Multivariate abrupt change detectors

Sarra Houidi, F. Auger, Houda Ben Attia Sethom, Laurence Miègeville, Dominique Fourer

Multivariate abrupt change detectors

S. Houidi [1,2], F. Auger[1], H. Ben Attia Sethom[2], L. Miègeville[1], D. Fourer[3]

[1] University of Nantes, Institut de Recherche en Énergie Électrique de Nantes Atlantique (IREENA), EA 4642, Saint-Nazaire, France

[2] University of Tunis El Manar, École Nationale d'Ingénieurs de Tunis, LR11ES15, Laboratoire des Systèmes Électriques 1002, Tunis, Tunisia

[3] University of Évry-Val-d'Essonne/Paris-Saclay, Informatique, Bio-informatique et Systèmes Complexes (IBISC), EA 4526, Courcouronnes, France

August 5, 2019

Contents

Email addresses: sarra.houidi@etu.univ-nantes.fr, francois.auger@univ-nantes.fr, houda.benattia@enicarthage.rnu.tn, laurence.miegeville@univ-nantes.fr, dominique.fourer@univ-evry.fr

1 Introduction

In this report, we aim at deriving three abrupt change detectors' decision functions: the Bayesian Information Criterion (BIC), the Cumulative Sum (CUSUM) and the Hotelling T^2 test. The three expressions are given by considering both univariate and multivariate cases. We first start by mathematically formulating the detection problem.

2 Hypothesis testing framework

Three detection algorithms capable of determining step-changes in signals are studied in this report. These changes can be detected by comparing the mean of the current observations with the mean of previous observations. Let $X_{n,p} = (x_{n_a}, x_{n_a+1}, \ldots, x_m, \ldots, x_n)$ with $x_m \in \mathbb{R}^p$, $\forall m \in \{n_a, \ldots, n\}$, be a realization of a Gaussian process corresponding to a $(p \times w)$ matrix of the last $w = n - n_a + 1$ samples of a p-dimensional time-series at the current time instant n. Each signal sample x_m corresponds to a vector of p features such as $x_m = (x_{m,1}, \ldots, x_{m,j}, \ldots, x_{m,p})^{\mathrm{T}}$, where $x_{m,j}$ is the value of feature $j \in \{1, \ldots, p\}$ at time instant m. Each vector x_m is assumed to follow a multivariate Gaussian distribution $\mathcal{N}_p(\mu, \Sigma)$, where $\mu \in \mathbb{R}^p$ is the mean vector and $\Sigma \in \mathbb{R}^{p \times p}$ is the semidefinite covariance matrix. An abrupt change occuring at a change time instant $n_a < n_c < n$ is modeled by an instantaneous modification of the statistical parameters (i.e. mean vector and/or covariance matrix). Two hypothses mut be considered:

$$\mathrm{H}_0 \;:\; x_{n_a}, \ldots, x_n \sim \mathcal{N}_p(\mu_0, \Sigma_0) \tag{1}$$

$$\begin{aligned} \mathrm{H}_1 \;:\; & x_{n_a}, \ldots, x_{n_b} \sim \mathcal{N}_p(\mu_{1a}, \Sigma_{1a}) \\ & x_{n_c}, \ldots, x_n \sim \mathcal{N}_p(\mu_{1b}, \Sigma_{1b}) \end{aligned} \tag{2}$$

with $n_b = n_c - 1$. The "without change" hypothesis H_0 supposes that, on both sides of n_c, the signal follows the same distribution (1), that is a multivariate Gaussian distribution with a mean vector μ_0 and a covariance matrix Σ_0. On the opposite, the "with change" hypothesis H_1 supposes that a change of distribution occurs at n_c (2). Before the change, the signal samples $x_{n_a}, \ldots, x_{n_b}$ follow a multivariate Gaussian distribution with a mean vector μ_{1a} and a covariance matrix Σ_{1a}. After the change, $x_{n_c}, \ldots, x_n$ follow a multivariate Gaussian distribution with a mean vector corresponding to $\mu_{1b} \neq \mu_{1a}$ and a covariance matrix equal to $\Sigma_{1b} \neq \Sigma_{1a}$. A detector performs a hypothesis test for each potential change point in a signal. An on-line approach is followed by using a sliding window over the signal. Therefore, at each time instant n, a decision between H_0 and H_1 (i.e. a decision "to reject H_0 in favor of H_1") is made by comparing a decision function g_n to an a dimensional threshold h [1].

$$\text{decide } \mathrm{H}_1 \quad \text{if} \quad g_n > h \tag{3}$$

$$\text{decide } \mathrm{H}_0 \quad \text{if} \quad g_n \leq h \tag{4}$$

In what follows three abrupt change detectors in their univariate and multivariate versions are studied: the BIC, the CUSUM and the Hotelling T^2 test.

3 Detectors' decision functions

In what follows, each decision function is presented in both, the multivariate and the univariate cases. For each detector, we assume that the samples x_m are taken from independent and identically distributed ($i.i.d$) Gaussian random vectors. It is also assumed that an abrupt change may occur at $n_c = n - \frac{w}{2} + 1$, the estimated covariance matrices are invertible and the sliding window length is $w \geq 2(p+1)$.

