Journal of Statistical Planning and Inference 143 (2013) 2151–2159

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Application of copulas to multivariate control charts

Ghislain Verdier *

Laboratoire de Mathématiques et de leur Applications, Université de Pau et des Pays de l'Adour, 64000 Pau, France

ARTICLE INFO

Article history:
Received 24 April 2013
Received in revised form
7 May 2013
Accepted 8 May 2013
Available online 21 May 2013

Keywords:
Statistical Process Control
Copulas modeling
Density estimation
Level sets
Multivariate control chart

ABSTRACT

The most popular multivariate control chart for monitoring the mean of a distribution is probably the Hotelling T^2 rule. Unfortunately, this rule relies on the assumption that the distribution under control is Gaussian, which is rarely true in practice. The objective of this paper is to propose a new approach for the non-normal multivariate case. It consists in the construction of a tolerance region obtained from a density level set estimation. The method follows a "plug-in" approach in which the density of the observations is previously estimated. This estimation is conducted using copulas modeling, an increasingly popular tool in multivariate modeling.

1. Introduction

For many years, Statistical Process Control is crucial in many industries. Due to the high competition between companies, it is necessary to use statistical methods in order to improve the production. Control charts introduced by Shewhart in the 1930s are part of these methods. The objective of a control chart is to detect, as quickly as possible, the change from an "in-control" to an "out-of-control" state of the system under supervision.

Let $X_1, X_2, \ldots$ be independent random variables. Under the normal operating mode (called hypothesis H_0), the X's are generated from the distribution function H. For each new observation, one has to decide whether it was generated from H (the system is still under control) or from $G \neq H$ (hypothesis H_1, the system is out of control). In the classic control charts context, the decision rule is constructed from a learning sample constituted of observations collected during a normal operating mode.

In most cases, several characteristics are involved in the supervision of a system (X is a vector), and these characteristics are often correlated. It is therefore necessary to consider multivariate control charts (Bersimis et al., 2007; Montgomery, 1996). Most of multivariate detection procedures encountered in the literature are based on a multinormality assumption on the distribution H. It is the case for example for the famous Hotelling T^2 rule (Hotelling, 1947), probably the most used rule in the industry for the detection of a change in the mean of the observations. However, the distribution of X under H_0 is often non-Gaussian and the Hotelling T^2 is no longer adapted.

Several alternatives can be considered. In the parametric case, Baillo and Cuevas (2006) fit a mixture model to the data before using a density level set estimation approach. In the same paper, they also develop a nonparametric approach in which the density of the observations under control is estimated by a kernel method. Extensive works were conducted on rules based on the notion of data depth (see for example Liu et al., 2004). He and Wang (2007) and Verdier and Ferreira (2011) use

* Tel.: +33 559407145.
E-mail address: ghislain.verdier@univ-pau.fr

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.05.005

a k-nearest neighbor approach by declaring a new observation "out-of-control" if its cumulative distance to its k nearest neighbors in the learning sample is greater than a fixed limit.

The method presented in this paper is inspired by the density level set estimation approach (Baillo and Cuevas, 2006; Cadre et al., 2013). It is a "plug-in" approach in which the density of the observations is estimated beforehand. A tolerance region is therefore defined. This region is no more than a density level set chosen in such a way that, under H_0, the probability that an observation falls outside this region is α, i.e. the false alarm rate fixed by the user.

Contrary to Baillo and Cuevas (2006) in which the density is estimated by fitting a normal mixture model to the data or by using a nonparametric kernel estimator, we propose to use copulas in order to model the multivariate distribution. This method has the advantage of being more flexible than a Gaussian mixture modeling, and avoids the problem of the curse of dimension encountered with nonparametric approaches.

These last years, copulas modeling (see for example Nelsen, 2006) has been increasingly used in many fields such as Finance, Actuarial science, etc. Copulas modeling is based on a representation from Sklar (1959) who states that the cumulative distribution function H can be uniquely written in the form

$$H(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}, \tag{1}$$

with d the dimension of X, $C : [0,1]^d \to [0,1]$ the copula and $F_1, \dots, F_d$ the marginal cumulative distribution functions of X.

