A NOTE ON TESTING FOR MULTI-MODALITY WITH DEPENDENT DATA

K. S. Chan[1]
University of Iowa, Iowa City, U.S.A.

Howell Tong
London School of Economics, U.K. & University of Hong Kong, Pokfulam, Hong Kong

May, 2002

SUMMARY

We have proposed a test for multi-modality with dependent data by resampling from a suitably constructed transition probability kernel, which includes Silverman's test with independent data as a special case. As far as we know, this is the first test of its kind for dependent data. We have extended some theoretical properties of Silverman's test with iid data to weakly dependent data, and also discussed the robustness of Silverman's test against departure from independence.

Some Key Words: Markov chain; Smoothed Autoregressive Bootstrap; Gaussian kernels; Silverman's test; Uniform ergodicity.

1 INTRODUCTION

Our interest in multi-modality is motivated by our study of dynamical systems. Now, a dynamical system may have multiple equilibria. In a random environment, they may exhibit themselves in the form of a multi-modal probability distribution, as a reflection of a mixing of distributions (with

[1] *Address for correspondence:* Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242, USA. (E-mail: kchan@stat.uiowa.edu).

locations around the equilibria) effected by the system dynamics. Besides insights on the underlying dynamics, multi-modality has important relevance to forecasting. For example, the sample mean may actually be a poor long-term forecast! The shape of the (1-dimensional) stationary marginal probability density function (pdf) of a time series $\{Y_t\}$ often sheds useful insights on the dynamics of the underlying process. Clearly, a bimodal marginal density implies that the underlying process is not a Gaussian linear process. Multi-modality can easily arise. Consider a first order threshold autoregressive (TAR) model with s regimes, all lag-1 coefficients being zero, and with Gaussian innovations, as a stylised form of a multi-equilibria dynamical system subject to Gaussian white noise, whose stationary marginal density is ordinarily a mixture of s normal densities; Jalali and Pemberton (1995). This result generalizes the well-known fact that for iid data from a mixing of s distinct populations, the number of modes is s if the populations are well separated. More generally, in a stochastic dynamical system, the number of modes in the stationary density provides important information on the number of local operating modes for the underlying dynamical process.

In the particular case of independent data, the problem of testing for multi-modality in the density has received much attention in the literature; Good and Gaskins (1980), Silverman (1981), Titterington et al. (1985), Hartigan and Hartigan (1985), Müller and Sawitzki (1991), Fisher et al. (1994), and Cheng and Hall (1998). Of particular relevance here is Silverman (1981), who has proposed a bootstrap test for testing H_0: the density has m modes versus the alternative hypothesis H_1 that the density has more than m modes.

As far as we are aware, there has not been much discussion in the literature on the testing for multi-modality with dependent data. In his contribution to the discussion of Hinkley, Tong (1988) has illustrated the potential of Silverman's bootstrap test for dependent data but without giving any justification.

In this note, we generalize Silverman's bootstrap test to the case of dependent data that are generated by a stationary pth order Markov process. Now, the null hypothesis is H_0 : the (1-dimensional) stationary marginal pdf has m modes, and the alternative hypothesis is H_1: the marginal pdf admits more than m modes. In section 2, we introduce a new bootstrap method suitable for finite order Markov processes. We shall call the new bootstrap method the smoothed autoregressive bootstrap. The basic idea of the smoothed autoregressive bootstrap is to estimate the true transition probability kernel using a product Gaussian kernel smoothing scheme, and then generate bootstrap series from the Markov process with the estimated transition probability kernel

as if it were the true one. The bandwidth of the kernel smoothing scheme is set to be the smallest bandwidth for which the estimated unidimensional marginal pdf of the data has m modes. The existence of such a critical bandwidth, denoted as $\hat{h}$, is ensured by a result of Silverman (1981). Bootstrap series are then normalized to match the observed sample variance, and they are used to approximate the p-value, $P(\hat{h} \geq \hat{h}_{\text{obs}})$, of the observed critical bandwidth, evaluated under a least favourable hypothesis consistent with the null hypothesis. We show that the bootstrap test developed here is robust with respect to the Markov order. In particular, Silverman's bootstrap test for multi-modality is robust against departure from the independence assumption. See Efron and Tibshirani (1993) for an introduction to bootstrap. For dependent data, there are in the literature a number of bootstrap methods, e.g. Künsch (1989), Lall and Sharma (1996), Politis and Romano (1994), Rajarshi (1990), and some recent but unpublished work by Chan, Tong and Stenseth, and Paparoditis and Politis. The bootstrap method proposed here is closely related to the methods proposed by Lall and Sharma (1996), Chan, Tong and Stenseth, and Paparoditis and Politis, but it seems to be particularly suitable for the purpose of testing for multi-modality in the marginal density. Silverman (1983) and Mammen et al. (1992) obtained some theoretical properties of Silverman's test with iid data. For a weakly dependent stationary process with sufficiently fast decaying mixing rate, Neumann (1998) showed the existence of a richer probability space on which a copy of the weakly dependent process can be approximated by a sequence of iid random variables so that their kernel density estimators of the 1-dimensional marginal pdf are asymptotically equivalent; see below for the exact result. Using this result, we extend in section 3 some of the theoretical properties of Silverman's test with iid data to weakly dependent data. All proofs are collected in an appendix. In section 4, we illustrate the bootstrap test with some simulated and real data. Finally, we give some concluding remarks in section 5.

2 TESTING FOR MULTI-MODALITY

Let $\{Y_t\}$ be a pth order stationary homogeneous Markov process, that is, $X_t = (Y_{t-p}, \cdots, Y_{t-1})^T$ is a stationary homogeneous Markov chain. We now motivate the introduction of the smoothed autoregressive bootstrap. The (stationary) $(p+1)-$dimensional joint pdf of $Z_t = (X_t^T, Y_t)^T$ can be estimated by the following kernel estimate:

$$\hat{f}_Z(z; h) = \sum g_{p+1}(z - Z_t; h)/m, \tag{1}$$

where all summations are from $p+1 \leq t \leq n$ unless stated otherwise, n is the sample size, $m = n-p$ and $h > 0$ is the bandwidth. The $(p+1)-$dimensional kernel function $g_{p+1}(z;h) = g_{p+1}(z/h)/h^{p+1}$. In this paper, we adopt a product Gaussian kernel, that is, the k-dimensional kernel $g_k(w) = \prod_{j=1}^{k} \psi(w_j)$, where $w = (w_1, \ldots, w_k)^T$ and ψ is the standard normal density. Notice that $g_{p+1}(z) = g_1(y)g_p(x)$, where $x = (y_1, \ldots, y_p)^T$ and $z = (x,y)^T$. This kernel estimate has a number of interesting properties. By marginalising $\hat{f}_Z$ to the y and the x components, it yields the kernel estimates of their corresponding densities:

$$\hat{f}_Y(y;h) = \hat{f}(y;h) = \sum g_1(y - Y_t; h)/m \quad \text{and} \tag{2}$$

$$\hat{f}_X(x;h) = \hat{f}(x;h) = \sum g_p(x - X_t; h)/m. \tag{3}$$

Consequently, the conditional pdf $f_{Y|X}(y|x) = f(y|x)$ can be estimated by

$$\hat{f}(y|x;h) = \sum g_1(y - Y_t; h) g_p(x - X_t; h) / \sum_s g_p(x - X_s; h) \tag{4}$$

$$= \sum w(x,t;h) g_1(y - Y_t; h), \tag{5}$$

where the weights $w(x,t;h) \propto g_p(x - X_t; h)$.

