ILLINOIS JOURNAL OF MATHEMATICS
Volume 43, Number 3, 1999

MAXIMUM PRODUCT OF SPACINGS METHOD: A UNIFIED FORMULATION WITH ILLUSTRATION OF STRONG CONSISTENCY

YONGZHAO SHAO AND MARJORIE G. HAHN

ABSTRACT. A simple unified formulation for generating an estimator of a univariate distribution (not necessarily continuous) is proposed. It is a variant of the maximum likelihood method and also a generalization of the *maximum product of spacings* (*MPS*) *method* for estimating continuous univariate distributions. The general formulation is then applied to monotone hazard rate families which arise frequently in applications. It is shown that any asymptotic MPS estimator for any family of distributions with monotone hazard rate is always strongly consistent (whether the setting is parametric or not). The MPS estimator of a distribution function with a monotone hazard rate can be derived explicitly and is asymptotically minimax for the Kolmogorov-Smirnov type loss functions.

1. Introduction

A typical statistical problem is to make inferences about a distribution F_{θ_0} based on some random sample $\{X_1, X_2, \dots, X_n\}$, which can be regarded as independent real-valued random variables having common, but unknown, distribution function F_{θ_0}. The common practice is to assume that F_{θ_0} is in a certain family of distributions, say in $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$. To make inferences about F_{θ_0}, one can apply the *probability integral transform* (PIT) to the data, i.e., for each $F_\theta \in \mathcal{F}$,

$$F_\theta : \{X_1, X_2, \dots, X_n\} \to \{F_\theta(X_1), F_\theta(X_2), \dots, F_\theta(X_n)\}. \tag{1.0}$$

If F_{θ_0} is continuous, it is well known that $\{F_{\theta_0}(X_1), F_{\theta_0}(X_2), \dots, F_{\theta_0}(X_n)\}$ are uniform random variables. When F_{θ_0} is known to be continuous, Cheng and Amin (1983) and independently Ranneby (1984) proposed the *maximum product of spacings method*, which takes the estimator of F_{θ_0} to be the $F_{\hat{\theta}_n}$ which maximizes the product of spacings, i.e.,

$$\prod_{i=0}^{n}[F_{\hat{\theta}_n}(X_{i+1,n}) - F_{\hat{\theta}_n}(X_{i,n})] = \sup_{F_\theta \in \mathcal{F}} \prod_{i=0}^{n}[F_\theta(X_{i+1,n}) - F_\theta(X_{i,n})]. \tag{1.1}$$

When F_{θ_0} is not known to be continuous, there might be ties in the data. Suppose that in the n observations $\{X_i, 1 \le i \le n\}$, there are m distinct values $\{Y_{j,m}, 1 \le j \le m\}$. Let $Y_{0,n} \equiv -\infty$. Let l_j denote the number of observations in $(Y_{j-1,n}, Y_{j,n}]$, i.e.,

---

Received April 22, 1998.
1991 Mathematics Subject Classification. Primary 62F10, 62G05; Secondary 62F12, 62G20.
The authors were partially supported by grants from the National Science Foundation.

exactly l_j of the observations $\{X_i, 1 \le i \le n\}$ are equal to $Y_{j,n}$. Since the observations $\{X_i, 1 \le i \le n\}$ are i.i.d., each of the l_j tied observations has the same probability of occurrence. The probability is about $\frac{F_\theta(Y_{j,n})-F_\theta(Y_{j-1,n})}{l_j}$ according to the model F_θ. Define $m_\theta \equiv m$ if $F_\theta(Y_{m,n}) = 1$; and $m_\theta \equiv m+1$ otherwise. Then $F_\theta(Y_{m_\theta,n}) \equiv 1$. In this case, it is natural to maximize the following product of spacings:

$$\sup_{F_\theta \in \mathcal{F}} \prod_{j=1}^{m_\theta} \left(\frac{F_\theta(Y_{j,n}) - F_\theta(Y_{j-1,n})}{l_j} \right)^{l_j}; \tag{1.2}$$

i.e., one can take the distribution $F_{\hat{\theta}_n}$ which maximizes the product of spacings in (1.2) as the estimator of F_{θ_0}. We will call this estimation technique the general maximum product of spacings (MPS) method and an estimator so obtained will be called an MPS estimator.

