Bernoulli **4**(2), 1998, 185–201

Discretely observing a white noise change-point model in the presence of blur

JACQUES ISTAS[1]* and HENRIK STRYHN[2]

[1]*I.N.R.A., Laboratoire de Biométrie, Domaine de Vilvert, F-78352 Jouy-en-Josas Cedex, France*
[2]*Department of Mathematics and Physics, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Copenhagen, Denmark*

In discretely observed diffusion models, inference about unknown parameters in a smooth drift function has attracted much interest of late. This paper deals with a diffusion-type change-point model where the drift has a discontinuity across the point of change, analysed in detail in continuous time by Ibragimov and Hasminskii. We consider discrete versions of this model with integrated or blurred observations at a regular lattice. Asymptotic convergence rates and limiting distributions are given for the maximum likelihood change-point estimator when the observation noise and the lattice spacing simultaneously decrease. In particular, it is shown that the continuous and discrete model convergence rates are generally equal only up to a constant; under specific conditions on the blurring function this constant equals unity, and the normalized difference between the estimators tends to zero in the limit.

Keywords: asymptotic distribution; blur; maximum likelihood; orthogonalization

1. Introduction

The continuous-time, Gaussian white noise change-point model in $T = [0, 1]$ is

$$\mathrm{d}V_t = f(t, \theta)\,\mathrm{d}t + \varepsilon \mathrm{d}W_t, \qquad t \in T, \tag{1}$$

where (W_t) is a standard Brownian motion, the drift has the form

$$f(t, \theta) = f_1(t)\mathbf{1}\{t \leqslant \theta\} + f_2(t)\mathbf{1}\{t > \theta\}, \tag{2}$$

and f_1, f_2 are known functions on T, with the change-point parameter $\theta \in T$ and the true change-point $\theta^* \in (0, 1)$. The drift $f(t, \theta)$ is discontinuous at $t = \theta^*$, that is, $\Delta \doteq f_1(\theta^*) - f_2(\theta^*) \neq 0$. The maximum likelihood (ML) estimator of θ has been studied by, among others, Ibragimov and Hasminskii (1981), Kutoyants (1984) and Korostelev (1987). Recently generalizations of (1) to two-dimensional time have attracted interest in the context of image analysis; see Hasminskii and Lebedev (1990), Rudemo and Stryhn (1994) and Stryhn (1994).

*To whom correspondence should be addressed.

1350–7265 © 1998 Chapman & Hall

Discrete observations of (1) are taken at equidistant lattice points $s \in S_n \subset T$, where n denotes the number of observations. The value $\bar{V}_s$ at s is a weighted integral of continuous-time observations in a neighbourhood of s,

$$\bar{V}_s = \int_T \varphi(t-s)\,\mathrm{d}V_t = \int_T \varphi(t-s) f(t, \theta)\,\mathrm{d}t + \varepsilon \int_T \varphi(t-s)\,\mathrm{d}W_t, \tag{3}$$

where $\varphi = \varphi_h$ (h is bandwidth parameter) is a kernel function satisfying $\int \varphi(u)\,\mathrm{d}u = 1$. (Some technical details of the model are deferred to Section 3.) The particular choice $\varphi(u) = n\mathbf{1}\{u \in [-1/(2n), 1/(2n)]\}$ leads to completely integrated observations of (1), or in effect to increments $V_{k/n} - V_{(k-1)/n}$ of (V_t). Our motivation for the blurring function φ stems from applications in signal or image processing where a continuous signal or scene is digitized on a regular grid by some recording device. The value produced at each recording point can intuitively be thought of as a weighted average of the signal intensity in the vicinity of that point. This formalism agrees with a standard way of modelling image blur by convolution with a point spread function; see, for example, Rosenfeld and Kak (1982, Chapter 7). We take φ to have bounded support – in view of the bounded observation interval T this seems most natural.

In model (1) we study asymptotics when $\varepsilon \to 0$. Two variants, Δ fixed or $\Delta \to 0$ in the limit, are considered; the latter corresponds to a low signal-to-noise ratio setting and is termed here decreasing jump-size asymptotics. In model (3) we let simultaneously $n \to \infty$, $\varepsilon \to 0$ and $h \to 0$ under suitable links between the rates. The limiting distributions are expressed in terms of functionals of a Brownian motion with triangular drift. That is, we consider the process $(B_u)_{u \in \mathbb{R}}$ given by $B_u = -|u|/2 + W_u$ where (W_u) is a two-sided Brownian motion. The argmax of (B_u) is well defined and its distribution, denoted here by F_{tri}, has been given by, among others, Bhattacharya and Brockwell (1976).

Denote by $\hat{\theta}_\varepsilon$ and $\bar{\theta}_\varepsilon$ the ML change-point estimator based on the continuous and discrete models (1) and (3), respectively. This paper gives limiting results of the following type:

$$\psi_\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta^*) \xrightarrow{\mathscr{L}} F_{\mathrm{tri}},\ \bar{\psi}_\varepsilon^{-1}(\bar{\theta}_\varepsilon - \theta^*) \xrightarrow{\mathscr{L}} F_{\mathrm{tri}} \qquad \text{as } \varepsilon \to 0, \tag{4}$$

for suitable convergence rates ψ_ε and $\bar{\psi}_\varepsilon$. Note that the discrete model estimator and rate depend on n and h as well as on ε, even if suppressed in the notation. Conditions under which the rates coincide are established, and in such cases it is furthermore shown that

$$\psi_\varepsilon^{-1}(\hat{\theta}_\varepsilon - \bar{\theta}_\varepsilon) \xrightarrow{\mathbb{P}} 0 \qquad \text{as } \varepsilon \to 0. \tag{5}$$

Inference about drift parameters in discretely observed diffusion models has been much studied of late for the case where the drift f is a smooth function (differentiable of some order) of t, θ and V_t; see, for instance, Laredo (1990), Genon-Catalot and Jacod (1993), Bibby and Sørensen (1995), and Pedersen (1995). To our knowledge no previous studies have been undertaken for discontinuous drift, not even in the present simple case where f does not depend on V_t. For smooth f, results like (4) and (5) are shown in Laredo (1990), within a more general framework of asymptotic sufficiency of the discrete observation model. In comparison, the discontinuous problem has faster convergence rate, of order ε^2 instead of ε, and a non-Gaussian limit F_{tri}.

The model (3) differs from the usual discrete change-point models by the integral of $f(t, \theta)$ across the change point for observations close to θ^*. For example, if $f_i \equiv \mu_i$, $i = 1, 2$, integrated observations $\bar{V}_s$ are i.i.d. $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ to the left and right of θ^* respectively, except for the $s \in S_n$ with $s - \theta^*$ inside the range of φ; in this case the mean value is a weighted average of μ_1 and μ_2. Moreover, the noise σ^2 will usually be held constant in asymptotical analysis, but our results correspond in fact to increasing variance of individual observations $\bar{V}_s$. This explains why asymptotic limits in terms of F_{tri} are feasible here not only for decreasing jump-size asymptotics (cf. Yao 1987) but also for fixed jump-size asymptotics, contrasting the standard i.i.d. change-point problem where lattice effects dominate in fixed asymptotics; see, for instance, Hinkley (1970).

This paper is organized as follows. All proofs are gathered together in the concluding Section 5. To begin, we briefly review the continuous model and state the main local assumptions on f_1, f_2 used throughout. Section 3 presents the discrete model in full detail and gives the asymptotic results of the ML change-point estimator. In Section 4 we summarize the relation between continuous and discrete model estimators and state the conditions under which the ML estimators are equivalent in the sense of (5). Also, extension of the treatment to Bayesian estimators is briefly discussed.

