



HAL
open science

carte de controle non paramétrique fondée sur les statistiques des prédécesseurs pour le suivi de données de survie

Jean-Christophe Turlot, Christian Paroissin, Narayanashramy Balakrishnan

► **To cite this version:**

Jean-Christophe Turlot, Christian Paroissin, Narayanashramy Balakrishnan. carte de controle non paramétrique fondée sur les statistiques des prédécesseurs pour le suivi de données de survie. QUALITA2013, Mar 2013, Compiègne, France. hal-00823138

HAL Id: hal-00823138

<https://hal.science/hal-00823138>

Submitted on 16 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CARTES DE CONTROLE NON PARAMETRIQUES FONDEES SUR LES STATISTIQUES DES PREDECESSEURS POUR LE SUIVI DE DUREES DE VIE.

Jean-Christophe Turlot ⁽¹⁾ Christian Paroissin ⁽¹⁾ Narayanaswamy Balakrishnan ⁽²⁾

⁽¹⁾ *Université de Pau et des Pays de l'Adour, Laboratoire de Mathématiques et de leurs applications-UMR CNRS 5142, Avenue de l'Université, 64013 Pau cedex, France, cparoiss@univ-pau.fr, turlot@univ-pau.fr*

⁽²⁾ *Mc Master University, Department of Mathematics and Statistics, Hamilton, Ontario, Canada L8S 4K1, bala@mcmaster.ca*

Résumé. On propose des cartes de contrôle non paramétriques fondées sur la statistique des prédécesseurs et plus généralement sur les statistiques de placement. Ces cartes de contrôle sont dédiées plus précisément aux distributions de durée de vie. Il s'agit de tester $G=F$ contre $G>F$, où F est la distribution des observations sous contrôle et G la distribution des observations d'un échantillon à tester. De telles cartes de contrôle ne semblent pas encore avoir été étudiées. Dans cette communication, nous présentons deux cartes de contrôle fondées sur les statistiques de placement: ces deux types de cartes se ramènent à des versions pondérées de statistiques de placements. Pour ces deux cartes, on calcule les valeurs critiques ou limites de contrôle ainsi que la période opérationnelle moyenne sous l'hypothèse nulle, mais aussi la puissance pour une des deux familles d'alternatives proposées par Lehmann. Des résultats numériques complètent la présentation.

Mots-clés. Statistique des prédécesseurs, statistiques de placements, alternatives de Lehmann, cartes de contrôle, statistiques de Wilcoxon et Mann-Whitney

Abstract. We propose some non-parametric one-sided control charts based on precedence statistic or related ones. These control charts are designed especially for lifetime distributions. Hence we want to test whether $G=F$ or $G>F$ where F is the distribution of an unit under control and G is the distribution of an unit of a test sample. Such kind of control charts have not be studied previously in the literature. In this talk, we will present two control charts, based on placement statistics : these two charts can be viewed as weighted versions of placement statistics. For these two control charts we compute critical values and unconditional average run length (ARL) under the null assumption and also under one of the two different Lehmann alternatives. Numerical computations are also provided.

Keywords. Precedence statistic, placement statistics, Lehmann alternatives, control charts, Wilcoxon, Mann-Whitney statistics

1. Introduction

Le contrôle statistique de processus permet de piloter la qualité d'un processus de fabrication. Lorsque les variations des caractéristiques à contrôler sont considérées comme naturelles (fluctuations de faible amplitude, attribuables à de multiples causes...), on dit que le processus est sous contrôle. Une carte de contrôle doit permettre de détecter le plus rapidement possible une dérive en signalant que le procédé est hors contrôle pour une cause assignable. Les cartes de contrôle usuelles et très largement utilisées font appel à la loi normale pour modéliser les fluctuations des mesures lorsque le procédé est sous contrôle (cartes de Shewhart, cartes CUSUM, cartes EWMA, ...). Lorsque cette hypothèse est inappropriée, la carte opère de manière incorrecte : elle peut conduire à une probabilité de fausse alarme bien supérieure à la spécification ou encore à

une faible probabilité de détection d'une réelle dérive. Les cartes non paramétriques, outre le fait que la loi de probabilité des variations du processus n'a pas besoin d'être connue, présentent l'avantage d'être robustes. Elles sont peu sensibles aux données atypiques, à l'asymétrie ou à la présence de queues de probabilité élevée dans la distribution sous jacente. Pour une revue de littérature sur les cartes sur les cartes non paramétriques, on pourra consulter l'article de Chakraborti et Graham (2007).

