_n \subseteq \overline{A_\epsilon}$. If $\mathcal{A}_n$ is a finite set, then at least one of its elements is a MPS estimator. If $\mathcal{A}_n$ is not a finite set, there are $\{F_k : k \in \boldsymbol{N}\}$ in $\mathcal{A}_n$ such that $\lim_{k\to\infty} S_n(F_k) = \sup_{P_F \in \bar{\mathcal{P}}} S_n(F)$. By considering subsequences if necessary, we can assume $\lim_{k\to\infty} P_{F_k} = P_\infty \in \overline{A_\epsilon} \subset \mathcal{P}$. By (3.0), $f_{P_\infty}(x) = \lim_{r\to 0} S_{P_\infty}(x,r)$. So

$$\overline{\lim_{k\to\infty}}\, f_k(x) \le \lim_{r\to 0} S_{P_\infty}(x,r) = f_{P_\infty}(x).$$

By the dominated convergence theorem,

$$S_n(F_\infty) \ge \lim_{k\to\infty} S_n(F_k) = \sup_{P_F \in \bar{\mathcal{P}}} S_n(F).$$

Thus, P_∞ is a MPS estimator for the sample with size n. □

Remark 3.3. Under the usual conditions to prove consistency of AMLE (e.g. Wald (1949), Le Cam (1953)), consistency of parameters in the Euclidean metric on the parameter space is equivalent to consistency of the probability measures in the total variation distance (see Landers and Rogge (1972)). More specifically, under the conditions of Wald (1949), if θ_0 is the true parameter, the estimates $\theta_n \to \theta_0$ in the Euclidean distance if and only if $P_{\theta_n} \to P_{\theta_0}$ in total variation distance. So, under general conditions, usual consistency in the parametric situation can be deduced from consistency of probability measures in total variation norm.

Consistency depends on the topology. A sequence of estimates can be consistent in one topology but inconsistent in some stronger topology. The weak(-star) topology can be a natural choice for the space of probability measures in the absence of additional information concerning the structure of the family of measures.

For example, if the true distribution is known to be in a finite set, then consistency of the MPS estimator can be obtained in a stronger metric, e.g. the total variation distance. Also, if there is information about the geometrical shape of the distribution (e.g. family of increasing densities), then consistency of the MPS estimator may be obtained in a stronger metric than the weak metric. In these situations the weak metric can be essentially equivalent to stronger metrics such as the total variation norm, as can be seen in the next section.

4. Universal consistency for unimodal distributions

Distributions with unimodal densities are the most commonly used continuous distributions in practice. We start by looking at a special case of unimodal distributions, those having monotone densities. Notice that the arguments in this section work for parametric models as well as for nonparametric models.

Example 4.1. (Families of nondecreasing densities) Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from $P_0 \in \mathcal{P}$, where $\mathcal{P}$ is a family of probability measures with densities which are nondecreasing on $(a, b]$, where $-\infty \leq a < b < \infty$. Since a nondecreasing function is continuous almost everywhere, there is no loss of generality in assuming the densities are right continuous. Let $\mathcal{F}$ be the family of nondecreasing and right continuous subdensities on $(a, b]$ and define a Levy-type distance ρ on $\mathcal{F}$ by

$$\rho(g, h) \equiv \inf\{\varepsilon > 0 : g((x-\varepsilon) \vee a+) - \varepsilon \leq h(x) \leq g((x+\varepsilon) \wedge b) + \varepsilon, \text{ for all } x \in \boldsymbol{R}\}.$$

Then $(\mathcal{F}, \rho)$ is a compact metric space (see Reiss (1973)) in which $\lim_{n\to\infty} \rho(f_n, f) = 0$ if and only if $\lim_{n\to\infty} f_n(x) = f(x)$ at all continuity points of f. The closure $\bar{\mathcal{P}}$ of $\mathcal{P}$ with respect to $\rho(P, Q) \equiv \rho(f_P, f_Q)$ is thus also compact. Note that the function $S_P(x, r) = \sup\{f_Q(x) : \rho(Q, P) \leq r, Q \in \mathcal{P}\}$ is nondecreasing for x and is upper semicontinuous for x and for P. Consequently, Corollary 3.1 and Scheffé's Theorem imply that any AMPS estimator is consistent with respect to L^1 or the total variation norm. □

Similarly, for families with nonincreasing densities on $[a, b)$, $-\infty < a < b \leq \infty$, every AMPS sequence is consistent with respect to the total variation distance. But the same is not true for the MLE. Example 4.7 of Reiss (1973) gives a nonincreasing density family in which the MLE exists but is not consistent.