The MPS method does not require the distributions to have densities. The simple unified formulation of maximizing (1.2) is well defined for any univariate distribution. It generalizes the MPS method defined in (1.1) for continuous distributions and it is asymptotically equivalent to the maximum likelihood (ML) method under very general conditions. Cheng and Amin (1983) and Ranneby (1984) showed that the MPS method can produce consistent estimators in the three-parameter lognormal and Weibull families as well as normal mixture models, while it is known that the ML method breaks down for such models due to the unboundedness of the likelihood function. Unboundedness difficulties do not arise in the MPS method since the product of spacings is always bounded. The general asymptotic behavior of the MPS method has attracted the attention of many researchers. Among the results, Shao and Hahn (1997) establishes that in any unimodal distribution family the asymptotic MPS estimator of the underlying unimodal density is L^1 consistent universally without any further conditions (in parametric or nonparametric settings). In contrast, many counterexamples exist for consistency of the maximum likelihood estimators (MLEs) in unimodal families. Moreover, it is well known that a necessary regularity condition for the consistency of the ML method is the local dominance condition in Perlman (1972) (see also Le Cam (1953) and Wang (1985)). The general consistency theorems for the MPS method obtained in Shao (1997) require much weaker conditions than those Wald-type conditions for the MLE. In particular, the local dominance type conditions are not necessary for consistency of the MPS estimators.

The remainder of the paper is organized as follows: Section 2 contains some general remarks. Section 3 provides proof of consistency for families of finitely many distributions. Section 4 deals with estimation of distributions with monotone hazard rate which are widely used in applications. Any asymptotic MPS estimator for any family of distributions with monotone hazard rate is shown to be always consistent (parametric or nonparametric). The MPS estimator of a cumulative distribution function with a monotone hazard rate is derived explicitly and is asymptotically minimax for the Kolmogorov-Smirnov type loss.

2. Remarks on the unified formulation of the MPS method

The unified formulation in Section 1 can be reformulated more precisely as follows: Let $X_1, \ldots, X_n$ be an i.i.d. sample from F_{θ_0} in $\mathcal{F}$, where $\mathcal{F} = \{F_\theta\colon \theta \in \Theta\}$ is any family of distribution functions defined on the real line $\mathbf{R}$. Let $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ denote an ordering of the sample. Define $Y_{0,n} \equiv -\infty$. For $1 < k$, define

$$Y_{k,n} \equiv \min\left\{X_{j,n}\colon X_{j,n} > Y_{k-1,n}\right\} \quad \text{and} \quad m \equiv \min\{j\colon Y_{j,n} = X_{n,n}\}. \tag{2.1}$$

Define $m_\theta \equiv m$ if $F_\theta(Y_{m,n}) = 1$; and $m_\theta \equiv m+1$ otherwise. Then $F_\theta(Y_{m_\theta,n}) \equiv 1$. Then for any $F_\theta \in \mathcal{F}$, $0 \equiv F_\theta(Y_{0,n}) \le F_\theta(Y_{1,n}) \le F_\theta(Y_{2,n}) \le \cdots \le F_\theta(Y_{m_\theta,n}) \equiv 1$. For $j = 1, \ldots, m_\theta$, let $l_j = \#\{i\colon X_i = Y_{j,n},\ 1 \le i \le n\}$. Define

$$\mathcal{P}_n(F_\theta, X) \equiv \prod_{j=1}^{m_\theta} \left(\frac{F_\theta(Y_{j,n}) - F_\theta(Y_{j-1,n})}{l_j}\right)^{l_j}. \tag{2.2}$$

The *maximum product of spacings (MPS) estimator* of F_{θ_0} is defined to be $F_{\hat\theta_n}$ where $\hat\theta_n = \hat\theta_n(X_1, \ldots, X_n)$ such that

$$\mathcal{P}_n(F_{\hat\theta_n}, X) = \sup_{F_\theta \in \mathcal{F}} \mathcal{P}_n(F_\theta, X). \tag{2.3}$$

We also say that the MPS estimator of the parameter θ_0 is $\hat\theta_n$ when (2.3) holds. A sequence $F_{\hat\theta_n}$ will be called an *asymptotic MPS (AMPS) estimator* of F_{θ_0}, if

$$\varliminf_{n\to\infty} \frac{\mathcal{P}_n(F_{\hat\theta_n}, X)}{\mathcal{P}_n(F_{\theta_0}, X)} \ge C \quad \text{for some } C > 0. \tag{2.4}$$

Remark 2.0. Notice that Ranneby (1984) and Shao and Hahn (1994) called the MPS method the "maximum spacing method", and its estimates the "MSP estimate" or "MSE" respectively. On the other hand, Cheng and Amin (1983) used the name "maximum product of spacings (MPS) method", which is adopted here.