2. La statistique des prédécesseurs.

Pour l'élaboration d'une carte de contrôle, on dispose d'une réalisation d'un échantillon de référence $X=(X_1, \dots, X_m)$ d'assez grande taille (dans notre application, $m=50$) et de l'observation d'une suite de lots $sY^h=(Y_1^h, \dots, X_n^h)$ prélevés durant la production ($h=1:H$), la taille des lots étant usuellement plus petite que pour l'échantillon de référence (dans l'application, $n=5$). On notera que les tailles d'échantillon ne pourront pas être plus petites que dans un cadre gaussien (carte de contrôle de Shewart, etc.). Pour les illustrations numériques, on a choisi $m=50$ et $n=5$.

Soit F la fonction de répartition de la variable aléatoire X dont est issu le premier échantillon et G la fonction de répartition de la variable aléatoire Y dont est issu l'un des échantillons à tester. Le problème de test posé, lorsqu'il s'agit de problèmes de durée de vie, s'écrit :

$$H_0 : "G = F" \text{ contre } H_1 : "G > F"$$

où F et G ont des distributions continues. La statistique des prédécesseurs $P_{(b)}$ consiste à compter le nombre d'observations de l'échantillon test Y^h inférieures à la b -ième observation dans l'échantillon ordonné de référence X .

Soient $X_{1:m} < X_{2:m} < \dots < X_{m:m}$ et $Y_{1:n} < Y_{2:n} < \dots < Y_{n:n}$ les statistiques d'ordre :

$$\forall b \in \{1, 2, \dots, m\}, \quad P_b = \sum_{i=1}^n I_{\{-\infty, X_{b:m}\}}(Y_i)$$

On a l'identité $\{P_b \geq r\} \Leftrightarrow \{Y_{r:n} < X_{b:m}\}$. Lorsque le processus est sous contrôle :

$$\forall i \in \{1, 2, \dots, n\}, \quad \Pr_{sc}[P_b = i] = \frac{\binom{b+i-1}{i} \binom{m+m-b-i}{n-i}}{\binom{m+n}{n}}$$

(voir par exemple, Balakrishnan et Ng, 2006). On en déduit la valeur critique $c=c(b, \alpha)$ du test, voir la figure 1 pour une illustration avec $\alpha=1\%$ ou $0,5\%$ et différentes valeurs de b (on notera que la valeur critique dépend aussi des tailles d'échantillon m et n). Dans la suite de l'article, on ne regardera que le cas où $\alpha=1\%$.

La puissance du test $P_{(b)}$ dépend de la loi alternative. En statistique non paramétrique, il est d'usage de considérer les familles d'alternatives de Lehmann.

$$(1) \quad G(x) = [F(x)]^\gamma$$

$$(2) \quad G(x) = 1 - [1 - F(x)]^\gamma$$

Les lois de type puissance sont stables pour la famille (1) tandis que la loi de Weibull, et plus généralement les lois de durée de vie à taux de risque proportionnels, sont stables pour la famille (2). Etant intéressé par le suivi de durées de vie, on se placera donc uniquement dans le cas de la

seconde alternative de Lehmann. Alors, sous cette hypothèse, $G > F$ équivaut à $\gamma > 1$ (et $G = F$ équivaut à $\gamma = 1$). La probabilité que la carte ne fournisse pas de signal, alors que le processus est hors contrôle, est définie par :

