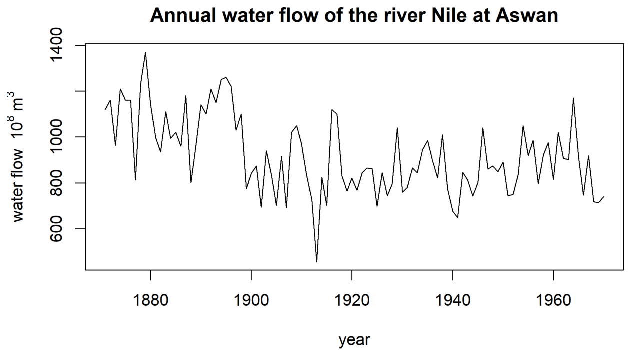
Sharp multiscale change detection for time series

Johann Köhne (University of Göttingen)
Fabian Mies (Delft University of Technology)

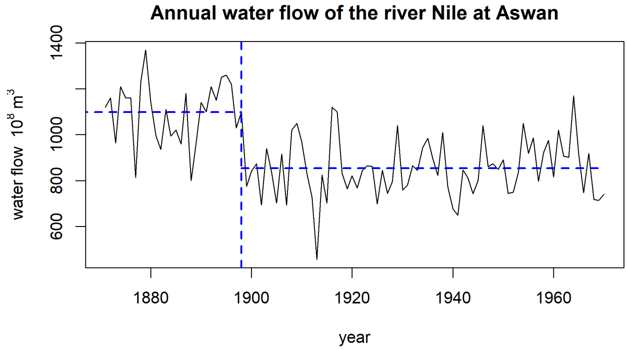
11th Bernoulli-IMS World Congress
2024-08-13, Bochum

Changepoint example



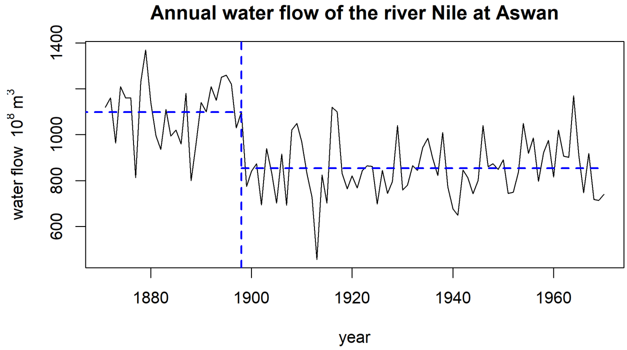
H_0 : the process is stationary
H_1 : the process is not stationary