DEFINITION 4.1. A function $f : \boldsymbol{R} \to [0, \infty]$ is called *unimodal* if f is nondecreasing on $(-\infty, M_f)$ and nonincreasing on (M_f, ∞) for some number M_f, which is called a *mode* of f. The set of all modes of f is called the *modal interval* I_f of f, and the center of the modal interval I_f is denoted by μ_f.

THEOREM 4.1. *Any AMPS sequence is consistent for a unimodal density in* L^1 *distance.*

PROOF. Since AMPS estimates are independent of the versions of the densities and since each unimodal density is continuous almost everywhere, one can assume the unimodal density to be right continuous on $(-\infty, M_f)$ and left continuous on (M_f, ∞), i.e. upper semicontinuous. Without loss of generality we assume $f(M_f) = \infty$. Define

$$\mathcal{F}_0 \equiv \left\{ f : f \text{is unimodal, upper semicontinuous, } \int_{\boldsymbol{R}} f d\lambda \leq 1, f(0) = \infty \right\}.$$

For each $f \in \mathcal{F}_0$ define

$$f^*(x) = \begin{cases} f(x) & x < 0 \\ f(0) & x \geq 0 \end{cases} \quad \text{and} \quad f^{**}(x) = \begin{cases} f(-x) & x < 0 \\ f(0) & x \geq 0 \end{cases}.$$

Let $\mathcal{F}_0^* = \{f^* \mid f \in \mathcal{F}_0\}$, $\mathcal{F}_0^{**} = \{f^{**} \mid f \in \mathcal{F}_0\}$, and define ρ as in Example 4.1. Then $(\mathcal{F}_0^*, \rho)$ and $(\mathcal{F}_0^{**}, \rho)$ are both compact metric spaces. For $f, g \in \mathcal{F}_0$, let

$$d_0(f, g) \equiv \max\{\rho(f^*, g^*), \rho(f^{**}, g^{**})\}.$$

By Theorem 2.6 of Reiss (1973), $(\mathcal{F}_0, d_0)$ is a compact metric space.

Note that for any unimodal density $f(x), f(x + M_f)$ has zero as a mode. Let

$$\mathcal{F} \equiv \left\{ f : f \text{ is unimodal, upper semicontinuous, } \int_{\boldsymbol{R}} f d\lambda \leq 1, f(M_f) = \infty \right\}.$$

For f and g in $\mathcal{F}$, define $d(f, g) \equiv \max\{d_0(f(\cdot + M_f), g(\cdot + M_g)), |M_f - M_g|\}$. Since M_f and M_g are always finite, d is a well-defined metric on $\mathcal{F}$. In fact,

LEMMA 4.1. *$(\mathcal{F}, d)$ is a locally compact metric space and a subset $\mathcal{K}$ of $\mathcal{F}$ is relatively compact if and only if $\{M_f \mid f \in \mathcal{K}\}$ is bounded. Furthermore, for f_n and f in $\mathcal{F}$, $n \in \boldsymbol{N}$, the following two assertions are equivalent:*

(i) $\lim_{n\to\infty} d(f_n, f) = 0$;

(ii) $\lim_{n\to\infty} |M_{f_n} - M_f| = 0$ *and* $\lim_{n\to\infty} f_n(x) = f(x)$ *at each continuity point* $x \neq M_f$ *of* f.

PROOF. See Theorems 2.10, 2.11 and 2.12 of Reiss (1973). □

$\bar{\mathcal{P}}$ with respect to the vague topology in Theorem 3.1 generally may not be compact with respect to the total variation distance. However, it suffices to prove that the AMPS estimates live on a compact subset when sample sizes are large.