2. Continuous model

The model (1) and the likelihood function are well defined under the weak assumption that f_1, f_2 are L_2-integrable over T. For the purposes of analysis we make the following further assumptions:

(i) $\Delta = f_1(\theta^*) - f_2(\theta^*) \neq 0$.

(ii) In a fixed neighbourhood N^* of θ^* the difference $\Delta_f(t) \doteq f_1(t) - f_2(t)$ satisfies, for fixed (F) and decreasing (D) jump-size asymptotics respectively,

$$\text{F:}\quad \Delta_f \text{ is continuous,}$$

$$\text{D:}\quad \Delta_f \text{ is } C^1 \text{ with } \Delta^* = \sup_{t \in N^*} |\Delta_f'(t)| \, (<\infty).$$

(iii) $\forall t \in T \, \forall h > 0$: $\displaystyle\int_{(t, t+h] \cap T} \Delta_f^2(u)\, \mathrm{d}u > 0$.

Note that in decreasing jump-size asymptotics not only Δ but the entire function $\Delta_f(\cdot)$ is subject to change in the limit. However, for simplicity we suppress the dependence on ε of Δ_f in our notation.

Condition (i) is essential for the statistical problem. Intuitively, the reason for the stronger smoothness assumption in (ii) for decreasing jump-size asymptotics is that we need the error in approximating Δ_f locally by Δ to be negligible relative to Δ; the condition is formulated throughout in terms of the first-order derivative Δ^*. Finally, (iii) is an identifiability condition ensuring uniqueness of the ML estimator, included here mostly for convenience.

Theorem A. *For model* (1), *define the rate of convergence*

$$\psi_\varepsilon = \varepsilon^2/\Delta^2,$$

and assume the above conditions (i)–(iii) as well as the following relations for fixed (F) *and decreasing* (D) *jump-size asymptotics, respectively:*

$$\text{F:}\quad \varepsilon \to 0,\ \Delta \neq 0, \text{ and D:}\quad \varepsilon \to 0,\ \Delta \to 0,\ \varepsilon/\Delta \to 0,\ \varepsilon^2\Delta^*/\Delta^3 \to 0.$$

Then we have that

$$\psi_\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta^*) \xrightarrow{\mathscr{L}} F_{\mathrm{tri}} \text{ as } \varepsilon \to 0.$$

Remark. The result is well known for fixed asymptotics (cf. Ibragimov and Hasminskii 1981, Section VII.2; or Kutoyants 1984, Section 3.5); see also our proof section. Recall that F_{tri} is the argmax distribution of the two-sided Brownian motion with triangular drift $(B_u)_{u\in\mathbb{R}}$.

3. Discrete model

We supply some details of the discrete model (3). The observation lattice S_n consists of n equidistant and symmetrically positioned points in T,

$$S_n = \left\{\frac{1/2}{n}, \frac{3/2}{n}, \ldots, \frac{n-1/2}{n}\right\}.$$

Let $\varphi \in L^2(\mathbb{R})$ be a function (kernel) with bounded support satisfying $\int_{\mathbb{R}}\varphi(u)\,\mathrm{d}u = 1$, and introduce the bandwidth parameter h by defining $\varphi_h(u) = h^{-1}\varphi(u/h)$. Without loss of generality, we take $\operatorname{supp}(\varphi) \subseteq [-\frac{1}{2}, \frac{1}{2}]$. For $s \in S_{n,h} = \{s \in S_n\colon s + h\operatorname{supp}(\varphi) \subseteq T\}$ the discrete model takes the form

$$\bar{V}_s = \int_T \varphi_h(t-s)\,\mathrm{d}V_t = \int_T \varphi_h(t-s) f(t,\theta)\,\mathrm{d}t + \varepsilon\int_T \varphi_h(t-s)\,\mathrm{d}W_t. \tag{6}$$

By the definition of $S_{n,h}$ the intervals in (6) do not exceed T. For incompletely recorded points in $S_n \backslash S_{n,h}$ some modification of the model to be described below, is desirable. Note that when $nh \leqslant 1$ we have $S_{n,h} = S_n$; furthermore, the supports of $\varphi_h(\cdot - s)$ are disjoint so that the model (6) has in fact independent observations. We calculate for $s, s' \in S_{n,h}$ with $s = (k-\frac{1}{2})/n$ and $s' = (k'-\frac{1}{2})/n$,

$$\mathrm{E}\bar{V}_s = \int_T \varphi_h(t-s) f(t,\theta)\,\mathrm{d}t \doteq \mu_s^\theta, \tag{7}$$

$$\operatorname{var}(\bar{V}_s) = (\varepsilon^2/h)\int_{\mathbb{R}} \varphi^2(u)\,\mathrm{d}u \doteq \sigma^2, \tag{8}$$

$$\operatorname{cov}(\bar{V}_s, \bar{V}_{s'}) = (\varepsilon^2/h)\int_{\mathbb{R}} \varphi(u)\varphi(u + (s - s')/h)\,\mathrm{d}u$$

$$= (\varepsilon^2/h)\int_{\mathbb{R}} \varphi(u)\varphi(u + \lambda(k - k'))\,\mathrm{d}u, \tag{9}$$

with

$$nh = \lambda^{-1} \text{ or } \lambda = (nh)^{-1}, \qquad 0 < \lambda < \infty. \tag{10}$$

These formulae show that the covariance structure of (6) is stationary on $S_{n,h}$ and depends only on (n, h) through λ. Two conclusions are drawn.

First, we choose to keep λ *fixed* in the asymptotic limit, thereby not altering the range of dependence in $(\bar{V}_s)$. This seems most natural for a study of the model, although the cases $nh \to 0$ or $nh \to \infty$ may also be of interest.

Second, we are led to consider extensions of (6) to S_n which maintain the stationary covariance. Clearly, one may simply discard the points outside $S_{n,h}$ or extend the continuous model at both end-points by some small intervals to achieve fully recorded observations in S_n. Note that by (10) the number of points in $S_n \backslash S_{n,h}$ is bounded in the limit. From a statistical point of view we find an (*ad hoc*) selection of data points somewhat unsatisfactory. We take instead the well-known solution from similar settings (for example, Ibragimov and Hasminskii 1981) to extend the models (1) and (6) periodically around T. That is, we let $\breve{W}_t = W_{t-[t]}$, $f(t, \theta) = f(t - [t], \theta)$, and $\breve{V}_t = V_{t-[t]}$, for $t \in \mathbb{R}$, where $[t]$ denotes the integer part of t, and define for $s \in S_n$,

$$\bar{V}_s = \int_{\mathbb{R}} \varphi_h(t - s)\,\mathrm{d}\breve{V}_t = \int_{\mathbb{R}} \varphi_h(t - s) f(t, \theta)\,\mathrm{d}t + \varepsilon\int_{\mathbb{R}} \varphi_h(t - s)\,\mathrm{d}\breve{W}_t. \tag{11}$$

In effect, the missing contributions to integrals over $(-\frac{1}{2}h, 0)$ are provided by observations from (1) in $(1 - \frac{1}{2}h, 1)$, and vice versa. Equation (7) is extended in the obvious way: $\mu_s^\theta = \int_{\mathbb{R}} \varphi_h(t - s) f(t, \theta)\,\mathrm{d}t$. The vector $(\bar{V}_s)_{s \in S_n}$ has a periodic extension to $\{(k - \frac{1}{2})/n;\ k \in \mathbb{Z}\}$ with stationary covariance matrix (under the weak condition $h \leqslant \frac{1}{2}$). The hereby introduced and somewhat counter-intuitive dependence between observations at opposite ends of the interval we consider to be of no practical importance from a statistical point of view, because the change point θ^* is located in the interior of T, and, as our results will show, the estimator belongs with probability close to one to a decreasing interval around θ^*. We shall exploit the periodic nature of the model to orthogonalize the covariance matrix under suitable conditions on φ (see the remark to Theorem 1 below).

