On change-points tests based on two-samples U-Statistics for weakly dependent observations

Michel Harel

INSPE (Université de Limoges, France)
IMT (Université de Toulouse, France)
E-mail: michel.harel@unilim.fr

Joint work with Joseph Ngatchou-Wandji (Université de Lorraine, France),
Echarif Elharfaoui (Université Chouaib Doukkali, Maroc)

Quatrième Rencontre Poitiers - Bordeaux - ASMA 2020

Vendredi 11 décembre 2020

Outline

1 - Introduction and Motivations

- We are interested in detecting possible differences between the distributions of real-valued random variables $X_1, X_2, \ldots, X_n$.

- For any $i = 1, 2, \ldots, n$, let F_i be the cdf of X_i.
  - We aim to checking possible differences between the F_i's.
  - We restrict ourselves to checking if there exists only one index i_0 for which F_{i_0} and F_{i_0+1} are different.

- We study this problem by testing the hypothesis $\mathcal{H}_0$ against the alternative $\mathcal{H}_1$, defined respectively by

$$\mathcal{H}_0 : F_1(x) = F_2(x) = \ldots = F_n(x), \; x \in \mathbb{R}$$

$$\mathcal{H}_1 : \exists \; \lambda_0 \in (0,1) : F_1(x) = F_2(x) = \ldots = F_{[n\lambda_0]}(x) = F(x),$$
$$x \in \mathbb{R} \text{ and } F_{[n\lambda_0]+1}(x) = \ldots = F_n(x) = G(x), \; x \in \mathbb{R}, \text{ and}$$
$$\exists \; x_0 \in \mathbb{R} \text{ such that } F(x_0) \neq G(x_0) \text{ and } \theta(F,F) \neq \theta(F,G).$$

- Examples: Figure 1 exhibits the chronograms of some time series each of size 200, owning a change-point at $t = 100$.

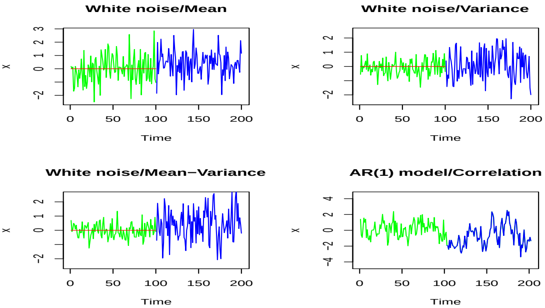

Figure 1. First row : change in the mean and change in the variance of a shifted white noise. Second row : change in both the mean and the variance of a shifted white noise, and change in the correlation of an AR(1) model.

- For cdf Q and R, denote by $\theta(Q,R)$ the following real number

$$\theta(Q,R) = \int\int h(x,y)dQ(x)dR(y).$$

- In order to evaluate the capacity of the tests to detect weak changes, we also consider the local alternatives $\mathcal{H}_{1,n}$ of the form

$$\mathcal{H}_{1,n} : \exists\ \lambda_0 \in (0,1) : F_1(x) = F_2(x) = \ldots = F_{[n\lambda_0]}(x) = F(x),$$
$$\text{and } F_{[n\lambda_0]+1}(x) = \ldots = F_n(x) = G(x),\ x \in \mathbb{R},\ \exists\ x_0 \in \mathbb{R}$$
$$\text{such that } F(x_0) \neq G(x_0) \text{ and } \theta(F,G) = \theta(F,F) + n^{-1/2}A,$$
$$\text{for some } A \in \mathbb{R}^*.$$

- Particular examples of local alternatives $\mathcal{H}_{1,n}$ are those for which there exists a constant B such that : $G(x) = F(x + n^{-1/2}B)$ and the kernel function h is twice differentiable with finite integral $\int\int(\partial h(x,y)/\partial y)dF(x)dG(y)$.
and bounded second-order derivatives $\partial^2 h(x,y)/\partial^2 y$.

- In the purpose of solving our testing problem, the tests we are going to use are based on the following statistics

$$T_{1,n} = \max_{1\le k\le n-1} \left| n^{-3/2} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left\{ h(X_i, X_j) - \theta_n(F,F) \right\} \right| \quad (1)$$

$$T_{2,n} = \frac{1}{n} \sum_{1\le k\le n-1} \left\{ n^{-3/2} \sum_{i=1}^{k} \sum_{j=k+1}^{n} \left\{ h(X_i, X_j) - \theta_n(F,F) \right\} \right\}^2 \quad (2)$$

where $\theta_n(F,F)$ is a consistent estimator.

