The Annals of Statistics
1991, Vol. 19, No. 3, 1471–1495

THE ASYMPTOTIC BEHAVIOR OF SOME NONPARAMETRIC CHANGE-POINT ESTIMATORS[1]

BY L. DÜMBGEN

Universität Heidelberg

Consider a sequence $X_1, X_2, \ldots, X_n$ of independent random variables, where $X_1, X_2, \ldots, X_{n\theta}$ have distribution P, and $X_{n\theta+1}, X_{n\theta+2}, \ldots, X_n$ have distribution Q. The change-point $\theta \in (0,1)$ is an unknown parameter to be estimated, and P and Q are two unknown probability distributions. The nonparametric estimators of Darkhovskh and Carlstein are imbedded in a more general framework, where random seminorms are applied to empirical measures for making inference about θ. Carlstein's and Darkhovskh's results about consistency are improved, and the limiting distributions of some particular estimators are derived in various models. Further we propose asymptotically valid confidence regions for the change point θ by inverting bootstrap tests. As an example this method is applied to the Nile data.

1. Introduction. Consider the following change-point model: For $n = 2, 3, 4, \ldots$ let P_n and Q_n be two probability distributions on a measurable space E, and let $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ be independent random variables. These variables $X_{i,n}$ have distribution P_n for $i \le n\theta_n$ and Q_n otherwise. The change point θ_n is some unknown number in $T_n := \{1/n, 2/n, \ldots, (n-1)/n\}$. For increasing n this number θ_n remains bounded away from 0 and 1, and for simplicity we assume that $\theta_n \to \theta \in (0,1)$ as $n \to \infty$.

This model has been studied extensively under various aspects. One problem is to test whether there exists a change in the underlying distribution at an unknown time point. Wolfe and Schechtman (1984) summarize nonparametric procedures (especially tests) for the change-point problem. More recently Csörgő and Horvath (1988) discuss nonparametric tests based on U-statistic type processes, and Romano (1989) proposes randomization tests and bootstrap tests for a change point.

If the presence of a change point is assumed, it is of interest to estimate the change point θ_n and to construct confidence regions. The present paper is concerned with these two problems. They have been studied mostly in parametric models for P_n and Q_n; see, for instance, Hinkley (1970), Cobb (1978), Worsley (1986) and Siegmund (1988). Two references for nonparametric estimators are Darkhovskh (1976) and Carlstein (1988).

In Section 2 the estimators of Carlstein and Darkhovskh are imbedded in a more general framework, where we apply random seminorms to certain empirical measures for making inference about θ. We give rates of convergence

Received December 1989; revised July 1990.

[1]Part of this work supported by the Deutsche Forschungsgemeinschaft.

AMS 1980 *subject classifications*. Primary 62G05, 62G15.

Key words and phrases. Change-point, nonparametric estimation, Kolmogorov–Smirnov, random seminorm, Mann–Whitney, bootstrap confidence sets.

under general assumptions. For example, suppose that $P_n = P$ and $Q_n = Q$ for all n. In parametric models, where P and Q are specified up to an unknown finite-dimensional parameter, the maximum likelihood estimator for θ_n is typically consistent with rate $O_p(n^{-1})$; see, for instance, Hinkley (1970) or Hinkley and Hinkley (1970). It follows from Theorem 1 that in general the nonparametric estimators are consistent with this rate $O_p(n^{-1})$, too.

In Section 3 we consider four particular seminorms and present the limiting distribution of the corresponding estimators in a general model for the sequences (P_n) and (Q_n). Further we propose a modified nonparametric estimator which corresponds to a Bayes estimator in parametric models. The nonparametric estimators are compared with semiparametric estimators in a mean shift model; see Section 4.

The results of Section 3 are applied in Section 5, where we propose bootstrap confidence sets for θ_n. These confidence sets are asymptotically valid and their size tends stochastically to 0. Finally in Section 6 this procedure is applied to the Nile data; see Cobb (1978).

Section 7 contains the more technical proofs.

2. The nonparametric estimators. At the beginning we describe a general procedure for estimating θ_n. The special estimators of Carlstein (1988) and Darkhovskh (1976) are given later; see Examples II and III. For every hypothetical change point $t \in T_n$, consider the empirical distributions

$$ {}^tP_n := \frac{1}{nt}\sum_{i=1}^{nt} \delta_{X_{i,n}} \quad \text{and} \quad P_n^t := \frac{1}{n(1-t)}\sum_{i=nt+1}^{n} \delta_{X_{i,n}}. $$

The (pointwise) expectation of these empirical measures ${}^tP_n$ and P_n^t is denoted by ${}^t\Pi_n$ and Π_n^t respectively. Then the difference $P_n^t - {}^tP_n$ estimates the signed measure

$$ \Pi_n^t - {}^t\Pi_n = \big[(\theta_n/t) \wedge ((1-\theta_n)/(1-t))\big](Q_n - P_n). $$

If t is near 0 or 1, this is a poor estimator. It seems reasonable to introduce weights $w(t) := t^{1/2}(1-t)^{1/2}$ and to consider the measures

$$ D_n^t := w(t)\big(P_n^t - {}^tP_n\big), \qquad t \in T_n. $$

These signed measures D_n^t estimate

$$ \Delta_n^t := w(t)\big(\Pi_n^t - {}^t\Pi_n\big) = \rho_n(t)(Q_n - P_n), $$

where

$$ \rho_n(t) := (1-\theta_n)\big[t/(1-t)\big]^{1/2} \wedge \theta_n\big[(1-t)/t\big]^{1/2}. $$

On the interval $[0, \theta_n]$ this function ρ_n is strictly increasing, and on $[\theta_n, 1]$ it is strictly decreasing. It attains its maximum in θ_n with $\rho_n(\theta_n) = w(\theta_n)$. Now we choose a seminorm N_n on the space $\mathscr{M}$ of all finite signed measures on E and compute $N_n(D_n^t)$ as a measure of difference between ${}^tP_n$ and P_n^t. In particular, we consider the estimator

$$\hat{\theta}_n := \arg\max\big(N_n(D_n^t) : t \in T_n\big),$$

which corresponds to the maximum likelihood estimator in parametric models. In some situations there are better nonparametric estimators than $\hat{\theta}_n$ corresponding to a certain Bayes estimator in a parametric context; see Section 3. However, it is convenient to study first $\hat{\theta}_n$. If some prior information about θ_n is available, one can use a truncated estimator

$$\hat{\theta}_n := \arg\max\big(N_n(D_n^t) : t \in T_n, t_0 \le t \le t_1\big),$$

where $0 < t_1 < \theta < t_1 < 1$. The seminorm N_n may be deterministic or random, as we shall see in some examples below. For a better understanding of the regularity conditions (2.1) and (2.2), it is useful to compare the observed process $(N_n(D_n^t)$: $t \in T_n)$ with the theoretical process $(N_n(\Delta_n^t)$: $t \in T_n)$. The latter has a simple structure:

$$N_n(\Delta_n^t) = \rho_n(t) N_n(Q_n - P_n), \qquad t \in T_n.$$

Hence $\theta_n = \arg\max(N_n(\Delta_n^t)$: $t \in T_n)$ and $\max\{N_n(\Delta_n^t)$: $t \in T_n\} = w(\theta_n) N_n(Q_n - P_n)$. Roughly speaking, $\hat{\theta}_n$ is a reasonable estimator for θ_n, if the maximal difference of these two processes $N_n(D_n^{(\cdot)})$ and $N_n(\Delta_n^{(\cdot)})$ is small compared to $N_n(Q_n - P_n)$. According to the triangle inequality, $|N_n(D_n^t) - N_n(\Delta_n^t)| \le N_n(D_n^t - \Delta_n^t)$. To ensure that these quantities are small enough, we make the following assumption:

(2.1) There is a Vapnik–Cervonenkis class $\mathscr{D}$ of measurable subsets of E such that

$$N_n(\nu) \le \|\nu\| := \sup\{|\nu(D)| : D \in \mathscr{D}\}, \qquad \forall\, n \ge 2, \forall\, \nu \in \mathscr{M}.$$

[For the definition of Vapnik–Cervonenkis (VC) classes of sets and for examples, see Pollard (1984), Chapter 2.] On the other hand, $N_n(Q_n - P_n)$ has to be large enough. In view of Theorem 1 we make the following assumption about the sequences (N_n), (P_n) and (Q_n):

(2.2) There is a constant $C_0 > 0$ and a sequence (γ_n) in $\mathbb{R}^+$ such that

$$\Pr\big(N_n(Q_n - P_n) \ge C_0 \gamma_n^{-1}\big) \to 1 \quad \text{and}$$
$$\gamma_n (\log\log n)^{1/2} n^{-1/2} \to 0 \quad \text{as } n \to \infty.$$

If we use the truncated estimator, we only have to require $\gamma_n n^{-1/2} \to 0$ in (2.2). An important special case of (2.2) is:

(2.3) There is a constant $C_0 > 0$ such that

$$\Pr\big(N_n(Q_n - P_n) \ge C_0\big) \to 1 \quad \text{as } n \to \infty.$$

For some suitable seminorms N_n condition (2.3) is automatically fulfilled, if $P_n = P$ and $Q_n = Q$ for all $n \ge 2$ with two fixed distributions $P \ne Q$; see below.

Now we can state our first result.

THEOREM 1. *If the seminorms N_n and the distributions P_n and Q_n meet the requirements* (2.1) *and* (2.2), *then*

$$\hat{\theta}_n - \theta_n = O_p(\gamma_n^2 n^{-1}).$$

This theorem is a direct consequence of the following stronger result which is also needed for later purposes.

PROPOSITION 1. *Suppose that the seminorms N_n and the distributions P_n and Q_n meet the requirements* (2.1) *and* (2.2). *Then there is a constant $C_1 > 0$ such that*

$$\liminf_{n\to\infty} \Pr(A_n(d)) \to 1 \quad \textit{as } d \to +\infty,$$

where

$$A_n(d) := \{N_n(D_n^t) - N_n(D_n^{\theta_n}) \le -C_1\gamma_n^{-1}|t - \theta_n|, \forall\, t \in T_n \setminus T_n(d)$$
$$\textit{and } N_n(D_n^{\theta_n}) \ge C_1\gamma_n^{-1}\},$$

$$T_n(d) := \{t \in T_n: |t - \theta_n| \le d\gamma_n^2 n^{-1}\}.$$

We end this section with some examples.

