ECONSTOR

WWW.ECONSTOR.EU

Der Open-Access-Publikationsserver der ZBW – Leibniz-Informationszentrum Wirtschaft
The Open Access Publication Server of the ZBW – Leibniz Information Centre for Economics

Steland, Ansgar

Working Paper

NP-optimal kernels for nonparametric sequential detection rules

Technical Report // Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen, No. 2004,09

Provided in cooperation with:
Technische Universität Dortmund

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der
Leibniz-Gemeinschaft

NP-Optimal Kernels for Nonparametric Sequential Detection Rules

Ansgar Steland[1]

Fakultät für Mathematik

Ruhr-Universität Bochum, Germany

ansgar.steland@ruhr-uni-bochum.de

ABSTRACT. An attractive nonparametric method to detect change-points sequentially is to apply control charts based on kernel smoothers. Recently, the strong convergence of the associated normed delay associated with such a sequential stopping rule has been studied under sequences of out-of-control models. Kernel smoothers employ a kernel function to downweight past data. Since kernel functions with values in the unit interval are sufficient for that task, we study the problem to optimize the asymptotic normed delay over a class of kernels ensuring that restriction and certain additional moment constraints. We apply the key theorem to discuss several important examples where explicit solutions exist to illustrate that the results are applicable.

Keywords: Control charts, financial data, nonparametric regression, quality control, statistical genetics.

INTRODUCTION

In many areas of human activities processes are monitored continuously in time in order to detect a change in process behavior. The resulting data generating process is called time series and statistical methods for analyzing the sequentially arriving data are necessary for making a decision. Historically, the need to monitor the quality of industrial production processes gave rise to the development of appropriate statistical procedures. The early methods of statistical process control were designed without taking into account the sequential nature of data generating. For a recent introduction to the field see Antoch and Jarušková (2002). Later they have been refined to provide appropriate tools for sequential decision making in many areas. For example, capital markets produce huge streams of data consisting of prices, returns, etc. Financial analysts can apply and benefit from sequential methods to decide whether such time series are in accordance with an *in-control model* describing normal market behavior or whether there is a change to an *out-of-control model* which should trigger further actions. Sequential decision procedures for financial time series take into account the fact that these time series are in most cases correlated.

An interesting new field for sequential methods is the analysis of time series data obtained in the course of genome experiments, which are still quite expensive. For example, suppose from each experiment (case) we may calculate some quantity summarizing the evidence of a gene-related effect. There are approximately 30000 genes. Hence, there is a strong interest in collecting large data sets of genetic data for large samples of individuals in order to detect genes which are responsible for diseases as soon as possible. It is expected that new therapies can be developed for many cost intensive diseases as cancer. From an economic viewpoint it is therefore reasonable to apply sequential methods to such problems in order to reduce the costs.

In this article a nonparametric method for deciding about the change-point is proposed. The advantage of nonparametric methods is that they can be applied without stating explicitly a parametric probability distribution for the quantity of interest. Thus, they avoid the doubtful practice to determine a parametric model on the base of observed (random) data. A nonparametric method consists of a control statistic, which reflects (a functional of) the process mean in a nonparametric way, and a control limit (critical value). The control statistic is compared with the control limit to judge whether the data speaks in favor of the in-control or out-of-control model.

Recently, a couple of nonparametric results on the asymptotic normed delay of stopping times based on Nadaraya-Watson type estimators have been obtained, extending classic

results dating back to Brodsky and Darkhovsky (1993). Fortunately, the results hold true for a rich class of dependent time series. It was shown that under certain regularity assumptions the normed delay converges to a deterministic quantity, which depends on the smoothing kernel used to weight the data, and a function defining the out-of-control model. This result suggests to study the mathematical problem of optimal kernel selection in order to minimize the asymptotic normed delay. Related sequential versions of kernel-weighted averages have also been discussed by Brodksy and Darkhovsky (1993), Ferger (1994b, 1994c, 1995), among others. For U-statistic approaches see Ferger (1994) , Gombay (1995), Huskova (2002), and the references given in these articles.