LEMMA 4.2. *Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from $F_0 \in \mathcal{F}$ where $\mathcal{F}$ is a family of probability distributions with unimodal densities. If $\{M_f\}$ is unbounded, then for each $D > 0$ there exist $C > 0$ and $N_C > 0$ such that whenever $n \geq N_C$,*

$$\sup_{|M_f| > C} R_n(F) = \sup_{|M_f| > C} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F(X_{j,n}) - F(X_{j-1,n})}{F_0(X_{j,n}) - F_0(X_{j-1,n})} \leq -D < 0.$$

PROOF. Without loss of generality, it may be assumed that F_0 has zero as a mode. Since F_0 is tight, $\forall \varepsilon > 0$, $\exists A > 0$, such that

$$\delta_0(A) \equiv \int_{-\infty}^{-A} dF_0(x) + \int_A^{\infty} dF_0(x) < \varepsilon.$$

Let B be any (big) positive integer and let $C \equiv (2B+1)A$. When $M_f > C$, $f(A) \leq \frac{1}{2BA}$, so

$$\int_{-A}^{A} f(x)dx \leq \frac{1}{B}.$$

Using the nondecreasing order statistics of the sample, let $J_C \equiv \{j : |X_{j,n}| > A\}$. Then

$$\begin{aligned}\sup_{M_f>C} R_n(F) &= \sup_{M_f>C} \frac{1}{n+1} \sum_{j=1}^{n+1} \log \frac{F(X_{j,n}) - F(X_{j-1,n})}{F_0(X_{j,n}) - F_0(X_{j-1,n})} \\ &= \sup_{M_f>C} \frac{1}{n+1} \left(\sum_{j\in J_C} + \sum_{j\notin J_C} \right) \log \frac{F(X_{j,n}) - F(X_{j-1,n})}{F_0(X_{j,n}) - F_0(X_{j-1,n})}.\end{aligned}$$

$$\begin{aligned}&\sup_{M_f>C} \frac{1}{n+1} \sum_{j\in J_C} \log \frac{F(X_{j,n}) - F(X_{j-1,n})}{F_0(X_{j,n}) - F_0(X_{j-1,n})} \\ &\leq \frac{1}{n+1} \sum_{j\in J_C} \log \left\{ \frac{1}{|J_C|} (F_0(X_{j,n}) - F_0(X_{j-1,n}))^{-1} \right\} \\ &= \frac{|J_C|}{n+1} \log \frac{n+1}{|J_C|} - \frac{1}{n+1} \sum_{j\in J_C} \log\{(n+1)(F_0(X_{j,n}) - F_0(X_{j-1,n}))\} \\ &\to -\delta_0(A)(\log \delta_0(A) - \gamma) \quad \text{when} \quad n \to \infty, \text{ since } \lim_{n\to\infty} \frac{|J_C|}{n+1} = \delta_0(A).\end{aligned}$$

$$\begin{aligned}&\sup_{M_f>C} \frac{1}{n+1} \sum_{j\notin J_C} \log \frac{F(X_{j,n}) - F(X_{j-1,n})}{F_0(X_{j,n}) - F_0(X_{j-1,n})} \\ &\leq \sup_{M_f>C} \frac{1}{n+1} \sum_{j\notin J_C} \log \left\{ \frac{B^{-1}}{(n+1-|J_C|)} (F_0(X_{j,n}) - F_0(X_{j-1,n}))^{-1} \right\} \\ &= \frac{n+1-|J_C|}{n+1} \log \frac{n+1}{n+1-|J_C|} \cdot B^{-1} \\ &\quad - \frac{1}{n+1} \sum_{j\notin J_C} \log\{(n+1)(F_0(X_{j,n}) - F_0(X_{j-1,n}))\} \\ &\to -(1-\delta_0(A))[\log[(1-\delta_0(A))B] - \gamma].\end{aligned}$$