EXAMPLE I. Let $\mathscr{D}$ be a VC class of measurable subsets of $\mathbb{R}^p$ and let N_n be the corresponding seminorm $\|\cdot\|$. Here condition (2.1) is obviously fulfilled. With $C_0 := 1$ and $\gamma_n := \|Q_n - P_n\|^{-1}$, requirement (2.2) is also met, provided that $\|Q_n - P_n\|^{-1} = o((\log\log n)^{-1/2} n^{1/2})$. For example,

$$\mathscr{D}_0 := \{\{y \in \mathbb{R}^p: \langle y, x\rangle \ge u\}: x \in \mathbb{R}^p, u \in \mathbb{R}\}$$

($\langle\cdot,\cdot\rangle$ denotes the standard inner product) is a VC class such that the corresponding seminorm $\|\cdot\|$ is even a norm on $\mathscr{M}$. In this case (2.3) is valid automatically if $P_n = P$ and $Q_n = Q$ for all $n \ge 2$ with $P \ne Q$.

Maybe the family $\mathscr{D}$ is too big to compute the quantities $\|D_n^t\|$ exactly. A possible way out is to choose a subset $\mathscr{D}^{(n)}$ of $\mathscr{D}$ by a random mechanism and to compute $N_n(\nu) := \sup\{|\nu(D)|: D \in \mathscr{D}^{(n)}\}$ instead of $\|\nu\|$. Beran and Millar (1986) study such a stochastic approximation for $\|\cdot\|$ corresponding to $\mathscr{D}_0$ above.

In Examples II and III and in Sections 3–5, we consider real-valued variables $X_{i,n}$. For notational convenience we identify arbitrary finite signed measures ν on the line with the corresponding cdf, that is, $\nu(x)$ stands for $\nu(-\infty, x]$, $x \in \mathbb{R}$. Moreover the seminorms N_n discussed there may be regarded as seminorms on $D(\mathbb{R})$, the space of cadlag functions f on the line such that $f(x) \to 0$ as $|x| \to +\infty$.

EXAMPLE II (Estimators of Carlstein). Carlstein generally uses seminorms N_n such that

$$\int |\nu(x)| \hat{R}_n(dx) \le N_n(\nu) \le \sup\{|\nu(x)|: x \in \mathbb{R}\}, \tag{2.4}$$

where $\hat{R}_n$ is the empirical measure $n^{-1}\sum_{i=1}^{n}\delta_{X_{i,n}}$. This measure estimates the mixture distribution $R_n := \theta_n P_n + (1-\theta_n)Q_n$. Examples for seminorms N_n satisfying (2.4) are given by

$$N_{n,r}(\nu) := \left(\int |\nu(x)|^r \hat{R}_n(dx)\right)^{1/r}, \qquad r \ge 1.$$

Generally it follows from (2.4) together with Tshebyshev's inequality that

$$N_n(Q_n - P_n) \ge \int |(Q_n - P_n)(x)| R_n(dx) + O_p(n^{-1/2}).$$

In order to verify condition (2.2), with a given sequence (γ_n), it is therefore sufficient to show that

$$\liminf_{n\to\infty} \gamma_n \int |(Q_n - P_n)(x)| R_n(dx) > 0.$$

For example, if $P_n = P$ and $Q_n = Q$ for all n, (2.3) is valid. For

$$\int |(Q-P)(x)| R_n(dx) = \int |(Q-P)(x)| R(dx) + o(1).$$

where $R := \theta P + (1-\theta)Q$. The limit is strictly positive because

$$(P+Q)\{x \in \mathbb{R}: P(x) \ne Q(x)\} > 0 \quad \text{whenever } P \ne Q.$$

EXAMPLE III [Darkhovskh's estimator (modified)]. Darkhovskh uses the Mann–Whitney statistic

$$M_n(t) := \frac{1}{n^2 t(1-t)} \sum_{i \le nt < j} I(X_{i,n} \le X_{j,n})$$

as a measure of difference between ${}^tP_n$ and P_n^t. He essentially proposes the estimator

$$\hat{\theta}_n^* := \arg\max(|M_n(t) - 1/2|: t \in T_n, t_0 \le t \le t_1),$$

where $0 < t_0 < \theta < t_1 < 1$. This estimator is consistent, if $P_n = P$ and $Q_n = Q$ for all n with continuous cdf's P and Q such that $\int P(x)Q(dx) \ne 1/2$.

We consider a modified version of Darkhovskh's estimator such that no prior information about θ_n, P_n, Q_n is necessary. For arbitrary $f \in D(\mathbb{R})$ we use the symmetrized function

$$f(x^*) := (f(x-) + f(x))/2, \qquad x \in \mathbb{R}.$$

If P' and Q' are probability measures on the line, $\int P'(x^*)Q'(dx) = 1 - \int Q'(x^*)P'(dx)$ and $\int P'(x^*)P'(dx) = 1/2$. Then we define

$$N_{n,D}(\nu) := \left| \int \nu(x^*) \hat{R}_n(dx) \right|$$

and

$$\hat{\theta}_{n,D} := \arg\max\left(N_{n,D}\left(D_n^t\right): t \in T_n \right).$$

One can show that $w(t)|M_n(t)| = N_{n,D}(D_n^t)$ for all $t \in T_n$ almost surely, if P_n and Q_n are continuous; see Dümbgen (1990). According to Tshebyshev's inequality,

$$\begin{aligned} N_{n,D}(Q_n - P_n) &= \left| \int (Q_n - P_n)(x^*) R_n(dx) \right| + O_p(n^{-1/2}) \\ &= \left| \int P_n(x^*) Q_n(dx) - 1/2 \right| + O_p(n^{-1/2}). \end{aligned}$$

Hence if we want to verify condition (2.2), with a given sequence (γ_n), we just have to show that

$$\liminf_{n \to \infty} \gamma_n \left| \int P_n(x^*) Q_n(dx) - 1/2 \right| > 0.$$

3. Limiting distributions. For the rest of this paper we consider the particular seminorms $\|\cdot\|$, $N_{n,1}$, $N_{n,2}$, $N_{n,D}$ on $D(\mathbb{R})$, where

$$\|\nu\| := \sup\{|\nu(x)|: x \in \mathbb{R}\}$$

is the Kolmogorov–Smirnov norm, and $N_{n,1}$, $N_{n,2}$, $N_{n,D}$ have been defined in Section 2. All objects such as test statistics, confidence sets etc. depend on which seminorm N_n is used. If we want to set off this dependency, we use the subscripts $K, (1), (2), D$. Carlstein (1988) compared the estimators $\hat{\theta}_{n,K}$, $\hat{\theta}_{n,(1)}$, $\hat{\theta}_{n,(2)}$ in a simulation study, and there seemed to be no significant difference between them. In Section 4 we consider a mean shift model and come to a similar conclusion. The estimator $\hat{\theta}_{n,D}$ of Darkhovskh is of special interest because it is easier to compute than the estimators of Carlstein.

The different models for the sequences (P_n) and (Q_n) considered later are a special case of the following one.

General model. There are continuous cdf's P and Q on the line such that both $\|P_n - P\|$ and $\|Q_n - Q\|$ tend to 0 as $n \to \infty$. There is a sequence (γ_n) in $\mathbb{R}^+$ and a continuous function h on $\mathbb{R}$ such that

$$\gamma_n(\log\log n)^{1/2} n^{-1/2} \to 0 \quad \text{and} \quad \|\gamma_n(Q_n - P_n) - h\| \to 0.$$

We give some examples.

EXAMPLE I (Standard model). $P_n = P$ and $Q_n = Q$ for all $n \geq 2$ with different continuous cdf's P and Q on $\mathbb{R}$. Here one may take $\gamma_n := 1$ and $h := Q - P$.

EXAMPLE II (Mean shift model). $P_n(dx) = P(dx) = f(x)\,dx$ and $Q_n(dx) = f(x - \gamma_n^{-1})\,dx$ for every $n \geq 2$, where f is a uniformly continuous probability density on $\mathbb{R}$, and (γ_n) is a sequence as in the general model such that $\gamma_n \to +\infty$. Here $\gamma_n(Q_n - P_n) = \gamma_n(P(\cdot - \gamma_n^{-1}) - P)$ tends uniformly to $h := -f$.

EXAMPLE III (Exponential families). Consider a simple exponential family $P_{(a)}(dx) := \exp(ax - \psi(a))P_{(0)}(dx)$ with a continuous cdf $P_{(0)}$ on the line. Here $a^{-1}(P_{(a)} - P_{(0)})(x)$ converges uniformly in x to

$$h(x) := \int I(y \leq x)(y - m)P_{(0)}(dy)$$

as $a \to 0$, where $m := \int yP_{(0)}(dy)$. This limit $h(x)$ is continuous with $h(m) \leq h(x) < 0$ for $0 < P_{(0)}(x) < 1$. If one chooses a sequence (γ_n) as in the general model with $\gamma_n \to +\infty$, then the distributions $P_n := P_{(0)}$ and $Q_n := P_{(\gamma_n^{-1})}$ meet the requirements of the general model.

Under the assumptions of the general model, there is a constant $\mu \geq 0$ such that

$$\gamma_n N_n(Q_n - P_n) = \mu + o_p(1).$$

If $\mu > 0$, condition (2.2) holds, and Theorem 1 implies that $\hat{\theta}_n - \theta_n = O_p(\gamma_n^2 n^{-1})$. Specifically we have

$$\mu_K = \|h\|, \qquad \mu_{(1)} = \int |h|\,dR, \qquad \mu_{(2)}^2 = \int h^2\,dR, \qquad \mu_D = \left|\int h\,dR\right|,$$

where we define $R := \theta P + (1 - \theta)Q$; see also Section 2. (Note that $\|R_n - R\| \to 0$.)

