Statistica Sinica **11**(2001), 1159-1174

# AN ALTERNATIVE UNIMODAL DENSITY ESTIMATOR WITH A CONSISTENT ESTIMATE OF THE MODE

Mary C. Meyer

*University of Georgia*

*Abstract:* The traditional maximum likelihood unimodal density estimator (Grenander (1956)) pieces together two isotonic density estimators at a known mode. It is discontinuous at the mode, and does not directly adapt to the case of unknown mode. This paper presents an alternative unimodal density estimator in the form of a generalized isotonic regression on a partial order which is continuous at the mode, and moves easily to the case of unknown mode. A penalized version is introduced to control the spiking at the mode, and is proved to be consistent everywhere. It is shown that the penalized estimator also provides a consistent estimate of the mode. Simulation results compare the penalized estimator to other nonparametric estimators in the literature in terms of Hellinger distance, the squared error loss of the estimate of the mode, and the height at the mode. Two important advantages of the new estimator are that it provides the density estimate and the mode estimate simultaneously, and that it is "fully automatic," that is, no pre-grouping or bounds on density height are necessary to prevent spiking.

*Key words and phrases:* Maximum likelihood, maximum penalized likelihood, nonparametric density estimation, unimodal density estimation.

## 1. Introduction

Let $x_1 < x_2 < \cdots < x_n$ be an ordered random sample from an unknown unimodal density $f$ with cumulative distribution function $F$ and known mode $\mu \in (x_{m-1}, x_m]$. The nonparametric maximum likelihood estimator (NMLE) maximizes the product $\prod f(x_i)$ restricted to the class of unimodal densities.

Grenander (1956) showed that the NMLE $\bar{f}$ is a step function with constant values on the intervals $[x_1, x_2), [x_2, x_3), \ldots, [x_{m-1}, \mu), (\mu, x_m], (x_m, x_{m+1}], \ldots,$ $(x_{n-1}, x_n]$, and zero otherwise. The value at $\mu$ can be taken to be the average of $\bar{f}(\mu-)$ and $\bar{f}(\mu+)$. Prakasa Rao (1969) showed that the NMLE is consistent for $x \neq \mu$ and derived its asymptotic distribution. Woodroofe and Sun (1993) showed that the estimate at the mode is inconsistent and presented a penalized MLE (the PMLE) which is consistent everywhere.

It is difficult to use this estimator (penalized or unpenalized) in the case of an unknown mode, because the likelihood is unbounded if $\mu$ is allowed to vary. Several ways to overcome this problem have been suggested. The modified likelihood approach puts the mode at an observation $x_m$ and ignores the value of the estimate at the mode, so that the expression to maximize is $\prod_{i\neq m} g(x_i)$, the modified likelihood. This also gives an estimate of the mode, as we can loop though the data to find $m$ to give the largest modified likelihood. The "plug-in" approach is to estimate the mode first, to get $M_n$ say, and to plug this value in for $\mu$ in the known mode solution. Wegman (1970) showed that if $M_n$ is consistent, then $\bar{f}_{n,\mu}$ and $\bar{f}_{n,M_n}$ have the same limit distribution at $x \neq \mu$. Wegman (1970) also suggested using a modal interval of fixed length $\epsilon$. This eliminates the spiking problem since the density can not be larger than $1/\epsilon$, but it is not consistent unless the underlying $f$ also has a modal interval of length at least $\epsilon$.

Bickel and Fan (1996) propose a linear spline MLE with pregrouping to control spiking. The pregrouping technique imposes a grid on the real line: the "modified empirical distribution function" has jumps only at the partition points, with jump heights corresponding to the numbers of observations falling into the partition interval. If the lengths of the partition intervals are $o(n^{-1/2})$, then the pregrouping MLE performs similarly to the plug-in MLE.

In the next section we show that an alternative partial ordering forces the step function to have constant value on $(x_{m-1}, x_m]$, which solves the problem of unknown mode and provides for an estimate of the mode. A term is added to the likelihood to penalize spiking, producing the penalized alternative unimodal density estimator. Local and global consistency results are provided. The estimator of the mode is shown to be consistent in Section 3. In Section 4, the density and mode estimators are compared to existing estimators through simulations. The proofs of theorems and propositions that are longer than a few lines are relegated to Section 5.

## 2. Alternative Penalized Estimator

The NMLE of Grenander is a step function, with jumps at the data and at the known mode $\mu$. It is the result of separate monotone regressions on $(-\infty, \mu)$ and $(\mu, \infty)$ and thus is not continuous at $\mu$, nor does the method move easily to the case where $\mu$ is unknown. The alternative nonparametric MLE (ANMLE) proposed here is also a step function, but with steps only at the data, not at the mode. It is the result of a single isotonic regression on a partial order, and moves easily to the case where $\mu$ is unknown.

Assuming the (known) mode $\mu$ is in $(x_{m-1}, x_m]$, the alternative partial ordering has "upper sets" $\mathcal{U} = \{u_1, \ldots, m, \ldots u_2\}$ and "lower sets" $\mathcal{L} = \{2, \ldots, l_1\} \cup \{l_2, \ldots, n\}$ where $l_1 < m$ and $l_2 > m$; or $l_1 = l_2 = m$. The solution provided by generalized isotonic regression is $\tilde{f}(x) = \tilde{\theta}_j$, for $x \in (x_{j-1}, x_j]$. The values of the estimate are

$$\tilde{\theta}_j = \max_{\mathcal{U}:j\in\mathcal{U}} \min_{\mathcal{L}:j\in\mathcal{L}} \frac{(l_1 - u_1 + 1)_+/n + (u_2 - l_2 + 1)_+/n}{(x_{l_1} - x_{u_1-1})_+ + (x_{u_2} - x_{l_2-1})_+}, \tag{1}$$

where $(x)_+ = \max(0, x)$. Here $u_1$, $u_2$, $l_1$, and $l_2$ are understood to define upper and lower sets. For a treatment of generalized isotonic regression, see Robertson, Wright, and Dykstra (1988) (henceforth called RWD), Section 1.5. Note that both the numerator and denominator are always greater than 0, since both $\mathcal{L}$ and $\mathcal{U}$ must contain $j$. At the modal interval, we have

$$\tilde{\theta}_m = \max_{u_1\le m\le u_2} \frac{(u_2 - u_1 + 1)/n}{x_{u_2} - x_{u_1-1}}. \tag{2}$$

Woodroofe and Sun (1993) showed that the traditional NMLE for a nonincreasing density on $(\mu, \infty)$ is inconsistent at the mode and proposed a penalized estimator which they showed to be consistent everywhere. The ANMLE is also inconsistent at $\mu$. The result is similar to that of Woodroofe and Sun; the proof is more complicated by the fact that $m$ is a random variable (see Section 5).