The smoothed autoregressive bootstrap is obtained by drawing realizations from the pth order Markov process with $\hat{f}(y|x;h)$ as its transition probability kernel. The bootstrap series will be denoted as $\{Y_1^*, \ldots, Y_n^*\}$, and we write $X_t^* = (Y_{t-p}^*, \cdots, Y_{t-1}^*)^T$. To implement the smoothed autoregressive bootstrap, we need to specify the order of the Markov process, the kernel bandwidth and the initial values. The Markov order may be consistently estimated by a cross-validation method based on prediction (Cheng and Tong, 1992). As a by-product of their method, it provides an estimate of the optimal bandwidth. However, for the purpose of testing for multi-modality, we determine instead the critical bandwidth which is the smallest bandwidth, $\hat{h}$, for which $\hat{f}_Y(\cdot)$ has m modes, as specified by the null hypothesis. The existence of $\hat{h}$ is guaranteed by a result in Silverman (1981), which states that for Gaussian kernels, the number of modes in the function $\hat{f}_Y(y;h)$ is a right-continuous non-increasing function in h. Setting the bandwidth parameter in the smoothed autoregressive bootstrap to $\hat{h}$ results in a least favourable Markov process consistent with the null hypothesis. We denote the critical bandwidth estimated from the data by $\hat{h}_{\text{obs}}$. There are two natural ways to start the smoothed autoregressive bootstrap. First, the starting values of the smoothed autoregressive bootstrap may be fixed at the observed initial values $Y_1, \cdots, Y_p$. Alternatively, the starting values may be randomized, for example, by setting them to be $Y_I, \cdots, Y_{I+p-1}$,

where I has a uniform distribution over $\{1, 2, \ldots, n-p\}$. The latter scheme will be adopted in the simulations and examples discussed in section 4.

We now extend Silverman's approach for testing H_0 against H_1 for pth order Markov data as follows:

(1) Let $\{Y_1^*, \ldots, Y_n^*\}$ be a smoothed autoregressive bootstrap series which is generated from the Markov process with $\hat{f}_{Y|X}(\cdot|\cdot; \hat{h}_{\text{obs}})$ as its transition probability kernel.

(2) Normalize the bootstrap series so that its variance equals that of the observed data. For simplicity, $\{Y_1^*, \ldots, Y_n^*\}$ denote the normalized series.

(3) Find $\hat{h}^*$ for the bootstrap series.

Replicate steps (1-3), say, B times to get $\hat{h}_1^*, \ldots, \hat{h}_B^*$ and compute the approximate p-value= $\#\{\hat{h}_i^* \geq \hat{h}_{\text{obs}}\}/B$ which estimates the probability $P(\hat{h} \geq \hat{h}_{\text{obs}})$, evaluated under a least favourable Markov process under H_0. Much computing resource can be saved by making use of the fact that $\hat{h}^* \geq \hat{h}_{\text{obs}}$ if and only if the density $\hat{f}_Y(\cdot; \hat{h}_{\text{obs}})$ based on the bootstrap series has more than m modes.

As iid data can be regarded as zero-th order Markov data, the above test includes Silverman's bootstrap test as a special case. Effectively, in this case $w(x, t; h)$ reduces to $\frac{1}{m}$, independent of x. In section 4, the new test method will be illustrated with some simulated and real time series data. It will be seen that the test for multi-modality that is designed for Markov data turns out to be rather insensitive to the Markov order in the case of weak temporal dependence. Some justification for this phenomenon will be provided in the appendix. In particular, for weakly dependent data, the approximate p-value hardly changes when different Markov orders are assumed for the smoothed autoregressive bootstrap. In other words, Silverman's test for multi-modality is robust against departures from the independence assumption within the framework of smoothed autoregressive bootstrap.

3 SOME THEORETICAL PROPERTIES

For iid data, some theoretical properties of Silverman's test have been available; see Silverman (1983) and Mammen et al. (1992). We aim to show below that many of these properties hold for weakly dependent data. First, we study some theoretical properties of the smoothed autoregressive bootstrap, with the bandwidth parameter fixed at $h > 0$. We shall prove in the appendix that the smoothed autoregressive bootstrap defined by the transition density $\hat{f}_{Y|X}(\cdot|\cdot; h)$ is uniformly

ergodic. Hence it is asymptotically stationary and admits a unique (1-dimensional) marginal stationary density, denoted as $\pi_Y(\cdot;h)$. Moreover, $\hat{f}_Y(\cdot;h)$ can be shown to differ from $\pi_Y(\cdot;h)$ with a sufficiently fast decaying error rate so that the critical bandwidth as determined by $\hat{f}_Y(\cdot;h)$ and that by $\pi_Y(\cdot;h)$ share similar behaviour asymptotically.

Let $y = y_{p+1}, x = (y_1, \cdots, y_p)^T$ and $w = (y_2, \cdots, y_{p+1})$. It follows from the definition of $\hat{f}_{Y|X}(y|x;h)$ that

$$\begin{aligned}
\int_{-\infty}^{\infty} \hat{f}_{Y|X}(y_{p+1}|x;h)\hat{f}_X(x;h)dy_1 &= \int \hat{f}_Z(y_1,\ldots,y_{p+1};h)dy_1 \\
&= \int \sum_{t=p+1}^{n} g_1(y_1 - Y_{t-p};h)g_p(w - X_{t+1};h)dy_1/m \\
&= \sum_{t=p+1}^{n} g_p(w - X_{t+1};h)/m \\
&= \sum_{t=p+1}^{n} g_p(w - X_t;h)/m + [g_p(w - X_{n+1};h) - g_p(w - X_{p+1};h)]/m \\
&= \hat{f}_X(w;h) + [g_p(w - X_{n+1};h) - g_p(w - X_{p+1};h)]/m. \qquad (6)
\end{aligned}$$