Nadaraya-Watson type smoothers employ a kernel function which is used to downweight past data. Clearly, for that purpose weights from the unit interval $[0, 1]$ are sufficient. Therefore, in this article we aim to minimize the asymptotic normed delay when constraining the class of admissible kernels to weighting functions with values in $[0, 1]$ which satisfy certain (moment) constraints. It turns out that this problem is related to the classic problem of optimal hypothesis testing dating back to Neyman and Pearson (1933). The resulting NP-optimal kernel differs substantially from the unconstrained optimal kernel derived in Steland (2003a) where essentially the class of bounded densities has been considered. Indeed, for many out-of-control models of practical interest the optimal kernel is given by an uniform kernel, and can also be calculated explicitly. However, in general its support is not necessarily a connected set and is obtained by comparing the out-of-control model defining function with a certain polynomial.

The organization of the paper is as follows. Section 1 introduces the data generating process in detail. The monitoring procedure is explained in Section 2. A brief review of the normed delay and its asymptotic behavior under the setting of this article is provided in Section 3. In Section 4 we discuss the problem of NP-optimal kernel choice as outlined above and, finally, Section 5 provides some examples where NP-optimal kernels can be calculated explicitly using elementary calculus.

1. Model

We consider a model for the time series data which is parameterized by a real valued parameter h. That parameter will be used as a scaling parameter in the out-of-control model and as a sample size parameter in the control statistic. Here we follow the approach to study a *sequence* of out-of-control models. The sequence of out-of-control models for the process mean is derived from a defining function $m_0(t)$ corresponding to $h = 1$ and

has the property that the normed delay of the monitoring procedure converges, as h tends to infinity. To reach this goal, detecting the out-of-control model is made harder as the sample size increases. That approach, which is related to the concept of local alternatives in classic hypothesis testing, will allow us to study the performance properties analytically. The function $m_0(t)$ defining the of out-of-control model is assumed satisfy $m_0(t) = 0$ if $t \leq 0$ and $m_0(t) > 0$ if $t > 0$. For each $h > 0$ we consider the out-of-control model

$$m_h(t) = m_0 \left(\frac{t - t_q}{h} \right) \begin{cases} = 0, & t \leq t_q, \\ > 0, & t > t_q, \end{cases}$$

for the process mean. Here t_q is the change-point in continuous time. The stochastic model for the process in continuous time is now given by

$$Y_h(t) = m_h(t) + \epsilon(t)$$

where $\epsilon(t)$ denotes the random error term with expectation 0, i.e., $E(\epsilon(t)) = 0$ for each t.

The process $Y(t)$ is now sequentially observed at discrete and equidistant time points $t_n = n$, $n = 1, 2, \ldots$, yielding the sequentially arriving observations $Y_{n,h} = Y_h(n)$, $n = 1, 2, \ldots$ For many applications the assumption that the process is observed at equidistant time points is not restrictive. Otherwise, one may use more general time designs as proposed by Steland (2003a) or apply re-sampling schemes as considered by Giot (2000) for financial data.

The model equation for the observations $Y_{n,h}$ takes the form

$$Y_{n,h} = m_{n,h} + \epsilon_n, \qquad n \in \mathbb{N},$$

where

$$m_{n,h} = m_h(n) = m_0([n - q]/h)$$

and $q = \lfloor t_q \rfloor + 1$ is the change-point in discrete time. $\epsilon_n = \epsilon(n)$ is a time series of error terms with expectation 0 for each n. We assume that $\{\epsilon_n\}$ is strictly stationary and α-mixing. Let us briefly recall the notion of α-mixing time series, which is a common approach to deal nonparametrically with correlated time series. It allows to obtain results which are not restricted to a certain time series model. Suppose we are given two events A and B which are determined by the time series. Let k denote the lag between these events. If, for example, $A = \{\epsilon_3 \leq 0\}$ and $B = \{\max_{i \geq 8} \epsilon_i > 1\}$, k equals 5. If we approximate the true probability $P(A \cap B)$ by $P(A)P(B)$, which assumes that A and B are independent, we make an error, namely $|P(A \cap B) - P(A)P(B)|$. A time series is called α-mixing, if that error is not larger than $\alpha(k)$ and if $\alpha(k)$ converges to 0, as k tends to ∞. This means, the

dependence between events becomes negligible if the the distance is very large. Formally, $\alpha(k)$ is defined by