The same argument works for $M_f < -C$, so

$$\begin{aligned}\overline{\lim_{n\to\infty}} \sup_{|M_f|>C} R_n(F) \leq &-2(1-\delta_0(A))(\log[(1-\delta_0(A))B] - \gamma) \\ &- 2\delta_0(A)(\log \delta_0(A) - \gamma) + 1\end{aligned}$$

where $\delta_0(A) < \varepsilon$ and ε is an arbitrary small positive number. Since $\varepsilon \log \varepsilon \to 0$ when $\varepsilon \downarrow 0$ and $\lim_{B\to\infty,\varepsilon\to 0} \log[(1-\delta_0(A))B] = \infty$, the right-hand side of the inequality goes to $-\infty$ when $\varepsilon \downarrow 0$ and $B \uparrow \infty$. This proves the lemma. □

COMPLETION OF THE PROOF OF THEOREM 4.1. Lemmas 4.1 and 4.2 justify consideration of AMPS estimates only on a compact subset of $\bar{\mathcal{P}}$ when $\bar{\mathcal{P}}$ is not compact with respect to d. By Lemma 4.1, the conditions of Corollary 3.1 hold. Thus, for any positive ε, there is a positive integer N so that when $n \geq N$, the AMPS estimator falls in an ε-neighborhood (in metric d) of the true density provided the true density has a unique mode M_f. However, the spacings do not depend on versions of the density. If the true density has a modal interval I_f with positive length $|I_f|$, i.e., the true density has infinitely many unimodal versions, then for any positive ε and n large enough, the AMPS estimator falls in the union of the open ε-neighborhoods in the metric d of all the versions of the true density. The union is a big open set in the metric d with diameter no less than $|I_f|$. However, the union is an ε-neighborhood of the true density in the metric ρ. Consequently, the AMPS sequence converges to the true density and hence also in L^1 distance by Theorem 2.17 of Reiss (1973). Moreover, the modes M_{f_n} of this AMPS sequence satisfy $\lim_{n\to\infty} \inf\{|M_{f_n} - m| : m \in I_f\} = 0$ a.s. □

Remark 4.1. Families of unimodal densities include many of the families of distributions in practice. Many counterexamples to consistency of the MLE and AMLE involve unimodal densities, such as the three-parameter log normal distributions, the three-parameter Weibull distribution, three-parameter Gamma distributions, those discussed in Smith (1985) and in Examples 3.1 and 3.2, and some distributions discussed in Le Cam (1990). In particular, the regularity conditions in Bahadur (1971) and Wang (1985) fail to hold for these situations. By Theorem 4.1, they cannot be counterexamples to consistency of the AMPS estimates. The proof of Theorem 4.1 can also be modified to establish consistency of AMPS estimates for families of monotone failure rates, or families of U-shaped densities and families of multimodal densities in which the modes are fixed.

Remark 4.2. Motivated by problems arising from reliability theory and survival analysis, much research effort has been devoted to the nonparametric estimation of a cdf with a unimodal density. Grenander (1956) first obtained the nonparametric MLE for a distribution with a monotone decreasing density on $[0, \infty)$, which is the least concave majorant of the empirical distribution function. Kiefer and Wolfowitz (1976) proved that the Grenander estimator is asymptotically efficient in the sense of minimax. Thorough asymptotic and nonasymptotic analyses for the Grenander estimator have been done by Groeneboom (1985) and Birgé (1989), respectively. Prakasa Rao (1969) considered the nonparametric estimation of a unimodal distribution with known mode. However, the likelihood function can be unbounded when the mode of the underlying distribution is unknown. (Recall that a monotone density is regarded as a special unimodal density.) Thus the nonparametric maximum likelihood method breaks down when the mode of the distribution is unknown. No other good estimates of a distribution function with a unimodal density (when the mode may not be known) have been proposed except the empirical distribution function, which is not necessarily unimodal. Shao (1996) obtained explicitly the maximum product of spacings estimator for a distribution with a unimodal density. The MPS estimator is always unimodal and has a simple explicit representation. Moreover, when the mode is known, the MPS estimator is asymptotically efficient in the sense of minimax. When the mode is unknown, Theorem 4.1 asserts the uniform consistency of the MPS estimator.