For the asymptotic analysis of model (6) we make the further assumption (iv) on f_1, f_2 and the assumptions in (v) on the kernel φ:

(iv) Δ_f is continuous on T, and there exist $\alpha_1 > 0$ and $\alpha_2 < \infty$ such that, for fixed (F) and decreasing (D) jump-size asymptotics, respectively,

F: $0 < \alpha_1 \leqslant \Delta_f(t) \leqslant \alpha_2$,

D: $0 < \alpha_1\Delta \leqslant \Delta_f(t) \leqslant \alpha_2\Delta$, and moreover $\omega(\Delta_f, r) = \Delta o(1)$ as $r \to 0$,

where $\omega(g, \cdot)$ is the modulus of continuity, $\omega(g, r) = \sup_{t,t' \in T: |t-t'| \leqslant r} |g(t) - g(t')|$.

(v) $\varphi \in L^{\infty}(\mathbb{R})$ has bounded support (taken to lie in $[-\frac{1}{2}, \frac{1}{2}]$) and $\int_{\mathbb{R}} \varphi(u)\,\mathrm{d}u = 1$, and the Fourier transform $\hat{\varphi}$ satisfies the condition

$$\exists c_1 > 0,\ c_2 < \infty\colon\ c_1 \leqslant \sum_{k\in\mathbb{Z}} |\hat{\varphi}(\omega + 2k\pi/\lambda)|^2 \leqslant c_2. \tag{12}$$

In (iv) we have, without loss of generality, taken $\Delta > 0$. The condition is valid locally around θ^* as a consequence of (ii), assuming in the decreasing jump-size case also $\Delta^* = O(\Delta)$. The extension to the entire interval is technically convenient for controlling the behaviour of the likelihood function away from θ^*.

Theorem 1. *For model* (6) *with* $nh \equiv \lambda^{-1}$, *for fixed* $0 < \lambda < \infty$, *periodically extended as in* (11) *if* $\lambda < 1$, *define the rate of convergence*

$$\bar{\psi}_\varepsilon = c_{\varphi,\lambda}\, \varepsilon^2/\Delta^2,$$

with the constant $c_{\varphi,\lambda}$ *given by*

$$c_{\varphi,\lambda} = \lambda \sum_{p\in\mathbb{Z}} \int_{\mathbb{R}} \varphi(u)\varphi(u + \lambda p)\,\mathrm{d}u,$$

and assume conditions (i)–(v) as well as the following relations between the models for fixed (F) and decreasing (D) jump-size asymptotics, respectively:

$$\text{F:}\quad \varepsilon \to 0,\ n \to \infty,\ \Delta \neq 0,\ \varepsilon^2 n \to \infty,$$

$$\text{D:}\quad \varepsilon \to 0,\ n \to \infty,\ \Delta \to 0,\ \varepsilon/\Delta \to 0,\ \varepsilon^2 n/\Delta^2 \to \infty,\ \varepsilon^2\Delta^*/\Delta^3 \to 0.$$

Then we have that

$$\bar{\psi}_\varepsilon^{-1}(\bar{\theta}_\varepsilon - \theta^*) \xrightarrow{\mathscr{L}} F_{\mathrm{tri}} \qquad \text{as } \varepsilon \to 0.$$

Remark. We indicate the main idea of the proof for dependent observations, which may be of independent interest. The covariance (9) of (6) is essentially given by L^2-products of the vectors $(\varphi(\cdot - \lambda k);\ k \in \mathbb{Z})$. Conditions under which such systems can be orthogonalized are summarized in Lemma 1 in Section 5; for the present it suffices that under condition (12) there exists a sequence $(a_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ by which we can define a kernel $\tilde{\varphi}$ as

$$\tilde{\varphi}(u) = \sum_{k\in\mathbb{Z}} a_k \varphi(u - \lambda k), \qquad u \in \mathbb{R}, \tag{13}$$

and such that the system $(\tilde{\varphi}(\cdot - \lambda k);\ k \in \mathbb{Z})$ is an orthogonal expansion of $(\varphi(\cdot - \lambda k);\ k \in \mathbb{Z})$. The expansion is utilized to transform (6) into an independent model for an orthogonalized observation vector $(\tilde{V}_s)_{s\in S_n}$,

$$\tilde{V}_s \doteq \sum_{k\in\mathbb{Z}} a_k \bar{V}_{s+k/n} = \int_{\mathbb{R}} \tilde{\varphi}_h(t - s) f(t, \theta)\,\mathrm{d}t + \varepsilon \int_{\mathbb{R}} \tilde{\varphi}_h(t - s)\,\mathrm{d}\check{W}_t. \tag{14}$$

Details are given in Section 5. The periodic extension of (6) is essential for the orthogonal expansion of the model.

Theorem 2. *Under the assumptions of Theorem* 1 *and whenever* $\psi_\varepsilon = \bar{\psi}_\varepsilon$, *we have that*

$$\psi_\varepsilon^{-1}(\hat{\theta}_\varepsilon - \bar{\theta}_\varepsilon) \xrightarrow{\mathbb{P}} 0 \qquad \textit{as } \varepsilon \to 0.$$

Remark. The conditions under which $\psi_\varepsilon = \bar{\psi}_\varepsilon$ are explained in detail in the next section.

4. Discussion and conclusions

In this section we elaborate on our conditions on the kernel φ and summarize the main message of Theorems 1–2, and also make a brief remark on extension of the results from ML to Bayesian estimators.

The conditions on φ are weak and essentially needed for the orthogonalization procedure; in particular, in the independent case $\varphi \in L^2(\mathbb{R})$ is sufficient, and (12) cancels. The condition (12) on the Fourier transform of φ expresses that $(\varphi(\cdot - \lambda k);\, k \in \mathbb{Z})$ is a Riesz basis for the subspace spanned by these functions. That is, for all $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ and $k_1, \ldots, k_m \in \mathbb{Z}$, the inequality

$$c_1 \sum_{i=1}^m \alpha_i^2 \leqslant \int_{\mathbb{R}} \left[\sum_{i=1}^m \alpha_i \varphi(u - \lambda k_i) \right]^2 \mathrm{d}u \leqslant c_2 \sum_{i=1}^m \alpha_i^2,$$

holds, with the same constants c_1, c_2 as in (12). Since φ is compactly supported the integral can be expressed in terms of the elements of the covariance matrix (9). The lower bound is the critical one, and it states loosely that no sequence of non-degenerate linear combinations of the $(\bar{V}_s)_{s \in S_n}$ can have a degenerate limit.

The convergence rates ψ_ε and $\bar{\psi}_\varepsilon$ coincide if and only if $c_{\varphi,\lambda} = 1$. For $\lambda \geqslant 1$ and generally for independent observations, we have $\lambda^{-1} c_{\varphi,\lambda} = \int_{\mathbb{R}} \varphi^2(u)\,\mathrm{d}u \geqslant 1$, by Hölder's inequality, and with equality only if $\varphi(u) = \mathbf{1}\{u \in [-\frac{1}{2}, \frac{1}{2}]\}$. For $\lambda < 1$ and independent observations one may ask if $c_{\varphi,\lambda} = 1$ is possible. The answer is affirmative; take any compactly supported wavelet ϕ (Daubechies 1988) with $\operatorname{supp}(\phi) = [-\lambda/2, \lambda/2]$, say, and rescale to $[-\frac{1}{2}, \frac{1}{2}]$ by $\varphi(u) \doteq \lambda^{-1}\phi(u/\lambda)$. In fact, the conditions on φ are close to being equivalent to $(\phi(\cdot - k);\, k \in \mathbb{Z})$ defining a multiresolution analysis (Meyer 1990, Chapter II).