The key for understanding the limiting behavior of estimators, confidence sets etc. is the limiting behavior of the process $(\tilde{W}_n(t)\colon t \in T_n)$, where

$$\tilde{W}_n(t) := n\left(N_n(D_n^t)^2 - N_n(D_n^{\theta_n})^2\right).$$

For example, one may write

$$\hat{\theta}_n = \arg\max\left(\tilde{W}_n(t) : t \in T_n\right).$$

We distinguish between two special cases of the general model.

MODEL 1. P and Q are different distributions and $\gamma_n = 1$ for all $n \geq 2$.

MODEL 2. $P = Q$ and $\gamma_n \to \infty$ as $n \to \infty$.

With respect to Model 2 we assume that the process $\tilde{W}_n(\cdot)$ is extended to a function in $D[0,1]$ via $\tilde{W}_n(s) := \tilde{W}_n([ns]/n)$, $\tilde{W}_n(0) := \tilde{W}_n(1) := -nN_n(D_n^{\theta_n})^2$. In the sequel let X_z ($z \in \mathbb{Z}$, the set of integers) be independent random variables having distribution P for $z \le 0$ and Q otherwise. Further let $(W(s)\colon s \in \mathbb{R})$ be a two-sided Brownian motion on the line; that is, $(W(s)\colon s \ge 0)$ and $(W(-s)\colon s \ge 0)$ are two independent Brownian motions on $\mathbb{R}_0^+$. Then we have the following result, where the subscript A generally stands for $K, (1), (2), D$.

PROPOSITION 2.

(I) *The limiting behavior of $\tilde{W}_n(\cdot)$ in Model 1: Under some regularity conditions, see* (III) *below, there is stochastic process $(W_A(z)\colon z \in \mathbb{Z})$ such that for arbitrary integers $d > 0$ the vector*

$$\left(\tilde{W}_{n,A}(\theta_n + n^{-1}z)\colon -d \le z \le d\right)$$

converges in distribution to

$$\left(2\mu_A W_A(z) - \mu_A^2|z|\colon -d \le z \le d\right)$$

as $n \to \infty$. More precisely, we have $W_A(0) := 0$ and

$$W_A(z) := \begin{cases} \displaystyle\sum_{i=z+1}^{0} \left[g_A(X_i) - E_{X\sim P}\, g_A(X)\right], & \text{for } z < 0,\\ \displaystyle -\sum_{i=z+1}^{0} \left[g_A(X_i) - E_{X\sim Q}\, g_A(X)\right], & \text{for } z > 0, \end{cases}$$

where $g_A\colon \mathbb{R} \to \mathbb{R}$ is a certain bounded function.

(II) *The limiting behavior of $\tilde{W}_n(\cdot)$ in Model 2: Under some regularity conditions* [*see* (III) *below*] *for arbitrary $d > 0$ the process*

$$\left(\tilde{W}_{n,A}(\theta_n + \gamma_n^2 n^{-1}s)\colon s \in [-d, d]\right)$$

converges in $D[-d,d]$ in distribution to

$$\left(2\mu_A\sigma_A W(s) - \mu_A^2|s|\colon s \in [-d,d]\right)$$

as $n \to \infty$, where

$$\sigma_A^2 := \operatorname{Var}_{X\sim R}(g_A(X)).$$

(III) *Regularity conditions and definitions*:

A = K: *We generally assume that there is a unique $x_0 \in \mathbb{R}$ such that*

$$|h(x_0)| = \|h\| > 0.$$

The function g_K is given by

$$g_K(x) := \operatorname{sign}(h(x_0))I(x \le x_0).$$

A = (1): *We generally assume that the function $|h|$ is strictly positive $(P + Q)$-almost everywhere. The function $g_{(1)}$ is given by*

$$g_{(1)}(x) := \int \operatorname{sign}(h(y))I(x \le y)R(dy).$$

A = (2): *In Model 2 we have to assume that* $\mu_{(2)} > 0$. *We generally define*

$$g_{(2)}(x) := \mu_{(2)}^{-1} \int h(y) I(x \le y) R(dy).$$

A = D: *We generally assume that* $\mu_D > 0$. *The function* g_D *is given by*

$$g_D(x) := -\operatorname{sign}\left(\int h\,dQ\right) R(x).$$

REMARK. The variances $\sigma_A^2 = \operatorname{Var}_{X \sim R}(g_A(X))$ are explicitly given by

$$\sigma_K^2 = R(x_0)(1 - R(x_0)) \le 1/4,$$

$$\sigma_{(1)}^2 = \int\int \operatorname{sign}(h(y))\operatorname{sign}(h(z))[R(y \wedge z) - R(y)R(z)] R(dy) R(dz) \le 1/12,$$

$$\sigma_{(2)}^2 = \mu_{(2)}^{-2} \int\int h(y) h(z)[R(y \wedge z) - R(y)R(z)] R(dy) R(dz) \le \pi^{-2},$$

$$\sigma_D^2 = 1/12.$$

The upper bound for $\sigma_{(2)}^2$ is the largest eigenvalue of the symmetric kernel $K(y, z) := R(y \wedge z) - R(y)R(z)$ in $L_2(R)$; see Shorack and Wellner (1986), Section 5.2-3.

Now we can easily derive the limiting distribution of the estimators $\hat{\theta}_{n,A}$. Theorem 1 and Proposition 2 together imply the following theorem.

THEOREM 2.

(I) *In Model 1 the variable* $n(\hat{\theta}_{n,A} - \theta_n)$ *converges in distribution to*

$$\arg\max\left(2\mu_A W_A(z) - \mu_A^2 |z| : z \in \mathbb{Z}\right),$$

provided the latter is unique almost surely.

(II) *In Model 2 the variable* $n\gamma_n^{-2}(\hat{\theta}_{n,A} - \theta_n)$ *converges in distribution to*

$$\left(\sigma_A^2/\mu_A^2\right)\arg\max(W(s) - |s|/2 : s \in \mathbb{R}).$$

REMARK 1. In the case of Model 2 we utilized the fact that for arbitrary $\alpha > 0$ the process $(\alpha W(s): s \in \mathbb{R})$ is distributed as $(W(\alpha^2 s): s \in \mathbb{R})$. An explicit formula for the cdf of the limiting distribution is given by Siegmund (1986).

REMARK 2. One can quite easily show that the variable

$$\arg\max\left(2\mu_K W_K(z) - \mu_K^2 |z| : z \in \mathbb{Z}\right)$$

is not unique in general. For $A = (1), (2), D$ the variable $\arg\max(2\mu_A W_A(z) - \mu_A^2|z|: z \in \mathbb{Z})$ is unique almost surely, if the distributions of the single summands of $W_A(\cdot)$ are continuous.

(a) This is fulfilled for $A = D$, because $R(x) = \theta P(x) + (1 - \theta)Q(x)$ is monotone increasing, and for $x \le x'$ we have

$$R(x) = R(x') \quad \text{if and only if } (P + Q)[x, x'] = 0.$$

(b) The distributions of the summands of $W_A(\cdot)$ are continuous for $A = (1), (2)$, if the assumptions of Theorem 2 for $A = (1)$ hold: There are countably many open intervals (a_i, b_i) in $\mathbb{R}$ such that $\operatorname{sign}(h)$ is constant and nonzero on each interval (a_i, b_i), and the union of these intervals (a_i, b_i) has $(P + Q)$-measure 2. Then the functions $g_{(r)}(\cdot)$, $r = 1, 2$, are monotone on each subinterval (a_i, b_i) and for $a_i \le x \le x' \le b_i$ we have

$$g_{(r)}(x) = g_{(r)}(x') \quad \text{if and only if } (P + Q)[x, x'] = 0.$$

Consequently for arbitrary $w \in \mathbb{R}$ the set $\{y \in \mathbb{R}: g_{(r)}(y) = w\}$ is a union of countably many intervals with $(P + Q)$-measure 0.

In order to motivate the estimator $\hat{\theta}_{n,A}^{(B)}$ defined below, consider the parametric change-point model

$$\big(\tilde{W}(s) := 2\sigma\mu W(s) + \mu^2\tau - \mu^2|s - \tau| \colon s \in \mathbb{R}\big),$$

where $\sigma, \mu > 0$ are known constants and $\tau \in \mathbb{R}$ is an unknown parameter. Our estimators $\hat{\theta}_n$ correspond to the estimator

$$\hat{\tau} := \operatorname{arg\,max}\big(\tilde{W}(s) \colon s \in \mathbb{R}\big) =_{\mathscr{L}} (\sigma^2/\mu^2) \operatorname{arg\,max}\big(W(s) - |s|/2 \colon s \in \mathbb{R}\big).$$

It is known [see Ibragimov and Has'minskii (1981), Chapter 7] that the formal Bayes estimator

$$\hat{\tau}^{(B)} := \int s \exp\big(\tilde{W}(s)/(2\sigma^2)\big)\, ds \Big/ \int \exp\big(\tilde{W}(s)/(2\sigma^2)\big)\, ds$$

$$=_{\mathscr{L}} \tau + (\sigma^2/\mu^2) \int s \exp\big(W(s) - |s|/2\big)\, ds \Big/ \int \exp\big(W(s) - |s|/2\big)\, ds$$

is superior to $\hat{\tau}$ in that $\operatorname{Var}(\hat{\tau}^{(B)})$ is about 0.73 times $\operatorname{Var}(\hat{\tau})$. Thus, for example, one could consider

$$\hat{\theta}_n^{(B)} := \sum_{t \in T_n} t \exp\Big(nN_n\big(D_n^t\big)^2/(2\hat{\sigma}_n^2)\Big) \Big/ \sum_{t \in T_n} \exp\Big(nN_n\big(D_n^t\big)^2/(2\hat{\sigma}_n^2)\Big)$$

$$= \theta_n + \frac{\int_{(-\theta_n, 1-\theta_n)} s \exp\big(\tilde{W}_n(\theta_n + s)/(2\hat{\sigma}_n^2)\big)\, ds}{\int_{(-\theta_n, 1-\theta_n)} \exp\big(\tilde{W}_n(\theta_n + s)/(2\hat{\sigma}_n^2)\big)\, ds}$$

$$= \theta_n + \gamma_n^2 n^{-1} \frac{\int_{(-n\gamma_n^{-2}\theta_n,\, n\gamma_n^{-2}(1-\theta_n))} s \exp\big(\tilde{W}_n(\theta_n + \gamma_n^2 n^{-1}s)/(2\hat{\sigma}_n^2)\big)\, ds}{\int_{(-n\gamma_n^{-2}\theta_n,\, n\gamma_n^{-2}(1-\theta_n))} \exp\big(\tilde{W}_n(\theta_n + \gamma_n^2 n^{-1}s)/(2\hat{\sigma}_n^2)\big)\, ds}$$

as an improved nonparametric estimator, where $\hat{\sigma}_n^2$ is an estimator for σ_A^2. In fact one can deduce from Propositions 1 and 2 the following result.