**Theorem 1.** $\frac{\tilde{\theta}_m}{f(\mu)} \Rightarrow \Delta$, *as* $n \to \infty$ *where* $\Delta = \sup_{k>0, j<0, k>j} \frac{k-j}{Y_{j+1}+\cdots+Y_k}$, *and* $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ *are i.i.d. standard exponential random variables.*

Note that $P(\Delta > 1) = 1$ by the Strong Law of Large Numbers. To get the penalized alternative estimator (APMLE), we include a penalty term in the likelihood function to reduce the estimate at the mode, and introduce a Lagrange multiplier $\gamma$ to constrain the area to be unity. For penalty parameter $\alpha > 0$, we want to maximize:

$$\sum_{i=2}^{n} \log\theta_i - n\alpha\theta_m - n\gamma\sum_{i=2}^{n}(x_i - x_{i-1})\theta_i = \sum_{i=2}^{n}[\log\theta_i - nw_i\theta_i] \tag{3}$$

where $w_i = \gamma(x_i - x_{i-1})$ for $i \neq m$ and $w_m = \alpha + \gamma(x_m - x_{m-1})$.

The APMLE (denoted $\hat{f}(x) = \hat{\theta}_j$ for $x \in I_j$) is equivalent to the unpenalized estimator of a sample with the data moved slightly: if each $x_i$ is multiplied by $\gamma$, and the value $\alpha$ is added to the new $x_m$, the resulting unpenalized estimator is identical to the penalized estimator for the original data. At the mode, this is

$$\hat{\theta}_m(\gamma) = \max_{u_1\le m\le u_2} \frac{(u_2 - u_1 + 1)/n}{\alpha + \gamma(x_{u_2} - x_{u_1-1})}, \tag{4}$$

which is decreasing and convex in $\gamma$.

From Theorem 1.3.6 of RWD, we have $\sum_{i=2}^{n}\hat{\theta}_i w_i = 1$ or

$$\alpha\hat{\theta}_m(\gamma) + \gamma\sum_{i=2}^{n}(x_i - x_{i-1})\hat{\theta}_i(\gamma) = 1.$$

The area constraint is $\sum_{i=2}^{n}(x_i - x_{i-1})\hat{\theta}_i(\gamma) = 1$, so that we must solve the equation $\alpha\hat{\theta}_m(\gamma) + \gamma = 1$. Define $\phi_{u_1,u_2}(\gamma) = 1 - \frac{\alpha(u_2-u_1+1)/n}{\alpha+\gamma(x_{u_2}-x_{u_1-1})}$. Then we must solve the following for $\gamma$: $\gamma = \min_{u_1\le m\le u_2}\phi_{u_1,u_2}(\gamma)$.

Each $\phi_{u_1,u_2}$ is increasing and concave, with $\phi_{u_1,u_2}(1) < 1$ and $\phi_{u_1,u_2}(0) > 0$ for $u_2 - u_1 + 1 < n$, so each equation $\gamma = \phi_{u_1,u_2}(\gamma)$ has a unique solution. Therefore,

$$\hat{\gamma} = \min_{u_1\le m\le u_2}\Big\{\frac{1}{2}(1 - \frac{\alpha}{x_{u_2} - x_{u_1-1}}) + \Big[(\frac{\alpha}{2(x_{u_2} - x_{u_1-1})})^2 + \frac{\alpha}{2(x_{u_2} - x_{u_1-1})}$$
$$(1 - \frac{2(u_2 - u_1 + 1)}{n}) + \frac{1}{4}\Big]^{1/2}\Big\}$$

For convenience of notation, $\hat{\theta}$ without an argument is meant to be $\hat{\theta}(\hat{\gamma})$. For the following proof of consistency, we write $\hat{\theta}_{m,n}$ for the above $\hat{\theta}_m$ to emphasize the dependence on $n$. We require

$$0 < f(\mu) < \infty,\ 0 < \alpha = \alpha_n \searrow 0, \text{ and } n\alpha \nearrow \infty, \tag{5}$$

and that $f$ be continuous at the mode. Note that the index $m$ of the modal interval depends on $n$ and the data. We ought to write $m(n)$ or $m_n$ but that would make the notation too cumbersome. Under (5), we get consistency at the mode. The proof of the following is in Section 5.

**Proposition 1.** *For any* $0 < \gamma_0 < 1$, $p - \lim_{n\to\infty}\sup_{\gamma_0\le\gamma\le 1}|\gamma\hat{\theta}_{m,n}(\gamma) - f(\mu)| = 0$.

**Corollary 1.** $p - \lim_{n\to\infty}\frac{1-\hat{\gamma}_n}{\alpha} = f(\mu)$.

**Proof.** To apply the proposition, we need to show that there is a $\gamma_0 > 0$ such that with probability approaching 1, $\hat{\gamma}_n > \gamma_0$. If $\alpha\hat{\theta}_{m,n}(1/2) < 1/2$, then $\hat{\gamma}_n > 1/2$. Since $P\{\alpha\hat{\theta}_{m,n}(1/2) \ge 1/2\} \to 0$ as $n \to \infty$, $1-\hat{\gamma}_n = \alpha\hat{\theta}_{m,n}(\hat{\gamma}_n) \le \alpha\hat{\theta}_{m,n}(1/2) \to^p 0$ and therefore$(1 - \hat{\gamma}_n)/\alpha \to^p f(\mu)$.

**Corollary 2.** $p$–$\lim_{n\to\infty}\hat{\theta}_{m,n}(\hat{\gamma}_n)/f(\mu) = 1$.

**Proof.** We have $\hat{\theta}_{m,n}(\hat{\gamma}_n) = (1 - \hat{\gamma}_n)/\alpha$.

To show consistency for $x \neq \mu$, we first prove that the unpenalized estimator is consistent away from the mode, then that the penalized estimator is not very different from the unpenalized estimator for $x \neq \mu$. In the following, we write

$u_1(n)$ and $u_2(n)$ for the previous $u_1$ and $u_2$, to emphasize dependence on $n$. The first step is to prove that the length of the modal interval for the unpenalized estimator goes to zero (see Section 5 for proof).

**Proposition 2.** *Let $f$ be strictly unimodal with mode $\mu$, and let $\tilde{\theta}$ be defined as in* (1)*. Let $x_{u_1(n)-1}$ and $x_{u_2(n)}$ define the modal interval as in* (2)*. Then $x_{u_2(n)} \to^p \mu$ and $x_{u_1(n)-1} \to^p \mu$.*

Let $\tilde{F}_n$ be the distribution function for the unpenalized alternative unimodal estimator, and let $\tilde{f}_n(x) = \tilde{\theta}_{k,n}$ for $x \in (x_{k-1}, x_k]$. It is easy to see that $\tilde{F}_n$ is the least concave majorant of $F_n$ for $x \geq x_{u_2(n)}$ and the greatest convex minorant of $F_n$ for $x \leq x_{u_1(n)}$. The following lemma is analogous to "Marshall's Lemma" associated with the traditional estimator.

