OPTIMAL SEQUENTIAL KERNEL DETECTION FOR DEPENDENT PROCESSES

Ansgar Steland[1]
Fakultät für Mathematik
Ruhr-Universität Bochum, Germany
ansgar.steland@ruhr-uni-bochum.de

Date: September 2003.

[1]Address of correspondence: Ansgar Steland, Ruhr-Universität Bochum, Fakultät für Mathematik, Mathematik 3 NA 3/71, Universitätsstr. 150, D-44780 Bochum, Germany.

Abstract. In many applications one is interested to detect certain (known) patterns in the mean of a process with smallest delay. Using an asymptotic framework which allows to capture that feature, we study a class of appropriate sequential nonparametric kernel procedures under local nonparametric alternatives. We prove a new theorem on the convergence of the normed delay of the associated sequential detection procedure which holds for dependent time series under a weak mixing condition. The result suggests a simple procedure to select a kernel from a finite set of candidate kernels, and therefore may also be of interest from a practical point of view. Further, we provide two new theorems about the existence and an explicit representation of optimal kernels minimizing the asymptotic normed delay. The results are illustrated by some examples.

Keywords: Enzyme kinetics, financial econometrics, nonparametric regression, statistical genetics, quality control.

Introduction

A classical problem of sequential analysis is to detect a location shift in an univariate time series by a binary decision procedure. More generally, we aim at testing sequentially whether the deterministic drift function vanishes (in-control or null model) or is equal to an out-of-control or alternative model. There are various important fields where such methods can be applied. We shall first briefly describe some fields of applications which motivated the topics discussed in this article.

Sequential methods are applied for a long time in quality control and statistical process control, where interest focuses on detecting the first time point where a production process fails. Failures of a machine may produce jumps in the sequence of the observed quality characteristic, whereas wastage may result in smooth but possibly nonlinear changes of the mean. In recent years there has been considerable interest in methods for dependent time series.

An active area is the on-line monitoring of sequential data streams from capital markets. Indeed, an analysts task is to detect structural changes in financial data as soon as possible in order to trigger actions as portfolio updates or hedges. Thus, methods designed to support sequential decision making are in order.

A further potential field of application is the analysis of microarray time series data consisting of gene expression levels of genes. Down- or upregulated genes can have important interpretations, e.g., when characterizing cancer cells, and the sequential detection of such level changes from time series could be of considerable value.

In biology sequential methods may be useful to study the temporal evolution of enzyme kinetics in order to detect time points where a reaction starts or exceeds a prespecified threshold. Often it is possible to associate certain (worst case) temporal patterns with phenomena, e.g., symptoms or reactions to stimuli, which are of biological interest to detect. In order to understand complex biological systems it may be useful to estimate such change points sequentially instead of applying a posteriori methods, since the behavior of the real biological system depends only on the past.

A basic model to capture level changes as motivated by the above application areas is as follows. Suppose we are observing a possibly non-stationary stochastic process, $\{\widetilde{Y}(t) : t \in \mathcal{T}\}$, in continuous or discrete time $\mathcal{T} = [0, \infty)$ with $E|\widetilde{Y}(t)| < \infty$. Consider the following decomposition of the process in a possibly non-homogenous drift $m(t) = E\widetilde{Y}(t)$ and an error process $\{\widetilde{\epsilon}_t : t \in \mathcal{T}\}$,

$$\widetilde{Y}(t) = m(t) + \widetilde{\epsilon}_t, \qquad (t \in \mathcal{T}).$$

Using the terms of statistical process control, we will say that the process is in-control, if $m(t) = 0$ for each $t \in \mathcal{T}$, and we are interested in kernel control charts, i.e., sequential kernel-based smoothing methods, to detect the first time point where the process gets out-of-control. Clearly, from a testing point of view we are sequentially testing the null hypothesis $H_0 : m = 0$ against the alternative $m \neq 0$. A common approach to the problem is to define a stopping rule (stopping time), N, based on some statistic that estimates at each time point t a functional of $\{m(s) : s \leq t\}$. Having defined a stopping rule, the stochastic properties of the associated delay, defined as N minus the change-point, are of interest.

Well-known stopping rules rely on CUSUM-, EWMA-, or Shewhart-type control charts which can often be tuned for the problem at hand. These proposals are motivated by certain optimality criteria and have been studied extensively in the literature. First publications are due to Page (1954, 1955), Girshick and Rubin (1952). Optimality properties of the CUSUM procedure in the sense of Lorden (1971), i.e., minimizing the conditional expectation of the delay given the least favorable event before the change-point was first shown by Moustakides (1986) and, using Bayesian arguments, by Ritov (1990, 1997) and Yakir (1997). For a discussion of EWMA control schemes see Schmid and Schoene (1997). Yakir, Krieger, and Pollak (1999) studied first order optimality of the CUSUM and Shiryayev-Roberts procedures to detect a change in regression. Their result deals with optimal stopping rules in the sense that the expected delay is minimal subject to a constraint on the average run length to a false alarm. However, that result is restricted to independent and normally distributed observations. A kernel-based a posteori procedure for detecting multiple change points which is in the spirit of the present article has been studied by Hušková and Slaby (1997) and Grabovsky, Horváth and Hušková (2000). For reviews we refer to Hušková (1991) and Antoch, Hušková, and Jarušková (2002).

The present paper provides an asymptotic analysis with local alternatives, which holds for a rich class of strongly mixing dependent processes. Our stopping rule uses a Priestley-Chao type kernel regression estimate which relies on a weighted sum of past observations without assuming knowledge of the alternative regression function or an estimate of it. The theory and application of such smoothing methods is nicely described in Hart (1997) or Härdle (1990). However, it is important to note that framework and assumptions of the present paper are different from classical nonparametric regression. Whereas in nonparametric regression it is assumed that the bandwidth h tends to 0 such that $nh \to \infty$, n denoting the (fixed) sample size, and $\max t_i - t_{i-1} \to 0$, as $n \to \infty$, our monitoring approach works with $h \to \infty$ and $t_i - t_{i-1} \geq \Delta$ for all i.

