Change-Point Analysis in a correlated Gaussian sequence.

Clément Chesseboeuf
Joint work with Hermine Biermé and Farida Enikeeva.

LMA-Université de Poitiers

Colloque JPS 2016

Change point analysis framework

- A family of random variables $(Y_k)_{1 \leqslant k \leqslant N}$.
- A parametric family of distributions $\{P_\theta, \ \theta \in \Theta\}$.
- A location parameter $\tau^\star \in]0, 1[$.

From the point of view of testing, the problem takes the form :

- The sequence is identically distributed :

$$P_{Y_k} = P_{\theta_1}, \quad 1 \leqslant k \leqslant N. \qquad (\mathcal{H}_0)$$

- There is a change in the sequence of distributions :

$$\begin{cases} P_{Y_k} = P_{\theta_1}, & 1 \leqslant k \leqslant [N\tau^\star], \\ P_{Y_k} = P_{\theta_2}, & [N\tau^\star] < k \leqslant N. \end{cases} \qquad (\mathcal{H}_1)$$

Related issues

- Change point detection :

$$\text{test } \mathcal{H}_0 \text{ against } \mathcal{H}_1.$$

- Change point estimation $\hat{\tau}_N$, consistency :

$$\hat{\tau}_N \underset{N\to\infty}{\longrightarrow} \tau^\star.$$

- Assumption on the process (Y_k).
  - Nature of the distribution : Gaussian, i.i.d...
  - Nature of the change parameter : mean, variance...

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. 1997

Assumption on the process

Global assumptions :

- Centered Gaussian random variables.
- Change in variance parameter.

Specifically, our framework is :

- $\left(X_1^n(k)\right)_{(k,n)}$ and $\left(X_2^n(k)\right)_{(k,n)}$ two triangular matrix of random variables such that

$$(i \in \{1,2\},\ n \in \mathbb{N}) \implies X_i^n(\cdot) \text{ is a stationnary Gaussian process.}$$

- Let $\tau^\star \in]0,1[$ (the change point) and $n \in \mathbb{N}$. The observed process is

$$Y_k^n = X_1^n(k)\,\mathbb{1}_{\{k \leqslant [n\tau^\star]\}} + X_2^n(k)\,\mathbb{1}_{\{k > [n\tau^\star]\}}. \tag{1}$$

Covariance structure

As we work in a Gaussian framework, everything is specified by the covariance structure.

- For each $i \in \{1,2\}$ and $n \in \mathbb{N}$, we use the notation

$$r_i^n(l) = \mathbb{E}\left(X_i^n(k+l)\, X_i^n(k)\right), \quad i \in \{1,2\} \tag{2}$$

  and

$$\sigma_{n,i}^2 = r_i^n(0), \quad i \in \{1,2\}.$$

- The covariance between the two processes is

$$r_n(k,l) = \mathbb{E}\left(X_1^n(k)\, X_2^n(l)\right).$$

Assumption on the covariance

- **(Change of variance)** There is a real $\mathbf{a} \neq \mathbf{1}$, such that

$$\frac{\sigma^2_{n,2}}{\sigma^2_{n,1}} \xrightarrow[n\to\infty]{} a. \tag{3}$$

- **(Constraint on right and left covariances)** For all $n \in \mathbb{N}$, $r_1^n(\cdot)$ and $r_2^n(\cdot)$ belong to $\ell^2(\mathbb{Z})$. Furthermore, there exist r_1 and $r_2 \in \ell^2(\mathbb{Z})$ such that

$$\frac{r_1^n}{\sigma^2_{n,1}} \xrightarrow[n\to\infty]{\ell^2} r_1 \quad \text{and} \quad \frac{r_2^n}{\sigma^2_{n,2}} \xrightarrow[n\to\infty]{\ell^2} r_2. \tag{4}$$

- **(Constraint on mixed covariance)**

$$r_n(k,l) \leqslant C \left| r_1^n(k-l)\, r_2^n(k-l) \right|^{\frac{1}{2}}, \quad \text{if } \; k \leqslant [n\tau^\star] < l \tag{5}$$

Estimation of $\tau^\star$

In change point analysis, the change point estimator is often the *argmax* of a contrast function.