- Denote by $[x]$ the integer part of any real number x. Noting that for any $k \in \{1, \dots, n-1\}$, there exists $\lambda_* \in [0,1]$ such that $k = [\lambda_* n]$, one can write, at least asymptotically,

$$T_{1,n} = \sup_{\lambda\in[0,1]} |Z_n(\lambda)|$$

$$T_{2,n} = \int_{\lambda\in[0,1]} {Z_n}^2(\lambda) d\lambda,$$

where Z_n stands for the following stochastic process

$$Z_n(\lambda) = n^{-3/2} \sum_{i=1}^{[n\lambda]} \sum_{j=[n\lambda]+1}^{n} \{h(X_i, X_j) - \theta_n(F, F)\}, \ 0 \le \lambda \le 1. \tag{3}$$

- The asymptotic distribution of a related process has been studied in the literature (Račkauskas and Wendler (2020), Csörgő and Horváth (1988) and by Dehling et al. (2015))
- These conditions are alleviated here and our study is done in a Skorohod space.
- Furthermore, besides the Kolmogorov-Smirnov type test usually studied in the literature
- We study a Cramer-von Mises version which has the advantage that its theoretical critical value can be approximated for any kernel h.

- We restrict our study to the classical case of one change-point detection.
- But our results can be generalized to multi-change-points detection which we postpone to a future paper.
- In Section 2, we define useful quantities such as the test statistics, and we list some assumptions.

- In Section 3 we study the asymptotic properties of our tests statistics under the null hypothesis, under a sequence of local alternatives and under fixed alternatives.
- Practical considerations are presented and discussed in Section 4.

2 - General definitions and assumptions

- Define the following *U*-statistic U_n with kernel h, and the following functions

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} h(X_i, X_j)$$

$$h_1^{(1)}(x) = \int h(x,y)dF(y) - \theta(F,F)$$

$$h_2^{(1)}(y) = \int h(x,y)dF(x) - \theta(F,F)$$

$$h_1^{(2)}(x) = \int h(x,y)dG(y) - \theta(F,G)$$

$$h_2^{(2)}(y) = \int h(x,y)dF(x) - \theta(F,G)$$

$$g^{(1)}(x,y) = h(x,y) - h_1^{(1)}(x) - h_2^{(1)}(y) + \theta(F,F)$$

$$g^{(2)}(x,y) = h(x,y) - h_1^{(2)}(x) - h_2^{(2)}(y) + \theta(F,G).$$

- Consider the Hoeffding's decomposition of U_n under $\mathcal{H}_0$

$$U_n = \theta(F, F) + U_{n,1}^{(1)} + U_{n,2}^{(1)} + U_n^{(2)}, \tag{4}$$

where

$$U_{n,1}^{(1)} = n^{-1} \sum_{i=1}^{n} h_1^{(1)}(X_i)$$

$$U_{n,2}^{(1)} = n^{-1} \sum_{i=1}^{n} h_2^{(1)}(X_i)$$

$$U_n^{(2)} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \left[h(X_i, X_j) - h_1^{(1)}(X_i) - h_2^{(1)}(X_j) \right] + \theta(F, F).$$

- Also, define the following real numbers

$$\sigma_{kl} = \mathbb{E}\left[h_k^{(1)}(X_1) h_l^{(1)}(X_1) \right] + 2 \sum_{j=1}^{\infty} \mathrm{Cov}\left(h_k^{(1)}(X_1), h_l^{(1)}(X_{1+j}) \right)$$

with $k, l = 1, 2$.

- We will assume that the sequence $\{X_i\}_{i\in\mathbb{N}}$ is absolutely regular with the rate

$$\beta(n) = \mathcal{O}(\tau^n), \quad 0 < \tau < 1, \tag{5}$$

where

$$\beta(k) = \sup_{n\in\mathbb{N}} \max_{1\leq j\leq n-k} \mathbb{E}\left[\sup_{A\in\mathcal{A}_{j+k}^{\infty}} \left|P(A \mid \mathcal{A}_0^j) - P(A)\right|\right],$$

with $\mathcal{A}_i^j$ standing for the σ-algebra generated by $X_i, \ldots, X_j, \; i, j \in \mathbb{N}\cup\{\infty\}$.

