doi:10.5705/ss.202015.0435

Subsampling for General Statistics under Long Range Dependence with application to change point analysis

Annika Betken and Martin Wendler

Ruhr-Universität Bochum and Universität Greifswald

Supplementary Material

This document contains additional material. Generalizations of the robust, self-normalized change point test by Betken (2016) for data with ties or with multiple change points are presented in Section and S1 amd S2. The technical lemmas given in Section S3 are needed for the proof of Theorem 1 in Section S4.

S1 A Modified Change Point Test for Data with Ties

If the distribution of $X_i = G(\xi_i)$ is not continuous, there is a positive probability that $X_i = X_j$ for some $i \neq j$, so there might be ties in the sample. We propose to use the following test statistic based on the modified ranks $\tilde{R}_i = \sum_{j=1}^{n}(1_{\{X_j<X_i\}} + \frac{1}{2}1_{\{X_j=X_i\}})$:

$$\tilde{T}_n(\tau_1, \tau_2) := \max_{k\in\{\lfloor n\tau_1\rfloor,\ldots,\lfloor n\tau_2\rfloor\}} \frac{\left|\sum_{i=1}^{k}\tilde{R}_i - \frac{k}{n}\sum_{i=1}^{n}\tilde{R}_i\right|}{\left\{\frac{1}{n}\sum_{t=1}^{k}\tilde{S}_t^2(1,k) + \frac{1}{n}\sum_{t=k+1}^{n}\tilde{S}_t^2(k+1,n)\right\}^{1/2}},$$

where

$$\tilde{S}_t(j,k) = \sum_{h=j}^{t}\left(\tilde{R}_h - \frac{1}{k-j+1}\sum_{i=j}^{k}\tilde{R}_i\right).$$

To be able to apply subsampling, we need $\tilde{T}_n$ to converge in distribution, which we will show now:

Lemma 1. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a stationary sequence of centered standard Gaussian variables with covariance function $\gamma(k) = k^{-D}L_\gamma(k)$ for a $D \in (0,1)$ and a slowly varying function L_γ. Let $X_i = G(\xi_i)$ for a function G, piecewise monotone on finitely many pieces. Then $\tilde{T}_n(\tau_1,\tau_2) \Rightarrow T$ for some random variable T.*

Proof. Let $h(x,y) = 1_{\{G(x)<G(y)\}} + \frac{1}{2}1_{\{G(x)=G(y)\}} - \frac{1}{2}$. We define the modified Wilcoxon process $(\tilde{W}_n(\lambda))_{\lambda\in[0,1]}$ by

$$\tilde{W}_n(\lambda) := \frac{1}{nd_n}\sum_{i=1}^{[n\lambda]}\sum_{j=[n\lambda]+1}^{n} h(\xi_i,\xi_j)$$

with $d_n = \sqrt{\mathrm{Var}(\sum_{i=1}^n \xi_i)}$. From Theorem 2.2 in Dehling, Rooch and Wendler (2017), we have the weak convergence of this process $\tilde{W}_n$ to the limit process W with

$$W(\lambda) = -(1-\lambda)Z(\lambda)\int\varphi(x)d\tilde{h}(x) - \lambda(Z(1)-Z(\lambda))\int\Big(\int\varphi(y)dh(x,y)(y)\Big)\varphi(x)dx.$$

Here, Z is a fractional Brownian motion, φ is the density function of the standard normal distribution and $\tilde{h}(x) = E[h(x,\xi_i)]$. Following the proof of Theorem 1 in Betken (2016), we can express $\tilde{T}_n$ as a function of $\tilde{W}_n$:

$$T_n(\tau_1,\tau_2) = \sup_{\tau_1\le\lambda\le\tau_2}\frac{\left|\tilde{W}_n(\lambda)\right|}{\left\{\int_0^\lambda(\tilde{W}_n(t)-\frac{c_n(t)}{c_n(\lambda)}\tilde{W}_n(\lambda))^2dt+\int_\lambda^1(\tilde{W}_n(t)-\frac{1-c_n(t)}{1-c_n(\lambda)}\tilde{W}_n(\lambda))^2dt\right\}^{1/2}}.$$