In the dependent observation case we have $\lambda^{-1} c_{\varphi,\lambda} = \int_{\mathbb{R}} \tilde{\varphi}^2(u)\,\mathrm{d}u$, by the formula in Lemma 1 in Section 5. Let $\phi(u) = \lambda\varphi(\lambda u)$ as above and denote by $(\tilde{\phi}(\cdot - k);\, k \in \mathbb{Z})$ the orthogonalization of $(\phi(\cdot - k);\, k \in \mathbb{Z})$; then $\tilde{\varphi}(u) = \lambda^{-1}\tilde{\phi}(u/\lambda)$, and $c_{\varphi,\lambda} = 1 \Leftrightarrow \int \tilde{\phi}^2(u)\,\mathrm{d}u = 1$. We say in this case that the function ϕ permits an ortho*normal* expansion.

In conclusion, the relation between asymptotics for the ML estimators in continuous and discrete models (1) and (11) respectively can be summarized as follows.

Corollary 1. *Under the assumptions of Theorem* 1, $\hat{\theta}_\varepsilon$ *and* $\bar{\theta}_\varepsilon$ *are asymptotically equivalent in the sense of Theorem* 2 *only in the following cases:*

disjoint supports ($\lambda \geqslant 1$): *integrated observations*, *i.e.*, $h = 1/n$ *and*

$$\varphi(u) = \mathbf{1}\{u \in [-\tfrac{1}{2}, \tfrac{1}{2}]\},$$

non-disjoint supports ($\lambda < 1$): *there exists an ortho*normal *expansion of the system*

$$(\phi(\cdot - k);\ k \in \mathbb{Z}).$$

Otherwise the asymptotic limits of $\hat{\theta}_\varepsilon$ *and* $\bar{\theta}_\varepsilon$ *are similar, but the convergence rate of* $\bar{\theta}_\varepsilon$ *relative to that of* $\hat{\theta}_\varepsilon$ *differs by a constant greater than one.*

For the sake of completeness we make a few remarks on Bayesian estimation. In the continuous-time model, for instance, the Bayesian estimator $\hat{\theta}_{\varepsilon,p}$ based on a positive, continuous prior density π on $\Theta = [0, 1]$ and a p-power loss function $t \mapsto |t - \theta^*|^p$ is defined as the minimizer of $t \mapsto \int (\varepsilon^{-2}|t-\theta|)^p L(\theta;\ V)\pi(\theta)\,\mathrm{d}\theta$, where $L(\cdot;\ V)$ is the likelihood function. Generally, the Bayesian estimator is consistent with the same rate of convergence as the ML estimator under *weaker* conditions on the statistical problem (Ibragimov and Hasminskii 1981, Section I.10), and our Theorems A, 1 and 2 should carry over to Bayesian estimation. The limiting distribution F_{tri} is accordingly replaced by the distribution of the minimizer of $v \mapsto \int_{\mathbb{R}} |v - u|^p \exp\{B_u\}\,\mathrm{d}u$. We refer to Ibragimov and Hasminskii (1981, Sections VII.2–3) for a discussion of this distribution and the fact that the asymptotic efficiency of $\hat{\theta}_\varepsilon$ relative to $\hat{\theta}_{\varepsilon,2}$ is about 0.74; the exact value has recently been calculated by Rubin and Song (1995).

5. Proofs

Proof of Theorem A

We give only a brief sketch to motivate the assumptions on f_1, f_2 for decreasing jump-size asymptotics. The main line of the proof is similar to that of Theorem 1 or the development in Stryhn (1994).

The log-likelihood ratio $\ell(\theta) = \log L(\theta;\ V) = \log(\mathrm{d}\mathbb{P}_\theta/\mathrm{d}\mathbb{P}_{\theta^*})(V)$ is

$$\ell(\theta) = -\tfrac{1}{2}\varepsilon^{-2}\int_T [f(t;\ \theta) - f(t;\ \theta^*)]^2\,\mathrm{d}t + \varepsilon^{-1}\int_T [f(t;\ \theta) - f(t;\ \theta^*)]\,\mathrm{d}W_t,$$

which for $\theta \geqslant \theta^*$ can be written

$$\ell(\theta) = -\tfrac{1}{2}\varepsilon^{-2}\int_{(\theta^*,\theta]} [f_1(t) - f_2(t)]^2\,\mathrm{d}t + \varepsilon^{-1}\int_{(\theta^*,\theta]} [f_1(t) - f_2(t)]\,\mathrm{d}W_t.$$

We rescale the process by letting $X_u = \ell(\theta)$ for $\theta = \theta^* + \psi_\varepsilon u$, and calculate by use of condition (ii), for $0 \leqslant u \leqslant K$ such that $\theta \in N^*$,

$$
\begin{aligned}
\mathrm{E}X_u = \mathrm{E}\ell(\theta) &= -\tfrac{1}{2}\varepsilon^{-2}\int_{(\theta^*,\theta]}[f_1(t)-f_2(t)]^2\,\mathrm{d}t \\
&= -\tfrac{1}{2}\varepsilon^{-2}\int_{(\theta^*,\theta]}[\Delta^2+2\Delta\Delta_f(\theta_t)(t-\theta^*)+\Delta_f^2(\theta_t)(t-\theta^*)^2]\,\mathrm{d}t \\
&= -\tfrac{1}{2}\varepsilon^{-2}\{\Delta^2(\theta-\theta^*)+\Delta\Delta^*(\theta-\theta^*)^2O(1)+(\Delta^*)^2(\theta-\theta^*)^3O(1)\} \\
&= -u/2+o(1) \qquad \text{as } \varepsilon\to 0,
\end{aligned}
$$

by the assumed asymptotic relations. Also, $\mathrm{var}(\ell(\theta)) = -2\mathrm{E}\ell(\theta)$, and the process $(\ell(\theta);\ \theta\in T)$ has independent increments. □

Proof of Theorem 1, independent case

The proof proceeds in two steps. First, we define a rescaled log-likelihood ratio process $(\bar{X}_u)$ to obtain $\bar{X}\xrightarrow{\mathrm{w}}B$ with respect to weak convergence on compact sets (cf. Neuhaus 1971); recall that $B=(B_u)$ is a two-sided Brownian motion with triangular drift. Second, we show that the normalized estimator $\bar{\psi}_\varepsilon^{-1}(\bar{\theta}_\varepsilon-\theta^*)$ is bounded in probability. Combining these two assertions with the continuous mapping theorem applied to the argmax functional restricted to compact intervals, the desired result follows. We consider below only decreasing jump-size asymptotics; the derivation for fixed asymptotics is entirely similar.

