Confidence Sets for a Changepoint Via Randomization Methods

By

Lutz Dümbgen

Technical Report No. 364
March 1993

Work supported by the Miller Institute for Basic Research in Science,
University of California, Berkeley

Department of Statistics
University of California
Berkeley, California 94720

CONFIDENCE SETS FOR A CHANGEPOINT VIA RANDOMIZATION METHODS

Lutz Dümbgen

ABSTRACT

Let $X(i)$, $i = 1, 2, \ldots, n$, be independent random variables with unknown distributions P for $i \leq n\theta$ and Q for $i > n\theta$. We investigate confidence sets for the unknown changepoint $\theta \in (0,1)$, which are based on randomization tests. In a simple parametric model for P and Q these tests are chosen to be Bayes-optimal in a certain sense. Then we imitate this method in a nonparametric framework. Asymptotic properties of the confidence sets are derived under weak conditions allowing that θ tends to zero or one and P is getting closer to Q.

Keywords and phrases: changepoint, randomization, permutation test, confidence set, Bayes-optimal, nonparametric

Work supported by the Miller Institute for Basic Research in Science, University of California at Berkeley

Author's address: Institut für angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany

1. Introduction

For each $n = 2, 3, \ldots$ let $X_n = (X_n(1), X_n(2), \ldots, X_n(n))$ be a vector of n independent random variables with values in a measurable space $\mathbf{X}$. Suppose that $X_n(i)$ has distribution P_n for $i \leq n\theta_n$ and Q_n otherwise, where P_n, Q_n are unknown, different probability measures on $\mathbf{X}$, and the changepoint θ_n is an unknown number in $\Theta_n := \{1/n, 2/n, \ldots, (n-1)/n\}$. The problem treated here is to find a confidence set for θ_n.

There is an extensive literature on this problem for models, where P_n and Q_n are assumed to be in a specified parametric family of distributions. Siegmund (1988) gives a good overview and references to other related work. Much less is known about nonparametric confidence sets. One possible method, which uses bootstrap tests, is described in Dümbgen (1991), but it relies on asymptotic theory. Alternatively we investigate parametric and nonparametric confidence sets that are both based on the classical method of randomization tests; see also Worsley (1986) and Siegmund (1986, 1988): Let $\mathcal{P}$ be a class of distributions containing P_n and Q_n. For each $\tau \in \Theta_n$ let $S_n^{(\tau)} = S_n^{(\tau)}(X_n)$ be a sufficient statistic for the restricted model, where $\theta_n = \tau$ and $P_n, Q_n \in \mathcal{P}$. Then consider a version $\mathbb{P}_n^{(\tau)}(\cdot|s)$ of $\mathcal{L}(X_n|S_n^{(\tau)} = s, \theta_n = \tau)$. For a given test statistic $T_n = T_n(X_n)$, one can compute the p-values $\hat{p}_n(\tau) = \hat{p}_n(\tau, X_n)$, where

$$\hat{p}_n(\tau, x) \;:=\; \mathbb{P}_n^{(\tau)}(T_n \geq T_n(x) | S_n^{(\tau)}(x)) \; .$$

Then

$$\hat{C}_n \;=\; \hat{C}_n(X_n) \;:=\; \{\tau \in \Theta_n : \hat{p}_n(\tau) > \alpha\}$$

defines a confidence set for θ_n with level $\alpha \in (0, 1/2)$. Note that this set is not necessarily an interval. Now the problem is to find suitable test statistics T_n and to study asymptotic properties of the corresponding sets $\hat{C}_n$.

Most papers on estimators or confidence sets use restrictive conditions on $P_n - Q_n$ and θ_n. One typical assumption is that θ_n is bounded away from 0 and 1, while $P_n - Q_n$ stays fixed or tends to 0 at a slow rate. A goal of the present paper is to relax such restrictions.

In section 2 we consider a simple normal shift model and derive a particular class of confidence sets, which are Bayes-optimal in a certain sense. Various asymptotic properties of these sets are presented.

Motivated by the parametric methods in section 2, we propose nonparametric confidence sets in section 3. They are based on permutation tests and use a formal Bayes-test statistic. The validity is now guaranteed without any restrictions on P_n and Q_n. These sets have similar asymptotic properties as the parametric confidence sets of section 2. An interesting reference in this context is Romano (1989), who discusses permutation tests of the hypothesis $P_n = Q_n$.

The results of sections 2 and 3 are proved in section 4.

2. The simple normal shift model

In this section we assume that $P_n = \mathcal{N}(\mu_n, 1)$ and $Q_n = \mathcal{N}(\nu_n, 1)$ with unknown means $\mu_n, \nu_n \in \mathbf{R}$. Thus X_n has an n-variate Gaussian distribution $\mathcal{N}(m, I)$ with mean vector $m = m(\theta_n, \mu_n, \nu_n)$ and identity covariance matrix I; generally $m(\tau, a, b)$ denotes the vector in $\mathbf{R}^n$ with the first $n\tau$ coordinates equal to a and the remaining $n - n\tau$ coordinates equal to b.

At first let us discuss briefly what can be expected from any confidence set

$C_n = C_n(X_n)$ with level α, where the size of C_n is measured by

$$\mathrm{dist}(C_n, \theta_n) \;:=\; \max_{t \in C_n} |t - \theta_n| \;.$$

For any fixed $\tau \in \Theta_n$ one can view $\mathbf{1}\{\tau \notin C_n\}$ as a test of the hypothesis $\theta_n = \tau$. Thus $\mathbb{P}\{\tau \notin C_n\}$ can not exceed the power of the Neyman-Pearson test of (τ, a, b) vs. (θ_n, μ_n, ν_n) with level α, which is given by

$$\Phi\Big(\|m(\theta_n, \mu_n, \nu_n) - m(\tau, a, b)\| + \Phi^{-1}(\alpha)\Big) \;\leq\; \Phi\Big(\|m(\theta_n, \mu_n, \nu_n) - m(\tau, a, b)\|\Big) \;,$$

for any fixed $a, b \in \mathbf{R}$ (Φ is the cdf of $\mathcal{N}(0,1)$). But with

$$\Delta_n \;:=\; \sqrt{n}(\mu_n - \nu_n)$$

and $k(s,t) := s \wedge t - st$, $k(t) := k(t,t)$ for $0 \leq s, t \leq 1$ one can show that the minimum of $\|m(\theta_n, \mu_n, \nu_n) - m(\tau, a, b)\|^2$ over all $a, b \in \mathbf{R}$ equals

$$|\tau - \theta_n| k(\tau)^{-1} k(\tau, \theta_n)\, \Delta_n^2 \;\leq\; \begin{cases} k(\theta_n)\, \Delta_n^2 \;, \\ |\tau - \theta_n|\, \Delta_n^2 \;. \end{cases}$$

Therefore a necessary condition for $\mathrm{dist}(C_n, \theta_n)$ to tend to 0 in probability is given by

$$k(\theta_n)\, \Delta_n^2 \;\to\; \infty \;.$$

Furthermore the best possible result (in terms of rates of convergence) one can expect is that

$$\mathrm{dist}(C_n, \theta_n) \;=\; O_p(\Delta_n^{-2}) \;.$$

Note that this lower bound for the size of C_n does not depend on θ_n. Another interesting conclusion for $\#C_n$, the cardinality of C_n, is that

$$\mathbb{E}(\#C_n) \;\geq\; (n-1)\Phi(-k(\theta_n)^{1/2}|\Delta_n|) \;.$$

Therefore, if $\Delta_n^2 \to \infty$, a necessary condition for $\mathbb{E}(\#C_n)$ to be of order $O(n\Delta_n^{-2})$ is given by

$$k(\theta_n)\,\Delta_n^2/\log(1/k(\theta_n)) \;\geq\; 2 \;+\; o(1)\;.$$

This follows from the known asymptotic expansion $\Phi(-x) = \exp(-x^2/2)/x(1+o(1))$ as $x \to \infty$.