Changepoint example

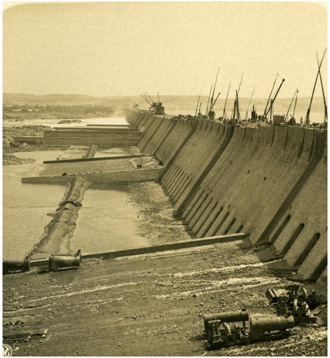


H_0 : the process is stationary
H_1 : the process is not stationary

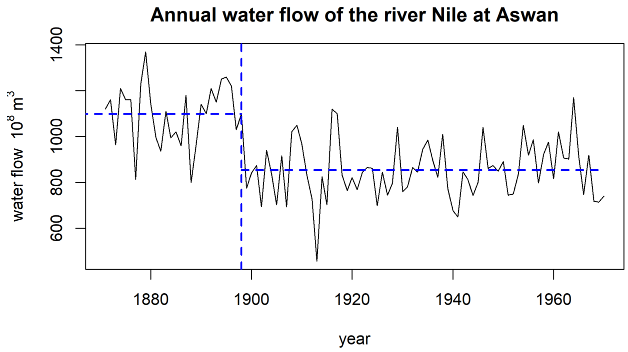
Changepoint example



H_0 : the process is stationary
H_1 : the process is not stationary



Changepoint example

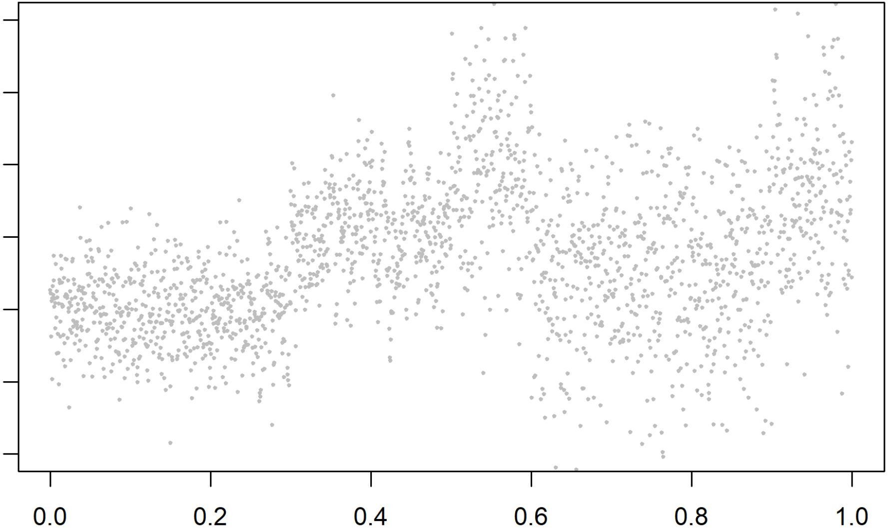


H_0 : the process is stationary
H_1 : the process is not stationary

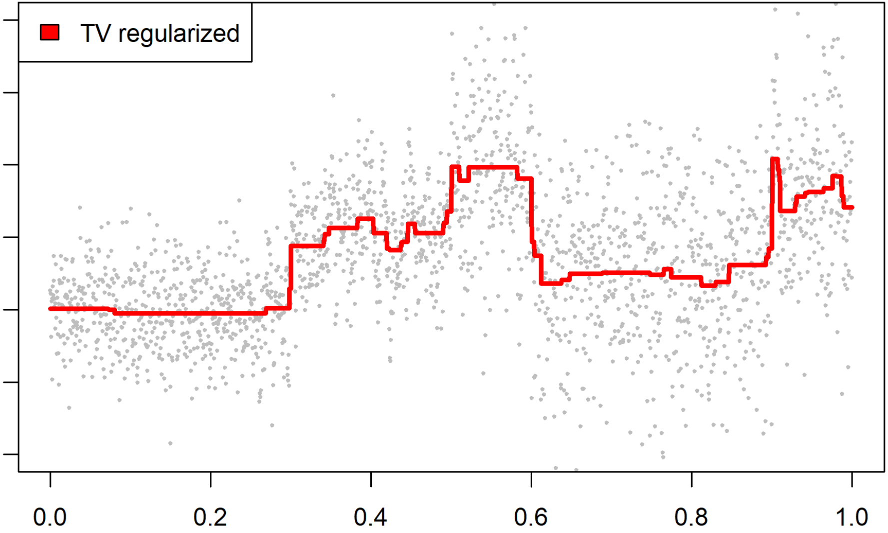
Statistically similar applications:

- Trend changes in house prices
- Volatility change in a portfolio
- Faults in a wind turbine
- Quality degradation in a manufacturing system

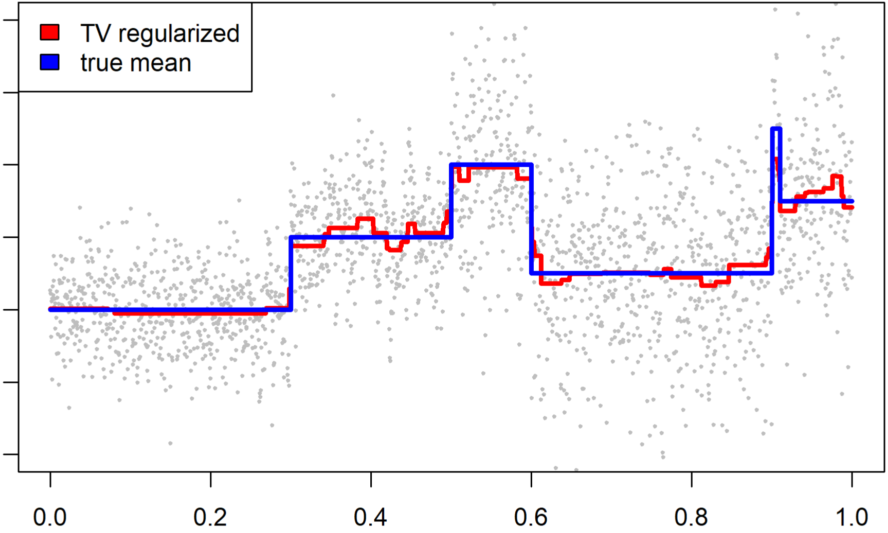
How many structural breaks?