Note that $c_n(\lambda)$ converges to λ uniformly, so we have the asymptotic equivalence

$$T_n(\tau_1,\tau_2) \approx \sup_{\tau_1\le\lambda\le\tau_2}\frac{\left|\tilde{W}_n(\lambda)\right|}{\left\{\int_0^\lambda(\tilde{W}_n(t)-\frac{t}{\lambda}\tilde{W}_n(\lambda))^2dt+\int_\lambda^1(\tilde{W}_n(t)-\frac{1-t}{1-\lambda}\tilde{W}_n(\lambda))^2dt\right\}^{1/2}}.$$

By the continuous mapping theorem, we get

$$
\begin{aligned}
&T_n(\tau_1, \tau_2) \\
&\Rightarrow \sup_{\tau_1 \le \lambda \le \tau_2} \frac{|W(\lambda)|}{\left\{ \int_0^\lambda (W(t) - \frac{t}{\lambda} W(\lambda))^2 dt + \int_\lambda^1 (W(t) - \frac{1-t}{1-\lambda} W(\lambda))^2 dt \right\}^{1/2}} =: T.
\end{aligned}
$$

□

S2 A Test for Multiple Change Points

For testing the alternative hypothesis of two change points, we suggest to use the test statistic $T_n(\tau_1, \tau_2, \varepsilon) = \sup_{(k_1,k_2) \in \Omega_n(\tau_1,\tau_2,\varepsilon)} |G_n(k_1, k_2)|$. Some calculations yield

$$
\begin{aligned}
&G_n(k_1, k_2) \\
&= \frac{\left|\tilde{W}_n(\lambda_1,\lambda_2)\right|}{\left\{ \int\limits_0^{\lambda_1} \left(\tilde{W}_n(r,\lambda_2) - \frac{r}{\lambda_1} \tilde{W}_n(\lambda_1,\lambda_2)\right)^2 dr + \int\limits_{\lambda_1}^{\lambda_2} \left(\tilde{W}_n(r,\lambda_2) - \frac{\lambda_2 - r}{\lambda_2 - \lambda_1} \tilde{W}_n(\lambda_1,\lambda_2)\right)^2 dr \right\}^{\frac{1}{2}}} \\
&+ \frac{|W_n^*(\lambda_2,\lambda_1)|}{\left\{ \int\limits_{\lambda_1}^{\lambda_2} \left(W_n^*(r,\lambda_1) - \frac{r-\lambda_1}{\lambda_2-\lambda_1} W_n^*(\lambda_2,\lambda_1)\right)^2 dr + \int\limits_{\lambda_2}^{1} \left(W_n^*(r,\lambda_1) - \frac{1-r}{1-\lambda_2} W_n^*(\lambda_2,\lambda_1)\right)^2 dr \right\}^{\frac{1}{2}}} \\
&\qquad + o_P(1),
\end{aligned}
$$

where

$$
\tilde{W}_n(\lambda, \tau) := W_n(\lambda, \lambda) - W_n(\lambda, \tau), \quad W_n^*(\lambda, \tau) := W_n(\lambda, \lambda) - W_n(\tau, \lambda)
$$

with

$$
W_n(\lambda, \tau) = \sum_{i=1}^{\lfloor n\lambda \rfloor} \sum_{j=\lfloor n\tau \rfloor + 1}^{n} \left(1_{\{X_i \le X_j\}} - \frac{1}{2} \right), \ 0 \le \lambda \le \tau \le 1.
$$