This suggests that $\hat{f}_X(\cdot;h)$ equals the p-dimensional stationary pdf of $X_t^* = (Y_{t-p}^*, \cdots, Y_{t-1}^*)^T$ up to an additive error of $O_p(1/n)$, under the L^1-norm. Let $\hat{f}_Y^{(s)}(\cdot;h)$ and $\pi_Y^{(s)}(\cdot;h)$ be the sth derivatives of $\hat{f}_Y(\cdot;h)$ and $\pi_Y(\cdot;h)$ respectively. More importantly, it can be shown that for a fixed integer $s \geq 0$,

$$\sup_y |\hat{f}_Y^{(s)}(\cdot;h) - \pi_Y^{(s)}(y;h)| = O(h^{-s-1}n^{-1}). \qquad (7)$$

It may be argued that it is more relevant to determine the critical bandwidth as the smallest bandwidth for which $\pi_Y(\cdot;h)$ has m modes. Unfortunately, this is not computationally tractable because $\pi_Y(y;h)$ is typically unknown. However, (7) implies that the critical bandwidth determined by $\hat{f}_Y(\cdot;h)$ and that by $\pi_Y(\cdot;h)$ share similar asymptotic behaviour in the sense that Theorem 3.1 below holds if the critical bandwidth is determined as the smallest bandwidth for which $\pi_Y(\cdot\ h)$ has m modes. This justifies determining the bandwidth using $\hat{f}_Y(\cdot;h)$, which is computationally tractable. The key to lifting the theoretical properties of Silverman's test with iid data to weakly dependent stationary process is a result due to Neumann (1998): Under suitable conditions including fast decaying mixing rate and conditions on the conditional densities, a richer probability space can be constructed on which there exist a copy of $\{Y_t\}$ and a sequence of iid variables, say, $\{W_t\}$ such that both sequences have identical 1-dimensional marginal distributions, and the two sets

$\{Y_1, \cdots, Y_n\}$ and $\{W_1, \cdots, W_n\}$ overlap to a large extent; consequently, the (1-dimensional) kernel density estimators based on Y's are well approximated by those based on the W's. Specifically, assume the following two conditions.

(A1) The process $\{Y_t\}$ is absolutely regular with geometric decaying β-mixing coefficient, i.e., there exist positive constants C_1, C_2 such that $\beta(k) \le C_1 \exp(-C_2 k)$ where

$$\beta(k) = \sup_i E\{\sup_{V \in \mathcal{F}_{i+k}^n} [P(V|\mathcal{F}_1^i) - P(V)\};$$

$\mathcal{F}_i^j$ denotes the σ-algebra generated by $Y_i, Y_{i+1}, \cdots, Y_j$.

(A2) Let $f_{Y_i|\mathcal{F}_j^{i-1}}$ be the conditional pdf of Y_i given $Y_j, \cdots, Y_{i-1}$. The conditional densities are uniformly bounded, and uniformly Lipschitz in the sense that there exits a constant C_3 such that $\sup_i \sup_{F \in \mathcal{F}_1^{i-1}} \{|f_{Y_i|F}(v) - f_{Y_i|F}(v')|\} \le C_3 |v - v'|, \forall v, v'$.

Note that the preceding two assumptions imply the validity of (A1)-(A3) in Neumann (1998), owing to the fact that the Y's form a homogeneous pth order Markov process. Let $\hat{f}_Y^{(s)}(\cdot; h)$ denote the sth-derivative of the kernel density estimator based on $Y_1, \cdots Y_n$, and $\tilde{f}_Y^{(s)}(\cdot; h)$ denote that based on $W_1, \ldots, W_n$. Under (A1) and (A2), it holds that there exist constants $2 < \gamma < 2.5, C$ and for any arbitrary large but fixed constant λ, with probability exceeding $1 - n^{-\lambda}$,

$$|\hat{f}_Y^{(s)}(y; h) - \tilde{f}_Y^{(s)}(y; h)| \le C h^{-s} [n^{-\gamma/5} + (nh)^{-1}] \log(n). \tag{8}$$

This error rate decays sufficiently fast to allow the transfer of many theoretical results of Silverman's test with iid data to weakly dependent data.

Finally, we need a set of assumptions on $\pi_Y(\cdot)$, the true (1-dimensional) stationary density of $\{Y_t\}$, which are essentially identical to those assumed by Mammen et al. (1992).

(A3) The density π_Y is bounded with a bounded support $[a, b]$. It is twice continuously differentiable with a Lipschitz continuous second derivative; $\pi_Y^{(1)}(a+) > 0$ and $\pi_Y^{(1)}(b-) < 0$. The density has j local maxima $v_0 < v_2 < \cdots < v_{2j-2}$ and $j-1$ local minima $v_1 < v_3 < \cdots < v_{2j-3}$. Furthermore, $\pi_Y^{(2)}(y) \ne 0$ and $\pi_Y(y) > 0$ whenever $\pi_Y^{(1)}(y) = 0$.

Note that (A3) implies that the true process $\{Y_t\}$ is not a least favorable pth order Markov process consistent with the null hypothesis. Thanks to (8), we can now state the main results which are lifted from those derived by Silverman (1983) and Mammen et al. (1992) for iid data. (Recall that

the true number of modes in π_Y is j and the null hypothesis is $H_0 : j = m$ with the alternative $H_1 : j > m$; $\hat{h}$ is the critical bandwidth for which $\hat{f}_Y$ has m modes.)

Theorem 3.1 Assume that (A1-A3) hold.

(I) Under H_1, there exists a constant $K > 0$ dependent on π_Y and m such that $P(\hat{h} > K) \to 1$ as $n \to \infty$.

(II) Assume that $j \leq m$, then for $\alpha_n \to 0$ and $\beta_n \to \infty$, $P(\alpha_n n^{-1/5} \leq \hat{h} \leq \beta_n n^{-1/5}) \to 1$.

(III) Under H_0, $n^{1/5}\hat{h}$ converges in distribution to $V = \sup_p \frac{\pi_Y(v_p)}{|\pi_Y^{(2)}(v_p)|^{2/5}} V_p$ where the supremum is over all the local maxima and V's are iid distributed according to some universal distribution.

(IV) Under H_0, for $\alpha_n \to 0$ and $\beta_n \to \infty$, $P(\alpha_n n^{-1/5} \leq \hat{h}^* \leq \beta_n n^{-1/5} | Y_1, \cdots, Y_n) \to 1$, in probability. □

In particular, when $j \leq m$, the critical bandwidth is of the order $O_p(n^{-1/5})$. (IV) implies that under H_0, the probability of rejecting H_0 is asymptotically bounded away from 0 for the interesting case that the true marginal density is in the interior of H_0. Furthermore, as argued by Fisher et al. (1994), (III) suggests that Silverman's test is asymptotically conservative under the null hypothesis in that $P(\text{p-value} < \alpha) < \alpha$. Finally, we note that the issue of the optimal order of the bandwidth for kernel conditional density estimator is not well understood. Recent work by Bashtannyk and Hyndman (2001) considered the case of iid multivariate data, and found that for a bivariate normal reference distribution, the optimal bandwidth is $O_p(n^{-1/6})$ for conditional density estimators, which is larger than the $O_p(n^{-1/5})$ rate for the critical bandwidth.