- Define the function q as

$$q(\tau) = \tau(1 - \tau).$$

- The *contrast function* is

$$J_n(\tau) = q\left(\frac{[n\tau]}{n}\right)\left(\frac{1}{[n\tau]}\sum_{k=1}^{[n\tau]}(Y_k^n)^2 - \frac{1}{n-[n\tau]}\sum_{l=[n\tau]+1}^{n}(Y_l^n)^2\right) \tag{6}$$

- The change point estimator is

$$\hat{\tau}_n = \underset{\tau\in]0,1[}{\operatorname{argmax}} \left|J_n(\tau)\right|. \tag{7}$$

Why is $\hat{\tau}_n$ a consistent estimator ?

- Asymptotically, the contrast mean $\left|\mathbb{E}\left(J_n(\tau)\right)\right|$ is

$$\left|\mathbb{E}\left(J_n(\tau)\right)\right| \underset{n\to\infty}{\sim} \begin{cases} \tau\left(1-\tau^\star\right)\left|\sigma_{n,1}^2-\sigma_{n,2}^2\right| & \tau \leqslant \tau^\star, \\ \left(1-\tau\right)\tau^\star\left|\sigma_{n,1}^2-\sigma_{n,2}^2\right| & \tau > \tau^\star. \end{cases} \tag{8}$$

  This is a *hat function* with a unique maximum at $\tau^\star$.

- Assume that we have

$$\frac{1}{\sigma_{n,1}^2}\|J_n - \mathbb{E}\left(J_n(\cdot)\right)\|_\infty \underset{n\to\infty}{\longrightarrow} 0, \tag{9}$$

  then

$$\hat{\tau}_n \underset{n\to\infty}{\longrightarrow} \underset{\tau\in]0,1[}{\operatorname{argmax}} \left|\mathbb{E}\left(J_n(\tau)\right)\right| = \tau^\star$$

Functional convergence theorem

Theorem

Each function of the problem is a random variable in the Skorohod space $D(0,1)$ (càdlàg functions). Let $\mathcal{D}$ denoted the convergence in law on this space, then

$$\frac{\sqrt{n}}{\sigma^2_{n,1}}\left(J_n(\tau) - \mathbb{E}\left(J_n(\tau)\right)\right) \xrightarrow[n\to\infty]{\mathcal{D}} B_{\tau^\star}(\tau). \quad (10)$$

Remark

- The limit process $B_{\tau^\star}$ is close to a Brownian-bridge (It is truly the case if $\tau^\star \geqslant 1$), and is continuous.
- the functional $\|\cdot\|_\infty$ is continuous on $\mathcal{C}([0,1])$.

Finally, this result leads to

$$\frac{1}{\sigma^2_{n,1}}\|J_n - \mathbb{E}\left(J_n(\cdot)\right)\|_\infty \xrightarrow[n\to\infty]{\mathbb{P}} 0. \quad (11)$$

and the consistency of $\hat{\tau}_n$.

Description of $B_{\tau^\star}$

- Recall that

$$\frac{\sigma_{n,2}^2}{\sigma_{n,1}^2} \xrightarrow[n\to\infty]{} a \neq 1 \quad \text{and} \quad \frac{r_i^n}{\sigma_{n,1}^2} \xrightarrow[n\to\infty]{\ell^2} r_i \quad i \in \{1,2\}. \tag{12}$$

- Let

$$\sigma_i^2 = 2\sum_{\mathbb{Z}} r_i(k)^2 \quad i \in \{1,2\}. \tag{13}$$

- Let W be a standard Brownian motion and define

$$W_{\tau^\star}(\tau) = \begin{cases} \sigma_1\, W(\tau) & \tau \leqslant \tau^\star, \\ \sigma_1\, W(\tau^\star) + a\,\sigma_2\,(W(\tau) - W(\tau^\star)) & \tau > \tau^\star. \end{cases} \tag{14}$$

Then

$$B_{\tau^\star}(\tau) = W_{\tau^\star}(\tau) - \tau\, W_{\tau^\star}(1). \tag{15}$$

Remark

$$(\tau^\star \geqslant 1 \;\text{ or }\; a\sigma_2 = \sigma_1) \Longrightarrow B_{\tau^\star} \text{ is a Brownian bridge.}$$

Change of Hurst exponent of fractional Brownian motion

- Let H_1 and $H_2 \in]0,1[$ be two Hurst exponents. Let W_{H_1} and W_{H_2} be two fBms.