Define $d_n^2 := \mathrm{Var}(\sum_{j=1}^n H_r(\xi_j))$, where H_r denotes the r-th order Hermite polynomial and r designates the Hermite rank of the class of functions $\left\{1_{\{G(\xi_i)\leq x\}} - F(x),\ x \in \mathbb{R}\right\}$. It can be shown that $\frac{1}{nd_n}W_n(\lambda,\tau)$ converges in distribution to

$$\{(1-\tau)Z_r(\lambda) - \lambda(Z_r(1) - Z_r(\tau))\}\frac{1}{r!}\int J_r(x)dF(x),\ 0 \leq \lambda \leq \tau \leq 1,$$

where Z_r is an r-th order Hermite process with Hurst parameter $H := \max\{1 - \frac{rD}{2}, \frac{1}{2}\}$ and where $J_r(x) = \mathrm{E}\left(H_r(\xi_i)1_{\{G(\xi_i)\leq x\}}\right)$.

As a result, under the hypothesis the limiting distribution of $T_n(\tau_1,\tau_2,\varepsilon)$ is given by $T(r,\tau_1,\tau_2,\varepsilon) = \sup_{\tau_1\leq\lambda<\tau\leq\tau_2,\ \tau-\lambda\geq\varepsilon} G_r(\lambda,\tau)$ with

$$\begin{aligned}
&G_r(\lambda,\tau)\\
&= \frac{\left|Z_r(\lambda) - \frac{\lambda}{\tau}Z_r(\tau)\right|}{\left\{\int\limits_0^\lambda \left(Z_r(t) - \frac{t}{\lambda}Z_r(\lambda)\right)^2 dt + \int\limits_\lambda^\tau \left(Z_r(t) - \frac{t-\lambda}{\tau-\lambda}Z_r(\tau) - \frac{\tau-t}{\tau-\lambda}Z_r(\lambda)\right)^2 dt\right\}^{\frac{1}{2}}}\\
&+ \frac{\left|Z_r(\tau) - \frac{1-\tau}{1-\lambda}Z_r(\lambda) - \frac{\tau-\lambda}{1-\lambda}Z_r(1)\right|}{\left\{\int\limits_\lambda^\tau \left(Z_r(t) - \frac{\tau-t}{\tau-\lambda}Z_r(\lambda) - \frac{t-\lambda}{\tau-\lambda}Z_r(\tau)\right)^2 dt + \int\limits_\tau^1 \left(Z_r(t) - \frac{1-t}{1-\tau}Z_r(\tau) - \frac{t-\tau}{1-\tau}Z_r(1)\right)^2 dt\right\}^{\frac{1}{2}}}.
\end{aligned}$$

S3 Auxiliary Results

Assumption 2. *$X_n = G(\xi_n)$ for a measurable function G and a stationary, Gaussian process $(\xi_n)_{n\in\mathbb{N}}$ with covariance function*

$$\gamma(k) := \mathrm{Cov}(\xi_1, \xi_{1+k}) = k^{-D}L_\gamma(k)$$

such that the following conditions hold:

1. *$D \in (0,1]$ and L_γ is a slowly varying function with*

$$\max_{\tilde{k}\in\{k+1,\ldots,k+2l'-1\}}\left|L_\gamma(k) - L_\gamma(\tilde{k})\right| \leq K\frac{l'}{k}\min\{L_\gamma(k), 1\}$$

for a constant $K < \infty$ *and all* $l' \in \{l_k, \ldots, k\}$.

2. $(\xi_n)_{n\in\mathbb{N}}$ *has a spectral density* f *with* $f(x) = |x|^{D-1}L_f(x)$ *for a slowly varying function* L_f *bounded away from* 0 *on* $[0, \pi]$ *such that* $\lim_{x\to 0} L_f(x) \in (0, \infty]$ *exists.*