Remark 2.1. Throughout the remainder of the paper we make the basic assumption *that the family $\mathcal{F} = \{F_\theta\colon \theta \in \Theta\}$ of distributions under consideration has the property that each sample is contained in the support of F_θ for each $\theta \in \Theta$.* If not, in the definition given above, (2.2) should be multiplied on the right-hand side by $\prod_{i=1}^n I_{\mathrm{supp}(F_\theta)}(X_i)$ since if $\{X_1, \ldots, X_n\} \not\subset \mathrm{supp}(F_\theta)$, F_θ is not qualified to be a candidate for the estimator of the true distribution. Here we assume $x \in \mathrm{supp}(F_\theta)$ if and only if $F_\theta(x+\varepsilon) > F_\theta(x-)$ for every positive ε. Since none of the ideas change, this assumption simplifies our expressions.

Remark 2.2. Clearly

$$\sum_{j=1}^{m_\theta} \left(\frac{F_\theta(Y_{j,n}) - F_\theta(Y_{j-1,n})}{l_j}\right) \cdot l_j \equiv 1$$

and $n \leq l_1 + \ldots + l_{m_\theta} \leq n + 1$. So

$$\mathcal{P}_n(F_\theta, X) = \prod_{j=1}^{m_\theta} \left(\frac{F_\theta(Y_{j,n}) - F_\theta(Y_{j-1,n})}{l_j} \right)^{l_j} \leq \frac{1}{n^n}.$$

If $\mathcal{F}$ contains the empirical distribution function $\mathbf{F}_n$, then it is easy to see that $\mathcal{P}_n(\mathbf{F}_n, X) = \frac{1}{n^n}$. Thus, the empirical distribution is an MPS estimator for the underlying true distribution function if $\mathcal{F}$ contains all the univariate distribution functions.

Remark 2.3 (Continuous distributions). When the distribution functions are all continuous there are no ties in the data (with probability 1). So $F_\theta(Y_{i,n}) = F_\theta(X_{i,n})$, $l_i = 1$ for $i = 1, \ldots, n < m_\theta = n + 1$, $F_\theta(Y_{0,n}) \equiv 0$, and $F_\theta(Y_{n+1,n}) \equiv 1$. Thus,

$$\mathcal{P}_n(F_\theta, X) = \prod_{j=1}^{n+1} [F_\theta(X_{j,n}) - F_\theta(X_{j-1,n})]. \tag{2.5}$$

Hence the MPS estimator is defined as $F_{\hat{\theta}_n}$ such that

$$F_{\hat{\theta}_n} = \arg \sup_{F_\theta \in \mathcal{F}} \sum_{j=1}^{n+1} \log[F_\theta(X_{j,n}) - F_\theta(X_{j-1,n})].$$

This is the formulation (1.1) proposed by Cheng and Amin (1983) and Ranneby (1984).

Theoretically ties in the data can happen with probability 0, however, in practice, this may happen more often than expected due to rounding errors. If there are ties in the data, (2.2) seems the most natural modification of (2.5) to handle ties in the data coming from continuous distributions. (2.2) will give results asymptotically equivalent to those of using (2.5) when there is no tie in the data. For this reason and for simplicity of exposition, we will assume that there is no tie in the data if the underlying distribution is continuous.