The discrete-time log-likelihood ratio $\bar{\ell}(\theta)=\log L(\theta;\ \bar{V})-\log L(\theta^*;\ \bar{V})$ takes the form

$$
\bar{\ell}(\theta) = -\tfrac{1}{2}\sigma^{-2}\sum_{s\in S_n}(\mu_s^\theta-\mu_s^{\theta^*})^2+\varepsilon\sigma^{-2}\sum_{s\in S_n}(\mu_s^\theta-\mu_s^{\theta^*})\int_{\mathbb{R}}\varphi_h(t-s)\,\mathrm{d}\check{W}_t. \tag{15}
$$

The process $(\bar{\ell}(\theta);\ \theta\in T)$ is Gaussian with

$$
\mathrm{var}(\bar{\ell}(\theta)) = \sigma^{-2}\sum_{s\in S_n}(\mu_s^\theta-\mu_s^{\theta^*})^2 = -2\mathrm{E}\bar{\ell}(\theta), \tag{16}
$$

$$
\mathrm{cov}(\bar{\ell}(\theta),\bar{\ell}(\theta')) = \sigma^{-2}\sum_{s\in S_n}(\mu_s^\theta-\mu_s^{\theta^*})(\mu_s^{\theta'}-\mu_s^{\theta^*}). \tag{17}
$$

Define a rescaled log-likelihood ratio $(\bar{X}_u)_{u\in\mathbb{R}}$ by letting $\theta=\theta^*+\bar{\psi}_\varepsilon u$ and

$$
\bar{X}_u = \bar{\ell}(\theta) = \bar{\ell}(\theta^*+\bar{\psi}_\varepsilon u).
$$

Take fixed $K>0$, and consider $0\leqslant u,\ u'\leqslant K$. Using (2), we have

$$
\mu_s^\theta-\mu_s^{\theta^*} = \int_{\theta^*}^{\theta}\varphi_h(t-s)[f_1(t)-f_2(t)]\,\mathrm{d}t, \qquad s\in S_{n,h},
$$

which, by property (ii) for $\theta\in N^*$, can be written

$$
\mu_s^\theta-\mu_s^{\theta^*} = [\Delta+\Delta^*(\theta-\theta^*)O(1)]\int_{\theta^*}^{\theta}\varphi_h(t-s)\,\mathrm{d}t.
$$

Since the sums over S_n in (16) are virtually restricted to lattice arguments s with $\theta^* - h/2 \leqslant s \leqslant \theta + h/2$ (and similarly for (17)), the moments of $\bar{X}$ can be expanded as

$$\begin{aligned}
\mathrm{E}\bar{X}_u &= -\tfrac{1}{2}\sigma^{-2}[\Delta + \Delta^*(\theta - \theta^*)O(1)]^2[n(\theta - \theta^*) + O(1)] \\
&= -u/2 + \Delta^2\varepsilon^{-2}hO(1) + \Delta^*\varepsilon^2/\Delta^3 O(1) \qquad \textit{as } \varepsilon \to 0, \\
\operatorname{cov}(\bar{X}_u, \bar{X}_{u'}) &= \sigma^{-2}[\Delta + \Delta^*(\theta - \theta^*)O(1)]^2[n(\min(\theta, \theta') - \theta^*) + O(1)] \\
&= \min(u, u') + \Delta^2\varepsilon^{-2}hO(1) + \Delta^*\varepsilon^2/\Delta^3 O(1) \quad \textit{as } \varepsilon \to 0.
\end{aligned}$$

In the main, these formulae show that $\bar{X}$ converges weakly on compact sets to B.

For the second part, we lead off by rewriting equation (15) as

$$\bar{X}_u = m(u) + \kappa(u) = m(u)[1 + \kappa(u)/m(u)], \qquad u \neq 0,$$

where $m(u) = \mathrm{E}\bar{X}_u$ is strictly negative for $u \neq 0$, by property (i). We will show that $\limsup \mathbb{P}(\inf_{u>K} \kappa(u)/m(u) \leqslant -(1 - \delta/2)) \to 0$ as $K \to \infty$ for some fixed $\delta \in (0, 1)$. This is sufficient for the desired argmax boundedness of $\bar{X}$, because the argmax cannot be taken at negative values of $\bar{X}_u$. Introduce the notation $\bar{Y} = (\bar{Y}_u)$, with

$$\bar{Y}_u = -\kappa(u)/m(u) = 2\varepsilon \frac{\sum_s(\mu_s^\theta - \mu_s^{\theta^*}) \int_{\mathbb{R}} \varphi_h(t - s)\, \mathrm{d}\check{W}_t}{\sum_s(\mu_s^\theta - \mu_s^{\theta^*})^2}.$$

Then the process $\bar{Y}$ is centred and Gaussian with $\operatorname{var}(\bar{Y}_u) = 4\varepsilon^2\lambda^{-1}c_{\varphi,\lambda}/[h\sum_s(\mu_s^\theta - \mu_s^{\theta^*})^2]$. We will first prove that $\limsup_{\varepsilon\to 0} \mathrm{E}\sup_{u>K}\bar{Y}_u \leqslant 1 - \delta$ for K large enough. Denote by φ_+ and φ_- the integrals of the positive and negative parts of φ, and by $\|\varphi\|_1$ the L^1-norm of φ; we have $\varphi_+ + \varphi_- = 1$ and $\varphi_+ - \varphi_- = \|\varphi\|_1$. Then we obtain, using condition (iv), the following bounds on $\mu_s^{\theta'} - \mu_s^\theta$, valid for $\theta' \geqslant \theta$,

$$|\mu_s^{\theta'} - \mu_s^\theta| \leqslant \|\varphi\|_1\alpha_2\Delta,$$

and furthermore, for $\theta + h/2 \leqslant s \leqslant \theta' - h/2$,

$$|\mu_s^{\theta'} - \mu_s^\theta| \geqslant \Delta_f(s) + \omega(\Delta_f, h)\varphi_- - \omega(\Delta_f, h)\varphi_+ \geqslant \alpha_1\Delta - ||\varphi||_1\omega(\Delta_f, h).$$

Now, we calculate, for $u' = \bar{\psi}_\varepsilon^{-1}(\theta' - \theta^*) \geqslant u = \bar{\psi}_\varepsilon^{-1}(\theta - \theta^*) \geqslant K$,

$$\mathrm{E}(\bar{Y}_{u'} - \bar{Y}_u)^2 = \frac{4\varepsilon^2 c_{\varphi,\lambda}}{\lambda h} \frac{\sum_s(\mu_s^{\theta'} - \mu_s^\theta)^2}{\sum_s(\mu_s^\theta - \mu_s^{\theta^*})^2 \sum_s(\mu_s^{\theta'} - \mu_s^{\theta^*})^2},$$

and insert the above expressions to obtain

$$\begin{aligned}
\mathrm{E}(\bar{Y}_{u'} - \bar{Y}_u)^2 &\leqslant \text{const.} \frac{\varepsilon^2 c_{\varphi,\lambda}}{\lambda h\Delta^2} \frac{n(\theta' - \theta)}{n(\theta - \theta^*)n(\theta' - \theta^*)} \\
&= \text{const.} \frac{u' - u}{uu'} = \text{const.}\, \mathrm{E}(Y_{u'} - Y_u)^2,
\end{aligned}$$

where $Y_u = W_u/u$ for Brownian motion (W_u). Thus, we can apply Sudakov–Fernique's inequality (Adler 1990, Theorem 2.9),

$$\limsup \mathrm{E} \sup_{u>K} \bar{Y}_u \leqslant \text{const.} \, \mathrm{E} \sup_{u>K} Y_u$$

$$\leqslant 1 - \delta \text{ for } K \text{ large enough.}$$

This calculation allows us to employ Borell's inequality (Adler 1990, Theorem 2.1) to yield

$$\begin{aligned}
\mathbb{P}\{\inf_{u>K} \kappa(u)/m(u) \leqslant -(1-\delta/2)\} &= \mathbb{P}\{\sup_{u>K} \bar{Y}_u \geqslant 1 - \delta/2\} \\
&\leqslant 2 \exp\{-\tfrac{1}{2}(1 - \delta/2 - \mathrm{E} \sup_{u>K} \bar{Y}_u)^2 / \sup_{u>K} \operatorname{var}(\bar{Y}_u)\} \\
&\leqslant 2 \exp\left\{ -\frac{\lambda h}{8 c_{\varphi,\lambda} \varepsilon^2} (\delta/2)^2 \sum_{\theta^* - h/2}^{\theta^* + \bar{\psi}_\varepsilon K - h/2} (\mu_s^{\theta^* + \bar{\psi}_\varepsilon K} - \mu_s^{\theta^*})^2 \right\} \\
&= 2 \exp\{-\delta^2 K/32 + o(1)\} \qquad \text{as } \varepsilon \to 0,
\end{aligned}$$

by an expansion similar to that in the first part of the proof. Thus the left-hand side probability tends to zero as $K \to \infty$ as desired, in fact at an exponential rate. □

Proof of Theorem 1, dependent case

The main idea of the proof is outlined in the remark after Theorem 1. The construction is based on the following lemma. Introduce, for any real-valued function f and any $p \in \mathbb{N}$, the notation $f_p = \int f^p(u)\,\mathrm{d}u$.

