Constancy of distributions: nonparametric monitoring of probability distributions over time

Nils Lid Hjort
Department of Mathematics and Statistics
University of Oslo
P.O. Box 1053 Blindern
N-0316 Oslo 3
Norway
nils@math.uio.no

Alex J. Koning
Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738
NL-3000 DR Rotterdam
The Netherlands
koning@few.eur.nl

Econometric Institute Report EI 2001-50

Abstract

In this paper we study stochastic processes which enable monitoring the possible changes of probability distributions over time. These processes may in particular be used to test the null hypothesis of no change. The monitoring processes are bivariate functions, of time and position at the measurement scale, and are approximated with zero mean Gaussian processes under the constancy hypothesis. One may then form Kolmogorov–Smirnov or other type of tests as functionals of the processes. To study null distributions of the resulting tests, we employ KMT-type inequalities to derive Cramér-type deviation results for (bootstrapped versions of) such tests statistics.

1 Introduction and summary

Assume that independent data are available for each of n consecutive occasions, perhaps measurements of some quantity taken on separate dates. The null hypothesis to be tested here is that of

$$H_0 \ : \ F_1 = F_2 = \ldots = F_n, \tag{1}$$

where F_i is the cumulative distribution function specifying the distribution of data $X_{i,1}, \ldots, X_{i,m_i}$ on occasion i. We shall refer to $X_{i,1}, \ldots, X_{i,m_i}$ as the i^{th} subsample. Together, the subsamples form the full sample. We shall denote the size $m_1 + \cdots + m_n$ of the full sample by m. Although it is not reflected in notation, remark that m depends on n, and tends to infinity as n tends to infinity.

In this framework, with a natural order underlying the sequence of data sets, typically by time, we are not interested in all kinds of alternatives to H_0. We rather focus on those alternative explanations that have to do with changes over time, like a shift from one parameter value to another at a certain stage, or some smooth trend change, and so on. In yet other words, the test statistics to be constructed do rely on the original ordering of the n data sets, and are typically not invariant under permutations of these sets.

Our framework, and the methods we develop, aim at being able to monitor quantitative phenomena and their potential changes over time, and should find use in fields like meteorology and climatology [is the temperature increasing?], finance [does the income distribution change in a society?], human socio-behaviour [do people move more than they used to?], and education [are there more lazy students than before?]. In Section 4 we illustrate our methods on data from speedskating championships 1970–2000.

When the cumulative distribution functions F_i belong to some parametric family, then the null hypothesis (1) may be reformulated as

$$H_0 \ : \ \theta_1 = \theta_2 = \ldots = \theta_n, \tag{2}$$

where θ_i is a parameter specifying the distribution of data $X_{i,1}, \ldots, X_{i,m_i}$ on occasion i. In Hjort and Koning (2001) tests of the null hypothesis (2) are investigated for the situation were the θ_i's are finite-dimensional and the m_i's are all equal to 1.

In this paper we take the opposite view, and consider the problem of testing (1) when the cumulative distribution functions F_i are not assumed belong to a certain parametric family. Our aim is to construct so-called monitoring processes, which represent the information contained in the n subsamples with respect to the validity of (1). Graphical displays of monitoring processes should yield useful "diagnostic plots", and functionals of the monitoring process should yield consistent tests of the null hypothesis (1). We present approximations of monitoring processes by means of Gaussian processes. The exponential inequalities describing these approximations are subsequently used in deriving deviation results [that is, a result describing the extreme tail of the distribution of a statistic] for test statistics related to the monitoring processes.

We shall study two different types of monitoring processes. The first type of monitoring process is related to the empirical distribution function, and was proposed in Section 2.6 in Csörgő and Horváth (1997). However, for reasons as given in section 1.2 in Silverman (1986), it may sometimes be more appropriate to respecify (1) as

$$H_0 \ : \ f_1 = f_2 = \ldots = f_n, \tag{3}$$

where f_i is the probability density function corresponding to F_i. In recognition of this fact, we propose a second type of monitoring process, related to the kernel density estimator.

Distribution estimation techniques are of use in an early stage of a statistical analysis as explanatory devices for checking the validity of model assumptions on which later stages of the statistical analysis will be based. In situations where model assumptions incorporate model constancy over time [leading to the use of full sample statistical techniques], violation of the hypothesis (1) may have serious consequences for the statistical analysis as a whole. The methods presented in this paper provide a safeguard against these consequences.

The focus of this paper is on obtaining null hypothesis results. It should be noted that for a full appraisal of the monitoring processes the behaviour of the monitoring process under the alternative hypothesis should also be studied. This will be the subject of a second paper [Hjort and Koning (2002)].

The outline of the paper is as follows. In Section 2 we introduce the monitoring processes, and study their behaviour under the null hypothesis with the aid of exponential inequalities. In Section 3 we use the inequalities of Section 2 to develop deviation results for tests of constancy. Section 4 analyzes speedskating data with the aid of monitoring processes. The proofs gathered in Section 6 draw on the technical results presented in Section 5.

2 Monitoring processes

2.1 Notation and preliminaries

In this section we introduce several monitoring processes, and provide Gaussian approximations under the null hypothesis. In particular, our intention is to show that there exists a non-negative constant κ such that the random variables Δ_n governing these Gaussian approximations belong to a class $\mathcal{C}_0(\kappa)$. This class, which is inspired by the KMT-inequality [cf. Inequality 1 in Section 5], is defined below.

Definition 1 *Let $\mathcal{P}_0$ the class of probability measures corresponding to the null hypothesis (1). A random variable Δ_n is said to belong to the class $\mathcal{C}_0(\kappa)$ [notation: $\Delta_n \in \mathcal{C}_0(\kappa)$] if positive constants c_1–c_4 exist, not depending on n, such that for every $0 < y < c_4 m$*

$$\sup_{P \in \mathcal{P}_0} P\left(|\Delta_n| > (c_1 \log m + y)^\kappa\right) \leq c_2 e^{-c_3 y}.$$

Since $\mathcal{C}_0(\kappa) \subset \mathcal{C}_0(\kappa')$ for $\kappa < \kappa'$, a requirement $\Delta_n \in \mathcal{C}_0(\kappa)$ becomes more stringent as κ decreases: ideally, κ should be as small as possible.

There are a two simple "arithmetic rules" available for the class $\mathcal{C}_0(\kappa)$: if $\Delta_n \in \mathcal{C}_0(\kappa)$ and $\Delta'_n \in \mathcal{C}_0(\kappa')$, then $\Delta_n + \Delta'_n \in \mathcal{C}_0(\kappa \vee \kappa')$ and $\Delta_n \cdot \Delta'_n \in \mathcal{C}_0(\kappa + \kappa')$.

The class $\mathcal{C}_0(\kappa)$ is related to some familiar concepts in probability theory: if $\Delta_n \in \mathcal{C}_0(\kappa)$ for all n, then there exists a positive constant c [for instance, $c = \left(c_1 + 2c_3^{-1}\right)^\kappa$, as readily can be seen by taking y equal to $2c_3^{-1} \log m$], such that

$$\sup_{P \in \mathcal{P}_0} \sum_{n=1}^{\infty} P\left(\Delta_n > c\,(\log m)^\kappa\right) < \infty,$$

and hence the Borel–Cantelli lemma yields that $\Delta_n \leq c\,(\log m)^\kappa$ infinitely often, almost surely, for every $P \in \mathcal{P}_0$; this implies that $\Delta_n/\,(\log m)^\kappa$ remains bounded in probability, uniformly in $P \in \mathcal{P}_0$. Moreover, for any sequence r_n such that $(\log m)^\kappa/r_n$ tends to zero, it follows that

$$\frac{\Delta_n}{r_n} = \frac{\Delta_n}{(\log m)^\kappa} \cdot \frac{(\log m)^\kappa}{r_n} \to 0 \quad \text{almost surely}$$

as $n \to \infty$, for every $P \in \mathcal{P}_0$; this implies $\Delta_n/r_n \to 0$ in probability, uniformly in $P \in \mathcal{P}_0$. In view of the last fact, we may interpret the results in this section as refinements of strong approximations of monitoring processes.

Throughout this paper, the subsample sizes m_i are allowed to be random, and are conveniently represented by the random distribution function

$$\mu_n(t) = m^{-1} \sum_{i=1}^{[nt]} m_i, \quad t \in [0,1].$$

For ease of exposition, within our framework the subsamples are observed at equidistant time instances. However, our results still hold even if these time instances are random, as long as Condition 1 is fulfilled.

Condition 1 *There exist a distribution function μ, a sequence s_n tending to ∞ as $n \to \infty$, and a constant $\alpha \geq 1$ such that*

$$s_n^2 \sup_{t\in[0,1]} |\mu_n(t) - \mu(t)| \in \mathcal{C}_0\Big(\alpha - \tfrac{1}{2}\Big)\,.$$

In industrial statistics, situations where the m_i's are generally larger than 1 are quite usual, as many manufacturing process create several products at the same time ["batch processes"]; the special case where the m_i's are all equal to one is referred as individual observations [cf. Does and Koning (2000)]. Observe that if every m_i is equal to a common value k, then we have $\mu_n(t) = [nt]/n$ and $m = nk$, and hence Condition 1 holds with $s_n = m^{1/2}$ and $\alpha = 1$.

In other circumstances, one may have that the m_i's result from m i.i.d. multinomial experiments. As one may interpret $\mu_n(i/n)$ as the value at point i/n of an empirical distribution function based on m independent observations having support on the interval $[0,1]$, the DKW-Inequality [cf. Inequality 3 in Section 5] yields that Condition 1 holds with $s_n = m^{1/4}$ and $\alpha = 1$.

In what follows we shall often discuss the situation where Condition 1 holds with $(\log m)^\alpha/s_n$ tending to zero as n tends to infinity. Note that this imposes a rather mild lower bound on the rate at which s_n tends to infinity.

2.2 The basic process

The monitoring processes we will consider have in common that they are all related to the basic process $A_n(t,x)$, defined by

$$A_n(t,x) = m^{-1/2} \sum_{i=1}^{[nt]} \sum_{j=1}^{m_i} \left(1_{\{X_{i,j} \le x\}} - F(x) \right) \quad t \in [0,1],\ x \in I\!R, \tag{4}$$

where F denotes the [unknown] common distribution function under the null hypothesis. Later results for monitoring processes will be derived by employing this relation. In this paragraph we present the fundamental result Theorem 1, in which under the hypothesis (1) the basic process $A_n(t,x)$ is approximated by means of a zero mean Gaussian process with covariance function (5). Its proof is deferred to Section 6.

Theorem 1 *If Condition 1 holds, then there exists a sequence of zero mean Gaussian processes $U_n(t,x)$ with covariance function*

$$\mu(t \wedge t') \{F(x \wedge x') - F(x)F(x')\}, \tag{5}$$

such that

$$\left(s_n \vee \frac{m^{1/2}}{\log m} \right) \sup_{t \in [0,1]} \sup_{x \in \boldsymbol{R}} |A_n(t,x) - U_n(t,x)| \in \mathcal{C}_0\left(\left(\frac{\alpha}{2} + \frac{1}{4} \right) \wedge 1 \right). \tag{6}$$

If $(\log m)^\alpha / s_n$ tends to zero as $n \to \infty$, then (6) yields [the random variable Δ_n on the left-hand side of (6) belongs to $\mathcal{C}_0(\alpha)$ since $\alpha \ge 1$ by Condition 1; this implies $\Delta_n / \left(s_n \vee m^{1/2} / \log m \right) \to 0$ almost surely]

$$\sup_{t \in [0,1]} \sup_{x \in \boldsymbol{R}} |A_n(t,x) - U_n(t,x)| \to 0 \quad \text{almost surely;}$$

that is, the Gaussian process $U_n(t,x)$ strongly approximates the basic process $A_n(t,x)$. As the processes $U_n(t,x)$ are identically distributed, this implies that the basic process $A_n(t,x)$ converges in distribution to a Gaussian process with covariance function (5).