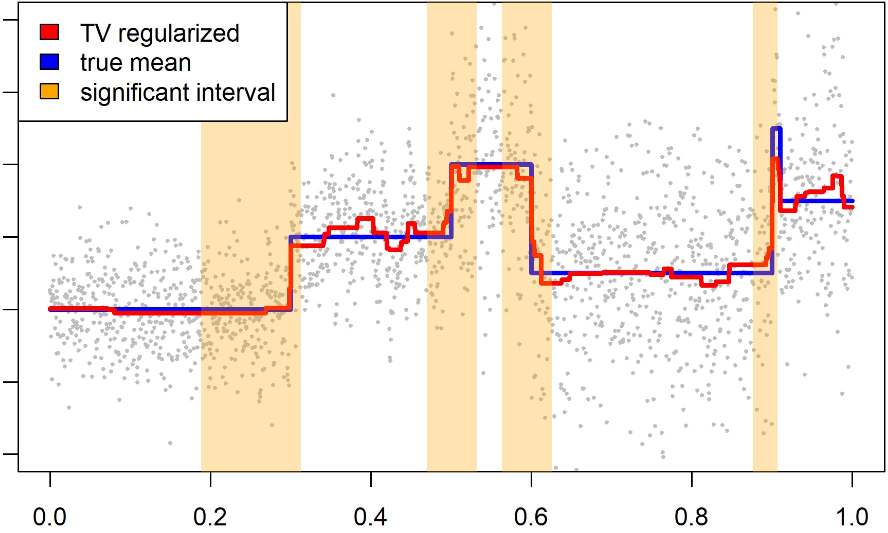
How many structural breaks?



How many structural breaks?



How many structural breaks?



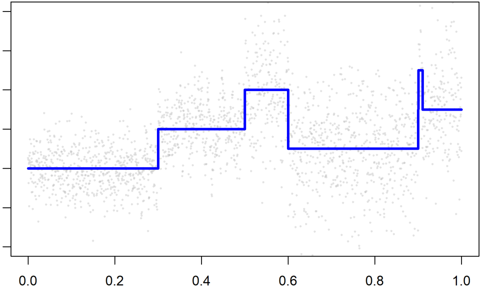
Multiple changes in mean

Multiple change point problem

We observe $X_1, \ldots, X_n$ where

$$X_t = \epsilon_t + \mu_t,$$
$$\mu_t = \sum_{k=1}^{m} \mu^{(k)} \mathbb{1}(t \in (\tau_{k-1}, \tau_k]).$$

for centered error terms ϵ_t (typically iid).



Multiple changes in mean

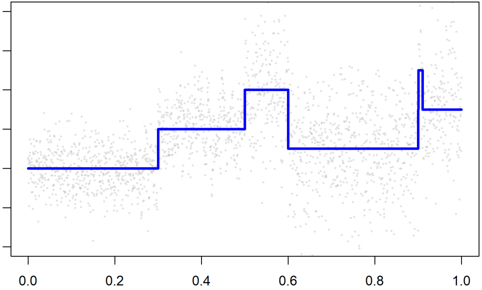
Multiple change point problem

We observe $X_1, \dots, X_n$ where

$$X_t = \epsilon_t + \mu_t,$$

$$\mu_t = \sum_{k=1}^{m} \mu^{(k)} \mathbb{1}(t \in (\tau_{k-1}, \tau_k]).$$

for centered error terms ϵ_t (typically iid).



- $\Delta_k = \mu_k - \mu_{k-1}$ (size of change)
- $L_k = |\tau_k - \tau_{k-1}| \wedge |\tau_{k+1} - \tau_k|$ (length of change)

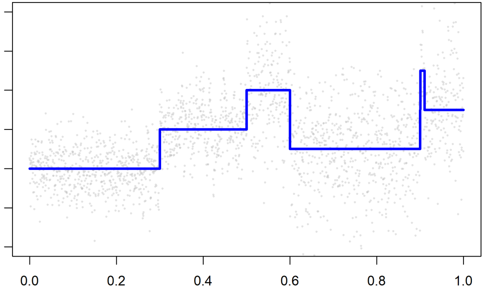
Multiple changes in mean

Multiple change point problem

We observe $X_1, \ldots, X_n$ where

$$X_t = \epsilon_t + \mu_t,$$
$$\mu_t = \sum_{k=1}^{m} \mu^{(k)} \mathbb{1}(t \in (\tau_{k-1}, \tau_k]).$$

for centered error terms ϵ_t (typically iid).