Sequential smoothing procedures, where a regression estimate is evaluated at the current observation, have been studied for various change-point problems, e.g. to monitor the derivative of a process mean (Schmid and Steland, 2000). Note also that they are implicitly applied in classical (fixed sample) nonparametric regression at the boundary. Of course, it is of special interest to study the simultaneous effect of both the kernel and the alternative drift on the asymptotic normed delay of the associated stopping rule. We prove a limit theorem addressing this question for general mixing processes. We then ask how to optimize the procedure w.r.t. the smoothing kernel for certain regression alternatives. It turns out that an explicit representation of the optimal kernel can be derived for arguments not exceeding the associated asymptotic optimal delay. For simple location shifts first results for the normed delay have been obtained by Brodsky and Darkhovsky (1993, 2000) for sequential kernel smoothers as studied here. When jumps are expected, jump-preserving estimators as discussed in Lee (1983), Chiu et. al (1998), Rue et al. (2002), and Pawlak and Rafajłowicz (2000, 2001) are an attractive alternative, since smoothers tend to smooth away jumps. Convergence results for the normed delay of jump-preserving stopping rules have been studied in Steland (2002a), where upper bounds for the asymptotic normed delay are established. For a Bayesian view on the asymptotic normed delay and optimal prior choice see Steland (2002b). On-line monitoring has been recently reviewed by Antoch and Jarušková (2002) and Frisén (2003). We also refer to Siegmund (1985).

We shall now explain the asymptotic framework of our approach more detailed. In order to evaluate a detection procedure we will consider local alternatives which converge to the in-control model as the (effective) sample size of the procedure tends to infinity. Simultaneously, the false-alarm rate will tend to 0. The local nonparametric alternatives studied here are given as a parameterized family of drift functions,

$$m(t) = m(t; h) = \mathbf{1}(t \geq t_q^*) m_0([t - t_q^*]/h),$$

where $t_q^* \in \mathcal{T}$ stands for the change-point assumed to be fixed but unknown, and $h > 0$ is a bandwidth parameter of the detection procedure introduced below determining the amount of past data used by the procedure. We assume $h \in \mathcal{H}$ for some countable and unbounded set $\mathcal{H} \subset \mathbb{R}_0^+$. m_0 denotes the generic model alternative inducing the sequence of local alternatives. We assume that m_0 is a piecewise Lipschitz continuous function. Our asymptotics will assume $h \to \infty$. Consequently, for each fixed $t \in \mathcal{T}$ we have $m_0(t; h) \to m_0(0)$, as $h \to \infty$, if m_0 is continuous in 0. In this sense, $m(t; h)$ defines a sequence of of local alternative if $m_0(0) = 0$. As we shall see below, h coincides with the bandwidth parameter determining the sample size of the kernel smoother on which the sequential detection procedure is based on. It turns out that the rate of convergence of the local alternative has to be related to the bandwidth parameter in this fashion to obtain a meaningful convergence result.

Assume the process is sampled at a sequence of fixed ordered time points, $\{t_n : n \in \mathbb{N}\}$, inducing a sequence of observations $\{Y_n : n \in \mathbb{N}\}$. Put $m_{nh} = m(t_n; h)$ and $\epsilon_n = \widetilde{\epsilon}(t_n)$ to obtain

$$Y_{nh} = m_{nh} + \epsilon_n, \qquad (n \in \mathbb{N}).$$

Let q denote the integer ensuring $t_q = \lfloor t_q^* \rfloor + 1$. Then

$$m_{nh} = \mathbf{1}(t_n \geq t_q) m_0([t_n - t_q]/h), \qquad n \in \mathbb{N}.$$

We do not assume that the time design $\{t_n\}$ becomes dense in some sense. In contrary, we use time points having a fixed minimal distance, and for simplicity we shall assume $t_n = n$ for all $n \in \mathbb{N}$. More general time designs will be discussed at the end of Section 2.

The organization of the paper is as follows. Section 1 provides basic notation, assumptions, and the definition of the kernel detection procedure. The limit theorem for the normed delay is established in Section 2. The result holds for a wide class of generic alternatives satisfying a mild integrability condition, provided that the smoothing kernel is Lipschitz continuous. Section 3 provides the result on the optimal kernel choice which minimizes the asymptotic normed delay. We provide both an existence theorem and a stronger representation theorem. Due to the close relationship of the optimal kernel and the generic alternative, this results requires both the kernel and the regression alternative to be continuous. We illustrate the results by a couple of examples.

1. Sequential kernel detection and assumptions

We consider the following sequential Priestley-Chao type kernel smoother

$$\widehat{m}_{nh} = \sum_{i=1}^{n} K_h(t_i - t_n) Y_{ih}, \tag{1}$$

$n \in \mathbb{N}$. We call $\widehat{m}_{nh}$ a *sequential* smoother, since at the n-th time point the Priestley-Chao type estimator $t \mapsto \sum K_h(t_i - t) Y_{ih}$ is only evaluated for $t = t_n$. Here and in the sequel $K_h(z) = K(z/h)/h$ denotes the rescaled version of a smoothing kernel K required to be a centered, symmetric, and Lipschitz continuous probability density. The associated $\{0, 1\}$-valued sequential decision rule is given by

$$d_{nh} = \mathbf{1}(|\widehat{m}_{nh}| > c), \tag{2}$$

or, $d_{nh} = \mathbf{1}(\widehat{m}_{nh} > c)$ (one-sided version), i.e., a signal is given if (the absolute value of) $\widehat{m}_{nh}$ exceeds a prespecified non-negative threshold c.

The corresponding stopping time is given by

$$N_h = \inf\{n \in \mathbb{N} : d_{nh} = 1\}$$

with $\inf \emptyset = \infty$. In addition, define the normed delay

$$\rho_h = \max\{N_h - t_q, 0\}/h.$$

If the kernel vanishes outside the interval $[-1, 1]$, the effective sample size of the detection procedure is equal to h. Then ρ_h is simply the delay expressed as a percentage of the effective sample size.

In this paper we will measure the efficiency of a decision procedure by the asymptotic behavior of its associated normed delay. We confine ourselves to stopping times meaning that decisions at time n only depend on $Y_1, \ldots, Y_n$.