From Eq. (1) it is possible to express the density h of X and therefore to construct the control chart from the density level set approach. In this paper, the estimation of C is obtained from the classic maximum pseudolikelihood method (Genest et al., 1995), and two approaches are used in order to estimate the marginal distributions F_i: a parametric method (maximum-likelihood estimation) and a nonparametric one based on kernel estimation and used by Liebscher (2005) and Bouezmarni and Rombouts (2009) in the framework of copulas modeling.

The paper is organized as follows. The density level set estimation to construct a multivariate control chart is presented in the following section. Then, the copulas modeling method and the estimation of h are considered. A convergence result is obtained in Section 4, followed by a part devoted to comparisons of different multivariate control charts through simulations.

2. Control charts based on density level set estimation

Let $X = (X^{(1)}, \dots, X^{(d)})^T$ be a random vector of $\mathbb{R}^d$ with continuous marginal cumulative distribution functions $F_1, \dots, F_d$. Let H and h respectively the cumulative distribution function and density of X. The aim of this paper is to decided if a new observation, saying X_{test}, has been generated from h. The decision will be taken given a set of i.i.d. observations $X_1, \dots, X_n$ (named the learning sample) drawn from h.

Following Baillo and Cuevas (2006), a multivariate tolerance region is defined. If X_{test} falls outside this tolerance region, the system is declared out-of-control.

This tolerance region is constructed such as the probability of false alarm (i.e. to consider an observation drawn from h as an out-of-control) is equal to a specified level α. It can be written in the following manner:

$$D(c_\alpha) = \{x \in \mathbb{R}^d \,|\, h(x) \ge c_\alpha\}, \tag{2}$$

where c_α satisfies

$$I = P[X \in D(c_\alpha)] = \int_{D(c_\alpha)} h(x)\, dx = 1 - \alpha. \tag{3}$$

The estimation of $D(c_\alpha)$ is referred as the problem of density level set estimation. There is an abundant literature in the case where the threshold c_α of (2) is not relied to a probability α through constraint (3). But recent papers (Cadre et al., 2013; Park et al., 2010) consider constraint (3) and provide convergence results of plug-in methods based on nonparametric estimation of h.

In the sequel, the unknown function h is estimated by h_n. The value c_α will be estimated by c_n, which satisfies

$$I_n = \int_{\{h_n \ge c_n\}} h_n(x)\, dx = 1 - \alpha \quad \text{a.s.} \tag{4}$$

Therefore, the estimated tolerance region will be

$$D_n(c_n) = \{x \in \mathbb{R}^d \,|\, h_n(x) \ge c_n\}.$$

But, by construction,

$$P(h_n(X) \ge c_n) = 1 - \alpha,$$

and therefore c_n is the α–quantile of $h_n(X)$. Cadre et al. (2013) propose to estimate c_n by the α-quantile of the empirical distribution of $h_n(X_1), \dots, h_n(X_n)$. This estimator is noted $\hat{c}_n$. In order to improve the estimation, we propose to consider an empirical quantile computed on a large number N of realisations obtained from the estimated density h_n. For this, we defined $\hat{F}_{h_n(X)}$, the distribution function of a random sample of $h_n(X)$. This sample is noted $h_n(X_{1,n}),\dots,h_n(X_{N,n})$ and we have

$$\hat{F}_{h_n(X)}(y)=\frac{1}{N}\sum_{i=1}^{N}\mathbf{I}(h_n(X_{i,n})\le y),$$

where $\mathbf{I}(\cdot)$ is the indicator function.

Therefore, the estimation of c_α is obtained by

$$\hat{c}_{N,n}=\inf\{c\in\mathbb{R}:\hat{F}_{h_n(X)}(c)\ge\alpha\}.$$

3. Estimation of *h* through copulas modeling

Copulas modeling is based on the following result from Sklar (1959):

Theorem 1. Sklar *Let H a d-dimensional distribution function with continuous marginal cumulative distribution functions* $F_1,\dots,F_d$. *Then there exists a unique function* $C:[0,1]^d\to[0,1]$ *such that*

$$H(x_1,\dots,x_d)=C\{F_1(x_1),\dots,F_d(x_d)\},\quad \forall x=(x_1,\dots,x_d)\in\mathbb{R}^d.$$

C, which is called a copula, is a d-dimensional cumulative distribution function with standard uniform margins. C captures the dependence of X.