Lemma 1. *Let $\phi \in L^2(\mathbb{R})$ and assume that $(\phi(\cdot - k);\ k \in \mathbb{Z})$ is a Riesz basis of* $\operatorname{span}(\phi(\cdot - k);\ k \in \mathbb{Z})$, *or equivalently that the Fourier transform $\hat{\phi}$ of ϕ satisfies*

$$\exists c_1 > 0,\ c_2 < \infty\colon\ c_1 \leqslant \sum_{k \in \mathbb{Z}} |\hat{\phi}(\omega + 2k\pi)|^2 \leqslant c_2.$$

Then there exists an orthogonal expansion of $(\phi(\cdot - k);\ k \in \mathbb{Z})$, that is, a (real-valued) sequence $(a_k) \in \ell^2(\mathbb{Z})$ such that the function $\tilde{\phi}\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\tilde{\phi}(u) = \sum_{k \in \mathbb{Z}} a_k \phi(u - k), \qquad u \in \mathbb{R},$$

has the following properties:

1. $\operatorname{span}(\phi(\cdot - k);\ k \in \mathbb{Z}) = \operatorname{span}(\tilde{\phi}(\cdot - k);\ k \in \mathbb{Z})$;
2. $\int_{\mathbb{R}} \tilde{\phi}(u - k)\tilde{\phi}(u - k')\,\mathrm{d}u = 0$ *for* $k \neq k'$.

Furthermore, if $\phi \in L^1(\mathbb{R})$ with $\int_{\mathbb{R}} \phi(u)\,\mathrm{d}u \neq 0$, then $\tilde{\phi} \in L^1(\mathbb{R})$ can be chosen to satisfy the following properties:

3. $\int_{\mathbb{R}} \tilde{\phi}(u)\,\mathrm{d}u = 1$;
4. $\int_{\mathbb{R}} \tilde{\phi}^2(u)\,\mathrm{d}u = \sum_{p \in \mathbb{Z}} \int_{\mathbb{R}} \phi(u)\phi(u + p)\,\mathrm{d}u / \phi_1^2 \geqslant 1$.

Finally, if ϕ has bounded support and belongs to $L^\infty(\mathbb{R})$, then

5. $\forall p > 0\ \exists C_p > 0$: $|\tilde{\phi}(u)| \leqslant C_p(1 + |u|)^{-p}$.

Proof. The lemma relies on standard orthonormalization techniques as in Meyer (1990, Chapter II). For completeness a brief sketch of proof is included here.

Define

$$\widehat{(\tilde{\phi})}(\omega) = \hat{\phi}(\omega) \Bigg/ \left[\sum_{k\in\mathbb{Z}} |\hat{\phi}(\omega + 2k\pi)|^2\right]^{1/2}.$$

By Fourier inversion, there exists a sequence (a_k) in $\ell^2(\mathbb{Z})$ such that

$$\tilde{\phi}(u) = \sum_{k\in\mathbb{Z}} a_k \phi(u - k).$$

Also, $\sum_{k\in\mathbb{Z}} |\widehat{(\tilde{\phi})}(\omega + 2k\pi)|^2 \equiv 1$, which by the Poisson summation formula is equivalent to

$$\int_{\mathbb{R}} \tilde{\phi}(u)\tilde{\phi}(u - k)\,\mathrm{d}u = \delta_{0,k},$$

where $\delta_{k,k'}$ equals 1 for $k = k'$, and 0 otherwise. Thus properties 1 and 2 have been proved, and in fact $(\tilde{\phi}(\cdot - k);\ k \in \mathbb{Z})$ is an orthonormal basis.

Next, by expanding $\tilde{\phi}$ in the identity $\sum_{k\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}kx} \int_{\mathbb{R}} \tilde{\phi}(u)\tilde{\phi}(u - k)\,\mathrm{d}u = \tilde{\phi}_2\ \ (= 1)$, one obtains the following relation, valid for all x:

$$|A(x)|^2\Phi(x) = \tilde{\phi}_2, \qquad \text{with } A(x) = \sum_{k\in\mathbb{Z}} a_k \mathrm{e}^{\mathrm{i}kx} \text{ and } \Phi(x) = \sum_{k\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}kx} \int_{\mathbb{R}} \phi(u)\phi(u - k)\,\mathrm{d}u. \tag{18}$$

In particular, it follows by insertion for $x = 0$ that $0 < A(0) = \sum_{k\in\mathbb{Z}} a_k = \tilde{\phi}_1/\phi_1$. Thus, we can scale the function $\tilde{\phi}$ and the sequence (a_k) to obtain property 3 and, accordingly, for the normalized function $\tilde{\phi}$,

$$\tilde{\phi}_2 = \Phi(0)/\phi_1^2,$$

$$\tilde{\phi}_2 = \left(\phi_1 \Bigg/ \left[\sum_{k\in\mathbb{Z}} |\hat{\phi}(2k\pi)|^2)\right]^{1/2}\right)^{-2} = \sum_{k\in\mathbb{Z}} |\hat{\phi}(2k\pi)|^2/\phi_1^2 \geqslant |\hat{\phi}(0)|^2/\phi_1^2 = 1.$$

Finally, when ϕ is compactly supported the function Φ in (18) is C^∞, and in particular $A(x)$ is C^∞ at 0. Therefore, for all $p \geqslant 0$, $\lim_{k\to\infty} k^p a_k = 0$ and, using also the boundedness of ϕ, we conclude that the rate of decrease of $\tilde{\phi}(u)$ at infinity is faster than every power of u. $\square$

We use the results of the lemma with the function ϕ given by $\phi(u) = \lambda\varphi(\lambda u)$, and let $\tilde{\varphi}(u) = \lambda^{-1}\tilde{\phi}(u/\lambda)$. Thus we obtain an orthogonal expansion $(\tilde{\varphi}(\cdot - \lambda k);\ k \in \mathbb{Z})$ of the system $(\varphi(\cdot - \lambda k);\ k \in \mathbb{Z})$ under the condition that $(\varphi(\cdot - \lambda k);\ k \in \mathbb{Z})$ is a Riesz basis, which is equivalent to (12). Clearly $\phi \in L^1(\mathbb{R})$ with $\phi_1 = 1$. Also, $\tilde{\varphi}_2 = \tilde{\phi}_2/\lambda$. Boundedness of ϕ and φ are equivalent.