Throughout the paper we shall assume that $\{\varepsilon_n\}$ is a stationary α mixing process in discrete time $\mathbb{N}$. Recall that α mixing (strongly mixing) means that $\alpha(k) \to 0$, if $k \to \infty$, where $\alpha(k)$ denotes the α-mixing coefficient defined by

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{k}^{\infty}} |P(A \cap B) - P(A)P(B)|.$$

Here $\mathcal{F}_k^l = \sigma(\varepsilon_k, \ldots, \varepsilon_l)$ stands for the σ-field induced by the random variables $\varepsilon_k, \ldots, \varepsilon_l$, $-\infty \le k \le l \le \infty$. Recall that α-mixing is a weak notion of dependence which is implied by

β- and ρ-mixing. For a general discussion of mixing coefficients and related limit theorems we refer to Bosq (1996). The regularity assumptions on the mixing coefficients of $\{\varepsilon_n\}$ will be given later. In addition, we assume $\{\varepsilon_n\}$ satisfies Cramer's condition, i.e.,

$$Ee^{c_1|\varepsilon_1|} < \infty$$

for some positive constant c_1.

The smoothing kernel K used to define the weighting scheme is taken from the class

$$\mathcal{K} = \left\{ K : \mathbb{R} \to [0, \infty) : \int K(s)\, ds = 1, K(s) = K(-s) \right\} \cap \mathrm{Lip}$$

of all symmetric probability densities on the real line which are Lipschitz continuous, i.e., there exists a Lipschitz constant L_K ensuring

$$|K(z_1) - K(z_2)| \leq L_K |z_1 - z_2|, \qquad (z_1, z_2 \in \mathbb{R}).$$

For our optimality results we will have to impose further conditions which will restrict the class $\mathcal{K}$.

Finally, we also need the following conditions. It is assumed that $m_0 : [0, \infty) \to \mathbb{R}$ is non-negative and satisfies, jointly with $K \in \mathcal{K}$, the following integrability condition,

$$\left| \int_0^x K(s - x) m_0(s)\, ds \right| < \infty \qquad (\forall x > 0).$$

2. Asymptotics for the normed delay

In this section we establish both an assertion about the in-control false-alarm rate and a limit theorem for the normed delay for general local nonparametric alternatives under dependent sampling.

We need the following specialized large deviation result for the control statistic $\widehat{m}_{nh}$. A related large deviation result for (unweighted) sums of random variables satisfying Cramer's condition can be found in Bosq (1996, Th. 1.4). For our purposes we need the following specialised version for mixing time series.

Define $S_{nh} = \sum_{i=1}^n K([t_i - t_n]/h)\epsilon_i$, $n \in \mathbb{N}$, $h \in \mathcal{H}$. For two real sequences (a_h) and (b_h) with $b_h \neq 0$ for sufficiently large h, we write $a_h \sim b_h$ if $a_h/b_h \to 1$, as $h \to \infty$ and $a_h \sim b_h$ up to a constant if $a_h/b_h \to c$, $h \to \infty$, for some constant c.

Theorem 2.1. *Assume $n/h \sim \zeta$ with $0 < \zeta < \infty$. Then the following assertions hold true.*

(i) *For each $x > 0$*

$$P(S_{nh} > xh) = O\left(\frac{n}{\sqrt{h}} e^{-c_2\sqrt{h}}\right) + O(n\alpha(\sqrt{h})) + O(\sqrt{h}e^{-c_1 h}) = o(1),$$

as $h \to \infty$, provided $\lim_{k\to\infty} k^2\alpha(k) = 0$.

(ii) *If $\sum_k k^2\alpha(k) < \infty$, then for each $x > 0$*

$$\sum_{h\in\mathcal{H}} P(S_{nh} > xh) < \infty$$

implying

$$P(S_{nh} > xh, i.o.) = 0.$$

Remark 2.1. *By construction of the stopping rule, Theorem 2.1 also makes an assertion about the in-control false-alarm rate. Note that our setting implies that the rate converges to 0, as the effective sample size h tends to infinity.*

Proof. Put $S_h = S_{nh}$. Fix $0 < \gamma < 1$. Note that $n/h \sim \zeta$. Partition the set $\{1, \ldots, n\}$ in blocks of length $l(h) = \lfloor(\zeta h)^{1/2}\gamma\rfloor$ yielding $b(h) = \lfloor n/l(h)\rfloor$ blocks. Note that $l(h) \sim h^{1/2}$ and $b(h) \sim h^{1/2}$ up to constants. We have

$$\begin{aligned} S_h &= \sum_{r=1}^{l(h)} S_h^{(r)} + R_h \\ S_h^{(r)} &= \sum_{k=1}^{b(h)} K([t_{kb(h)+r} - t_n]/h)\epsilon_{kb(h)+r} \\ R_h &= \sum_{i=b(h)l(h)+1}^{n} K([t_i - t_n]/h)\epsilon_i. \end{aligned}$$

W.l.o.g. we can assume $b(h)l(h) = n$, since $P[R_h > xh] = O(l(h)e^{-c_1 h})$ for some constant $c_1 > 0$. Next observe that

$$P[S_h > xh] \le \sum_{r=1}^{b(h)} P[S_h^{(r)} > (xh)/b(h)].$$

Markov's inequality, Cramer's condition, the strong mixing property, and Volonski and Rosanov (1959) provide for each $r = 1, \dots, l(h)$

$$\left| P\left[S_h^{(r)} > \frac{xh}{b(h)} \right] - e^{-t(xh)/b(h)} \prod_{k=1}^{b(h)} E e^{tK([t_{kl(h)+r} - t_n]/h])\epsilon_k} \right|$$
$$\leq 16(b(h)-1)\alpha(l(h)).$$

It is well-known that Cramer's condition holds iff. there are constants $g > 0$ and $T > 0$ such that $Ee^{t\epsilon_1} \leq e^{gt^2}$ for all $|t| \leq T$ and $g > (1/2)E\epsilon_1^2$ (Petrov (1975), Lemma III.5). Thus,

$$e^{-tx} \prod_{k=1}^{b(h)} E e^{tK([t_{kl(h)+r} - t_n]/h)\epsilon_i}$$
$$\leq \exp\left\{ gt^2 \sum_{k=1}^{b(h)} K([t_{kl(h)+r} - t_n]/h) - txh/b(h) \right\}$$