**Lemma 1.** $\sup_{x \geq x_{u_2(n)}} \mid \tilde{F}_n(x) - F(x) \mid \leq \sup_{x \geq x_{u_2(n)}} \mid F_n(x) - F(x) \mid$, $\sup_{x \leq x_{u_1(n)-1}} \mid \tilde{F}_n(x) - F(x) \mid \leq \sup_{x \leq x_{u_1(n)-1}} \mid F_n(x) - F(x) \mid$.

**Proof.** For the first statement, let $\epsilon_n = \sup_{x \geq x_{u_2(n)}} \mid F_n(x) - F(x) \mid$. Since $F(x) + \epsilon_n$ is concave and majorizes $F_n(x)$ for $x \geq x_{u_2(n)}$, we have $F(x) - \epsilon_n \leq F_n(x) \leq \tilde{F}_n(x) \leq F(x) + \epsilon_n$. The proof for the second statement is similar, so that the unpenalized distribution function is consistent for $x$ on either side of the modal interval. Next, we show that the unpenalized density estimator is consistent for $x \neq \mu$.

**Proposition 3.** *Let $f$ be strictly unimodal with mode $\mu$, and let $\tilde{\theta}$ be defined as in* (1)*. For every $x \neq \mu$, $\tilde{f}_n(x) \to f(x)$, in probability as $n \to \infty$. Further, $\mid \tilde{f}_n(x) - f(x) \mid = O_p(n^{-1/4})$.*

We obtain the global consistency of $\hat{f}_n$ by showing that $\hat{f}_n$ and $\tilde{f}_n$ do not differ substantially away from the mode. Let $h(f,g)$ be the Hellinger distance between the densities $f$ and $g$, that is,

$$h^2(f,g) = \int \left(\sqrt{f} - \sqrt{g}\right)^2 dx = 2 - 2\int \sqrt{fg}dx. \tag{6}$$

**Theorem 2.** $h^2(\tilde{f}_n, \hat{f}_n) \leq \alpha[\tilde{f}_n(\mu) - \hat{f}_n(\mu)]$

The corollary follows by noting that $\hat{f}_n(\mu) - \tilde{f}_n(\mu)$ is stochastically bounded.

**Corollary 3.** *If $\alpha = \alpha_n \to 0$ as $n \to \infty$, then $h^2(\tilde{f}_n, \hat{f}_n) \to 0$ in probability.*

## 3. Estimation of the Mode

If the mode is unknown, estimates $\hat{f}_{(i)}$ of the density can be obtained by assuming that the mode is in the interval $(x_{i-1}, x_i]$, for $i = 2, \ldots, n$, and the estimate chosen which maximizes the likelihood. This gives an estimate of the mode. If $\hat{m}$ maximizes $\hat{f}_{(i)}$ over $i$, we can estimate the mode to be the midpoint of the unimodal interval, or $\hat{\mu}_n = (x_{\hat{u}_1(n)-1} + x_{\hat{u}_2(n)})/2$. We show this is consistent and, in Section 4, we compare it to the other mode estimators.

Suppose $x_1, \ldots, x_n$ is a sorted random sample from $f$, and $F_n$ is the empirical distribution function. Let $h(f, g)$ be the Hellinger distance between the densities $f$ and $g$, as in (6).

**Theorem 3.** *Let $f$ be continuous and strictly unimodal with mode $\mu$, and $\int \sqrt{f} dx < \infty$. Let $\hat{\theta}$ be the penalized alternative density estimator, with $\alpha_n$ is as in* (5)*. Then*

$$h^2(f, \hat{f}_n) \to 0. \tag{7}$$

This theorem and Lemma 3 of Section 5 give the following.

**Corollary 4.** $\hat{\mu}_n \to^p \mu$.

## 4. Simulations

We compare both the density estimator and the mode estimator to other nonparametric estimators through simulations, with data simulated from four densities.

1. the standard gaussian density,
2. the standard exponential density,
3. an asymmmetric density ($f(x) = 2[\exp(-x)I\{x > 0\} + \exp(2x)I\{x < 0\}]/3$),
4. the gamma density with shape parameter 2, shifted so the mode is at the origin.

For the choice of the penalty parameter $\alpha$, we refer to the results of Woodroofe and Sun (1993). For the derivation of the asymptotic distribution of the value of the PMLE at the mode, they require certain choices of $\alpha$ that depend on the derviatives of the underlying density at the mode. This suggests the use of $\alpha = n^{-3/5}$ for densities 1 and 4, and $\alpha = n^{-2/3}$ for 2 and 3. We calculate the APMLE for both choices. The Fortran 90 code for the APMLE can be found at www.stat.uga.edu/∼mmeyer/abs10.html.

One thousand samples from each of the densities were generated, for each of three sample sizes: n=40, 100, and 200. For each of these 12,000 samples, four estimates of the density were calculated, using the APMLE with both choices of $\alpha$, the kernel estimator, and the modified likelihood method. For the kernel estimates, we use

$$\hat{f}_K(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right),$$

where $K(\cdot)$ is the standard normal density, $h = 1.06sn^{-1/5}$, and $s$ is the sample standard deviation. See Figure 1 for each estimation method. The true density is shown as a dashed line in each plot.