THEOREM 3. *Let $\hat{\sigma}^2_{n,A} = \hat{\sigma}^2_{n,A}(X_{1,n}, \dots, X_{n,n})$ be an estimator converging in probability to a certain constant $\phi^2_A > 0$.*

Then in Model 1 the variable $n(\hat{\theta}^{(B)}_{n,A} - \theta_n)$ converges in distribution to

$$\sum_{z \in \mathbb{Z}} z \exp\left(\overline{W}_A(z)\right) \Big/ \sum_{z \in \mathbb{Z}} \exp\left(\overline{W}_A(z)\right),$$

where $\overline{W}_A(z) := \mu_A \phi_A^{-2} W_A(z) - \mu_A^2 \phi_A^{-2}|z|/2$.

In Model 2 the variable $n\gamma_n^{-2}(\hat{\theta}^{(B)}_{n,A} - \theta_n)$ converges in distribution to

$$\left(\phi_A^2/\mu_A^2\right) \int s \exp\left(\overline{W}(s)\right) ds \Big/ \int \exp\left(\overline{W}(s)\right) ds,$$

where $\overline{W}(s) := (\sigma_A/\phi_A)W(s) - |s|/2$.

We do not give a rigorous proof of Theorem 3. Just note that Proposition 1 implies that on the event $A_n(d)$:

$$\begin{aligned}\tilde{W}_n(t) &= n\left(N_n(D_n^t) + N_n(D_n^{\theta_n})\right)\left(N_n(D_n^t) - N_n(D_n^{\theta_n})\right) \\ &\leq -C_1^2 n\gamma_n^{-2}|t - \theta_n|, \qquad \forall\, t \in T_n \setminus T_n(d).\end{aligned}$$

Hence one may approximate $\hat{\theta}_n^{(B)}$ by

$$\theta_n + \gamma_n^2 n^{-1} \frac{\int_{(-d,d)} s \exp\left(\tilde{W}_n(\theta_n + \gamma_n^2 n^{-1}s)/(2\hat{\sigma}_n^2)\right) ds}{\int_{(-d,d)} \exp\left(\tilde{W}_n(\theta_n + \gamma_n^2 n^{-1}s)/(2\hat{\sigma}_n^2)\right) ds}$$

with a large fixed $d > 0$ and then apply Proposition 2.

Generally we have $\sigma_D^2 = 1/12$, and even $\sigma_{(1)}^2 = 1/12$, provided that h is strictly positive (negative) $(P + Q)$-almost everywhere. Thus one could take $\hat{\sigma}^2_{n,D} := 1/12$ and (under some circumstances) $\hat{\sigma}^2_{n,(1)} := 1/12$. Alternatively one could generally take the plug-in estimates

$$\hat{\sigma}^2_{n,A} := \mathrm{Var}_{X \sim \hat{R}_n}\left(\hat{g}_{n,A}(X)\right),$$

where $\hat{g}_{n,A}(x)$ is defined as $g_A(x)$ with h, R replaced by $P_n^{\hat{\theta}_n} - {}^{\hat{\theta}_n}P_n, \hat{R}_n$ respectively. This would lead to a two-step procedure where the estimator $\hat{\theta}_n$ is used to get an improved estimator $\hat{\theta}_n^{(B)}$. Note that $g_A(x)$ remains unchanged if h is replaced by αh, $\alpha > 0$. Moreover it follows from Lemma 2 that

$$\begin{aligned}\left\| {}^{\hat{\theta}_n}P_n - P_n \right\| &\leq O_p(n^{-1/2}) + \left\| {}^{\hat{\theta}_n}\Pi_n - P_n \right\| \\ &= O_p(n^{-1/2}) + \hat{\theta}_n^{-1}\left(\hat{\theta}_n - \theta_n\right)^+ \|Q_n - P_n\| \\ &= O_p(n^{-1/2}), \\ \left\| P_n^{\hat{\theta}_n} - Q_n \right\| &= O_p(n^{-1/2}),\end{aligned} \tag{3.1}$$

and this implies that

$$\left\| \gamma_n\left(P_n^{\hat{\theta}_n} - {}^{\hat{\theta}_n}P_n\right) - h \right\| \to 0 \quad \text{in probability}.$$

Using this result and the fact that $\|\hat{R}_n - R\| = o_p(1)$, one can deduce that the plug-in estimator for σ_A^2 is consistent.

4. Comparison with semiparametric estimators in a mean shift model. For simplicity assume that the variables $X_{i,n}$ are normally distributed with common variance σ^2 and known means 0 or γ_n^{-1} for $i \le n\theta_n$ or $i > n\theta_n$ respectively. Then the maximum likelihood estimator for θ_n may be represented as

$$\hat{\theta}_{n,H} := \arg\max\left(\tilde{W}_{n,H}(t): t \in T_n\right),$$

where

$$\tilde{W}_{n,H}(t) := \sum_{i=nt+1}^{n\theta_n} \sigma^{-2}(X_{i,n} - EX_{i,n}) - \sigma^{-2} n\gamma_n^{-2}|t - \theta_n|/2$$

[$\sum_{i=k+1}^{m}(\cdot) := -\sum_{i=m+1}^{k}(\cdot)$ for $m < k$]; see Hinkley (1970), who also considers the case of unknown variance and unknown means. The Bayes estimator with respect to the uniform prior on T_n is given by

$$\hat{\theta}_{n,H}^{(B)} := \sum_{t \in T_n} t \exp\left(\tilde{W}_{n,H}(t)\right) \Big/ \sum_{t \in T_n} \exp\left(\tilde{W}_{n,H}(t)\right).$$

Now one can easily compare the nonparametric estimators $\hat{\theta}_{n,A}$ and $\hat{\theta}_{n,A}^{(B)}$ with $\hat{\theta}_{n,H}$ and $\hat{\theta}_{n,H}^{(B)}$ respectively in the mean shift model given at the beginning of Section 3. We additionally assume that f is even with $f(0) > f(x)\ \forall\, x \ne 0$ and that $\sigma^2 := \mathrm{Var}_P(X) < +\infty$. Then $n\gamma_n^{-2}(\hat{\theta}_{n,H} - \theta_n)$ converges in distribution to

$$\sigma^2 \arg\max(W(s) - |s|/2: s \in \mathbb{R}),$$

and $n\gamma_n^{-2}(\hat{\theta}_{n,H}^{(B)} - \theta_n)$ converges in distribution to

$$\sigma^2 \int s \exp(W(s) - |s|/2)\, ds \Big/ \int \exp(W(s) - |s|/2)\, ds.$$

It follows from Theorems 2 and 3 that the corresponding nonparametric estimators have the same limiting distribution up to a scaling factor, namely:

$$E_A^{-1}\sigma^2 \arg\max(W(s) - |s|/2: s \in \mathbb{R})$$

and

$$E_A^{-1}\sigma^2 \int s \exp(W(s) - |s|/2)\, ds \Big/ \int \exp(W(s) - |s|/2)\, ds,$$

where $E_A := \sigma^2\sigma_A^{-2}\mu_A^2$. Here it is assumed that the estimators $\hat{\sigma}_{n,A}^2$ in the definition of $\hat{\theta}_{n,A}^{(B)}$ are consistent for σ_A^2. In this special model the quantities E_A are given by

$$E_K = 4\sigma^2 f(0)^2, \qquad E_{(1)} = E_D = 12\sigma^2 \int f(x)^2\, dx,$$

$$E_{(2)} = \sigma^2\left(\int f(x)^3\, dx\right)^2 (\mathrm{Var}_P[G(X)])^{-1}, \qquad G(x) := \int_{-\infty}^{x} f(y)^2\, dy.$$

TABLE 1
Relative asymptotic efficiencies

Distribution	Normal	Logistic	Double exponential
E_K	$2/\pi \approx 0.637$	$\pi^2/12 \approx 0.822$	2
$E_{(1)} = E_D$	$3/\pi \approx 0.955$	$\pi^2/9 \approx 1.097$	1.5
$E_{(2)}$	$* \approx 0.914$	$\pi^2/9 \approx 1.097$	$5/3 \approx 1.667$

$* = (3\pi[\pi^{-1} \arcsin((5/6)^{1/2}) - 1/4])^{-1}$.

They are a measure of relative asymptotic efficiency of the nonparametric estimators relative to the parametric ones and have been computed for the densities of the standard normal, logistic and double exponential distribution; see Table 1. For the calculations the reader is referred to Dümbgen (1990). Note that E_K is the Pitman efficiency of the sign test relative to the t test, and E_D is the Pitman efficiency of the Mann–Whitney test relative to the t test; see Lehmann (1975). It turns out that none of the maximum type estimators $\hat{\theta}_{n,A}$ is clearly superior to the other ones, and the same holds true for the formal Bayes estimators $\hat{\theta}_{n,A}^{(B)}$.