Minimizing the r.h.s. w.r.t. t gives the upper bound

$$\begin{cases} e\left(-\frac{x^2h^2}{b(h)^2} \frac{1}{2gC(r)} \right), & (xh)/b(h) \leq gTC(r) \\ e\left(-\frac{xh}{b(h)} \frac{T}{2} \right), & (xh)/b(h) > gTC(r) \end{cases}$$

where $C(r) = \sum_{r=1}^{b(h)} K([t_{kl(h)+r} - t_n]/h)^2$. Observe that the timepoints

$$t_{kl(h)+r}, k = 1, \dots, b(h)$$

form an equidistant partition of an interval converging to $(-\zeta, 0)$. The size of the partition equals $l(h)/h \sim h^{-1/2}$ up to a constant. Therefore, using $\int_{-\zeta}^0 K(s)^2\, ds = \int_0^\zeta K(s)^2\, ds$,

$$\left| \frac{l(h)}{h} \sum_{k=1}^{b(h)} K([t_{kb(h)+r} - t_n]/h)^2 - \int_0^\zeta K(s)^2\, ds \right| = O\left(\frac{l(h)}{h} \right).$$

Consequently, $[l(h)/h)]^{-1}C(r)$ is bounded away from 0 for large enough h. Thus, uniformly in $r = 1, \dots, b(h)$,

$$P[S_h^{(r)} > (xh)/b(h)] = O(e^{-ch^{1/2}}) + O(h^{1/2}\alpha(h^{1/2})),$$

for some constant $c > 0$, yielding

$$\begin{aligned}P[S_h > xh] &\leq \sum_{r=1}^{b(h)} P[S_h^{(r)} > (xh)/b(h)] + P[R_h > x] \\ &= O(b(h)e^{-ch^{1/2}}) + O(b(h)h^{1/2}\alpha(h^{1/2})) + O(l(h)e^{-c_1 h}) \\ &= O\left(\frac{n}{\sqrt{h}}e^{-c_2\sqrt{h}}\right) + O(n\alpha(\sqrt{h})) + O(\sqrt{h}e^{-c_1 h}).\end{aligned}$$

Therefore, the mixing condition

$$\lim_{k\to\infty} k^2\alpha(k) = 0$$

ensures

$$P[S_h > xh] = o(1), \quad \text{as } h \to \infty.$$

Finally, the above estimates and

$$\sum_k k^2\alpha(k) < \infty$$

yield $\sum_{h\in\mathcal{H}} P[S_h > xh] < \infty$, and an application of Borel-Cantelli provides

$$P[S_h > xh \ i.o.] = 0.$$

□

We may now formulate our main result on the strong law of large numbers for the normed delay. Define

$$\rho_0 = \inf\left\{\rho > 0 : \int_0^\rho K(s-\rho)m_0(s)ds = c\right\}. \tag{3}$$

Theorem 2.2. *Let $K \in \mathcal{K}$ be a given kernel and m_0 be a piecewise Lipschitz continuous generic alternative with $m_0(0) = 0$ and either $m_0 \geq 0$ or $m_0 \leq 0$ such that (3) exists and $0 < \rho_0 < \infty$. Then*

$$\rho_h \stackrel{a.s.}{\to} \rho_0,$$

as $h \to \infty$, provided that $\sum_k k^4\alpha(k) < \infty$.

Remark 2.2. *The proof even shows complete convergence.*

Proof. W.l.o.g. we assume $m_0 \geq 0$. Let $\varepsilon > 0$. We shall estimate $P[\rho_h - \rho_0 > \varepsilon]$ and $P[\rho_h < \rho_0 - \varepsilon]$. Put $n(h) = \lfloor (\rho_0 + \varepsilon)h \rfloor$. Then we have

$$\begin{aligned}
P[\rho_h - \rho_0 > \varepsilon] &= P[N_h > (\rho_0 + \varepsilon)h] \\
&\leq P[|\widehat{m}_{n(h),h}| \leq c] \\
&= P\left[\left|\sum_{i=1}^{n(h)} K_h(t_i - t_n)[m_{ih} + \epsilon_i]\right| \leq c\right] \\
&\leq P\left[\left|\sum_{i=1}^{n(h)} K_h(t_i - t_n)\epsilon_i\right| > c - \left|\sum_{i=q}^{n(h)} K_h(t_i - t_n)m_{ih}\right|\right],
\end{aligned}$$

where in the last step we used the fact that $Y_{ih} = \varepsilon_i$ if $1 \leq i < q = t_q$ and $Y_{ih} = m_0([t_i - t_q]/h) + \epsilon_i$ if $q \leq i \leq n(h)$. For the following argument we may assume that m_0 is Lipschitz continuous, since otherwise one may argue on subintervals. We have

$$\begin{aligned}
\sum_{i=q}^{n(h)} K_h(t_i - t_n)m_{ih} &= h^{-1} \sum_{i=q}^{n(h)} K([i - n(h)]/h) m_0([i - q]/h) \\
&= h^{-1} \sum_{i=q}^{n(h)} K(i/h - \rho_0 - \varepsilon) m_0(i/h) + O(h) \\
&= \int_0^{\rho_0 + \varepsilon} K(s - \rho_0 - \varepsilon) m_0(s) ds + O(h^{-1}) \\
&= \int_0^{\rho_0} K(s - \rho_0) m_0(s) ds + O(\varepsilon) + O(h^{-1}),
\end{aligned}$$

since $q/h \to 0$ and $n(h)/h \to \rho_0 + \varepsilon$, as $h \to \infty$, and K is Lipschitz continuous. Recalling the definition of ρ_0, there exists a constant $\kappa > 0$, which depends on ε, with $c - |\sum_{i=q}^{n(h)} K_h(t_i - t_n)m_{ih}| \geq \kappa > 0$, yielding

$$P[\rho_h - \rho_0 > \varepsilon] \leq P[S_{n(h)} > \kappa h].$$

We may now apply Theorem 2.1 (ii) with $\zeta = \rho_0 + \varepsilon$ to conclude that