$$\alpha(k) = \sup_{A \in \mathcal{F}_0^n, B \in \mathcal{F}_{n+k}^\infty} |P(A \cap B) - P(A)P(B)|,$$

where $\mathcal{F}_i^j = \sigma(X_i, \ldots, X_j)$ denotes the σ-field (information set) of the random variables $X_i, \ldots X_j$. We assume that $\{\epsilon_n\}$ is α-mixing with mixing coefficients $\{\alpha(k) : k \in \mathbb{N}\}$ satisfying

$$(1) \qquad\qquad \sum_{k=0}^\infty k^2 \alpha(k) < \infty.$$

This assumption is not restrictive. In many cases the sequence of mixing coefficients satisfies $\alpha(k) \le Cq^{-\beta}$ for some $q \in (0, 1)$ and some $\beta > 0$ which implies (1). Indeed, many time series models, in particular Markovian processes, ARMA models and GARCH models, satisfy assumption (1) under fairly weak conditions on the model parameters. Therefore, working with mixing conditions as (1) has the advantage that the results are valid for a rich class of time series covering many classic models as special cases. For a detailed discussion of mixing coefficients and their properties we refer to Bosq (1996).

At first glance it seems to be restrictive to consider only a change-point model for monitoring the process mean. However, many other monitoring problems can be transformed appropriately and solved by applying our monitoring procedure for the mean. We briefly sketch some important problems, which are worth to be investigated in greater detail.

Example 1.1. *In financial econometrics it is important to detect changes in the dispersion of financial returns $\{R_t\}$, since the dispersion determines the risk associated with an investment in the underlying asset. In order to detect risk increases one may monitor the sequence of centered squared returns, $Y_t = R_t^2 - E_0(R_t^2)$, or the centered absolute value of the returns, i.e., $Y_t = |R_t| - E_0|R_t|$. Here E_0 means that the expectation is calculated assuming the in-control model is true. When considering returns over very short time horizons, e.g., minutes, also called* high-frequent *data, intraday seasonality can be a severe problem. Intraday seasonality, i.e., a fixed temporal pattern of the dispersion due to events as opening or closing of certain stock exchanges, can be taken into account by analyzing $R_i/\phi(s_i)$, where $\phi(s_i)$ is the deterministic intraday seasonal component for time s_i, which is defined as the expected volatility conditioned on time-of-day (Andersen and Bollerslev, 1997).*

Example 1.2. *Suppose we want to monitor a time series $\{X_n\}$ with mean 0 and common variance $\sigma^2 = Var(X_n)$ in order to detect a change in the lag-1 autocorrelation. In this*

case one may monitor the sequence $Y_n = X_n X_{n-1}$, $n \geq 1$, since $E(Y_n) = \gamma_n(1)$ is the lag-1 autocovariance. An appropriate change-point model is to assume that $\gamma_n(1) = \gamma_0(1)$ for all $n < q$ and $\gamma_n(1) = \gamma_0(1) + \Delta$ for $n \geq q$. If $\Delta \neq 0$, q is the change-point where the autocovariance increases. This means, by monitoring $X_n X_{n-1}$ for a change in the mean we may solve the problem.

Example 1.3. *Assume we aim to detect changes in the mean or variance of a time series $\{X_n\}$ using an univariate control scheme. Let μ_0 and σ_0^2 denote the in-control mean and variance, respectively. Suppose the out-of-control model states that mean and variance are given by μ_1 and σ_1^2. In this case we can monitor the sequence $Y_n = (Y_n - \mu_0)^2 / \sigma_0^2$. The in-control mean of Y_n equals 1, whereas the expectation of Y_n under the out-of-control model is given by $\frac{\sigma_1^2}{\sigma_0^2} + \frac{(\mu_1 - \mu_0)^2}{\sigma_0^2}$.*