$$P = \Pr_{HC} [Y_{c:n} > X_{b:m}] \quad F < G$$

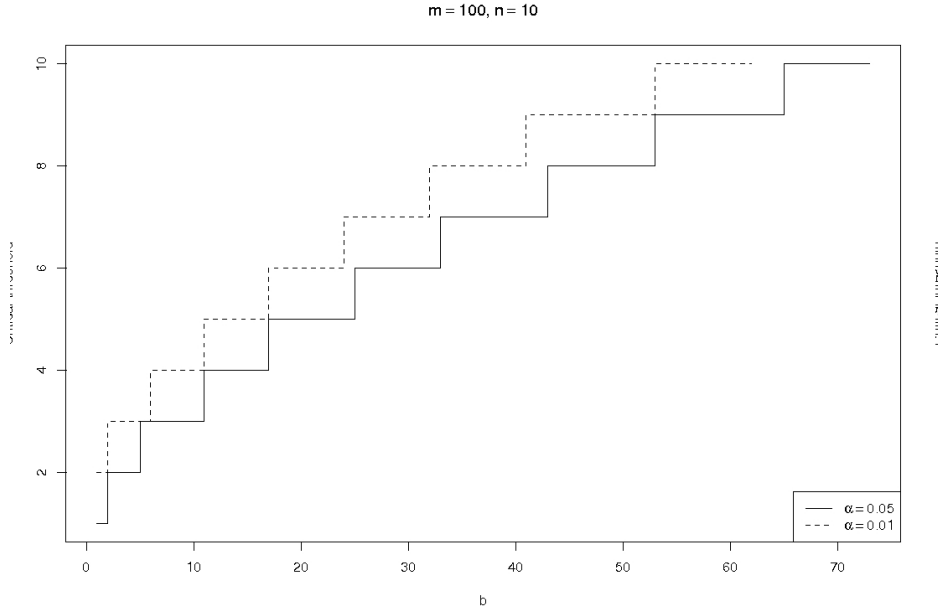


Figure 1 : seuils pour $m=100$, $n=10$ et $\alpha=1\%$ (trait pointillé) ou $0,5\%$ (trait continu).

Elle peut être explicitée en toute généralité :

$$P = Cste \times \sum_{k=0}^{n-c} (-1)^k \frac{\binom{n-c}{k}}{c+k} \int_0^1 [G(F^{-1}(u))]^{c+k} u^{b-1} (1-u)^{m-b} du$$

où la constante $Cste$ est égale à :

$$Cste^{-1} = B(c, n+1-c)B(b, m-b+1)$$

(voir, par exemple, Chakraborti & al., 2004 ou Balakrishnan & al., 2010). Les puissances sont donc explicites dans le cadre des deux alternatives de Lehmann. En particulier, sous une alternative de type Lehmann (2), on a $GF^{-1}(u) = 1 - (1-u)^\gamma$ et P peut alors s'exprimer comme une somme de fonctions beta :

$$P = Cste \times \sum_{k=0}^{n-c} (-1)^k \frac{\binom{n-c}{k}}{c+k} \sum_{h=0}^{c+k} (-1)^h \binom{c+k}{h} B(b, m + \gamma h + 1 - b).$$

2.1 La recherche du meilleur test des prédécesseurs

Considérons l'ensemble $\{P_1, \dots, P_m\}$ des tests des prédécesseurs de probabilité de fausse alarme égale au seuil nominal α . Le test des prédécesseurs optimal (Lin et Sukhatme, 1992) est le test des prédécesseurs dont la puissance est maximale pour une alternative donnée. Dans le cadre de l'alternative de Lehmann de type (2), le meilleur test des prédécesseurs est défini par :

$$b_\gamma = \arg \max_b \Pr_{HC, \gamma} (Y_{c(b, \alpha):n} < X_{b:m})$$

Afin de pouvoir comparer rigoureusement les puissances des tests en fonction de b , on a recours à des tests randomisés (tests mixtes). La figure 2 donne la puissance du test des prédécesseurs en fonction de b pour $m=100$ et $n=10$, sous l'hypothèse alternative de Lehmann de type (2) pour différentes valeurs de γ .