Now we derive an explicit version of $\hat{C}_n$. With $S_n(t) := \sum_{1\leq i\leq nt} X_n(i)$, the statistic $S_n^{(\tau)} := (S_n(\tau), S_n(1) - S_n(\tau))$ is sufficient and complete for the restricted model, where $\theta_n = \tau$. Therefore any confidence set C_n with exact level α satisfies the condition

$$\int 1\{\tau \in C_n(x)\}\mathbb{P}_n^{(\tau)}(dx|s) \;=\; 1-\alpha \quad \text{for Lebesgue} - \text{almost all } s \in \mathbf{R}^2\;.$$

We want to minimize the Bayes-risk

$$R^{(\tau)}(C_n) \;:=\; \int 1\{\tau \in C_n(x)\}\,M^{(\tau)}(dx)$$

among all confidence sets with exact level α, where

$$M^{(\tau)} \;:=\; \int \mathcal{N}\Big(m(t, a+(1-t)b, a-tb), I\Big)\,1\{t \neq \tau\}\,\mathcal{U}_n(dt)\,H(da)\,db$$

for some finite measure H on the line and $\mathcal{U}_n := n^{-1}\sum_{t\in\Theta_n}\delta_t$. In other words, H is a prior for the mean $\theta_n\mu_n + (1-\theta_n)\nu_n$ of $n^{-1}S_n(1)$ (which provides no information about θ_n), $\mathcal{U}_n$ (restricted to $\Theta_n \setminus \{\tau\}$) is a prior for θ_n, and Lebesgue measure is a noninformative prior for $\mu_n - \nu_n$. This Bayes-risk is finite, which is not obvious but can be shown quite easily. The density f of $M^{(\tau)}$ with respect to $\mathcal{N}(0, I)$ exists and has the form

$$f(X_n) \;=\; T_n'\,g(S_n^{(\tau)})$$

for some function $g > 0$, where

$$T_n' \;:=\; \int k(t)^{-1/2}\exp(W_n(t))\,1\{t \neq \tau\}\,\mathcal{U}_n(dt)\;,$$

$$\begin{aligned} T_n &:= \int k(t)^{-1/2} \exp(W_n(t)) \mathcal{U}_n(dt) , \\ W_n(t) &:= k(t)^{-1} D_n(t)^2/2 , \\ D_n(t) &:= \sqrt{n}^{-1}(S_n(t) - tS_n(1)) . \end{aligned}$$

Furthermore, since

$$\mathcal{N}(0,I) = \int \mathbb{P}_n^{(\tau)}(\cdot|s) \, \mathcal{N}\left(0, \begin{pmatrix} \tau & 0 \\ 0 & 1-\tau \end{pmatrix}\right)(ds) ,$$

the Bayes-risk $R^{(\tau)}(C_n)$ can be written as

$$\int \int T_n'(x) 1\{\tau \in C_n(x)\} \, \mathbb{P}_n^{(\tau)}(dx|s) \, g(s) \, \mathcal{N}\left(0, \begin{pmatrix} \tau & 0 \\ 0 & 1-\tau \end{pmatrix}\right)(ds) .$$

Therefore the confidence set $\hat{C}_n$, which is defined as in section 1 with the particular test statistic T_n above is Bayes-optimal among all confidence sets with exact level α (note that T_n' and T_n differ by a function of $S_n^{(\tau)}$ only).

In the above derivation one could certainly replace $\mathcal{U}_n$ with any other finite prior for θ_n. From now on we consider the test statistic

$$T_n := \int k(t)^{\beta} \exp(W_n(t)) \mathcal{U}_n(dt) ,$$

where β is any fixed number in $[-1, \infty)$. The resulting confidence sets $\hat{C}_n$ have the following asymptotic properties:

Theorem 1a: *Suppose that*

$$\begin{aligned} k(\theta_n) \, \Delta_n^2 / \log\log n &\to \infty \quad \text{if } \beta = -1 , \\ k(\theta_n) \, \Delta_n^2 / \log(1/k(\theta_n)) &\to \infty \quad \text{if } \beta > -1 . \end{aligned}$$

Then

$$\operatorname{dist}(\hat{C}_n, \theta_n) = O_p(\Delta_n^{-2}) .$$

There are two interesting special cases: Suppose first that $\mu_n - \nu_n = \text{const} \neq 0$. Then $\mathrm{dist}(\hat{C}_n, \theta_n)$ is of order $O_p(n^{-1})$, provided that

$$((n\theta_n) \wedge (n - n\theta_n))/\log\log n \rightarrow \infty \quad \text{if } \beta = -1 ,$$
$$((n\theta_n) \wedge (n - n\theta_n))/\log n \rightarrow \infty \quad \text{if } \beta > -1 .$$

If $\theta_n \rightarrow \theta \in (0,1)$, then $\mathrm{dist}(\hat{C}_n, \theta_n)$ is of order $O_p(\Delta_n^{-2})$, provided that

$$\Delta_n^2/\log\log n \rightarrow \infty \quad \text{if } \beta = -1 ,$$
$$\Delta_n^2 \rightarrow \infty \quad \text{if } \beta > -1 .$$

The limiting behavior of $\hat{C}_n$ can be described as follows: Let $\hat{p}_n(r) := 0$ for $r \in [-\infty, 0] \cup [1, \infty]$ and $\hat{p}_n(r) := \hat{p}_n([nr]/n)$ for $r \in [0,1]$ (the same type of extension is used for any other process on Θ_n). Further let $(Z(r))_{r \in \mathbf{R}}$ be a two-sided Brownian motion on the line; i.e. $(Z(r))_{r \geq 0}$ and $(Z(-r))_{r \geq 0}$ are two independent Brownian motions.

Theorem 2a: *Suppose that the assumptions of Theorem 1a hold with $\mu_n - \nu_n \rightarrow 0$. Then the process*

$$(\hat{p}_n(\theta_n + \Delta_n^{-2} r))_{r \in [-\infty, \infty]}$$

converges in distribution in $D[-\infty, \infty]$ to the process

$$(\hat{p}(r))_{r \in [-\infty, \infty]} ,$$

where $\hat{p}(-\infty) := \hat{p}(\infty) := 0$, and

$$\begin{aligned} \hat{p}(r) &:= H\Big(\exp(-W(r)) \int \exp(W(t))\, dt\Big) , \\ H(r) &:= \mathbb{P}\Big\{\int \exp(W(t))\, dt \geq r\Big\} , \\ W(r) &:= Z(r) - |r|/2 \quad \textit{for } r \in \mathbf{R} . \end{aligned}$$

An explicit formula for H is given by Siegmund (1988). For our purposes one only needs to know that H is continuous.

If $\mu_n - \nu_n \to \delta \neq 0$ one can obtain a similar result for the process $(\hat{p}_n(\theta_n + j/n))_{j=0,\pm1,\pm2,\ldots}$. Here the corresponding limit process $(\hat{p}^*(j))_{j=0,\pm1,\pm2,\ldots}$ has the form

$$\begin{aligned}
\hat{p}^*(j) &:= H^*\Big(\exp(-W^*(j)) \sum_{-\infty<i<\infty} \exp(W^*(i))\Big) , \\
H^*(r) &:= \mathbb{P}\Big\{ \sum_{-\infty<i<\infty} \exp(W^*(i)) \geq r \Big\} , \\
W^*(j) &:= \delta Z(j) - \delta^2 |j|/2 .
\end{aligned}$$

Hence $\hat{C}_n$ behaves similarly as the optimal shift equivariant confidence set C_5 in Siegmund (1988).

3. Nonparametric confidence sets

Here we make no parametric assumptions on P_n and Q_n. Similarly as in section 2 define

$$S_n(t) := \sum_{1\leq i\leq nt} \delta_{X_n(i)} .$$

Again the statistic $S_n^{(\tau)} := (S_n(\tau), S_n(1) - S_n(\tau))$ is known to be sufficient for the restricted model when $\theta_n = \tau$ and P_n, Q_n are arbitrary. An explicit version of $\mathbb{P}_n^{(\tau)}(\cdot|S_n^{(\tau)})$ can be described as follows: For $\tau \in \Theta_n$ let $\Pi_n^{(\tau)}$ be uniformly distributed on the set of all permutations π of $\{1, 2, \ldots, n\}$ such that $\pi(i) \leq n\tau$ for all $i \leq n\tau$, and let $\Pi_n^{(\tau)}$ and X_n be independent. Then

$$\mathcal{L}(X_n|S_n^{(\tau)}, \theta_n = \tau) = \mathcal{L}(\Pi_n^{(\tau)} X_n|X_n) ,$$

where $\Pi_n^{(\tau)} X_n := \left(X_n(\Pi_n^{(\tau)}(1)), \ldots, X_n(\Pi_n^{(\tau)}(n))\right)$.

As for the choice of T_n, let $\|\cdot\|_n$ be a seminorm on the space of finite signed measures on $\mathbf{X}$, which can be a function of the random measure $S_n(1)$. Then we define

$$\begin{aligned} T_n &:= \int k(t)^\beta \exp(W_n(t)) \mathcal{U}_n(dt) , \\ W_n(t) &:= k(t)^{-1} \|D_n(t)\|_n^2/2 , \\ D_n(t) &:= \sqrt{n}^{-1}(S_n(t) - tS_n(1)) \end{aligned}$$

for some fixed $\beta \geq -1$.

An essential technical requirement is that $\|\cdot\|_n$ is bounded by a Kolmogorov-Smirnov type norm $\|\cdot\|_{\mathcal{F}}$:

$$\|\cdot\|_n \leq \|\cdot\|_{\mathcal{F}} \quad \text{almost surely} .$$

More precisely, $\|m\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |m(f)|$, where $\mathcal{F}$ is a countable family of measurable functions $f : \mathbf{X} \to [0, c]$, $0 < c < \infty$, and there are constants $A, B > 0$ such that the covering numbers

$$\min\left\{k \in \{1, 2, \ldots\} : \exists f_1, \ldots, f_k \text{ with } \min_{1 \leq i \leq k} P((f - f_i)^2) \leq u^2 \ \forall f \in \mathcal{F}\right\}$$

are bounded by Au^{-B} for all $u \in (0, 1]$ and arbitrary probability measures P on $\mathbf{X}$. Different examples for $\|\cdot\|_n$ and $\|\cdot\|_{\mathcal{F}}$ can be found in Dümbgen (1991). When $\mathbf{X} = \mathbf{R}$ one might take

$$\|m\|_{nD} := \left|\sqrt{3} \int m(x)\, n^{-1} S_n(1)(dx)\right|$$

(proposed by Darkhovskiy, 1976), where $m(x) := m(-\infty, x) - m(x, \infty)$ is a symmetrized cdf of m. This seminorm $\|\cdot\|_{nD}$ is bounded by $\sqrt{12}$ times the usual Kolmogorov-Smirnov norm on the line.