Let $\tau^\star \in]0,1[$ and $n \in \mathbb{N}$. Assume that we observe the process

$$X_k^n = \begin{cases} W_{H_1}(\frac{k}{n}) & k \leqslant [n\tau^\star], \\ W_{H_1}(\frac{[n\tau^\star]}{n}) + W_{H_2}(\frac{k-[n\tau^\star]}{n}) & k > [n\tau^\star]. \end{cases} \quad (16)$$

We want to estimate the parameter $\tau^\star$. To this aim we use the sequence of increments

$$Y_k^n = X_k^n - X_{k-1}^n. \quad (17)$$

Remark

If $H_i < \frac{3}{4}$, then Y fulfill the set of assumptions.

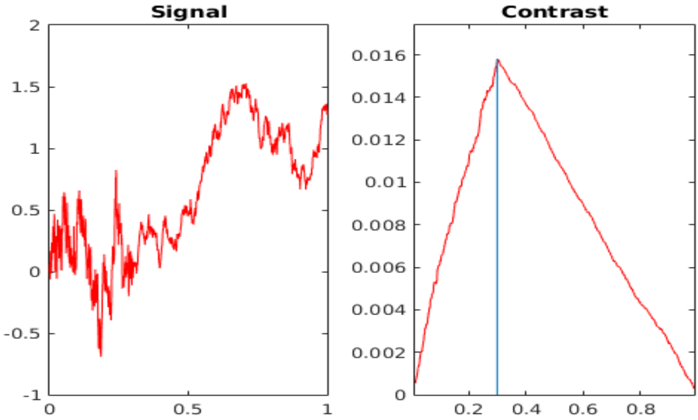

FIGURE : $N = 1024$, $H_1 = 0.3$, $H_2 = 0.5$, $\tau^\star = 0.3$, $error = 0.0012$.

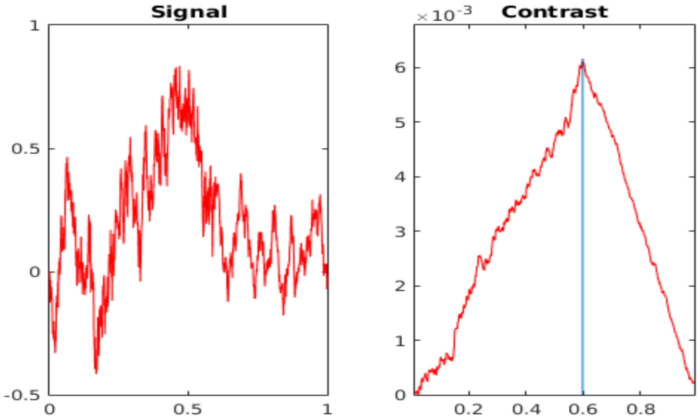

FIGURE : $N = 1024,\ \ H_1 = 0.4,\ \ H_2 = 0.5,\ \tau^\star = 0.6,\ \ error = 0.0014.$

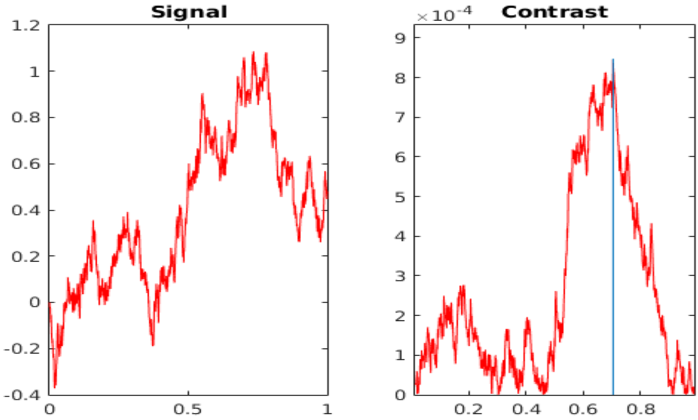

FIGURE : $N = 1024$, $H_1 = 0.49$, $H_2 = 0.5$, $\tau^\star = 0.6$, $error = -0.1051$.

Thank you for your attention.

- P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., 1999.
- J. Bardet and I. Kammoun. « Detecting abrupt changes of the long-range dependence or the self-similarity of a Gaussian process ». In: *C. R. Acad. Sci. Paris, Ser. I 340.* (2008).
- M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability and Statistics, 1997.
- M. Lavielle and C. Ludeña. « The multiple change-points problem for the spectral distribution ». In: *Bernoulli 6, no. 5* 5 (2000), pp. 845–869.
- I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus : from Stein's method to universality*. Cambridge tracts in mathematics. Cambridge University Press, 2012.
- P. R. Bertrand, M. Fhima and A. Guillin. « Off-line detection of multiple change points with the Filtered Derivative with p-Value method ». In: *Sequential Analysis* (Feb. 2010), p. 26.