- $\Delta_k = \mu_k - \mu_{k-1}$ (size of change)
- $L_k = |\tau_k - \tau_{k-1}| \wedge |\tau_{k+1} - \tau_k|$ (length of change)

Detectability (Verzelen et al., 2023)

Changepoint τ_k is detectable if and only if

$$E_k = \Delta_k \sqrt{L_k} \gg \sqrt{\log \frac{n}{L_k}}$$

Multiscale scan statistic for changepoints

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \left[\overbrace{\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right|}^{\text{scan statistic for interval } (u,v]} - \overbrace{c_\alpha \sqrt{\log \frac{e\,n}{|v-u|}}}^{\text{multiple testing correction}} \right]$$

Multiscale scan statistic for changepoints

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \Bigg[\overbrace{\sup_{(u,v]\subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right|}^{\text{scan statistic for interval } (u,v]} - \overbrace{c_\alpha \sqrt{\log \frac{e\, n}{|v-u|}}}^{\text{multiple testing correction}} \Bigg]$$

If c_α is chosen suitably:

- $P(T_n > 0) \leq \alpha$ under H_0,
- $P(T_n > 0) \to 1$ if there is at least one detectable change

Multiscale scan statistic for changepoints

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \Bigg[\overbrace{\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right|}^{\text{scan statistic for interval } (u,v]} - \overbrace{c_\alpha \sqrt{\log \frac{e\,n}{|v-u|}}}^{\text{multiple testing correction}} \Bigg]$$

If c_α is chosen suitably:

- $P(T_n > 0) \leq \alpha$ under H_0,
- $P(T_n > 0) \to 1$ if there is at least one detectable change

Critical value c_α depends on the error terms ϵ_t

- Dümbgen and Spokoiny (2001) for Gaussian white noise model
- Frick et al. (2014) for changepoints with Gaussian errors and constraint $|I| \gg \log(n)^3$

Multiscale scan statistic for changepoints

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \Bigg[\overbrace{\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right|}^{\text{scan statistic for interval } (u,v]} - \overbrace{c_\alpha \sqrt{\log \frac{e\,n}{|v-u|}}}^{\text{multiple testing correction}} \Bigg]$$

If c_α is chosen suitably:

- $P(T_n > 0) \leq \alpha$ under H_0,
- $P(T_n > 0) \to 1$ if there is at least one detectable change

Critical value c_α depends on the error terms ϵ_t

- Dümbgen and Spokoiny (2001) for Gaussian white noise model
- Frick et al. (2014) for changepoints with Gaussian errors and constraint $|I| \gg \log(n)^3$
- Verzelen et al. (2023) consider sub-Gaussian errors and $|I| \gg \log(n)$

Multiscale scan statistic for changepoints

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \Bigg[\overbrace{\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right|}^{\text{scan statistic for interval } (u,v]} - \overbrace{c_\alpha \sqrt{\log \frac{e\,n}{|v-u|}}}^{\text{multiple testing correction}} \Bigg]$$

If c_α is chosen suitably:

- $P(T_n > 0) \leq \alpha$ under H_0,
- $P(T_n > 0) \to 1$ if there is at least one detectable change

Critical value c_α depends on the error terms ϵ_t

- Dümbgen and Spokoiny (2001) for Gaussian white noise model
- Frick et al. (2014) for changepoints with Gaussian errors and constraint $|I| \gg \log(n)^3$
- Verzelen et al. (2023) consider sub-Gaussian errors and $|I| \gg \log(n)$
- Dette et al. (2020) allow for stationary dependence, but with the constraint $|I| \gg n^{2/3}$

Multiscale scan statistic for changepoints

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \Bigg[\overbrace{\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right|}^{\text{scan statistic for interval } (u,v]} - \overbrace{c_\alpha \sqrt{\log \frac{e\,n}{|v-u|}}}^{\text{multiple testing correction}} \Bigg]$$

If c_α is chosen suitably:

- $P(T_n > 0) \leq \alpha$ under H_0,
- $P(T_n > 0) \to 1$ if there is at least one detectable change

Critical value c_α depends on the error terms ϵ_t

- Dümbgen and Spokoiny (2001) for Gaussian white noise model
- Frick et al. (2014) for changepoints with Gaussian errors and constraint $|I| \gg \log(n)^3$
- Verzelen et al. (2023) consider sub-Gaussian errors and $|I| \gg \log(n)$
- Dette et al. (2020) allow for stationary dependence, but with the constraint $|I| \gg n^{2/3}$

Our contribution

Derive a procedure which is

- feasible
- optimal
- for dependent, nonstationary errors

via novel asymptotic theory.