2.3 Monitoring cumulative distribution functions

The basic process $A_n(t,x)$ is unfit for use as a monitoring process, as it depends on the unknown cumulative distribution function $F(x)$. In this paragraph we consider monitoring the null hypothesis (1) by means of the process

$$B_n(t,x) = \frac{1}{\sqrt{m}} \sum_{i=1}^{[nt]} m_i \left(\hat{F}_i(x) - \bar{F}_n(x) \right), \quad t \in [0,1],\ x \in I\!R, \tag{7}$$

where

$$\hat{F}_i(x) = \frac{1}{m_i} \sum_{j=1}^{m_i} 1_{\{X_{i,j} \le x\}}$$

is the empirical estimator of $F(x)$ in the i^{th} subsample, and

$$\bar{F}_n(x) = \frac{1}{m} \sum_{i=1}^{n} m_i \hat{F}_i(x) = \frac{1}{m} \sum_{i=1}^{n} \sum_{j=1}^{m_i} 1_{\{X_{i,j} \le x\}}$$

is the empirical estimator of $F(x)$ in the full sample. In Section 2.6 in Csörgő and Horváth (1997) a multivariate version of $B_n(t,x)$ is used to detect change point alternatives.

Lemma 1 *If Condition 1 holds, then under the hypothesis (1) there exists a sequence of zero mean Gaussian processes $V_n(t,x)$ with covariance function*

$$\{\mu(t \wedge t') - \mu(t)\mu(t')\} \{F(x \wedge x') - F(x)F(x')\}, \tag{8}$$

such that

$$\left(s_n \vee \frac{m^{1/2}}{\log m}\right) \sup_{t \in [0,1]} \sup_{x \in \boldsymbol{R}} |B_n(t,x) - V_n(t,x)| \in \mathcal{C}_0(\alpha). \tag{9}$$

If $(\log m)^{\alpha}/s_n$ tends to zero as $n \to \infty$, then Lemma 1 yields that the Gaussian process $V_n(t,x)$ strongly approximates the monitoring process $B_n(t,x)$. As the processes $V_n(t,x)$ are identically distributed, this implies that $B_n(t,x)$ converges in distribution to a zero mean Gaussian process with covariance function (8) [see also Theorem 2.6.1 in Csörgő and Horváth (1997), p. 153].

We have that $V_n(t,x)$ is equal in distribution to $\Gamma(\mu(t), F(x))$, where $\Gamma(w,u)$ is a zero mean Gaussian process with covariance function $\{w \wedge w' - ww'\}\{u \wedge u' - uu'\}$. In literature, the Gaussian process $\Gamma(w,u)$ is called the Wiener pillow [Piterbarg (1996), p. 137; inspired by the fact that $\Gamma(w,u) = 0$ almost surely for all (w,u) on the boundary of the unit square], the completely tucked Brownian sheet [van der Vaart and Wellner (1996), p. 368] or the tied-down Kiefer process [Csörgő and Horváth (1997), p. 320]. We shall refer to Γ as the Brownian pillow. One may view the Brownian pillow as a two-parameter generalization of the Brownian bridge.

Weighing provides a convenient way of strengthening properties of the monitoring process. Lemma 2 describes the behaviour of the weighted monitoring process

$$C_n(t,x) = L(t)B_n(t,x) - \int_0^t B_n(v,x)\, dL(v), \quad t \in [0,1],\ x \in I\!R. \tag{10}$$

Condition 2 *There exist a finite constant $c_5 > 0$ such that $L(t)$ is bounded by c_5, and has variation bounded by c_5.*

Lemma 2 *Let $L(t)$ satisfy Condition 2, and let $V_n(t,x)$ be the zero mean Gaussian process approximating $B_n(t,x)$ in Lemma 1. There exists a sequence of zero mean Gaussian processes*

$$W_n(t,x) = L(t)V_n(t,x) - \int_0^t V_n(v,x)\,dL(v), \quad t \in [0,1],\ x \in I\!R,$$

with covariance function

$$\left\{\int_0^{t\wedge t'} L(v)^2 d\mu(v) - \int_0^t L(v)d\mu(v)\int_0^{t'} L(v)d\mu(v)\right\}\{F(x\wedge x') - F(x)F(x')\}, \tag{11}$$

such that

$$\left(s_n \vee \frac{m^{1/2}}{\log m}\right) \sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}} |C_n(t,x) - W_n(t,x)| \in \mathcal{C}_0(\alpha). \tag{12}$$

If $(\log m)^\alpha / s_n$ tends to zero as $n \to \infty$, then Lemma 2 yields that the Gaussian process $W_n(t,x)$ strongly approximates the monitoring process $C_n(t,x)$. As the processes $W_n(t,x)$ are identically distributed, this implies that $C_n(t,x)$ converges in distribution to a zero mean Gaussian process with covariance function (11).

2.4 Monitoring probability density functions

The process $B_n(t,x)$ usually provides a satisfactory way of monitoring the null hypothesis (1). However, for reasons as given in Section 1.2 in Silverman (1986), a probability density function may often describe the distribution of a random variable more appropriately than a cumulative distribution function. In this paragraph we consider monitoring the null hypothesis (3) by means of the process

$$B_{n,h}(t,x) = h^{1/2} m^{-1/2} \sum_{i=1}^{[nt]} m_i \left(\hat{f}_i(x) - \bar{f}_{n,h}(x)\right), \quad t \in [0,1],\ x \in I\!R,$$

where

$$\hat{f}_i(x) = \frac{1}{m_i h}\sum_{j=1}^{m_i} K\left(\frac{X_{i,j}-x}{h}\right)$$

is the kernel density estimator in subsample i, and

$$\bar{f}_{n,h}(x) = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K\left(\frac{X_{i,j}-x}{h}\right) = \frac{1}{m}\sum_{i=1}^{n} m_i \hat{f}_i(x)$$

is the full sample kernel density estimator under the null hypothesis (3). Here, $K(x)$ is a symmetric density, and h a smoothing parameter. Observe that we use the same smoothing parameter h for each density estimator $\hat{f}_i$.

Let f denote the common probability density function under the null hypothesis (3); that is, the derivative of $F(x)$. Introduce

$$\sigma_h(x,x') = h^{-1}\int K\left(\frac{v-x}{h}\right)K\left(\frac{v-x'}{h}\right)f(v)dv - h\mathcal{E}\bar{f}_{n,h}(x)\mathcal{E}\bar{f}_{n,h}(x'),$$

where

$$\mathcal{E}\bar{f}_{n,h}(x) = h^{-1}\int K\left(\frac{v-x}{h}\right)f(v)dv = \int K(u)f(x+hu)du.$$

One may interpret $\sigma_h(x,x')/h$ as the covariance function of the full sample estimator $\bar{f}_{n,h}(x)$. Observe that in general $\mathcal{E}\bar{f}_{n,h}(x)$ does not coincide with $f(x)$; hence, kernel density estimators may be biased.

Condition 3 *The kernel function $K(x)$ is a symmetric probability density function satisfying*

$$\int |K'(x)|\,dx < c_6,$$

where $K'(x)$ denotes the derivative of $K(x)$, and c_6 is a finite constant.

Lemma 3 *If Conditions 1 and 3 hold, then under the hypothesis (3) there exists a sequence of zero mean Gaussian processes $V_{n,h}\,(t,x)$ with covariance function*

$$\{\mu\,(t\wedge t') - \mu(t)\mu(t')\}\,\sigma_h(x,x'), \tag{13}$$

such that

$$\left(s_n \vee \frac{m^{1/2}}{\log m}\right) h^{1/2} \sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}} |B_{n,h}\,(t,x) - V_{n,h}\,(t,x)| \in \mathcal{C}_0(\alpha)\,. \tag{14}$$

The proof of Lemma 3 exploits the relation between the the density estimator and the empirical process. This relation was noticed already in Bickel and Rosenblatt (1973), a seminal paper in density estimation. However, the powerful machinery of Komlós, Major and Tusnády (1975) became available later, and was used in the context in density estimation in Csörgő and Révész (1981) [Theorem 6.1.1, p. 223], and in Konakov and Piterbarg (1983). The idea of using strong approximation in the context of density estimation traces back to Rosenblatt (1971).

If $(\log m)^\alpha/s_n$ tends to zero as $n\to\infty$, then Lemma 3 yields for fixed and positive h that the Gaussian process $V_{n,h}\,(t,x)$ strongly approximates the monitoring process $B_{n,h}\,(t,x)$. As the processes $V_{n,h}\,(t,x)$ are identically distributed, this implies that $B_{n,h}\,(t,x)$ converges in distribution to a zero mean Gaussian process with covariance function (13).

Lemma 3 continues to hold if h is replaced by h_n which tends to zero as n tends to infinity. In this case, Lemma 3 yields that the Gaussian process $V_{n,h_n}\,(t,x)$ strongly approximates the monitoring process $B_{n,h_n}(t,x)$ if $(h_n)^{-1/2}(\log m)^\alpha/s_n$ tends to zero as $n \to \infty$.

However, for h_n tending to zero, the process $V_{n,h_n}(t,x)$ [and hence $B_{n,h_n}(t,x)$] does not have a limit in distribution. To clarify this, introduce

$$J(c) = \int K\left(u + \frac{c}{2}\right) K\left(u - \frac{c}{2}\right) du,$$

and write

$$\mathcal{E}\bar{f}_{n,h}(x) = \int K(u) f(hu + x)\, du \approx f(x) + \tfrac{1}{2}h^2 \int u^2 K(u) du f''(x),$$

$$\begin{aligned}
&\sigma_h(x,x') + h\mathcal{E}\bar{f}_{n,h}(x)\mathcal{E}\bar{f}_{n,h}(x') \\
&\quad = \int K\left(u + \frac{x'-x}{2h}\right) K\left(u - \frac{x'-x}{2h}\right) f\left(hu + \frac{x'+x}{2}\right) du \\
&\quad \approx J\left(\frac{x-x'}{h}\right) f\left(\frac{x'+x}{2}\right)
\end{aligned}$$

for h close to zero. It follows that $\sigma_h(x,x')$ tends to $\sigma_0(x,x') = J(0)f(x)1_{\{x=x'\}}$ for h tending to zero. The structure of $\sigma_0(x,x')$ implies that a Gaussian process with covariance function $\mu(t \wedge t')\,\sigma_0(x,x')$ cannot have continuous sample paths. Continuity of sample paths is a key condition in the study of Gaussian processes [cf. Ledoux and Talagrand (1991), Chapter 12].

Although the process $B_{n,h_n}(t,x)$ does not have a limit in distribution, Lemma 3 nevertheless yields that for every n there exists a Gaussian process which nearly has the same distribution as $B_{n,h_n}(t,x)$. This underlines the usefulness of strong approximation methods in density estimation.

Lemma 4 describes the behaviour of the weighted monitoring process

$$C_{n,h}(t,x) = L(t)B_{n,h}(t,x) - \int_0^t B_{n,h}(v,x)\, dL(v), \quad t \in [0,1],\ x \in I\!R. \tag{15}$$

Lemma 4 *Let $L(t)$ satisfy Condition 2, and let $V_{n,h}(t,x)$ be the zero mean Gaussian process approximating $B_{n,h}(t,x)$ in Lemma 3. There exists a sequence of zero mean Gaussian processes*

$$W_{n,h}(t,x) = L(t)V_{n,h}(t,x) - \int_0^t V_{n,h}(v,x)\, dL(v), \quad t \in [0,1],\ x \in I\!R,$$

with covariance function

$$\left\{\int_0^{t\wedge t'} L(v)^2 d\mu(v) - \int_0^t L(v) d\mu(v) \int_0^{t'} L(v) d\mu(v)\right\} \sigma_h(x,x'), \tag{16}$$

such that

$$\left(s_n \vee \frac{m^{1/2}}{\log m}\right) h^{1/2} \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} |C_{n,h}(t,x) - W_{n,h}(t,x)| \in \mathcal{C}_0(\alpha)\,. \tag{17}$$

If $(\log m)^{\alpha}/s_n$ tends to zero as $n \to \infty$, then Lemma 4 yields for fixed and positive h that the Gaussian process $W_{n,h}(t,x)$ strongly approximates the monitoring process $C_{n,h}(t,x)$. As the processes $W_{n,h}(t,x)$ are identically distributed, this implies that $C_{n,h}(t,x)$ converges in distribution to a zero mean Gaussian process with covariance function (16).