The density h of X can be written, for all $x=(x_1,\dots,x_d)\in\mathbb{R}^d$,

$$h(x_1,\dots,x_d)=c(F_1(x_1),\dots,F_d(x_d))f_1(x_1)\cdots f_d(x_d),\tag{5}$$

where

$$c(u_1,\dots,u_d)=\frac{\partial^d}{\partial u_1\dots\partial u_d}C(u_1,\dots,u_d).$$

3.1. Copula adjustment

As it is classically done in copulas modeling, we consider a parametric family $\mathcal{C}=\{C_\theta;\theta\in\Theta\}$ in order to model the dependence between the d components of X. Below are presented two examples of such families

- the Frank family

$$C_\theta^F(u_1,\dots,u_d)=-\frac{1}{\theta}\log\left(1+\frac{\prod_{i=1}^{d}(e^{-\theta u_i}-1)}{(e^{-\theta}-1)^{d-1}}\right),\quad \theta\in\Theta_F:=\mathbb{R}^+.$$

- the Clayton family

$$C_\theta^C(u_1,\dots,u_d)=\left(\sum_{i=1}^{d}u_i^{-\theta}-(d-1)\right)^{-1/\theta},\quad \theta\in\Theta_C:=\mathbb{R}_*^+.$$

The estimation of the unknown parameter θ_0 is done by the maximum pseudolikelihood approach, studied notably by Genest et al. (1995). The estimator obtained is noted $\hat{\theta}_n$.

3.2. Marginal distributions adjustment

In the sequel, two methods for the adjustment of the marginal distributions will be considered.

*Method*1: A parametric approach is used. For each of the d components of X, the marginal distribution is supposed to lie in a parametric family indexed by a parameter $\theta^{(j)}$, which is estimated by the maximum likelihood method.

*Method*2: The marginal distributions are estimated by nonparametric estimators. Following Liebscher (2005), a nonparametric kernel estimator of f_j, the density of the j-th component of X is used

$$\hat{f}_j(z)=\frac{1}{nb}\sum_{i=1}^{n}K\left(\frac{z-X_i^{(j)}}{b}\right),$$

where K is a kernel and b its bandwidth. Note that Bouezmarni and Rombouts (2009) propose an extension of this approach considering a local linear kernel. According to the authors, this local linear kernel is more robust to the boundary bias problem.

3.3. Estimation of the density h and its properties

The density h of X is expressed as in Eq. (5). An estimator of h, in the parametric context of *Method*1 is given by

$$h_n^{(1)}(x_1,\dots,x_d)=c_{\hat{\theta}_n}(F_1(x_1|\hat{\theta}^{(1)}),\dots,F_j(x_j|\hat{\theta}^{(j)}))f_1(x_1|\hat{\theta}^{(1)})\dots f_j(x_j|\hat{\theta}^{(j)}) \tag{6}$$

for all $x=(x_1,\dots,x_d)\in\mathbb{R}^d$.

In the nonparametric case, two estimators of h can be considered. In the first one, the marginal cdf $F_j, j=1,\dots,d$ of Eq. (5) are replaced by $\hat{F}_j$, the empirical distributions

$$\hat{F}_j(x_j)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{I}(X_i^{(j)}\le x_j).$$

The estimator is then given by

$$h_n^{(2)}(x_1,\dots,x_d)=c_{\hat{\theta}_n}(\hat{F}_1(x_1),\dots,\hat{F}_d(x_d))\hat{f}_1(x_1)\dots\hat{f}_d(x_d). \tag{7}$$

Another estimator can be considered if the marginal cdf $F_j, j=1,\dots,d$ are replaced by the nonparametric cdf $F_{\hat{f}_j}$, obtained from the nonparametric estimations

$$F_{\hat{f}_j}(x_j)=\int_{-\infty}^{x_j}\hat{f}_j(t)\,dt.$$

The third estimator is then given by

$$h_n^{(3)}(x_1,\dots,x_d)=c_{\hat{\theta}_n}(F_{\hat{f}_1}(x_1),\dots,F_{\hat{f}_d}(x_d))\hat{f}_1(x_1)\dots\hat{f}_d(x_d). \tag{8}$$

Under a strong assumption on c (satisfied for example by the Fairlie–Gumbel–Morgenstern family of copulas) and general conditions on the kernel K, the bandwidth b and the estimator $\hat{\theta}_n$, Liebscher (2005) obtains the uniform convergence of $h_n^{(2)}$

$$\sup_{x\in\mathbb{R}^d}|h_n^{(2)}(x)-h(x)|\to 0 \quad \text{a.s.} \tag{9}$$

By using a result of Devroye and Wagner (1979) who obtained the uniform convergence of $F_{\hat{f}_j}$, it is straightforward, following the proof of Theorem 2.1 of Liebscher (2005) to extend the result (9) to $h_n^{(3)}$.