4 SIMULATIONS AND EXAMPLES

Some aspects of the finite sample behaviour of the test proposed in the preceding section are now examined via Monte Carlo methods. We consider a first order threshold autoregressive models (Tong, 1990) of the following form:

$$Y_t = \begin{cases} -\Delta + \varepsilon_t & \text{if } Y_{t-1} \leq 0\ , \\ \Delta + \varepsilon_t & \text{otherwise.} \end{cases}$$

where the ε_ts are independent and identically normally distributed with zero mean and unit variance; furthermore, ε_t is independent of $Y_{t-1}, Y_{t-2}, \cdots$. It can be checked that the stationary marginal pdf of (Y_t) is that of an equally weighted mixture of $N(-\Delta, 1)$ and $N(\Delta, 1)$. Furthermore, it can

be easily verified that the density of $\frac{1}{2}N(-\Delta, 1) + \frac{1}{2}N(\Delta, 1)$ is bimodal if and only if $\Delta > 1$. The arguments of the proof of Theorem A.1 can be adapted to show that the above threshold model is uniformly ergodic and hence of short memory. Assumption (A3) is false because the stationary marginal pdf has infinite support, although this problem can be fixed by employing truncated normally distributed noise.

Stationary time series data of size 100 are simulated from the preceding TAR model with Δ ranging from 0.0–2.7 with increments equal to 0.3, augmented by the case $\Delta = 1.0$, the least favourable case. With each simulated time series, bootstrap p-values, based on 200 bootstraps, are computed for each of the nine hypotheses defined by two factors, namely, the hypothesized number of modes in the density taking values from 1 to 3, and the assumed Markov order of $\{Y_t\}$ being from 0 to 2. The experiment is replicated 200 times. Table 1 reports the empirical rejection rates of the tests with nominal 5% type I error rates. Recall that the density of Y_t is bimodal if and only if $\Delta > 1$. This explains the jump in the rejection rates of the hypothesis of unimodality of the density the moment the parameter Δ becomes larger than 1.

The upper panel of Table 1 is for the case when the Markov order is (incorrectly) set as 0, that is, the Silverman test assuming iid data. Note that when $\Delta < 1$ and hence when the null hypothesis of unimodality is true, the Silverman test tends to under-reject the hypothesis of unimodality compared with the 5% nominal level. Under-rejection remains for the case $\Delta = 1$, in which case (A3) is invalidated by the zero second derivative of the marginal density at the mode. On the contrary, while the extended Silverman test, with Markov order higher than 1, also exhibits under-rejection of unimodality under the null hypothesis of unimodal stationary density, its empirical rejection rate seems closer to 5% in the least favourable case when $\Delta = 1$. Recall that a test is said to be similar if the probability of type I error is constant under the null composite hypothesis. It is of interest to note that Lehmann and Stein (1948, Theorem 1) proved that in testing a composite null hypothesis against a simple alternative, most powerful tests may sometimes be obtained in the form of some likelihood-ratio test based on some least favourable distribution if the similarity requirement is replaced by the more general requirement that the probability of type I error be not more than the stated significance level of the test; indeed, for the likelihood-ratio test they devised, the probability of type I error is strictly less than the significance level if the true model under the null hypothesis is "distinct" from the least favourable distribution. The situation may be best illustrated with the following simple example: consider iid data $Y_1, \cdots, Y_n$ from $N(\mu, 1)$ and

testing $H_0 : \mu \leq 0$ vs the alternative $H_1 : \mu > 0$. We may calibrate the test statistic $T = \bar{X}$ via the least favourable distribution $\mu = 0$. In this case, the test is certainly conservative for $\mu < 0$ but it is uniformly most powerful. This example is perhaps too simple compared with the test of modes contemplated here as the latter has a much more complex set of least favourable distributions than the singleton $\mu = 0$ for the former problem. Nevertheless, the preceding simple example suggests that the conservative nature of the (extended) Silverman test under the null unimodal hypothesis in this TAR model need not be surprising nor undesirable, although the Silverman test is not a likelihood-ratio test. We note that the Silverman test is not always conservative as shown in the second example below, when the model is nearly non-stationary.

The power of the test for one mode versus two modes in the density appears to increase from $\Delta = 1$ to 2.1 and then decrease when Δ exceeds 2.1. The decrease in the power of the test owes to the fact that when Δ exceeds 2.1, the process tends to remain in one regime for an extended period as the noise, being $N(0, 1)$, exceeds 1.96 with no more than 2.5% probability. Hence, for small sample size, the process may disguise itself as unimodal with probability increasing with Δ. These simulation results suggest the following. (i) The Silverman's test and its extension to dependent data are generally conservative. (ii) The new tests are insensitive to the Markov order, and the Silverman test appears to be slightly more powerful than its extension. (iii) The tests are reasonably powerful for detecting the number of modes in the density.

In the second experiment, we consider the stationary first order autoregressive model with Gaussian noise sequence, that is, $Y_t = \phi Y_{t-1} + \varepsilon_t$. Six autoregressive parameter values, namely, $\phi = \pm 0.5, \pm 0.9$ and ± 0.95 are used in the experiment. Here, the true marginal densities are Gaussian. Therefore, we only consider the case of testing whether or not the density is unimodal, with the Markov order set as 0 or 1. The experimental setup is similar to the previous experiment, with sample size equal to 100, bootstrap p-values based on 200 bootstraps and replication size equal to 200. The empirical rejection rate of the nominal 5% tests are reported in Table 2. Note that the closer the autoregressive coefficient is to 1 in magnitude, the nearer the process is to non-stationarity. Results from this experiment reinforce the preceding conclusion that for weakly dependent data, the Silverman's test and its generalization developed here have similar empirical sizes and are generally conservative. The tests become liberal when the process is nearly non-stationary, whence (A1) is nearly violated. The size distortion problem seems less severe with the test based on the smoothed autoregressive bootstrap than with the Silverman's test, which is a

Table 1: Empirical rejection rates of the tests, with nominal size of 5 %, on the number of modes in the densities of time series data generated from a first order threshold autoregressive model. The stationary marginal density of the threshold model is bimodal if and only if $\Delta > 1$.

