Change-point detection method for the prediction of dreaded events during online monitoring of lung transplant patients

Nassim Sahki[1] - **Anne Gégout-Petit**[1] - **Sophie Wantz-Mézières**[1]

[1] Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France

Context

- Survival after lung transplantation is about 80% at 1 year and 50% at 6 years.
- The two main complications responsible for deaths in lung transplant patients are infection and/or rejection.

Main objective

- Test the monitoring of lung transplant patients by connected sensors ;
- Propose a methodology for real-time prediction of a serious event (infection and/ or rejection) via the change-point detection in the evolution of the multivariate signals collected by these connected sensors.

Clinical test & Health data

- AP-HP (Assistance Publique-Hôpitaux de Paris) launches the EOLE-VAL Test (duration= 2 years, observation= 6 months, patients number≃25) at Bichat Hospital.

- Health data come from the real-time medical surveillance of some **respiratory health parameters** (physiological and spirometry) of lung transplant patients by **connected objects**.

— **Physiological** : • Skin temperature • Pulse oximeter oxygen saturation (SpO_2) • Heart rate • Respiratory rate • Physical activity • Sleep quality

— **Spirometry** : • Forced Expiratory Volume in 1 second (FEV1).

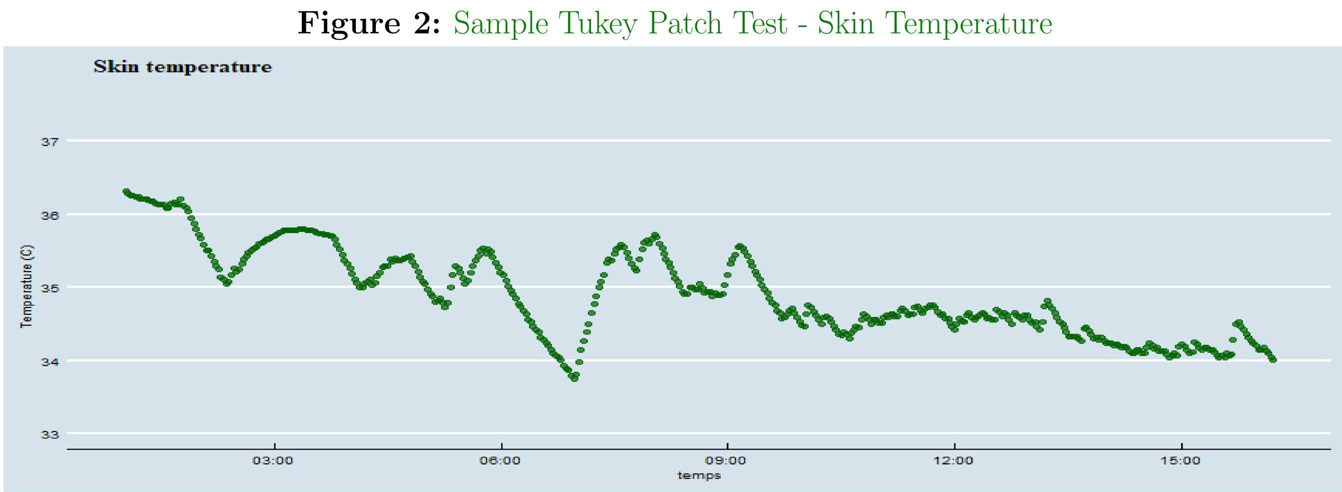
Figure 1: Connected objects

Tukey Patch Thermometer

Watch Pulse Oximeter

SmartOne Spirometer

Figure 2: Sample Tukey Patch Test - Skin Temperature



Online change-point detection

- The application context places us in the sequential framework where the series $\{x_t\}_{t=1,\dots,n} = \{x_1, \dots, x_n\}$ is sequentially observed until time n, not fixed.
- The challenge here is to minimize the average detection delay "*ADD*" while maintaining a given probability of false alarm "α".
- Statistically, the problem of change-point detection is to sequentially test for each new observation x_n, the hypotheses :

$$\begin{cases} H_{0,n} : v > n, & X_t \sim f_0(\cdot) \quad \forall\, t = 1, \dots, n \\ H_{1,n} : \exists\, v \leq n, & X_t \sim f_0(\cdot) \quad \forall\, t = 1, \dots, (v-1) \\ & X_t \sim f_1(\cdot) \quad \forall\, t = v, \dots, n \end{cases} \tag{1}$$

- Change-point detection here is based on the choice of a recursive statistic and the **threshold** it must reach to signal a detection.