5. Bootstrap confidence sets. In Model 2 we have a parametric limiting distribution of the process $\tilde{W}_n(\cdot)$ depending only on the parameters μ_A and σ_A. In Model 1, however, the distribution of the discrete-time process $W_A(\cdot)$ depends on the (infinite-dimensional) parameter (θ, P, Q), and it cannot be approximated by a continuous-time Gaussian process. For that reason we do not use the asymptotic results for Model 2 for making inferences about θ. Alternatively, we propose a bootstrap procedure which works in both models: For every hypothetical change point $t \in T_n$, we test the hypothesis "$\theta_n = t$" on the nominal level α using the statistic

$$M_n(t) := \max_{s \in T_n} \tilde{W}_n(s,t),$$

where

$$\tilde{W}_n(s,t) := n\Big(N_n(D_n^s)^2 - N_n(D_n^t)^2\Big).$$

This test statistic corresponds to the likelihood ratio statistic in parametric models. For example, in exponential families one can compute the exact hypothetical distribution of the likelihood ratio statistic conditional on a sufficient statistic; see, for instance, Siegmund (1988). Here we estimate the hypothetical (unconditional) distribution $L_n(t)$ of $M_n(t)$ by

$$\hat{L}_n(t) := \mathscr{L}\big(M_n(t, X_1^*(t), X_2^*(t), \dots, X_n^*(t)) | X_{1,n}, X_{2,n}, \dots, X_{n,n}\big),$$

where $M_n(t, X_1^*(t), X_2^*(t), \dots, X_n^*(t))$ is computed as $M_n(t)$ using resampling variables $X_1^*(t), X_2^*(t), \dots, X_n^*(t)$. Given $(X_{1,n}, X_{2,n}, \dots, X_{n,n})$, these variables $X_i^*(t)$ are independent, having distribution ${}^tP_n$ for $i \le nt$ and P_n^t otherwise. Then we reject the hypothesis "$\theta_n = t$", if

$$\hat{L}_n(t)[M_n(t), +\infty) \le \alpha.$$

The confidence set $K_n(\alpha)$ consists of all $t \in T_n$, where the bootstrap test did not reject. This confidence set has asymptotic coverage probability not less than $1 - \alpha$ if

$$\limsup_{n \to \infty} \Pr\Big(\hat{L}_n(\theta_n)[M_n(\theta_n), +\infty) \le \alpha\Big) \le \alpha. \tag{5.1}$$

We shall see in Theorem 4 that $L_n(\theta_n)$ converges weakly to a certain distribution L_0 under the assumptions of Proposition 2. Then (5.1) holds, if $\hat{L}_n(\theta_n)$ also converges to L_0 in probability, $L_0\{0\} < 1 - \alpha$ and L_0 restricted to $(0, +\infty)$ is continuous. The convergence of $\hat{L}_n(\theta_n)$ follows from a more general result: Let (α_n) be an arbitrary sequence of positive numbers tending to 0. It follows from Lemma 2 that

$$\max_{|t-\theta_n| \le \alpha_n} \big\| {}^tP_n - P \big\| = o_p(1), \qquad \max_{|t-\theta_n| \le \alpha_n} \big\| P_n^t - Q \big\| = o_p(1),$$

$$\max_{|t-\theta_n| \le \alpha_n} \big\| \gamma_n\big(P_n^t - {}^tP_n\big) - h \big\| = o_p(1).$$

Consequently there is a sequence of events A_n (depending on $X_{1,n}, X_{2,n}, \dots, X_{n,n}$) such that $\Pr(A_n)$ tends to 1, and on these events A_n the pairs $({}^tP_n, P_n^t)$, $|t - \theta_n| \le \alpha_n$, as well as (P_n, Q_n) meet the requirements of the general model with P, Q, (γ_n) and h unchanged.

THEOREM 4. *Suppose that the assumptions of Proposition* 2 *hold. Then in Model* 1 *the variable* $M_{n,A}(\theta_n)$ *converges in distribution to*

$$\max\{2\mu_A W_A(z) - \mu_A^2 |z| \colon z \in \mathbb{Z}\}.$$

In Model 2, $M_{n,A}(\theta_n)$ *converges in distribution to*

$$2\sigma_A^2 \max\{W(s) - |s|/2 \colon s \in \mathbb{R}\}.$$

Theorem 4 follows essentially from Theorem 1 and Proposition 2. In Model 1 the limiting distribution $L_{0,A}$ is not continuous, because $\max\{2\mu_A W_A(z) - \mu_A^2|z| \colon z \in \mathbb{Z}\} = 0$ with positive probability. However, if the summands of $(W_A(z)\colon z \in \mathbb{Z})$ are continuously distributed, $L_{0,A}$ restricted to $(0, +\infty)$ is continuous; see also the remarks following Theorem 2. We end this section with a result for the size of the confidence sets $K_n(\alpha)$.

THEOREM 5. *Suppose that the assumptions of Proposition* 2 *hold. Then both in Models* 1 *and* 2 *we have for any* $\alpha \in (0,1)$:

$$\max_{t \in K_n(\alpha)} |t - \theta_n| = O_p\big(\gamma_n^2 n^{-1}\big).$$

6. The Nile data. As an example for applying bootstrap confidence sets, the Nile data have been analyzed; see also Cobb (1978) and Carlstein (1988).

TABLE 2
Bootstrap tests, "observed levels"

i	$s_{(1)}(i)$	$s_{(2)}(i)$	$s_D(i)$
1–23	0.000	0.000	0.000
24	0.001	0.001	0.001
25	0.011	0.006	0.007
26	0.069	0.057	0.060
27	0.197	0.190	0.189
28	1.000	1.000	1.000
29	0.081	0.080	0.082
30	0.035	0.021	0.026
31	0.020	0.008	0.014
32	0.004	0.001	0.002
33	0.005	0.001	0.002
34	0.003	0.001	0.002
35	0.001	0.000	0.001
36–99	0.000	0.000	0.000

The data are measurements of the annual volume of the Nile river at Assuan in the years 1871 to 1970. The change-point estimators $\hat{I}_A := 100\hat{\theta}_{100,A}$ applied to these 100 data yield $\hat{I}_A = 28$ (corresponding to the year 1898), where $A = (1), (2), D$. The "observed levels"

$$s_A(i) := \hat{L}_{100}(i/100)[M_{100,A}(i/100), +\infty),$$

have been estimated in 10,000 Monte Carlo simulations; see Table 2. For example 95% bootstrap confidence sets for I are given by

$$K_{100,(1)}(0.05) = K_{100,(2)}(0.05) = K_{100,D}(0.05) = [26, 29].$$

7. Proofs. At first we derive a maximal inequality for mixed empirical distributions.

LEMMA 1. *Let $\mathscr{F}$ be a family of measurable functions f: $E \to [-1, 1]$ such that the class $\mathscr{F}^*$ of all subgraphs $\{(x, t) \in E \times \mathbb{R}$: $0 \le t \le f(x)$ or $f(x) \le t \le 0\}$, $f \in \mathscr{F}$, is a VC class. Then there are constants $K_1, K_2 > 0$ depending only on the discriminating polynomial p of $\mathscr{F}^*$ such that for all $m \in \mathbb{N}$ and for arbitrary independent E-valued random variables $X_1, X_2, \dots, X_m$:*

$$\Pr\left(\left\| m^{-1/2} \sum_{i=1}^{m} \left[\delta_{X_i} - \mathscr{L}(X_i)\right]\right\|_{\mathscr{F}} \ge \eta\right) \le K_1 \exp(-K_2\eta^2), \quad \forall \eta > 0, \tag{7.1}$$

where $\|\nu\|_{\mathscr{F}} := \sup\{|\int f\,d\nu|: f \in \mathscr{F}\}$, $\nu \in \mathscr{M}$.

PROOF. Let m be an arbitrary fixed natural number. With the two symmetrizations described in 2.3 of Pollard (1984), one can show that for arbitrary

$\eta \geq 8^{1/2}$:

$$\Pr\left(\left\| m^{-1/2} \sum_{i=1}^{m} \left[\delta_{X_i} - \mathscr{L}(X_i)\right] \right\|_{\mathscr{F}} \geq \eta \right)$$
$$\leq 4 \Pr\left(\left\| m^{-1/2} \sum_{i=1}^{m} S_i \delta_{X_i} \right\|_{\mathscr{F}} \geq \eta/4 \right),$$

where $S_1, S_2, \ldots, S_m$ is a Rademacher sequence independent from $X_1, X_2, \ldots, X_m$. By conditioning on $X_1, X_2, \ldots, X_m$ one recognizes that it suffices to show

(7.2) There are constants $K_1, K_2, \eta_0 > 0$, depending only on the polynomial p, such that for arbitrary points $x_1, x_2, \ldots, x_m \in E$ and $P^* := m^{-1/2}\sum_{i=1}^{m} S_i \delta_{x_i}$:

$$\Pr(\|P^*\|_{\mathscr{F}} \geq \eta) \leq K_1 \exp(-K_2 \eta^2), \qquad \forall\, \eta \geq \eta_0.$$

Let ρ denote the L_2-norm of the probability measure $m^{-1}\sum_{i=1}^{m} \delta_{x_i}$. Then the covering numbers

$$N(u) := \min\left\{k \in \mathbb{N}: \exists\, f_1, f_2, \ldots, f_k \in \mathscr{F} \text{ s.t. } \min_{1 \leq i \leq k} \rho(f - f_i) \leq u \ \forall\, f \in \mathscr{F}\right\},$$

$0 < u \leq 1$, are bounded by Au^{-B}, where $A, B > 0$ are constants depending only on p; see 2.5 of Pollard (1984), Approximation Lemmas 25 and 36. In addition, the paths of the process $(\int f\, dP^*: f \in \mathscr{F})$ are continuous with respect to ρ. Thus Hoeffding's (1963) inequality and the Chaining Lemma [see 7.2 in Pollard (1984)] yield

$$\Pr\left(\exists\, f, f' \in \mathscr{F} \text{ s.t. } \left|\int (f - f')\, dP^*\right| > J(\varepsilon) \text{ and } \rho(f - f') \leq \varepsilon\right) \leq 2\varepsilon$$

for all $\varepsilon \in (0, 1)$, where $J(\varepsilon) := 26\int_0^{\varepsilon}[2\log(A^2/u^{2B+1})]^{1/2}\, du$. Consequently

$$\Pr(\|P^*\|_{\mathscr{F}} \geq \eta) \leq 2\varepsilon + 2A\varepsilon^{-B} \exp\left(-(\eta - J(\varepsilon))^2/2\right) \tag{7.3}$$
$$\forall\, \varepsilon \in (0, 1) \text{ with } 0 < J(\varepsilon) < \eta.$$

If we take $\varepsilon = \varepsilon(\eta) := \exp(-C\eta^2)$, where $0 < BC < 1/2$, then $J(\varepsilon) \to 0$ as $\eta \to \infty$, and there are $K_1, K_2, \eta_0 > 0$ such that

$$2\varepsilon + 2A\varepsilon^{-B} \exp\left(-(\eta - J(\varepsilon))^2/2\right) \leq K_1 \exp(-K_2 \eta^2), \qquad \forall\, \eta \geq \eta_0. \quad \square$$

We introduce some notation for quantities, which will play an important role in the sequel:

(a) The differences $D_n^t - \Delta_n^t$ are denoted by B_n^t.
(b) For $0 \leq s \leq t \leq 1$ we define

$$S_n(s, t) := \sum_{ns < i \leq nt} \left[\delta_{X_i} - \mathscr{L}(X_i)\right] \quad \text{and} \quad S_n(t, s) := -S_n(s, t).$$

Then we can write

$$B_n^t = [t/(1-t)]^{1/2} n^{-1} S_n(t,1) - [(1-t)/t]^{1/2} n^{-1} S_n(0,t),$$

$${}^tP_n - {}^t\Pi_n = (nt)^{-1} S_n(0,t), \quad P_n^t - \Pi_n^t = (n(1-t))^{-1} S_n(t,1).$$

The next lemma contains some auxiliary inequalities for $S_n(\cdot,\cdot)$ and $B_n^{(\cdot)}$, where we utilize Lemma 1 for the special case $\mathscr{F} = \{I_D: D \in \mathscr{D}\}$.