$$\sum_h P(\rho_h - \rho_0 > \varepsilon) < \infty.$$

To estimate $P[\rho_h < \rho_0 - \varepsilon]$ note that

$$\{\rho_h < \rho_0 - \varepsilon\} = \{q \leq N_h \leq q + \lfloor (\rho_0 + \varepsilon)h \rfloor\},$$

since $N_h \geq q$ on $\{\rho_h < \rho_0 - \varepsilon\}$ by definition of ρ_h. We have

$$
\begin{aligned}
& P[q \leq N_h \leq q + \lfloor (\rho_0 - \varepsilon)h \rfloor] \\
& \quad \leq P\left[\max_{q \leq k \leq q + \lfloor (\rho_0-\varepsilon)h \rfloor} |\widehat{m}_{kh}| > c \right] \\
& \quad \leq \sum_{k=q}^{q+\lfloor (\rho_0-\varepsilon)h \rfloor} P\left[\left| \sum_{i=1}^{k} K_h(t_i - t_k)\varepsilon_i \right| > c - \left| \sum_{i=q}^{q+\lfloor (\rho_0-\varepsilon)h \rfloor} K_h(t_i - t_k) m_{ih} \right| \right].
\end{aligned}
$$

First note that

$$
c - \sum_{i=1}^{q} K_h(t_i - t_k) m_{ih} \to \int_{\rho_0 - \varepsilon}^{\rho_0} K(s - \rho_0) m_0(s)\, ds > 0.
$$

Further, $q \leq n \leq q + \lfloor (\rho_0 - \varepsilon)h \rfloor$ implies $n/h \to \rho_0 - \varepsilon$, as $n, h \to \infty$. An application of Theorem 2.1 (i) to each summand with $n = k$, noting that there are $O(h)$ summands and, of course, $k = O(h)$, we see that

$$
\begin{aligned}
P[q \leq N_h \leq q + \lfloor (\rho_0 - \varepsilon)h \rfloor] &= O(h\sqrt{h} e^{-c_2\sqrt{h}}) + O(h^2 \alpha(\sqrt{h})) + O(h\sqrt{h} e^{-c_1 h}) \\
&= o(1),
\end{aligned}
$$

as $h \to \infty$, provided $k^4 \alpha(k) = o(1)$. Further, $\sum_k k^4 \alpha(k) < \infty$ implies

$$
\sum_h P[\rho_h - \rho_0 < -\varepsilon] < \infty
$$

yielding complete convergence,

$$
\sum_h P[|\rho_h - \rho_0| > \varepsilon] < \infty, \quad \text{for every } \varepsilon > 0,
$$

which implies a.s. convergence (e.g. Karr (1993), Prop. 5.7). □

We close this section with a brief discussion of more general time designs. For some applications it may be possible and reasonable to determine at each time point the time points $t_{n1}, \ldots, t_{nn}$ where observations are taken. For example, one may start with monthly observations and reduce the distance between successive observations to ensure that the most recent data points are daily measurements. Note that such a thinning effect can not be obtained by a smoothing kernel.

Remark 2.3. *Assume F_T is a d.f. with support $[0,1]$ possessing a density f_T. Suppose at the n-th time point we may select the time points where observations are taken. We assume that*

$$t_{ni} = nF_T^{-1}(i/n), \qquad i = 1, \dots, n.$$

Clearly, the choice $t_{ni} = i$ corresponds to the uniform distribution $F_T(s) = s$, $s \in [0,1]$. When using skewed time designs, we can ensure that more recent observations dominate the sample of size n. It is straightforward to check that the proofs of Theorem 2.1 and Theorem 2.2 also work for that choice of time points. In this case we obtain

$$\begin{aligned} \sum_{i=1}^n K_h(t_{ni} - t_{nn}) m(t_{ni}/h) &\to \int_0^\zeta K(\zeta(F_T^{-1}(s/\zeta) - 1)) m_0(F_T^{-1}(s/\zeta))\, ds \\ &= \zeta \int_0^1 K(\zeta(s-1)) m_0(s) f_T(s)\, ds, \end{aligned}$$

if $n/h \sim \zeta$, yielding

$$\rho_h \stackrel{a.s.}{\to} \inf\left\{ \rho > 0 : \rho \int_0^1 K(\rho(s-1)) m_0(s) f_T(s)\, ds = c \right\},$$

as $h \to \infty$, under the conditions of Theorem 2.2.

3. Optimal kernels

The result of the previous section suggests the following kernel selection procedure. Suppose we are given a finite set $\{K_1, \dots, K_M\}$ of candidate kernels. Then we may choose the kernel which minimizes the corresponding asymptotic normed delay ρ^*. For an example where this selection rule was successfully applied to a real data set see Steland (2002c).

However, the natural question arises how to optimize the asymptotic normed delay with respect to the smoothing kernel K. It turns out that for the setting studied in this paper a meaningful result can be obtained. The canonical solution of the functional optimization problem can be given explicitly for arguments not exceeding the asymptotic optimal delay. Indeed, the optimal kernel is equal to a composition of the generic alternative and a time-reversal transformation which depends on the optimal asymptotic normed delay.

Although in this paper we assume that m is continuous at $t = t_q$, let us briefly discuss the discontinuous classical change-point model given by $m(s) = a\mathbf{1}(s \geq t_q), a > 0$ for all $s \in \mathcal{T}$. Brodsky and Darkhovsky (1993, Th. 4.2.8) have shown that the normed delay of the regular stopping rule (2) converges with probability 1, i.e., $\rho_h \overset{a.s.}{\to} \rho_0$, as $h \to \infty$, where the constant ρ_0 is given by

$$\rho_0 = \inf \left\{ \rho : \int_0^\rho K(s) ds = c/a \right\}.$$

If $F_K(z) = \int_{-\infty}^z K(s)\, ds,\ z \in \mathbb{R}$, denotes the associated distribution function, we have the explicit solution $\rho_0(K) = F_K^{-1}(1/2 - c/a)$. It is easy to show that for every $c > 0$ there exists a symmetric kernel with unit variance and bounded support such that the functional ρ_0 vanishes.