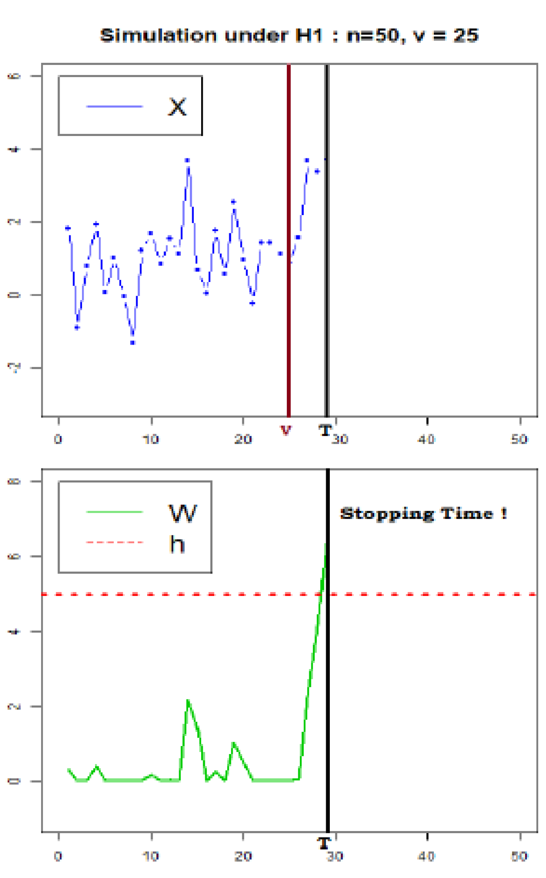
⇒ CUSUM statistics of Page based on the score S_t :

$W_t(\delta, q) = \max\{0, W_{t-1}(\delta, q) + S_t(\delta, q)\},\ t \geq 1, W_0(\delta, q) = 0$

- The score function $S_t(\delta, q; X_1, \dots, X_t)$ of Tartakovsky & al. (2012) is calculated according to the observations and the detection objective :

$\delta = (\mu_1 - \mu_0)/\sigma_0$, $q = \sigma_0/\sigma_1$ respectively the minimum change on the mean and on the variance that we want to detect. μ_0, σ_0^2 and μ_1, σ_1^2 the mean, the variance of the pre-change and the post-change regimes.

⇒ The traditional method suggested for setting a **constant threshold** is based on Wald inequality, after fixing the tolerated false alarm rate "α", while respecting : $h_\alpha \leq -\ln(\alpha)$.



⇒ Margavio & al. (1995) suggest a **conditional instantaneous threshold** by controlling the false conditional alarm rate at each instant of the trajectory.