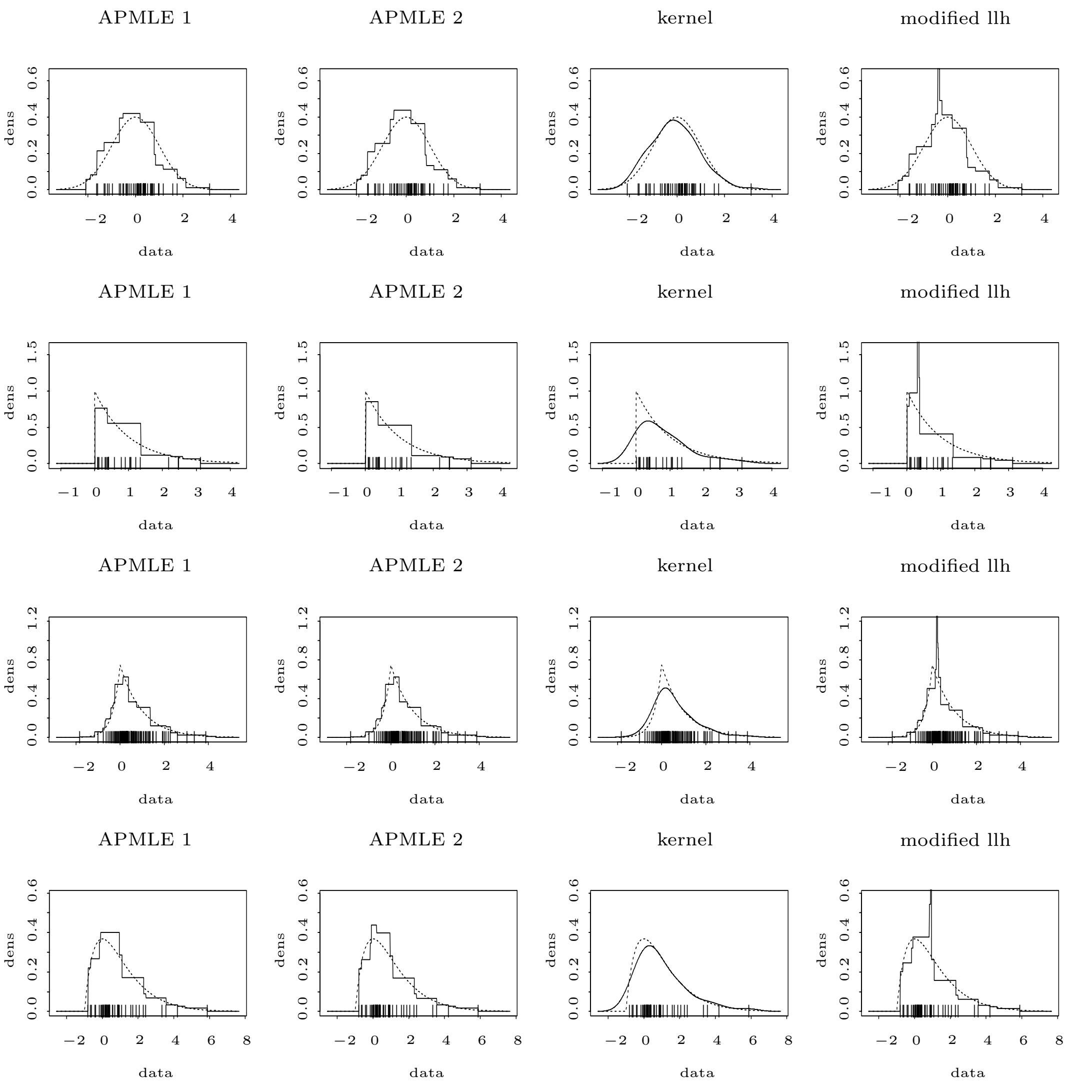


Figure 1. Typical Density Estimators shown with True Density.

The first table contains the average $h^2(g, f)$ for estimates $g$ and underlying densities $f$. We see that for the Gaussian density, the kernel estimator performs well, but has trouble with peaks or finite support. This can also be seen in the figure. The Hellinger distance for the APMLE is not much affected by the choice of $\alpha$.

Table 1

| Hellinger Distance Results (mean, SD) | | | | | |
|---|---|---|---|---|---|
| Density | $n$ | APMLE, $\alpha = n^{-3/5}$ | APMLE, $\alpha = n^{-2/3}$ | Kernel | Modified |
| 1 | 40 | 0.078 (1.1e-3) | 0.078 (1.0e-3) | 0.022 (4.3e-4) | 0.12 (2.0e-3) |
| | 100 | 0.039 (4.3e-4) | 0.039 (4.3e-4) | 0.012 (2.3e-4) | 0.056 (8.5e-4) |
| | 200 | 0.023 (2.2e-4) | 0.024 (2.3e-4) | 0.0074 (1.3e-4) | 0.040 (4.6e-4) |
| 2 | 40 | 0.078 (9.4e-4) | 0.077 (9.4e-4) | 0.19 (1.2e-3) | 0.10 (1.7e-3) |
| | 100 | 0.038 (4.3e-4) | 0.037 (4.2e-4) | 0.17 (7.0e-4) | 0.054 (3.2e-3) |
| | 200 | 0.023 (2.4e-4) | 0.022 (2.6e-4) | 0.15 (4.6e-4) | 0.031 (1.1e-3) |
| 3 | 40 | 0.085 (8.9e-4) | 0.083 (8.9e-4) | 0.046 (7.0e-4) | 0.11 (2.3e-3) |
| | 100 | 0.043 (4.0e-4) | 0.042 (3.8e-4) | 0.029 (3.7e-4) | 0.060 (1.9e-3) |
| | 200 | 0.026 (2.1e-4) | 0.025 (2.1e-4) | 0.020 (2.2e-4) | 0.035 (1.2e-3) |
| 4 | 40 | 0.073 (9.2e-4) | 0.073 (9.4e-4) | 0.089 (8.6e-4) | 0.011 (1.7e-3) |
| | 100 | 0.037 (4.5e-4) | 0.037 (4.2e-4) | 0.065 (4.8e-4) | 0.055 (1.1e-3) |
| | 200 | 0.022 (2.2e-4) | 0.022 (2.3e-4) | 0.052 (3.0e-4) | 0.0452 (1.2e-3) |

Five estimates of the mode are compared, using APMLE with both choices of $\alpha$, the kernel estimator, the Robertson-Cryer method, and a "Bayes" estimate. The Bayes estimate is taken from Bickel and Fan (1996), and is calculated by weighting the data by the modified likelihood:

$$\hat{m}_B = \sum_i \frac{L(i)}{\sum_j L(j)} x_i,$$

where $L(i)$ is the likelihood of the MLE with the $i$th datapoint deleted. Robertson and Cryer (1974) proposed a procedure for estimating the mode which chooses a sequence of subintervals where the sample points are the most dense. Given a random sample from an unknown unimodal distribution, integers $\{k_i\}$ and l, the first step is to find the smallest interval containing $k_1$ observations, say $(l_1, r_1)$. Next, find the smallest subinterval of $(l_1, r_1)$ containing $k_2$ observations, and continue until the interval has less than $l$ observations. Robertson and Cryer give conditions on the $l$ and $\{k_i\}$, depending on the sample size $n$, that ensure consistency. For the penalized alternative estimator, if the estimated modal interval includes either $x_1$ or $x_n$, we take the mode to be that endpoint. The mode estimate comparisons are summarized in Table 2. Again, the kernel estimator is best for the Gaussian density, but the AMPLE is a good choice for the other three densities. Note that for the Gamma density, the kernel has smaller SEL but greater bias compared to the AMPLE. The APMLE mode estimate seems not to be sensitive to the choice of $\alpha$.