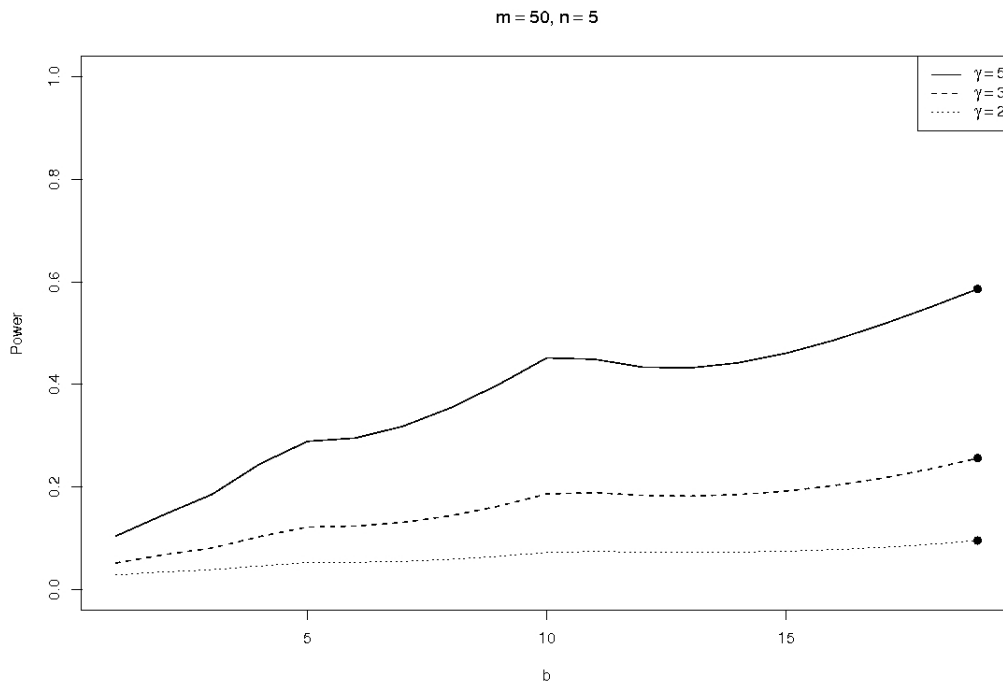


Figure 2 : puissance de $P_{(b)}$ en fonction de b pour l'alternative Lehmann (2) avec $\gamma=2, 3$ et 5 .

Pour m et n fixés, en partant de plusieurs valeurs de γ , on constate que le choix de b ne semble pas dépendre de l'alternative γ : la meilleure valeur de b est voisine de $m/2$ sous l'alternative de type Lehmann (2) et toutes (ou presque) les observations de l'échantillon à tester sont prises en compte.

2.3 La période opérationnelle moyenne

L'espérance du nombre h d'échantillons nécessaires pour déclarer le procédé hors contrôle (Période Opérationnelle Moyenne – *POM* – ou Average Run Length – *ARL*) à partir de l'instant où le procédé n'est plus sous contrôle et l'instant où le procédé est signalé hors contrôle par la carte peut être calculé :

$$ARL(F,G) = \frac{1}{B(b, m-b+1)} \int_0^1 \frac{1}{1 - \Pr_{F,G}(u)} u^{b-1} (1-u)^{m-b} du$$

avec $\Pr_{F,G}(u) = \Pr(V_{c:n} \geq G(F^{-1}(u)/u_{b:m} = u))$ où $(V_{i:n}), (U_{j:m})$ font référence à la statistique d'ordre associée à la loi uniforme pour des observations indépendantes. De même que pour le calcul de la puissance, on peut construire une version randomisée afin d'atteindre exactement un ARL_0 (i.e. l' ARL lorsque le procédé est sous contrôle) donné. La figure 3 illustre les résultats obtenus

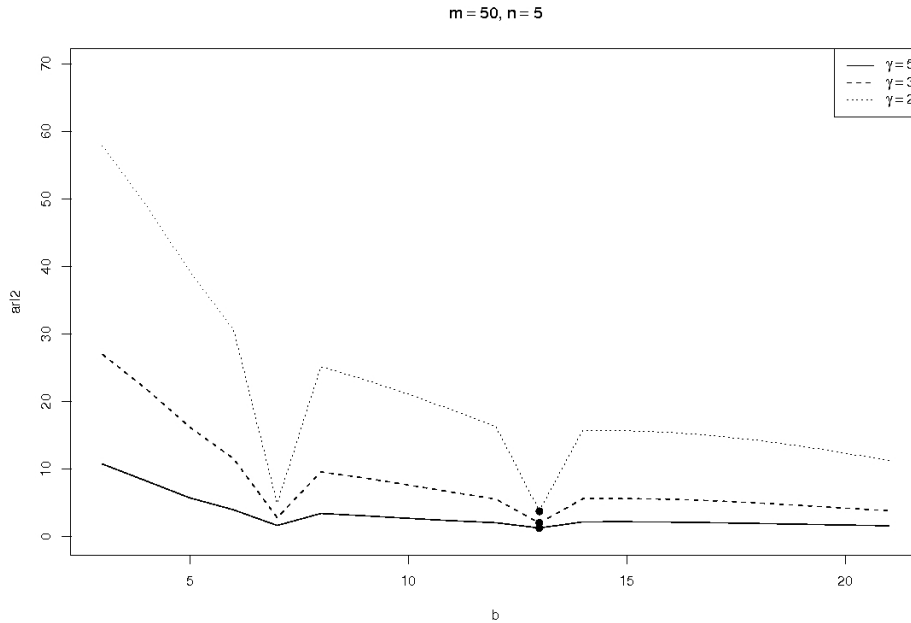


Figure 3 : ARL associé à la statistique $P_{(b)}$ en fonction de b pour l'alternative Lehmann (2) avec $\gamma=2, 3$ et 5 .