Markov order set to be 0											
mode	$\Delta = 0.0$	0.3	0.6	0.9	1	1.2	1.5	1.8	2.1	2.4	2.7
1	0.010	0.010	0.005	0.025	0.005	0.065	0.185	0.525	0.635	0.500	0.295
2	0.035	0.035	0.030	0.010	0.030	0.040	0.030	0.015	0.030	0.005	0.025
3	0.060	0.025	0.030	0.025	0.060	0.025	0.020	0.020	0.010	0.030	0.040
Markov order set to be 1											
mode	$\Delta = 0.0$	0.3	0.6	0.9	1	1.2	1.5	1.8	2.1	2.4	2.7
1	0.005	0.020	0.000	0.020	0.035	0.075	0.150	0.375	0.500	0.435	0.285
2	0.045	0.030	0.025	0.010	0.030	0.030	0.025	0.030	0.040	0.025	0.040
3	0.040	0.035	0.030	0.030	0.050	0.015	0.015	0.015	0.010	0.035	0.025
Markov order set to be 2											
mode	$\Delta = 0.0$	0.3	0.6	0.9	1	1.2	1.5	1.8	2.1	2.4	2.7
1	0.005	0.015	0.000	0.020	0.035	0.070	0.140	0.385	0.435	0.420	0.275
2	0.045	0.030	0.020	0.010	0.020	0.020	0.020	0.030	0.045	0.025	0.035
3	0.055	0.030	0.030	0.035	0.035	0.010	0.010	0.020	0.020	0.040	0.035

Table 2: Empirical rejection rates of the tests, with nominal size of 5 %, on the unimodality of the density of time series data generated from a first order autoregressive Gaussian model.

Markov order set to be 0						
mode	$\phi =$ -0.95	-0.90	-0.50	0.50	0.90	0.95
1	0.215	0.105	0.03	0.015	0.095	0.225
Markov order set to be 1						
mode	$\phi =$ -0.95	-0.90	-0.50	0.50	0.90	0.95
1	0.18	0.065	0.015	0.020	0.075	0.135

merit of the extended Silverman test. However, the case when $|\phi| = 0.95$ implies rather strong temporal dependence of the underlying process. Over the range from -0.9 to 0.9, the empirical sizes of the new tests based on the smoothed autoregressive bootstrap are comparable with the nominal 5%. Hence, we recommend the use of the new tests for testing for multi-modality with weakly dependent Markov data.

Next, we apply the test to two well-known time series data sets: the logarithmically transformed lynx data and the sunspot data; see Tong(1990) for the data listing and relevant background information. Previous studies (Cheng and Tong, 1992 and Yao and Tong, 1992) have suggested that the lynx data may be approximately an order two or three process, whereas the sunspot data may be adequately modelled as an order four process. Hence, we set the maximum order to be three and four respectively for the lynx data and the sunspot data. The test results are summarized in Table 3, based on which there is evidence that the density of the lynx data is bimodal, but the sunspot data appear to be unimodal. All bootstrap p-values reported in Tables 3 – 4 are based on 1,000 bootstraps.

Finally, we illustrate the test with the daily Hang Seng index from November, 24, 1969 to November, 28, 1997. This is a fairly long time series consisting of 6909 daily closing prices of the Hong Kong stock exchange market. The raw data show an exponential growth pattern. To induce approximate stationarity, we consider the first difference of the logarithmically transformed Hang Seng indices. See Figure 1 for the time series plot which displays three possible "outliers": the two largest crashes on the 26th of October, 1987 and the fifth of June, 1989, as well as the largest re-bounce on the 29th of October, 1997. As it is unlikely that the transformed Hang Seng series is stationary over such a long time span, the series is divided into five periods for

Table 3: Illustration of the tests for multi-modality with the log lynx data and the sunspot data.

log lynx data						
mode	$\hat{h}$	p-value (order=0)	p-value (order=1)	p-value (order=2)	p-value (order=3)	
1	0.684	0.014	0.017	0.015	0.014	
2	0.335	0.429	0.437	0.469	0.435	
3	0.203	0.788	0.775	0.790	0.794	
4	0.179	0.637	0.663	0.638	0.645	
sunspot data						
mode	$\hat{h}$	p-value (order=0)	p-value (order=1)	p-value (order=2)	p-value (order=3)	p-value (order=4)
1	11.9	0.377	0.330	0.285	0.311	0.288
2	8.33	0.710	0.649	0.647	0.606	0.622
3	8.14	0.371	0.296	0.261	0.238	0.270
4	7.00	0.375	0.306	0.341	0.309	0.318

further study; see Table 4. We have employed the Cheng-Tong (Cheng and Tong, 1992) order determination method which chooses the order which minimizes the non-parametric cross validatory sum of squared first step prediction errors. According to this criterion, over the whole time span, the first differences of the log Hang Seng data may be an order 1 process, with order zero being extremely competitive. This conclusion is consistent with the view that should the Hong Kong stock market be fully efficient, the log-transformed Hang Seng indices should behave as a random walk. Table 4 summarizes the test results on the number of modes in the density of the differences of the log-transformed Hang Seng index, over the five time periods. Interestingly, the tests suggest that over the first three periods, the density of the transformed Hang Seng indices is unimodal, but it may be tri-modal over the last two periods. Notice that the three "outliers" occur over the last two periods. We subsequently deleted these outliers from the data, and repeated the analysis. Without the largest crashes in the fourth period, the density now appears to be unimodal. When the highest re-bounce in the final period is removed, the density also appears to be unimodal. This example is instructive in demonstrating that the Silverman's test and its extension developed here are sensitive to outliers. Therefore we suggest that it may be worthwhile exploring the potential of these tests for outlier detection.