For h_n tending to zero as n tends to infinity, Lemma 3 yields that the Gaussian process $W_{n,h_n}(t,x)$ strongly approximates the monitoring process $C_{n,h_n}(t,x)$ if $(h_n)^{-1/2}(\log m)^{\alpha}/s_n$ tends to zero as $n \to \infty$. However, the process $W_{n,h_n}(t,x)$ [and hence $C_{n,h_n}(t,x)$] does not have a limit in distribution.

2.5 Bootstrapped versions of monitoring processes

The bootstrapped version of a monitoring process is obtained by replacing each of the original random variables $X_{i,j}$ by a random variable X^*_{ij}, where the X^*_{ij}'s together form a random sample of length m drawn from the cumulative distribution function $\bar{F}_n$.

Lemma 5 *Let $C^*_n(t,x)$ and $C^*_{n,h}(t,x)$ be the bootstrapped versions of $C_n(t,x)$ and $C_{n,h}(t,x)$. If Conditions 1–3 hold, then there exist a sequence of zero mean Gaussian processes $W'_n(t,x)$ with covariance function (11) and a sequence of zero mean Gaussian processes $W'_{n,h}(t,x)$ with covariance function (16) such that*

$$\left(s_n \vee m^{1/4}\right) \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left|C^*_n(t,x) - W'_n(t,x)\right| \in \mathcal{C}_0(\alpha)\,, \tag{18}$$

$$\left(s_n \vee m^{1/4}\right) h^{1/2} \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left|C^*_{n,h}(t,x) - W'_{n,h}(t,x)\right| \in \mathcal{C}_0(\alpha)\,. \tag{19}$$

Although they have a common distribution, the processes $W_n(t,x)$ [cf. Lemma 2] and $W'_n(t,x)$ typically do not coincide. As similar remark holds for the processes $W_{n,h}(t,x)$ [cf. Lemma 4] and $W'_{n,h}(t,x)$.

If $(\log m)^{\alpha}/s_n$ tends to zero as $n \to \infty$, then Lemma 5 yields that the Gaussian process $W'_n(t,x)$ strongly approximates the bootstrapped monitoring process $C^*_n(t,x)$. As $W'_n(t,x)$ is equal in distribution to $W_n(t,x)$, it follows that the bootstrap "works" in the sense that $C^*_n(t,x)$ and $C_n(t,x)$ share the same limiting distribution. For fixed and positive h, a similar argument shows that $C^*_{n,h}(t,x)$ and $C_{n,h}(t,x)$ share the same limiting distribution.

3 Tests of constancy

3.1 Notation and preliminaries

The objective in this section is to establish deviation results for tests of constancy which are derived from monitoring processes. In this paragraph we describe a general framework for obtaining deviation results.

Consider a statistical test which rejects the null hypothesis for large values of a test statistic T_n. We shall say that the test statistic T_n obeys a deviation result if

$$\lim_{n\to\infty}(y_n)^{-2}\log\sup_{P\in\mathcal{P}_0} P\left(T_n > y_n\right) = -a/2 \tag{20}$$

holds for certain sequences y_n such that $y_n \to \infty$. The strength of the deviation result is determined by the class of sequences which are allowed. Chernoff type deviation results allow sequences $y_n = o\left(m^{1/2}\right)$, and are relevant for the computation of exact Bahadur efficiency [cf. Bahadur (1960)]; Cramér type deviation results allow sequences $y_n = o\left(m^{1/6}\right)$, and are relevant for the computation of intermediate efficiency [cf. Kallenberg (1983)]; moderate deviation results allow sequences $y_n = O\left((\log m)^{1/2}\right)$, and are relevant for the computation of weak intermediate efficiency [cf. Kallenberg (1983)] and Bayes risk efficiency [cf. Rubin and Sethuraman (1965)].

Lemma 6 *Suppose the test statistic T_n' satisfies (i)–(iii) below.*

(i) There exists a random variable $\tilde{T}_n'$ with distribution not depending on n, and for every $P \in \mathcal{P}_0$ there exists $\nu_0 = \nu_0(P)$ such that

$$\lim_{y\to\infty} y^{-2}\log\sup_{P\in\mathcal{P}_0} P\left(\nu_0^{-1}\tilde{T}_n' \geq y\right) = -a/2.$$

(ii) There exists a sequence of positive constants r_n such that $r_n/(\log m) \to \infty$ and $r_n = O\left(m^{1/2}\right)$ as $n \to \infty$, and a positive constant $\kappa \geq \frac{1}{2}$ such that

$$r_n\nu_0^{-1}\left|T_n' - \tilde{T}_n'\right| \in \mathcal{C}_0(\kappa)\,.$$

(iii) There exist a statistic $\hat{\nu}$ and a non-negative constant $\kappa' \geq 1-\kappa$ such that

$$r_n\left|\nu_0/\hat{\nu} - 1\right| \in \mathcal{C}_0(\kappa')\,.$$

Then there exists a positive constant a such that the test statistic $T_n = \hat{\nu}^{-1}T_n'$ satisfies the deviation result (20) for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left((r_n)^{1/(2\kappa+2\kappa'-1)}\right)$ as $n \to \infty$.

One of the features of Lemma 6 is the use of exponential inequalities to derive deviation results. Examples of deviation results obtained via exponential inequalities [but in a simple null hypothesis setting] may be found in Inglot and Ledwina (1990, 1993) and in Koning (1992, 1994).

Inspection of the proof of Lemma 6 reveals that κ' may be taken equal to 0 if $\hat{\nu}$ coincides with ν_0.

3.2 A general approach for sublinear tests

In this paragraph we briefly outline the verification of the conditions of Lemma 6 for test statistics based on the monitoring processes $C_n(t,x)$ and $C_{n,h}(t,x)$.

Let $D([0,1]\times I\!R)$ denote the space of real-valued functions defined on $[0,1]\times I\!R$ which are cadlag in both components, and let $T : D([0,1]\times I\!R) \to I\!R^+$ be a functional which is positive-homogeneous [that is, $T(c\xi) = cT(\xi)$ for every constant $c > 0$ and every $\xi \in D([0,1]\times I\!R)$] and Lipschitz [that is, there exists a constant $c_7 > 0$ such that $|T(\xi) - T(\xi')| \le c_7 \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} |\xi(t,x) - \xi'(t,x)|$ for every $\xi, \xi' \in D([0,1]\times I\!R)$].

If we set

$$T_n' = T(C_n) \quad \tilde{T}_n' = T(W_n),$$

then we have by Lemma 2 that (12) holds under Condition 1. Since

$$\left|T_n' - \tilde{T}_n'\right| \le c_7 \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} |C_n(t,x) - W_n(t,x)|,$$

this yields

$$\left(s_n \vee \frac{m^{1/2}}{\log m}\right)\left|T_n' - \tilde{T}_n'\right| \in \mathcal{C}_0(\alpha),$$

and hence condition (ii) of Lemma 6 is satisfied with $r_n = s_n \vee m^{1/2}/\log m$ and $\kappa = \alpha$. Thus, it only remains to show that conditions (i) and (iii) of Lemma 6 are satisfied. In this respect, we note that if the functional T is not only positive-homogeneous and Lipschitz, but sublinear as well [that is, $T(\xi+\xi') \le T(\xi) + T(\xi')$ for every $\xi, \xi' \in D([0,1]\times I\!R)$], then condition (i) of Lemma 6 may be verified along the lines of the proof of Theorem 5.2 in Borell (1975) [see also Inequality 1 in Koning and Protassov (2001)]; in this case we should set ν_0 proportional to $\|T\|_{\mathcal{H}}$, and a equal to $\left(\nu_0^{-1}\|T\|_{\mathcal{H}}\right)^{-2}$, where $\|T\|_{\mathcal{H}} = \sup_{\xi\in\mathcal{O}_{\mathcal{H}}} T(\xi)$, and $\mathcal{O}_{\mathcal{H}}$ is the unit ball in the reproducing kernel Hilbert space $\mathcal{H}$ belonging to the Brownian pillow W_1 [cf. Section III.2 in Adler (1990)].

Similarly, if we set

$$T_n' = T(C_{n,h}) \quad \tilde{T}_n' = T(W_{n,h}),$$

where T is positive-homogeneous and Lipschitz, then under Condition 1 it follows by Lemma 4 that (ii) of Lemma 6 is satisfied with $r_n = s_n \vee m^{1/2}/\log m$ and $\kappa = \alpha$. Again, it only remains to show that conditions (i) and (iii) of Lemma 6 are satisfied. Also, if the functional T is sublinear as well, then condition (i) of Lemma 6 may be verified along the lines of the proof of Theorem 5.2 in Borell (1975). In this case we should set ν_0 proportional to $\|T\|_{\mathcal{H}_h}$, and a equal to $\left(\nu_0^{-1}\|T\|_{\mathcal{H}_h}\right)^{-2}$, where $\mathcal{H}_h$ is the reproducing kernel Hilbert space belonging to $W_{1,h}$.

3.3 Supremum type tests

To illustrate the general approach described in the previous paragraph, we now verify conditions (i) and (iii) of Lemma 6 for the special case where T takes the form

$$T(\xi) = \sup_{x \in \boldsymbol{R}} S\left(\xi\left(\cdot, x\right)\right),$$

where

$$S\left(\xi_2\right) = \sup_{v \in V} \sqrt{Q_v(\xi_2, \xi_2)} \tag{21}$$

for every $\xi_2 \in D\left(I\!\!R\right)$; here V is some index set, and Q_v is a symmetric bounded bilinear form on $D\left(I\!\!R\right)$ for every $v \in V$ [see also Koning and Protasov (2001)]. Without loss of generality, we shall confine ourselves to test statistics of the form $\sup_{x \in \boldsymbol{R}} S\left(B_n\left(\cdot, x\right)\right)$ or $\sup_{x \in \boldsymbol{R}} S\left(B_{n,h}\left(\cdot, x\right)\right)$ [observe that $\sup_{x \in \boldsymbol{R}} S\left(C_n\left(\cdot, x\right)\right)$ may be expressed as $\sup_{x \in \boldsymbol{R}} S_L\left(B_n\left(\cdot, x\right)\right)$ for a convenient choice of S_L].

Typical examples of S are the Kolmogorov functional S_{Kol}, the Cramér-von Mises functional S_{CvM} and the Andersen-Darling functional S_{AD}, respectively defined by

$$S_{\text{Kol}}(\xi_2) = \sup_{x \in \boldsymbol{R}} \left|\xi_2(x)\right|,$$

$$S_{\text{CvM}}(\xi_2) = \left\{\int_0^1 \left(\xi_2(v)\right)^2 d\mu(v)\right\}^{1/2},$$

$$S_{\text{AD}}(\xi_2) = \left\{\int_0^1 \frac{\left(\xi_2(v)\right)^2}{\mu(v)(1-\mu(v))} d\mu(v)\right\}^{1/2}.$$

For every S of the form (21) there is an associated positive constant $a_S = \sup_{\xi \in \mathcal{O}_{\mathcal{H}}} S\left(\xi\right)$, where $\mathcal{O}_{\mathcal{H}}$ is the unit ball in the reproducing kernel Hilbert space $\mathcal{H}$ belonging to the covariance function $\mu\left(t \wedge t'\right) - \mu(t)\mu(t')$. In particular, we have $a_{S_{\text{Kol}}} = 4$, $a_{S_{\text{CvM}}} = \pi^2$ and $a_{S_{\text{AD}}} = 2$ [cf. Koning and Protasov (2001)].

Lemma 7 provides the necessary additional results for $\sup_{x \in \boldsymbol{R}} S\left(B_n\left(\cdot, x\right)\right)$.

Lemma 7 *The random variable $\tilde{T}_n' = \sup_{x \in \boldsymbol{R}} S\left(V_n\left(\cdot, x\right)\right)$ satisfies condition (i) of Lemma 6 with $\nu_0 = 1$ and $a = 4a_S$.*

As an example, suppose that Condition 1 holds with $\alpha = 1$ and $s_n = m^{1/2}/\log m$. Applying Lemma 6 with $\hat{\nu} = \nu_0 = 1$ yields that $\sup_{x \in \boldsymbol{R}} S\left(C_n\left(\cdot, x\right)\right)$ satisfies the deviation result (20) for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left(m^{1/2}/\log m\right)$ as $n \to \infty$; the constant a is given by Lemma 7. This deviation result comprises a Cramér type deviation result, and only just falls short of being a Chernoff type result.