2. Monitoring Procedure

The monitoring procedure consists of a control statistic which is calculated at each time point n using past and current observations. Given a smoothing kernel $K(z)$ define

$$(2) \qquad \widehat{m}_{n,h} = \sum_{i=1}^{n} K_h(i - n) Y_{ih},$$

where $K_h(z) = h^{-1} K(z/h)$ denotes the rescaled kernel. $K(z) \geq 0$ is used as a weighting function to downweight observations which are far away from the current observation Y_n. To ensure the downweighting property $K(z)$ is usually chosen as a unimodal function satisfying $K(z_1) \geq K(z_2)$, if $z_1 \leq z_2$. We may assume that $K(z)$ is symmetric around 0. Further, typically one uses densities with mean 0, variance 1. Then the rescaled version $K_h(z)$ has mean 0 and variance h^2. In particular, if $K(z)$ has support $(-1, 1)$, i.e., $K(z) = 0$ if $|z| \geq 1$, the most recent h observations are used to calculate the control statistic. Thus, we will call h the *effective sample size*. In the more general case that $K(z)$ has support $(-a, a)$, the number of observations actually employed by the control statistic is ha. In both cases, h measures the effective sample size (up to a constant).

At each time point the control statistic is now compared with a control limit c, and a signal is given at the first time point where the control limit is exceeded. The corresponding stopping time is defined by

$$N_h = \inf\{n \in \mathbb{N} : \widehat{m}_{n,h} > c\}.$$

The stopping time is used in practice to trigger further actions. For an example where credit spreads are monitored using this approach see Steland (2002b).

Finally, define the normed delay as

$$\rho_h = \frac{\max(N_h - q, 0)}{h}.$$

Note that $100 \cdot \rho_h$ expresses the delay of the procedure as a percentage of the effective sample size.

3. A Review of the Asymptotic Normed Delay

Under regularity assumptions discussed to some extent below it is known that ρ_h converges to a functional of K and m_0, as the parameter h tends to ∞. For simple shift functions this problem has been studied by Brodsky & Darkhovsky (1993). For general sequences of out-of-control models almost sure convergence under the mixing condition (1) has been shown by Steland (2003a) for kernel smoothers as (2). Results about its asymptotic distribution have been established by Steland (2003b). Related results for jump-preserving methods as proposed by Pawlak and Rafajłowicz (2000, 2001) can be found in Steland (2002a). Jump-preserving monitoring schemes use a control statistic which is designed to react quickly to (large) shifts. Due to their averaging property, control procedures as the CUSUM or EWMA scheme, or the nonparametric method studied here, have the unpleasant property that large shifts are smoothed away to some degree, in particular if they last only for a few time points. Jump-preserving control statistics use only those data points to calculate a location statistic, whose values are similar to the current observation Y_n. It has been shown that under quite general conditions such detection rules are able to detect shifts with no delay, whereas for procedures based on arithmetic means or weighted averages as the EWMA control chart or the kernel control chart studied here, there is always a positive probability that a shift is not detected immediately, even for very large shifts. Simulations indicate that jump-preserving charts can be preferable when one is interest in detecting large shifts as soon as possible, in particular if they last only for a few time points. For details we refer to Pawlak, Rafajłowicz and Steland (2003) and the references given there.

For the setting considered in this article, in Steland (2003a) it has been shown that

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

as $h \to \infty$, where ρ_0 is defined by

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

7

provided the mixing condition (1) is satisfied, the kernel K is Lipschitz continuous, and the pair (K, m_0) satisfies the integrability condition

$$(4) \qquad I(x) = \int_0^x K(s - x) m_0(s) \, ds < \infty, \qquad (\forall x > 0),$$

Clearly, (4) holds true, if $m_0, K \in L_1$.

For jump-preserving monitoring procedures as discussed above an upper bound for ρ_h has been established, which is obtained from ρ_0 by modifying the control limit c. For details see Steland (2002a, 2002c) and the references given there. Therefore, the results of this article also apply to jump-preserving control charts in the sense that the upper bound is minimized.

4. NP-Optimal Kernels

Let us now discuss the question how to choose the kernel K defining the weighting scheme of the sequential stopping rule in order to minimize the asymptotic normed delay (3) under the constraint $0 \leq K(z) \leq 1$ and additional (moment) constraints. In this case the search for optimal kernels can be reduced to an optimization problem similar as the Neyman-Pearson Lemma of optimal hypothesis testing, thus giving rise to the notion of NP-optimality.