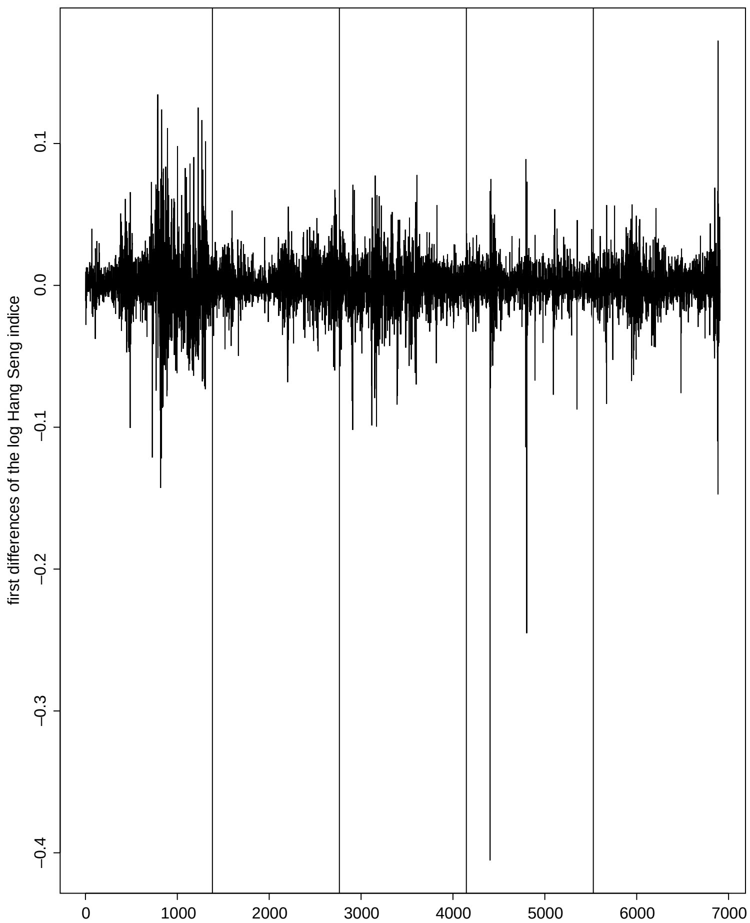

Figure 1: Time series plot of the first differences of the log Hang Seng indices. Artificial vertical lines have been super-imposed on the diagram to indicate the five sub-periods over which the densities are studied.

5 CONCLUSION

Because of (8), kernel density estimation of the marginal pdf is robust to weak serial dependence in data. Consequently, despite having incorporated the serial dependence structure, the extended Silverman test need not yield higher power with weakly dependent data than the Silverman test; indeed the robustness of the kernel marginal density estimation coupled with the additional source of randomness due to Markov transitions may cause the slight reduction in the power of the extended Silverman test, as reported in Table 1. However, (8) fails with strongly dependent data, and Table 2 suggests that for such data, the extended Silverman test may have lesser size distortion problem than the Silverman test. Thus, it is of interest to explore alternative approaches of testing the number of modes that may be rendered more powerful by properly incorporating the serial dependence structure. The smoothed autoregressive bootstrap ordinarily generates a uniformly ergodic Markov chain, which implies that it may not be appropriate for data with long memory. A second problem is to devise a test of the number of modes in the density with long-memory data. The number of modes in the density is often determined via a sequence of hypothesis testing. This approach requires correction to ensure a fixed overall error rate. It is interesting to explore alternative approaches to the sequential testing approach. Another problem is to assess the impact of order determination on the extended Silverman test. Finally, following Cheng and Hall (1998), it is of interest to develop other approaches for calibrating the Silverman test for size and power improvement.

6 ACKNOWLEDGEMENTS

K.S. Chan was partially supported by the NSF grant DMS 9504798 and a faculty scholar award from the University of Iowa. H. Tong was partially supported by the Engineering and Physical Science Research Council, UK (grant GR/L16385) and the European Community Human Mobility (grant ERB CHRX-CT-94-0693). We thank Kai W. Ng for providing us the Hang Seng index data. We also thank the Editor and an anonymous referee for helpful comments.

Table 4: Tests for multi-modality for the first differences of the log-transformed Hang Seng daily indices for 5 periods from November, 24, 1969 to November, 28, 1997.

order=0										
	1st period: 1–1382		2nd period: 1383–2764		3rd period: 2765–4146		4th period: 4147–5528		5th period: 5529–6908	
mode	$\hat{h}$	p-value	$\hat{h}$	p-value	$\hat{h}$	p-value	$\hat{h}$	p-value	$\hat{h}$	p-value
1	.01123	.569	.00522	.422	.00798	.398	.07257	.000	.03224	.000
2	.01066	.154	.00395	.453	.00631	.269	.04526	.000	.01738	.031
3	.00805	.166	.00354	.279	.00413	.827	.01222	.130	.01117	.045
4	.00569	.516	.00308	.226	.00399	.546	.00827	.266	.00449	.531
5	.00456	.727	.00273	.167	.00394	.217	.00723	.172	.00416	.291
order=1										
	1st period: 1-1382		2nd period: 1383-2764		3rd period: 2765-4146		4th period: 4147-5528		5th period: 5529-6908	
mode	$\hat{h}$	p-value	$\hat{h}$	p-value	$\hat{h}$	p-value	$\hat{h}$	p-value	$\hat{h}$	p-value
1	.01123	.528	.00522	.386	.00798	.437	.07257	.000	.03224	.000
2	.01066	.150	.00395	.451	.00631	.289	.04526	.000	.01738	.069
3	.00805	.181	.00354	.268	.00413	.813	.01222	.157	.01117	.041
4	.00569	.513	.00308	.234	.00399	.527	.00827	.318	.00449	.517
5	.00456	.736	.00273	.159	.00394	.212	.00723	.173	.00416	.365
with three outliers deleted from the data										
	4th period, order=0		5th period, order=0				4th period, order=1		5th period, order=1	
mode	$\hat{h}$	p-value	$\hat{h}$	p-value			$\hat{h}$	p-value	$\hat{h}$	p-value
1	.01228	.204	.01739	.070			0.01228	0.191	0.01739	0.058
2	.00824	.431	.01112	.047			0.00824	0.625	0.01112	0.043
3	.00725	.265	.00445	.661			0.00725	0.427	0.00445	0.694
4	.00647	.106	.00413	.439			0.00647	0.167	0.00413	0.469
5	.00564	.052	.00394	.185			0.00564	0.069	0.00394	0.228

References

[1] Bashtannyk D. M. and Hyndman R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, **36**, 279-298.

[2] Chan, K. S. (1993). A review of some limit theorems of Markov chains & their applications. In " Dimension, Estimation and Models" edited by H. Tong, 108-135. *World Scientific.*

[3] Cheng, B. and Tong, H. (1992). Consistent non-parametric order determination and chaos—with discussion. *J. R. Statist. Soc., Series B* **54**, 427-49 and 451-74.

[4] Cheng, M. and Hall, P. (1998). Calibrating the excess mass and dip tests of modality. Journal of the Royal Statistical Society, Series B, Methodological, **60**, 579-589.

[5] Efron, B. and Tibshirani R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

[6] Fisher, N. I., Mammen, E. and Marron, J. S. (1994). Testing for multi-modality. *Computational Statistics and Data Analysis*, **18**, 499-512.

[7] Good, I. J. and Gaskins, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data (C/R: p56-73). *Journal of the American Statistical Association*, **75**, 42-56.

[8] Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, **13**, 70-84.

[9] Jalali, A. and Pemberton, J. (1995). Mixture models for time series. *Journal of Applied Probability*, **32**, 123-138.

[10] Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annal of Statistics*, **17**, 1217-1241.

[11] Lall, U. and Sharma, A. (1996). A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series. *Water Resources Research*, **32**, 3, 679-93.

[12] Lehmann, E. L. and Stein C. (1948). Most powerful tests of composite hypothesis I. Normal distributions. *Annals of Mathematical Statistics*, **19**, 495-516.