Lemma 2. *Under Assumption 2, there is a constant* $K_D < \infty$*, such that for all* $x_1, \ldots, x_l \in \mathbb{R}$ *with* $\mathrm{Var}(\sum_{i=1}^{l} x_i\xi_i) = 1$

$$\sum_{i=1}^{l} x_i^2 \le K_D.$$

Proof. Recall that we can rewrite the covariances as

$$\gamma(k) = \int_{-\pi}^{\pi} \mathrm{e}^{ik\lambda} f(\lambda)d\lambda$$

and that the spectral density f can be written as $f(\lambda) = L_f(|\lambda|)|\lambda|^{D-1}$. By our assumptions $L_f(x) \ge C_{\min}$ for a constant $C_{\min} > 0$, so that we can conclude that

$$\begin{aligned}
1 =&\mathrm{Var}\Big(\sum_{i=1}^{l} x_i\xi_i\Big) = \sum_{1\le j,k\le l} x_j x_k \gamma(j-k)\\
=&\sum_{1\le j,k\le l} x_j x_k \textstyle\int_{-\pi}^{\pi} \mathrm{e}^{i(j-k)\lambda} f(\lambda)d\lambda = \sum_{1\le j,k\le l} x_j x_k \int_{-\pi}^{\pi} \mathrm{e}^{i(j-k)\lambda} L_f(|\lambda|)|\lambda|^{D-1}d\lambda\\
=&2\textstyle\int_{0}^{\pi} \sum_{1\le j,k\le l} x_j x_k \mathrm{e}^{i(j-k)\lambda} L_f(\lambda)\lambda^{D-1}d\lambda = 2\int_{0}^{\pi}\Big|\sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda}\Big|^2 L_f(\lambda)\lambda^{D-1}d\lambda\\
\ge&2C_{\min}\pi^{D-1}\int_0^{\pi}\Big|\sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda}\Big|^2 d\lambda.
\end{aligned}$$

We rewrite the integrand as

$$\Big|\sum_{j=1}^{l} x_j \mathrm{e}^{-ij\lambda}\Big|^2 = \sum_{1\le j,k\le l} x_j x_k \mathrm{e}^{-ij\lambda}\mathrm{e}^{ik\lambda} = \sum_{j=1}^{l} x_j^2 + \sum_{j\ne k} x_j x_k \mathrm{e}^{-i(j-k)\lambda}$$

$$
\begin{aligned}
&= \sum_{j=1}^{l} x_j^2 + \sum_{j<k} x_j x_k \left(\mathrm{e}^{-i(j-k)\lambda} + \mathrm{e}^{-i(k-j)\lambda}\right) \\
&= \sum_{j=1}^{l} x_j^2 + 2\sum_{j<k} x_j x_k \cos((k-j)\lambda) \\
&= \sum_{1\leq j,k\leq l} x_j x_k \cos((k-j)\lambda).
\end{aligned}
$$

As a result, we have

$$
\begin{aligned}
\int_0^{\pi} \left| \sum_{j=1}^{l} x_j \,\mathrm{e}^{-ij\lambda} \right|^2 d\lambda &= \int_0^{\pi} \sum_{1\leq j,k\leq l} x_j x_k \cos((k-j)\lambda) d\lambda \\
&= \sum_{1\leq j,k\leq l} x_j x_k \int_0^{\pi} \cos((k-j)\lambda) d\lambda \\
&= \sum_{j=1}^{l} x_j^2 \int_0^{\pi} \cos(0) d\lambda + \sum_{j\neq k} x_j x_k \int_0^{\pi} \cos((k-j)\lambda) d\lambda \\
&= \pi \sum_{j=1}^{l} x_j^2.
\end{aligned}
$$

All in all, this yields

$$
1 = \mathrm{Var}\left(\sum_{i=1}^{l} x_i \xi_i\right) \geq 2C_{\min}\pi^{D-1} \int_0^{\pi} \left| \sum_{j=1}^{l} x_j \,\mathrm{e}^{-ij\lambda} \right|^2 d\lambda = 2C_{\min}\pi^{D} \sum_{j=1}^{l} x_j^2.
$$

Therefore, the statement of the lemma holds with $K_D = 1/(2C_{\min}\pi^D)$.