Table 2

| Mode Estimation Results (mean, ASEL) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dens | $n$ | APMLE, $\alpha_1$ | | APMLE, $\alpha_2$ | | Kernel | | R-C | | Bayes | |
| 1 | 40 | -4.1e-3 | (0.18) | -8.7e-3 | (0.20) | -4.2e-3 | (0.11) | -5.4e-3 | (0.18) | 0.062 | (0.16) |
| | 100 | 0.020 | (0.11) | -0.014 | (0.15) | 0.014 | (0.068) | 0.027 | (0.14) | 0.025 | (0.11) |
| | 200 | -3.8e-3 | (0.084) | -0.012 | (0.11) | -1.3e-3 | (0.046) | -2.7e-3 | (0.098) | -5.4e-3 | (0.081) |
| 2 | 40 | 0.12 | (0.051) | 0.16 | (0.074) | 0.52 | (0.29) | 0.34 | (0.16) | 0.36 | (0.18) |
| | 100 | 0.13 | (0.045) | 0.14 | (0.046) | 0.47 | (0.23) | 0.24 | (0.083) | 0.27 | (0.095) |
| | 200 | 0.11 | (0.030) | 0.12 | (0.031) | 0.43 | (0.19) | 0.19 | (0.052) | 0.21 | (0.058) |
| 3 | 40 | 0.21 | (0.091) | 0.20 | (0.084) | 0.28 | (0.097) | 0.21 | (0.093) | 0.23 | (0.11) |
| | 100 | 0.16 | (0.043) | 0.14 | (0.041) | 0.24 | (0.066) | 0.15 | (0.53) | 0.17 | (0.059) |
| | 200 | 0.13 | (0.026) | 0.12 | (0.025) | 0.21 | (0.049) | 0.11 | (0.031) | 0.13 | (0.034) |
| 4 | 40 | 0.19 | (0.34) | 0.23 | (0.34) | 0.39 | (0.24) | 0.30 | (0.29) | 0.35 | (0.32) |
| | 100 | 0.17 | (0.16) | 0.17 | (0.20) | 0.30 | (0.13) | 0.19 | (0.20) | 0.23 | (0.17) |
| | 200 | 0.13 | (0.12) | 0.14 | (0.16) | 0.24 | (0.080) | 0.15 | (0.15) | 0.19 | (0.13) |

Table 3 contains comparisons of the heights at the mode of the APMLE estimates with the kernel estimates. Here the choice of $\alpha$ is important. The smaller value allows more peaking at the mode, which makes it more appropriate for densities 2 and 3. The kernel estimates are too short for the peaked densities.

Table 3

| Mode Height Estimation Results (mean, ASEL) | | | | | | | |
|---|---|---|---|---|---|---|---|
| Density | $n$ | APMLE, $\alpha_1$ | | APMLE, $\alpha_2$ | | Kernel | |
| 1 | 40 | 0.42 | (4.2e-3) | 0.44 | (6.9e-3) | 0.38 | (3.3e-3) |
| ht=0.3989 | 100 | 0.41 | (2.2e-3) | 0.44 | (4.4e-3) | 0.38 | (1.8e-3) |
| | 200 | 0.41 | (1.7e-3) | 0.44 | (3.5e-3) | 0.38 | (1.2e-3) |
| 2 | 40 | 0.65 | (0.14) | 0.70 | (0.11) | 0.49 | (0.26) |
| ht=1.0 | 100 | 0.69 | (0.10) | 0.75 | (0.073) | 0.53 | (0.23) |
| | 200 | 0.73 | (0.077) | 0.78 | (0.054) | 0.55 | (0.20) |
| 3 | 40 | 0.54 | (0.053) | 0.59 | (0.036) | 0.45 | (0.098) |
| ht = 0.75 | 100 | 0.58 | (0.036) | 0.62 | (0.025) | 0.47 | (0.081) |
| | 200 | 0.59 | (0.028) | 0.63 | (0.018) | 0.49 | (0.067) |
| 4 | 40 | 0.38 | (3.9e-3) | 0.41 | (6.8e-3) | 0.30 | (6.3e-3) |
| ht=0.3679 | 100 | 0.39 | (2.3e-3) | 0.41 | (4.7e-3) | 0.32 | (3.3e-3) |
| | 200 | 0.038 | (1.5e-3) | 0.41 | (3.9e-3) | 0.33 | (2.3e-3) |

## 5. Proofs

**Proof of Theorem 1.** Since $F(x_1)$, $F(x_2), \ldots, F(x_n)$ are equal in distribution to the order statistics from a random sample from a uniform density, we have $[F(x_1),\ldots,F(x_n)] =^{\mathcal{D}} [\Gamma_1,\ldots,\Gamma_n]/\Gamma_{n+1}$, where $\Gamma_i = Y_1+\cdots+Y_i$, $Y_i \sim^{i.i.d.}$ $Exp(1)$. Since $m = \#\{k \le n : X_k \le \mu\}$, $m$ is a binomial random variable with parameters $F(\mu)$ and $n$. The conditional joint distribution of the random variables $F(x_k)-F(x_j)$, with $1 \le j < m \le k \le n$, given $m$, may be described as follows. Let

$$F_-(x) = \frac{F(x)}{F(\mu)},\ x \le \mu \ \text{ and } \ F_+(x) = \frac{F(x)-F(\mu)}{1-F(\mu)},\ x \ge \mu,$$

so that $F(x_k)-F(x_j) = F(\mu)\,[1-F_-(x_j)]+[1-F(\mu)]\,[F_+(x_k)]$ for $1 \le j < m \le k \le n$. Now the conditional joint distribution of $x_1,\ldots,x_{m-1}$, given $m$, is the same as the distribution of order statistics of a sample of size $m-1$ from $F_-$; and, similarly, the conditional joint distribution of $x_m,\ldots,x_n$ is the same as the distribution of order statistics of a sample of size $n-m+1$ from $F_+$. So the conditional joint distribution of $F(x_k)-F(x_j)$, $1 \le j < m \le k \le n$, given $m$, is the same as that of

$$F(\mu) \times \frac{Y_{j-m}+\cdots+Y_0}{Y_{-m+1}+\cdots+Y_0} + [1-F(\mu)] \times \frac{Y_1+\cdots+Y_{k-m+1}}{Y_1+\cdots+Y_{n-m+1}},$$

$1 \le j < m \le k \le n$, where $\ldots, Y_{-1}, Y_0, Y_1, \ldots$ are i.i.d. standard exponential random variables. Note that

$$\frac{nF(\mu)}{Y_{-m+1}+\cdots+Y_0} \to 1, \tag{8}$$

$$\frac{n[1-F(\mu)]}{Y_1+\cdots+Y_{n-m+1}} \to 1 \tag{9}$$

in (unconditional) probability as $n \to \infty$, since $(Y_{-m+1}+\cdots+Y_0)/m \to 1$ and $(Y_1+\cdots+Y_{n-m+1})/(n-m+1) \to 1$, and $m/n \to F(\mu)$. Let

$$\Delta_n^d = \sup_{j<m,k\ge m,k-j\ge d} \frac{(k-j)/n}{F(x_k)-F(x_j)}. \tag{10}$$

**Lemma 2.** $\Delta_n^d \to^{\mathcal{D}} \Delta^d$, *where*

$$\Delta^d = \sup_{k\ge 0,j<0,k-j\ge d} \frac{k-j}{Y_{j+1}+\cdots+Y_k}.$$

**Proof.** Let $G_n(m;y) = P\{\Delta_n^d \le y \mid m\}$ and let $G$ denote the distribution of $\Delta^d$. Then it follows directly from (8) and (9) that $G_n(m;\cdot) \to^{\mathcal{D}} G$ with probability 1, in the Levy metric say. To show that the unconditional distributions converge, let $H_n(y) = P(\Delta_n^d \leq y)$ and let $h$ be any a bounded continuous function. Then

$$\int_{\mathbb{R}} h dH_n = \int \left[ \int_{\mathbb{R}} h(y) G_n(m; dy) \right] dP$$

and the integrand converges boundedly to $\int_{\mathbb{R}} h dG$ with probability 1.