LEMMA 2.

For $n \geq 2$ and $0 \leq s \leq t \leq 1$,

$$\Pr\Big(\max_{s \leq u \leq t} \| S_n(s,u) \| \geq \eta [n(t-s)]^{1/2}\Big) \leq 2K_1 \exp(-K_2 \eta^2/4), \quad \forall\, \eta > 0. \tag{7.4}$$

There is a constant $K_3 > 0$ depending only on $\mathscr{D}$ such that

$$\max_{t \in T_n} \| B_n^t \| \leq K_3 (\log\log n)^{1/2} n^{-1/2} \tag{7.5}$$

with probability tending to 1.

PROOF. With a standard argument for processes with independent increments, one can show that for $0 \leq s \leq t \leq 1$,

$$\Pr\Big(\max_{s \leq u \leq t} \| S_n(s,u) \| \geq \varepsilon\Big) \leq 2 \max_{s \leq u \leq t} \Pr(\| S_n(u,t) \| \geq \varepsilon/2), \qquad \forall\, \varepsilon > 0.$$

Applying (7.1), we get (7.4). The global bound for $\|B_n^t\|$ may be proved as follows: One can quite easily show that

$$\Pr\Big(\max_{t \in T_n} \| (nt)^{-1/2} S_n(0,t) \| \geq \varepsilon\Big)$$
$$\leq \sum_{0 \leq j < \log_2(n)} \Big[\Pr(\| 2^{-j/2} S_n(0, 2^j/n) \| \geq \varepsilon/2)$$
$$+ \Pr\Big(\max_{1 \leq i \leq 2^j} \| 2^{-j/2} S_n(2^j/n, (2^j+i)/n) \| \geq \varepsilon/2\Big)\Big]$$

for all $\varepsilon > 0$. Then (7.1) and (7.4) lead to

$$\Pr\Big(\max_{t \in T_n} \| (nt)^{-1/2} S_n(0,t) \| \geq \varepsilon\Big) \leq 3K_1 (\log 2)^{-1} \log n \exp(-K_2 \varepsilon^2/16)$$

for all $\varepsilon > 0$. Similarly one can treat $\max_{t \in T_n} \|[n(1-t)]^{-1/2} S_n(t,1)\|$, and (7.5) follows. □

PROOF OF PROPOSITION 1. First of all (7.4) implies that

$$| N_n(D_n^{\theta_n}) - w(\theta_n) N_n(Q_n - P_n) | \leq \| B_n^{\theta_n} \| = O_p(n^{-1/2}).$$

Hence, according to (2.2), $N_n(D_n^{\theta_n}) \geq C_1 \gamma_n^{-1}$ with probability tending to 1 for arbitrary $C_1 \in (0, w(\theta) C_0)$. It is not very difficult to show that there is a constant $C > 0$ such that $\rho_n(\theta_n) - \rho_n(t) \geq |t - \theta_n| C$ for every $n \geq 2$ and for arbitrary $t \in T_n$. With this inequality one can deduce:

$$\text{There is a constant } C' > 0 \text{ such that}$$
$$\Pr\Big(N_n(D_n^t) - N_n(D_n^{\theta_n}) \leq \| B_n^t - B_n^{\theta_n} \| - |t - \theta_n| \gamma_n^{-1} C' \text{ for all } t \in T_n\Big) \to 1. \tag{7.6}$$

For the quantity $N_n(D_n^t)$ equals

$$\begin{aligned} N_n\big(B_n^t &- B_n^{\theta_n} + B_n^{\theta_n} + \rho_n(t) w(\theta_n)^{-1} \Delta_n^{\theta_n}\big) \\ &= N_n\big(B_n^t - B_n^{\theta_n} + (\rho_n(\theta_n) - \rho_n(t)) w(\theta_n)^{-1} B_n^{\theta_n} + \rho_n(t) w(\theta_n)^{-1} D_n^{\theta_n}\big) \\ &\leq \| B_n^t - B_n^{\theta_n} \| + (\rho_n(\theta_n) - \rho_n(t)) w(\theta_n)^{-1} \| B_n^{\theta_n} \| + \rho_n(t) w(\theta_n)^{-1} N_n\big(D_n^{\theta_n}\big). \end{aligned}$$

Consequently

$$\begin{aligned} N_n&(D_n^t) - N_n(D_n^{\theta_n}) \\ &\leq \| B_n^t - B_n^{\theta_n} \| - \gamma_n^{-1}(\rho_n(\theta_n) - \rho_n(t))\big(C_0 - 2w(\theta_n)^{-1} \gamma_n \| B_n^{\theta_n} \|\big) \\ &= \| B_n^t - B_n^{\theta_n} \| - \gamma_n^{-1}(\rho_n(\theta_n) - \rho_n(t))\big(C_0 - o_p(1)\big) \end{aligned}$$

for all $t \in T_n$ with probability tending to 1; see (2.2). Hence (7.6) is valid with $C' = CC_0'$ for arbitrary $C_0' \in (0, C_0)$.

The term $\|B_n^t - B_n^{\theta_n}\|$ in (7.6) can be globally bounded by $2K_3(\log\log n)^{1/2} n^{-1/2}$; see (7.5). Since $2K_3(\log\log n)^{1/2} n^{-1/2} = o(\gamma_n^{-1})$, (7.6) implies the following:

$$\begin{aligned} &\text{There is a constant } C'' > 0 \text{ such that} \\ &\Pr\big(N_n(D_n^t) - N_n(D_n^{\theta_n}) \leq -|t - \theta_n| \gamma_n^{-1} C'' \\ &\qquad\qquad \text{for all } t \in T_n \setminus [a, 1-a]\big) \to 1 \\ &\text{for arbitrary fixed } a \in (0, 1/2). \end{aligned} \tag{7.7}$$

Thus it suffices to consider $\|B_n^{(\cdot)} - B_n^{\theta_n}\|$ on compact subintervals of (0,1 1). The assertion of Proposition 1 follows, if we show:

$$\begin{aligned} &\text{For arbitrary constants } C''' > 0 \text{ and } a \in (0, 1/2) \\ &\liminf_{n \to \infty} \Pr\big(\| B_n^t - B_n^{\theta_n} \| \leq |t - \theta_n| \gamma_n^{-1} C''' \\ &\qquad\qquad \forall\, t \in T_n \cap [a, 1-a] \setminus T_n(d)\big) \to 1 \\ &\text{as } d \to +\infty. \end{aligned} \tag{7.8}$$

With the partial sums $S_n(\cdot, \cdot)$ one can write in detail

$$\begin{aligned} B_n^t - B_n^{\theta_n} = &\Big([t/(1-t)]^{1/2} - [\theta_n/(1-\theta_n)]^{1/2}\Big) n^{-1} S_n(\theta_n, 1) \\ &+ \Big([(1-\theta_n)/\theta_n]^{1/2} - [(1-t)/t]^{1/2}\Big) n^{-1} S_n(0, \theta_n) \\ &+ \big(w(t)^{-1} - w(\theta_n)^{-1}\big) n^{-1} S_n(t, \theta_n) \\ &+ w(\theta_n)^{-1} n^{-1} S_n(t, \theta_n). \end{aligned}$$

For arbitrary $a \in (0, 1/2)$ the functions $[t/(1-t)]^{1/2}$, $[(1-t)/t]^{1/2}$ and $w(t)^{-1}$ are Lipschitz continuous on $[a, 1-a]$. Hence there is a constant $C > 0$ such that for all $n \geq 2$ and for all $t \in T_n \cap [a, 1-a]$:

$$\left\| B_n^t - B_n^{\theta_n} - w(\theta_n)^{-1} n^{-1} S_n(t, \theta_n) \right\| \leq C|t - \theta_n| \max_{s \in [0,1]} \left\| n^{-1} S_n(s, \theta_n) \right\|.$$

With the maximal inequality (7.4) it follows that:

$$\text{For every } a \in (0, 1/2)$$
$$\left\| B_n^t - B_n^{\theta_n} - w(\theta_n)^{-1} n^{-1} S_n(t, \theta_n) \right\| \leq |t - \theta_n| O_p(n^{-1/2}) \tag{7.9}$$
$$\text{for all } t \in T_n \cap [a, 1-a],$$

where $O_p(n^{-1/2})$ denotes a random variable not depending on $t \in T_n$. Hence (7.8) would follow from:

$$\text{For arbitrary constants } C''' > 0$$
$$\limsup_{n \to \infty} \Pr\big(\| S_n(t, \theta_n) \| > n|t - \theta_n| \gamma_n^{-1} C''' \tag{7.10}$$
$$\text{for some } t \in T_n \setminus T_n(d) \big) \to 0$$
$$\text{as } d \to +\infty.$$