□

Lemma 3. *Under Assumption 2, there are constants $K'_D < \infty$ and $l_0 \in \mathbb{N}$ such that*

$$
\left| \sum_{i=1}^{l} x_i \right| \leq K'_D l^{D/2}
$$

for all $l \geq l_0$ and $x_1, \ldots, x_l \in \mathbb{R}$ with $\mathrm{Var}\left(\sum_{i=1}^{l} x_i \xi_i\right) = 1$.

Proof. The statement of the proof is equivalent to the existence of a constant $C > 0$, such that for all $x_1, \ldots, x_l \in \mathbb{R}$ with $\sum_{i=1}^{l} x_i = 1$, we have

$$\operatorname{Var}\left(\sum_{i=1}^{l} x_i \xi_i\right) \geq C l^{-D}.$$

Let $x_1^\star, \ldots, x_l^\star \in \mathbb{R}$ with $\sum_{i=1}^{l} x_i^\star = 1$ be the values that minimize $\operatorname{Var}\left(\sum_{i=1}^{l} x_i^\star \xi_i\right)$. Then $\hat{\mu}_\xi(\xi_1, \ldots, \xi_n) := \sum_{i=1}^{l} x_i^\star \xi_i$ is the best linear unbiased estimator for $\mu := E(\xi_1)$. For a process $(\zeta_n)_{n \in \mathbb{N}}$ with spectral density $f_\zeta(x) = \frac{1}{2\pi} |1 - e^{ix}|^{D-1}$, we have

$$\operatorname{Var}\left(\hat{\mu}_\zeta(\zeta_1, \ldots, \zeta_n)\right) \geq C_1 l^{-D}$$

for a constant $C_1 > 0$ by a Corollary of Theorem 5.1 in Adenstedt (1974). We rewrite the spectral density f_ζ of $(\zeta_n)_{n \in \mathbb{N}}$ with the help of the spectral density f of $(\xi_n)_{n \in \mathbb{N}}$ as

$$f_\zeta(x) = f(x) \frac{\left|1 - e^{ix}\right|^{D-1}}{2\pi |x|^{D-1} L_f(x)}.$$

Note that the function g with $g(x) = \frac{\left|1 - e^{ix}\right|^{D-1}}{2\pi |x|^{D-1} L_f(x)}$ is bounded, as we assumed that L_f is bounded away from 0. Hence, we have

$$\operatorname{Var}\left(\hat{\mu}_\xi(\xi_1, \ldots, \xi_n)\right) \geq \frac{1}{g(0)} \operatorname{Var}\left(\hat{\mu}_\zeta(\zeta_1, \ldots, \zeta_n)\right) \geq C l^{-D}$$

for all $l \geq l_0$ by Lemma 4.4 in Adenstedt (1974). □

The next Lemma deals with the ρ-mixing coefficient, which is defined in the following way: Let $\mathcal{A}, \mathcal{B}$ be two σ-fields. Then

$$\rho(\mathcal{A}, \mathcal{B}) := \sup \operatorname{corr}(X, Y),$$

where the supremum is taken over all $\mathcal{A}$-measurable random variables X and $\mathcal{B}$-measurable random variables Y. For details we recommend the book of Bradley (2007).

Lemma 4. *Under Assumption 2, there are constants* $C_1, C_2 < \infty$ *such that*

$$\rho(k,l) := \rho\big(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l+1 \leq j \leq k+2l)\big) \\ \leq C_1 \, (k/l)^{-D} \, L_\gamma(k) + C_2 l^2 k^{-D-1} \max\{L_\gamma(k), 1\}$$

for all $k \in \mathbb{N}$ *and all* $l \in \{l_k, \ldots, k\}$*.*

Proof. Kolmogorov and Rozanov (1960) proved that there exist real numbers $a_1, \ldots, a_l$, $b_1, \ldots, b_l$ such that