The nonparametric confidence sets $\hat{C}_n$ have similar asymptotic properties as the parametric ones of section 2. With

$$\Delta_n \;:=\; \sqrt{n}(P_n - Q_n)$$

the following result holds:

Theorem 1b: *Suppose that*

$$k(\theta_n)\, \|\Delta_n\|_n^2 / \log\log n \;\to_p\; \infty \quad \textit{if } \beta = -1\;,$$
$$k(\theta_n)\, \|\Delta_n\|_n^2 / \log(1/k(\theta_n)) \;\to_p\; \infty \quad \textit{if } \beta > -1\;.$$

Then

$$\mathrm{dist}(\hat{C}_n, \theta_n) \;=\; O_p(\|\Delta_n\|_n^{-2})\;.$$

Note that $\|\Delta_n\|_n$ is random in general. But it can often be approximated by a nonrandom number. For instance it follows from Tshebyshev's inequality that

$$\sqrt{3}\int \Delta_n(x)\, n^{-1} S_n(1)(dx) \;=\; \delta_n + O_p(1)\;, \tag{1}$$

where

$$\delta_n \;:=\; \sqrt{3}\int \Delta_n(x) P_n(dx)\;.$$

Here is a result about the limiting distribution of $\hat{C}_n$ for the particular seminorm $\|\cdot\|_{nD}$. The proof in section 4 could be extended to other seminorms; see also Dümbgen (1991).

Theorem 2b: *Suppose that P_n, Q_n converge weakly to a common continuous distribution P on the real line and*

$$k(\theta_n)\, \delta_n^2 / \log\log n \;\to\; \infty \quad \textit{if } \beta = -1\;,$$

$$k(\theta_n)\,\delta_n^2/\log(1/k(\theta_n)) \;\to\; \infty \quad \textit{if } \beta > -1\ .$$

Then the process $(\hat{p}_n(\theta_n + \delta_n^{-2}r))_{r\in[-\infty,\infty]}$ *converges in distribution in* $D[-\infty,\infty]$ *to the process* $(\hat{p}(r))_{r\in[-\infty,\infty]}$ *defined in Theorem 2a.*

In particular suppose that P_n, Q_n are normal distributions as in section 2, where $\mu_n - \nu_n$ tends to zero. Then $n^{-1}(\mu_n - \nu_n)^{-2}\delta_n^2 \to 3/\pi \approx 0.955$. Consequently the nonparametric p-values $\hat{p}_n(\theta_n + r)$ behave asymptotically as the parametric p-values $\hat{p}_n(\theta_n + 3r/\pi)$.

4. Proofs

One can prove the preceding results in a common framework: The quantities μ_n, P_n, ν_n, Q_n, Δ_n as well as the random variables $S_n(t)$, $D_n(t)$ are viewed as points in a normed linear space $(\mathbf{B}, \|\cdot\|)$. In the normal shift model $\mathbf{B} = \mathbf{R}$ and $\|\cdot\|_n := \|\cdot\| := |\cdot|$, whereas in the nonparametric model $\mathbf{B}$ is the space of bounded functions on $\mathcal{F}$ and $\|\cdot\| := \|\cdot\|_{\mathcal{F}}$. In order to distinguish between the cases $\beta = -1$ and $\beta > -1$ we use superscripts $(\cdot)^{(=)}$ and $(\cdot)^{(>)}$ respectively for T_n and other related quantities.

4.1. Auxiliary results, I

In this part we regard $\|\cdot\|_n$ and $D_n(\theta_n)$ as fixed and write

$$D_n(t) \;=\; k(t,\theta_n)Y_n \;+\; Z_n(t) \quad , \quad Y_n \;:=\; k(\theta_n)^{-1}D_n(\theta_n)\ ,$$

so that $Z_n(\theta_n) = 0$. The following quantities play a crucial role:

$$L_n \;\;:=\;\; \max_{t\in\Theta_n} \Big(k(t)\log\log(1/k(t))\Big)^{-1}\|Z_n(t)\|^2\ ,$$

$$\begin{aligned} M_n(\sigma) &:= \sigma^{-1/2} \max_{t\in\Theta_n:|t-\theta_n|\le\sigma} \|Z_n(t)\| \quad \text{and} \\ N_n(\sigma) &:= \sigma^{1/2} \max_{t\in\Theta_n:|t-\theta_n|\ge\sigma} |t-\theta_n|^{-1}\|Z_n(t)\| \quad \text{for} \quad \sigma > 0 \ . \end{aligned}$$

Here is a crude but useful bound:

Proposition 1: *If* $L_n = O_p(1)$, *then*

$$T_n^{(=)} = O_p\Big((\log n)^{L_n+1} \exp(2W_n(\theta_n))\Big)$$

and

$$T_n^{(>)} = O_p\Big(k(\theta_n)^{(\beta+1)/4} \exp(2W_n(\theta_n)) + \exp(4k(\theta_n)^{1/2}W_n(\theta_n))\Big) \ .$$

Proof of Proposition 1: One can write

$$W_n(t) = \|\rho(t,\theta_n)Y_n + k(t)^{-1/2}Z_n(t)\|_n^2/2 \ ,$$

where

$$\rho(t,\theta) := k(t)^{-1/2}k(t,\theta) = \Big[(1-\theta)(t/(1-t))^{1/2}\Big] \wedge \Big[\theta((1-t)/t)^{1/2}\Big] \ .$$

The function $\rho(\cdot,\theta)$ is strictly increasing on $(0,\theta]$ and strictly decreasing on $[\theta,1)$ with $\rho(\theta,\theta) = k(\theta)^{1/2}$. By the triangle inequality,

$$W_n(t) \le \rho(t,\theta_n)^2\|Y_n\|_n^2 + \log\log(1/k(t))L_n \quad \forall\, t \in \Theta_n \ . \tag{2}$$

In particular,

$$W_n(t) \le 2W_n(\theta_n) + \log\log(1/k(t))L_n \le 2W_n(\theta_n) + \log\log(2n)L_n \ ,$$

and thus

$$\begin{aligned} T_n^{(=)} &\le \int k(t)^{-1}\,\mathcal{U}_n(dt)\,(\log(2n))^{L_n} \exp(2W_n(\theta_n)) \\ &= O(\log n)(\log(2n))^{L_n} \exp(2W_n(\theta_n)) \ . \end{aligned}$$

On the other hand one can easily show that $\rho(t,\theta_n)^2 \leq 2k(\theta_n)^{3/2}$, if $\theta_n \leq 1/2$ and $t \geq k(\theta_n)^{1/2}$, or, if $\theta_n \geq 1/2$ and $t \leq 1 - k(\theta_n)^{1/2}$. Consequently,

$$\begin{aligned}
T_n^{(>)} &\leq \int k(t)^\beta \log(1/k(t))^{L_n} \exp(\rho(t,\theta_n)^2 \|Y_n\|_n^2)\mathcal{U}_n(dt) \\
&\leq O_p(1) \int k(t)^{(\beta-1)/2} \exp(\rho(t,\theta_n)^2\|Y_n\|_n^2)\mathcal{U}_n(dt) \\
&\leq O_p(1) \int_{\Theta_n \cap [0,k(\theta_n)^{1/2}]} k(t)^{(\beta-1)/2}\mathcal{U}_n(dt)\ \exp(2W_n(\theta_n)) \\
&+ O_p(1) \int k(t)^{(\beta-1)/2}\mathcal{U}_n(dt)\ \exp(2k(\theta_n)^{3/2}\|Y_n\|_n^2) \\
&= O_p(k(\theta_n)^{(\beta+1)/4})\exp(2W_n(\theta_n)) + O_p(1)\exp(4k(\theta_n)^{1/2}W_n(\theta_n)) \quad \Box
\end{aligned}$$

The bounds in Proposition 1 are useful for small values of $k(\theta_n)$ and moderate values of $W_n(\theta_n)$. However, if $W_n(\theta_n)$ is sufficiently large, one can approximate $W_n(t) - W_n(\theta_n)$ by

$$\tilde{W}_n(t) := \|Y_n\|_n \|k(\theta_n)Y_n - 2^{-1}|t-\theta_n|Y_n + Z_n(t)\|_n - k(\theta_n)\|Y_n\|_n^2$$

and T_n by $k(\theta_n)^{\beta+1}W_n(\theta_n)^{-1}\exp(W_n(\theta_n))$ times

$$\tilde{T}_n := \|Y_n\|_n^2 \int_{\Theta_n^{(o)}} \exp(\tilde{W}_n(t))\mathcal{U}_n(dt)/2 ,$$

where $\Theta_n^{(o)} := \{t \in \Theta_n : |t-\theta_n| \leq 2k(\theta_n)\}$:

Proposition 2: *Suppose that $L_n = O_p(1)$, $M_n(\sigma_n) = O_p(1)$ and $N_n(\sigma_n) = O_p(1)$ for any fixed sequence of numbers $\sigma_n > 0$. Further suppose that*

$$\begin{aligned}
k(\theta_n)\|Y_n\|_n^2/\log\log n &\to \infty \quad \text{if } \beta = -1 , \\
k(\theta_n)\|Y_n\|_n^2/\log(1/k(\theta_n)) &\to \infty \quad \text{if } \beta > -1 .
\end{aligned}$$

If $n^{-1}\|Y_n\|_n^2 \to \infty$, then

$$T_n = n^{-1}k(\theta_n)^\beta \exp(W_n(\theta_n))(1+o_p(1)) .$$

On the other hand, if $n^{-1}\|Y_n\|_n^2 = O(1)$, *then*

$$\begin{aligned} \tilde{T}_n &= O_p(1) \ , \quad \tilde{T}_n^{-1} \ = \ O_p(1) \quad \text{and} \\ T_n &= k(\theta_n)^{\beta+1} W_n(\theta_n)^{-1} \exp(W_n(\theta_n)) \tilde{T}_n (1 + o_p(1)) \ . \end{aligned}$$

Proof of Proposition 2: At first some useful inequalities are listed that can be proved with elementary calculations: For arbitrary $t, \theta \in (0,1)$,

$$\begin{aligned} |k(t) - k(\theta)| &\leq |t - \theta| \ , \qquad (3) \\ \rho(\theta,\theta) - \rho(t,\theta) &\leq k(\theta)^{-1/2}|t-\theta| \quad \text{and} \\ |\rho(\theta,\theta) - \rho(t,\theta) - k(\theta)^{-1/2}|t-\theta|/2| &\leq k(\theta)^{-3/2}|t-\theta|^2 \ . \end{aligned}$$

Further, let

$$\eta \ := \ (t \vee \theta(1 - t \wedge \theta))/(t \wedge \theta(1 - t \vee \theta)) \ \geq \ 1 \ .$$

Then,

$$\begin{aligned} \rho(t,\theta)^2 \ = \ \eta^{-1}k(\theta) \ , \quad 1 - \eta^{-1} \ \leq \ k(\theta)^{-1}|t-\theta| \ \leq \ \eta - 1 \ , \qquad (4) \\ \eta^{-1} \ \leq \ k(t)^{-1}k(t,\theta) \ \leq \ 1 \quad \text{and} \quad \eta^{-1} \ \leq \ k(t)^{-1}k(\theta) \ \leq \ \eta \ . \end{aligned}$$

Now let $\lambda > 1$ and $\gamma, \gamma_n > 0$ be arbitrary fixed numbers such that $\gamma_n \to \infty$. The set Θ_n is split into the two subsets $\Theta_n(\lambda)$ and $\Theta_n \setminus \Theta_n(\lambda)$, where $\Theta_n(\lambda)$ is the set of all $t \in \Theta_n$ such that

$$\lambda^{-1} \ \leq \ (t(1-\theta_n))/(\theta_n(1-t)) \ \leq \ \lambda \ .$$

Then (2) and (4) imply that

$$\begin{aligned} \int_{\Theta_n \setminus \Theta_n(\lambda)} k(t)^{-1} \exp(W_n(t)) \mathcal{U}_n(dt) \ &\leq \ O_p\Big((\log n)^{L_n+1} \exp(2\lambda^{-1}W_n(\theta_n))\Big) \qquad (5) \\ \int_{\Theta_n \setminus \Theta_n(\lambda)} k(t)^{\beta} \exp(W_n(t)) \mathcal{U}_n(dt) \ &= \ O_p\Big(\exp(2\lambda^{-1}W_n(\theta_n))\Big) \quad \text{for } \beta > -1 \ . \end{aligned}$$

Furthermore

$$\max_{t\in\Theta_n(\lambda)} \|D_n(t) - k(t,\theta_n)Y_n\| \;=\; O_p(k(\theta_n)^{1/2}) \,. \tag{6}$$

For $D_n(t) - k(t,\theta_n)Y_n$ equals $Z_n(t)$, and $\|Z_n(t)\|$ is not greater than

$$(\lambda-1)^{1/2}k(\theta_n)^{1/2}M_n((\lambda-1)k(\theta_n))$$

for all $t \in \Theta_n(\lambda)$, by (4).

Now the set $\Theta_n(\lambda)$ itself is split into two subsets $\Theta_n(\lambda,\bar\gamma)$ and $\Theta_n(\lambda)\setminus\Theta_n(\lambda,\bar\gamma)$, where $\Theta_n(\lambda,\bar\gamma)$ is the set of all $t\in\Theta_n(\lambda)$ with $|t-\theta_n| \le \bar\gamma\|Y_n\|_n^{-2}$. On the one hand,

$$W_n(t) - W_n(\theta_n) \;\le\; -|t-\theta_n|\lambda^{-1}(1+o_p(1))\|Y_n\|_n^2/2 \quad \forall t\in\Theta_n(\lambda)\setminus\Theta_n(\lambda,\gamma_n)\,. \tag{7}$$

For $W_n(t) - W_n(\theta_n)$ is not greater than

$$\begin{aligned}
&(\rho(t,\theta_n)\|Y_n\|_n + k(t)^{-1/2}\|Z_n(t)\|_n)^2/2 \;-\; k(\theta_n)\|Y_n\|_n^2/2\\
&\le\; (\rho(t,\theta_n) + k(t)^{-1/2}|t-\theta_n|\epsilon_n)^2\|Y_n\|_n^2/2 \;-\; k(\theta_n)\|Y_n\|_n^2/2\\
&=\; -|t-\theta_n|(k(t)^{-1}k(t,\theta_n)(1-2\epsilon_n) - k(t)^{-1}|t-\theta_n|\epsilon_n^2)\|Y_n\|_n^2/2\\
&\le\; -|t-\theta_n|\lambda^{-1}(1-2\epsilon_n-\lambda^2\epsilon_n^2)\|Y_n\|_n^2/2
\end{aligned}$$

for all $t\in\Theta_n(\lambda)\setminus\Theta_n(\lambda,\gamma_n)$, provided that $\epsilon_n := \gamma_n^{-1/2}N_n(\gamma_n\|Y_n\|_n^{-2}) < 1/2$; the last displayed inequality is a consequence of (4). In particular, if $\kappa_n := \lambda^{-1}(1-2\epsilon_n-\lambda^2\epsilon_n^2)/2 > 0$, then

$$\begin{aligned}
&\int_{\Theta_n(\lambda)\setminus\Theta_n(\lambda,\gamma_n)} k(t)^\beta \exp(W_n(t))\,\mathcal{U}_n(dt)\\
&\le\; \lambda^\beta k(\theta_n)^\beta \exp(W_n(\theta_n)) \int_{\Theta_n(\lambda)\setminus\Theta_n(\lambda,\gamma_n)} \exp(-\kappa_n|t-\theta_n|\|Y_n\|_n^2)\,\mathcal{U}_n(dt)\\
&\le\; 2\lambda^\beta k(\theta_n)^\beta \exp(W_n(\theta_n))\,\exp(-\gamma_n\kappa_n)\,n^{-1}(1-\exp(-\kappa_n n^{-1}\|Y_n\|_n^2))^{-1}\,.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\int_{\Theta_n(\lambda)\setminus\Theta_n(\lambda,\gamma_n)} k(t)^\beta \exp(W_n(t))\,\mathcal{U}_n(dt) \\
&=\; o_p\Big(k(\theta_n)^\beta(\|Y_n\|_n^2\wedge n)^{-1}\exp(W_n(\theta_n))\Big)\,.
\end{aligned} \tag{8}$$

For the moment suppose that $n^{-1}\|Y_n\|_n^2$ tends to infinity. T_n is obviously not smaller than

$$n^{-1}k(\theta_n)^\beta \exp(W_n(\theta_n)) .$$

The bounds in (5) are of smaller order than that, provided that $\lambda > 2$. Together with (8), where $\gamma_n := n^{-1}\|Y_n\|_n^2/2$, one can deduce that T_n is not greater than $n^{-1}k(\theta_n)^\beta \exp(W_n(\theta_n))(1 + o_p(1))$. This establishes the first part of Proposition 2, and for the rest of this proof we assume that $n^{-1}\|Y_n\|_n^2$ is bounded.