4. Properties

In Cadre et al. (2013) the convergence in probability of $\hat{c}_n$ to c_α is obtained supposing that $\hat{c}_n$ is defined as the α−quantile of the empirical distribution of $h_n(X_1),\dots,h_n(X_n)$. In order to improve the estimation procedure, we propose in this paper to generate N realisations from h_n, noted $X_{1,n},\dots,X_{N,n}$, and define the new estimator of c_α, noted $\hat{c}_{N,n}$, as the α−quantile of $h_n(X_{1,n}),\dots,h_n(X_{N,n})$. More precisely,

$$\hat{c}_{N,n}=\inf\{c\in\mathbb{R}:\hat{F}_{h_n(X)}(c)\ge\alpha\},$$

with

$$\hat{F}_{h_n(X)}(y)=\frac{1}{N}\sum_{i=1}^{N}\mathbf{I}(h_n(X_{i,n})\le y).$$

In the sequel, for the convergence relatively to n and N, we adopt the classic approach in which we set $N=\gamma n$, with γ a constant. Let $\|\cdot\|$ be the usual Euclidean norm and $\|\cdot\|_\infty$ be the uniform norm

$$\|f\|_\infty=\sup_{x\in\mathbb{R}^d}|f(x)|.$$

We follow the paper of Cadre et al. (2013). Let $\mathcal{H}_0$ the following set:

$$\mathcal{H}_0=\left\{c\in]0;\sup_{x\in\mathbb{R}^d}h(x)[:\inf_{\{h=c\}}\|\nabla h\|=0\right\}.$$

Assumption 1.

1. The density h is of class C^2 with a bounded Hessian matrix, and $h(x)\to 0$ as $\|x\|\to\infty$.

2. $\mathcal{H}_0$ has Lebesgue content 0.
3. $\lambda(\{h=c\})=0$ for all $c>0$.

See Cadre et al. (2013) for explanation concerning Assumption 1.

Theorem 2. *Suppose that h satisfies Assumption 1 and that*

$$\sup_{x\in\mathbb{R}^d}|h_n(x)-h(x)|\to 0 \quad \text{p.s.}$$

Then, for almost all $c\in]0;1[$,

$$\hat{c}_{N,n}\to c_\alpha \quad \textit{in probability for } N\to\infty.$$

The proof is given in Appendix.

5. Simulation trials

In this section the performance of the control charts based on copulas modeling is compared to two classic multivariate control charts: the Hotelling T^2 rule and a control chart based on data depth (see for example Liu et al., 2004). The computations are made using R, and particularly the copula package (Kojadinovic and Yan, 2010).

The Hotelling T^2 rule is no more than a density level set approach in which the unknown distribution H is supposed to be a multivariate Gaussian distribution, with unknown mean and covariance matrix (see Montgomery, 1996). The data depth control chart is a nonparametric approach which declares an observation out-of-control if it is too deep relatively to the learning sample. The depth is measured through the simplicial depth (Liu, 1995).

Concerning the control charts by copulas, the density h of the observations is only estimated using $h_n^{(1)}$ and $h_n^{(3)}$, since the estimator $h_n^{(2)}$ gives very similar results with $h_n^{(3)}$. The simulations are conducted with $\alpha=2\%$ and $N=2000$.

In practice, the choice of a copula family is not always simple. Therefore, in order to make the procedure more automatic, several families of copulas are considered. A goodness-of-fit test (Kojadinovic and Yan, 2011) is done and we decide, arbitrarily, to choose the family which presents the greater p-value. For the nonparametric estimation of the margins, Gaussian kernels are used and the bandwidth is obtained by the plug-in method proposed by Wand and Jones (1994).

5.1. Multivariate normal case

In this subsection, the four control charts are applied on a multivariate normal distribution of dimension 2 with the identity covariance matrix. Under H_0, the mean of the distribution is equal to $\mu_0=(0,0)$ and under H_1, the mean is equal to $\mu_1=(2,2)$.