Contribution

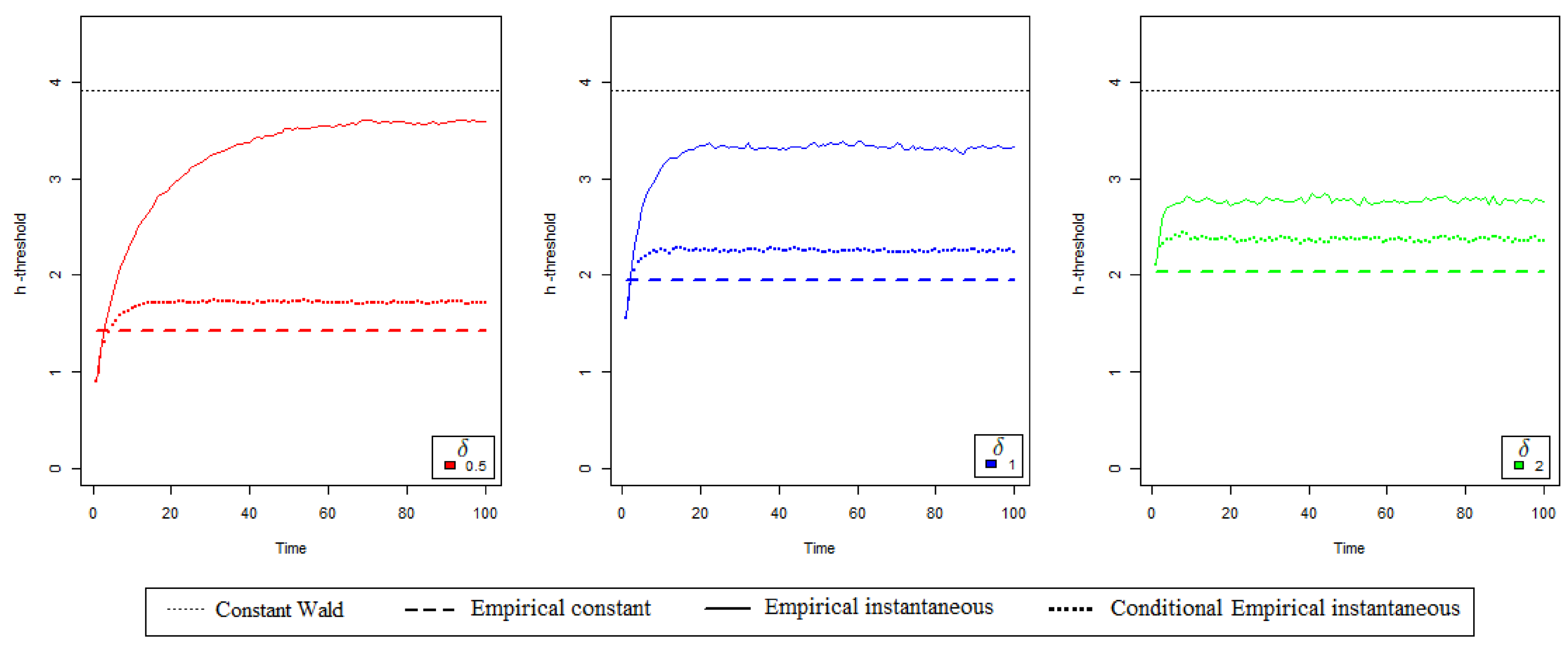
⇒ We propose new detection thresholds : the empirical constant, the empirical instantaneous and the empirical instantaneous dynamic ;

⇒ The thresholds are built by an empirical method which consists in performing simulations of the statistic $W_t(\delta, q)$ under the pre-change regime and constructing the threshold by the empirical quantile of the law of statistics, as following :

1. **Empirical constant threshold** is the quantile of the maximum values of the simulated statistics obtained along the trajectory.
2. **Empirical instantaneous threshold** is the quantile of the values of the simulated statistics obtained at each time of trajectory.
3. **Empirical instantaneous dynamic threshold** consists to use the previous instantaneous threshold and adapt it to the behavior of the statistics (data-driven). It moves in time when the statistic returns to its initial value (zero).