As for the approximation $\tilde{T}_n$, note first that

$$\Theta_n(3) \subset \Theta_n^{(o)} \quad \text{and} \quad \Theta_n(\infty,\tilde{\gamma}) \subset \Theta_n\Big((1-\tilde{\gamma}W_n(\theta_n)^{-1}/2)^{-1}\Big)$$

whenever $0 < \tilde{\gamma} \le 2W(\theta_n)$. This is a direct consequence of (4). In particular, $\Theta_n(\infty,\gamma) \subset \Theta_n^{(o)}$ for sufficiently large n. Now one can show that

$$\begin{aligned} \|Y_n\|_n^2 \int_{\Theta_n^{(o)}\setminus\Theta_n(\infty,\gamma_n)} \exp(\tilde{W}_n(t))\mathcal{U}_n(dt) &= o_p(1) \quad \text{and} \qquad (9) \\ \left(\|Y_n\|_n^2 \int_{\Theta_n^{(o)}\cap\Theta_n(\infty,\gamma)} \exp(\tilde{W}_n(t))\mathcal{U}_n(dt)\right)^{\pm 1} &= O_p(1) . \end{aligned}$$

For $|\tilde{W}_n(t) + |t-\theta_n|\|Y_n\|_n^2/2|$ is not greater than $\|Y_n\|_n\|Z_n(t)\|_n$ for all $t \in \Theta_n^{(o)}$, and

$$\|Y_n\|_n\|Z_n(t)\|_n \le \begin{cases} \gamma^{1/2}M_n(\gamma\|Y_n\|_n^{-2}) , & \text{if } t \in \Theta_n(\infty,\gamma) , \\ \gamma_n^{-1/2}N_n(\gamma_n\|Y_n\|_n^{-2})|t-\theta_n|\|Y_n\|_n^2 , & \text{if } t \notin \Theta_n(\infty,\gamma_n) . \end{cases}$$

In particular, (9) implies the boundedness of $\tilde{T}_n$ and $\tilde{T}_n^{-1}$.

Finally,

$$\max_{t\in\Theta_n(\lambda,\gamma_n)} |W_n(t) - W_n(\theta_n) - \tilde{W}_n(t)| = o_p(1) \quad \text{if } \gamma_n^2 W_n(\theta_n)^{-1} \to 0 . \qquad (10)$$

For $W_n(t) - W_n(\theta_n) - \tilde{W}_n(t)$ can be written as

$$\begin{aligned} & k(\theta_n)^{1/2}\|Y_n\|_n \Big(\|\rho(t,\theta_n)Y_n + k(t)^{-1/2}Z_n(t)\|_n - \\ & \quad \|k(\theta_n)^{1/2}Y_n - 2^{-1}k(\theta_n)^{-1/2}|t-\theta_n|Y_n + k(\theta_n)^{-1/2}Z_n(t)\|_n\Big) \\ & + \Big(\|\rho(t,\theta_n)Y_n + k(t)^{-1/2}Z_n(t)\|_n - k(\theta_n)^{1/2}\|Y_n\|_n\Big)^2/2 , \end{aligned}$$

and thus its absolute value is not greater than

$$\begin{aligned}
& k(\theta_n)^{1/2}\|Y_n\|_n^2 \, |\rho(\theta_n,\theta_n) - \rho(t,\theta_n) - k(\theta_n)^{-1/2}|t-\theta_n|/2| \\
+ & \ \|Y_n\|_n \, |k(t)^{-1/2}k(\theta_n)^{1/2} - 1| \, \|Z_n(t)\|_n \\
+ & \ (\rho(\theta_n,\theta_n) - \rho(t,\theta_n))^2\|Y_n\|_n^2 \ + \ k(t)^{-1}\|Z_n(t)\|_n^2 \\
\leq & \ (1 + \lambda^{1/2}\epsilon_n + 1 + \lambda\epsilon_n^2)\gamma_n^2 k(\theta_n)^{-1}\|Y_n\|_n^{-2} \ ,
\end{aligned}$$

where $\epsilon_n := \gamma_n^{-1/2}M_n(\gamma_n\|Y_n\|_n^{-2})$; see (3) and (4). One can use (10) for showing that

$$\begin{aligned}
& \int_{\Theta_n(\lambda,\gamma_n)} k(t)^\beta \exp(W_n(t))\,\mathcal{U}_n(dt) \qquad (11) \\
= & \ \ k(\theta_n)^{\beta+1}W_n(\theta_n)^{-1}\exp(W_n(\theta_n))\,\bar{T}_n\,(1+o_p(1)) \quad \text{if } \gamma_n^2 W_n(\theta_n)^{-1} \to 0 \ .
\end{aligned}$$

For $\Theta_n(\infty,\gamma_n)$ is a subset of $\Theta_n(\lambda_n)$, where $\lambda_n := (1 - \gamma_n W_n(\theta_n)^{-1}/2)^{-1} \to 1$; in particular, $\Theta_n(\lambda,\gamma_n) = \Theta_n(\infty,\gamma_n) \subset \Theta_n^{(o)}$ for sufficiently large n. Thus

$$k(t)^\beta \exp(W_n(t) - W_n(\theta_n)) \ = \ k(\theta_n)^\beta \exp(\tilde{W}_n(t))(1 + r_n(t)) \ ,$$

where $\max_{t\in\Theta_n(\lambda,\gamma_n)} |r_n(t)| = o_p(1)$, by (4) and (10). Finally, $k(\theta_n)^\beta \exp(\tilde{W}_n(t))$ can be written as $k(\theta_n)^{\beta+1}W_n(\theta_n)^{-1}\|Y_n\|_n^2 \exp(\tilde{W}_n(t))/2$, and one can deduce (11) from (9).

The inequalities (5), (8) and (11) with $\lambda > 2$ yield the last assertion in Proposition 2 □

Here is a result that can be used to verify the assumptions about L_n, M_n and N_n in Propositions 1 and 2:

Lemma 1: *Let $(V(t))_{t\in\Theta_n}$ be a* B*-valued stochastic process such that*

$$\mathbb{P}\Big\{\max_{t\in\Theta_n : t\leq\sigma} \|V(t)\| \geq \eta\Big\} \ \leq \ K\exp(-L\sigma^{-1}\eta^2) \quad \forall\, \sigma,\eta > 0 \ ,$$

where $K \geq 1$, $L > 0$. Then

$$\mathbb{P}\Big\{\max_{t\in\Theta_n: t\leq \exp(-2)} (t\log\log(1/t))^{-1/2}\|V(t)\| \geq \eta\Big\} \;\leq\; CK\exp(-L\eta^2/C) \quad \text{and}$$

$$\mathbb{P}\Big\{\sigma^{1/2}\max_{t\in\Theta_n: t\geq\sigma} t^{-1}\|V(t)\| \geq \eta\Big\} \;\leq\; CK\exp(-L\eta^2/C)$$

for all $\sigma, \eta > 0$, where $C > 0$ is a universal constant.

Proof of Lemma 1: The function $h(t) := (t\log\log(1/t))^{1/2}$ is nondecreasing on $(0, \exp(-2)]$. Therefore $\mathbb{P}\Big\{\max_{t\in\Theta_n: t\leq\exp(-2)} h(t)^{-1}\|V(t)\| \geq \eta\Big\}$ is not greater than

$$\begin{aligned}
&\sum_{0\leq i<(\log n-2)/\log 2} \mathbb{P}\Big\{\max_{t\in\Theta_n: 2^i\leq nt\leq 2^{i+1}} \|V(t)\| \geq h(2^i/n)\eta\Big\} \\
&\leq\; K \sum_{0\leq i<(\log n-2)/\log 2} \exp(-L\eta^2 h(2^i/n)^2 n 2^{-i-1}) \\
&=\; K \sum_{0\leq i<(\log n-2)/\log 2} (\log n - i\log 2)^{-L\eta^2/2} \\
&\leq\; K(\log 2)^{-1}\int_{2-\log 2}^{\log n} x^{-L\eta^2/2}\,dx \\
&\leq\; K(2/\log 2 - 1)(L\eta^2/2 - 1)^{-1}(2-\log 2)^{-L\eta^2/2}\,,
\end{aligned}$$

provided that $L\eta^2/2 > 1$. This yields the first assertion. As for the second part,

$$\begin{aligned}
&\mathbb{P}\Big\{\max_{t\in\Theta_n: t\geq\sigma} t^{-1}\|V(t)\| \geq \sigma^{-1/2}\eta\Big\} \\
&\leq\; \sum_{i=1}^{\infty}\mathbb{P}\Big\{\max_{t\in\Theta_n: i\sigma\leq t\leq(i+1)\sigma} \|V(t)\| \geq i\sigma^{1/2}\eta\Big\} \\
&\leq\; K\sum_{i=1}^{\infty}\exp(-Li^2\eta^2/(i+1)) \\
&\leq\; K/(\exp(L\eta^2/2) - 1) \quad \Box
\end{aligned}$$

4.2. Auxiliary results, II

The p-values $\hat{p}_n$ can be represented as follows: For each $\tau \in \Theta_n$ let $D_n^{(\tau)} =$

$(D_n^{(\tau)}(t))_{t\in\Theta_n}$ be a stochastic process defined on the same probability space as X_n such that

$$D_n^{(\tau)}(\tau) \;=\; D_n(\tau) \quad \text{and} \quad \mathcal{L}(D_n^{(\tau)}|X_n) \;=\; \mathbb{P}_n^{(\tau)}(D_n|S_n^{(\tau)})\,. \tag{12}$$

For any statistic $G_n = G_n(D_n, \theta_n)$ let $G_n^{(\tau)} := G_n(D_n^{(\tau)}, \tau)$. Then

$$\hat{p}_n(\tau) \;=\; \mathbb{P}(T_n^{(\tau)} \geq T_n|X_n)\,.$$

Explicitly, in the normal shift model let B be a Brownian bridge, which is independent from X_n. Then $D_n =_{\mathcal{L}} (k(t,\theta_n)\Delta_n + B(t))_{t\in\Theta_n}$, and one may define

$$D_n^{(\tau)}(t) \;:=\; k(\tau)^{-1}k(t,\tau)D_n(\tau) + Z^{(\tau)}(t) \;=\; k(t,\tau)Y_n^{(\tau)} + Z^{(\tau)}(t)\,,$$

where $Z^{(\tau)}(t) := B(t) - k(\tau)^{-1}k(t,\tau)B(\tau)$. The validity of (12) follows essentially from the fact that $B(\tau)$ and $Z^{(\tau)}$ are independent.