In particular, if $\mathcal{F}$ consists of all the continuous distribution functions, then it is easy to see that the "Fisher predictive distribution", i.e., $F_{\hat{\theta}_n}$ in $\mathcal{F}$ satisfying $F_{\hat{\theta}_n}(X_{i,n}) = i/(n+1)$ for $1 \leq i \leq n$, is an MPS estimator. A special case is the *Pyke's modified empirical* $F_{\hat{\theta}_n}$ which is uniformly distributed on each spacing and satisfies $F_{\hat{\theta}_n}(X_{i,n}) = i/(n+1)$. More specifically, *Pyke's modified empirical* $F_{\hat{\theta}_n}$ on a finite interval $(a, b]$ is defined as follows:

$$F_{\hat{\theta}_n}(x) = \frac{1}{n+1} \sum_{i=1}^{n+1} \left(i - 1 + \frac{x - X_{i-1,n}}{X_{i,n} - X_{i-1,n}} \right) I(X_{i-1,n} < x \leq X_{i,n}) \tag{2.6}$$

where $X_{0,n} = a$, $X_{n+1,n} = b$ and $I(\cdot)$ is the indicator function. Note that any "Fisher predictive distribution" $F_{\hat{\theta}_n}$ is strongly consistent by the Glivenko-Cantelli theorem and the fact that

$$\sup_x |F_{\hat{\theta}_n}(x) - \mathbf{F}_n(x)| = \frac{1}{n+1}, \tag{2.7}$$

where $\mathbf{F}_n(x)$ is the empirical distribution function.

The formulation of the MPS method does not require the distributions to have densities. The following example might give some flavor for this method.

Example 2.1 (Estimation of median). Let $X_1, X_2, \cdots, X_n$ be i.i.d. from F_0 in $\mathcal{F}$, the class of all *continuous* univariate probability distributions, each with a unique median. Since the MPS estimator is a "Fisher predictive distribution" which satisfies

$$F_{\hat{\theta}_n}(X_{i,n}) = \frac{i}{n+1}, \quad 1 \leq i \leq n,$$

it is strongly consistent, i.e., $\sup_{x\in\mathbf{R}} |F_{\hat{\theta}_n}(x) - F_0(x)| \to 0$ as $n \to +\infty$. Furthermore, suppose we want to estimate the *median* $\mathcal{M}_0 = \mathcal{M}(F_0) \equiv \inf\{x\colon F_0(x) \geq \frac{1}{2}\}$. Regard the median as a functional $\mathcal{M}$ on $\mathcal{F}$, i.e.,

$$\mathcal{M}(F_\theta) \equiv \inf\left\{x\colon F_\theta(x) \geq \frac{1}{2}\right\}, \quad \forall F_\theta \in \mathcal{F}. \tag{2.8}$$

Then the median $\mathcal{M}$ is a continuous functional, in the sense that $\forall\epsilon > 0$, $\exists\delta > 0$, such that

$$\sup_{t\in\mathbf{R}} |F_\theta(t) - F(t)| < \delta \quad \Rightarrow \quad |\mathcal{M}(F_\theta) - \mathcal{M}(F)| \leq \epsilon \tag{2.9}$$

(see Pollard (1984), p. 7). Since the MPS estimator is obtained by maximizing a function, it has the invariance property when applied to functionals. Consequently, the MPS estimator $\mathcal{M}(F_{\hat{\theta}_n})$ for $\mathcal{M}_0$ is strongly consistent, i.e., $\lim_{n\to\infty} \mathcal{M}(F_{\hat{\theta}_n}) = \mathcal{M}_0$. In fact, $\mathcal{M}(F_{\hat{\theta}_n}) = X_{\frac{n+1}{2},n}$ if n is odd; $\mathcal{M}(F_{\hat{\theta}_n}) \in [X_{\frac{n}{2},n}, X_{\frac{n}{2}+1,n}]$ if n is even. For comparison, notice that it is not clear how to estimate $\mathcal{M}_0$ in the formulation of the classical MLE since $\mathcal{F}$ is not a dominated family.

Remark 2.4 (Discrete distributions). When the distributions are all purely discrete with finitely many cell probabilities $(\pi_1, \ldots, \pi_k)$, i.e., multinomial, the MPS maximizes the quantity $\prod_{i=1}^k \left(\frac{\pi_i}{n_i}\right)^{n_i}$ over $\pi = (\pi_1, \ldots, \pi_k)$. This is equivalent to the classical MLE which maximizes the quantity $\pi_1^{n_1} \cdots \pi_k^{n_k}$. Consequently, the MPS method can also be viewed as a variant of the MLE in this sense.