**Proof of the Theorem.** Fix $d \geq 1$. For each $n$, choose $l_1$ and $l_2$ such that $l_1 \leq m \leq l_2$ and $d = l_2 - l_1 + 1$. Then

$$\begin{aligned}
&\min_{l_1 \leq u_1 \leq m \leq u_2 \leq l_2} \frac{F(x_{u_2}) - F(x_{u_1 - 1})}{x_{u_2} - x_{u_1 - 1}} \times \max_{l_1 \leq u_1 \leq m \leq u_2 \leq l_2} \frac{(u_2 - u_1 + 1)/n}{F(x_{u_2}) - F(x_{u_1 - 1})} \\
&\leq \tilde{\theta}_{m,n} = \max_{u_1 \leq m \leq u_2} \frac{F(x_{u_2}) - F(x_{u_1 - 1})}{x_{u_2} - x_{u_1 - 1}} \times \frac{(u_2 - u_1 + 1)/n}{F(x_{u_2}) - F(x_{u_1 - 1})} \\
&\leq f(\mu) \max_{u_1 \leq m \leq u_2} \frac{(u_2 - u_1 + 1)/n}{F(x_{u_2}) - F(x_{u_1 - 1})} = f(\mu) \Delta_n^1.
\end{aligned}$$

Letting first $n \to \infty$ then $d \to \infty$, we have $\frac{\tilde{\theta}_{m,n}}{f(\mu)} \to^{\mathcal{D}} \Delta^1 = \Delta$.

**Proof of Proposition 1.** Choose $0 < \gamma_0 < 1$. Since $\gamma \hat{\theta}_{m,n}(\gamma)$ is increasing in $0 < \gamma \leq 1$, we have

$$\begin{aligned}
&P\Big\{ \sup_{\gamma_0 \leq \gamma \leq 1} \mid \gamma \hat{\theta}_{m,n}(\gamma) - f(\mu) \mid \geq \epsilon \Big\} \\
&\leq P\{\hat{\theta}_{m,n}(1) - f(\mu) \geq \epsilon\} + P\{\gamma_0 \hat{\theta}_{m,n}(\gamma_0) - f(\mu) < -\epsilon\}
\end{aligned}$$

for all $\epsilon > 0$. So it suffices to show

$$p\text{-}\lim_{n \to \infty} [\hat{\theta}_{m,n}(1) - f(\mu)]_+ = 0 \tag{11}$$

and

$$p\text{-}\lim_{n \to \infty} [\gamma \hat{\theta}_{m,n}(\gamma) - f(\mu)]_- = 0, \text{ for all } \gamma_0 \leq \gamma \leq 1. \tag{12}$$

Let $k = k(n)$ be the smallest integer so that $k/n \geq \sqrt{\alpha}$. For each n, choose indices $u_1$ and $u_2$ such that $u_1 \leq m \leq u_2$ and $u_2 - u_1 + 1 = k$. Conditional on $m$, we have $(F(x_{u_2}) - F(x_{u_1 - 1}))/((u_2 - u_1 + 1)/n)$ is equal in distribution to

$$\frac{\frac{nF(\mu)}{Y_{-m+1} + \cdots + Y_0} [Y_{u_1 - m} + \cdots + Y_0] + \frac{n[1 - F(\mu)]}{Y_1 + \cdots + Y_{n-m}} [Y_1 + \cdots + Y_{u_2 - m + 1}]}{u_2 - u_1 + 1},$$

which converges in probability to 1 as $n \to \infty$ then $k \to \infty$, using (8) and (9). Since $(u_2 - u_1 + 1)/n \leq \sqrt{\alpha} + 1/n \to 0$, we have $(u_2 - u_1 + 1)n/[(x_{u_2} - x_{u_1 - 1})] \to^p f(\mu)$ for the specially chosen sequence of $u_1$ and $u_2$. Using again that $\alpha = o(k/n)$,

$p$–$\lim_{n\to\infty}[\gamma(u_2-u_1+1)/n]/[\alpha+\gamma(x_{u_2}-x_{u_1-1})]=f(\mu)$, $\forall\gamma_0\le\gamma\le 1$, for the sequence of $u_1$ and $u_2$. Now by (4),

$$\gamma\hat{\theta}_{m,n}(\gamma)=\max_{u_1\le m\le u_2}\frac{\gamma(u_2-u_1+1)/n}{\alpha+\gamma(x_{u_2}-x_{u_1-1})}\ge\frac{\gamma(u_2-u_1+1)/n}{\alpha+\gamma(x_{u_2}-x_{u_1-1})}, \qquad (13)$$

for all $u_1$, $u_2$ such that $u_1\le m\le u_2$, and so $p-\lim_{n\to\infty}[\gamma\hat{\theta}_{m,n}(\gamma)-f(\mu)]_-=0,\forall\gamma_0\le\gamma\le 1$. To show (11), note that

$$\begin{aligned}\hat{\theta}_{m,n}(1)&=\max_{u_1\le m\le u_2}\left(\frac{(u_2-u_1+1)/n}{F(x_{u_2})-F(x_{u_1-1})}\times\frac{F(x_{u_2})-F(x_{u_1-1})}{\alpha+(x_{u_2}-x_{u_1-1})}\right)\\&\le\max_{u_1\le m\le u_2}\left(\frac{(u_2-u_1+1)/n}{F(x_{u_2})-F(x_{u_1-1})}\times\frac{f(\mu)(x_{u_2}-x_{u_1-1})}{\alpha+(x_{u_2}-x_{u_1-1})}\right).\end{aligned}$$

Fix $d\ge 1$ and for each $n>d$, choose $l_1$ and $l_2$ such that $l_1\le m\le l_2$ and $d=l_2-l_1+1$. Then

$$\begin{aligned}&\max_{l_1\le u_1\le m\le u_2\le l_2}\left(\frac{(u_2-u_1+1)/n}{F(x_{u_2})-F(x_{u_1-1})}\times\frac{f(\mu)(x_{u_2}-x_{u_1-1})}{\alpha+(x_{u_2}-x_{u_1-1})}\right)\\&\le\Delta_n^d\times\max_{l_1\le u_1\le m\le u_2\le l_2}\frac{nf(\mu)(x_{u_2}-x_{u_1-1})}{n\alpha+n(x_{u_2}-x_{u_1-1})}\to^p 0,\end{aligned}$$

since $n\alpha\to\infty$, and $\Delta_n^d$ and $\max_{l_1\le u_1\le m\le u_2\le l_2}n(x_{u_2}-x_{u_1-1})$ are stochastically bounded in $n=d+1,d+2,\ldots$. To see the latter, note that

$$\begin{aligned}\max_{l_1\le u_1\le m\le u_2\le l_2}n(x_{u_2}-x_{u_1-1})&=n(x_{l_2}-x_{l_1-1})\\&=\frac{x_{l_2}-x_{l_1-1}}{F(x_{l_2})-F(x_{l_1-1})}\times n[F(x_{l_2})-F(x_{l_1-1})].\end{aligned}$$

The first term is close to $1/f(\mu)$ and the second is of order $d$.