In order to simplify the notation, we fix $n \geq 2$ and consider independent random variables $Y_1, Y_2, Y_3, \ldots$ and $Y_1^*, Y_2^*, Y_3^*, \ldots$, where the Y_i have distribution P_n, and the Y_i^* have distribution Q_n. With these variables we define

$$Z_m := \sum_{i=1}^m (\delta_{Y_i} - P_n) \quad \text{and} \quad Z_m^* := \sum_{i=1}^m (\delta_{Y_i^*} - Q_n).$$

Then

$$\Pr\big(\| S_n(t, \theta_n) \| > n|t - \theta_n| \gamma_n^{-1} C''' \text{ for some } t \in T_n \setminus T_n(d) \big)$$
$$\leq \Pr\Big(\max_{m \geq d\gamma_n^2} m^{-1} \| Z_m \| \geq \gamma_n^{-1} C''' \Big) + \Pr\Big(\max_{m \geq d\gamma_n^2} m^{-1} \| Z_m^* \| \geq \gamma_n^{-1} C''' \Big).$$

It is known that the sequences $(m^{-1}\|Z_m\|: m \in \mathbb{N})$ and $(m^{-1}\|Z_m^*\|: m \in \mathbb{N})$ are reverse submartingales; see Pollard (1984), pages 21 and 22. Thus it follows from Chow's (1960) inequality that

$$\Pr\Big(\max_{m \geq d\gamma_n^2} m^{-1} \| Z_m \| \geq \gamma_n^{-1} C''' \Big) \leq \gamma_n C'''^{-1} E\big(m_0^{-1} \| Z_{m_0} \| \big),$$

where $m_0 := \min\{m \in \mathbb{N}: m \geq d\gamma_n^2\}$. The same inequality holds for Z_m^* in place of Z_m. According to (7.1), the right-hand side of the preceding inequality is bounded by

$$\gamma_n C'''^{-1} m_0^{-1/2} K_1 \int \exp(-K_2 x^2)\, dx \leq d^{-1/2} C'''^{-1} K_1 \int \exp(-K_2 x^2)\, dx,$$

which tends to 0 as $d \to +\infty$, and (7.10) follows. □

PROOF OF PROPOSITION 2. At first we derive a stochastic expansion for $\tilde{W}_n(\cdot)$: For arbitrary integers $d > 0$,

$$\tilde{W}_n(t) = \mu\big(N_n(2\gamma_n^{-1}S_n(t,\theta_n) - n\gamma_n^{-2}|t-\theta_n|h + H_n) \\ -N_n(H_n)\big) + r_n(t), \tag{7.11}$$

where

$$\mu := \Pr - \lim_{n\to\infty} \gamma_n N_n(Q_n - P_n), \qquad H_n := 2n\gamma_n^{-2}\theta(1-\theta)(h+m_n),$$

$$m_n \in D(\mathbb{R}), \qquad \|m_n\| = o_p(1) \quad \text{and} \quad \max_{t\in T_n(d)} |r_n(t)| = o_p(1).$$

PROOF OF (7.11). $\tilde{W}_n(t)$ may be written as

$$\tilde{W}_n(t) = 2nN_n(D_n^{\theta_n})\big(N_n(D_n^t) - N_n(D_n^{\theta_n})\big) - n\big(N_n(D_n^t) - N_n(D_n^{\theta_n})\big)^2.$$

Now we need some auxiliary inequalities for the difference $N_n(D_n^t) - N_n(D_n^{\theta_n})$, which equals $N_n(B_n^t - B_n^{\theta_n} + \Delta_n^t - \Delta_n^{\theta_n} + D_n^{\theta_n}) - N_n(D_n^{\theta_n})$. Since $\|B_n^{\theta_n}\| = O_p(n^{-1/2})$,

$$N_n(D_n^{\theta_n}) = \gamma_n^{-1}w(\theta_n)\mu + o_p(\gamma_n^{-1}) = O_p(\gamma_n^{-1}). \tag{7.12}$$

According to (7.9),

$$\max_{t\in T_n(d)} \big\|B_n^t - B_n^{\theta_n} - w(\theta_n)^{-1}n^{-1}S_n(t,\theta_n)\big\| = O_p(\gamma_n^2 n^{-3/2}), \tag{7.13a}$$

$$\max_{t\in T_n(d)} \|B_n^t - B_n^{\theta_n}\| = O_p(\gamma_n n^{-1}), \tag{7.13b}$$

where (7.13b) follows by applying (7.4) to $S_n(\cdot,\cdot)$ in (7.13a). Since $\Delta_n^t - \Delta_n^{\theta_n}$ is equal to $(\rho_n(t) - \rho_n(\theta_n))(Q_n - P_n)$, a simple Taylor expansion of ρ_n leads to

$$\max_{t\in T_n(d)} \big\|\Delta_n^t - \Delta_n^{\theta_n} + 2^{-1}w(\theta_n)^{-1}|t-\theta_n|(Q_n - P_n)\big\| = O_p(\gamma_n^3 n^{-2}), \tag{7.14a}$$

$$\max_{t\in T_n(d)} \|\Delta_n^t - \Delta_n^{\theta_n}\| = O_p(\gamma_n n^{-1}). \tag{7.14b}$$

Now one can easily deduce from these inequalities that

$$W_n(t) = \mu\big(N_n(2\gamma_n^{-1}S_n(t,\theta_n) - n\gamma_n^{-2}|t-\theta_n|h + H_n) \\ -N_n(H_n)\big) + r_n(t), \tag{7.15}$$

where $H_n := 2n\gamma_n^{-2}w(\theta_n)\gamma_n^{-1}D_n^{\theta_n}$ and $\max_{t\in T_n(d)}|r_n(t)| = o_p(1)$.

Finally the expansion (7.11) follows from

$$\big\|w(\theta_n)\gamma_n D_n^{\theta_n} - \theta(1-\theta)h\big\| = o_p(1).$$

So far we did not use particular properties of the different seminorms N_n. If we define

$$\nu_n(t) := 2\gamma_n^{-1}S_n(t,\theta_n) - n\gamma_n^{-2}|t-\theta_n|h \quad \text{and} \quad \beta_n := 2\theta(1-\theta)n\gamma_n^{-2},$$

we have

$$\max_{t \in T_n(d)} \|\nu_n(t)\| = O_p(1), \qquad \beta_n \to +\infty,$$

and (7.11) shows that $\tilde{W}_n(t)$ can be approximated locally in θ_n by

$$\beta_n \mu\left[N_n\left(\beta_n^{-1}\nu_n(t) + h + m_n\right) - N_n(h + m_n)\right].$$

Now we approximate this by a certain linear functional of $\nu_n(t)$ for $A = K, (1), (2), D$ separately.

(a) In the case $N_n = \|\cdot\|$, one can quite easily see that there is a sequence (ε_n) in $\mathbb{R}^+$, which converges to 0, such that

$$\|\beta_n^{-1}\nu_n(t) + h + m_n\| = \sup_{x \in M_n} \left|\left(\beta_n^{-1}\nu_n(t) + h + m_n\right)(x)\right| \quad \text{for all } t \in T_n(d), \tag{7.16}$$

with probability tending to 1 as $n \to \infty$, where $M_n := \{|h| > |h(x_0)| - \varepsilon_n\}$. The continuity of P and Q implies that

$$\sup_{\substack{t \in T_n(d), \\ x \in M_n}} \left|\gamma_n^{-1}S_n(t, \theta_n)(x) - \gamma_n^{-1}S_n(t, \theta_n)(x_0)\right| \to 0 \text{ in probability.} \tag{7.17}$$

This can be shown using a Skorohod imbedding for the empirical processes under consideration; see Dümbgen (1990) for a detailed proof. But then (7.11), (7.16) and (7.17) together yield

$$\tilde{W}_{n,K}(t) = \text{sign}(h(x_0))\mu_K 2\gamma_n^{-1}S_n(t, \theta_n) - \mu_K^2\gamma_n^{-2}n|t - \theta_n| + r_n(t), \tag{7.18$_K$}$$

where $\max_{t \in T_n(d)}|r_n(t)| = o_p(1)$.

(b) For $N_n = N_{n,1}$ we have

$$\begin{aligned} &N_{n,1}\left(\beta_n^{-1}\nu_n(t) + h + m_n\right) - N_{n,1}(h + m_n) \\ &\quad = \int\left(\left|\beta_n^{-1}\nu_n(t) + h + m_n\right| - |h + m_n|\right) d\hat{R}_n. \end{aligned}$$

If we replace this term by

$$\beta_n^{-1}\int \text{sign}(h)\nu_n(t)\, d\hat{R}_n,$$

the resulting error is bounded absolutely by

$$4\beta_n^{-1}\|\nu_n(t)\|\int I\left(|h| \le \|m_n\| + \beta_n^{-1}\|\nu_n(t)\|\right) d\hat{R}_n,$$

because

$$\big|\, |a + b + c| - |b + c| - \text{sign}(c)a \,\big| \le 4|a|I(|c| \le |a| + |b|)$$

for arbitrary $a, b, c \in \mathbb{R}$. Since $|h| > 0$ R-almost everywhere and $\hat{R}_n$ converges weakly to R in probability,

$$\int I(|h| \le \varepsilon_n)\, d\hat{R}_n = o_p(1)$$

whenever $\varepsilon_n \downarrow 0$. Consequently

$$(7.18_{(1)}) \qquad \begin{aligned} \tilde{W}_{n,(1)}(t) &= 2\mu_{(1)}\gamma_n^{-1} \int \operatorname{sign}(h) S_n(t,\theta_n)\, d\hat{R}_n \\ &\quad - \mu_{(1)}^2 \gamma_n^{-2} n|t-\theta_n| + r_n(t), \end{aligned}$$

where $\max_{t \in T_n(d)} |r_n(t)| = o_p(1)$.

(c) In the case $N_n = N_{n,2}$ we have $N_{n,2}(h+m_n)^2 = \mu_{(2)}^2 + o_p(1)$ and one easily sees that

$$\begin{aligned} &N_{n,2}\big(\beta_n^{-1}\nu_n(t) + h + m_n\big) \\ &\qquad = N_{n,2}(h+m_n) + \mu_{(2)}^{-1}\beta_n^{-1} \int \nu_n(t) h\, d\hat{R}_n + o_p\big(\beta_n^{-1}\|\nu_n(t)\|\big). \end{aligned}$$

Hence

$$(7.18_{(2)}) \quad \tilde{W}_{n,(2)}(t) = 2\gamma_n^{-1} \int S_n(t,\theta_n) h\, d\hat{R}_n - \mu_{(2)}^2 \gamma_n^{-2} n|t-\theta_n| + r_n(t),$$

where $\max_{t \in T_n(d)} |r_n(t)| = o_p(1)$.