To formulate the formal optimization problem and its solution we need some further notation. Let $\text{Lip}\,(\mathbb{R})$ denote the class of Lipschitz continuous functions on the real line. Define the following classes of kernels.

$$\widetilde{\mathcal{K}} = \{K \in L_1(\mathbb{R}) : K \in \text{Lip}\,(\mathbb{R}), \; K(s) = K(-s) \; \forall s \in \mathbb{R}, \|K\|_\infty < \infty\}$$

$$\mathcal{K} = \{K \in L_1(\mathbb{R}) : K \in \text{Lip}\,(\mathbb{R}), \; K(s) = K(-s) \; \forall s \in \mathbb{R}, 0 \leq K \leq 1\}.$$

For kernels from the class $\widetilde{\mathcal{K}}$ the problem to optimize $\rho_0 = \rho_0(K)$ w.r.t. K for a given m_0 has been studied in Steland (2003a). Note that optimization over $\widetilde{\mathcal{K}}$ is an essentially unconstrained problem. Indeed, the symmetry condition can be dropped, since in principle the kernel K used for $\widehat{m}_{n,h}$ can be chosen as a one-sided kernel. But the constraint $0 \leq K \leq 1$, which suffices to downweight observations far away from the current one, yields a substantially different solution.

To define the notion of NP-optimality let us introduce some additional notation. Denote by

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

8

the *reachable set at time ρ* where

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

Note that if a control limit c is an element of the reachable set, i.e., $c \in R(\rho)$, there is a kernel $K \in \mathcal{K}$ so that ρ is the corresponding asymptotic normed delay.

Definition 1. *Let c be a given control limit. A pair $(K^*, \rho^*) \in \mathcal{K} \times [0, \infty)$ is called* NP-optimal, *if*

$$\rho^* = \inf\{\rho > 0 : c \in R(\rho)\} \qquad and \qquad I(K^*, \rho^*) = c.$$

In this situation K^* is called NP-optimal kernel *and ρ^* optimal asymptotic normed delay.*

Crucial to the following investigation is the following criterion for NP-optimality.

Lemma 1. *(Steland 2003a, 3.1). If (K^*, ρ^*) is NP-optimal, we have*

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

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

where the second formula follows by integration by parts and symmetry of K. □

Thus, to find optimal pairs (K^*, ρ^*) over the class $\mathcal{K}$ under additional (moment) constraints we consider the following problem.

(P) For $\rho^* = \inf\{\rho > 0 : c \in R(\rho)\}$ maximize the functional

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

over the class $\mathcal{K}$ under the additional (moment) constraints

$$(5) \qquad \int_{-\infty}^{\infty} K(s) \, ds \;=\; 1,$$

$$(6) \qquad \int_{-\infty}^{\infty} s K(s) \, ds \;=\; 0,$$

$$(7) \qquad \int_{-\infty}^{\infty} s^2 K(s) \, ds \;=\; 1,$$

9

The problem (P) can be solved using the same techniques from linear functional analysis which are also applied to prove the Neyman-Pearson Lemma on optimal hypothesis testing, see Neyman and Pearson (1933), Lehmann (1986), and Rao (1973). A proof of the following theorem is given in the appendix.

Theorem 1. *NP-optimal kernels K^* of the optimization problem (P) satisfy*

$$K^*(s) = \begin{cases} 0, & m(\rho^* - s) < k_1 + k_2 s + k_3 s^2, \\ 1, & m(\rho^* - s) > k_1 + k_2 s + k_3 s^2, \\ a, & m(\rho^* - s) = k_1 + k_2 s + k_3 s^2, \end{cases}$$

for $s \in [0, \rho^*]$, where $a \in (0, 1)$. The constants k_1, k_2, and k_3 are choosen to ensure that the contraints (5), (6), and (7) are satisfied. ρ^* satisfies the equation

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

The theorem asserts that a NP-optimal kernel is essentially an indicator function $\mathbf{1}_A$ of a set $A = A(m_0)$ whose shape strongly depends on the generic alternative m_0. A is obtained by comparing a translated version of the generic alternative with a polynomial of degree 2. Therefore, optimal kernels are not necessarily indicator functions on a connected set.