3.1 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) of $X_{n,p}$ under hypothesis H_i, $i \in \{0,1\}$ is defined as a maximum likelihood criterion penalized by the model complexity [3, 4], proportional to the number M_i of free parameters to be estimated:

$$\mathrm{BIC}_n(\mathrm{H_i}) = \max_{\mu,\Sigma} \ln(\mathcal{L}_{\mathrm{n,i}}) - \frac{\lambda}{2} \mathrm{M_i} \ln(\mathrm{w}) \tag{5}$$

where $\mathcal{L}_{n,i}$ is the data likelihood function under hypothesis $\mathrm{H_i}$, $i \in \{0,1\}$ defined as the joint Probability Density Function (PDF) of the observed data and considered as a function of the statistical parameters μ and Σ. The scalar λ is a penalty factor, ideally equal to 1 [3].

3.1.1 Criterion derivation for a multivariate case

For a multi dimensional signal, under H_0, the number of free parameters M_0 in (5) corresponds to the sum of the dimension p of the mean vector μ_0 plus the $p(p+1)/2$ variances and covariances to be estimated from the symmetric $(p \times p)$ covariance matrix Σ_0, resulting in $M_0 = p(p+3)/2$. Under H_1, the number of free parameters are the p ones from the $(p \times 1)$ mean vectors μ_{1a}, and the p ones from the $(p \times 1)$ mean vector μ_{1b} plus the $p(p+1)/2$ ones from the $(p \times p)$ covariance matrices Σ_{1a}, and the $p(p+1)/2$ ones from Σ_{1b}, so $M_1 = p(p+3) = 2M_0$.

Assuming that the samples x_m are taken from *i.i.d* Gaussian random vectors, the likelihood functions of $X_{n,p}$ under H_0 and H_1 correspond to the joint PDFs which are, by independance of the x_m values, equal to the product of the samples' PDFs, such as:

$$\mathcal{L}_{n,0} = \prod_{m=n_a}^{n} p_{\mu_0,\Sigma_0}(x_m) \tag{6}$$

$$= \prod_{m=n_a}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}} \det(\Sigma_0)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_m-\mu_0)^T\Sigma_0^{-1}(x_m-\mu_0)\right) \tag{7}$$

$$\mathcal{L}_{n,1} = \prod_{m=n_a}^{n_b} p_{\mu_{1a},\Sigma_{1a}}(x_m) \prod_{m=n_c}^{n} p_{\mu_{1b},\Sigma_{1b}}(x_m) \tag{8}$$

$$= \prod_{m=n_a}^{n_b} \frac{1}{(2\pi)^{\frac{p}{2}}} \det(\Sigma_{1a})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_m-\mu_{1a})^T\Sigma_{1a}^{-1}(x_m-\mu_{1a})\right) \tag{9}$$

$$\times \prod_{m=n_c}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}} \det(\Sigma_{1b})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_m-\mu_{1b})^T\Sigma_{1b}^{-1}(x_m-\mu_{1b})\right) \tag{10}$$

The two expressions are maximized when considering the Maximum Likelihood Estimators (MLEs) of the mean vectors and covariance matrices [3, 4]:

$$(\hat{\mu}_0, \hat{\Sigma}_0) = \underset{\mu_0,\Sigma_0}{\operatorname{argmax}} \mathcal{L}_{n,0} \quad \text{and} \quad (\hat{\mu}_{1a}, \hat{\mu}_{1b}, \hat{\Sigma}_{1a}, \hat{\Sigma}_{1b}) = \underset{\mu_{1a},\mu_{1b},\Sigma_{1a},\Sigma_{1b}}{\operatorname{argmax}} \mathcal{L}_{n,1} \tag{11}$$

such as:

$$\hat{\mu}_{1a} = \frac{1}{n_c-n_a}\sum_{m=n_a}^{n_b} x_m, \qquad \hat{\Sigma}_{1a} = \frac{1}{n_c-n_a}\sum_{m=n_a}^{n_b} (x_m-\hat{\mu}_{1a})(x_m-\hat{\mu}_{1a})^{\mathrm{T}} \tag{12}$$

$$\hat{\mu}_{1b} = \frac{1}{n-n_b}\sum_{m=n_c}^{n} x_m, \qquad \hat{\Sigma}_{1b} = \frac{1}{n-n_b}\sum_{m=n_c}^{n} (x_m-\hat{\mu}_{1b})(x_m-\hat{\mu}_{1b})^{\mathrm{T}} \tag{13}$$

$$\hat{\mu}_0 = \frac{1}{w}\sum_{m=n_a}^{n} x_m = \frac{n_c-n_a}{w}\hat{\mu}_{1a} + \frac{n-n_b}{w}\hat{\mu}_{1b}, \qquad \hat{\Sigma}_0 = \frac{1}{w}\sum_{m=n_a}^{n} (x_m-\hat{\mu}_0)(x_m-\hat{\mu}_0)^{\mathrm{T}} \tag{14}$$