For the sake of completeness, Table 1 lists the approximate upper percentage points of $\sup_{x \in \boldsymbol{R}} S\left(V_n\left(\cdot, x\right)\right)$ reported in Koning and Protasov (2001) .

Lemma 8 provides the necessary additional results for $\sup_{x \in \boldsymbol{R}} S\left(B_{n,h}\left(\cdot, x\right)\right)$.

Lemma 8 *For $P \in \mathcal{P}_0$, define $\nu_0 = \nu_0(P)$ by $\nu_0^2 = \sup_{x \in \boldsymbol{R}} \sigma_h(x,x)$. Define the estimator $\hat{\nu}$ by*

$$\hat{\nu}^2 = \sup_{x \in \boldsymbol{R}} \frac{1}{mh} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ K\left(\frac{X_{i,j} - x}{h}\right) - \bar{f}_{n,h}(x) \right\}^2 .$$

Assume Condition 3 holds, and assume that there exists a positive constant c_8 such that $c_8 \nu_0 \geq 1$ for every $P \in \mathcal{P}_0$.

(i) The random variable $\tilde{T}_n' = \sup_{x \in \boldsymbol{R}} S\left(V_{n,h}\left(\cdot, x\right)\right)$ satisfies condition (i) of Lemma 6 with $\nu_0^{-1} = \sup_{x \in \boldsymbol{R}} \sigma_h(x,x)$ and $a = a_S$.

(ii) There exist positive constants c_{17}–c_{19} such that

$$\sup_{P \in \mathcal{P}_0} P\left(h\left|\nu_0/\hat{\nu} - 1\right| > m^{-1/2} y\right) \leq c_{17} \exp\left\{-c_{18} y\right\}$$

for $0 < y < \left(\min(1, c_{19} h)\right)^2 m$.

Observe that Lemma 8 (ii) implies that the statistic $\hat{\nu}$ satisfies condition (iii) of Lemma 6 for $\kappa' = \frac{1}{2}$ and $r_n = m^{1/2}$.

As an example, suppose that Condition 1 holds with $\alpha = 1$ and $s_n = m^{1/2}/\log m$. Let $\hat{\nu}$ be as defined in Lemma 8. Applying Lemma 6 yields that $\sup_{x \in \boldsymbol{R}} S\left(B_{n,h}\left(\cdot, x\right)\right)$ satisfies the deviation result (20) for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left(m^{1/4}/(\log m)^{1/2}\right)$ as $n \to \infty$. The constant a is given by Lemma 8. Observe that this deviation result comprises a Cramér type deviation result, but stretches less far than the corresponding deviation result for $\sup_{x \in \boldsymbol{R}} S\left(C_n\left(\cdot, x\right)\right)$.

3.4 Bootstrap tests

When using test statistics of the type $T\left(C_n\right)$ or $T\left(C_{n,h}\right)$, one is hampered by the limited knowledge available in present literature about the distribution of a functional of a general zero mean Gaussian process; for instance, there is even no known formula for the distribution function of the supremum of a Kiefer process [cf. Csörgő and Horváth (1997), p. 101].

In this paragraph we resort to the bootstrapped versions of these tests as “ the use of the bootstrap either relieves the analyst from having to do complex mathematical derivations, or in some instances provides an answer where no analytical answer can be obtained” [Efron and Tibshirani (1993), p. 394]. The bootstrapped version of the test based on $T\left(C_n\right)$ is obtained by using the conditional distribution of $T\left(C_n^*\right)$ given $\bar{F}_n$ [instead of the unconditional distribution of $T\left(C_n\right)$] to determine the achieved significance level. The bootstrapped version of the test based on $T\left(C_{n,h}\right)$ is obtained in a similar manner.

If T is a sublinear Lipschitz functional, then Lemma 5 yields that the bootstrap "works" in the sense that $T\left(C_n^*\right)$ and $T\left(C_{n,h}^*\right)$ have the same limiting distribution as $T\left(C_n\right)$ and $T\left(C_{n,h}\right)$, respectively.

More importantly, the bootstrap relieves us from the task of verifying condition (iii) of Lemma 6 [the proof of Lemma 8 (iii) shows that this task may well be formidable], as the bootstrap procedure implicitly estimates ν_0: the bootstrapped version of the test based on $\nu_0^{-1} T\left(C_n\right)$ is equivalent to the bootstrapped version of the test based on $T\left(C_n\right)$, and the bootstrapped version of the test based on $\nu_0^{-1} T\left(C_{n,h}\right)$ is equivalent to the bootstrapped version of the test based on $T\left(C_{n,h}\right)$, This means that we may apply Lemma 6 with $\hat{\nu} = \nu_0$; observe that condition (iii) of Lemma 6 now holds trivially.

Conditions (i) and (ii) of Lemma 6 may be verified in the same manner as before, with Lemma 5 taking over the role of Lemma 2 and Lemma 4. This only has a minor effect on our results [in fact, we now should take r_n equal to $s_n \vee m^{1/4}$ instead of $s_n \vee m^{1/2} / \log m$].

As an example, suppose that Condition 1 holds with $\alpha = 1$ and $s_n = m^{1/2} / \log m$. Applying Lemma 6 yields that the bootstrapped version of $\sup_{x \in \boldsymbol{R}} S\left(C_n(\cdot, x)\right)$ satisfies the deviation result (20) for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left(m^{1/4}\right)$ as $n \to \infty$; the constant a is given by Lemma 8. Although this deviation result comprises a Cramér type deviation result, it stretches less far than the corresponding deviation result for $\sup_{x \in \boldsymbol{R}} S\left(C_n(\cdot, x)\right)$.

Lemma 6 also yields that the bootstrapped version of $\nu_0^{-1} \sup_{x \in \boldsymbol{R}} S\left(B_{n,h}(\cdot, x)\right)$ satisfies the deviation result (20) for all sequences y_n such that $y_n \to \infty$ and $y_n = o\left(m^{1/4}\right)$ as $n \to \infty$; the constant a is given by Lemma 8. This deviation result comprises a Cramér type deviation result, and is slightly better than the corresponding deviation result for $\sup_{x \in \boldsymbol{R}} S\left(B_{n,h}(\cdot, x)\right)$.

4 An application to speedskating data

Speedskating world allround championships are annual events consisting of four distances 500m, 5000m, 1500m and 10000m [in that order]. There are limitations on the number of participants on the 10k distance. In the years 1970–1992 a maximum of 16 participants were allowed. Due to the 10k selection rules, some of these 16 participants may have some distance results missing. For instance, in 1992 no 5k results were recorded for Johansen and Søndrål, and no 10k results were recorded for Bos and Tröger. In 1993 the 10k selection rules were altered, lowering the number of 10k participants to 12. As participants with missing distance results were excluded from the data, in total $m = 441$ observations were recorded during the period 1970–2000. Each of these obervations consisted of a 0.5k, a 5k, a 1.5k and a 10k result.

Over the years the results on those four distances have improved considerably, due to changes in professionalism, training methods, environment [indoor skating rinks], material ["klapschaats"]. Amazingly, the 10k times are now about two full minutes faster than in 1970, and the 5k times similarly a minute faster.

It is invariably interesting, and sometimes fruitful, to predict tomorrow's performances based on today's achievements; such exercises can sometimes determine race strategies. The Dutch coach Ab Krook used the simple prediction rule that the 10k time would be close to twice the 5k time plus 20 seconds. Krook's rule may be viewed as a statement with respect to the expectation of the "endurance variable"

$$y = [\text{10k result}] - 2\,[\text{5k result}]\,.$$

In this example we investigate whether the "endurance distribution", the distribution of the endurance variable y, has remained constant throughout the period 1970–2000. In Figure 1 we have plotted Krook's variable versus the year in which the speed skating world championship event took place.

In Figure 2 the monitoring process $B_n(t,x)$ for cumulative distribution functions is displayed. The statistic $\sup_{x\in\boldsymbol{R}} S_{\mathrm{Kol}}(B_n(\cdot,x))$ takes the value 0.795. According to Table 1 the null hypothesis should not be rejected at the 0.05 level. However, 100.000 bootstrap replications yield an attained significance level of 0.0181, indicating that the null hypothesis should be rejected at the 0.05 level. As both Table 1 and the bootstrap are fundamented on asymptotic methods, we should doubt whether asymptotic methods indeed apply here: although $m = 441$, we only have $n = 31$.

The statistics $\sup_{x\in\boldsymbol{R}} S_{\mathrm{CvM}}(B_n(\cdot,x))$ and $\sup_{x\in\boldsymbol{R}} S_{\mathrm{AD}}(B_n(\cdot,x))$ take the value 0.313 and 0.735, respectively. According to Table 1 the null hypothesis should not be rejected at the 0.10 level, in agreement with the attained significance level of 0.2174 and 0.2028 [100.000 bootstrap replications], respectively.

In Hjort and Koning (2002) the use of bandwidth $h = .75s$ for monitoring probability density functions is advocated, where

$$s^2 = \frac{1}{m-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(X_{i,j} - \bar{X}_n\right)^2, \quad \bar{X}_n = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i} X_{i,j}.$$

For the speed skating data s takes the value 15.485.

In Figure 3 the monitoring process $B_{n,.75s}(t,x)$ for probability density functions is displayed. The statistics $\sup_{x\in\boldsymbol{R}} S_{\mathrm{Kol}}(B_{n,.75s}(\cdot,x))$, $\sup_{x\in\boldsymbol{R}} S_{\mathrm{CvM}}(B_{n,.75s}(\cdot,x))$ and $\sup_{x\in\boldsymbol{R}} S_{\mathrm{AD}}(B_{n,.75s}(\cdot,x))$ take the values 0.040, 0.018 and 0.041, respectively. The attained significance levels are 0.1491, 0.4147 and 0.4143 [100.000 bootstrap replications], respectively, indicating that the null hypothesis should not be rejected at the 0.05 level.

It is interesting and perhaps surprising that the endurance variable y appears to not have changed its distribution over the past 30 years, in spite of the drastic changes the 5k and 10k times have experienced over this time interval.

5 Some technical results

In this section some technical results with respect to the sequential uniform empirical process and Gaussian processes are collected. These results will be used in Section 6.

Let $Y_1, Y_2, \ldots$ be independent standard uniform random variables. The sequential uniform empirical process is defined at stage m as

$$m^{-1/2} \sum_{\ell=1}^{[mw]} \left(1_{\{Y_\ell \le u\}} - u\right) \quad \text{for } w \in [0,1],\ u \in [0,1],$$

where $Y_1, Y_2, \ldots$ are independent standard uniform random variables. As m tends to infinity, the sequential uniform empirical process converges weakly to a Kiefer process; that is, a zero mean Gaussian process $\Psi(w,u)$ with covariance function $w \wedge w' \{u \wedge u' - uu'\}$ [Müller (1970)]. In Kiefer (1972) the first strong approximation for the sequential uniform empirical process was given, which was subsequently refined in Csörgő and Révész (1975), and given its final form in Komlós, Major and Tusnády (1975). It is known that Inequality 1 holds with $c_9 = 76$, $c_{10} = 2.028$ and $c_{11} = 1/4$ [cf. Csörgő and Horváth (1993), p. 150].

Inequality 1 (KMT-inequality) *Given independent standard uniform random variables $Y_1, Y_2, \ldots$, there exists a sequence of Kiefer processes $\Psi_n(w,u)$ such that*

$$P\left(\max_{k=1,2,\ldots,m} \sup_{u\in[0,1]} \left|\sum_{\ell=1}^{k} \left(1_{\{Y_\ell \le u\}} - u\right) - \Psi_n\left(\frac{k}{m}, u\right)\right| > (c_9 \log m + y)\log m\right)$$
$$< c_{10} \exp\{-c_{11} y\}$$

for all $y > 0$ and n, where c_9–c_{11} are positive absolute constants.

Inequality 1 is in fact a slightly weakened version of the original KMT-inequality, which in addition states that there exists an “underlying” Kiefer process $\Psi(w,u)$ such that $\Psi_n(w,u) = m^{-1/2}\Psi(mw,u)$ for every n.

Lemma 9 *There exist positive constants c_{12},c_{13} such that*

$$P\left(\sup_{w \le r} \sup_{u\in[0,1]} |\Psi_n(w,u)| > y\sqrt{r}\right) \le c_{12} \exp\{-c_{13} y^2\} \tag{22}$$

for every $y > 0$ and $r > 0$.