(d) One easily shows that

$$N_{n,D}\big(\beta_n^{-1}\nu_n(t) + h + m_n\big) = \beta_n^{-1} \operatorname{sign}\Big(\int h\, dQ\Big) \int \nu_n(t)(x^*)\, \hat{R}_n(dx)$$

for all $t \in T_n(d)$ with probability tending to 1 as $n \to \infty$. Hence

$$(7.18_D) \qquad \begin{aligned} \tilde{W}_{n,D}(t) &= \mu_D \operatorname{sign}\Big(\int h\, dQ\Big) 2\gamma_n^{-1} \int S_n(t,\theta_n)(x^*)\, \hat{R}_n(dx) \\ &\quad - \mu_D^2 \gamma_n^{-2} n|t-\theta_n| + r_n(t), \end{aligned}$$

where $\max_{t \in T_n(d)} |r_n(t)| = o_p(1)$.

Starting from (7.18_A), we prove Proposition 2:

(a) In the case $A = K$ Proposition 2 is just an easy consequence of (7.18_K), where in Model 2 we use Donsker's invariance principle for sums of iid rv's.

(b) The case $A = D$ is also quite simple: The term

$$\gamma_n^{-1} \int S_n(t,\theta_n)(x^*) h(x)\, \hat{R}_n(dx) = -\gamma_n^{-1} \int \hat{R}_n(x^*) h(x)\, S_n(t,\theta_n)(dx)$$

in (7.18_D) can be approximated by

$$-\gamma_n^{-1} \int R_n(x^*) h(x)\, S_n(t,\theta_n)(dx)$$

because $\|\hat{R}_n - R_n\| = O_p(n^{-1/2})$ and the measure $S_n(t,\theta_n)$ has total variation not greater than $2D\gamma_n^2$ for all $t \in T_n(d)$. The latter integral is equal to

$$\gamma_n^{-1} \sum_{i=nt+1}^{n\theta_n} [R_n(X_{i,n}{}^*) - ER_n(X_{i,n}{}^*)].$$

One can show quite easily that $\mathscr{L}_{P_n}[R_n(X^*)]$ and $\mathscr{L}_{Q_n}[R_n(X^*)]$ converge weakly to $\mathscr{L}_P[R(X)]$ and $\mathscr{L}_Q[R(X)]$ respectively. Especially the boundedness of R_n and R imply that in Model 2 both $\mathrm{Var}_{P_n}[R_n(X^*)]$ and $\mathrm{Var}_{Q_n}[R_n(X^*)]$ tend to $\mathrm{Var}_P[P(X)]$. Hence Proposition 2 follows for $A = D$.

(c) For $A = (1)$ and (2) we have to work a bit harder. The arguments in both cases are very similar, so we only consider $A = (1)$. The integral

$$\gamma_n^{-1} \int \operatorname{sign}(h) S_n(t, \theta_n)\, d\hat{R}_n$$

in $(7.18_{(1)})$ is equal to

$$\gamma_n^{-1} \sum_{i=nt+1}^{n\theta_n} \left[\hat{g}_{n,(1)}(X_{i,n}^*) - \int \hat{g}_{n,(1)}\, d\mathscr{L}(X_{i,n})\right],$$

where $\hat{g}_{n,(1)}(x) := \int \operatorname{sign}(h) I(x \le \cdot)\, d\hat{R}_n$. With

$$g_{n,(1)}(x) := \int \operatorname{sign}(h) I(x \le \cdot)\, dR_n,$$

one can deduce from Lemma 1 that

$$\|\hat{g}_{n,(1)} - g_{n,(1)}\| = O_p(n^{-1/2}), \tag{7.19}$$

where $\mathscr{F} := \{\operatorname{sign}(h) I(x \le \cdot) \colon x \in \mathbb{R}\}$. Thus $\hat{g}_{n,(1)}$ may be approximated by $g_{n,(1)}$. Finally one can show quite easily that $\mathscr{L}_{P_n}[g_{n,(1)}(X)]$ and $\mathscr{L}_{Q_n}[g_{n,(1)}(X)]$ converge weakly to $\mathscr{L}_P[g_{(1)}(X)]$ and $\mathscr{L}_Q[g_{(1)}(X)]$ respectively. In Model 2 the corresponding variances converge to $\mathrm{Var}_P[g_{(1)}(X)]$. Hence we obtain Proposition 2 for $A = (1)$. □

Proof of Theorem 5. For $t \in T_n$ define

$$\hat{q}_n(t) := \max\{r \in \mathbb{R} \colon \hat{L}_n(t) \ge \alpha\}.$$

Then $t \in K_n(\alpha)$ implies $M_n(t) \le \hat{q}_n(t)$. The variable $M_n(t)$ itself is not less than $-W_n(t) = -n[N_n(D_n^t) - N_n(D_n^{\theta_n})][N_n(D_n^t) + N_n(D_n^{\theta_n})]$. Since $N_n(D_n^t) + N_n(D_n^{\theta_n}) \ge N_n(D_n^{\theta_n}) = \gamma_n^{-1}[w(\theta_n)\mu + o_p(1)]$, Proposition 1 yields

$$\liminf_{n \to \infty} \Pr\left(n\gamma_n^{-2}|t - \theta_n| \le C_2 \hat{q}_n(t)\ \forall\, t \in K_n(\alpha)/T_n(d)\right) \to 1 \tag{7.20}$$

as $d \to +\infty$, where $C_2 > 0$ is some suitable constant. Now we first need a global bound for $\hat{q}_n(t)$:

$$\begin{aligned} &\text{There is a constant } C' \text{ such that}\\ &\hat{q}_n(t) \le C' n^{1/2} (\log\log n)^{1/2} \gamma_n^{-1} \quad \text{for all } t \in T_n\\ &\text{with probability tending to 1.} \end{aligned} \tag{7.21}$$

Accept this for the moment. Then it follows from (7.20) and (7.21) that

$$\max_{t \in K_n(\alpha)} |t - \theta_n| = O_p\big(\gamma_n (\log\log n)^{1/2} n^{-1/2}\big).$$

Hence we only have to consider $|t - \theta_n| \le \alpha_n$ in (7.20) with a suitable sequence (α_n) tending to 0. As mentioned in Section 5, $\hat{L}_n(t)$ converges weakly to L_0 uniformly for $|t - \theta_n| \le \alpha_n$ (in probability). Hence

$$\max_{|t-\theta_n| \le \alpha_n} |\hat{q}_n(t)| = O_p(1),$$

and Theorem 5 follows.

Proof of (7.21). Remember that $\hat{L}_n(t)$ is equal to the distribution of $M_n(\theta_n)$, where θ_n, P_n, Q_n replaced by $t, {}^tP_n, P_n^t$ respectively. Let $\tau_n := \max_{s \in T_n} \|B_n^s\|$. Then

$$\begin{aligned}\tilde{W}_n(s) &\le n\Big[\big(N_n(\Delta_n^s) + \tau_n\big)^2 - \big(N_n(\Delta_n^{\theta_n}) - N_n(B_n^{\theta_n})\big)^2\Big] \\ &\le 4n\tau_n N_n\big(\Delta_n^{\theta_n}\big) + n\tau_n^2.\end{aligned}$$

It follows from (7.5) that τ_n is not greater than $K_3(\log\log n)^{1/2} n^{-1/2}$ with probability tending to 1. Since the proof of (7.5) did not depend on the special sequences $(\theta_n), (P_n), (Q_n)$, there is an integer n_0 such that

$$\hat{q}_n(t) \le 4K_3 n^{1/2} (\log\log n)^{1/2} \|D_n^t\| + K_3^2 \log\log n \quad \text{for all } t \in T_n \text{ and } n \ge n_0.$$

But

$$\|D_n^t\| \le \|Q_n - P_n\| + K_3 n^{-1/2} (\log\log n)^{1/2} = O\big(\gamma_n^{-1}\big) + K_3 n^{-1/2} (\log\log n)^{1/2}$$

for all $t \in T_n$ with probability tending to 1. Hence (7.21) follows together with $\log\log n = o(n^{1/2}(\log\log n)^{1/2}\gamma_n^{-1})$. □

Acknowledgments. I want to thank Werner Ehm, who led my attention to Carlstein's paper and asked for rates of consistency, and I sincerely thank my Ph.D advisor Professor D. W. Müller for his encouragement and advice. I am also grateful to the referees, especially that they proposed considering improved (formal Bayes) estimators.

REFERENCES

Beran, R. and Millar, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.* **14** 431–443.

Carlstein, E. (1988). Nonparametric estimation of a change-point. *Ann. Statist.* **16** 188–197.

Chow, Y. S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* **11** 107–111.

Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika* **65** 243–251.

Csörgő, M. and Horvath, L. (1988). Invariance principles for changepoint problems. *J. Multivariate Anal.* **27** 151–168.

DARKHOVSKH, B. S. (1976). A non-parametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory Probab. Appl.* **21** 178–183.

DÜMBGEN, L. (1990). The asymptotic behavior of some nonparametric changepoint estimators. Dissertation, Univ. Heidelberg.

HINKLEY, D. V. (1970). Inference about the changepoint in a sequence of random variables. *Biometrika* **57** 1–17.

HINKLEY, D. V. and HINKLEY, E. A. (1970). Inference about the changepoint in a sequence of binomial variables. *Biometrika* **57** 477–488.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation, Asymptotic Theory.* Springer, New York.

LEHMANN, E. L. (1975). *Nonparametrics.* Holden-Day, San Francisco.

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.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14** 361–404.

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Internat. Statist. Rev.* **56** 31–48.

WOLFE, D. A. and SCHECHTMAN, E. (1984). Nonparametric statistical procedures for the changepoint problem. *J. Statist. Plann. Inference* **9** 389–396.

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
367 EVANS HALL
BERKELEY, CALIFORNIA 94720