Therefore, in the sequel we assume that m_0 is a non-constant function. It will turn out that we now obtain solutions with non-vanishing optimal asymptotic normed delay. This allows to define an ordering relation on the set of admissible generic alternatives by comparing the optimal asymptotic normed delays. Anticipating the relationship between the optimal kernel and m_0, we assume that m_0 is Lipschitz continuous. For the optimality result of this section we also have to assume the following stronger regularity assumptions on the class of admissible kernels.

(K1) $\mathcal{K}$ is a class of uniformly Lipschitz continuous probability densities with Lipschitz constant L, i.e.,

$$\sup_{K \in \mathcal{K}} |K(z_1) - K(z_2)| \leq L|z_1 - z_2| \qquad (z_1, z_2 \in \mathbb{R}).$$

(K2) The class $\mathcal{K}$ is uniformly bounded, i.e.,

$$\|\mathcal{K}\|_\infty = \sup_{K \in \mathcal{K}} \|K\|_\infty \leq C_{\mathcal{K}} < \infty$$

holds true for some constant $C_{\mathcal{K}}$.

Define the mapping $I : \mathcal{K} \times [0, \infty) \to \mathbb{R}$,

$$I(K, \rho) = \int_0^\rho K(s - \rho) m_0(s) ds,$$

and denote by

$$\mathcal{R}(\rho) = \{ I(K, \rho) : K \in \mathcal{K} \}$$

the *reachable set at time* ρ. It is clear that $\mathcal{R}(\rho)$ is closed when $\mathcal{K}$ is equipped with the uniform topology induced by the supnorm.

Definition 3.1. *A pair* $(K^*, \rho^*) \in \mathcal{K} \times [0, \infty)$ *is* optimal, *if*

$$\rho^* = \inf\{\rho : c \in \mathcal{R}(\rho)\}$$

and K^* *ensures that*

$$I(K^*, \rho^*) = c.$$

For fixed $K \in \mathcal{K}$ define

$$\Psi(\rho) = I(K, \rho) = \int_0^\rho K(s - \rho) m_0(s)\, ds.$$

We will assume that there exists a positive $R \in \mathbb{R}$ such that Ψ is a strictly increasing function on $[0, R)$. Note that Ψ is continuous since K is Lipschitz continuous by assumption. We have the following theorem on the existence of optimal kernels.

Theorem 3.1. *Assume there exists* $K_1 \in \mathcal{K}$ *and some* $\rho_1 \geq 0$ *with*

$$I(K_1, \rho_1) = \int_0^{\rho_1} K_1(s - \rho_1) m_0(s) ds = c.$$

Then there exists an optimal kernel $K^* \in \mathcal{K}$*, i.e.,*

$$I(K^*, \rho^*) = c$$

where $\rho^* = \inf\{\rho : c \in \mathcal{R}(\rho)\}$ *is the optimal asymptotic normed delay.*

Remark 3.1. *For many generic alternatives* m_0 *it should be a trivial task to verify the condition of Theorem 3.1 holds true.*

Proof. By assumption we have $c \in \mathcal{R}(\rho_1)$. Let $\rho^* = \inf\{\rho : c \in \mathcal{R}(\rho)\}$. Then $0 \leq \rho^* \leq \rho_1$. We shall show $c \in \mathcal{R}(\rho^*)$. Then, by definition of $\mathcal{R}(\rho^*)$, there exists an optimal kernel $K^* \in \mathcal{K}$ with $I(K^*, \rho^*) = c$. $c \in \mathcal{R}(\rho^*)$ is a consequence of the following continuity argument. There exists a non-increasing sequence $\{\rho_n\}$ with $\rho_n \to \rho^*$ and an associated sequence $\{K_n\} \subset \mathcal{K}$ with

$$I(K_n, \rho_n) = c \in \mathcal{R}(\rho_n).$$

Since $K_n \in \mathcal{K}$, $I(K_n, \rho^*) \in \mathcal{R}(\rho^*)$. We have

$$\begin{aligned}
&|I(K_n, \rho^*) - I(K_n, \rho_n)| \\
&= \left| \int_0^{\rho^*} K_n(s-\rho^*) m_0(s) ds - \int_0^{\rho_n} K_n(s-\rho_n) m_0(s) ds \right| \\
&= \left| \int_0^{\rho^*} [K_n(s-\rho^*) - K(s-\rho_n)] m_0(s)\, ds \right. \\
&\qquad \left. - \int_{\rho^*}^{\rho_n} K_n(s-\rho_n) m_0(s)\, ds \right| \\
&\le L|\rho^* - \rho_n| \int_0^{\rho^*} m_0(s)\, ds \\
&\qquad + \left[\int_{\rho^*}^{\rho_n} K_n^2(s-\rho_n)\, ds \right]^{1/2} \left[\int_{\rho^*}^{\rho_n} m_0^2(s)\, ds \right]^{1/2}.
\end{aligned}$$

Clearly, by assumption (K2), $\int_{\rho^*}^{\rho_n} K_n^2(s-\rho_n)\, ds \le C_{\mathcal{K}}^2 |\rho^* - \rho_n|$. Therefore,

$$|I(K_n, \rho^*) - I(K_n, \rho_n)| = o(1),$$

as $n \to \infty$, yielding $I(K_n, \rho^*) \to c$, as $n \to \infty$. Since $I(K_n, \rho^*) \in \mathcal{R}(\rho^*)$ for all n and since $\mathcal{R}(\rho^*)$ is closed, we obtain

$$c = \lim_{n\to\infty} I(K_n, \rho^*) \in \mathcal{R}(\rho^*).$$

□

The following Lemma provides an useful characterization of each optimal pair (K^*, ρ^*) and is crucial to calculate optimal kernels.