$$\rho\big(\sigma(\xi_i, 1 \leq i \leq l), \sigma(\xi_j, k+l+1 \leq j \leq k+2l)\big) = \operatorname{Cov}\Big(\sum_{i=1}^{l} a_i \xi_i, \sum_{j=1}^{l} b_j \xi_{k+l+j}\Big)$$

and $\operatorname{Var}\big(\sum_{i=1}^{l} a_i \xi_i\big) = \operatorname{Var}\big(\sum_{j=1}^{l} b_j \xi_{k+l+j}\big) = 1$. The triangular inequality yields

$$\Big|\operatorname{Cov}\Big(\sum_{i=1}^{l} a_i \xi_i, \sum_{j=1}^{l} b_j \xi_{k+l+j}\Big)\Big| \\ \leq \Big|\sum_{i=1}^{l} a_i \sum_{j=1}^{l} b_j\Big| \, |\gamma(k)| + \sum_{i=1}^{l} \sum_{j=1}^{l} |a_i||b_j| \, |\gamma(k) - \gamma(k+l+j-i)| \, .$$

We will treat the two summands on the right hand side separately. For the first term, it follows by Lemma 3 that

$$\Big|\sum_{i=1}^{l} a_i \sum_{j=1}^{l} b_j\Big| \, |\gamma(k)| = \Big|\sum_{i=1}^{l} a_i\Big| \Big|\sum_{j=1}^{l} b_j\Big| \, |\gamma(k)| \leq K_d'^2 l^D L_\gamma(k) k^{-D}.$$

Before we deal with the second summand, we observe that by Hölder's inequality and Lemma 2

$$\sum_{i=1}^{l} |a_i| \leq \sqrt{l \sum_{i=1}^{l} a_i^2} \leq \sqrt{K_D}\sqrt{l} \quad \text{and} \quad \sum_{j=1}^{l} |b_j| \leq \sqrt{l \sum_{j=1}^{l} b_j^2} \leq \sqrt{K_D}\sqrt{l}.$$

Due to Assumption 2

$$\sup_{|k-\tilde{k}| \leq 2l-1} \Big|L_\gamma(k) - L_\gamma(\tilde{k})\Big| \leq K \frac{l}{k}$$

for some constant K. Consequently, for all $\tilde{k} \in \{k+1, \ldots, k+2l-1\}$

$$\begin{aligned}
\left|\gamma(k) - \gamma(\tilde{k})\right| &\leq L_\gamma(k) \left|k^{-D} - \tilde{k}^{-D}\right| + |L_\gamma(k) - L_\gamma(\tilde{k})|\tilde{k}^{-D} \\
&\leq L_\gamma(k) \left(k^{-D} - (k+2l-1)^{-D}\right) + |L_\gamma(k) - L_\gamma(\tilde{k})|k^{-D} \\
&\leq C_d k^{-D-1} l L_\gamma(k) + K \frac{l}{k} k^{-D} \max\{L_\gamma(k), 1\} \\
&\leq C_3 k^{-D-1} l \max\{L_\gamma(k), 1\}
\end{aligned}$$

for some constants C_d, C_3. Combining this with the bounds for $\sum_{i=1}^{l} |a_i|$, $\sum_{j=1}^{l} |b_j|$, we finally arrive at

$$\begin{aligned}
\sum_{i=1}^{l} |a_i| \sum_{j=1}^{l} |b_j| \, |\gamma(k) - \gamma(k+l+j-i)| &\leq K_D l \max_{\tilde{k} \in \{k+1, \ldots, k+2l-1\}} \left|\gamma(k) - \gamma(\tilde{k})\right| \\
&= K_D C_3 k^{-D-1} l^2 \max\{L_\gamma(k), 1\}.
\end{aligned}$$

□

S4 Proof of the Main Result

Before we give the proof of our main theorem, let us recall our assumptions on the test statistic and the block length:

Assumption 1. *$(X_n)_{n\in\mathbb{N}}$ is a stochastic process and $(T_n)_{n\in\mathbb{N}}$ is a sequence of statistics such that $T_n \Rightarrow T$ in distribution as $n \to \infty$ for a random variable T with distribution function F_T.*