- We recall that $\{X_i\}_{i\in\mathbb{N}}$ is absolutely regular or β-mixing if $\beta(k) \longrightarrow 0$ as $n \to \infty$.

- We also assume that $\{X_i\}_{i\in\mathbb{N}}$ is stationary and ergodic.

- We consider $(Y_i)_{1 \leq i \leq n}$ a sequence of stationary ergodic and absolute regular random variables with the same rate as the sequence $(X_i)_{1 \leq i \leq n}$, that is with rate (5).
  We assume the cdf of the Y_i's is G. For any $i, j \in \mathbb{N}$, the absolute regular dependence between Y_i and Y_j is the same as the dependence between X_i and X_j.
- We assume a geometrical mixing rate by convenience.
  We believe our results can be established as well for arithmetical mixing rates to be found.

3 - Asymptotics

Theorem 1

Under $\mathcal{H}_0$, if $\max\{E\left[|h(X_i,X_j)|^{2+\delta}\right], \int\int_{\mathbb{R}^2}|h(x,y)|^{2+\delta}dF(x)dF(y)\}<\infty$ for some $\delta>0$ and the absolute regularity condition (5) is satisfied, then for any $k,l=1,2$, $\sigma_{kl}<\infty$.

If in addition $\sigma_{kl}>0$, $1\le k,l\le 2$, then the sequence of processes of $\{Z_n(\lambda); 0\le\lambda\le 1\}_{n\in\mathbb{N}}$ converges in distribution towards a zero-mean Gaussian process with representation

$$Z(\lambda)=(1-\lambda)W_1(\lambda)+\lambda(W_2(1)-W_2(\lambda)),\ 0\le\lambda\le 1,$$

where $\{W_1(\lambda),W_2(\lambda)\}_{0\le\lambda\le 1}$ is a two-dimensional zero-mean Brownian motion with covariance kernel matrix with entries $\mathrm{Cov}(W_k(s),W_l(t))=min(s,t)\sigma_{kl},\ k,l=1,2.$

Remark 1 The covariance kernel of the Gaussian process Z defined in Theorem 1 is given for all $s, t \in [0,1]$ by

$$\begin{aligned}
\Delta(s,t) &= Cov(Z(s), Z(t)) \\
&= \sigma_{11}[(1-s)(1-t)\min s,t] + \sigma_{22}[st(1-s-t+\min s,t)] \\
&+ \sigma_{12}[t(1-s)(s-\min s,t) + s(1-t)(t-\min s,t)]. \quad (6)
\end{aligned}$$

Theorem 2

Under $\mathcal{H}_{1,n}$, if
$\mathbb{E}\left[|h(X_i, X_j)|^{2+\delta}\right]$, $\mathbb{E}\left[|h(Y_i, Y_j)|^{2+\delta}\right]$, $\mathbb{E}\left[|h(X_i, Y_j)|^{2+\delta}\right]$,
$\int\int_{\mathbb{R}^2} |h(x, y)|^{2+\delta} dF(x)dF(y)$, $\int\int_{\mathbb{R}^2} |h(x, y)|^{2+\delta} dG(x)dG(y)$, and
$\int\int_{\mathbb{R}^2} |h(x, y)|^{2+\delta} dF(x)dG(y)$ are finite for some $\delta > 0$, if condition (5) holds and for any $k, l = 1, 2$, $\sigma_{kl} > 0$, then the sequence of processes $\{Z_n(\lambda); 0 \leq \lambda \leq 1\}_{n\in\mathbb{N}}$ converges in distribution towards a Gaussian process $\widetilde{Z}$ with mean $(1 - \lambda)\lambda A$ and representation

$$\widetilde{Z}(\lambda) = (1 - \lambda)\lambda A + Z(\lambda), \quad 0 \leq \lambda \leq 1,$$

where $\{Z(\lambda)\}_{0\leq\lambda\leq1}$ is the zero-mean Gaussian process defined in Theorem 1.