Acknowledgements

The authors are indebted to Professors Ronald Pyke, Kathryn M. Roeder, Richard L. Smith, Jane-Ling Wang and Jon A. Wellner for helpful discussions. We also thank the referees for their careful reading and for valuable suggestions that have improved the paper.

References

Bahadur, R. R. (1971). Some limit theorems in statistics, *Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia.

Basu, D. (1955). An inconsistency of the method of maximum likelihood, *Ann. Math. Statist.*, **26**, 144–145.

Birgé, R. (1989). The Grenander estimator: a nonasymptotic approach, *Ann. Statist.*, **17**, 1532–1549.

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 Stephens, M. A. (1989). A goodness-of-fit test using Moran's statistic and estimated parameters, *Biometrika*, **76**, 385–392.

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

Chung, K. L. (1974). *A Course in Probability Theory*, Academic Press, New York.

Cox, D. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Cressie, N. (1976). On the logarithms of high order spacings, *Biometrika*, **63**, 343–355.

Darling, D. A. (1953). On a class of problems related to random division of an interval, *Ann. Math. Statist.*, **24**, 239–253.

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

Gebert, J. R. and Kale, B. K. (1969). Goodness of fit tests based on discriminatory information, *Statist. Hefte*, **3**, 192–200.

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

Groeneboom, P. (1985). Estimating a monotone density, *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II (eds. L. Le Cam and R. A. Olshen), 539–555, Wadsworth, Monterey, California.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of Fifth Berkeley Symposium*, **1**, 221–233.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Statist.*, **27**, 887–906.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions, *Z. Wahrsch. Verw. Gebiete*, **36**, 73–85.

Landers, D. and Rogge, L. (1972). Characterization of the topologies used in the theory of maximum likelihood estimation, *Z. Wahrsch. Verw. Gebiete*, **21**, 197–200.

Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, *University of California Publications in Statistics*, **1**, 277–329.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer, New York.

Le Cam, L. (1990). Maximum likelihood: an introduction, *International Statistics Review*, **59**, 153–171.

Lehmann, E. L. (1991). *Theory of Point Estimation*, Wadsworth and Brooks/Cole, Pacific Grove, California.

Lind, N. C. (1994). Information theory and maximum product of spacings estimation, *J. Roy. Statist. Soc. Ser. B*, **56**, 341–343.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS regional conference series in probability and statistics, Vol. 5, IMS, Philadelphia.

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

Pfanzagl, J. (1988). Consistency of maximum likelihood estimation for certain nonparametric families, in particular: mixtures, *J. Statist. Plann. Inference*, **19**, 137–158.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density, *Sankhyā Ser. A*, **31**, 32–36.

Pyke, R. (1965). Spacings, *J. Roy. Statist. Soc. Ser. 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.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

Roeder, K. (1990). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies, *J. Amer. Statist. Assoc.*, **85**, 617–624.

Roeder, K. (1992). Semiparametric estimation of normal mixture densities, *Ann. Statist.*, **20**, 929–943.

Shao, Y. (1996). On nonparametric estimation of a distribution function using the maximum product of spacings method (preprint).

Shao, Y. and Hahn, M. G. (1994). Maximum spacing estimates: a generalization and improvement of maximum likelihood estimates I, *Probability in Banach Spaces*, *9*, Progress in Probability, Vol. 35 (eds. J. Hoffmann-Jorgensen, J. Kuelbs and M. B. Marcus), 417–431, Birkhauser, Boston.

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. (1996). On a distribution-free test of fit for continuous distribution functions, *Scand. J. Statist.*, **23**, 63–73.

Shorack, G. R. (1972). The best test of exponentiality against Gamma alternatives, *J. Amer. Statist. Assoc.*, **67**, 213–214.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases, *Biometrika*, **72**, 67–92.

Titterington, D. M. (1985). Comment on "estimating parameters in continuous univariate distribution", *J. Roy. Statist. Soc. Ser. B*, **47**, 115–116.

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. (1970). Maximum likelihood estimation of a unimodal density function, *Ann. Math. Statist.*, **41**, 457–471.