Then, we have:

$$\mathrm{BIC_n(H_0)} = \ln(\mathcal{L}_{n,0}) - \frac{\lambda}{4}p(p+3)\ln(w) \tag{15}$$

$$= \ln\left(\left(\frac{1}{(2\pi)^p\det(\hat{\Sigma}_0)}\right)^{\frac{w}{2}} \exp\left(-\frac{1}{2}\sum_{m=n_a}^{n}(x_m-\hat{\mu}_0)^{\mathrm{T}}\hat{\Sigma}_0^{-1}(x_m-\hat{\mu}_0)\right)\right) - \frac{\lambda}{4}p(p+3)\ln(w) \tag{16}$$

$$= -\frac{1}{2}\sum_{m=n_a}^{n}(x_m-\hat{\mu}_0)^T\hat{\Sigma}_0^{-1}(x_m-\hat{\mu}_0) - \frac{wp}{2}\ln(2\pi) - \frac{w}{2}\ln(\det(\hat{\Sigma}_0)) - \frac{\lambda}{4}p(p+3)\ln(w) \tag{17}$$

and

$$\mathrm{BIC_n(H_1)} = \ln(\mathcal{L}_{n,1}) - \frac{\lambda}{2}p(p+3)\ln(w) \tag{18}$$

$$= \ln\left(\left(\frac{1}{(2\pi)^p\det(\hat{\Sigma}_{1a})}\right)^{\frac{n_c-n_a}{2}} \left(\frac{1}{(2\pi)^p\det(\hat{\Sigma}_{1b})}\right)^{\frac{n-n_b}{2}} \exp\left(-\frac{1}{2}\left(\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^{\mathrm{T}}\hat{\Sigma}_{1a}^{-1}(x_m-\hat{\mu}_{1a})\right.\right.\right.$$

$$\left.\left.\left. + \sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})^{\mathrm{T}}\hat{\Sigma}_{1b}^{-1}(x_m-\hat{\mu}_{1b})\right)\right)\right) - \frac{\lambda}{2}p(p+3)\ln(w) \tag{19}$$

$$\tag{20}$$

Since:

$$\sum_{m=n_a}^{n} (x_m - \hat{\mu}_0)^{\mathrm{T}}\hat{\Sigma}_0^{-1}(x_m - \hat{\mu}_0) = \sum_{m=n_a}^{n} \mathrm{tr}\left[(x_m - \hat{\mu}_0)^{\mathrm{T}}\hat{\Sigma}_0^{-1}(x_m - \hat{\mu}_0)\right] \quad (21)$$

$$= \sum_{m=n_a}^{n} \mathrm{tr}\left[\hat{\Sigma}_0^{-1}(x_m - \hat{\mu}_0)(x_m - \hat{\mu}_0)^{\mathrm{T}}\right] \quad (22)$$

$$= \mathrm{tr}\left[\hat{\Sigma}_0^{-1}\sum_{m=n_a}^{n} (x_m - \hat{\mu}_0)(x_m - \hat{\mu}_0)^{\mathrm{T}}\right] \quad (23)$$

$$= \mathrm{tr}\left[w\mathrm{I}\right] = wp \quad (24)$$

likewise for $\sum_{m=n_a}^{n_b} (x_m - \hat{\mu}_{1a})^{\mathrm{T}}\hat{\Sigma}_{1a}^{-1}(x_m - \hat{\mu}_{1a})$ and $\sum_{m=n_c}^{n} (x_m - \hat{\mu}_{1b})^{\mathrm{T}}\hat{\Sigma}_{1b}^{-1}(x_m - \hat{\mu}_{1b})$, then the BIC values under H_0 and H_1 are respectively:

$$\mathrm{BIC}_n(H_0) = -\frac{wp}{2}\ln(2\pi) - \frac{w}{2}\ln(\det(\hat{\Sigma}_0)) - \frac{wp}{2} - \frac{\lambda}{4}p(p+3)\ln(w) \quad (25)$$

$$\mathrm{BIC}_n(H_1) = -\frac{wp}{2}\ln(2\pi) - \frac{(n_c - n_a)}{2}\ln(\det(\hat{\Sigma}_{1a})) - \frac{(n - n_b)}{2}\ln(\det(\hat{\Sigma}_{1b})) - \frac{wp}{2} - \frac{\lambda}{2}p(p+3)\ln(w) \quad (26)$$

The BIC variation for a given value of n_c is given by:

$$\Delta\mathrm{BIC}_n = \mathrm{BIC_n(H_1)} - \mathrm{BIC_n(H_0)} \quad (27)$$

$$= \frac{w}{2}\ln(\det(\hat{\Sigma}_0)) - \frac{(n_c - n_a)}{2}\ln(\det(\hat{\Sigma}_{1a})) - \frac{(n - n_b)}{2}\ln(\det(\hat{\Sigma}_{1b})) - \frac{\lambda}{4}p(p+3)\ln(w) \quad (28)$$

If $\Delta\mathrm{BIC}_n > 0$, the model of two Gaussians is favored (i.e. the signal can be segmented into two parts at n_c). Consequently, the decision function can be expressed as:

$$g_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h \quad \text{with} \quad g_n = \Delta\mathrm{BIC_n}, \quad (29)$$

$$= \ln\left(\frac{\det(\hat{\Sigma}_0)^{\frac{w}{2}}}{\det(\hat{\Sigma}_{1a})^{\frac{n_c - n_a}{2}}\det(\hat{\Sigma}_{1b})^{\frac{n - n_b}{2}}}\right) - \frac{\lambda}{4}p(p+3)\ln(w) \quad (30)$$