3. Estimation for families of finitely many distributions

Since the definition of the generalized MPS method in Section 2 is simple, does not depend on density versions, and does not involve taking limits, it is potentially applicable to any univariate family of distributions. Thus, it is instructive to see how strong consistency of the MPS estimator might be proven. This can be easily illustrated for a finite family of distributions each with at most finitely many atoms.

LEMMA 3.1. *If $\mathcal{F}$ is a finite family of distributions, each of which has at most finitely many atoms, then maximizing $\mathcal{P}_n(P_\theta, X)$ yields a strongly consistent MPS estimator.*

Proof. Any distribution function F can be written as a unique convex combination of a discrete and a continuous one, $F = \alpha F_d + (1-\alpha)F_c$ with $\alpha = \sum_j P(X = z_j)$, where the summation is over all atoms z_j of the distribution F. Let Θ be a finite set and let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample with common unknown distribution in $\mathcal{F} = \{F_\theta\colon \theta \in \Theta,\ F_\theta = F_{\theta d} + F_{\theta c},\ F_{\theta d} \text{ multinomial}\}$. Let F_{θ_0} be the true distribution function with jumps at $Z = \{z_1, \ldots, z_k\}$. For simplicity of notation, assume that $F_{\theta c}(x)$ has subdensity $f_\theta(x)$ and let $\alpha = \sum_{j=1}^k (F_{\theta_0}(z_j) - F_{\theta_0}(z_j-))$. Then, when n is large, for any $F_\theta \neq F_{\theta_0}$,

$$\log\left[\frac{\mathcal{P}_n(P_\theta, X)}{\mathcal{P}_n(P_{\theta_0}, X)}\right] \approx \sum_{j=1}^k \ell_j \log \frac{F_\theta(z_j) - F_\theta(z_j-)}{F_{\theta_0}(z_j) - F_{\theta_0}(z_j-)} + \sum_{X_{i,n} \notin Z} \log \frac{F_\theta(X_{i,n}) - F_\theta(X_{i-1,n})}{F_{\theta_0}(X_{i,n}) - F_{\theta_0}(X_{i-1,n})}.$$

By the strong law of large numbers (SLLN) and the Information-type Inequality in Shao and Hahn (1995),

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log\left[\frac{\mathcal{P}_n(P_\theta, X)}{\mathcal{P}_n(P_{\theta_0}, X)}\right] \le \sum_{j=1}^k (F_{\theta_0}(z_j) - F_{\theta_0}(z_j-)) \log \frac{F_\theta(z_j) - F_\theta(z_j-)}{F_{\theta_0}(z_j) - F_{\theta_0}(z_j-)} + (1-\alpha)\int_{\mathbf{R}} \left(\log \frac{f_\theta(x)}{f_{\theta_0}(x)}\right) \frac{f_{\theta_0}(x)}{\int_{\mathbf{R}} f_{\theta_0}(x)dx} dx.$$

By Jensen's Inequality,

$$\begin{aligned}\overline{\lim_{n\to\infty}} \frac{1}{n} \log\left[\frac{\mathcal{P}_n(P_\theta, X)}{\mathcal{P}_n(P_{\theta_0}, X)}\right] &\le \alpha \log\left(\frac{\sum_{j=1}^k (F_\theta(z_j) - F_\theta(z_j-))}{\alpha}\right) + (1-\alpha)\log \frac{\int_{\mathbf{R}} f_\theta(x)dx}{\int_{\mathbf{R}} f_{\theta_0}(x)dx} \\ &\le \log \sum_{j=1}^k (F_\theta(z_j) - F_\theta(z_j-)) + \log \int_{\mathbf{R}} f_\theta(x)dx < 0.\end{aligned}$$

Since the family is finite, after finitely many steps, the MPS estimator becomes the true value F_{θ_0}. The assumption that $F_{\theta c}$ has a subdensity is not essential. Generalization can be made as in the proof of the Information-type Inequality in Shao and Hahn (1995). □

Remark 3.1. General strong consistency theorems for $\mathcal{F}$ not finite can be formulated by imitating the strong consistency theorems in Shao and Hahn (1997) using the ideas in the proof of Lemma 3.1. It is relatively straightforward, thus will not be performed here.