Proof of Lemma 9 For each t, we may view $\Psi_n(w,\cdot)$ as a random element of the separable Banach space $C([0,1])$. Moreover, $r^{-1/2}\Psi_n(r,\cdot)$ is a Brownian bridge. Combining Proposition 2.6.1 in de la Peña and Giné (1999) and (2.2.11) in Shorack and Wellner (1986), p. 34, yields

$$P\left(\sup_{w\in[0,r]} \sup_{u\in[0,1]} |\Psi_n(w,u)| > y\sqrt{r}\right) \le 9P\left(\sup_{u\in[0,1]} |\Psi_n(r,u)| > y\sqrt{r}/30\right)$$
$$\le 9P\left(\sup_{u\in[0,1]} \left|r^{-1/2}\Psi_n(r,u)\right| > y/30\right) \le 18 \exp\left\{-y^2/450\right\}.$$

This completes the proof of Lemma 9. □

Lemma 10 *Let $\omega(\epsilon)$ be the modulus of continuity defined by*

$$\omega(\epsilon) = \sup_{u\in[0,1]} \sup_{0\le w-w'\le\epsilon} |\Psi_n(w,u) - \Psi_n(w',u)| \tag{23}$$

for $\epsilon > 0$, and let r_n be a sequence tending to infinity as $n \to \infty$, satisfying $\log r_n \le c\log m$ *for some constant $c \ge 2$. Then*

$$r_n\omega\left(\frac{(c\log m + y)^\kappa}{r_n^2}\right) \in \mathcal{C}_0\left(\frac{\kappa+1}{2}\right)$$

for every $\kappa \ge 0$.

Proof of Lemma 10 Let c_{12} and c_{13} be as in Lemma 9, and define $c_{14} = 64c_{12}$, $c_{15} = 4c_{13}/9$. Let $\epsilon > 0$, and let N denote the smallest integer satisfying $N \ge 16/\epsilon$. By a similar construction as employed in the proof of Inequality 14.1.1 in Shorack and Wellner (1986), p.536, it follows by Lemma 9 that

$$\begin{aligned} P\left(\omega(r) \ge z\sqrt{\epsilon}\right) &\le \sum_{j=1}^{N}\left\{P\left(\sup_{0\le w\le\epsilon}\sup_{u\in[0,1]} |\Psi_n(w,u)| \ge \frac{2}{3}z\sqrt{\epsilon}\right)\right. \\ &\left. +P\left(\sup_{0\le w\le 1/N}\sup_{u\in[0,1]} |\Psi_n(w,u)| \ge \frac{2}{3}z\frac{\sqrt{rN}}{4}\frac{1}{\sqrt{N}}\right)\right\} \\ &\le \sum_{j=1}^{N}\left\{c_{12}\exp\left\{-\frac{4c_{13}}{9}z^2\right\} + c_{12}\exp\left\{-\frac{4c_{13}}{9}z^2\frac{\epsilon N}{16}\right\}\right\} \\ &\le 2Nc_{12}\exp\left\{-c_{15}z^2\right\} \le \frac{c_{14}}{\epsilon}\exp\left\{-c_{15}z^2\right\}, \end{aligned}$$

since $N < 32/\epsilon$ by definition of N. Now, take ϵ equal to $(c\log m + y)^\kappa/r_n^2$ and z^2 equal to $y - (\log\epsilon)/c_{15}$. Since $-\log\epsilon \le 2\log r_n \le c\log m$ for $y \ge 1$, we have

$$z\sqrt{\epsilon} = \sqrt{\epsilon(y - (\log\epsilon)/c_{15})} \le (c\log m + y)^{(\kappa+1)/2}/r_n,$$

for every $y > (\log\epsilon)/c_{15}$, and hence

$$\begin{aligned} P\left(r_n\omega(\epsilon) \ge (c\log m + y)^{(\kappa+1)/2}\right) &\le \frac{c_{14}}{\epsilon}\exp\left\{-c_{15}\left(y - (\log\epsilon)/c_{15}\right)\right\} \\ &\le c_{14}\exp\left\{-c_{15}y\right\} \end{aligned}$$

for every $y > (\log\epsilon)/c_{15}$. This completes the proof of Lemma 10. □

Our main tool in handling Gaussian processes will be the inequality given in Borell (1975). The formulation below is taken from Samorodnitsky (1991). Observe that $P\left(\sup_{t\in M}|X(t)| > y\right)$ is bounded by $2P\left(\sup_{t\in M} X(t) > y\right)$. Moreover, observe that Inequality 2 is relevant for Kiefer processes as well, as (22) implies

$$\mathcal{E}\sup_{w\in[0,1]}\sup_{u\in[0,1]}|\Psi(w,u)| \le \int_0^\infty P\left(\sup_{w\in[0,1]}\sup_{u\in[0,1]}|\Psi(w,u)| > y\right)dy \le c_{16}, \tag{24}$$

with $c_{16} = \frac{1}{2}c_{12}(c_{13})^{-1/2}\sqrt{\pi}$.

Inequality 2 (Borell's inequality) *Let $\{X(t) : t \in M\}$ be a zero mean separable Gaussian process, and let σ^2 denote $\sup_{t\in M} \mathcal{E}(X(t))^2$. If $\mu = \mathcal{E} \sup_{t\in M} X(t)$ exists, then for any $y > \mu$,*

$$P\left(\sup_{t\in M} X(t) > y\right) \leq 2e^{y\mu/\sigma^2} e^{-y^2/2\sigma^2}.$$

The DKW-inequality [Dvoretzky, Kiefer and Wolfowitz (1956)] is our main tool for handling empirical processes. Below we present the extended version of Bretagnolle (1980) [cf. Inequality 25.1.2 in Shorack and Wellner (1986), p. 797] which allows the random variables $X_1, \ldots, X_m$ to have different distributions. In case these random variables have a common distribution, one may replace $2e \exp\{-2y^2\}$ by $2 \exp\{-2y^2\}$ [cf. Csörgő and Horváth (1993), p. 119].

Inequality 3 (DKW-inequality) *Let $X_1, \ldots, X_m$ be independent random variables, and let $F_\ell(x)$ denote the cumulative distribution function of X_ℓ. Then, for every $y > 0$,*

$$P\left(\sup_{x\in \boldsymbol{R}} \left| m^{-1/2} \sum_{\ell=1}^{m} \left(1_{\{X_\ell \leq x\}} - F_\ell(x)\right)\right| \geq y\right) \leq 2e \exp\{-2y^2\}.$$

6 Proofs

This section contains the proofs of Theorems 1, and Lemma's 1, 3, 5 and 6. The proofs make use of the technical results collected in Section 5. The proofs of Lemma's 2 and 4 are straightforward, and hence not included.

Proof of Theorem 1 Carefully form the pooled sample $X_1, \ldots, X_m$ by amalgamating the n subsamples, leaving the order "between" and "within" samples intact. Define the random variables $Y_1, \ldots, Y_m$ by $Y_\ell = F(X_\ell)$, and observe that $Y_1, \ldots, Y_m$ are independent random variables having a standard uniform distribution. Moreover, we may express $A_n(t, x)$ as $\tilde{U}_n(\mu_n(t), F(x))$, where

$$\tilde{U}_n(w, u) = m^{-1/2} \sum_{\ell=1}^{[mw]} \left(1_{\{Y_\ell \leq u\}} - u\right), \quad w \in [0, 1],$$

is the sequential uniform empirical process derived from $Y_1, \ldots, Y_m$. According to Inequality 1 there exists a Kiefer process $\Psi(w, u)$ such that

$$\frac{m^{1/2}}{\log m} \max_{k=1,2,\ldots,m} \sup_{(k-1)/m \leq w \leq k/m} \left|\tilde{U}_n(w, u) - m^{-1/2}\Psi(k, u)\right| \in \mathcal{C}_0(1).$$

Let $\omega(\epsilon)$ be as defined in Lemma 10, and observe that $m^{1/2}\omega\left(m^{-1}\right) \in \mathcal{C}_0\left(\frac{1}{2}\right)$ by Lemma 10. Since

$$\begin{aligned}
&\sup_{t\in[0,1]}\sup_{u\in[0,1]}\left|\tilde{U}_n(\mu_n(t),u)-\Psi_n(\mu_n(t),u)\right| \\
&\quad\le \sup_{u\in[0,1]}\max_{k=1,2,\dots,m}\sup_{(k-1)/m<w\le k/m}\left|\tilde{U}_n(w,u)-\Psi_n(w,u)\right| \\
&\quad\le \sup_{u\in[0,1]}\max_{k=1,2,\dots,m}\sup_{(k-1)/m<w\le k/m}\left|\Psi_n(w,u)-\Psi_n\left(\frac{k}{m},u\right)\right| \\
&\qquad+\sup_{u\in[0,1]}\max_{k=1,2,\dots,m}\left|\tilde{U}_n(k/m,u)-\Psi_n\left(\frac{k}{m},u\right)\right| \\
&\quad\le \omega\left(m^{-1}\right)+\sup_{u\in[0,1]}\max_{k=1,2,\dots,m}\left|\tilde{U}_n(k/m,u)-\Psi_n\left(\frac{k}{m},u\right)\right|,
\end{aligned}$$

we obtain

$$\left(\frac{m^{1/2}}{\log m}\right)\sup_{t\in[0,1]}\sup_{u\in[0,1]}\left|\tilde{U}_n(\mu_n(t),u)-\Psi_n(\mu_n(t),u)\right|\in\mathcal{C}_0(1)\,. \tag{25}$$

Moreover, we have

$$\begin{aligned}
&P\left(\sup_{t\in[0,1]}\sup_{u\in[0,1]}\left|\Psi_n(\mu_n(t),u)-\Psi_n(\mu(t),u)\right|>y_n\right) \\
&\qquad\le P\left(\sup_{t\in[0,1]}\left|\mu_n(t)-\mu(t)\right|>\epsilon\right)+P\left(\omega\left(\epsilon\right)\ge y_n\right)
\end{aligned} \tag{26}$$

for every $y_n > 0$, $\epsilon > 0$ and $P \in \mathcal{P}_0$. Now, take y_n equal to $(c\log m + y)^{\alpha/2+1/4}$, and ϵ equal to $(c\log m+y)^{\alpha-1/2}\left(s_n\right)^{-2}$. According to Condition 1 and Lemma 10 we have $s_n^2\sup_{t\in[0,1]}\left|\mu_n(t)-\mu(t)\right| \in \mathcal{C}_0(\alpha-1/2)$ and $s_n\omega\left(\epsilon\right)\in\mathcal{C}_0(\alpha/2+1/4)$, respectively. Thus, (26) implies

$$s_n\sup_{t\in[0,1]}\sup_{u\in[0,1]}\left|\Psi_n(\mu_n(t),u)-\Psi_n(\mu(t),u)\right|\in\mathcal{C}_0\left(\frac{\alpha}{2}+\frac{1}{4}\right). \tag{27}$$

Define $U_n\left(t,x\right)$ as $\Psi_n(\mu(t),F(x))$, $t\in[0,1]$, $x\in I\!\!R$. It is easily seen that the process $U_n\left(t,x\right)$ indeed is a zero mean Gaussian process with covariance function (5). Combining (25) and (27) yields (6), which completes the proof of Theorem 1. □

Proof of Lemma 1 Observe that $B_n\left(t,x\right)$ coincides with $A_n\left(t,x\right)-\mu_n(t)A_n\left(1,x\right)$. Let $U_n\left(t,x\right)$ be the Gaussian process occurring in Theorem 1, and define the process $V_n\left(t,x\right)$ by $V_n\left(t,x\right)=U_n\left(t,x\right)-\mu(t)U_n\left(1,x\right)$, $t\in[0,1]$, $x\in I\!\!R$. It is easily seen that $V_n\left(t,x\right)$ is a zero mean Gaussian process with covariance function (8).