Since we consider that an abrupt change occurs at $n_c = n - \frac{w}{2} + 1$, the decision function becomes:

$$g'_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h' \quad \text{with} \quad g'_n = \ln\left(\frac{\det(\hat{\Sigma}_0)^2}{\det(\hat{\Sigma}_{1a})\det(\hat{\Sigma}_{1b})}\right) \quad \text{and} \quad h' = \frac{4}{w}h + \frac{\lambda}{w}p(p+3)\ln(w) \quad (31)$$

3.1.2 Criterion derivation for a scalar case

For a one dimensional signal, the number of free parameters under H_0 corresponds to μ_0 and σ_0, so $M_0 = 2$. Under H_1, the number of free parameters is $\mu_{1a}, \sigma_{1a}, \mu_{1b}, \sigma_{1b}$, so $M_1 = 4$. The BIC criterion derivation for a one dimensional signal is given in [?] and corresponds to

$$g'_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h' \quad \text{with} \quad g'_n = \ln\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_{1a}\hat{\sigma}_{1b}}\right), \; h' = \frac{4}{w}\left(h + \lambda\ln(w)\right) \quad (32)$$

3.2 CUmulative Sum (CUSUM) algorithm

The CUSUM algorithm is a statistical test for the detection of a mean change in a Gaussian process [1]. It involves the calculation of a cumulative sum and works by tracking its deviations from a threshold value.

3.2.1 Criterion derivation for a multivariate case

In the case of a multi dimensional signal, when considering (1) and (2), in order to derive the CUSUM decision function, it is assumed that under H_0, $\mu_0 = \mu_{1a}$ and $\Sigma_0 = \Sigma_{1a}$. Under H_1, $\Sigma_{1b} = \Sigma_{1a}$. The decision rule is based on a maximization of the log-likelihood ratio such as:

$$g_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h, \quad \text{with} \quad g_n = \sum_{m=n_c}^{n} s_m \tag{33}$$

Indeed, the log likelihood ration over n_c corresponds to the cumulative sum of values s_m such that:

$$\begin{aligned}
\ln\left(\frac{\mathcal{L}_{n,1}}{\mathcal{L}_{n,0}}\right) &= \ln\left(\frac{\prod_{m=n_a}^{n_b} p_{\mu_{1a},\Sigma_{1a}}(x_m) \prod_{m=n_c}^{n} p_{\mu_{1b},\Sigma_{1a}}(x_m)}{\prod_{m=n_a}^{n} p_{\mu_{1a},\Sigma_{1a}}(x_m)}\right) && (34)\\
&= \sum_{m=n_c}^{n} \ln\left(\frac{p_{\mu_{1b},\Sigma_{1a}}(x_m)}{p_{\mu_{1a},\Sigma_{1a}}(x_m)}\right) && (35)\\
&= \sum_{m=n_c}^{n} s_m && (36)
\end{aligned}$$

Because x_m are i.i.d $\mathcal{N}_p(\mu, \Sigma)$, s_m becomes:

$$\begin{aligned}
s_m &= \ln\left(\frac{(2\pi)^{-\frac{p}{2}} \det(\Sigma_{1a})^{-\frac{1}{2}} \exp(-\frac{1}{2}(x_m - \mu_{1b})^{\mathrm{T}}\Sigma_{1a}^{-1}(x_m - \mu_{1b}))}{(2\pi)^{-\frac{p}{2}} \det(\Sigma_{1a})^{-\frac{1}{2}} \exp(-\frac{1}{2}(x_m - \mu_{1a})^{\mathrm{T}}\Sigma_{1a}^{-1}(x_m - \mu_{1a}))}\right) && (37)\\
&= \ln\left(\frac{\exp(-\frac{1}{2}(x_m - \mu_{1b})^{\mathrm{T}}\Sigma_{1a}^{-1}(x_m - \mu_{1b}))}{\exp(-\frac{1}{2}(x_m - \mu_{1a})^{\mathrm{T}}\Sigma_{1a}^{-1}(x_m - \mu_{1a}))}\right) && (38)\\
&= \frac{1}{2}(x_m - \mu_{1a})^{\mathrm{T}}\Sigma_{1a}^{-1}(x_m - \mu_{1a}) - \frac{1}{2}(x_m - \mu_{1b})^{\mathrm{T}}\Sigma_{1a}^{-1}(x_m - \mu_{1b}) && (39)\\
&= -\frac{1}{2}x_m^{\mathrm{T}}\Sigma_{1a}^{-1}\mu_{1a} - \frac{1}{2}\mu_{1a}^{\mathrm{T}}\Sigma_{1a}^{-1}x_m + \frac{1}{2}\mu_{1a}^{\mathrm{T}}\Sigma_{1a}^{-1}\mu_{1a} + \frac{1}{2}x_m^{\mathrm{T}}\Sigma_{1a}^{-1}\mu_{1b} + \frac{1}{2}\mu_{1b}^{\mathrm{T}}\Sigma_{1a}^{-1}x_m - \frac{1}{2}\mu_{1b}^{\mathrm{T}}\Sigma_{1a}^{-1}\mu_{1b} && (40)
\end{aligned}$$