[13] Mammen, E., Marron, J. S. and Fisher, N. I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probability Theory and Related Fields,* **91**, 115-132.

[14] Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability.* London: Springer Verlag.

[15] Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, **86**, 738-746.

[16] Neumann M. H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Annals of Statistics*, **26**, 2014-2048.

[17] Nummelin, E.(1984). *General irreducible Markov chains and non-negative operators.* Cambridge: Cambridge University Press.

[18] Politis, D. N. and Romano, J. P. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, **89**, 428, 1303-1313.

[19] Rajarshi, M. B. (1990). Bootstrap in Markov-sequences based on estimates of transition density. *Annals of the Institute of Statistical Mathematics*, **40**, 565-586.

[20] Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B*, **43**, 97-99.

[21] Silverman, B. W. (1983). Some properties of a test for multimodality based on kernel density estimates. *Probability, Statistics and Analysis*, (Kingman, J. F. C. and Reuter, G. E. H., eds.) 248-259. Cambridge: Cambridge University Press.

[22] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*, Wiley.

[23] Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability*, **22**, 587-611.

[24] Tong, H. (1988). Contribution to the discussion of the paper "Bootstrap methods" by D. V. Hinkley, *J. Roy. Statist. Soc., B*, **50**, 359-360.

[25] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

A APPENDIX

Henceforth, we adopt the convention that K denotes a constant that may vary at each of its occurrences. Let $\{Y_t^*\}$ be a pth order Markov process and $X_t^* = (Y_{t-p}^*, \cdots, Y_{t-1}^*)^T$, with the transition pdf given by $\hat{f}_{Y|X}(y|x;h)$ with a fixed bandwidth $h > 0$. We first derive some pertinent probabilistic properties of the Markov chain $\{X_t^*\}$. We will show below that $\{X_t^*\}$ has a unique stationary probability measure. As shown below, for $t \geq p$, the tth step transition probability measure of $\{X_t^*\}$ admits a pdf, hence the stationary probability measure of $\{X_t^*\}$ is absolutely continuous and whose pdf is denoted as $\pi_X(\cdot;h)$. By an abuse of notation, $\pi_X(\cdot;h)$ will also denote the stationary probability measure.

Theorem A.1 $\{X_t^*\}$ is an irreducible and aperiodic Markov chain. Moreover, it is uniformly ergodic: there exist two positive constants K and $\rho < 1$ such that the tth step transition probability measure $P^t(x,\cdot)$ converges geometrically fast to $\pi_X(\cdot;h)$. Specifically,

$$\|P^t(x,\cdot) - \pi_X(\cdot;h)\|_\tau \leq K\rho^t, \tag{A1}$$

where $\|\cdot\|_\tau$ denotes the total variation norm of the enclosed signed measure, that is, for any two probability measures ν and η, $\|\nu - \eta\|_\tau = 2\sup_A |\nu(A) - \eta(A)|$. □

Uniform ergodicity implies that $\{X_t^*\}$ is strongly mixing with an exponential decaying mixing rate. In particular, under suitable moment conditions, the law of large numbers and the central limit theorem holds for partial sums of functionals of $\{X_t^*\}$; see Nummelin (1984) and Meyn and Tweedie (1993). Thus, the empirical distribution of $\{X_1^*, \ldots, X_n^*\}$ will closely resemble that from independent resampling from the marginal distribution $\hat{f}_Y(\cdot)$ and hence the robustness of the test for multi-modality with respect to the order of the Markov process within the framework of smoothed autoregressive bootstrap. Finally, we note that Theorem A.1 holds if the Gaussian kernel is replaced by some kernel that is positive everywhere.

Proof of Theorem A.1

We shall make heavy use of the Markov chain techniques in the following proof. For a review of Markov chain techniques with applications to nonlinear time series, see Tong (1990) or

Chan (1993) which the reader may consult for unexplained notions and proofs or references of the results mentioned here. Let $\chi(y) = \min_{t=p+1,\ldots,n} g(y - Y_t; h)$. Then, χ is positive everywhere. Note that $\forall y, x, \hat{f}_{Y|X}(y|x;h) \geq \min_{t=p+1,\ldots,n} g_1(y - Y_t; h) = \chi(y) > 0$. Let $x_1 = (y_1, \ldots, y_p)^T$ and $x_2 = (y_{p+1}, \ldots, y_{2p})^T$. The pth step transitional pdf of $\{X_t^*\}$ equals

$$\begin{aligned} p(X_{2p}^* = x_2 | X_p^* = x_1) &= \hat{f}_{Y|X}(y_{2p}|y_p, \ldots, y_{2p-1}; h) \times \ldots \times \hat{f}_{Y|X}(y_{p+1}|y_1, \ldots, y_p; h) \quad \text{(A2)} \\ &\geq \prod_{j=p+1}^{2p} \chi(y_j) > 0. \quad \text{(A3)} \end{aligned}$$

Consequently, the mth step transitional pdf of $\{X_t^*\}$ is positive everywhere, for all $m \geq p$. Hence, $\{X_t^*\}$ is irreducible and aperiodic.

To verify that $\{X_t^*\}$ is uniformly ergodic, it follows from Theorem 6.1.5 in Nummelin (1984) that it suffices to show that the state space R^p is a small set, which is the case if the pth step transition pdf is uniformly bounded below by an integrable non-negative function. However, (A3) shows that the latter holds, and hence this completes the proof.

Proof of (7):

Inequality (A1) implies that if the initial pdf of X_{p+1}^* equals $\hat{f}_X(\cdot; h)$, the L^1-distance between the marginal pdf of X_t^*, denoted as $\hat{f}_X(\cdot; t, h)$ and $\pi_X(\cdot; h)$ is bounded by $K\rho^t$. But this implies that the marginal pdf of Y_t^*, denoted by $\hat{f}_Y(\cdot; t, h)$ differ from $\pi_Y(\cdot; h)$ by less than $K\rho^t$ in sup-norm because

$$\begin{aligned} |f_Y(y_t; t, h) - \pi_Y(y_t; h)| &\leq \int \hat{f}_{Y|X}(y_t|x_t) |\hat{f}_X(x_t; t, h) - \pi_X(x_t; h)| dx_t \\ &\leq Kh^{-1} \int |\hat{f}_X(x_t; t, h) - \pi_X(x_t; h)| dx_t \\ &\leq Kh^{-1}\rho^t, \quad \text{(A4)} \end{aligned}$$

because $\hat{f}_{Y|X}(y_t|x_t)$ is uniformly bounded by Kh^{-1}, for Gaussian kernel.