Theorem 3

We assume that under $\mathcal{H}_1$, the integrability conditions in Theorem 2 and condition (5) are satisfied, then

$$\frac{1}{\sqrt{n}} Z_n^*(t) \xrightarrow[n\to\infty]{a.s.} \begin{cases} \theta(F,F)t(\lambda_0 - t) + \theta(F,G)t(1-\lambda_0), & 0 \leq t \leq \lambda_0 \\ \theta(G,G)(t-\lambda_0)(1-t) + \theta(F,G)\lambda_0(1-t), & \lambda_0 \leq t < 1. \end{cases} \tag{7}$$

where

$$Z_n^*(t) = n^{-3/2} \sum_{i=1}^{[nt]} \sum_{j=[nt]+1}^{n} h(X_i, X_j),\ 0 \leq t \leq 1,$$

Theorem 4

Assume that the assumptions of Theorem 2 hold. Let $(Z(\lambda) : 0 \leq \lambda \leq 1)$ be the limiting process defined in Theorems 1 and 2, and Δ its covariance kernel. Then

i- Under $\mathcal{H}_0$, as n tends to infinity, one has the following convergence in distribution,

$$T_{1,n} \longrightarrow \sup_{\lambda \in [0,1]} |Z(\lambda)|$$

$$T_{2,n} \longrightarrow \sum_{j \geq 1} \zeta_j \chi_j^2,$$

where the χ_j^2's are iid chi-square random variables with one degree of freedom and the ζ_j's are standing for the eigen-values of the linear integral operator ∇ defined for any square integrable function τ on $[0,1]$ by

$$\nabla[\tau(\cdot)] = \int_{[0,1]} \Delta(\cdot, s)\tau(s)ds. \tag{8}$$

ii- Under $\mathcal{H}_{1,n}$, as n tends to infinity, one has the following convergence in distribution,

$$T_{1,n} \longrightarrow \sup_{\lambda \in [0,1]} |(1-\lambda)\lambda A + Z(\lambda)|$$

$$T_{2,n} \longrightarrow \sum_{j \geq 1} \zeta_j \chi_j^{*2},$$

where the χ_j^{*2}'s are iid non-central chi-square random variables with one degree of freedom and non-centrality parameters $\rho_j^2 \zeta_j^{-1}$ with the e_j's standing for the eigen-vectors of the integral operator ∇, associated with the eigen-value ζ_j, and

$$\rho_j = A \int_{[0,1]} \lambda(1-\lambda) e_j(\lambda) d\lambda.$$

iii- Under $\mathcal{H}_1$, as n tends to infinity, one has the following convergence in probability, $T_{1,n} \longrightarrow \infty, \quad T_{2,n} \longrightarrow \infty.$

Define σ by

$$\sigma = \mathrm{Var}(h_1^{(1)}(X_1)) + 2\sum_{j=1} \mathrm{Cov}(h_1^{(1)}(X_1), h_1^{(1)}(X_{1+j})).$$

Corollary 1

Assume that the assumptions of Theorem 2 hold, and that h is such that its associated $h_1^{(1)}$ and $h_2^{(1)}$ satisfy $h_1^{(1)}(x) = -h_2^{(1)}(x)$. Then

- i- Under $\mathcal{H}_0$, as n tends to infinity, one has the following convergence in distribution

$$T_{1,n} \longrightarrow \sigma \sup_{\lambda\in[0,1]} \left|W^0(\lambda)\right|$$

$$T_{2,n} \longrightarrow \sigma^2 \sum_{j\geq 1} \frac{1}{j^2\pi^2}\chi_j^2$$

ii- Under $\mathcal{H}_{1,n}$, as n tends to infinity, one has the following convergence in distribution

$$T_{1,n} \longrightarrow \sup_{\lambda\in[0,1]} \left|(1-\lambda)\lambda A + \sigma W^0(\lambda)\right|$$

$$T_{2,n} \longrightarrow \sum_{j\geq 1} \frac{1}{j^2\pi^2}\chi_j^{*2},$$

where W^0 is the Brownian bridge on $[0,1]$, the χ_j^2's and χ_j^{*2}'s are as in Theorem 4 but the non-centrality parameters are $2A^2\left\{2[1-(-1)^j]/j\pi\right\}^2\sigma^{-2}$.

Remark 2 It is easy to check that anti-symmetric kernels h are such that their associated $h_1^{(1)}$ and $h_2^{(1)}(x)$ satisfy the property $h_1^{(1)}(x) = -h_2^{(1)}(x)$.