Clearly, the result can be formulated for an arbitrary number of moment conditions. For sake of illustration let us consider the case where the optimization is only constrained by the (moment) condition

$$\int_{-\infty}^{\infty} K(s)ds = 1.$$

In this case NP-optimal kernels can be chosen as

(8)
$$K^*(s) = \begin{cases} 1, & m_0(\rho^* - s) > k_1, \\ 0, & m_0(\rho^* - s) \le k_1. \end{cases}$$

for $s \in [-\rho^*, \rho^*]$, where ρ^* is a solution of the nonlinear integral equation

(9)
$$\int_{[0,\rho^*] \cap \{s:m_0(\rho^*-s)>k_1\}} m_0(\rho^* - s)\, ds = c.$$

Whereas for a given defining function m_0 (8) and known ρ^* can be even solved graphically, it can be hard to solve equation (9) for to obtain ρ^* for a given control limit.

5. Examples

The assertion of the theorem, which solves the functional optimization problem, is quite involved, even for the slightly simplified problem given by equations (8) and (9). But for many explicit choices of m_0 the NP-optimal kernel and the optimal asymptotic normed delay can be calculated explicitly using elementary calculus. For illustration we provide three basic examples.

We start with a unit jump which decays exponentially fast.

Example 5.1. *Let*

$$m_0(t) = e^{-t}, \qquad t > 0.$$

Clearly,

$$m_0(\rho^* - t) > k_1 \Leftrightarrow t > \rho^* + \log(k_1).$$

Anticipating that $k_1 \in (0, 1)$, we obtain

$$K^*(t) = \begin{cases} 1, & \text{if } \rho^* + \log(k_1) \leq |t| \leq \rho^*, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\int_{-\infty}^{\infty} K^*(s)\, ds = -2\log(k_1)$ does not depend on ρ^*, the choice

$$k_1 = e^{-\frac{1}{2}}$$

ensure that K^* is a probability density. Furthermore,

$$\int_0^{\rho^*} K(\rho^* - s) m_0(s)\, ds = \int_{\rho^* + \log(k_1)}^{\rho^*} e^s\, ds = c$$

is satisfied if

$$\rho^* = \log \frac{c}{1 - k_1} \approx \log(c) + 0.932752$$

Next let us look at a simple $\cap$-shaped model for $m_0(t)$ located at $t = 1$.

Example 5.2. *Define $m_0(t) = 1 - (t - 1)^2$ if $t \in [0, 2]$, and $= 0$ otherwise. For $k_1 \in (0, 1)$ we have*

$$m_0(\rho^* - t) > k_1 \Leftrightarrow \rho^* - 1 - \sqrt{1 - k_1} < t < \rho^* - 1 + \sqrt{1 - k_1}.$$

Thus,

$$K^*(t) = \begin{cases} 1, & |t| \in [\rho^* - 1 - \sqrt{1 - k_1}, \rho^* - 1 + \sqrt{1 - k_1}], \\ 0, & \text{else,} \end{cases}$$

11

For $k_1 = 7/9$ corresponding to

$$K^*(t) = \begin{cases} 1, & |t| \in [\rho^* - 1 - 1/\sqrt{8}, \rho^* - 1 + 1/\sqrt{8}], \\ 0, & else. \end{cases}$$

we obtain $\int_{-\infty}^{\infty} K^*(s)\, ds = 1$. Further,

$$I(\rho^*) = \int_0^{\rho^*} K(\rho^* - s) m_0(s)\, ds = \int_0^{\rho^*} \mathbf{1}_{[1-1/\sqrt{8}, 1+\sqrt{8}]}(s)(1 - s^2)\, ds$$

is increasing for $\rho^* \in (1 - 1/\sqrt{8}, 1]$, and decreasing for $\rho^* \in (1, 1 + 1/\sqrt{8}]$ with maximal value $I_{\max} = \frac{12 - \sqrt{2}}{96}$. In particular, for $1 - 1/\sqrt{8} \le \rho^* \le 1 + 1/\sqrt{8}$ we have

$$I(\rho^*) = -\frac{13}{24} - \frac{1}{48\sqrt{2}} + r - \frac{r^3}{3}.$$

Thus, for $c \in (0, I_{\max}]$ we may solve $I(\rho^*) = c$ for ρ^* as shown in Figure 1. An analytic formula can also be obtained.