Lemma 3.1. *Assume $(\rho^*, K^*) \in [0, R] \times \mathcal{K}$ is optimal. Then*

$$\int_0^{\rho^*} K^*(s-\rho^*) m_0(s)\, ds = \sup_{K\in\mathcal{K}} \int_0^{\rho^*} K(\rho^* - s) m_0(s)\, ds$$

Proof. Assume there exists some $\widetilde{K} \in \mathcal{K}$ with

$$\int_0^{\rho^*} \widetilde{K}(\rho^* - s) m_0(s)\, ds > \int_0^{\rho^*} K^*(\rho^* - s) m_0(s)\, ds = c.$$

Since $\rho \mapsto \int_0^{\rho} \widetilde{K}(\rho - s) m_0(s)\, ds$ is strictly increasing and continuous for $\rho \in [0, \rho^*]$, there exists a ρ^{**} with $\rho^{**} < \rho^*$ such that

$$\int_0^{\rho^{**}} \widetilde{K}(\rho^{**} - s) m_0(s)\, ds = c,$$

implying that the pair (ρ^*, K^*) is not optimal which is a contradiction. $\square$

We are now in a position to formulate and prove the following result about the explicit representation of the optimal kernel for a given generic alternative.

Theorem 3.2. *In addition to the regularity assumptions of this section assume*

$$0 < \int_0^\infty m_0(s)\, ds < \infty$$

and that the set

$$\mathcal{S} = \left\{ \rho \geq 0 : \frac{\int_0^\rho m_0^2(s)\, ds}{\int_0^\infty m_0(s)\, ds} = c \right\}$$

is non-empty. Then the following conclusions hold true.

(i) *The optimal asymptotic normed delay is given by* $\rho^* = \inf \mathcal{S}$

(ii) *The optimal smoothing kernel* K^* *satisfies*

$$K^*(z) = \frac{m_0(\rho^* - |z|)}{2\int_0^\infty m_0(s)\, ds}, \qquad z \in [-\rho^*, \rho^*].$$

Proof. Let $K \in \mathcal{K}$ be an arbitrary candidate kernel. By the Cauchy–Schwarz inequality we have

$$\left| \int_0^{\rho^*} K(\rho^* - s) m_0(s)\, ds \right| \leq \left[\int_0^{\rho^*} K^2(\rho^* - s)\, ds \right]^{1/2} \left[\int_0^{\rho^*} m_0^2(s)\, ds \right]^{1/2}$$

with equality if and only if

$$K(\rho^* - s) = \lambda \cdot m_0(s), \qquad \forall s \in [0, \rho^*]$$

for some constant $\lambda \in \mathbb{R}$. Since $\int_0^\infty K(s) ds = 1/2$ and $m_0(s) = 0$ if $s < 0$,

$$\lambda^{-1} = 2 \int_0^\infty m_0(s)\, ds.$$

Therefore, since $\mathcal{K}$ is a symmetric class,

$$K^*(s) = \frac{m_0(\rho^* - |s|)}{2\int_0^\infty m_0(s)\, ds}, \qquad s \in [-\rho^*, \rho^*].$$

Finally, we obtain

$$c = \int_0^{\rho^*} K^*(\rho^* - s) m_0(s)\, ds = \frac{\int_0^{\rho^*} m_0^2(s)\, ds}{2\int_0^\infty m_0(s)\, ds}.$$

$\square$

4. Examples

Let us consider some special cases to illustrate the results.

Example 4.1. *For a truncated linear drift,*

$$m_0(t) = at\mathbf{1}_{[0,T]}(t),\ t \geq 0,$$

we obtain $\rho^* = \frac{3}{2}\frac{c}{a}$ *and*

$$K^*(z) = \frac{1}{T^2}\left(\frac{3c}{2a} - |z|\right), \qquad |z| \leq \frac{3c}{2a}$$

Example 4.2. *Assume the generic alternative is given by a truncated exponential drift*

$$m_0(t) = e^{\lambda t}\mathbf{1}_{[0,T]}(t), t \geq 0,$$

for some $\lambda \in \mathbb{R}$*, where* T *is a positive truncation constant. If* $T \geq \rho^*$*, we have*

$$c = \frac{\int_0^{\rho^*} m_0^2(s)\, ds}{\int_0^{\rho^*} m_0(s)\, ds} = e^{\lambda\rho^*} + 1.$$

Hence, the optimal asymptotic normed delay is given by

$$\rho^* = \frac{\ln(c-1)}{\lambda}.$$

The optimal kernel K^* *is given by*

$$K^*(z) = \frac{1}{2}\frac{e^{\lambda\rho^*}}{e^{\lambda T} - 1}\lambda e^{-\lambda|z|}.$$

Note that K^* *converges to the density of the Laplace distribution,* $(\lambda/2)e^{-\lambda|z|}$*, if* $\rho^* = T \to \infty$*. Hence, for exponential drifts exponential weighting schemes are asymptotically optimal in this sense.*

Example 4.3. *Usually, enzyme processes are described by the Michaelis-Menten framework. Exploiting the quasi-steady-state approximations, the enzyme kinetic can be summarized by the differential equation*

$$\frac{d[S]}{dt} = -\frac{v_{\max}[S]}{K_M + [S]}$$

with initial condition $[S](0) = [S_0]$*, where* $[S](t)$ *stands for the substrate concentration at time* t*,* K_M *denotes the Michaelis-Menten rate constant, and* $v_{\max}$ *is the maximal velocity.*

For further details we refer to Schnell and Mendoza (1997). The solution of the differential equation is given by

$$[S](t) = K_M W\left(\frac{[S_0]}{K_M}\exp\left(\frac{-v_{\max}t[S_0]}{K_M}\right)\right),$$

where W stands for Euler's omega function, the (principal branch of the) inverse of the function $x \mapsto x\exp(x)$ (Euler (1777), Corless et al. (1996)). The optimal kernel to detect the generic alternative

$$m_0(t) = (S_0 - [S])(t)\mathbf{1}_{[0,T]}(t)$$

is given by

$$K^*(z) = (S_0 - [S])(\rho^* - |z|)/C_S, \qquad |z| \le \rho^*.$$

Observing that $\int_a^b W(de^{\lambda}t)\,dt = \int_{de^{\lambda a}}^{de^{\lambda b}} W(y)y^{-1}\,dy$, $d, \lambda \in \mathbb{R}$, and using the formulas

$$\begin{aligned}\int_a^b W(y)/y\,dy &= \frac{1}{6}W(y)^2[3+2W(y)]\Big|_a^b,\\ \int_a^b W(y)^2/y\,dy &= \frac{1}{2}W(y)[2+3W(y)]\Big|_a^b,\end{aligned}$$

one may obtain explicit formulas for $C_S = 2\int_0^T (S_0 - [S])(t)\,dt$ and for the enumerator and denominator of the nonlinear equation

$$c = \frac{\int_0^{\rho^*} m_0^2(s)\,ds}{\int_0^{\rho^*} m_0(s)\,ds}.$$

Acknowledgments

The author is grateful to Prof. E. Rafajłowicz, Technical University of Wrocław, Poland, for helpful comments and a lecture on optimal kernel choice, and two anonymous referees for their valuable remarks.