In the nonparametric model let $D_n^{(\tau)}$ be defined as D_n with $\Pi_n^{(\tau)}X_n$ in place of X_n.

The following two results are essential in the proof of Theorems 1a-b and 2a-b:

$$\begin{aligned}&\left\|(n\theta_n)^{-1}S_n(\theta_n) - P_n\right\| \vee \left\|(n - n\theta_n)^{-1}(S_n(1) - S_n(\theta_n)) - Q_n\right\| \\ &= O_p(\sqrt{n}^{-1}k(\theta_n)^{-1/2})\end{aligned} \tag{13}$$

(in the normal shift model P_n and Q_n stand for μ_n and ν_n respectively). Moreover, there is a function $b : (0,\infty) \to [0,1]$ such that (for suitable versions of $\mathbb{P}(\cdot|X_n)$)

$$\begin{aligned}&\mathbb{P}(L_n^{(\tau)} \geq \eta|X_n) \vee \mathbb{P}(M_n^{(\tau)}(\sigma) \geq \eta|X_n) \vee \mathbb{P}(N_n^{(\tau)}(\sigma) \geq \eta|X_n) \\ &\leq\; b(\eta) \quad \forall \tau \in \Theta_n \quad \forall \sigma, \eta > 0 \quad \text{and } b(\eta) \to 0 \text{ as } \eta \to \infty.\end{aligned} \tag{14}$$

In the normal shift model, (13) is obvious, while (14) can be easily derived from Lemma 1. For it is well-known that the Brownian bridge B satisfies the assumptions of Lemma 1, and the process $Z^{(\tau)}$ can be represented as

$$\Big(Z^{(\tau)}(t\tau)\Big)_{t\in[0,1]} \;=\; \tau^{1/2}B^{(1)}\,, \quad \Big(Z^{(\tau)}(\tau+t(1-\tau))\Big)_{t\in[0,1]} \;=\; (1-\tau)^{1/2}B^{(2)} \quad (15)$$

with two independent Brownian bridges $B^{(1)}$ and $B^{(2)}$.

In the nonparametric model, (13) follows from a maximal inequality for empirical processes such as in Alexander (1984); see also Dümbgen (1991, Lemma 1). (14) follows from Lemma 1 and Lemma 2 below. Just note that conditional on $S_n^{(\tau)}$ the two processes $(Z_n^{(\tau)}(t))_{t\in\Theta_n:t<\tau}$ and $(Z_n^{(\tau)}(t))_{t\in\Theta_n:t>\tau}$ are independent and behave similarly as the processes $(\tau^{1/2}B_{n\tau}(t/\tau))_{t\in\Theta_n:t<\tau}$ and $\Big((1-\tau)^{1/2}B_{n-n\tau}((t-\tau)/(1-\tau))\Big)_{t\in\Theta_n:t>\tau}$ respectively, where $B_2, B_3, \ldots$ are defined as follows:

Let $x_n = (x_n(1), \ldots, x_n(n))$ be a fixed point in $\mathbf{X}^n$, let $R_n := n^{-1}\sum_{i=1}^n \delta_{x_n(i)}$, and let Π_n be uniformly distributed on the set of all permutations of $\{1, \ldots, n\}$. Then define

$$B_n(t) \;:=\; \sqrt{n}^{-1}\sum_{i=1}^{nt}(\delta_{\Pi_n x_n(i)} - R_n)\,.$$

Lemma 2: *There are constants $K, L > 0$ depending only on $\mathcal{F}$ such that*

$$\mathbb{P}\Big\{\max_{t\in\Theta_n:t\le\sigma}\|B_n(t)\|_{\mathcal{F}} \ge \sigma^{1/2}\eta\Big\} \;\le\; K\exp(-L\eta^2) \quad \forall\,\sigma,\eta > 0\,.$$

Proof of Lemma 2: Since $(B_n(t))_{t\in\Theta_n} =_{\mathcal{L}} (-B_n(1-t))_{t\in\Theta_n}$,

$$\mathbb{P}\Big\{\max_{t\le\sigma}\|B_n(t)\|_{\mathcal{F}} \ge \sigma^{1/2}\eta\Big\} \;\le\; 2\mathbb{P}\Big\{\max_{t\le 1/2}\|B_n(t)\|_{\mathcal{F}} \ge \sigma^{1/2}\eta\Big\}\,.$$

Hence one may assume without loss of generality that $\sigma \in \Theta_n \cap [0, 1/2]$. Now define $\epsilon := \sigma^{1/2}\eta$ and

$$\begin{aligned} A \;&:=\; \{\max_{t\le\sigma}\|B_n(t)\|_{\mathcal{F}} \ge \epsilon\}\,, \\ A_t \;&:=\; \{\|B_n(t)\|_{\mathcal{F}} \ge \epsilon \text{ and } \|B_n(s)\|_{\mathcal{F}} < \epsilon \text{ for } s < t\}\,. \end{aligned}$$

Then,

$$\mathbb{P}(A) \;\le\; \mathbb{P}\{\|B_n(\sigma)\|_{\mathcal{F}} \ge \epsilon/4\} \;+\; \sum_{t\in\Theta_n : t<\sigma} \mathbb{P}\Big\{A_t \cap \{\|B_n(\sigma)\|_{\mathcal{F}} < \epsilon/4\}\Big\}\,,$$

and one can show with the triangle inequality that $A_t \cap \{\|B_n(\sigma)\|_{\mathcal{F}} < \epsilon/4\}$ is a subset of

$$A_t \cap \Big\{\Big\|(B_n(\sigma) - B_n(t)) - (1-t)^{-1}(\sigma - t)(B_n(1) - B_n(t))\Big\|_{\mathcal{F}} \ge \epsilon/4\Big\}\,.$$

The event A_t is measurable with respect to $\Pi_n(1), \ldots, \Pi_n(nt)$, and conditional on $\Pi_n(1), \ldots, \Pi_n(nt)$ the random measure $(B_n(\sigma) - B_n(t)) - (1-t)^{-1}(\sigma - t)(B_n(1) - B_n(t))$ behaves similarly as $(1-t)^{1/2} B_{n-nt}((\sigma - t)/(1-t))$. Consequently the asserted inequality follows via Tshebyshev's inequality from the following one:

$$\begin{aligned}&\text{There exist } K', L' > 0 \text{ depending only on } \mathcal{F} \text{ such that}\\ &\mathbb{E}\Big(\exp(\lambda t^{-1}\|B_n(t)\|_{\mathcal{F}}^2)\Big) \;\le\; 1 + K'\lambda/(L' - \lambda) \quad \forall\, \lambda \in (0, L') \quad \forall\, t \in \Theta_n\,.\end{aligned} \tag{16}$$

The aforementioned maximal inequalities for empirical processes imply that (16) is true, if $B_n(t)$ is replaced with

$$\tilde{B}_n(t) \;:=\; \sqrt{n}^{-1} \sum_{i=1}^{nt} (\delta_{X_i} - R_n)\,,$$

where $X_1, \ldots, X_{nt}$ are independent with distribution R_n. But

$$\mathbb{E}\Big(h(B_n(t))\Big) \;\le\; \mathbb{E}\Big(h(\tilde{B}_n(t))\Big)$$

for arbitrary convex functions h on the linear span of $\delta_{x_{n(1)}}, \ldots, \delta_{x_{n(n)}}$, according to Theorem 4 of Hoeffding (1963); see LeCam (1986, Lemma 16.7.2) for an elegant proof $\square$