In the phase I of implementation, the control charts are constructed from a learning sample of size n of in-control data (two size are used: $n=250$ and $n=500$). For the control charts by copulas, five families of copulas are considered: Clayton, Frank, Plackett, Student and obviously the Normal copula family. The estimation of the margins of the parametric approach is made by testing three types of distribution: Cauchy, Student (df=2) and the Normal distribution. Then in phase II, the control charts are applied on 500 observations generated under H_0 and 500 observations generated under H_1.

The previous phases I and II are repeated 1000 times which allows to obtain an estimation of the false alarm rate and the probability of detection of each control charts considered. The results are presented in Table 1, in which M1 and M2 refer to the control charts by copulas obtained from, respectively, $h_n^{(1)}$ and $h_n^{(3)}$, and DD is the control chart based on data depth. The standard deviation of the number of detections is also given.

As expected, the Hotelling T^2 is the most efficient rule. For $n=250$ the false alarm rate is high for the control chart based on data depth whereas it is relatively close to the fixed rate $(\alpha=2\%)$ for the copulas modeling (Method 1). The performances

Table 1
Average number and standard deviation of the number of detections for the multinormal distribution.

Sample size	Mode		T^2	DD	M1	M2
$n=250$	H_0	Mean (%)	1.96	3.96	2.19	1.69
		St. dev.	3.87	7.43	4.40	4.00
	H_1	Mean (%)	57.40	64.12	57.95	52.96
		St. dev.	27.59	50.74	30.49	34.83
$n=500$	H_0	Mean (%)	2.02	2.30	2.13	1.68
		St. dev.	3.60	4.90	3.66	4.06
	H_1	Mean (%)	57.80	55.92	57.96	53.64
		St. dev.	20.88	46.18	23.93	26.97

of the data depth rule are better for a larger sample size. However, the Method 1 (M1) of copulas modeling remains more efficient. The results obtained by Method 2 (M2) are not satisfactory. As it could be seen in the sequel, it is probably due to the bandwidth parameter of the kernel estimations.

5.2. Multivariate non-normal case

The learning sample is now generated from a Franck copula of parameter $\theta = 5$ with Fisher margins $(X^{(i)} \sim F(10,15),\ i = 1,2)$. Two mean changes are considered: firstly a shift of $(-1,-1)$ and secondly a shift of $(+2,+2)$. For the control charts by copulas, six families of copulas are considered : Clayton, Frank, Gumbel, Plackett, Student and Normal. Four types of distribution are considered for the margins : Normal, Log-normal, Gamma and Fisher.

The results are presented in Table 2. Fig. 1 represents the learning sample of a run, and the corresponding tolerance region obtained for each of the four rules. The theoretical tolerance region is also represented in dashed line.

Obviously, the Hotelling T^2 is no longer suited and presents a high false alarm rate. For $n = 500$, the control chart based on data depth gives good results for the first change, but presents a very low rate of detection for the second. This is confirmed by Fig. 1, in which it is easy to see that the tolerance region of the data depth control chart is far from the theoretical tolerance region. On the contrary, the tolerance region obtained by Method 1 of copulas modeling is very close to the theoretical one.

Table 2
Average number and standard deviation of the number of detections for a non-multinormal distribution.

Sample size	Mode		T^2	DD	M1	M2
$n = 250$	H_0	Mean (%)	4.98	3.74	2.34	2.88
		St. dev.	7.17	7.34	5.22	5.34
	H_1 change 1	Mean (%)	7.74	77.89	71.13	58.89
		St. dev.	26.25	11.77	11.69	17.66
	H_1 change 2	Mean (%)	56.67	23.32	43.51	51.43
		St. dev.	55.89	74.81	64.07	67.93
$n = 500$	H_0	Mean (%)	4.86	2.17	2.11	2.48
		St. dev.	6.25	4.66	4.27	4.45
	H_1 change 1	Mean (%)	6.38	76.01	70.77	60.99
		St. dev.	19.21	11.38	11.43	14.55
	H_1 change 2	Mean (%)	55.76	13.96	42.52	47.70
		St. dev.	41.48	44.59	47.80	52.37