Finally, let us consider a Gaussian model for m_0 with two parameters $\Delta \in \mathbb{R}$ and $\tau > 0$. In this case the calculations are more involved but still elementary. It turns out that the asymptotic optimal normed delay does not depend on the parameter τ.

Example 5.3. *Define*

$$m_0(t) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{(t-\Delta)^2}{2\tau^2}}, \qquad t \ge 0,$$

where $\Delta \in \mathbb{R}^+$ *and* $\tau > 0$. *For* $k_1 < 1/\sqrt{2\pi\tau^2}$ *the inequality*

$$m_0(\rho^* - t) > k_1$$

is equivalent to

$$t \in I(\rho^*, \Delta, k_1, \tau) = [\rho^* - \Delta - i(\Delta, k_1), \rho^* - \Delta + i(\Delta, k_1)]$$

where

$$i(k_1, \tau) = \sqrt{2}\sqrt{\tau^2 \ln(\sqrt{2\pi\tau^2}k_1)}$$

Thus,

$$K^*(t) = \begin{cases} 1, & |t| \in I(\rho^*, \Delta, k_1, \tau), \\ 0, & |t| \notin I(\rho^*, \Delta, k_1, \tau). \end{cases}$$

Note that $i(k_1, \tau)$ *is injective in* τ *for* $\tau > 0$ *and* $k_1 < 1/\sqrt{2\pi\tau^2}$. *Hence, since* $|I(\Delta, k_1, \tau)| = 2i(k_1, \tau)$, *for given* τ *we can solve*

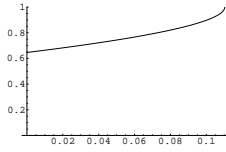$$\int_{-\infty}^{\infty} K^*(s)\, ds = 4i(k_1, \tau)$$

FIGURE 1. Optimal asymptotic normed delay ρ^* as a function of the threshold c.

for $k_1 = k_1(\tau)$ yielding

$$k_1(\tau) = \frac{e^{-32\tau^2}}{\sqrt{2\pi\tau^2}}.$$

This implies

$$i(k_1(\tau), \tau) = \sqrt{2}\sqrt{\tau^2 \log \exp(-32\tau^2)} = \frac{1}{4}.$$

Finally,

$$
\begin{aligned}
\int_0^{\rho^*} K^*(\rho^* - s)m_0(s)ds &= \int_0^{\rho^*} \mathbf{1}_{[\rho^*-\Delta-i(k_1(\tau),\tau),\rho^*-\Delta+i(k_1(\tau),\tau)]}(s)m_0(s)\,ds \\
&= \int_0^{\rho^*} \mathbf{1}_{[\Delta-1/4,\Delta+1/4]}(s)m_0(s)\,ds \\
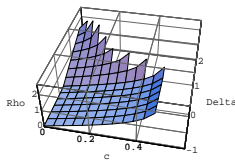&= \Phi(\min(\rho^*, \Delta + 1/4)) - \Phi(\max(0, \Delta - 1/4)))
\end{aligned}
$$

13

FIGURE 2. Optimal asymptotic normed delay ρ^* as a function of c and Δ.

where Φ denotes the cumulative distribution function of the standard normal distribution. Consequently, if $\rho^* \leq \Delta + 1/4$ we have to solve

$$\Phi(\rho^*) - \Phi(\max(0, \Delta - 1/4)) = c$$

which yields

$$\rho^* = \Phi^{-1}(\Phi(\max(0, \Delta - 1/4)) + c).$$

provided $0 \leq \Phi(\max(0, \Delta - 1/4)) + c \leq 1$. In particular, the optimal asymptotic normed delay does not depend on the scale parameter τ of the alternative. The solution is shown in Figure 2. Note that for small values of Δ and small control limits ρ^* can be reasonably approximated by a linear function of these variabels, but for larger values the shape is strongly nonlinear.