Since:

$$x_m^T\Sigma_{1a}^{-1}\mu_{1b} = \mu_{1b}^T\Sigma_{1a}^{-1}x_m \quad \text{and} \quad x_m^T\Sigma_{1a}^{-1}\mu_{1a} = \mu_{1a}^T\Sigma_{1a}^{-1}x_m \tag{41}$$

Then:

$$\begin{aligned}
s_m &= \frac{1}{2}\left[2\mu_{1b}^{\mathrm{T}}\Sigma_{1a}^{-1}x_m - 2\mu_{1a}^{\mathrm{T}}\Sigma_{1a}^{-1}x_m + (\mu_{1a}^{\mathrm{T}}\Sigma_{1a}^{-1}\mu_{1a} - \mu_{1b}^{\mathrm{T}}\Sigma_{1a}^{-1}\mu_{1b})\right] && (42)\\
&= (\mu_{1b} - \mu_{1a})^{\mathrm{T}}\Sigma_{1a}^{-1}x_m - \frac{1}{2}(\mu_{1b} + \mu_{1a})^{\mathrm{T}}\Sigma_{1a}^{-1}(\mu_{1b} - \mu_{1a}) && (43)\\
&= (\mu_{1b} - \mu_{1a})^{\mathrm{T}}\Sigma_{1a}^{-1}\left(x_m - \frac{\mu_{1b} + \mu_{1a}}{2}\right) && (44)
\end{aligned}$$

To calculate s_m, all the unknowns in its expression, such as the mean vectors $\hat{\mu}_{1b}$, $\hat{\mu}_{1a}$ and the covariance matrix $\hat{\Sigma}_{1a}$, are replaced by their MLEs expressed in (12), (13). Then, according to (33), the CUSUM decision function becomes:

$$\begin{aligned}
g_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h \quad \text{with} \quad g_n &= (\hat{\mu}_{1b} - \hat{\mu}_{1a})^T\hat{\Sigma}_{1a}^{-1}\left(\sum_{m=n_c}^{n} x_m - \frac{\hat{\mu}_{1b} + \hat{\mu}_{1a}}{2}\right) && (45)\\
&= (\hat{\mu}_{1b} - \hat{\mu}_{1a})^T\hat{\Sigma}_{1a}^{-1}\left((n_c - n + 1)\hat{\mu}_{1b} - (n_c - n + 1)\frac{\hat{\mu}_{1b} + \hat{\mu}_{1a}}{2}\right) && (46)\\
&= \frac{(n - n_b)}{2}(\hat{\mu}_{1b} - \hat{\mu}_{1a})^T\hat{\Sigma}_{1a}^{-1}(\hat{\mu}_{1b} - \hat{\mu}_{1a}) && (47)
\end{aligned}$$

If considering an abrupt change occurring inside a sliding window of w samples at $n_c = n - \frac{w}{2} + 1$, then the decision function can be expressed as follows, when considering (14) and (12):

$$g'_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h' \quad \text{where} \quad g_n = (\hat{\mu}_{1b} - \hat{\mu}_{1a})^T\hat{\Sigma}_{1a}^{-1}(\hat{\mu}_{1b} - \hat{\mu}_{1a}) \quad \text{and} \quad h' = \frac{4}{w}h \tag{48}$$

3.2.2 Criterion derivation for a scalar case

The CUSUM criterion derivation for a one dimensional signal is given in [?] and corresponds to:

$$g_n = \sum_{m=n_c}^{n} s_m \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h \quad \text{with} \quad s_m = (n - n_b)\frac{(\hat{\mu}_{1b} - \hat{\mu}_{1a})^2}{2\hat{\sigma}_{1a}^2} \tag{49}$$

Since we aim at detecting a mean change at $n_c = n - \frac{w}{2} + 1$, the CUSUM decision function corresponds to:

$$g'_n = \frac{(\hat{\mu}_{1b} - \hat{\mu}_{1a})^2}{\hat{\sigma}_{1a}^2} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} h' \quad \text{where} \quad h' = \frac{4}{w}h \tag{50}$$

3.3 Hotelling T² test

When testing for the difference between the means of two normally distributed samples with unknown variances but assumed equal, the most commonly used statistical test is the Hotelling T^2 test [5]. It aims at quantifying the difference between two Normal distributions using the mean and variance in the data.

3.3.1 Criterion derivation for a multivariate case

The Hotelling T^2 test is the multivariate extension of the Student's t-test [6]. In a Hotelling T^2 test, the difference between the $(p \times 1)$ mean vectors of two samples is considered [7]. In order to formulate the Hotelling T^2 test, we need to derive the maximum likelihood ratio test for both hypotheses H_0 and H_1. The likelihood ratio corresponds to:

$$\Lambda_n = \frac{\mathcal{L}_{n,1}}{\mathcal{L}_{n,0}} \tag{51}$$

which is maximized when considering the MLEs in (14), (12) and (13), such as:

$$\boldsymbol{\Lambda}_n = \frac{\left(\frac{1}{(2\pi)^p \det(\hat{\Sigma}_1)}\right)^{\frac{w}{2}} \exp\left(-\frac{1}{2}\left(\sum_{m=n_a}^{n_b}(x_m - \hat{\mu}_{1a})^{\mathrm{T}}\hat{\Sigma}_1^{-1}(x_m - \hat{\mu}_{1a}) + \sum_{m=n_c}^{n}(x_m - \hat{\mu}_{1b})^{\mathrm{T}}\hat{\Sigma}_1^{-1}(x_m - \hat{\mu}_{1b})\right)\right)}{\left(\frac{1}{(2\pi)^p \det(\hat{\Sigma}_0)}\right)^{\frac{w}{2}} \exp\left(-\frac{1}{2}\sum_{m=n_a}^{n}(x_m - \hat{\mu}_0)^{\mathrm{T}}\hat{\Sigma}_0^{-1}(x_m - \hat{\mu}_0)\right)} \tag{52}$$

where:

$$\hat{\Sigma}_1 = \frac{1}{w}\left(\sum_{m=n_a}^{n_b}(x_m - \hat{\mu}_{1a})(x_m - \hat{\mu}_{1a})^{\mathrm{T}} + \sum_{m=n_c}^{n}(x_m - \hat{\mu}_{1b})(x_m - \hat{\mu}_{1b})^{\mathrm{T}}\right) \tag{53}$$

$$= \frac{n_c - n_a}{w}\hat{\Sigma}_{1a} + \frac{n - n_b}{w}\hat{\Sigma}_{1b} \tag{54}$$

$$\tag{55}$$

Then (52) can be simplified using (24), such as:

$$\boldsymbol{\Lambda}_n = \left(\frac{\det(\hat{\Sigma}_0)}{\det(\hat{\Sigma}_1)}\right)^{\frac{w}{2}} \tag{56}$$

$$= \left(\frac{\det\left[\frac{1}{w}\sum_{m=n_a}^{n}(x_m - \hat{\mu}_0)(x_m - \hat{\mu}_0)^{\mathrm{T}}\right]}{\det\left[\frac{1}{w}\left(\sum_{m=n_a}^{n_b}(x_m - \hat{\mu}_{1a})(x_m - \hat{\mu}_{1a})^{\mathrm{T}} + \sum_{m=n_c}^{n}(x_m - \hat{\mu}_{1b})(x_m - \hat{\mu}_{1b})^{\mathrm{T}}\right)\right]}\right)^{\frac{w}{2}} \tag{57}$$

The above expression can be simplified, indeed:

$$\sum_{m=n_a}^{n_b}(x_m - \hat{\mu}_0)(x_m - \hat{\mu}_0)^{\mathrm{T}} = \sum_{m=n_a}^{n_b}(x_m - \hat{\mu}_{1a})(x_m - \hat{\mu}_{1a})^{\mathrm{T}} + (n_c - n_a)(\hat{\mu}_{1a} - \hat{\mu}_0)(\hat{\mu}_{1a} - \hat{\mu}_0)^{\mathrm{T}} \tag{58}$$

$$= \sum_{m=n_a}^{n_b}(x_m - \hat{\mu}_{1a})(x_m - \hat{\mu}_{1a})^{\mathrm{T}} + \frac{(n_c - n_a)(n - n_b)^2}{w^2}(\hat{\mu}_{1a} - \hat{\mu}_{1b})(\hat{\mu}_{1a} - \hat{\mu}_{1b})^{\mathrm{T}} \tag{59}$$

Liwewise, we have:

$$\sum_{m=n_c}^{n}(x_m-\hat{\mu}_0)(x_m-\hat{\mu}_0)^{\mathrm{T}} = \sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})(x_m-\hat{\mu}_{1b})^{\mathrm{T}} + \frac{(n-n_b)(n_c-n_a)^2}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})(\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}} \quad (60)$$

According to (54), we obtain:

$$\hat{\Sigma}_0 = \hat{\Sigma}_1 + \frac{(n_c-n_a)(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})(\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}} \quad (61)$$

Therefore, when substituting (61) in (57), we have:

$$\begin{aligned}
\mathbf{\Lambda}_n &= \left(\frac{\det(\hat{\Sigma}_1 + \frac{(n_c-n_a)(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})(\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}})}{\det(\hat{\Sigma}_1)}\right)^{\frac{w}{2}} && (62)\\
&= \left(\frac{\frac{w^2}{(n_c-n_a)(n-n_b)}\det(\hat{\Sigma}_1 + \frac{(n_c-n_a)(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})(\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}})}{\frac{w^2}{(n_c-n_a)(n-n_b)}\det(\hat{\Sigma}_1)}\right)^{\frac{w}{2}} && (63)\\
&= \left(\frac{\det(\hat{\Sigma}_1)\left(\frac{w^2}{(n_c-n_a)(n-n_b)} + (\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}}\hat{\Sigma}_1^{-1}(\hat{\mu}_{1b}-\hat{\mu}_{1a})\right)}{\frac{w^2}{(n_c-n_a)(n-n_b)}\det(\hat{\Sigma}_1)}\right)^{\frac{w}{2}} && (64)\\
&= \left(1 + \frac{(n_c-n_a)(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}}\hat{\Sigma}_1^{-1}(\hat{\mu}_{1b}-\hat{\mu}_{1a})\right)^{\frac{w}{2}} && (65)\\
&= (1+g_n)^{\frac{w}{2}} && (66)
\end{aligned}$$