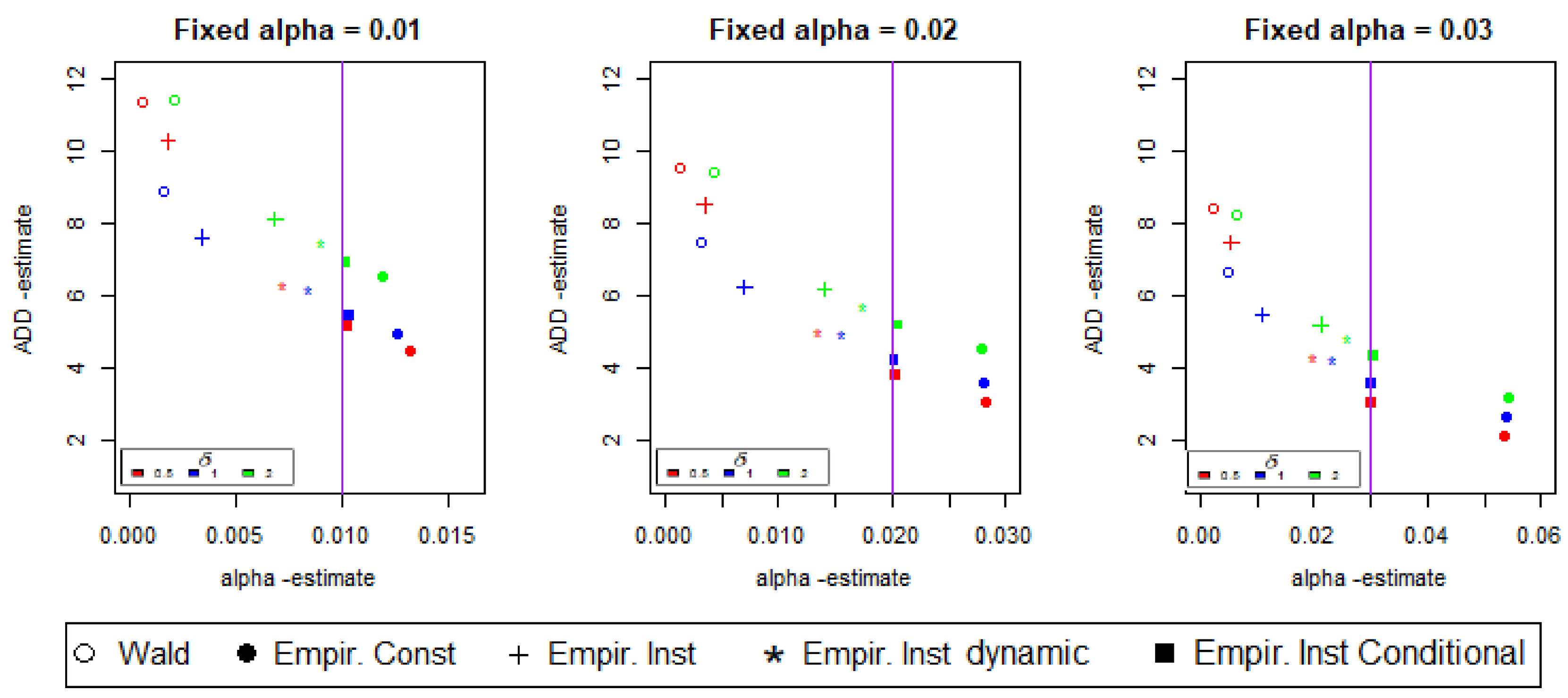
⇒ The thresholds depend on the chosen objective detection.

Figure 3: Comparison of the different empirical thresholds and that of Wald, built for $\alpha = 0.02$ and according to different detection objectives $\delta \in \{0.5, 1, 2\}$, $q = 1$, $\sigma_0 = 1$.



Thresholds performance

Figure 4: Simulation results under the pre-change regime (estimation of α) and under the post-change regime (estimation of *ADD*) obtained by the different detection thresholds and according to three different detection objectives on the mean $\delta \in \{0.5, 1, 2\}$, $q = 1$. We have the results for three different values of the tolerated false alarm rate α. The real change-point is of a level of $\delta^R = 1$.



- The results show that the empirical thresholds are faster than that of Wald.
- The best threshold is the conditional instantaneous because it makes a compromise between the detection delay and the false alarm level. It gives the best average detection delay while respecting the tolerated false alarm rate.

Perspectives

— Estimation of signal parameters (mean and variance) of the pre-change regime.

— Adaptation of the change-point detection methodology to the multivariate case.

— Application of proposed methodology to respiratory health data collected from lung transplant patients.

Références

[1] Thomas M Margavio, Michael D Conerly, William H Woodall, and Laurel G Drake. Alarm rates for quality control charts. *Statistics & Probability Letters*, 24(3) :219–224, 1995.

[2] Ewan S Page. Continuous inspection schemes. *Biometrika*, 41(1/2) :100–115, 1954.

[3] Alexander G Tartakovsky, Aleksey S Polunchenko, and Grigory Sokolov. Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing*, 7(1) :4–11, 2012.

[4] Abraham Wald. Sequential tests of statistical hypotheses. *The annals of mathematical statistics*, 16(2) :117–186, 1945.