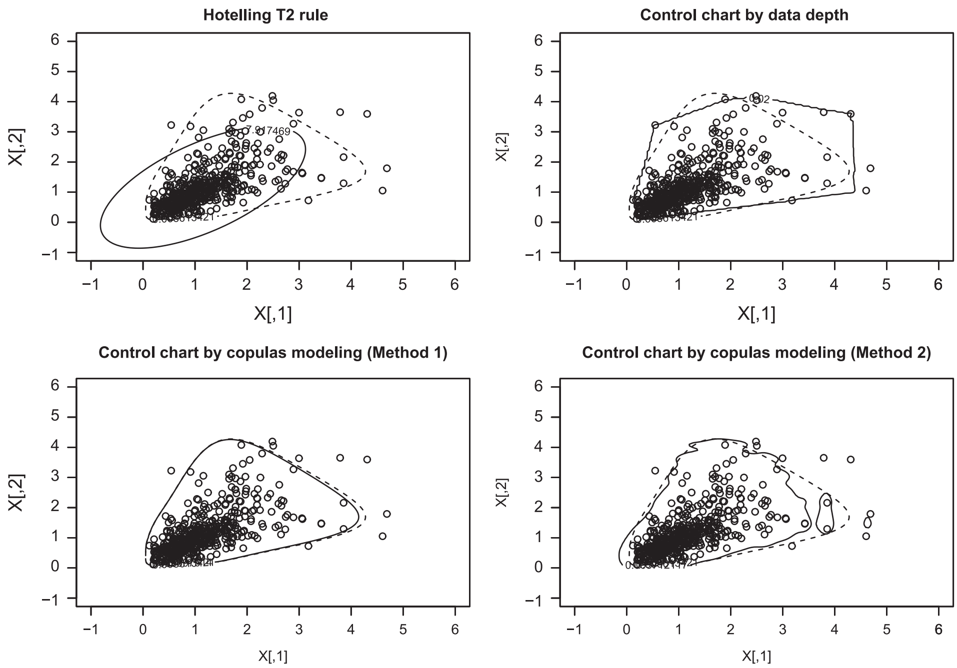

Fig. 1. Learning sample ($n = 500$) and tolerance regions obtained from the four rules, compared to the theoretical tolerance zone (dashed line).

Moreover, the empirical procedure consisting in choosing the copula family which obtains the largest p-value for the goodness-of-fit test presents a high success rate: 85% for $n=250$ and 94% for $n=500$.

Method 2 of copulas modeling gets no better results than in the multivariate Gaussian case. It seems clear from Fig. 1 that this is due to the choice of the smoothing parameter. An alternative can be the use of the local linear kernel estimator proposed by Bouezmarni and Rombouts (2009), more robust to the boundary problem.

6. Conclusion

The Hotelling T^2 control chart is probably the most used rule in industry for the problem of multivariate fault detection. Unfortunately, this rule relies on the assumption that the observations under control are Gaussian. When the T^2 is applied on non-normal multivariate observations, it can leads to a lot of false alarms and non-detections.

For several years, alternative approaches are proposed, based on parametric or nonparametric methods. The control chart proposed in this paper is based on copulas modeling. Copulas modeling is a very efficient tool for multivariate modeling and it is used in a large number of application domains: finance, actuarial science, etc. The control chart obtained is not subject to the curse of dimension of nonparametric approaches and remains relatively flexible thanks to a wide choice of copula families.

However, it has to be noted that we consider in this paper the problem of deciding if an observation X_{test} has been generated from a reference distribution F. If the objective is to detect a change in a dynamic system (from an unknown time t_c, the system is out of control and all the new observations are generated from a new distribution G), the control chart proposed in this paper can be called "without memory", like the Hotelling T^2 rule or the data depth rule used in Section 5. At the time step t, the decision of declaring the system "in-control" or "out-of-control" is taken only using the observation X_t (and obviously the learning sample). Control charts "without memory" are particularly well suited for the detection of large changes in the distribution F. But when we have to deal with a small change (for example a small variation of the process mean), control charts "with memory", like the MEWMA (a combination of the T^2 and an EWMA procedure (see Montgomery, 1996) for Gaussian observations or nonparametric approaches (Qiu, 2008; Zou and Tsung, 2011) will be preferable. The way to combine our method with, for example, an EWMA procedure in order to take into account the history of the process (the observations $X_{t-1}, X_{t-2}, \ldots$) and thus obtain a more efficient control chart in the case of a small change will probably be the subject of a future work.

Appendix A. Proof of Theorem 2

We will used in the sequel the Levy metric defined by

$$d_L(\varphi_1,\varphi_2)=\inf\{\theta>0 : \forall x\in\mathbb{R}, \varphi_1(x-\theta)-\theta\le\varphi_2(x)\le\varphi_1(x+\theta)+\theta\},$$

with φ_1 and φ_2 two real-valued functions on $\mathbb{R}$.