Note, when considering $u = \frac{w^2}{(n_c-n_a)(n-n_b)}$, $V = \hat{\Sigma}_1$, $\vec{w} = (\hat{\mu}_{1b}-\hat{\mu}_{1a})$:

$$\begin{aligned}
A &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} u & \vec{w}^T \\ -\vec{w} & V \end{bmatrix} && (67)\\
\det(A) &= \left\{ \begin{array}{l} \det(A_{11})\det(A_{22}-A_{21}A_{11}^{-1}A_{12}) \\ \det(A_{22})\det(A_{11}-A_{12}A_{22}^{-1}A_{21}) \end{array} \right. = \left\{ \begin{array}{l} u\det(V+\frac{1}{u}\vec{w}\vec{w}^T) \\ \det(V)(u+\vec{w}^TV^{-1}\vec{w}) \end{array} \right. && (68)
\end{aligned}$$

Thus,

$$\begin{aligned}
u\det(V+\frac{1}{u}\vec{w}\vec{w}^T) &= \det(V)(u+\vec{w}^TV^{-1}\vec{w}) && (69)\\
\frac{\det(V+\frac{1}{u}\vec{w}\vec{w}^T)}{\det(V)} &= 1+\frac{\vec{w}^TV^{-1}\vec{w}}{u} && (70)
\end{aligned}$$

Informally, this likelihood ratio aims at measuring the plausibility of H_0 relative to H_1. Therefore, if the likelihood ratio is sufficiently small, we might be inclined to reject H_0. According to Neyman Pearson Lemma [6], this is made possible by setting (66) less than $\gamma \in [0,1]$. Indeed, under H_0, we have $\mu_{1a} = \mu_{1b}$, so $\mathbf{\Lambda}_n = 1$ and under H_1, we have $\mu_{1a} \neq \mu_{1b}$, so $\mathbf{\Lambda}_n > 1$. This leads to:

$$g_n > \gamma^{\frac{2}{w}} - 1 = h \quad (71)$$

g_n corresponds to the decision function that is compared to an adjusted threshold h such as:

$$g_n \underset{H_0}{\overset{H_1}{\gtrless}} h \quad \text{with} \quad g_n = \left(\frac{w}{n_c-n_a} + \frac{w}{n-n_b}\right)^{-1}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}}\hat{\Sigma}_1^{-1}(\hat{\mu}_{1b}-\hat{\mu}_{1a}) \quad (72)$$

According to (54), the decision function g_n can also be expressed as:

$$g'_n \underset{H_0}{\overset{H_1}{\gtrless}} h' = h \times \frac{w^2}{(n_c-n_a)(n-n_b)} \quad (73)$$

$$\text{with} \quad g'_n = (\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}}(\frac{n_c-n_a}{w}\hat{\Sigma}_{1a} + \frac{n-n_b}{w}\hat{\Sigma}_{1b})^{-1}(\hat{\mu}_{1b}-\hat{\mu}_{1a}) \quad (74)$$

If an abrupt change occurs inside a sliding window of w samples at $n_c = n - \frac{w}{2} + 1$, the Hotelling's T^2 decision rules becomes:

$$g''_n \underset{H_0}{\overset{H_1}{\gtrless}} h'' = h' \times 2 \quad \text{with} \quad g'_n = (\hat{\mu}_{1b}-\hat{\mu}_{1a})^{\mathrm{T}}(\hat{\Sigma}_{1a}+\hat{\Sigma}_{1b})^{-1}(\hat{\mu}_{1b}-\hat{\mu}_{1a}) \quad (75)$$

3.3.2 Criterion derivation for a scalar case: the Student t-test

The student's t-test is a generalization of Hotelling's T^2 statistic, used in univariate hypothesis testing. As for the multivariate case, we need to derive the maximum likelihood ratio test which corresponds to:

$$\mathbf{\Lambda}_n = \frac{\mathcal{L}_{n,1}}{\mathcal{L}_{n,0}} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{\frac{w}{2}} \tag{76}$$

$$= \left(\frac{\frac{1}{w}\left(\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_0)^2+\sum_{m=n_c}^{n}(x_m-\hat{\mu}_0)^2\right)}{\frac{1}{w}\left(\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2+\sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})^2\right)}\right)^{\frac{w}{2}} \tag{77}$$

The above expression can be simplified, indeed:

$$\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_0)^2 = \sum_{m=n_a}^{n_b}((x_m-\hat{\mu}_{1a})+(\hat{\mu}_{1a}-\hat{\mu}_0))^2 \tag{78}$$

$$= \sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2+(n_c-n_a)(\hat{\mu}_{1a}-\hat{\mu}_0)^2 \tag{79}$$

$$= \sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2+(n_c-n_a)\left(\hat{\mu}_{1a}-\frac{(n_c-n_a)\hat{\mu}_{1a}+(n-n_b)\hat{\mu}_{1b}}{w}\right)^2 \tag{80}$$

$$= \sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2+\frac{(n_c-n_a)(n-n_b)^2}{w^2}(\hat{\mu}_{1a}-\hat{\mu}_{1b})^2 \tag{81}$$