Now, (6) implies that for $\int |\hat{f}_X(x; p+2, h) - \hat{f}_X(x; p+1, h)| dx \leq K/n$. Consequently, for $t > p$, $\int |\hat{f}_X(x; t+1, h) - \hat{f}_X(x; t, h)| dx \leq K/n$ because, with $w = (y_1, \cdots, y_p)^T$ and $x = (y_2, \cdots, y_{p+1})^T$,

$$\begin{aligned} \int |\hat{f}_X(x; t+1, h) - \hat{f}_X(x; t, h)| dx &\leq \int \hat{f}_{Y|X}(y_{p+1}|w) |\hat{f}_X(w; t, h) - \hat{f}_X(w; t-1, h)| dy_{p+1} dw \\ &\leq \int |\hat{f}_X(w, t, h) - \hat{f}_X(w; t-1, h)| dw. \end{aligned}$$

In turn, this implies that $\hat{f}_Y(y; t+1, h) - \hat{f}_Y(y; t; h)$ is uniformly bounded by $K(hn)^{-1}$, owing to the inequality that

$$|\hat{f}_Y(y; t+1, h) - \hat{f}_Y(y; t, h)| \leq \int \hat{f}_{Y|X}(y|x; h) |\hat{f}_X(x; t, h) - \hat{f}_X(x; t-1, h)| dx$$

$$\begin{aligned} &\leq \; Kh^{-1}\int |\hat{f}_X(x,t,h)-\hat{f}_X(x;t-1,h)|dx \\ &\leq \; K(nh)^{-1}. \end{aligned} \tag{A5}$$

Hence, $|\hat{f}_Y(y;h) - \pi_Y(y;h)| \leq \sum_{t=p+1}^{k-1} |\hat{f}_Y(y;t,h) - \hat{f}_Y(y;t+1,h)| + |\hat{f}_Y(y;k,h) - \pi_Y(y;h)| \leq Kh^{-1}(k/n + \rho^k)$, owing to (A4) and (A5). Letting $0 < \gamma < 1$ and $k = n^{1-\gamma}$, we have the desirable inequality $\sup_{y\in R} |\hat{f}_Y(y;h) - \pi_Y(y;h)| \leq Kh^{-1}n^{-\gamma}$. Similarly, it can be shown that $\hat{f}_Y^{(s)}(\cdot;h)$ differ from $\pi_Y^{(s)}(\cdot;h)$ by an error of order $O(h^{-(s+1)}n^{-\gamma})$, under the sup-norm.

Proof of (8):

Proposition 2.1 of Neumann (1998) states that with probability exceeding $1-O(n^{-\lambda})$, the cardinality of the intersection of an interval $[a,b]$ and the symmetric difference of the sets $\{Y_1,\cdots,Y_n\}$ and $\{W_1,\cdots,W_n\}$ is bounded by $O(\{[n^{1/2}(b-a)+1]\log(n)\})$, uniformly for all intervals with $b - a = O(n^C)$, for some constant $C > 0$. Based on this bound, Neumann (1998) gave a bound, that is sharper than (8), on the difference between $\hat{f}_Y(\cdot;h)$ and $\tilde{f}_Y(\cdot,h)$ for kernels with a bounded support. Here, we adapt Neumann's argument to Gaussian kernels. Let $0.5 > \epsilon > 0$ be a constant, and $I(\cdot)$ the indicator function. Consider the decomposition $\hat{f}_Y(y;h) = (mh)^{-1}\sum\psi[(y - Y_t)/h]I(|y-Y_t|/h < n^\epsilon) + (mh)^{-1}\sum\psi[(y-Y_t)/h]I(|y-Y_t|/h \geq n^\epsilon) = \hat{f}_{1,Y}(y;h) + \hat{f}_{2,Y}(y;h)$. Similarly, we can decompose $\tilde{f}_Y = \tilde{f}_{1,Y} + \tilde{f}_{2,Y}$. Now, with probability exceeding $1 - O(n^{-\lambda})$,

$$\begin{aligned} |\hat{f}_Y(y;h) - \tilde{f}_Y(y;h)| \;&\leq\; |\hat{f}_{1,Y}(y;h) - \tilde{f}_{1,Y}(y;h)| + |\hat{f}_{2,Y}(y;h) - \tilde{f}_{2,Y}(y;h)| \\ &\leq\; K[(nh)^{-1}(n^{1/2}n^\epsilon h + 1)\log(n) + h^{-1}\exp(-n^{2\epsilon}/2)], \end{aligned} \tag{A6}$$

owing to the preceding bound due to Neumann and the fact that $\psi[(y-Y_{t+1})/h]I(|y-Y_{t+1}|/h \geq n^\epsilon) \leq h^{-1}\exp(-n^{2\epsilon}/2)/\sqrt{2\pi}$, which implies (8) for the case $s = 0$. The proof for $s > 0$ is similar and hence omitted.

Proof of Theorem 3.1: The proof of (I) is similar to that of proposition 3 in Silverman (1983), thanks to (8). The proof of (II) can be adapted from that of Corollary 2.1 in Mammen et al. (1992) by noting that owing to (8), lemmas 4, 5 and 6 there continue to hold, with the modification of (4.2) in Mammen et al. (1992) to

$$B_s = \{\sup_t |X_n^{(s)}(t) - Y_n^{(s)}(t)| \geq c_1 \log(n) n^{-(\gamma-s)/5}\},$$

where $2 < \gamma < 2.5$ is the same constant in (8). The proof of lemma 2 requires some changes. Using the notations in that proof, we need to set a_n to be $c\log(n)n^{-(\gamma-2)/5}$. Then, arguing as

in the proof of lemma 2 in Mammen et al. (1992), it can be verified that with probability $>1 - O(\log^2(n) n^{-(\gamma-2)/5})$, if $|\tilde{Y}_n'(t)| \leq c\log(n) n^{-(\gamma-1)/5}$ or $|\hat{f}_h'(t)| < (c+c_1)\log(n) n^{-(\gamma-1)/5}$, then $|\tilde{Y}_n''(t)| > c\log(n) n^{-(\gamma-2)/5}$ or $|\hat{f}_h''(t)| < (c-c_1)\log(n) n^{-(\gamma-2)/5}$, which holds for all $t \in U_n$. In other words, $\tilde{Y}_n'$ and $\hat{f}_h'(t)$ are monotone if $|\tilde{Y}_n''(t)|$ or $|\hat{f}_h''(t)|$ are small. Then the rest of the proof is similar. These lemmas then imply the validity of (II). The proofs of (III) and (IV) are similar to those of Theorems 4 and 5 in Mammen et al. (1992) and hence omitted. Finally, we note that were the critical bandwidth be determined by the smallest bandwidth for which $\pi_Y(\cdot; h)$ has m modes, (I)-(IV) continue to hold owing to (7) and the fact that $\{X_t^*\}$ is uniformly ergodic. This observation makes formal the claim that the two methods of determining the critical bandwidth share identical behaviour asymptotically.