The derivation (14) is obtained as follows: for $s \in S_n$,

$$\begin{aligned}\tilde{V}_s = \sum_{k\in\mathbb{Z}} a_k \bar{V}_{s+k/n} &= \int_{\mathbb{R}} \sum_{k\in\mathbb{Z}} a_k \varphi_h(t - s - k/n)\,\mathrm{d}\check{V}_t \\ &= \int_{\mathbb{R}} h^{-1} \sum_{k\in\mathbb{Z}} a_k \varphi((t-s)/h) - \lambda k)\,\mathrm{d}\check{V}_t \\ &= \int_{\mathbb{R}} h^{-1}\tilde{\varphi}((t-s)/h)\,\mathrm{d}\check{V}_t = \int_{\mathbb{R}} \tilde{\varphi}_h(t-s)\,\mathrm{d}\check{V}_t,\end{aligned}$$

where $\tilde{\varphi}_h$ denotes the kernel obtained from $\tilde{\varphi}$ with bandwidth h. To calculate the covariance function of $(\tilde{V}_s)$ we use the formula below, valid for $g_1,\ g_2 \in L^2(\mathbb{R})$ such that the right-hand side is finite:

$$\mathrm{E}\int_{\mathbb{R}} g_1(t)\,\mathrm{d}\check{W}_t \int_{\mathbb{R}} g_2(t)\,\mathrm{d}\check{W}_t = \sum_{p\in\mathbb{Z}} \int_{\mathbb{R}} g_1(u) g_2(u+p)\,\mathrm{d}u.$$

In a similar fashion to (9) we have, using property 2 of Lemma 1,

$$\begin{aligned}\operatorname{cov}(\tilde{V}_s,\ \tilde{V}_{s'}) &= \varepsilon^2 \sum_{p\in\mathbb{Z}} \int_{\mathbb{R}} \tilde{\varphi}_h(t-s)\tilde{\varphi}_h(t-s'+p)\,\mathrm{d}t \\ &= (\varepsilon^2/h) \sum_{p\in\mathbb{Z}} \int_{\mathbb{R}} \tilde{\varphi}(u)\tilde{\varphi}(u + \lambda(k - k' + np))\,\mathrm{d}u \\ &= (\varepsilon^2/h)\tilde{\varphi}_2 \mathbf{1}\{\exists p\colon (k-k') + np = 0\} \\ &= \begin{cases} (\varepsilon^2/h)\tilde{\varphi}_2 & \text{for } (k-k') \text{ modulo } n = 0 \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

Since the transformation connecting $(\bar{V}_s)$ and $(\tilde{V}_s)$ is linear and bijective, the corresponding ML estimates of θ are identical. However, the kernel $\tilde{\varphi}$ is not necessarily compactly supported, and some additional arguments are necessary next to those already given for the independent case.

The transformed, discrete model (14) is Gaussian with

$$\mathrm{E}\tilde{V}_s = \int \tilde{\varphi}_h(t-s) f(t,\ \theta)\,\mathrm{d}t \doteq \tilde{\mu}_s^\theta$$

$$\operatorname{var}(\tilde{V}_s) = (\varepsilon^2/h)\int \tilde{\varphi}^2(u)\,\mathrm{d}u = \varepsilon^2\tilde{\varphi}_2/h \doteq \tilde{\sigma}^2.$$

The log-likelihood function is

$$\tilde{\ell}(\theta) = \log L(\theta;\, \tilde{V}) - \log L(\theta^*;\, \tilde{V})$$

$$= -\tfrac{1}{2}\tilde{\sigma}^{-2}\sum_{s\in S_n}(\tilde{\mu}_s^\theta - \tilde{\mu}_s^{\theta^*})^2 + \varepsilon\tilde{\sigma}^{-2}\sum_{s\in S_n}(\tilde{\mu}_s^\theta - \tilde{\mu}_s^{\theta^*})\int_{\mathbb{R}}\tilde{\varphi}_h(t-s)\,\mathrm{d}\check{W}_t,$$

from which we obtain the formula similar to (16),

$$\mathrm{E}\tilde{\ell}(\theta) = -\tfrac{1}{2}\operatorname{var}(\tilde{\ell}(\theta)) = -\tfrac{1}{2}\tilde{\sigma}^{-2}\sum_{s\in S_n}(\tilde{\mu}_s^\theta - \tilde{\mu}_s^{\theta^*})^2.$$

We proceed as in the independent case by rescaling $\tilde{\ell}$ as $\tilde{X}_u \doteq \tilde{\ell}(\theta^* + \bar{\psi}_\varepsilon u)$. By similar reasoning, it suffices for obtaining $\tilde{X} \xrightarrow{\mathrm{w}} B$ on compact sets to show that (for $\theta \geqslant \theta^*$)

$$\sum_{s\in S_n}\left\{\int_{\theta^*}^{\theta}\tilde{\varphi}_h(t-s)\,\mathrm{d}t\right\}^2 = \sum_{s\in S_n}\left\{\int_{(\theta^*-s)/h}^{(\theta-s)/h}\tilde{\varphi}(u)\,\mathrm{d}u\right\}^2 \sim n(\theta-\theta^*) \qquad \text{as } \varepsilon \to 0. \tag{19}$$

Determine for arbitrary $\delta > 0$ by property 3 in Lemma 1 a constant $K_0 > 0$ such that

$$\forall K \geqslant K_0\colon \left|\left(\int_{-K}^{K}\tilde{\varphi}(u)\,\mathrm{d}u\right)^2 - 1\right| \leqslant \delta. \tag{20}$$

Now define $S_{n,1} = \{s \in S_n\colon (\theta^* - s)/h < -K_0 \text{ and } (\theta - s)/h > K_0\}$. The number of lattice points in $S_{n,1}$ is $n(\theta - \theta^*) - 2nhK_0 + O(1) = n(\theta - \theta^*) + O(1)$, which, in combination with (20), yields

$$\left|\sum_{s\in S_{n,1}}\left\{\int_{\theta^*}^{\theta}\tilde{\varphi}_h(t-s)\,\mathrm{d}t\right\}^2 - n(\theta-\theta^*)\right| \leqslant \delta n(\theta-\theta^*) + O(1) \qquad \text{as } \varepsilon \to 0.$$

Next, we turn to the contribution to (19) from $s \notin S_{n,1}$. The number of lattice points in $(\theta^*, \theta^* + hK_0)$ and $(\theta - hK_0, \theta)$ is in each case equal to $nhK_0 + O(1) = O(1)$. The sums ($s < \theta^*$ and $s > \theta$) can be estimated using property 5 of Lemma 1 in the following manner, for $p > 3/2$:

$$\sum_{s<\theta^*}\left\{\int_{(\theta^*-s)/h}^{(\theta-s)/h}\tilde{\varphi}(u)\,\mathrm{d}u\right\}^2 \leqslant \text{const.}\sum_{s<\theta^*}\left\{\int_{(\theta^*-s)/h}^{\infty}(1+|u|)^{-p}\,\mathrm{d}u\right\}^2$$

$$= \text{const.}\sum_{s<\theta^*}\left\{1 + \frac{\theta^* - s}{h}\right\}^{-2(p-1)}$$

$$= \text{const.}\sum_{k=n(\theta^*-s)>0}\{1+\lambda k\}^{-2(p-1)} = O(1).$$

Finally, going through the calculations for argmax boundedness, only very few adaptations are necessary for non-compactly supported $\tilde{\varphi}_h$. For the (crucial) upper and lower bounds of $\tilde{\mu}_s^\theta - \tilde{\mu}_s^{\theta^*}$ the global assumption (iv) on f_1, f_2 in combination with the rapidly decreasing tails of $\tilde{\varphi}$ in property 5 of Lemma 1 are sufficient. The last point where the bounded support of φ has been utilized is for an easy lower bound on $\operatorname{var}(\bar{Y}_u)$ in Borell's inequality; however, calculations similar to (19) lead to a lower bound in terms of $(\theta^* + \bar{\psi}_\varepsilon K)$, and the same argument applies. □

Proof of Theorem 2

Take arbitrary $\nu > 0$ and $c > 0$, and let $I_\varepsilon = (\hat{\theta}_\varepsilon - c\psi_\varepsilon, \hat{\theta}_\varepsilon + c\psi_\varepsilon)$; we will show that

$$\limsup_{\varepsilon\to 0} \mathbb{P}(\psi_\varepsilon^{-1}|\hat{\theta}_\varepsilon - \bar{\theta}_\varepsilon| > c) = \limsup_{\varepsilon\to 0} \mathbb{P}(\bar{\theta}_\varepsilon \notin I_\varepsilon) \geqslant 1 - \nu.$$