As for (78), we have:

$$\sum_{m=n_c}^{n}(x_m-\hat{\mu}_0)^2 = \sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})^2+\frac{(n_c-n_a)^2(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^2 \tag{82}$$

and $\hat{\sigma}_0^2 = \frac{1}{w}\sum_{m=n_a}^{n}(x_m-\hat{\mu}_0)^2$ can also be expressed as follows:

$$\hat{\sigma}_0^2 = \frac{1}{w}\left(\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2+\sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})^2\right)+\frac{(n_c-n_a)(n-n_b)^2}{w^3}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^2 + \frac{(n_c-n_a)^2(n-n_b)}{w^3}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^2 \tag{83}$$

By analogy with (54), we have:

$$\hat{\sigma}_1^2 = \frac{1}{w}\left(\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2+\sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})^2\right) \tag{84}$$

$$= \frac{n_c-n_a}{w}\hat{\sigma}_{1a}^2+\frac{n-n_b}{w}\hat{\sigma}_{1b}^2 \tag{85}$$

$$\text{where} \quad \hat{\sigma}_{1a}^2 = \frac{1}{n_c-n_a}\sum_{m=n_a}^{n_b}(x_m-\hat{\mu}_{1a})^2 \quad \text{and} \quad \hat{\sigma}_{1b}^2 = \frac{1}{n-n_b}\sum_{m=n_c}^{n}(x_m-\hat{\mu}_{1b})^2 \tag{86}$$

Where $\hat{\sigma}_1^2$ is commonly referred as the pooled variance [7]. Therefore,

$$\hat{\sigma}_0^2 = \hat{\sigma}_1^2+\frac{(n_c-n_a)(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^2 \tag{87}$$

Thus, when substituing (87) in (77), we have:

$$\Lambda_n = \left(\frac{\hat{\sigma}_1^2+\frac{(n_c-n_a)(n-n_b)}{w^2}(\hat{\mu}_{1b}-\hat{\mu}_{1a})^2}{\hat{\sigma}_1^2}\right)^{\frac{w}{2}} \tag{88}$$

$$= \left(1+\frac{(n_c-n_a)(n-n_b)}{w^2\hat{\sigma}_1^2}\frac{(\hat{\mu}_{1b}-\hat{\mu}_{1a})^2}{\hat{\sigma}_1^2}\right)^{\frac{w}{2}} \tag{89}$$

$$= (1+g_n)^{\frac{w}{2}} \tag{90}$$

g_n corresponds to the Student's t-test and can be also expressed as:

$$g_n \quad = \quad \left(\frac{w}{n_c - n_a} + \frac{w}{n - n_b}\right)^{-1} \frac{(\hat{\mu}_{1b} - \hat{\mu}_{1a})^2}{\hat{\sigma}_1^2} \tag{91}$$

By setting the likelihood ratio less than $\lambda \in [0, 1]$ according to Neyman Pearson Lemma, we obtain:

$$(1 + g_n)^{\frac{w}{2}} < \gamma \quad \Rightarrow \quad g_n > \gamma^{\frac{2}{w}} - 1 = h \tag{92}$$

The Student t-test can be expressed as:

$$g'_n \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtreqless}} h' \quad \text{with} \quad g'_n \quad = \quad g_n \times \frac{w^2}{(n_c - n_a)(n - n_b)} = \frac{(\hat{\mu}_{1b} - \hat{\mu}_{1a})^2}{\hat{\sigma}_1^2} \tag{93}$$

$$= \quad \frac{(\hat{\mu}_{1b} - \hat{\mu}_{1a})^2}{\frac{n_c - n_a}{w}\hat{\sigma}_{1a}^2 + \frac{n - n_b}{w}\hat{\sigma}_{1b}^2} \tag{94}$$

$$\text{and} \quad h' \quad = \quad h \times \frac{w^2}{(n_c - n_a)(n - n_b)} \tag{95}$$

with h' is a threshold value adjusted according to some desired decision probabilities. When an abrupt change is occurring at $n_c = n - \frac{w}{2} + 1$, the Student's t-test decision rules becomes:

$$g''_n = \frac{(\hat{\mu}_{1b} - \hat{\mu}_{1a})^2}{\hat{\sigma}_{1a}^2 + \hat{\sigma}_{1b}^2} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtreqless}} h'' = 2h' \tag{96}$$

References

[1] M. Basseville, “Detecting changes in signals and systems - a survey,” Automatica, vol. 24, pp. 309-326, May 1988.

[2] J.D. Healy, “A Note on Multivariate CUSUM Procedures”,Technometrics, vol. 29, pp. 409–412, 1987.

[3] P. Delacourt, C. Wellekens, “DISTBIC: a speaker-based segmentation for audio data indexing,” In Speech Communication, vol. 32, pp. 111-126, September 2000.

[4] B. Zhou, J. H. L. Hansen,“Efficient audio stream segmentation via the combined T^2 Statistic and Bayesian Information Criterion,” IEEE Trans. on speech and audio processing, vol. 13, July 2005.

[5] D. Nikovski, A. Jain, “Fast adaptive algorithms for abrupt change detection”, Machine Learning Springer, vol. 79, No. 3, 283–306, December 2010.

[6] N. Giri,“On the likelihood ratio test of a normal multivariate testing problem”, The Annals of Mathematical Statistics, vol. 35, No. 1, pp. 181-189, March 1964.

[7] H. Hotelling, “A generalized T test and measure of multivariate dispersion”, Proc. of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California, Los Angeles and Berkeley, pp.23-41, 1951.