4. Estimation for distributions with monotone hazard rates

Families of distributions with *monotone hazard rate* are of considerable importance in reliability theory, survival analysis, and other applications. For a distribution function $F(x)$ with density $f(x)$ on $\mathbf{R}$, the *hazard rate* $q(x)$ (also called *failure rate*) is defined as

$$q(x) = \frac{f(x)}{1 - F(x)} \quad \text{for } F(x) < 1. \tag{4.1}$$

The importance of the monotone hazard (or failure) rate property and more references can be found in Barlow et al (1972) and Robertson et al (1988). The discussion below focuses solely on the family of distributions with decreasing failure rate since the other situations can be handled using the same ideas.

Definition 4.1. A distribution function F is said to have a *decreasing failure rate* (DFR) if $R(x) = \log[1 - F(x)]$ is convex on its support $[\beta, \infty)$, where $\beta > -\infty$.

If $F(x)$ has DFR and $F(\beta) > 0$, then F is absolutely continuous on $(\beta, +\infty)$. It is easy to check that the derivative of the absolutely continuous part must be decreasing on (β, ∞), although it may have a jump at β. Thus DFR is, in a sense, a generalization of the decreasing density assumption.

Remark 4.2. In a nonparametric setting, Grenander (1956) applies the maximum likelihood method to the distributions with decreasing densities very elegantly and obtains the MLE as the smallest majorant of the empirical distribution function. Similarly, explicit solutions for the MPS estimator have been obtained in Shao (1997). For example, the MLE for the family of all distributions with non-increasing densities on [0, 1] is the smallest concave majorant of the empirical distribution; the MPS estimator for the same family is the smallest concave majorant of Pyke's modified empirical distribution. Monotone density can be regarded as a special unimodal density. When the mode of the monotone density is unknown, the MLE is not proper because of unbounded likelihood, but the explicit MPS estimator still exists and is asymptotically minimax for the Kolmogorov-Smirnov type loss functions by the arguments of Kiefer and Wolfowitz (1976). The strong consistency theorem given below for the asymptotic MPS works for any family with DFR (parametric or not).

THEOREM 4.1. *Any asymptotic MPS estimator for any distribution with DFR is strongly consistent, whether the setting is parametric or not.*

Proof. For convenience, we assume that β is known. If the true underlying distribution has no jump at β, then the problem reduces to a family of decreasing densities. Then by Theorem 4.1 of Shao and Hahn (1997), any asymptotic MPS sequence is strongly consistent. Thus, assume the true underlying distribution has a jump at β. It can be assumed that

$$\beta = X_{1,n} = \cdots = X_{k,n} < X_{k+1,n} < \cdots < X_{n,n}$$

is an ordered sample from the true distribution F_{θ_0}. When maximizing

$$\mathcal{P}_n(F_\theta, X) = \prod_{j=1}^{m_\theta} \left(\frac{F_\theta(Y_{j,n}) - F_\theta(Y_{j-1,n})}{l_j} \right)^{l_j},$$

only those distributions having a jump at β play a role (since those without a jump at β will have $\mathcal{P}_n(F_\theta, X) \equiv 0$ for n large enough). By the strong law of large numbers and the proof of Lemma 3.1,

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \frac{1}{n}(\log \mathcal{P}_n(F_\theta, X) - \log \mathcal{P}_n(F_{\theta_0}, X)) \\
&\leq F_{\theta_0}(\beta) \log \frac{F_\theta(\beta)}{F_{\theta_0}(\beta)} + (1 - F_{\theta_0}(\beta)) \log \int_\beta^\infty \frac{f_\theta(x)}{1 - F_{\theta_0}(\beta)} \mathrm{dx} \\
&< 0 \quad \text{(by Jensen's Inequality)} .
\end{aligned}$$

Hence, when n is large enough, $\mathcal{P}_n(F_\theta, X) < \mathcal{P}_n(F_{\theta_0}, X)$. Define a distance on $\mathcal{F}$ as follows:

$$d(F_{\theta_1}, F_{\theta_2}) = |F_{\theta_1}(\beta) - F_{\theta_2}(\beta)| \vee \ell(F_{\theta_1}, F_{\theta_2})$$

where $\ell(F_{\theta_1}, F_{\theta_2})$ is the Lévy type distance between the derivatives of F_{θ_1} and F_{θ_2}, i.e.,

$$\ell(F_{\theta_1}, F_{\theta_2}) \equiv \inf\{\varepsilon > 0\colon f_{\theta_1}((x-\varepsilon)\vee\beta+)-\varepsilon \leq f_{\theta_2}(x) \leq f_{\theta_1}(x+\varepsilon)+\varepsilon,\ x \in (\beta, \infty)\}.$$