4 - Practical considerations

- Here, we apply our results to detecting a change in the mean and/or in the variance and/or in the correlation of data from some simple models.
- We sampled 1000 sets of $n = 200$ data $X_1, X_2, \ldots, X_n$ from the model

$$X_i = \begin{cases} \varepsilon_i & i = 1, \ldots, 100 \\ \mu + \rho X_{i-1} + \omega \varepsilon_i & i = 101, \ldots, 200 \end{cases} \quad (9)$$

where μ is a real number, ω is a positive number, the ε_i's are iid and for all $i = 1, \ldots, 200$, $\varepsilon_i \sim \mathcal{N}(0, 1)$, or $\varepsilon_i \sim \mathcal{T}(3)$ (Student distribution with 3 degrees of freedom), or $\varepsilon_i = \mathcal{E}_i - 1$ with $\mathcal{E}_i \sim \mathcal{E}(1)$ ($\mathcal{E}(1)$ exponential distribution with parameter 1).

- We first apply our Kolmogorov-Smirnov and Cramér-von Mises type tests to testing $\mu = 0$ against $\mu \neq 0$ for $\omega = 1$ and $\rho = 0$ (testing a change in the mean of a shifted white noise). Next, we apply the two tests to testing $\omega = 1$ against $\omega \neq 1$ for $\mu = 0$ and $\rho = 0$ (testing change in the variance of a white noise). Finally, we consider testing $\rho = 0$ against $\rho \neq 0$ for $\mu = 0$ and $\omega = 1$ (testing a change in the correlation of an AR(1) model).

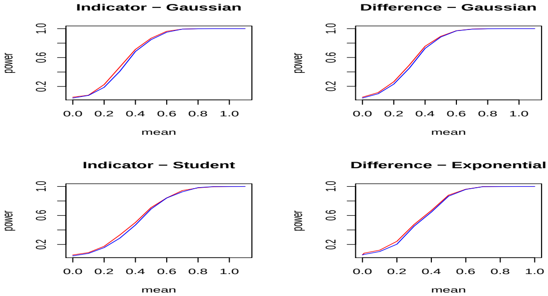

Figure 2. Empirical power of CM test (red color). Empirical power of KS test blue color. First row : change in the mean of a shifted Gaussian white noise respectively with the "indicator" and "difference" kernels. Second row : change in the mean of a shifted Student white noise with the "indicator" kernel, and change in the mean of a shifted centered exponential white noise with the "difference" kernel.

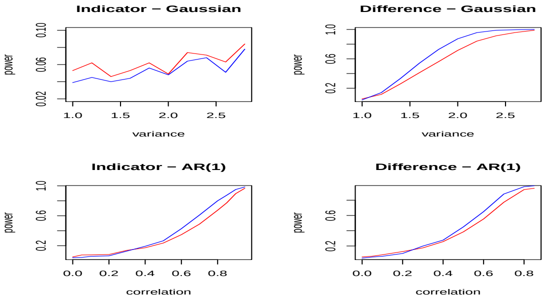

Figure 3. Empirical power of CM test (red color). Empirical power of KS test blue color. First row : change in the variance of a shifted Gaussian white noise respectively with the "indicator" and the "difference" kernels. Second row : change in the correlation of an AR(1) model respectively with the "indicator" and the "difference" kernels.

5 - Restricted bibliography

- Csörgő, M. and Horváth, L. (1988). Invariance principales for changepoint problems. *J. Multivariate Anal.* 27:151-168.
- Dehling H., Fried, R., Garcia, I., and Wendler, M. (2015). Change-Point Detection Under Dependence Based on Two-Sample U-Statistics. *Asymptotic Laws and Methods in Stochastics*. Fields Institute Communications, Springer, New York, NY, 76:195-220.
- Oodaira, H. and Yoshihara, K. (1972). Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition, *Kodai Mathematical Seminar Reports*, 24:259-269.
- Račkauskas, A. and Wendler, M. (2020). Convergence of U-processes in Hölder spaces with application to robust detection of a changed segment. *Statist. Papers*, 61:1409-1435.
- Yoshihara, K. (1976). Limiting behavior of U-statistics for stationary absolutely regular processes, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 35:237-252.

Thanks for your attention