Let c_{16} be as in (24), and observe that

$$\mathcal{E}\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|U_n\left(t,x\right)\right|\le\mathcal{E}\sup_{w\in[0,1]}\sup_{u\in[0,1]}\left|\Psi_n(w,u)\right|\le c_{16}. \tag{28}$$

As $\mathcal{E}\left\{U_n\left(t,x\right)\right\}^2$ remains bounded by $1/4$, Inequality 2 yields that $\sup_x\left|U_n\left(1,x\right)\right|$ belongs to $\mathcal{C}_0\left(\frac{1}{2}\right)$, which together with Condition 1 implies

$$s_n^2 \sup_{t\in[0,1]}\left|\mu_n(t)-\mu(t)\right|\sup_{x\in\boldsymbol{R}}\left|U_n\left(1,x\right)\right|\in\mathcal{C}_0(\alpha)\,. \tag{29}$$

Since

$$\begin{aligned}\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|B_n\left(t,x\right)-V_n\left(t,x\right)\right| &\le \sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|A_n\left(t,x\right)-U_n\left(t,x\right)\right|\\ &+\sup_{t\in[0,1]}\mu_n(t)\sup_{x\in\boldsymbol{R}}\left|A_n\left(1,x\right)-U_n\left(1,x\right)\right|\\ &+\sup_{t\in[0,1]}\left|\mu_n(t)-\mu(t)\right|\sup_{x\in\boldsymbol{R}}\left|U_n\left(1,x\right)\right|,\end{aligned}$$

Lemma 1 now follows by combining Theorem 1 and (29). □

Proof of Lemma 3 As $K(x)$ is a probability density function, we have that $K(x)$ tends to zero as x tends to $\pm\infty$, and hence

$$K\left(\frac{X_{i,j}-x}{h}\right)=-h^{-1}\int 1_{\{X_{i,j}\le v\}}K'\left(\frac{v-x}{h}\right)dv. \tag{30}$$

It follows that

$$\hat{f}_i(x)=-\frac{1}{h^2}\int\hat{F}_i(v)K'\left(\frac{v-x}{h}\right)dv,\ \ \text{and}\ \ \bar{f}_{n,h}(x)=-\frac{1}{h^2}\int\bar{F}_n(v)K'\left(\frac{v-x}{h}\right)dv,$$

which allows us to write

$$B_{n,h}\left(t,x\right)=-h^{-3/2}\int B_n\left(t,v\right)K'\left(\frac{v-x}{h}\right)dv.$$

Now, define the process $V_{n,h}\left(t,x\right)$ by

$$V_{n,h}\left(t,x\right)=-h^{-3/2}\int V_n\left(t,v\right)K'\left(\frac{v-x}{h}\right)dv,\quad t\in[0,1],\ x\in\mathbb{R},$$

where $V_n\left(t,x\right)$ is the Gaussian process approximating $B_n\left(t,x\right)$ in Lemma 1. Since $V_{n,h}\left(t,x\right)$ is merely a linear transformation of the zero mean Gaussian process $V_n\left(t,x\right)$, it follows that the former process is Gaussian; moreover, it is easily seen that $V_{n,h}\left(t,x\right)$ is zero mean. As the covariance function of $V_n\left(t,x\right)$ satisfies (8), the covariance function of $V_{n,h}\left(t,x\right)$ may be expressed as

$$\begin{aligned}&\mathcal{E}V_{n,h}\left(t,x\right)V_{n,h}(t',x')\\ &\quad=h^3\int\int\left\{\mathcal{E}V_n\left(t,v\right)V_n(t',v')\right\}K'\left(\frac{v-x}{h}\right)K'\left(\frac{v'-x'}{h}\right)dvdv'\\ &\quad=\left\{\mu\left(t\wedge t'\right)-\mu(t)\mu(t')\right\}h^{-3}\\ &\qquad\times\int\int\left\{F\left(v\wedge v'\right)-F(v)F(v')\right\}K'\left(\frac{v-x}{h}\right)K'\left(\frac{v'-x'}{h}\right)dvdv'.\end{aligned}$$

As we have

$$
\begin{aligned}
&\int\int F(v\wedge v')\,K'\left(\frac{v-x}{h}\right)K'\left(\frac{v'-x'}{h}\right)dvdv'\\
&\quad=\int\int\int_{-\infty}^{v\wedge v'} f(u)duK'\left(\frac{v-x}{h}\right)K'\left(\frac{v'-x'}{h}\right)dvdv'\\
&\quad=\int\int_u^\infty\int_u^\infty K'\left(\frac{v-x}{h}\right)K'\left(\frac{v'-x'}{h}\right)dvdv'f(u)du\\
&\quad=h^2\int K\left(\frac{u-x}{h}\right)K\left(\frac{u-x'}{h}\right)f(u)du
\end{aligned}
$$

and

$$
\int\int F(v)F(v')K'\left(\frac{v-x}{h}\right)K'\left(\frac{v'-x'}{h}\right)dvdv'=h^2\mathcal{E}\bar f_{n,h}(x)\mathcal{E}\bar f_{n,h}(x'),
$$

it follows that $V_{n,h}\,(t,x)$ is indeed a zero-mean Gaussian process with covariance function (13). Finally, because

$$
\begin{aligned}
&\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|B_{n,h}\,(t,x)-V_{n,h}\,(t,x)\right|\\
&\quad\le\left\{\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|B_n\,(t,x)-V_n\,(t,x)\right|\right\}\left\{\sup_{x\in\boldsymbol{R}}h^{-3/2}\int\left|K'\left(\frac{v-x}{h}\right)\right|dv\right\}\\
&\quad\le\left\{\sup_{t\in[0,1]}\sup_{x\in\boldsymbol{R}}\left|B_n\,(t,x)-V_n\,(t,x)\right|\right\}\left\{h^{-1/2}\int\left|K'(v)\right|dv\right\},
\end{aligned}
$$

Condition 3 yields that (14) follows from (9). This concludes the proof of Lemma 3. □

Proof of Lemma 5 By the argument given in Shorack (1982) [see also Section 23.1 in Shorack and Wellner (1986), p. 763], we may assume the existence of independent standard uniform random variables Y_{ij}^* such that $X_{ij}^*=\bar F_n^{-1}(Y_{ij}^*)$. We have that $A_n^*\,(t,x)$, the bootstrapped version of $A_n\,(t,x)$, coincides with $\tilde A_n\left(t,\bar F_n(x)\right)$, where

$$
\tilde A_n\,(t,u)=m^{-1/2}\sum_{i=1}^{[nt]}\sum_{j=1}^{m_i}\left(1_{\{Y_{ij}^*\le u\}}-u\right)\quad\text{for } u\in[0,1].
$$

It follows by Theorem 1 that there exists a zero mean Gaussian process $\tilde U_n'\,(t,u)$ with covariance function $\mu\,(t\wedge t')\,\{u\wedge u'-uu'\}$ such that

$$
\left(s_n\vee\frac{m^{1/2}}{\log m}\right)\sup_{t\in[0,1]}\sup_{u\in[0,1]}\left|\tilde A_n\,(t,u)-\tilde U_n'\,(t,u)\right|\in\mathcal{C}_0\left(\left(\frac{\alpha}{2}+\frac{1}{4}\right)\wedge 1\right),
$$

which implies

$$\left(s_n \vee \frac{m^{1/2}}{\log m}\right) \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left|\tilde{A}_n(t, F_n(x)) - \tilde{U}'_n(t, F_n(x))\right| \in \mathcal{C}_0\left(\left(\frac{\alpha}{2} + \frac{1}{4}\right) \wedge 1\right). \quad (31)$$

As $\tilde{U}'_n(t,u)$ is equal in distribution to $Z(\mu(t), u) - uZ(\mu(t), 1)$, where $Z(w,u)$ is a two-parameter Brownian motion on $[0,1]^2$ [that is, a mean zero Gaussian process with covariance function $(w \wedge w')(u \wedge u')$], we may write

$$\begin{aligned} P&\left(\sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left|\tilde{U}'_n\left(t, \bar{F}_n(x)\right) - \tilde{U}'_n(t, F(x))\right| > 2(\log m + y)\right) \\ &\leq P\left(\sup_{x\in\boldsymbol{R}} \left|\bar{F}_n(x) - F(x)\right| > \epsilon\right) + P\left(\epsilon \sup_{w\in[0,1]} |Z(w,1)| > \log m + y\right) \\ &+ P\left(\sup_{w\in[0,1]} \sup_{|u-u'|\leq\epsilon} |Z(w,u) - Z(w,u')| > \log m + y\right). \end{aligned} \quad (32)$$

Now, take ϵ equal to $m^{-1/2}(\log m + y)^{1/2}$, and remark that Inequality 3 yields

$$P\left(\sup_{x\in\boldsymbol{R}} \left|\bar{F}_n(x) - F(x)\right| > \epsilon\right) \leq P\left(m^{1/2} \sup_{x\in\boldsymbol{R}} \left|\bar{F}_n(x) - F(x)\right| > y^{1/2}\right) \leq 2e^{-2y}. \quad (33)$$

Note that $Z(\cdot, 1)$ is a Brownian motion on $[0,1]$. As $\mathcal{E} \sup_{w\in[0,1]} |Z(w,1)|$ is finite, Inequality 2 implies $\sup_{w\in[0,1]} |Z(w,1)| \in \mathcal{C}_0\left(\frac{1}{2}\right)$, and hence it follows that

$$m^{1/2}\epsilon \sup_{w\in[0,1]} |Z(w,1)| \in \mathcal{C}_0(1). \quad (34)$$

Moreover, a similar argument as given in the proof of Lemma 10 yields that

$$m^{1/4} \sup_{w\in[0,1]} \sup_{|u-u'|\leq\epsilon} |Z(w,u) - Z(w,u')| \in \mathcal{C}_0(1). \quad (35)$$

Define $U'_n(t,x)$ by $\tilde{U}'_n(t, F(x))$, and observe that $U'_n(t,x)$ is a zero mean Gaussian process with covariance function (5). Combining (31)–(35) yields that

$$\left(s_n \vee m^{1/4}\right) \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} |A^*_n(t,x) - U'_n(t,x)| \in \mathcal{C}_0\left(\left(\frac{\alpha}{2} + \frac{1}{4}\right) \wedge 1\right). \quad (36)$$

Next, consider $B^*_n(t,x)$, the bootstrapped version of $B_n(t,x)$. Since we may write $B^*_n(t,x) = A^*_n(t,x) - \mu_n(t) A^*_n(1,x)$, it follows as in the proof of Lemma 1 that for every n there exists a zero mean Gaussian process $V'_n(t,x) = U'_n(t,x) - \mu(t) U'_n(1,x)$ with covariance function (8) such that

$$\left(s_n \vee m^{1/4}\right) \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} |B^*_n(t,x) - V'_n(t,x)| \in \mathcal{C}_0(\alpha). \quad (37)$$

Similar reasoning as in the proof of Lemma 3 yields for $B^*_{n,h}(t,x)$, the bootstrapped version of $B_n(t,x)$, that for every n there exists a zero mean Gaussian process with covariance function (13) such that

$$\left(s_n \vee m^{1/4}\right) h^{1/2} \sup_{t\in[0,1]} \sup_{x\in\boldsymbol{R}} \left| B^*_{n,h}(t,x) - V'_{n,h}(t,x) \right| \in \mathcal{C}_0(\alpha)\,. \tag{38}$$

As we may write

$$C^*_n(t,x) = L(t) B^*_n(t,x) - \int_0^t B^*_n(v,x)\, dL(v),$$

$$C^*_{n,h}(t,x) = L(t) B^*_{n,h}(t,x) - \int_0^t B^*_{n,h}(v,x)\, dL(v),$$

(18) and (19) follow from (37) and (38). This completes the proof of Lemma 5. □

Proof of Lemma 6 As by (i) the distribution of $\tilde{T}'_n$ does not depend on n, (ii) implies that $\nu_0^{-1} T'_n$ converges in distribution to $\nu_0^{-1}\tilde{T}_1$ for every $P \in \mathcal{P}_0$. Moreover, (iii) implies that $\nu_0/\hat{\nu}$ converges in distribution to 1 for every $P \in \mathcal{P}_0$. It follows by Slutsky's Theorem that $\hat{\nu}^{-1} T_n = (\nu_0/\hat{\nu})(\nu_0^{-1} T_n)$ converges in distribution to ν_0^{-1} for every $P \in \mathcal{P}_0$.