3. Extension à deux versions pondérées des statistiques de placement

Les statistiques de placement sont définies par (M_1, M_2, \dots, M_b) où :

$$M_j = \sum_{i=1}^n I_{\{X_{j-1:n}, X_{j:n}\}}(Y_i) \quad \text{pour tout } 1 \leq j \leq b$$

La statistique des prédécesseurs d'ordre b est donc la somme des b premières statistiques de placement. Pour éviter des phénomènes masquant, Balakrishnan et Ng (2006) ont proposé différents tests fondés sur les statistiques de placements. Deux versions pondérées des statistiques de placements sont proposées.

3.1 La statistique pondérée P_b^*

On considère une première statistique pondérée :

$$P_b^* = \sum_{j=1}^b (m-j+1)M_j$$

Le processus est déclaré hors contrôle si $P_b^* \geq c$. Si le procédé est sous contrôle ($G=F$), la distribution de la statistique des placements est la suivante :

$$\Pr_{SC}(M_1 = m_1, M_2 = m_2, \dots, M_b = m_b) = \frac{\binom{m+n - \sum m_i - b}{m-b}}{\binom{m+n}{n}}$$

La valeur critique du test est définie par $c_{\alpha,b}$ vérifiant :

$$\sum_{D(c_{b,\alpha})} \Pr_{SC}(M_1 = m_1, M_2 = m_2, \dots, M_b = m_b) = \alpha$$

où $D(c_{b,\alpha}) = \{(m_1, m_2, \dots, m_b) / \sum (m-i+1)m_i \geq c\}$. On présente les valeurs critiques et la puissance de ce test pondéré pour différentes valeurs de b . Le gain de puissance semble être modéré en regard du test des prédécesseurs P_b (figure 4).

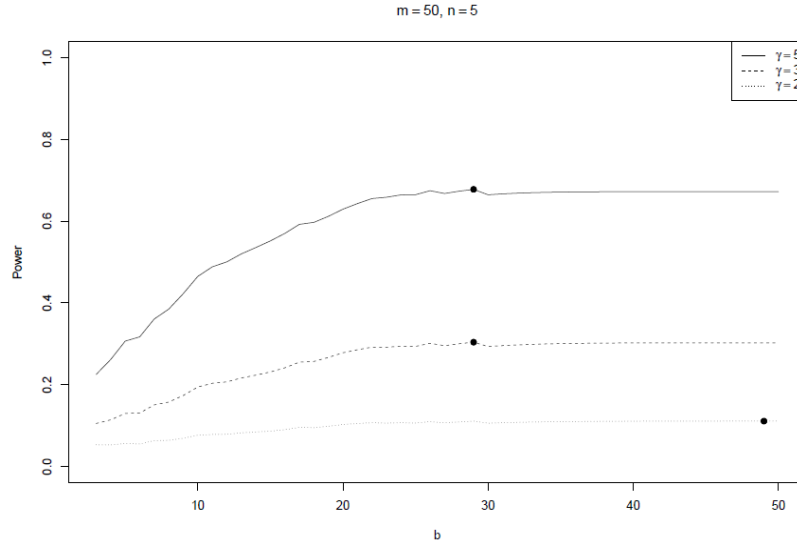


Figure 4 : puissance du test P_b^* pour l'alternative de type Lehmann2 en fonction de b .

3.2 La statistique de type Wilcoxon $P_{W_{Min,b}}$

On considère une seconde statistique pondérée :

$$P_{W_{Min,b}} = \sum_{j=1}^b (b-j+1)M_j$$

Il s'agit d'une statistique de type Wilcoxon (Balakrishnan et Ng, 2006) ; en particulier lorsque $b=m$, cette statistique définit un test équivalent à celui basé sur la statistique de Wilcoxon, $P_{W_{Min,m}} = W$.