Rewrite the statistic as a Hölder norm

$$T_n = \sup_{I\subset[1,n]} \inf_{\mu} \left[\sup_{(u,v]\subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right| - c_\alpha \sqrt{\log \frac{e\,n}{|v-u|}} \right]$$

Step (i): bound the test statistic under H_0 as

$$T_n \leq T_n^* = \sup_{0\leq u<v\leq 1} \frac{|S_n(v) - S_n(u)|}{\sqrt{|u-v|}} - c_\alpha \sqrt{\log \frac{e}{|u-v|}}, \qquad \text{where } S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \epsilon_t + \frac{u - \frac{\lfloor un \rfloor}{n}}{\sqrt{n}} \epsilon_{\lceil un \rceil}.$$

Rewrite the statistic as a Hölder norm

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \left[\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right| - c_\alpha \sqrt{\log \frac{e\, n}{|v-u|}} \right]$$

Step (i): bound the test statistic under H_0 as

$$T_n \leq T_n^* = \sup_{0 \leq u < v \leq 1} \frac{|S_n(v) - S_n(u)|}{\sqrt{|u-v|}} - c_\alpha \sqrt{\log \frac{e}{|u-v|}}, \qquad \text{where } S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \epsilon_t + \frac{u - \frac{\lfloor un \rfloor}{n}}{\sqrt{n}} \epsilon_{\lceil un \rceil}.$$

Step (ii): Reformulate the decision rule as

$$\overline{T}_n := \sup_{0 \leq u < v \leq 1} \frac{|S_n(v) - S_n(u)|}{\sqrt{|u-v| \log(e/|u-v|)}} > c_\alpha$$

Rewrite the statistic as a Hölder norm

$$T_n = \sup_{I \subset [1,n]} \inf_{\mu} \left[\sup_{(u,v] \subset I} \left| \frac{1}{\sqrt{|v-u|}} \sum_{t=u+1}^{v} (X_t - \mu) \right| - c_\alpha \sqrt{\log \frac{e\, n}{|v-u|}} \right]$$

Step (i): bound the test statistic under H_0 as

$$T_n \leq T_n^* = \sup_{0 \leq u < v \leq 1} \frac{|S_n(v) - S_n(u)|}{\sqrt{|u-v|}} - c_\alpha \sqrt{\log \frac{e}{|u-v|}}, \qquad \text{where } S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \epsilon_t + \frac{u - \frac{\lfloor un \rfloor}{n}}{\sqrt{n}} \epsilon_{\lceil un \rceil}.$$

Step (ii): Reformulate the decision rule as

$$\overline{T}_n := \sup_{0 \leq u < v \leq 1} \frac{|S_n(v) - S_n(u)|}{\sqrt{|u-v| \log(e/|u-v|)}} > c_\alpha$$

Step (iii): Rewrite as Hölder norm with modulus $\rho_0(h) = \sqrt{h(1 + \log h)}$

$$\overline{T}_n := \|S_n\|_{\rho_0} \xrightarrow{d} \|\sigma B\|_{\rho_0} \quad ?$$

Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$

Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$

Question 1: Is this just weak convergence?

Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$

Question 1: Is this just weak convergence?

- $\rho_0(h) = \sqrt{h \log(e/h)}$ is exactly the modulus of continuity of a Brownian motion
- Brownian motion belongs to C^{ρ_0}, but its probability measure is **not tight**.
- However, we do obtain weak convergence in the Hölder space C^{ρ}, for all $\rho(h) \gg \rho_0(h)$.

Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$

Question 2: What happens at $t < t_0$?

Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$

Question 2: What happens at $t < t_0$?

- Let $X_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$,
- and $X_t' = \sqrt{2}N_t \cdot X_t$, for $N_t \overset{\text{iid}}{\sim} \text{Ber}(0.5)$
- We have $S_n \xrightarrow[n\to\infty]{d} W$ and $S_n' \xrightarrow[n\to\infty]{d} W$, but X_t' has a larger sub-Gaussian norm.

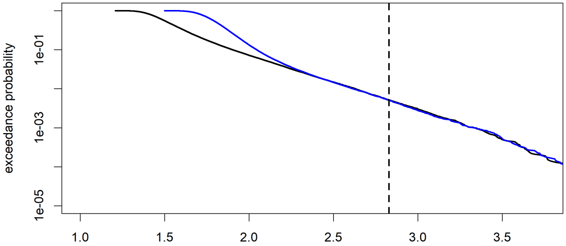
Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$



Distribution of $\|S_n\|_{\rho_0}$ ($n = 2^{15}$) with same variance, but different sub-Gaussian norm.