Proof of Theorem 2. Let us introduce the following notation:

$$D^l(c)=\{x\in\mathbb{R}^d : h(x)\le c\} \quad \text{and} \quad D_n^l(c)=\{x\in\mathbb{R}^d : h_n(x)\le c\},$$

and we note μ and μ_N, respectively the law of X and the empirical law obtained from the N realisations $X_{1,n},\ldots,X_{N,n}$. If we note $F_{h(X)}$ the cdf of $h(X)$, we obtain

$$F_{h(X)}(c)=\mu(D^l(c))$$

and

$$\hat{F}_{h_n(X)}(c)=\frac{1}{N}\sum_{i=1}^{N}\mathbf{I}(h_n(X_{i,n})\le c)=\mu_N(D_n^l(c)).$$

The proof of Proposition 3.4 of Cadre et al. (2013) can be adapted to our situation and leads to a bound for $d_L(F_{h(X)},\hat{F}_{h_n(X)})$

$$d_L(F_{h(X)},\hat{F}_{h_n(X)})\le\max(\|h_n-h\|_\infty, V_N), \tag{A.1}$$

with,

$$V_N=\sup_{c\ge 0}|\mu_N(D^l(c))-\mu(D^l(c))|.$$

From Assumption 1, $F_{h(X)}$ is a bijection from $]0;\sup_{x\in\mathbb{R}^d}h(x)[$ to $]0;1[$ (see Cadre et al., 2013). Let G its inverse function. By Lemma 3.1 of Cadre et al. (2013), G is almost everywhere differentiable.

We consider $\alpha\in]0;1[$ such that G is differentiable at α. Then $G(\alpha)=c_\alpha$. If we note G_N the pseudo-inverse of $\hat{F}_{h_n(X)}$, i.e.

$$G_N(s)=\inf\{c\ge 0 : \hat{F}_{h_n(X)}(c)\ge s\},$$

then $G_N(\alpha)=\hat{c}_{N,n}$.

Moreover, we have

$$d_L(F_{h(X)}, \hat{F}_{h_n(X)}) \leq 1,$$

since $0 \leq F_{h(X)}(c) \leq 1$ and $0 \leq \hat{F}_{h_n(X)}(c) \leq 1$ for all $c \in \mathbb{R}$.

By the property of the Levy metric

$$d_L(F_{h(X)}, \hat{F}_{h_n(X)}) = d_L(G, G_N).$$

Therefore by using Lemma 3.2 of Cadre et al. (2013) and inequality (A.1), we obtain for some positive constant D_1

$$|\hat{c}_{N,n} - c_\alpha| = |G_N(\alpha) - G(\alpha)| \leq D_1 d_L(F_{h(X)}, \hat{F}_{h_n(X)}) \leq D_1 \max(\|h_n - h\|_\infty, V_N). \tag{A.2}$$

The end of the proof is obtained by using the following extension of the Lemma 3.3 of Cadre et al. (2013).

Lemma 3. *V_N converges in probability, as N tends to infinity, to* 0.

Proof of Lemma 3. We note μ_n the probability measure associated to the estimation h_n of h, i.e. $\mu_n(A) = \int_A h_n(x)\,dx$.

In order to simplify the notation we note $\mathcal{A}$ the following collection of sets:

$$\mathcal{A} = \{D^l(c), c \geq 0\}.$$

Therefore,

$$V_N := \sup_{c \geq 0} |\mu_N(D^l(c)) - \mu(D^l(c))| = \sup_{A \in \mathcal{A}} |\mu_N(A) - \mu(A)| \leq \sup_{A \in \mathcal{A}} |\mu_N(A) - \mu_n(A)| + \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)|.$$

Then, for $\epsilon > 0$

$$P[V_N > \epsilon] \leq P\left[\sup_{A \in \mathcal{A}} |\mu_N(A) - \mu_n(A)| + \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| > \epsilon\right] \leq P\left[\sup_{A \in \mathcal{A}} |\mu_N(A) - \mu_n(A)| > \frac{\epsilon}{2}\right] + P\left[\sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| > \frac{\epsilon}{2}\right]. \tag{A.3}$$

But the Vapnik–Chervonenkis dimension of $\mathcal{A}$ is 1 (see Devroye et al., 1996). We can therefore applied the Vapnik–Chervonenkis inequality (see Devroye et al., 1996), which gives

$$P\left[\sup_{A \in \mathcal{A}} \left|\mu_N(A) - \mu_n(A)\right| > \frac{\epsilon}{2}\right] \leq D_2 N e^{-N(\epsilon/2)^2/32},$$

with D_2 a constant.