It is easy to show that $(\mathcal{F}, d)$ is a compact metric space. By the above arguments, for any $F_\theta \in \mathcal{F}$ which is different from F_{θ_0}, there exists some positive ε_θ such that

$$F_{\theta_0}(\beta) \log \frac{F_\theta(\beta) + \varepsilon_\theta}{F_{\theta_0}(\beta)} + (1 - F_{\theta_0}(\beta)) \log \int_\beta^\infty \frac{f_\theta(x) + \varepsilon_\theta}{1 - F_{\theta_0}(\beta)} \mathrm{dx} < 0.$$

So there exists some positive integer N_θ such that whenever $n > N_\theta$,

$$\mathcal{P}_n(G_\theta, X) < \mathcal{P}_n(F_{\theta_0}, X), \quad \forall G_\theta \in b_d(F_\theta, \varepsilon_\theta) \equiv \{G_\theta\colon d(G_\theta, F_\theta) < \varepsilon_\theta\}.$$

For any positive δ, $\mathcal{F} - b_d(F_{\theta_0}, \delta)$ is compact and $\{b_d(F_\theta, \varepsilon_\theta)\colon F_\theta \in \mathcal{F} - b_d(F_{\theta_0}, \delta)\}$ is an open cover. Thus, there exist finitely many small open balls which cover $\mathcal{F} - b_d(F_{\theta_0}, \delta)$. Hence when n is bigger than N, the largest of those finitely many N_θ's, necessarily

$$\mathcal{P}_n(F_\theta, X) < \mathcal{P}_n(F_{\theta_0}, X), \quad \forall F_\theta \notin b_d(F_{\theta_0}, \delta).$$

Maximizing $\mathcal{P}_n(F_\theta, X)$ over $F_\theta \in \mathcal{F}$ yields some distribution in the neighborhood of the true distribution, which implies strong consistency.

When β is unknown, the proof is similar to the proof of the unimodal densities with unknown modes as in Shao and Hahn (1997), which we do not repeat here. □

Roeder (1990, 1992) adapted the MPS method for simultaneously estimating parameters and providing goodness-of-fit tests in some semiparametric mixture models. The generalized formulation of the MPS method in Section 2 treats the parametric, nonparametric and semiparametric problems in a unified way. For instance, Theorem 4.1 remains valid for the estimation of a distribution function in a family which can be parametric, nonparametric or semiparametric, as long as they have monotone failure rates. In particular, when nothing is known about the underlying distribution function except that it has a monotone failure rate, the problem is in a semiparametric setting if we view the possible jump height of the distribution at its mode as a one-dimensional parameter and the monotone subdensity as infinite dimensional. We also have the following:

THEOREM 4.2. *The MPS estimator of a distribution with a monotone decreasing (or increasing) failure rate is, explicitly, the least concave majorant (or greatest convex minorant) of Pyke's modified empirical distribution function and is asymptotically minimax for the Kolmogorov-Smirnov type loss functions.*

Proof. If the underlying distribution F_0 has a *decreasing failure rate* (DFR), then $R(x) = \log[1 - F_0(x)]$ is convex on its support $[\beta, \infty)$, where $\beta > -\infty$. If $F_0(\beta) = 0$, then F_0 has a decreasing density on $[\beta, \infty)$. Remark 2.3 and Shao (1997) imply that the MPS estimator for F_0 is the least concave majorant (LCM) of the Pyke's modified empirical distribution (as given in (2.6)), i.e. the smallest concave distribution function which is no less than the Pyke's modified empirical distribution. When $F_0(\beta) > 0$, denote the number of ties of the data at β by b_n, then the MPS estimator for $F_0(\beta)$ is $b_n/(n+1)$, and the MPS estimator for $F_0(x)$ is still the LCM of the Pyke's modified empirical distribution. The asymptotic minimaxity of the MPS estimator for the Kolmogorov-Smirnov type loss functions follows directly from Kiefer and Wolfowitz (1976) and Shao (1997). The situation that F_0 has an *increasing failure rate* can be handled in the same way. □