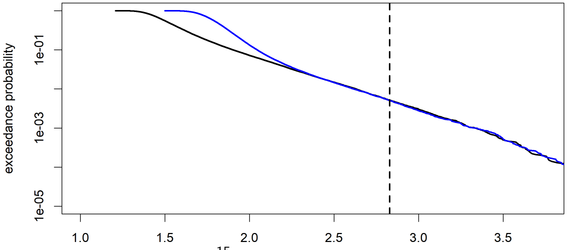
Asymptotic critical values

Theorem

- Let $S_n \xrightarrow{d} W$ in $C_0[0,1]$, for a Gaussian process W, and (Donsker's theorem)
- the increments of S_n are sub-Gaussian, $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$. (tightness)

Then there exists a $t_0 > 0$ such that

$$P\left(\|S_n\|_{\rho_0} > t\right) \to P\left(\|W\|_{\rho_0} > t\right) \quad \forall t \geq t_0.$$



Distribution of $\|S_n\|_{\rho_0}$ ($n = 2^{15}$) with same variance, but different sub-Gaussian norm.

Statistical implication:

- Use multiscale statistic with critical value c_α such that $P(\|W\|_{\rho_0} > c_\alpha) \leq \alpha$
- If α is small enough, the test will be asymptotically valid.

Bootstrapping critical values

What is the limit W?

If the errors ϵ_t are suitably ergodic and locally-stationary, then

$$S_n(u) \xrightarrow{d} W(u) = \int_0^u \sigma_s \, dB_s$$

and σ_s^2 is the local long-run-variance.

Bootstrapping critical values

Approximate W based on $\eta_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ as

$$W_n^*(u) = \frac{1}{\sqrt{n}} \sum_{t=b}^{\lfloor nu \rfloor} \left(\frac{1}{\sqrt{b}} \sum_{s=t-b+1}^{t} X_t - \widehat{\mu}_t \right) \cdot \eta_t,$$

What is the limit W?

If the errors ϵ_t are suitably ergodic and locally-stationary, then

$$S_n(u) \xrightarrow{d} W(u) = \int_0^u \sigma_s \, dB_s$$

and σ_s^2 is the local long-run-variance.

Bootstrapping critical values

Approximate W based on $\eta_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ as

$$W_n^*(u) = \frac{1}{\sqrt{n}} \sum_{t=b}^{\lfloor nu \rfloor} \left(\frac{1}{\sqrt{b}} \sum_{s=t-b+1}^{t} X_t - \widehat{\mu}_t \right) \cdot \eta_t,$$

and

$$T_n^* = \sup_{|u-v|>c_n} \frac{|W_n^*(u) - W_n^*(v)|}{\rho_0(|u-v|)}.$$

What is the limit W?

If the errors ϵ_t are suitably ergodic and locally-stationary, then

$$S_n(u) \xrightarrow{d} W(u) = \int_0^u \sigma_s \, dB_s$$

and σ_s^2 is the local long-run-variance.

Bootstrapping critical values

Approximate W based on $\eta_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ as

$$W_n^*(u) = \frac{1}{\sqrt{n}} \sum_{t=b}^{\lfloor nu \rfloor} \left(\frac{1}{\sqrt{b}} \sum_{s=t-b+1}^{t} X_t - \widehat{\mu}_t \right) \cdot \eta_t,$$

and

$$T_n^* = \sup_{|u-v|>c_n} \frac{|W_n^*(u) - W_n^*(v)|}{\rho_0(|u-v|)}.$$

What is the limit W?

If the errors ϵ_t are suitably ergodic and locally-stationary, then

$$S_n(u) \xrightarrow{d} W(u) = \int_0^u \sigma_s \, dB_s$$

and σ_s^2 is the local long-run-variance.