4.3. Proof of Theorems 1a-b

For any fixed number $\epsilon > 0$ let A_n be events in the underlying probability space such that $\mathbb{P}(A_n) \geq 1 - \epsilon + o(1)$. All subsequent statements are meant to hold along $(A_n)_n$. According to (13), the A_n can be chosen such that $(Y_n)_n$ meets the requirements of Proposition 2 and

$$\left\|(n\theta_n)^{-1}S_n(\theta_n) - P_n\right\| \vee \left\|(n - n\theta_n)^{-1}(S_n(1) - S_n(\theta_n)) - Q_n\right\| \;=\; o(\sqrt{n}^{-1}\|\Delta_n\|_n)\;;$$

in particular, $\|Y_n - \Delta_n\| = o(\|\Delta_n\|_n)$. Hence one has to show that $\mathrm{dist}(\hat{C}_n, \theta_n) = O(\|Y_n\|_n^{-2})$. In addition one may assume that the following four conditions hold:

$$L_n \;\leq\; \eta_1 \tag{17}$$

for some $\eta_1 > 0$ (by (14));

$$T_n \;\geq\; \eta_2\, k(\theta_n)^{\beta}(\|Y_n\|_n^2 \wedge n)^{-1}\exp(W_n(\theta_n)) \tag{18}$$

for some $\eta_2 > 0$ (by Proposition 2);

$$\max_{\tau\in\Theta_n(\lambda)} \|Y_n\|_n^{-1}\|Y_n^{(\tau)} - k(\tau)^{-1}k(\tau,\theta_n)Y_n\| \;\to\; 0\;, \tag{19}$$

for any fixed $\lambda > 4$ (by (4) and (6));

$$W_n(\tau) - W_n(\theta_n) \;\leq\; -\eta_3\,|\tau - \theta_n|\|Y_n\|_n^2 \quad \forall\,\tau \in \Theta_n(\lambda)\setminus\Theta_n(\lambda,\eta_4) \tag{20}$$

for some $\eta_3, \eta_4 > 0$ (by (7)).

Note that $\hat{C}_n$ is a subset of $\{\tau \in \Theta_n : \hat{q}_n(\tau) \geq T_n\}$, where $\hat{q}_n(\tau)$ stands for the quantile $\max\{r \in \mathbf{R} : \mathbb{P}(T_n^{(\tau)} \geq r|X_n) \geq \alpha\}$. According to (14) one may apply Proposition 1 to all processes $D_n^{(\tau)}$, $\tau \in \Theta_n$. Together with (2) this implies that

$$\hat{q}_n^{(>)}(\tau) \;\leq\; \eta_5\Big(k(\tau)^{(\beta+1)/4}\exp(2W_n(\tau)) + \exp(4k(\tau)^{1/2}W_n(\tau))\Big)$$

$$
\begin{aligned}
&\leq \eta_5 \left(k(\tau)^{(\beta+1)/4} \log(1/k(\tau))^{\eta_1} + \exp\left(4\eta_1 k(\tau)^{1/2} \log\log(1/k(\tau)) \right) \right) \\
&\qquad \cdot \exp(2\rho(\tau,\theta_n)^2 \|Y_n\|_n^2) \\
&\leq \eta_6 \exp(2\rho(\tau,\theta_n)^2 \|Y_n\|_n^2) \quad \text{and} \\
\hat{q}_n^{(=)}(\tau) &\leq \eta_5 (\log n)^{\eta_5} \exp(2W_n(\tau)) \\
&\leq \eta_6 (\log n)^{\eta_6} \exp(2\rho(\tau,\theta_n)^2 \|Y_n\|_n^2)
\end{aligned}
$$

for some $\eta_5, \eta_6 > 0$ and for all $n \geq n_1$ with a fixed integer n_1. Together with (4) and (18) this implies that

$$\hat{C}_n \subset \Theta_n(\lambda) \quad \forall n \geq n_2$$

for a suitable $n_2 \geq n_1$. But (19) and (4) show that one can apply Proposition 2 to all $D_n^{(\tau)}$, $\tau \in \Theta_n(\lambda)$, simultaneously for proving that

$$
\begin{aligned}
\hat{q}_n(\tau) &\leq \eta_7 k(\tau)^{\beta} (\|Y_n^{(\tau)}\|_n^2 \wedge n)^{-1} \exp(W_n(\tau)) \\
&\leq \eta_8 k(\theta_n)^{\beta} (\|Y_n\|_n^2 \wedge n)^{-1} \exp(W_n(\tau)) \quad \forall \tau \in \Theta_n(\lambda) \quad \forall n \geq n_3
\end{aligned}
$$

for some $\eta_7, \eta_8 > 0$ and some $n_3 \geq n_2$. Hence

$$\hat{C}_n \cap \Theta_n(\lambda) \subset \Theta_n\left(\lambda, \eta_4 \vee (\eta_3^{-1} \log(\eta_8/\eta_2))\right) \quad \forall n \geq n_3 ,$$

according to (18) and (20) $\square$

4.4. Proof of Theorems 2a-b

For an arbitrary fixed $\epsilon > 0$ let the events A_n be as in section 4.3, and again all subsequent statements are meant to hold along $(A_n)_n$. According to (1) one may assume that

$$\|\delta_n^{-1} \Delta_n\|_n^2 \rightarrow 1 , \tag{21}$$

where $\delta_n := \Delta_n$ in the normal shift model and $\delta_n := \sqrt{3}\int \Delta_n\, P_n(dx)$ in the non-parametric model. In particular, $\|\delta_n^{-1}Y_n\|_n^2 \to 1$. The proof of Theorems 1a-b shows that

$$\max_{|\tau|\geq\gamma_n} \hat{p}_n(\theta_n + \delta_n^{-2}\tau) \;\to\; 0 \quad \text{whenever } \gamma_n \to \infty\ .$$

Hence it suffices to show that $(\hat{p}_n(\theta_n + \delta_n^{-2}\tau))_{\tau\in[-\gamma,\gamma]}$ converges in distribution to $(\hat{p}(\tau))_{\tau\in[-\gamma,\gamma]}$ for any fixed $\gamma > 0$.

Let $\lambda_n := (1 - \gamma k(\theta_n)^{-1}\delta_n^{-2})^{-1}$. Then $1 < \lambda_n \to 1$, and $\Theta_n(\lambda_n)$ contains all $\tau \in \Theta_n$ with $|\tau - \theta_n| \leq \gamma\delta_n^{-2}$. It is shown below that the $(A_n)_n$ can be chosen such that

$$\mathcal{L}\Big((\tilde{W}_n^{(\tau)}(\tau + \delta_n^{-2}r))_{r\in[-\tilde{\gamma},\tilde{\gamma}]}\Big|X_n\Big) \;\to_w\; \mathcal{L}\Big((W(r))_{r\in[-\tilde{\gamma},\tilde{\gamma}]}\Big) \tag{22}$$
$$\text{uniformly in (u.i.) } \tau \in \Theta_n(\lambda_n) \quad \forall\, \tilde{\gamma} > 0.$$

Further one may assume that

$$T_n \;=\; k(\theta_n)^{\beta+1}W_n(\theta_n)^{-1}\exp(W_n(\theta_n))\,\tilde{T}_n\,(1 + o(1))\ ,$$
$$W_n(\tau) - W_n(\theta_n) - \tilde{W}_n(\tau) \;\to\; 0 \quad \text{u.i. } \tau \in \Theta_n(\lambda_n)\ ,$$

according to Proposition 2 and (10). It follows from (4) and (19) that

$$k(\tau)/k(\theta_n) \;\to\; 1\ , \quad W_n(\tau)/W_n(\theta_n) \;\to\; 1 \quad \text{and} \quad \|\delta_n^{-1}Y_n^{(\tau)}\|_n^2 \;\to\; 1 \tag{23}$$
$$\text{u.i. } \tau \in \Theta_n(\lambda_n).$$

Consequently $\hat{p}_n(\tau)$ can be written as

$$\mathbb{P}\Big(\exp(-W_n(\tau))k(\tau)^{-\beta-1}W_n(\tau)\,T_n^{(\tau)} \geq \exp(-\tilde{W}_n(\tau))\tilde{T}_n(1 + r_n(\tau))\Big|X_n\Big)\ ,$$

where $r_n(\tau) \to 0$ u.i. $\tau \in \Theta_n(\lambda_n)$. But now one can apply Proposition 2, (9), (22) and the Continuous Mapping Theorem to all processes D_n, $D_n^{(\tau)}$, $\tau \in \Theta_n(\lambda_n)$, for showing that

$$\Big(\exp(-\tilde{W}_n(\theta_n + \delta_n^{-2}\tau))\tilde{T}_n\Big)_{\tau\in[-\gamma,\gamma]}$$
$$\to_{\mathcal{L}} \ \Big(\exp(-W(\tau))\int \exp(W(t))\,dt\Big)_{\tau\in[-\gamma,\gamma]}$$

and

$$\mathcal{L}\Big(\exp(-W_n(\tau))k(\tau)^{-\beta-1}W_n(\tau)\,T_n^{(\tau)}\Big|X_n\Big) \ \to_w \ \mathcal{L}\Big(\int \exp(W(t))\,dt\Big)$$

u.i. $\tau \in \Theta_n(\lambda_n)$. Since H is continuous, this implies that $\hat{p}_n(\tau)$ can be uniformly approximated by $H\Big(\exp(-\tilde{W}_n(\tau))\tilde{T}_n\Big)$, and the desired result follows.