5. Concluding remarks

The simple formulation proposed in Section 2, although widely applicable, certainly should not be recommended blindly. The interest in general procedures is that they may provide reasonably good solutions, while better formulations and solutions can often be tailor-made for specific situations (since additional knowledge is available). We study the MPS estimator as a variant of the MLE with the hope that most of the asymptotic optimalities of the MLE in regular parametric cases (in the sense of Cramér) will follow and that the MPS method may work well when the ML method fails. The MPS method is intended for non-regular likelihood problems and certainly not an approach suggested to replace the maximum likelihood method in general. For other issues such as the calculation of confidence intervals and the handling of certain censored samples by the MPS method, the reader is referred to Cheng and Traylor (1995) and its ensuing discussions.

Acknowledgment. The authors are indebted to Professors Shaw-Hwa Lo and Jon A. Wellner for helpful discussions. The authors also would like to thank the referee for several valuable comments.

REFERENCES

1. R. E. Barlow, D. J. Bartholomew, and H. D. Brunk, *Statistical inference under order restrictions*. Wiley, New York, 1972.
2. R. C. H. Cheng and N. A. K. Amin, *Estimating parameters in continuous univariate distributions with a shifted origin*, J. Roy. Statist. Soc. B **45** (1983), 394–403.
3. R. C. H. Cheng and L. Traylor, *Non-regular maximum likelihood problems (with discussions)*, J. Roy. Statist. Soc. B **57** (1995), 3–44.
4. U. Grenander, *On the theory of mortality measurement, Part II*, Skan. Aktuarietidskr. **39** (1956), 125–153.
5. J. Kiefer and J. Wolfowitz, *Asymptotically minimax estimation of concave and convex distribution functions*, Z. Wahrsch. Verw. Gebiete. **36** (1976), 73–85
6. L. Le Cam, *On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates*, Univ. Calif. Public. in Statist. **1** (1953), 277–329.
7. M. D. Perlman, *On the strong consistency of approximate maximum likelihood estimators*, Proc. 6th Berkeley Symp. Math. Statist. Prob., University of California Press, 1972, pp. 221–233.
8. D. B. Pollard, *Convergence of stochastic processes*, Springer-Verlag, New York, 1984.
9. B. Ranneby, *The maximum spacing method. An estimation method related to the maximum likelihood method*, Scand. J. Statist. **11** (1984), 93–112.
10. T. Robertson, F. T. Wright, and R. L. Dykstra, *Order restricted statistical inference*, Wiley, New York, 1988.
11. K. Roeder, *Density estimation with confidence sets exemplified by superclusters and voids in the galaxies*, J. Amer. Statist. Assoc. **85** (1990), 617–624.
12. ———, *Semiparametric estimation of normal mixture densities*, Ann. Statist. **20** (1992), 929–943.
13. Y. Shao, *Consistency of the maximum product of spacings method and estimation of a unimodal distribution*, 1997, preprint.
14. Y. Shao and M. G. Hahn, "Maximum spacing estimates: A generalization and improvement of maximum likelihood estimates I" in *Probability in Banach spaces, Proc. 9th Internat. Conf. on Probability*

in Banach Spaces, Progress in Probability, vol. 35, Eds. J. Hoffmann-Jorgensen, J. Kuelbs and M. B. Marcus, Birkhauser, Boston, 1994, pp. 417–431.

15. ———, *Limit theorems for the logarithm of sample spacings*, Statist. Probab. Lett. **24** (1995), 121–132.
16. ———, *Strong consistency of the maximum product of spacings estimates with applications in nonparametrics and in estimation of unimodal densities*, Ann. Inst. Statist. Math., **51** (1) (1999).
17. A. Wald, *Note on the consistency of the maximum likelihood estimate*, Ann. Math. Stat. **20** (1949), 595–601.
18. J. L. Wang, *Strong consistency of approximate maximum likelihood estimators with applications to nonparametrics*, Ann. Statist. **13** (1985), 932–946.

Yongzhao Shao, Department of Statistics, Columbia University, New York, NY 10027
`yshao@stat.columbia.edu`

Marjorie G. Hahn, Department of Mathematics, Tufts University, Medford, MA 02155
`mhahn@emerald.tufts.edu`