Assumption 3. *Let $(l_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence of integers such that $l = l_n \to \infty$ as $n \to \infty$ and $l_n = \mathcal{O}\left(n^{(1+D)/2-\epsilon}\right)$ for some $\epsilon > 0$.*

Proof of Theorem 1. Let t be a point of continuity of F_T. In order to simplify notation, we write $N = n - l + 1$ and $T_{l,i} = T_l(X_i, \ldots, X_{i+l-1})$. The triangular inequality yields

$$|\hat{F}_{l,n}(t) - F_{T_n}(t)| \leq |\hat{F}_{l,n}(t) - F_T(t)| + |F_T(t) - F_{T_n}(t)|.$$

The second term on the right-hand side of the above inequality converges to zero because of Assumption 1. As L_2-convergence implies stochastic convergence, it suffices to show that $\mathrm{E}(|\hat{F}_{l,n}(t) - F_T(t)|^2) \longrightarrow 0$ in order to prove that the first term converges to zero, as well. We have

$$\begin{aligned}
&\mathrm{E}\left(|\hat{F}_{l,n}(t) - F_T(t)|^2\right)\\
&= \mathrm{E}\left(\hat{F}^2_{l,n}(t)\right) - \left(\mathrm{E}\,\hat{F}_{l,n}(t)\right)^2 + (F_T(t))^2 - 2F_T(t)\,\mathrm{E}\,\hat{F}_{l,n}(t) + \left(\mathrm{E}\,\hat{F}_{l,n}(t)\right)^2\\
&= \mathrm{Var}(\hat{F}_{l,n}(t)) + \left|\mathrm{E}\,\hat{F}_{l,n}(t) - F_T(t)\right|^2.
\end{aligned}$$

Furthermore, stationarity of the process $(X_n)_{n\in\mathbb{N}}$ and Assumption 1 imply

$$\mathrm{E}\,\hat{F}_{l,n}(t) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left(1_{\{T_{l,i}\leq t\}}\right) = P\left(T_{l,1} \leq t\right) = F_{T_l}(t) \xrightarrow{l\to\infty} F_T(t).$$

It remains to show that $\mathrm{Var}(\hat{F}_{l,n}(t)) \longrightarrow 0$. Again, it follows by stationarity of $(X_n)_{n\in\mathbb{N}}$ that

$$\begin{aligned}
&\mathrm{Var}\left(\hat{F}_{l,n}(t)\right)\\
&\quad= \frac{1}{N}\mathrm{Var}\left(1_{\{T_{l,1}\leq t\}}\right) + \frac{2}{N^2}\sum_{i=2}^{N}(N-i+1)\mathrm{Cov}\left(1_{\{T_{l,1}\leq t\}}, 1_{\{T_{l,i}\leq t\}}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad \leq \frac{2}{N}\sum_{i=1}^{N}\left|\mathrm{Cov}\left(1_{\{T_{l,1}\leq t\}}, 1_{\{T_{l,i}\leq t\}}\right)\right|.
\end{aligned}$$

Recall that by Assumption 3, we have $l \leq C_l n^{(1+D)/2-\epsilon}$ for some constants C_l and $\epsilon > 0$. For n large enough such that $l < \frac{1}{2}\lfloor n^{1-\epsilon/2}\rfloor$, we split the sum of covariances into two parts:

$$\frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{Cov}\left(1_{\{T_{l,1}\leq t\}}, 1_{\{T_{l,i}\leq t\}}\right)\right|$$