It remains to prove claim (22). For notational convenience we first consider the normal shift model: Here

$$\begin{aligned}
\tilde{W}_n^{(\tau)}(t) &= |Y_n^{(\tau)}|\,|k(\tau)Y_n^{(\tau)} + Z_n^{(\tau)}(t) - Y_n^{(\tau)}|t-\tau|/2| \ - \ k(\tau)Y_n^{(\tau)2} \\
&= Y_n^{(\tau)}Z_n^{(\tau)}(t) - Y_n^{(\tau)2}|t-\tau|/2 \ ,
\end{aligned}$$

provided that

$$|Y_n^{(\tau)}Z_n^{(\tau)}(t) - Y_n^{(\tau)2}|t-\tau|/2| \ \le \ k(\tau)Y_n^{(\tau)2} \ .$$

But $\min_{\tau\in\Theta_n(\lambda_n)} k(\tau)Y_n^{(\tau)2} \to \infty$ and

$$\begin{aligned}
&|Y_n^{(\tau)}Z_n^{(\tau)}(t) - Y_n^{(\tau)2}|t-\tau|/2| \\
&\quad\le \ \tilde{\gamma}^{1/2}|\delta_n^{-1}Y_n^{(\tau)}|\,M_n^{(\tau)}(\tilde{\gamma}\delta_n^{-2}) \ + \ \tilde{\gamma}\delta_n^{-2}Y_n^{(\tau)2}/2 \\
&\quad= \ O(1)M_n^{(\tau)}(\tilde{\gamma}\delta_n^{-2}) + O(1) \ , \\
&|Y_n^{(\tau)}Z_n^{(\tau)}(t) - Y_n^{(\tau)2}|t-\tau|/2 \ - \ \delta_n Z_n^{(\tau)}(t) + \delta_n^2|t-\tau|/2| \\
&\quad\le \ \tilde{\gamma}^{1/2}|\delta_n^{-1}Y_n^{(\tau)2} - 1|\,M_n^{(\tau)}(\tilde{\gamma}\delta_n^{-2}) \ + \ \tilde{\gamma}|\delta_n^{-2}Y_n^{(\tau)2} - 1|/2 \\
&\quad= \ o(1)M_n^{(\tau)}(\tilde{\gamma}\delta_n^{-2}) + o(1)
\end{aligned}$$

for all $t \in \Theta_n$ with $|t-\tau| \leq \tilde{\gamma}\delta_n^{-2}$; see (23). Together with (14) it follows that one may replace $W_n^{(\tau)}(t)$ with $\delta_n Z_n^{(\tau)}(t) - \delta_n^2|t-\tau|/2$ when checking (22). But one can deduce from (15) that for any fixed $\tilde{\gamma} > 0$,

$$(\delta_n Z_n^{(\tau)}(\tau + \delta_n^{-2}r))_{r\in[-\tilde{\gamma},\tilde{\gamma}]} \to_{\mathcal{L}} (Z(r))_{r\in[-\tilde{\gamma},\tilde{\gamma}]}$$

u.i. $\tau \in \Theta_n(\lambda_n)$, and (22) follows for the normal case.

As for the nonparametric model, note first that

$$\max_{\tau\in\Theta_n(\lambda_n)} \left\|(n\tau)^{-1}S_n(\tau) - P\right\| \vee \left\|(n-n\tau)^{-1}(S_n(1)-S_n(\tau)) - P\right\| \to 0 , \qquad (24)$$

where $\|\cdot\|$ stands for the Kolmogorov-Smirnov norm times $\sqrt{12}$. For $\|(n\tau)^{-1}S_n(\tau) - P\|$ can be approximated by

$$\|(n\tau)^{-1}S_n(\tau) - (n\theta_n)^{-1}S_n(\theta_n)\| = \sqrt{n}^{-1}\|(1-\tau)Y_n^{(\tau)} - (1-\theta_n)Y_n\| ,$$

and one can easily show that the right hand side tends to zero u.i. $\tau \in \Theta_n(\lambda_n)$; see (19). The measure $(n-n\tau)^{-1}(S_n(1)-S_n(\tau))$ can be treated analogously.

Similarly as in the normal shift model one can show that $\tilde{W}_n^{(\tau)}(t)$ may be replaced with

$$\delta_n\sqrt{3}\int Z_n^{(\tau)}(t)(x)\,R_n(dx) - \delta_n^2|t-\tau|/2$$

when checking (22); here R_n denotes $n^{-1}S_n(1)$. One can write

$$\int Z_n^{(\tau)}(t)(x)\,n^{-1}S_n(1)(dx) = \begin{cases} \sum_{nt<i\leq n\tau} \Pi_n^{(\tau)} x_n^{(\tau)}(i) , & \text{if } t \leq \tau , \\ \sum_{n\tau<i\leq nt} \Pi_n^{(\tau)} x_n^{(\tau)}(i) , & \text{if } t \geq \tau , \end{cases}$$

where

$$\sqrt{n}\,x_n^{(\tau)}(i) := \begin{cases} R_n(X_n(i)) - (n\tau)^{-1}\sum_{j\leq n\tau} R_n(X_n(j)) , & \text{if } i \leq n\tau , \\ -R_n(X_n(i)) + (n-n\tau)^{-1}\sum_{j>n\tau} R_n(X_n(j)) , & \text{if } i > n\tau . \end{cases}$$

These vectors $x_n^{(\tau)}$ have coordinates in $[-\sqrt{n}^{-1}, \sqrt{n}^{-1}]$, and both $\sum_{i\leq n\tau} x_n^{(\tau)}(i)$ and $\sum_{i>n\tau} x_n^{(\tau)}(i)$ are zero. Moreover one can deduce from (24) that both

$\tau^{-1}\sum_{i\le n\tau} x_n^{(\tau)}(i)^2$ and $(1-\tau)^{-1}\sum_{i>n\tau} x_n^{(\tau)}(i)^2$ converge to 1/3 u.i. $\tau \in \Theta_n(\lambda_n)$. Thus Lemma 3 below implies that the conditional distribution of

$$\Big(\delta_n\sqrt{3}\int Z_n^{(\tau)}(\tau+\delta_n^{-2}r)(x)\,R_n(dx)\Big)_{r\in[-\tilde\gamma,\tilde\gamma]}$$

given X_n converges weakly to $\mathcal{L}\Big((Z(r))_{r\in[-\tilde\gamma,\tilde\gamma]}\Big)$ u.i. $\tau\in\Theta_n(\lambda)$ for any fixed $\tilde\gamma>0$
□

In order to formulate Lemma 3 let the random permutation Π_n be as in section 4.2, let $x_n = (x_n(1),\ldots,x_n(n))$ be a vector in $\mathbf{R}^n$ such that $\sum_{i=1}^n x_n(i) = 0$ and $\sum_{i=1}^n x_n(i)^2 = 1$. Then define

$$b_n(t) \;:=\; \sum_{1\le i\le nt} \Pi_n x_n(i)\ , \quad t\in[0,1]\ .$$

Lemma 3: *Suppose that* $\max_{1\le i\le n}\gamma_n^{-1}x_n(i)^2 \to 0$, *where* $\gamma_n>0$ *are constants such that* $\gamma_n\to 0$ *and* $n\gamma_n\to\infty$. *Then*

$$(\gamma_n^{-1/2}b_n(r\gamma_n))_{r\in[0,\tilde\gamma]} \;\to_{\mathcal{L}}\; (Z(r))_{r\in[0,\tilde\gamma]} \quad \forall\,\tilde\gamma>0\ .$$

This can be proved with the techniques of Billingsley (1968, chapter 4), especially Theorems 19.3, 19.4 and 24.1. For a different method of proof see Dümbgen (1992).

Acknowledgements. I am grateful to Lucien LeCam for answering numerous questions and for interesting conversations during my stay at Berkeley. Helpful suggestions of an anonymous referee are gratefully acknowledged.

References

Alexander, K.S. (1984): Probability inequalities for empirical processes and a law of the iterated logarithm. *Annals of Probability* **12**, 1041-1067

Billingsley, P. (1968): *Convergence of Probability Measures.* Wiley, New York

Darkhovskiy, B.S. (1976): A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Prob. Appl.* **21**, 178-183

Dümbgen, L. (1991): The asymptotic behavior of some nonparametric changepoint estimators. *Annals of Statistics* **19**, 1471-1495

Dümbgen, L. (1992): Combinatorial stochastic processes. Preprint

LeCam, L. (1986): *Asymptotic methods in statistical decision theory.* Springer, New York

Pollard, D. (1984): *Convergence of Stochastic Processes.* Springer, New York

Romano, J. (1989): Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17**, 141-159

Siegmund, D. (1986): Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14**, 361-404

Siegmund, D. (1988): Confidence sets in changepoint problems. *Internat. Statist. Rev.* **56**, 31-48

Worsley, K.J. (1986): Confidence regions and tests for a changepoint in a sequence of exponential family random variables. *Biometrika* **73**, 91-104