Now, $\hat{\theta}_{m,n}(1)\ge\gamma_0\hat{\theta}_{m,n}(\gamma_0)\ge f(\mu)+o_p(1)$ since $\gamma\hat{\theta}_{m,n}(\gamma)$ is non-decreasing in $\gamma$. We have with probability approaching one,

$$\begin{aligned}\hat{\theta}_{m,n}(1)&\le\max_{u_1\le l_1\le m\le l_2\le u_2}\frac{f(\mu)(x_{u_2}-x_{u_1-1})}{\alpha+(x_{u_2}-x_{u_1-1})}\times\frac{(u_2-u_1+1)/n}{F(x_{u_2})-F(x_{u_1-1})}\\&\le f(\mu)\Delta_n^d\to^{\mathcal{D}} f(\mu)\Delta^d,\end{aligned}$$

by Lemma 2. Let $d\to\infty$ to get $p$–$\lim_{n\to\infty}[\hat{\theta}_{m,n}(1)-f(\mu)]_+=0$.

**Proof of Proposition 2.** For any $\delta>0$, define $l(n)=\min\{l:x_l\ge\mu+\delta\}$. Let $\Omega_n=\{x_{u_2(n)}>\delta\}$, a subset of the underlying probability space. For $\omega\in\Omega_n$, we have

$$\tilde{\theta}_{m,n}=\max_{u_1\le m(n),u_2\ge l(n)}\frac{F_n(x_{u_2})-F_n(x_{u_1-1})}{x_{u_2}-x_{u_1-1}}.$$

Now,

$$P\left\{\max_{u_1 \le m(n), u_2 \ge l(n)} \frac{F_n(x_{u_2}) - F_n(x_{u_1-1})}{x_{u_2} - x_{u_1-1}} < f(\mu)\right\} \to 1$$

by the convergence of the empirical distribution function, so $P\{[\tilde{\theta}_{m,n} > f(\mu)] \cap \Omega_n\} \to 0$.

Writing $P\{\tilde{\theta}_{m,n} > f(\mu)\} = P\{[\tilde{\theta}_{m,n} > f(\mu)] \cap \Omega_n\} + P\{[\tilde{\theta}_{m,n} > f(\mu)] \cap \Omega_n^C\}$ we have that the left hand side goes to one by Theorem 1, so $P\{\Omega_n^C\} \to 1$.

Since this is true for all $\delta > 0$, $p-\lim_{n\to\infty} x_{u_2(n)} = \mu$. The proof of the second statement is similar.

If $f$ is continuously differentiable near $x$, then $\max\{f(x-\delta) - f(x),\ f(x) - f(x+\delta)\} \le C\delta$ for some $C$. Then, setting $\delta = \sqrt{R_n}$ in (14) yields $\mid \tilde{f}_n(x) - f(x) \mid \le (C+2)\sqrt{R_n} = O_p(n^{-1/4})$ as $n \to \infty$.

**Proof of Proposition 3.** Suppose that $x > \mu$, let $0 < \delta < (x-\mu)/2$, and let $A_n$ be the event $A_n = \{x_{u_2} \ge (\mu + x)/2\}$. Then $P(A_n) \to 1$ as $n \to \infty$ by Proposition 3, and $A_n$ implies

$$f(x+\delta) \le \frac{F(x+\delta) - F(x)}{\delta} \le \frac{\tilde{F}_n(x+\delta) - \tilde{F}_n(x)}{\delta} + 2\frac{R_n}{\delta} \le \tilde{f}_n(x) + 2\frac{R_n}{\delta}$$

where $R_n = \sup_{x \ge x_{u_2}} \mid \tilde{F}_n(x) - F(x) \mid$. Similarly, $A_n$ implies $f(x-\delta) \ge \tilde{f}_n(x) - 2R_n/\delta$. Thus,

$$\mid \tilde{f}_n(x) - f(x) \mid \le \max\{f(x-\delta) - f(x), f(x) - f(x+\delta)\} + 2\frac{R_n}{\delta}. \tag{14}$$

Given $\epsilon > 0$, let $\delta$ be so small that $\max\{f(x-\delta) - f(x), f(x) - f(x+\delta)\} \le \epsilon/2$. Then $P\left\{\mid \tilde{f}_n(x) - f(x) \mid \ge \epsilon\right\} \le P\left\{R_n \ge \epsilon\delta/4\right\} + P(A_n^c)$, which approaches zero as $n \to \infty$.

**Proof of Theorem 2.** Since $\hat{f}_n$ maximizes the penalized likelihood,

$$\begin{aligned} 0 &\le l_n(\hat{f}_n) - l_n(\tilde{f}_n) \\ &= n\left[\int_{-\infty}^{\infty} \log \hat{f}_n dF_n - \int_{-\infty}^{\infty} \log \tilde{f}_n dF_n\right] - n\alpha[\hat{f}_n(\mu) - \tilde{f}_n(\mu)]. \end{aligned}$$

From Theorem 1.2.1 of Robertson, et al. (1988), $\tilde{f}_n$ increases or decreases only at values $x_k$ for which $\tilde{F}_n(x_k) = F_n(x_k)$. It follows that $\int_{-\infty}^{\infty} \log \tilde{f}_n dF_n = \int_{-\infty}^{\infty} \log \tilde{f}_n d\tilde{F}_n$, and, therefore, that

$$\begin{aligned} 0 &\le \int_{-\infty}^{\infty} \log \frac{\hat{f}_n}{\tilde{f}_n} d\tilde{F}_n + \int_{-\infty}^{\infty} (\tilde{F}_n - F_n) d \log \hat{f}_n - \alpha[\hat{f}_n(\mu) - \tilde{f}_n(\mu)] \\ &\le -h^2(\tilde{f}_n, \hat{f}_n) - \alpha[\hat{f}_n(\mu) - \tilde{f}_n(\mu)]. \end{aligned}$$

The final inequality follows by noting that $\tilde{F}_n - F_n \geq 0$ where $\hat{f}_n$ is decreasing and $\tilde{F}_n - F_n \leq 0$ where $\hat{f}_n$ is increasing, by writing $\log x = 2\log\sqrt{x} \leq 2(\sqrt{x}-1)$ for $0 < x < \infty$, and by the second expression for squared Hellinger distance.