Ainsi la famille des tests $\{P_{W_{Min,b}} : 1 \leq b \leq m\}$ englobe le test classique de Wilcoxon, Mann et

Whitney. Le processus est déclaré hors contrôle si $P_{W_{Min,b}} \geq c$ et la valeur critique du test est définie par $c_{\alpha,b}$ vérifiant :

$$\sum_{D(c_{b,\alpha})} \Pr_{cont}(M_1 = m_1, M_2 = m_2, \dots, M_b = m_b) = \alpha$$

où $D(c_{b,\alpha}) = \{(m_1, m_2, \dots, m_b) / \sum (b-i+1)m_i \geq c\}$. On présente les valeurs critiques et la puissance du test pondéré et du test de type Wilcoxon pour différentes valeurs de b (figure 4).

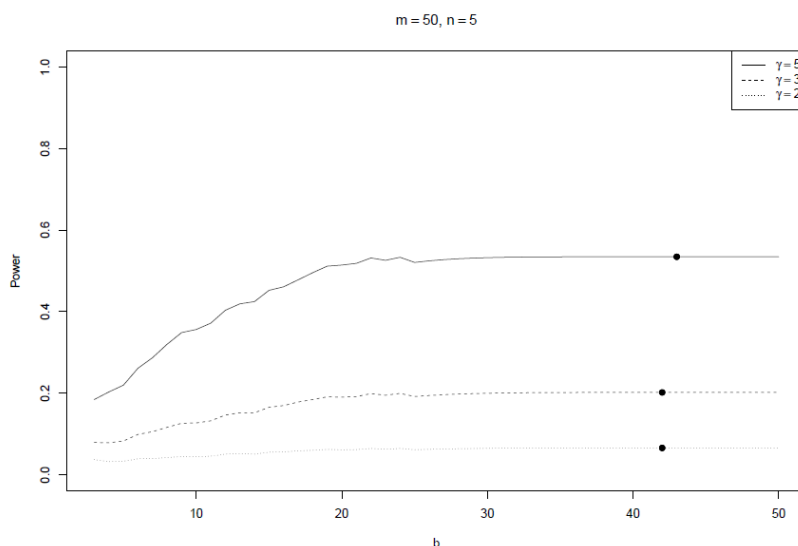


Figure 5 : puissance du test $P_{W_{\min,b}}$ pour l'alternative de type Lehmann2 en fonction de b .

3.3 Comparaisons avec le carte de contrôle basé sur la statistique des prédécesseurs

Le test de Wilcoxon Mann et Whitney est équivalent aux deux tests P_b^* et $P_{W_{\min,b}}$ lorsque $b=m$. On peut faire la conjecture de l'existence d'un test plus puissant que le test classique dans l'une de ces deux classes de tests pour un choix optimal de b lorsque l'on est en présence d'une alternative de type Lehmann (2). Cependant, le test de la médiane (P_b avec $b = m/2$), le test $P_{W_{\min,b}}$ ou P_b^* pour b^* compris entre $m/2$ et m , le test de Wilcoxon, Mann et Whitney ($b = m$) ont des puissances comparables. En se basant des calculs exactes et des illustrations numériques similaires, on a observé de meilleurs résultats pour ces deux dernières cartes de contrôle pour l'hypothèse de Lehmann de type (1). Ce résultat, en combinaison avec un autre établissant un lien entre les deux types d'hypothèses, nous donne espoir de proposer une carte de contrôle plus performante que celle de base étudiée dans la première partie.

Discussion

Dans cet article, nous proposons trois cartes de contrôle basées sur des fonctions simples du vecteur des statistiques de placement. La puissance et la période opérationnelle moyenne ont été calculées lorsque le processus est sous contrôle et sous l'hypothèse alternative de Lehmann de type (2). Les trois cartes de contrôle semblent donner des résultats analogues.

Bibliographie

- [1] Chakraborti. S et Graham M.A. (2007). Non parametric control charts. In *Encyclopedia of Statistics in Quality and Reliability*. Vol. 1; Ney-York. John Wiley, pp. 415-429.
- [2] Balakrishnan, N., Ng. H.K.T. (2006). Precedence Type Tests and Applications. Hoboken, NJ: John Wiley and sons.
- [3] Chakraborti. S., van Der Laan. P., van de Wiel, M.A. (2004). A class of distribution-free control charts. *J. Roy. Statist. Soc. Ser. C. Appl. Statist*, 53, pp 443-462.
- [4] N. Balakrishnan, I.S. Triantafillou, M.V. Koutras (2010). A distribution free Control Chart based on Order Statistics. *Communications in Statistics. Theory and methods*, 39, 3652-3677.
- [5] Lin G.H. & Sukhatme, S. (1992). On the choice of precedence tests. *Communications in Statistics. Theory and methods*, 21, 2949-2968.