To prove (20), take $\epsilon \in (0,1)$, and observe that (i) implies that there exists $y_\epsilon \geq 0$ such that

$$\sup_{P\in\mathcal{P}_0} P\left(\nu_0^{-1}\tilde{T}'_n \geq y^{1/2}\right) \leq \exp\{-c_3 y\}$$

for every $y \geq y_\epsilon$, where $c_3 = (1-\epsilon)a/2$. By taking $c_1 = 0$ and $c_2 = \exp\{c_3 y_\epsilon\}$, it follows that

$$\nu_0^{-1}\tilde{T}'_n \in \mathcal{C}_0\left(\tfrac{1}{2}\right).$$

Write

$$\left|\hat{\nu}^{-1} T'_n - \nu_0^{-1}\tilde{T}'_n\right| \leq \Delta_{n1} + \Delta_{n2}\left\{\nu_0^{-1}\tilde{T}'_n + \Delta_{n1}\right\},$$

where

$$\Delta_{n1} = \nu_0^{-1}\left|T'_n - \tilde{T}'_n\right|, \quad \Delta_{n2} = |\nu_0/\hat{\nu} - 1|\,.$$

Since $r_n\Delta_{n1} \in \mathcal{C}_0(\kappa)$ by condition (ii), and $r_n\Delta_{n2} \in \mathcal{C}_0(\kappa')$ by condition (iii), we obtain

$$r_n \left|\hat{\nu}^{-1} T'_n - \nu_0^{-1}\tilde{T}'_n\right| \in \mathcal{C}_0(\kappa + \kappa')\,.$$

Let β denote $(2\kappa + 2\kappa' - 1)^{-1}$, and observe that $\kappa' \geq 1 - \kappa$ implies $\beta \leq 1$. Let y_n be a sequence of scalars such that $y_n \to \infty$ and $y_n = o\left((r_n)^\beta\right)$ as $n \to \infty$, and let x_n denote $(r_n)^\beta y_n$. Since $x_n = o\left((r_n)^{2\beta}\right) = o(m)$, $\log m = o(x_n)$ and $(y_n)^2 = o(x_n)$ as $n \to \infty$, it follows that

$$\begin{aligned} P\left(r_n \left|\hat{\nu}^{-1} T'_n - \nu_0^{-1}\tilde{T}'_n\right| > (x_n)^{\kappa+\kappa'}\right) &\leq c_2 \exp\{-c_3 (x_n - c_1 \log m)\} \\ &= o\left(\exp\left\{-\tfrac{1}{2} a (y_n)^2\right\}\right) \end{aligned}$$

as $n \to \infty$. Now, for every $P \in \mathcal{P}_0$ we may bound $P\left(\hat{\nu}^{-1}T_n' > y_n\right)$ between

$$P\left(\nu_0^{-1}T_n' > y_n + (x_n)^{\kappa+\kappa'}/r_n\right) - P\left(r_n\left|\hat{\nu}^{-1}T_n' - \nu_0^{-1}\tilde{T}_n'\right| > (x_n)^{\kappa+\kappa'}\right)$$

and

$$P\left(\nu_0^{-1}T_n' > y_n - (x_n)^{\kappa+\kappa'}/r_n\right) + P\left(r_n\left|\hat{\nu}^{-1}T_n' - \nu_0^{-1}\tilde{T}_n'\right| > (x_n)^{\kappa+\kappa'}\right).$$

Since $(x_n)^{\kappa+\kappa'}/r_n = (r_n)^{\beta(\kappa+\kappa')}(y_n)^{\kappa+\kappa'-1}y_n/r_n = o\left((r_n)^{\beta(2\kappa+2\kappa'-1)}y_n\right) = o(y_n)$, this completes the proof of Lemma 6. □

Proof of Lemma 7 As the covariance function of $V_n\,(t,x)$ has product structure, it follows by Lemma 4 in Koning and Protassov (2001) that the reproducing kernel Hilbert space of the mean zero Gaussian process $V_n\,(t,x)$ is equal to the tensor product $\mathcal{H}_1 \circ \mathcal{H}_2$, where $\mathcal{H}_1$ is the reproducing kernel Hilbert space corresponding to the covariance function $\mu(t \wedge t') - \mu(t)\mu(t')$, and $\mathcal{H}_2$ is the reproducing kernel Hilbert space corresponding to the covariance function $x \wedge x' - xx'$.

Observe that $T = T_1 \circ T_2$, with $T_1 = S$, and T_2 defined by $T_2(\xi) = \sup_{x\in\boldsymbol{R}}|\xi(x)|$ for $\xi \in D\,(I\!R)$. We have $\|T_1\|_{\mathcal{H}_1} = 4$ [cf. the example following Lemma 3 in Koning and Protassov (2001)], and $\|T_2\|_{\mathcal{H}_2} = \|S\|_{\mathcal{H}_2} = a_S$. Combining Corollary 1 and Inequality 1 in Koning and Protassov (2001) yields that condition (i) of Lemma 6 holds with $\nu_0 = 1$ and

$$a = \|T\|_{\mathcal{H}_1\circ\mathcal{H}_2} = \|T_1\|_{\mathcal{H}_1} \cdot \|T_2\|_{\mathcal{H}_2} = 4a_S.$$

This completes the proof of Lemma 7. □

Proof of Lemma 8 As the covariance function of $V_{n,h}\,(t,x)$ has product structure, it follows by Lemma 4 in Koning and Protassov (2001) that the reproducing kernel Hilbert space of the mean zero Gaussian process $V_{n,h}\,(t,x)$ is equal to the tensor product $\mathcal{H}_1 \circ \mathcal{H}_2$, where $\mathcal{H}_1$ is the reproducing kernel Hilbert space corresponding to the covariance function $\mu(t \wedge t') - \mu(t)\mu(t')$, and $\mathcal{H}_2$ is the reproducing kernel Hilbert space corresponding to the covariance function $\sigma_h(x,x')$. Observe that $T = T_1 \circ T_2$, with T_1 and T_2 as in the proof of Lemma 7. Since $\|T_1\|_{\mathcal{H}_1} = \sup \sigma_h(x,x)$ and $\|T_2\|_{\mathcal{H}_2} = \|S\|_{\mathcal{H}_2} = a_S$, combining Corollary 1 and Inequality 1 in Koning and Protassov (2001) yields that condition (i) of Lemma 6 holds with $\nu_0^{-1} = \sup_{x\in\boldsymbol{R}}\sigma_h(x,x)$ and $a = a_S$. This completes the proof of Lemma 8 (i).

We continue with the proof of Lemma 8(ii). Observe that $\hat{\nu}^2 = \sup_{x\in\boldsymbol{R}}\hat{\sigma}_{n,h}(x,x)$, where $\hat{\sigma}_{n,h}(x,x)$ is defined by

$$\hat{\sigma}_{n,h}(x,x) = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{K\left(\frac{X_{i,j}-x}{h}\right) - \bar{f}_{n,h}(x)\right\}^2.$$

According to (30), we may write

$$\bar{f}_{n,h}(x) = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K\left(\frac{X_{i,j}-x}{h}\right) = -\frac{1}{mh^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int 1_{\{X_{i,j}\le v\}}K'\left(\frac{v-x}{h}\right)dv.$$

Hence, we have

$$\hat{\sigma}_{n,h}(x,x) + h\left(\bar{f}_{n,h}(x)\right)^2 = \frac{1}{mh}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{K\left(\frac{X_{i,j}-x}{h}\right)\right\}^2$$
$$= -\frac{2}{mh^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\int 1_{\{X_{i,j}\le v\}}K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv.$$

Similarly, integration by parts yields

$$\sigma_h(x,x) + h\left(\mathcal{E}\bar{f}_{n,h}(x)\right)^2 = -2h^{-2}\int F(v)K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv,$$
$$\mathcal{E}\bar{f}_{n,h}(x) = -\frac{1}{h^2}\int F(v)K'\left(\frac{v-x}{h}\right)dv. \tag{39}$$

Introduce

$$\Delta_n = m^{-1/2}\sup_{x\in\boldsymbol{R}}\left|m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le x\}} - F(x)\right|.$$

Since Condition 3 implies

$$\int|K'(u)|\,du < c_6, \quad 2\int K(u)\,|K'(u)|\,du < c_6^2,$$

and since (39) implies $\sup_{x\in\boldsymbol{R}}\mathcal{E}\bar{f}_{n,h}(x) \le c_6/h$, we may write

$$\begin{aligned} h\left|\hat{\nu}^2 - \nu_0^2\right| &\le h\sup_{x\in\boldsymbol{R}}|\hat{\sigma}_{n,h}(x,x) - \sigma_h(x,x)| \\ &\le 2c_6\Delta_{n1} + (\Delta_{n1})^2 + \Delta_{n2}, \\ &\le 3c_6\Delta_{n1} + \Delta_{n2} \quad \text{if} \quad \Delta_{n1} \le c_6, \end{aligned}$$

where

$$\begin{aligned} \Delta_{n1} &= h\sup_{x\in\boldsymbol{R}}\left|\bar{f}_{n,h}(x) - \mathcal{E}\bar{f}_{n,h}(x)\right| \\ &\le \sup_{x\in\boldsymbol{R}}\left|h^{-2}\int\left(\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le v\}} - F(v)\right)K'\left(\frac{v-x}{h}\right)dv\right| \\ &\le c_6 m^{-1/2}\Delta_n, \end{aligned}$$

and

$$\begin{aligned} \Delta_{n2} &= h\sup_{x\in\boldsymbol{R}}\left|\left\{\hat{\sigma}_{n,h}(x,x) + h\left(\bar{f}_{n,h}(x)\right)^2\right\} - \left\{\sigma_h(x,x) + h\left(\mathcal{E}\bar{f}_{n,h}(x)\right)^2\right\}\right| \\ &\le \sup_{x\in\boldsymbol{R}}\left|2h^{-2}\int\left(\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m_i}1_{\{X_{i,j}\le v\}} - F(v)\right)K\left(\frac{v-x}{h}\right)K'\left(\frac{v-x}{h}\right)dv\right| \\ &\le c_6^2 m^{-1/2}\Delta_n. \end{aligned}$$

Hence, we obtain

$$h\left|\hat{\nu}^2-\nu_0^2\right| \le (2c_6)^2 m^{-1/2}\Delta_n \quad \text{if} \quad \Delta_n \le m^{1/2}. \tag{40}$$

Since $(2c_8)^2\left|\hat{\nu}^2-\nu_0^2\right| \le 3$ implies

$$\left|(\hat{\nu}/\nu_0)^2-1\right| \le (\nu_0)^{-2}\left|\hat{\nu}^2-\nu_0^2\right| \le c_8^2\left|\hat{\nu}^2-\nu_0^2\right| \le \tfrac{3}{4},$$

and $\left|x^{-1/2}-1\right|$ is bounded by $4\left|x-1\right|$ for $x \ge \frac{1}{4}$, it follows that

$$|\nu_0/\hat{\nu}-1| \le 4\left|(\hat{\nu}/\nu_0)^2-1\right| \le (2c_8)^2\left|\hat{\nu}^2-\nu_0^2\right| \quad \text{if} \quad (2c_8)^2\left|\hat{\nu}^2-\nu_0^2\right| \le 3. \tag{41}$$

As Inequality 3 yields

$$\sup_{P\in\mathcal{P}_0} P\left(\Delta_n > y^{1/2}\right) \le 2e\exp\{-2y\}$$

for every $y > 0$, Lemma 8(ii) follows by combining (40) and (41) [take $c_{17} = 2e$, $c_{18} = 2/(4c_6c_8)^2$ and $c_{19} = 3/(4c_6c_8)^2$]. □

7 Acknowledgements

The speedskating data come from the authors' personal files, but can also be found at Jeroen Heijmans' website [http://weasel.student.utwente.nl/~speedskating/]. We thank C.F. de Vroege for his helpfulness with the data. Alex J. Koning is grateful for hospitality and financial support in connection with visits to the Department of Mathematics at the University of Oslo. Nils Lid Hjort is grateful to the Tinbergen Institute at the Erasmus University Rotterdam for hospitality and financial support.

References

[1] Adler, R.J. (1990). *An introduction to continuity, extrema, and related topics for general Gaussian processes*. Institute of Mathematical Statistics, Hayward CA.

[2] Bahadur, R.R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31**, 276–295.

[3] Bickel, P.J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics* **1**, 1071–1095. Corrections, ibid. **3**, 1370.