Consider the deviation between continuous- and discrete-time likelihoods,

$$Z(\theta) = \bar{\ell}(\theta) - \ell(\theta), \qquad \theta \in T.$$

The process $(Z(\theta);\, \theta \in T)$ is Gaussian with $Z(\theta^*) = 0$, and, for $\theta > \theta^*$, we have

$$\operatorname{var}(Z(\theta)) = \varepsilon^{-2}\left[\int_{(\theta^*,\theta]} (f_1 - f_2)^2\, \mathrm{d}\lambda - (\lambda h/c_{\varphi,\lambda}) \sum_{s\in S_n} (\mu_s^\theta - \mu_s^{\theta^*})^2\right] = 2\mathrm{E}\, Z(\theta),$$

$$\operatorname{cov}(Z(\theta),\, Z(\theta+\delta) - Z(\theta)) = -\sigma^{-2} \sum_{s\in S_n} (\mu_s^\theta - \mu_s^{\theta^*})(\mu_s^{\theta+\delta} - \mu_s^\theta).$$

By the expansions of the moments of X_u and $\bar{X}_u$ in the proofs of Theorems A and 1 respectively, and the fact that here $n\lambda h/c_{\varphi,\lambda} = c_{\varphi,\lambda}^{-1} = 1$, we have, for any $K > 0$ and $T_\varepsilon = [\theta^* - \psi_\varepsilon K,\, \theta^* + \psi_\varepsilon K]$,

$$\mathrm{E}\, Z(\theta),\, \operatorname{cov}(Z(\theta),\, Z(\theta')) = o(1) \qquad \text{as } \varepsilon \to 0, \text{ uniformly for } \theta, \theta' \in T_\varepsilon,$$

and consequently

$$\mathbb{P}(\sup_{\theta\in T_\varepsilon} |\bar{\ell}(\theta) - \ell(\theta)| > \eta) \to 0 \qquad \text{as } \varepsilon \to 0. \tag{21}$$

Since $\bar{\theta}_\varepsilon$ and $\hat{\theta}_\varepsilon$ are both consistent with rate ψ_ε (Theorems A and 1), we can fix $K > 0$ such that

$$\limsup_{\varepsilon\to 0} \mathbb{P}(\bar{\theta}_\varepsilon \in T_\varepsilon \text{ and } \hat{\theta}_\varepsilon \in T_\varepsilon) \geqslant 1 - \nu/3. \tag{22}$$

Furthermore, by the convergence of the rescaled log-likelihood of the continuous-time model (similarly to the proof of Theorem 1) we can fix $\eta > 0$ such that

$$\limsup_{\varepsilon\to 0} \mathbb{P}(\ell(\hat{\theta}_\varepsilon) - \sup_{\theta\notin I_\varepsilon} \ell(\theta) \geqslant 2\eta) \geqslant 1 - \nu/3. \tag{23}$$

Finally, we may take ε small enough to make the probability in (21) less than $\nu/3$.

After these preliminaries the main argument of the proof goes as follows. From the definition of $\bar{\theta}_\varepsilon$ and the set I_ε, we have

$$\{\bar{\theta}_\varepsilon \notin I_\varepsilon\} \subseteq \{\sup_{\theta\notin I_\varepsilon} \bar{\ell}(\theta) > \bar{\ell}(\hat{\theta}_\varepsilon)\},$$

and thus

$$\begin{aligned}\mathbb{P}(\bar{\theta}_\varepsilon \notin I_\varepsilon) &\leqslant \mathbb{P}(\bar{\theta}_\varepsilon \notin T_\varepsilon \text{ or } \hat{\theta}_\varepsilon \notin T_\varepsilon) + \mathbb{P}(\bar{\ell}(\hat{\theta}_\varepsilon) - \sup_{\theta \in T_\varepsilon \setminus I_\varepsilon} \bar{\ell}(\theta) < 0) \\ &\leqslant \mathbb{P}(\bar{\theta}_\varepsilon \notin T_\varepsilon \text{ or } \hat{\theta}_\varepsilon \notin T_\varepsilon) + \mathbb{P}(\sup_{\theta \in T_\varepsilon} |\bar{\ell}(\theta) - \ell(\theta)| > \eta) + \mathbb{P}(\ell(\hat{\theta}_\varepsilon) - \sup_{\theta \in T_\varepsilon \setminus I_\varepsilon} \ell(\theta) < 2\eta), \\ &\leqslant \nu/3 + \nu/3 + \nu/3 = \nu,\end{aligned}$$

where for the second inequality we used the fact that $\bar{\ell}(\hat{\theta}_\varepsilon) - \sup_{\theta \in T_\varepsilon \setminus I_\varepsilon} \bar{\ell}(\theta) < 0$ and $\sup_{\theta \in T_\varepsilon} |\bar{\ell}(\theta) - \ell(\theta)| \leqslant \eta$ imply $\ell(\hat{\theta}_\varepsilon) - \sup_{\theta \in T_\varepsilon \setminus I_\varepsilon} \ell(\theta) < 2\eta$. □

Acknowledgements

The research of the second author was supported by I.N.R.A. and the Danish Informatics Network in the Agricultural sciences, while visiting the Laboratoire de Biométrie in spring 1995.

References

Adler, R.J. (1990) *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, IMS Lecture Notes Monograph Ser. Hayward, CA: Institute of Mathematical Statistics.

Bhattacharya, P.K. and Brockwell, P.J. (1976) The minimum of an additive process with applications to signal estimation and storage theory. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **37**, 51–75.

Bibby, B.M. and Sørensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.

Daubechies, I. (1988) Orthonormal basis of compactly supported wavelets. *Comm. Pure Appl. Math.*, **41**, 909–996.

Genon-Catalot, V. and Jacod, J. (1993) On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. H. Poincaré, Probab. Statist.*, **29**, 119–151.

Hasminskii, R.Z. and Lebedev, V.S. (1990) On the properties of parametric estimators for areas of a discontinuous image. *Probl. Control Inform. Theory*, **19**, 375–385.

Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1–17.

Ibragimov, I.A. and Hasminskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.

Korostelev, A.P. (1987) On minimax estimation of a discontinuous signal. *Theory Probab. Appl.*, **32**, 727–730.

Kutoyants, Yu.A. (1984) *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann Verlag.

Laredo, C.F. (1990) A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *Ann. Statist.*, **18**, 1158–1171.

Meyer, Y. (1990) *Ondelettes: Ondelettes et Opérateurs I*. Paris: Hermann.

Neuhaus, G. (1971) On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.*, **42**, 1285–1295.

Pedersen, A.R. (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**, 55–71.

Rosenfeld, A. and Kak, A.C. (1982) *Digital Picture Processing*, 2nd edn. London: Academic Press.

Rubin, H. and Song, K.-S. (1995) Exact computation of the asymptotic efficiency of maximum likelihood estimators of a discontinuous signal in a Gaussian white noise. *Ann. Statist.*, **23**, 732–739.

Rudemo, M. and Stryhn, H. (1994) Boundary estimation for star-shaped objects. In E. Carlstein, H.-G. Müller and D. Siegmund (eds), *Change-Point Problems* IMS Lecture Notes Monograph Ser. Hayward, CA: Institute of Mathematical Statistics.

Stryhn, H. (1994) Spatial change point models applied to image segmentation. PhD thesis, The Royal Veterinary and Agricultural University, Copenhagen.

Yao, Y.-C. (1987) Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann. Statist.*, **15**, 1321–1328.

Received December 1995 and revised September 1996.