**Proof of Theorem 3**

**Lemma 3.** *For $f$ strictly unimodal with mode $\mu$, there exists a $\delta > 0$ such that for all unimodal $g$ with mode $\mu_0 \neq \mu$, we have*

$$h^2(f,g) > \delta \tag{15}$$

*where $\delta$ depends on $\mu_0$, $\mu$, and $f$.*

**Proof.** Suppose without loss of generality that $\mu_0 < \mu$, and let $\mathcal{G}$ be the set of unimodal densities with mode $\mu_0$. Since $f$ is strictly unimodal, we have $f(\mu_0) < f(\mu)$. Let $h_0^2(f,g) = \int_{\mu_0}^{\mu}(\sqrt{f}-\sqrt{g})^2 dx$. We show that, since all $g \in \mathcal{G}$ are nonincreasing on $[\mu_0,\mu]$, the function which minimizes $h_0^2(f,g)$ over $\mathcal{G}$ is the constant function $c$ where $\sqrt{c} = \int_{\mu_0}^{\mu}\sqrt{f}dx/(\mu-\mu_0)$. Expanding the square gives

$$\begin{aligned}&\int_{\mu_0}^{\mu}\left(\sqrt{f}-\sqrt{g}\right)^2 dx\\ &= \int_{\mu_0}^{\mu}\left(\sqrt{f}-\sqrt{c}\right)^2 dx + \int_{\mu_0}^{\mu}(\sqrt{c}-\sqrt{g})^2 dx - 2\int_{\mu_0}^{\mu}\left(\sqrt{f}-\sqrt{c}\right)(\sqrt{g}-\sqrt{c})\,dx\end{aligned}$$

and the last term is negative since it is the covariance between a non-increasing function and a non-decreasing one. This gives $\inf_{g\in\mathcal{G}} h^2(f,g) \geq h^2(f,c)$ and the lemma is proved with $\delta = h^2(f,c)/2 > 0$.

**Lemma 4.** *If $g$ is unimodal and $0 \leq g(x) \leq c$, $\forall x$, then*

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\log[g(x_i)+\epsilon] - \int\log[g+\epsilon]dF\Big| \leq 2\sup_x\Big|F_n(x)-F(x)\Big|\log\left(\frac{c+\epsilon}{\epsilon}\right). \tag{16}$$

**Proof.**

$$\begin{aligned}&\Big|\frac{1}{n}\sum_{i=1}^{n}\log[g(x_i)+\epsilon] - \int\log[g+\epsilon]dF\Big| = \Big|\int\log[g+\epsilon]d(F_n-F)\Big|\\ &= \Big|\int[F_n(x)-F(x)]d\log(g+\epsilon)\Big|\\ &\leq 2\sup_x|F_n(x)-F(x)|\,[\log(c+\epsilon)-\log(\epsilon)].\end{aligned}$$

**Lemma 5.** *We have*

$$\int\log\left(\frac{g+\epsilon}{f}\right)f dx \leq 2\sqrt{\epsilon}\int\sqrt{f}dx - h^2(f,g) \tag{17}$$

**Proof.**

$$\int \log\left(\frac{g+\epsilon}{f}\right) f dx \le 2\int\left(\sqrt{\frac{g+\epsilon}{f}}-1\right) f dx$$
$$= 2\int \sqrt{f}\left(\sqrt{g+\epsilon}-\sqrt{g}\right) dx - 2 + 2\int \sqrt{fg}dx$$

The lemma follows since the last two terms are $-h^2(f,g)$, and the function $\sqrt{y+\epsilon}-\sqrt{y}$ is maximized at $y=0$.

**Proof of the Theorem.** As in (3), let $l_n(g) = \sum_{i=1}^n \log g(x_i) - n\alpha_n g(\mu_g)$, where $\mu_g$ maximizes $g(x)$. Since $\hat{f}_n$ is the maximum likelihood estimate, we have

$$l_n(\hat{f}_n) \ge l_n(f),\ \forall n. \tag{18}$$

For the penalized estimate, we have

$$\hat{f}_n(x) \le \frac{1}{\alpha_n},\ \forall n \tag{19}$$

by (4). Let $\mathcal{U}_c$ be the set of unimodal densities $g$ with $g(x) \le c$. Then

$$\frac{1}{n}[l_n(g) - l_n(f)] \le \frac{1}{n}\sum_{i=1}^n \log[g(x_i)+\epsilon] - \int \log(g+\epsilon)dF$$
$$+\int \log\Big(\frac{g+\epsilon}{f}\Big)dF - \Big[\frac{1}{n}\sum_{i=1}^n \log f(x_i) - \int \log f dF\Big] - \alpha_n[g(x_{\hat{m}(n)}) - f(x_{\hat{m}(n)})].$$

By Lemmas 3 and 4 and the Law of Large Numbers, this is not greater than

$$2\sup_x |F_n(x)-F(x)| \log\left(\frac{c+\epsilon}{\epsilon}\right) + 2\sqrt{\epsilon}\int \sqrt{f}dx - h^2(f,g) + o_p(1) + \alpha_n f(\mu).$$

Let $c_n = 1/\alpha_n$ and $\epsilon_n = 1/n$. Then

$$\frac{1}{n}\left[l_n(\hat{f}_n) - l_n(f)\right] \le \sup_x |F_n(x)-F(x)| \log\left(1+\frac{n}{\alpha_n}\right)$$
$$+2n^{-\frac{1}{2}}\int \sqrt{f}dx - h^2(\hat{f}_n, f) + o_p(1) + \alpha_n f(\mu).$$

Rearranging the inequality and using (18) gives:

$$h^2(\hat{f}_n, f) \le \sup_x |F_n(x)-F(x)| \log\left(1+\frac{n}{\alpha_n}\right) + 2n^{-\frac{1}{2}}\int \sqrt{f}dx + o_p(1)$$

which goes to zero since $\sup_x |F_n(x)-F(x)|$ goes to zero like $n^{-1/2}$.

## References

Bickel, P. and Fan, J. (1996). Some problems of the estimation of unimodal densities, *Statist. Sinica* **6**, 23-45.

Grenander, P. (1956). On the theory of mortality measurement, part II. *Skand. Akt.* **39**, 125-153.

Groeneboom, P. (1985). Estimating a monotone density. In *Proc. Conf. in Honor of Jerzy Neyman and Jack Kiefer* (Edited by L. LeCam and R. Olshen), 539-555.

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

Robertson, T. and Cryer, J. (1974). An iterative procedure for estimating the mode. *J. Amer. Statist. Assoc.* **69**, 1012-1016.

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

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

Woodroofe, M. and Sun, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when $f$ is non-increasing. *Statist. Sinica* **3**, 501-515.

Department of Statistics, University of Georgia, Athens, GA 30602-1952, U.S.A.

E-mail: mmeyer@stat.uga.edu

(Received July 1999; accepted February 2001)