Bootstrap consistency

- If $b = b_n$ and c_n are chosen suitably, and
- $\frac{1}{n}\sum_{t=1}^n |\hat{\mu}_t - \mu_t|^2 = \mathcal{O}(n^{-\eta})$ for some $\eta > 0$,

then $(T_n^* | X_1, \ldots, X_n) \xrightarrow{d} \|W\|_{\rho_0}$

Bootstrapping critical values

Approximate W based on $\eta_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ as

$$W_n^*(u) = \frac{1}{\sqrt{n}} \sum_{t=b}^{\lfloor nu \rfloor} \left(\frac{1}{\sqrt{b}} \sum_{s=t-b+1}^{t} X_t - \widehat{\mu}_t \right) \cdot \eta_t,$$

and

$$T_n^* = \sup_{|u-v|>c_n} \frac{|W_n^*(u) - W_n^*(v)|}{\rho_0(|u-v|)}.$$

(i) Compute the multiscale test statistic T_n
(ii) Estimate μ_t nonparametrically
(iii) Sample T_n^* repeatedly as above, to obtain an approximate quantile $\hat{c}_\alpha$
(iv) Detect a change if $T_n > \hat{c}_\alpha$.

What is the limit W?

If the errors ϵ_t are suitably ergodic and locally-stationary, then

$$S_n(u) \xrightarrow{d} W(u) = \int_0^u \sigma_s \, dB_s$$

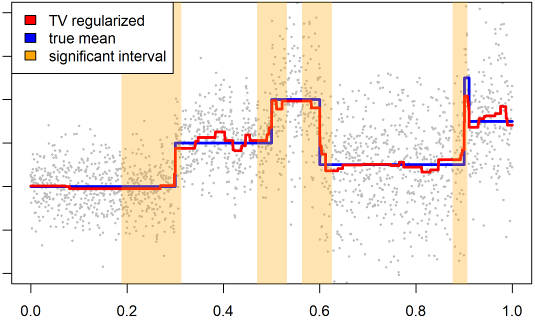
and σ_s^2 is the local long-run-variance.

Bootstrap consistency

- If $b = b_n$ and c_n are chosen suitably, and
- $\frac{1}{n} \sum_{t=1}^n |\hat{\mu}_t - \mu_t|^2 = \mathcal{O}(n^{-\eta})$ for some $\eta > 0$,

then $(T_n^* | X_1, \ldots, X_n) \xrightarrow{d} \|W\|_{\rho_0}$

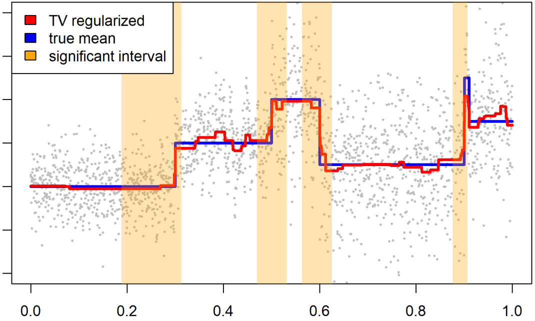
Intervals of significance



- As $T_n = \sup_I T_n(I)$ is in the form of a scan statistic, we also find interval estimates $\widehat{I}_1, \ldots, \widehat{I}_m$ for the changepoints, where $\boxed{T_n(\hat{I}_k) > \hat{c}_\alpha}$.
- As in Fryzlewicz (2023), we find that with probability at least $1 - \alpha$, every $\widehat{I}_k$ contains at least one changepoint:

$$\boxed{\liminf P\left(\hat{I}_j \cap \boldsymbol{\tau} \neq \emptyset \; \forall j = 1, \ldots, \hat{m}\right) \geq \alpha}$$

Intervals of significance



- As $T_n = \sup_I T_n(I)$ is in the form of a scan statistic, we also find interval estimates. $\widehat{I}_1, \ldots, \widehat{I}_m$ for the changepoints, where $\boxed{T_n(\hat{I}_k) > \hat{c}_\alpha}$.
- As in Fryzlewicz (2023), we find that with probability at least $1-\alpha$, every $\widehat{I}_k$ contains at least one changepoint:

$$\boxed{\liminf P\left(\hat{I}_j \cap \boldsymbol{\tau} \neq \emptyset \; \forall j = 1, \ldots, \hat{m}\right) \geq \alpha}$$

Proposition (Power)

If the k-th changepoint τ_k satisfies $\Delta_k \sqrt{L_k} \gg \sqrt{\log(n/L_k)}$, then

$$P(T_n(I) > \hat{c}_\alpha) \to 1.$$

Technicalities: Increments of the partial sum process

Recall: We require the interpolated partial sum process $S_n(u)$ to satisfy

(i) $S_n \Rightarrow W$ in $C_0(0,1)$

(ii) $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$

Technicalities: Increments of the partial sum process

Recall: We require the interpolated partial sum process $S_n(u)$ to satisfy

(i) $S_n \Rightarrow W$ in $C_0(0,1)$

(ii) $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$

Assumption (Local stationarity)

The noise process is given by

$$\epsilon_t = \epsilon_{t,n} = G_{t/n}(\eta_t, \eta_{t-1}, \ldots)$$

for $\eta_i \overset{\text{iid}}{\sim} U(0,1)$, and the mapping $u \mapsto G_u(\eta)$ has bounded variation, measured in $\|\cdot\|_{\psi_2}$.