For the second term of the right member of (A.3), we have

$$\forall A, \quad |\mu_n(A) - \mu(A)| = \left|\int_A h_n(x)\,dx - \int_A h(x)\,dx\right| \leq \int_A |h_n(x) - h(x)|\,dx \leq \int |h_n(x) - h(x)|\,dx.$$

But by Glick Theorem (see Glick, 1974), since the densities h_n are supposed to be converging almost surely to h, we have

$$\int |h_n(x) - h(x)| dx \to 0 \quad \text{p.s.},$$

which concludes the proof of the Lemma, and then, the proof of the Theorem. □

References

Baillo, A., Cuevas, A., 2006. Parametric versus nonparametric tolerance regions in detection problems. Computational Statistics 21 (3–4), 523–536.

Bersimis, S., Psarakis, S., Panaretos, J., 2007. Multivariate statistical process control charts: an overview. Quality and Reliability Engineering International 23, 517–543.

Bouezmarni, T., Rombouts, J., 2009. Semiparametric multivariate density estimation for positive data using copulas. Computational Statistics and Data Analysis 53 (6), 2040–2054.

Cadre, B., Pelletier, B., Pudlo, P., 2013. Estimation of density level sets with a given probability content. Journal of Nonparametric Statistics 25 (1), 261–272.

Devroye, L., Györffi, L., Lugosi, G., 1996. A Probabilistic Theory of Pattern Recognition. Springer, Berlin.

Devroye, L.P., Wagner, T.J., 1979. The L_1 convergence of kernel density estimates. The Annals of Statistics 7 (5), 1136–1139.

Genest, C., Ghoudi, K., Rivest, L.-P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distribution. Biometrika 82 (3), 543–552.

Glick, N., 1974. Consistency conditions for probability estimators and integrals of density estimators. Utilitas Mathematica 6, 61–74.

He, Q.P., Wang, J., 2007. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. IEEE Transactions on Semiconductor Manufacturing 20 (4), 345–354.

Hotelling, H., 1947. Multivariate quality control, illustrated by the air testing of sample bombsights. In: Eisenhart, C., Hastay, M.W., Wallis, W.A. (Eds.), Techniques of Statistical Analysis. McGraw-Hill, New York, pp. 111–184.

Kojadinovic, I., Yan, J., 2010. Modeling multivariate distributions with continuous margins using the copula R package. Journal of Statistical Software 34 (9), 1–20.

Kojadinovic, I., Yan, J., 2011. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. Statistics and Computing 21, 17–30.

Liebscher, E., 2005. Semiparametric density estimators using copula. Communications in Statistics—Theory and Methods 67, 318–348.

Liu, R.Y., 1995. Control charts for multivariate processes. Journal of the American Statistical Association 90 (432), 1380–1387.

Liu, R.Y., Singh, K., Teng, J.H., 2004. DDMA-charts: nonparametric multivariate moving average control charts based on data depth. Allgemeines Statistisches Archiv 88 (2), 235–258.

Montgomery, D., 1996. Introduction to Statistical Quality Control. Wiley, New York.
Nelsen, R., 2006. An Introduction to Copulas. Springer-Verlag, New York.
Park, C., Huang, J.Z., Ding, Y., 2010. A computable plug-in estimator of minimum volume sets for novelty detection. Operations Research 58 (5), 1469–1480.
Qiu, P., 2008. Distribution-free multivariate process control based on log-linear modeling. IIE Transactions 40, 664–677.
Sklar, A., 1959. Fonctions de répartition á n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.
Verdier, G., Ferreira, A., 2011. Adaptive mahalanobis distance and k-nearest neighbor rule for fault detection in semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing 24 (1), 59–68.
Wand, M., Jones, M., 1994. Multivariate plugin bandwidth selection. Computational Statistics 9, 97–116.
Zou, C., Tsung, F., 2011. A multivariate sign EWMA control chart. Technometrics 53, 84–97.