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475 ,,Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

[1] Antoch J. and Jarušková M. (2002). On-line statistical process control, in: *Multivariate Total Quality Control, Foundations and Recent Advances*, ed. Lauro C., Antoch J., and Vinzi, V.E., Physica, Heidelberg.

[2] Antoch J., Hušková M., and Jarušková M. (2002). Off-line statistical process control, in: *Multivariate Total Quality Control, Foundations and Recent Advances*, ed. Lauro C., Antoch J., and Vinzi, V.E., Physica, Heidelberg.

[3] Bosq D. (1996). *Nonparametric Statistics for Stochastic Processes.* Springer, New York.

[4] Brodsky B. E. and Darkhovsky B. S. (1993). *Nonparametric Methods in Change-Point Problems.* Mathematics and its applications, **243**, Kluwer Academic Publishers, Dordrecht.

[5] Brodsky B. E. and Darkhovsky B. S. (2000). *Non-Parametric Statistical Diagnosis Problems and Methods.* Mathematics and its applications, **509**, Kluwer Academic Publishers, Dordrecht.

[6] Corless R.M., Gonnet G.H., Hare D.E., Jeffrey D.J., and Knuth D.E. (1996). On the Lambert W function. *Advances in computational mathematics*, **5**, 329-359.

[7] Chiu C.K., Glad I.K., Marron J. (1998). Edge-preserving smoothers for image processing (with discussion). *Journal of the American Statistical Association*, **93**, 526-541.

[8] Euler L. (1777). De formulis exponentialibus replicatis, Leonhardi Euleri Opera Omnia, Ser. 1, Opera Mathematica **15** (1927), 268-297.

[9] Frisén M. (2003). Statistical surveillance. Optimality and methods. *International Statistical Review*, **71**, 2, 403-434.

[10] Girshick M.A. and Rubin H. (1952). A Bayes approach to a quality control model. *Annals of Mathematical Statistics*, **23**, 1, 114-125.

[11] Grabovsky I., Horváth L., and Hušková M. (2000). Limit theorems for kernel-type estimators for the time of change. *Journal of Statistical Planning and Inference*, **89**, 25-56.

[12] Hart J.D. (1997). *Nonparametric Smoothing and Lack of Fit Tests*, Springer, New York.

[13] Härdle W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

[14] Hušková M. (1991). Sequentially adaptive nonparametric procedures, In *Handbook of Sequential Analysis*, Gosh, B. K., and Sen. P. K. (eds.), ch. 20, Marcel Dekker, New York.

[15] Hušková M. and Slaby A. (2001). Permutation tests for multiple changes. *Kybernetika*, **37**, 5, 605-622.

[16] Karr A.F. (1993). *Probability*, Springer, New York.

[17] Lee J. (1983). Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, **24**, 255-269.

[18] Lorden G. (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, **42**, 1897-1908.

[19] Moustakides G. V. (1986). Optimal stopping times for detecting a change in distribution. *Annals of Statistics*, **14**, 1379-1388.

[20] Page E.S. (1954). Continuous inspection schemes. *Biometrika*, **1**, 100-115.

[21] Page E.S. (1955). A test for a change in a parameter occuring at an unknown point. *Biometrika*, **42**, 523-526.

[22] Pawlak M. and Rafajłoiwzc R. (2000). Vertically weighted regression - A tool for nonlinear data analysis and constructing control charts, **84**, 367-388. *Journal of the German Statistical Association*, **47**, 367-388.

[23] Pawlak M. and Rafajłowicz E. (2001). Jump preserving signal reconstruction using vertical weighting. *Nonlinear Analysis*, **47**, 327-338.

[24] Petrov V.V. (1975). *Sums of independent random variables.* Transl. from the Russ. by A. A. Brown, Springer, Berlin.

[25] Ritov Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics*, **18**, 1464-149.

[26] Rue H., and Chu C.K., and Godtliebsen F., and Marron J.S. (2002). M-smoother with local linear fit. *Journal of Nonparametric Statistics*, **14**, 155-168.

[27] Schmid W. and Schöne A. (1997). Some properties of the EWMA control chart in the presence of autocorrelation. *Annals of Statistics*, **25**, 3, 1277-1283.

[28] Schmid W. and Steland A. (2000). Sequential control of non-stationary processes by nonparametric kernel control charts. *Journal of the German Statistical Association*, **84**, 3, 315-336

[29] Schnell S. and Mendoza C. (1997). Closed form solution for time-dependent enzyme kinetics. *Journal of Theoretical Biology*, **187**, 207-212.

[30] Siegmund D. (1985). *Sequential Analysis.* Springer, New York.

[31] Steland A. (2002a). Nonparametric monitoring of financial time series by jump-preserving estimators. *Statistical Papers*, **43**, 361-377.

[32] Steland A. (2002b). A Bayesian view on detecting drifts by nonparametric methods. *Economic Quality Control*, **17**, 177-186.

[33] Steland A. (2002c). Sequential credit risk: Can we benefit from sequential nonparametric control? *Selected Papers Operation Research Proceedings 2002*, 337-344.

[34] Volkonskii V. A. and Rozanov (1959). Some limit theorems for random functions. *Theory of Probability and its Applications*, **4**, 178-197.

[35] Yakir B. (1997). A note on optimal decision detection of a change in distribution. *Annals of Statistics*, **25**, 5, 2117-2126

[36] Yakir B., Krieger A. M., and Pollak M. (1999). Detecting a change in regression: First-order optimality. *Annals of Statistics*, **27**, 6, 1896-1913.