Definition

For any $h \in \mathbb{N}$, define the sub-Gaussian physical dependence measure of $(\epsilon_t)_{t\in\mathbb{Z}}$ as

$$\delta_{\psi_2}(h) := \sup_u \left\| G_u(\eta_t) - G_u(\eta_{t,h}) \right\|_{\psi_2}.$$

- $\delta_{\psi_2}(h) = \mathcal{O}(h^{-2-\eta})$ is sufficient for $S_n \Rightarrow W$.

Technicalities: Increments of the partial sum process

Recall: We require the interpolated partial sum process $S_n(u)$ to satisfy

(i) $S_n \Rightarrow W$ in $C_0(0,1)$

(ii) $\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}$

Assumption (Local stationarity)

The noise process is given by

$$\epsilon_t = \epsilon_{t,n} = G_{t/n}(\eta_t, \eta_{t-1}, \ldots)$$

for $\eta_i \overset{\text{iid}}{\sim} U(0,1)$, and the mapping $u \mapsto G_u(\eta)$ has bounded variation, measured in $\|\cdot\|_{\psi_2}$.

Definition

For any $h \in \mathbb{N}$, define the sub-Gaussian physical dependence measure of $(\epsilon_t)_{t\in\mathbb{Z}}$ as

$$\delta_{\psi_2}(h) := \sup_u \|G_u(\eta_t) - G_u(\eta_{t,h})\|_{\psi_2}.$$

- $\delta_{\psi_2}(h) = \mathcal{O}(h^{-2-\eta})$ is sufficient for $S_n \Rightarrow W$.

Theorem

$$\|S_n(u) - S_n(v)\|_{\psi_2} \leq C\sqrt{|u-v|}\sum_{j=1}^{\infty}\sqrt{j}\delta_{\psi_2}(j)$$

Technicalities: Bootstrap consistency

Theorem

Suppose that

- the noise ϵ_t is locally stationary, and its physical dependence measure decays as $\delta_{\psi_2}(j) = \mathcal{O}(j^{-3})$, and
- the mean estimator satisfies $\frac{1}{n}\sum_{t=1}^{n} |\hat{\mu}_t - \mu_t|^2 = \mathcal{O}_{L_p}(n^{-\eta})$.

then the 2-Wasserstein distance between T_n and the (X-conditional) distribution of T_n^* is bounded as

$$d_{W_2}(T_n, T_n^*)^2 = \frac{\log(n)}{\rho_0(c_n)}\, \mathcal{O}_P\left(\left(\frac{b}{n}\right)^{\frac{1}{4}} + \frac{1}{\sqrt{b}} + n^{-\eta/4}\right).$$

Summary

Methodological findings:

(i) Asymptotic threshold c_α for multiscale statistic only depends on second moments

(ii) Sub-Gaussian variance determines for which significance levels α the asymptotic threshold is applicable.

(iii) Consider smallest scales of the data, but neglect them when bootstrapping.

Summary

Methodological findings:

(i) Asymptotic threshold c_α for multiscale statistic only depends on second moments

(ii) Sub-Gaussian variance determines for which significance levels α the asymptotic threshold is applicable.

(iii) Consider smallest scales of the data, but neglect them when bootstrapping.

Thank you for your attention!

Dette, H., Eckle, T., and Vetter, M. (2020). Multiscale change point detection for dependent data. *Scandinavian Journal of Statistics*, 47(4):1243–1274.

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale Testing of Qualitative Hypotheses. *The Annals of Statistics*, 29(1).

Frick, K., Munk, A., and Sieling, H. (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 495–580. Publisher: JSTOR.

Fryzlewicz, P. (2023). Narrowest Significance Pursuit: inference for multiple change-points in linear models. *Journal of American Statistical Association*. arXiv: 2009.05431.

Verzelen, N., Fromont, M., Lerasle, M., and Reynaud-Bouret, P. (2023). Optimal Change-Point Detection and Localization. *The Annals of Statistics*, 51(4):1586–1610. arXiv: 2010.11470.