[4] Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Inventiones math.* **30**, 207–216.

[5] Bretagnolle, J. (1980). Statistique de Kolmogorov-Smirnov pour un enchantillon nonequireparti. *Colloq. internat. CNRS* **307**, 39–44.

[6] Csörgő, M., Horváth, L. (1993). *Weighted approximations in probability and statistics*. Wiley, New York.

[7] Csörgő, M., Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley, New York.

[8] Csörgő, M., Révész, P. (1975). A new method to prove Strassen type laws of invariance principle. I, II. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **31**, 255–260; 261–269.

[9] Csörgő, M., Révész, P. (1981). *Strong approximations in probability and statistics*. Academic Press, New York.

[10] de la Peña, V.H., Giné, E. (1999). *Decoupling. From dependence to independence.* Springer-Verlag, New York.

[11] Does, R.J.M.M., Koning, A.J. (2000). CUSUM charts for preliminary analysis of individual observations. *Journal of Quality Technology* **32**, 122–132.

[12] Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics* **27**, 642–669.

[13] Efron B., Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Monographs on statistics and applied probability. Chapman and Hall, London.

[14] Hjort, N.L., Koning, A.J. (2001). Tests for the constancy of model parameters over time. *Journal of Nonparametric Statistics*, to appear.

[15] Hjort, N.L., Koning, A.J. (2002). Constancy of distributions: asymptotic efficiency of certain nonparametric tests of constancy. *In preparation*.

[16] Inglot, T. and Ledwina, T. (1990). On probabilities of excessive deviations for Kolmogorov-Smirnov, Crámer-von Mises and chi-square statistics. *Annals of Statistics* **18**, 1491–1495.

[17] Inglot, T. and Ledwina, T. (1993). Moderately large deviations and expansions of large deviations for some functionals of weighted empirical process. *Annals of Probability* **21**,1691–1705.

[18] Kallenberg, W.C.M. (1983). Intermediate efficiency, theory and examples. *Annals of Statistics* **11**, 170-182.

[19] Kiefer, J. (1972). Skorohod embedding of multivariate RV's and the sample DF. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **24**, 1–35.

[20] Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent RV's and sample D.F. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwante Gebiete* **32**, 111–131.

[21] Konakov, V.D., Piterbarg, V.I. (1983). Rate of convergence of distributions of maximal deviations of Gaussian processes and empirical density functions. *Theory of Probability and its Applications* **28**, 172–178.

[22] Koning, A.J. (1992). Approximation of stochastic integrals with applications to goodness-of-fit tests. *Annals of Statistics* **20**, 428–454.

[23] Koning, A.J. (1994). Approximation of the basic martingale. *Annals of Statistics* **22**, 565–579.

[24] Koning, A.J., Protasov, V. (2001). Tail behaviour of Gaussian processes with applications to the Brownian pillow. *Econometric Institute Report EI 2001-49*.

[25] Ledoux, M., Talagrand, M. (1991). *Probability in Banach spaces*. Springer-Verlag, Berlin.

[26] Müller, D.W. (1970). On Glivenko-Cantelli convergence. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **16**, 195–210.

[27] Piterbarg, V.I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields*. American Mathematical Society, Providence, RI.

[28] Rosenblatt, M. (1971). Curve estimation. *Annals of Mathematical Statistics* **42**, 1815–1842.

[29] Rubin, H., Sethuraman, J. (1965). Bayes risk efficiency. *Sankhyā Ser. A* **27**, 347–356.

[30] Samorodnitsky, G. (1991). Probability tails of Gaussian extrema. *Stochastic Processes and their Applications* **38**, 55–84.

[31] Shorack, G.R. (1982). Bootstrapping robust regression. *Communications in Statistics* **A11**, 961–972.

[32] Shorack, G.R., Wellner, J.A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.

[33] Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.

[34] van der Vaart, A.W., Wellner, J.A. (1996). *Weak convergence and empirical processes, with applications to statistics* Springer, New York.

$T(\Gamma)$	Percentage points 0.10	0.05	0.01
$\sup_{x\in\boldsymbol{R}} S_{\text{Kol}}\left(\Gamma\left(\cdot,x\right)\right)$	0.7741	0.8331	0.9563
$\sup_{x\in\boldsymbol{R}} S_{\text{CvM}}\left(\Gamma\left(\cdot,x\right)\right)$	0.4099	0.4510	0.5355
$\sup_{x\in\boldsymbol{R}} S_{\text{AD}}\left(\Gamma\left(\cdot,x\right)\right)$	0.9355	1.0260	1.2121

Table 1: Approximated upper percentage points for various random variables $T(\Gamma)$, where Γ is the Brownian pillow [Koning and Protasov (2001)].

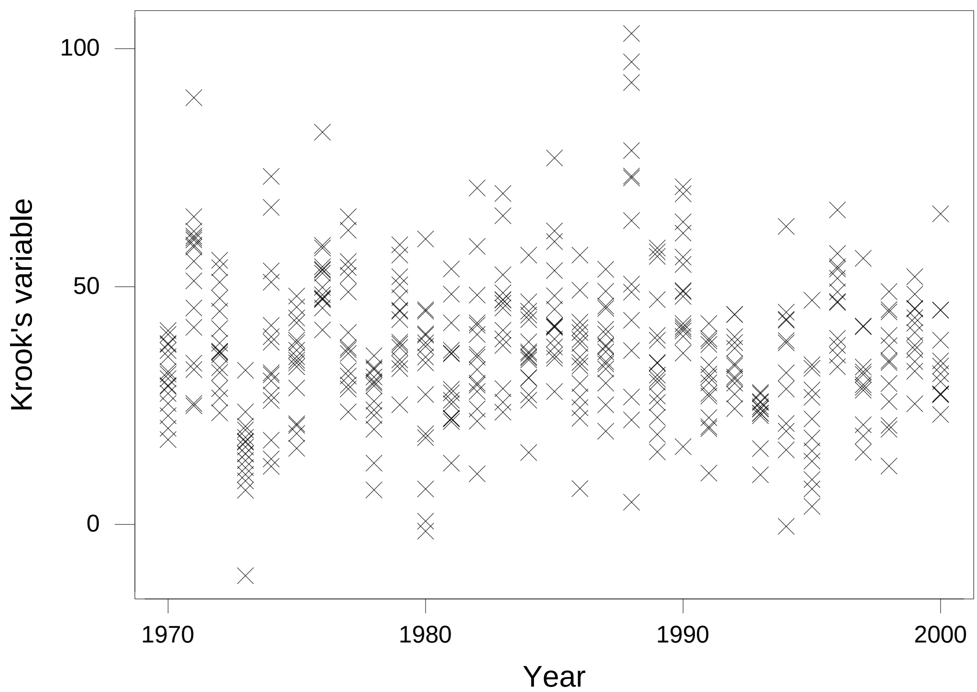

Figure 1: Plot of Krook's variable versus the year in which the event took place. The number m_i of participants depends on the relevant event: 12 [Calgary 1992, Hamar 1993, Gothenburg 1994, Baselga da Pine 1995, Inzell 1996, Nagano 1997, Heerenveen 1998, Hamar 1999, Milwaukee 2000], 14 [Heerenveen 1976, Gothenburg 1978, Oslo 1983, Alma Ata 1988, Heerenveen 1991], 15 [Inzell 1974, Heerenveen 1977, Oslo 1979, Heerenveen 1980, Oslo 1981, Assen 1982, Hamar 1985, Inzell 1986, Heerenveen 1987] or 16 [Oslo 1970, Gothenburg 1971, Oslo 1972, Deventer 1973, Oslo 1975, Gothenburg 1984, Oslo 1989, Innsbrück 1990].

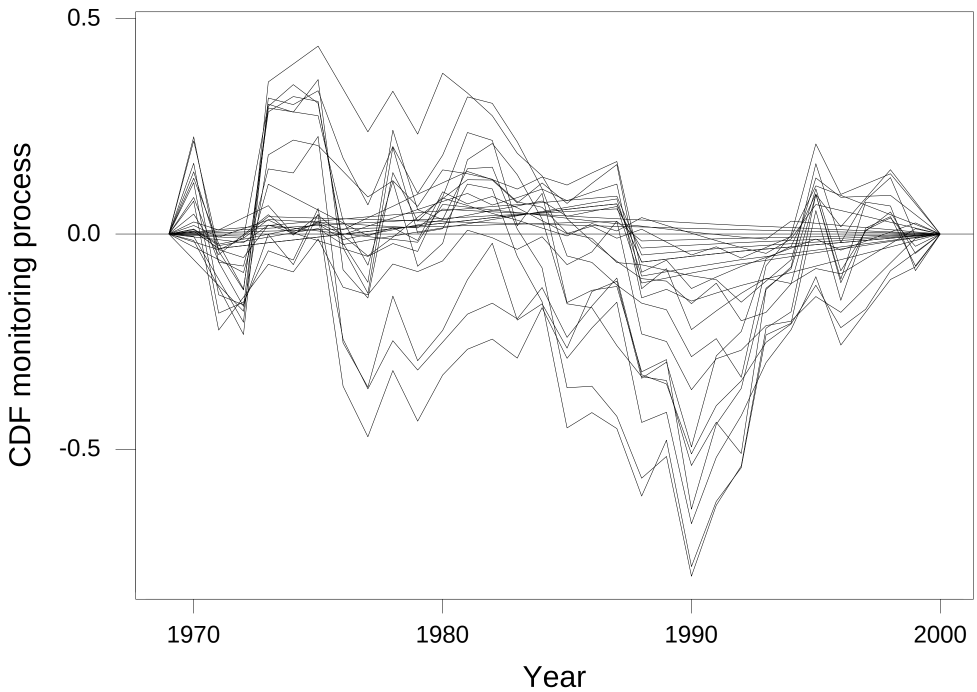

Figure 2: The monitoring process $B_n(t, x)$: plots of $B_n(t, x)$ versus t for various choices of x.

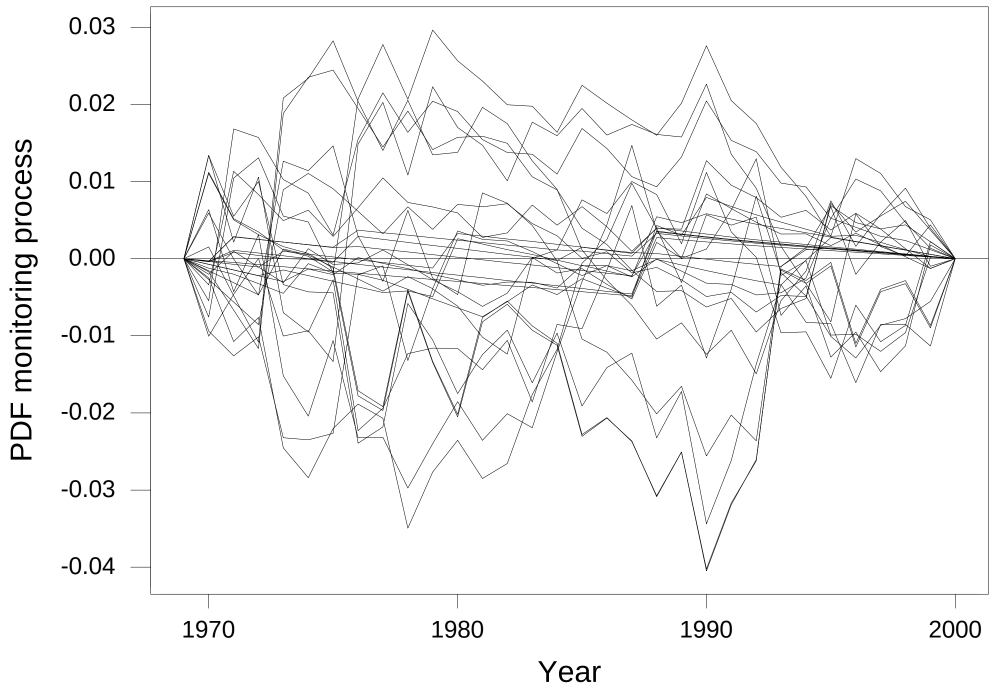

Figure 3: The monitoring process $B_{n,.75s}\,(t,x)$: plots of $B_{n,.75s}(\cdot,x)$ versus t for various choices of x.