$$
\begin{aligned}
&= \frac{1}{N}\sum_{i=1}^{\lfloor n^{1-\epsilon/2}\rfloor}\left|\operatorname{Cov}\left(1_{\left\{T_{l,1}\le t\right\}}, 1_{\left\{T_{l,i}\le t\right\}}\right)\right| \\
&+ \frac{1}{N}\sum_{i=\lfloor n^{1-\epsilon/2}\rfloor+1}^{N}\left|\operatorname{Cov}\left(1_{\left\{T_{l,1}\le t\right\}}, 1_{\left\{T_{l,i}\le t\right\}}\right)\right| \\
&\le \frac{\lfloor n^{1-\epsilon/2}\rfloor}{N} + \frac{1}{N}\sum_{k=\lfloor n^{1-\epsilon/2}\rfloor+1}^{N}\rho(\sigma(X_i, 1\le i\le l), \sigma(X_j, k\le j\le k+l-1)) \\
&\le \frac{\lfloor n^{1-\epsilon/2}\rfloor}{N} + \frac{1}{N}\sum_{k=\lfloor n^{1-\epsilon/2}\rfloor-l}^{N-l-1}\rho(k,l),
\end{aligned}
$$

where

$$
\rho(k,l) := \rho\big(\sigma(X_i, 1\le i\le l), \sigma(X_j, k+l+1\le j\le k+2l)\big).
$$

Obviously, the first summand converges to zero by Assumption 3. For the second summand note that as a consequence of Potter's Theorem (Theorem 1.5.6 in the book of Bingham, Goldie and Teugels (1987)), there is a constant C_L such that $L_\gamma(k) \le C_L k^{D\epsilon/2}$ for all $k \in \mathbb{N}$. This together with Lemma 4 yields

$$
\begin{aligned}
&\frac{1}{N}\sum_{k=\lfloor n^{1-\epsilon/2}\rfloor-l}^{N-l-1}\rho(k,l) \\
&\le C_L C_1 \frac{l^D}{N}\sum_{k=\lfloor n^{1-\epsilon/2}\rfloor/2}^{N-l-1} k^{-D}k^{D\epsilon/2} + C_L C_2 \frac{l^2}{N}\sum_{k=\lfloor n^{1-\epsilon/2}\rfloor/2}^{N-l-1} k^{-D-1}k^{D\epsilon/2} \\
&\le C_L C_1 C_l^D 2^{D(1-\epsilon/2)} n^{D(((1+D)/2-\epsilon)-(1-\epsilon/2)+\epsilon/2(1-\epsilon/2))} \\
&\quad + C_L C_2 C_l^2 2^{1+D(1-\epsilon/2)} n^{((1+D-2\epsilon)-(D+1)(1-\epsilon/2)+(1-\epsilon/2)D\epsilon/2)} \\
&\le C\left(n^{-D\left((1-D)/2+\epsilon^2/4\right)} + n^{-\epsilon(\frac{3}{2}-D+D\epsilon/4)}\right) \xrightarrow{n\to\infty} 0
\end{aligned}
$$

for some constant $C < \infty$. Thus, we have proved that $\operatorname{Var}(\hat{F}_{l,n}(t)) \to 0$ as $n \to \infty$ and that the first conjecture of Theorem 1 holds.

The second assertion of Theorem 1 follows from $F_{T_n}(t) - \hat{F}_{l,n}(t) \to 0$ in probability by the usual Glivenko-Cantelli argument for the uniform convergence of empirical distribution functions; see for example section 20 in the book of Billingsley (1995). □

Bibliography

Adenstedt, R.K. (1974). On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics* **2**, 1095 – 1107.

Betken, A. (2016). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *Journal of Time Series Analysis* **37**, 185 – 809.

Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons, Inc.

Bingham, N.H., Goldie, C.M., and Teugels, J.L. (1987). *Regular variation*. Cambridge University Press.

Bradley, R.C. (2007). *Introduction to strong mixing conditions*. Kendrick press.

Dehling, H., Rooch, A., and Wendler, M. (2017). Two-Sample U-Statistic Processes for Long-Range Dependent Data. *Statistics* **51**, 84 – 104.

Kolmogorov, A.N. and Rozanov, Y.A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theory of Probability & Its Applications* **5**, 204 – 208.