14

APPENDIX: PROOF OF THEOREM 1

First note that for any $K \in \mathcal{K}$ with (5), (6), and (7) by linearity we have

$$(10) \qquad \int_{-\infty}^{\infty} (\sum_{i=1}^{2} k_i s^i) K(s)\, ds = \sum_{i=0}^{2} \int_{-\infty}^{\infty} s^i K(s)\, ds.$$

Clearly, a kernel $K^* \in \mathcal{K}$ is NP-optimal if

$$\int_{0}^{\rho^*} m_0(\rho^* - s) K^*(s)\, ds \geq \int_{0}^{\rho^*} m_0(\rho^* - s) K(s)\, ds$$

for all $K \in \mathcal{K}$ satisfying (5), (6), and (7). Equivalently, NP-optimality holds true if

$$\int_{0}^{\rho^*} m_0(\rho^* - s)[K^*(s) - K(s)]\, ds \geq 0$$

for all $K \in \mathcal{K}$ with (5), (6), and (7). Observe that for $s \in [0, \rho^*]$ with $K^*(s) = 1$ we have

$$m(\rho^* - s) > \sum_{i=0}^{2} k_i s^i, \quad \text{and} \quad K^*(s) \geq K(s),$$

which implies

$$(11) \qquad [m(\rho^* - s) - \sum_{i=0}^{2} k_i s^i][K^*(s) - K(s)] \geq 0.$$

Similarly, for $s \in [0, \rho^*]$ with $K^*(s) = 0$ we have

$$m(\rho^* - s) < \sum_{i=0}^{2} k_i s^i, \quad \text{and} \quad K^*(s) \leq K(s),$$

15

again yielding (11). If $m(\rho^* - s) = \sum_{i=0}^{2} k_i s^i$, (11) holds true for any $a \in (0, 1)$. Consequently, using (10) we can conclude

$$\int_0^{\rho^*} m(\rho^* - s)[K^*(s) - K(s)]\,ds$$
$$= \int_0^{\rho^*} [m(\rho^* - s) - \sum_{i=0}^{2} k_i s^i][K^*(s) - K(s)]\,ds$$
$$\geq 0$$

for all $K \in \mathcal{K}$ satisfying (5), (6), and (7). Hence, (K^*, ρ^*) is NP-optimal. $\square$

REFERENCES

[1] Andersen, T. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, **4**, 115-158.

[2] Antoch J. and Jarušková D. (2002). On-line statistical process control. In: *Multivariate Total Quality Control, Foundations and Recent Advances.* Eds: Lauro C., Antoch J., and Vinzi V.E. Physica.

[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] Ferger, D. (1994a). On the power of nonparametric change-point tests. *Metrika*, **41**, 277-292.

[6] Ferger, D. (1994b). An extension of the Csörgö-Horvath functional limit theorem and its applications to changepoint problems. *Journal of Multivariate Analysis*, **51**, 2, 338-351.

[7] Ferger, D. (1995). Nonparametric tests for nonstandard change-point problems. *The Annals of Statistics*, **23**, 5, 1848-1861.

[8] Giot, P. (2000). Time transformations, intraday data and volatility models. *Journal of Computational Finance*, **4**, 31-62.

[9] Gombay, E. and Horvath, L. (1995). An application of the U-statitics to change-point analysis. *Acta Sci. Math (Szeged)*, **60**, 345-357.

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

[11] Lehmann, E. (1986). *Testing Statistical Hypotheses.* 2nd Ed. John Wiley and Sons, New York.

[12] Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc.*, **A 231**, 289-337.

[13] Pawlak, M. and Rafajłowicz, E. (2000). Vertically weighted regression - a tool for nonlinear data analysis and constructing control charts. *Journal of the German Statistical Society*, **84**, 367 - 388.

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

[15] Pawlak, M., Rafajłowicz, E., and Steland, A. (2003). Detecting jumps in time series - Nonparametric setting. *Journal of Nonparametric Statistics*, to appear.

[16] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[17] 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.

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

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

[20] Steland, A. (2002c). A Bayesian view on detecting drifts by nonparametric methods. *Economic Quality Control*, **17**, 2, 177-186.

[21] Steland, A. (2003a). Optimal sequential kernel smoothers under local nonparametric alternatives for dependent processes. *Journal of Statistical Planning and Inference*, under revision.

[22] Steland, A. (2003b). Sequential control of time series by functionals of kernel-weighted empirical processes under local alternatives. *Metrika*, to appear.