REVUE DE STATISTIQUE APPLIQUÉE

B. FREDENUCCI

# Maîtrise statistique des procédés pour un processus d'observations autorégressif d'ordre 1

*Revue de statistique appliquée*, tome 41, n° 1 (1993), p. 37-54

<http://www.numdam.org/item?id=RSA_1993__41_1_37_0>

*Rev. Statistique Appliquée,* 1993, **XLI** (1), 37-54

# MAÎTRISE STATISTIQUE DES PROCÉDÉS POUR UN PROCESSUS D'OBSERVATIONS AUTORÉGRESSIF D'ORDRE 1

B. Fredenucci

*Université Mendès-France, Grenoble*

## RÉSUMÉ

La maîtrise statistique des procédés a pour but l'obtention d'une production d'une qualité régulière au meilleur niveau possible. La principale méthode pour y parvenir consiste à détecter l'apparition de causes anormales dites assignables pour prévenir leurs effets sur la production. Habituellement, elle est appliquée à des procédés pour lesquels les observations successives sont indépendantes. Cependant, cette hypothèse n'est pas toujours réaliste. Cet article est consacré au cas où les observations successives sont supposées issues d'un processus autorégressif d'ordre 1. Dans cette situation, les performances de la règle de décision sont modifiées et nous proposons une méthode, utilisant les chaînes de Markov, qui permet de déterminer les temps moyens de première décision de réglage. Ici, nous étudions plus particulièrement le cas où la règle de décision prend en compte les deux dernières observations. Le programme qui figure en annexe permet d'évaluer les performances obtenues en fonction des décalages de la moyenne selon la valeur du coefficient d'autocorrélation du modèle et les valeurs choisies pour les limites de décision.

***Mots-clés :*** *Maîtrise Statistique des Procédés, processus autorégressif d'ordre 1, règle de décision, temps moyen de première décision de réglage.*

## SUMMARY

The aim of statistical process control is to get production with a regular quality at the best possible level. The method to get it is to detect as soon as possible the appearance of assignable causes in order to prevent their effects on the production. Usually, SPC is used for the processes for which the sequence of observations is independent. However, this hypothesis is not always a realistic one. This paper is about the case for which the successive observations are supposed to come from a first order autoregressive process. In this case, the performances of the decision's rule are modified and we give a method, using the Markov chains approach, which produces the average run length. Here we are especially studying the case of the decision rule which used the two consecutive last observations. The program in the appendix allows the evaluation of the rule's performances depending on the shift of the mean, on the autocorrelation coefficient and on the choiced limits of the decision'rule.

***Key-words :*** *SPC, first autoregressive process, decision'rule, average run length.*

## 1. Introduction

La maîtrise statistique des procédés ou SPC selon la terminologie anglo-saxonne propose des méthodes qui permettent l'amélioration de la qualité au stade de la production. Ses bases ont été établies, dès 1924, lorsque Walter Shewhart a introduit la technique des cartes de contrôle. Le but assigné au SPC est celui de détecter l'apparition de causes anormales, dites assignables, qui perturbent le procédé. Une fois détectée, on peut procéder à l'élimination de la cause, ou, selon le contexte, à l'élimination de ses effets par recentrage du procédé sur la valeur cible.

En vue d'améliorer la sensibilité de la méthode de Shewhart face à de faibles perturbations, on a très tôt proposé une règle de décision basée sur un double système de limites (limites de surveillance, limites de contrôle) qui permet de prendre en compte deux observations consécutives.

A la base de la maîtrise statistique des procédés et de l'évaluation de ses performances potentielles figure l'hypothèse selon laquelle les observations successives sont indépendantes. Or, cette hypothèse, dont la faiblesse avait été perçue par Wetherill, dès 1969, semble devoir être mise en cause pour une large gamme de procédés. Nous nous proposons ici d'étudier le cas d'un modèle où les observations successives sont produites selon un processus autorégressif d'ordre 1.

L'objectif de cette étude est d'établir une méthode pour déterminer les performances de la règle de décision dans cette situation plus générale. Du fait du type de règle de décision envisagé, l'évaluation des performances s'effectue en se référant à l'ARL, temps moyen nécessaire pour décider d'un premier réglage (Average Run Length). L'expression de l'ARL que nous obtenons n'étant pas directement calculable, nous procédons à une discrétisation de l'ensemble des observables qui permet d'obtenir une expression approchée mais calculable de l'ARL.

## 2. Le modèle autorégressif et la règle de décision

Le modèle proposé est celui d'un processus autorégressif d'ordre 1 :

$$Z_n = \theta Z_{n-1} + X_n, n \in \{1, 2, ...\}$$

où la variable $Z_n$ représente l'observation d'ordre $n$ et $\theta$ le coefficient d'auto-corrélation ($\theta \in [0, 1[$). Les variables $X_n$ sont indépendantes et distribuées selon la même loi normale d'espérance $(1-\theta)k$ et de variance $\sigma^2$. La variable initiale $Z_0$ est supposée être égale à la valeur $k$ de sorte que le processus $(Z_n)_{n \geqslant 1}$ est centré sur la valeur $k$. La valeur cible correspondant au bon état de marche est $k = 0$.

Les données naturelles concernant ce processus sont le coefficient d'auto-corrélation $\theta$ et la variance asymptotique de $Z_n$, notée $\sigma_z^2$. Ces deux paramètres peuvent être estimés à partir d'une série d'observations suffisamment longue à condition qu'aucune cause assignable ne soit intervenue.

La variance $\sigma_z^2$ est la variance de référence car elle correspond aux seules sources aléatoires de variabilité de $Z_n$. On peut alors en déduire la variance $\sigma^2$

compte tenu de ce que $\sigma_z^2 = \dfrac{\sigma^2}{1-\theta^2}$. Nous exprimerons par la suite le paramètre $k$ en unité de $\sigma_z$, c'est-à-dire à l'aide du paramètre $K$ défini par $k = K\sigma_z$.

Lorsque $\theta = 0$, on retrouve le modèle des observations successives indépendantes. Le fait que ce modèle ait longtemps prévalu est peut-être lié à la difficulté de repérer visuellement une structure AR (1) sauf pour de fortes valeurs de $\theta$.

Dans des contextes voisins, de récents articles proposent pour $Z_n$ une modélisation encore plus générale, de type processus ARMA [Mac Gregor (1990); Faltin-Tucker-Hahn (1990)], qui consiste en deux équations :

$$Z_n = m_n + Y_n$$
$$m_n = \theta m_{n-1} + X_n$$

où $m_n$, la moyenne, est un processus AR (1) alors que $X_n$ et $Y_n$ sont des variables indépendantes entre elles et dans le temps. Le modèle AR (1) proposé, dans le cadre duquel nous évaluons ensuite les performances, apparaît comme le ARMA particulier où la variance de $Y_n$ est négligeable par rapport à celle de $X_n$. Ceci peut correspondre à la situation où $Y_n$ représente un facteur de variabilité dont la variance peut être réduite par répétition des mesures.

En ce qui concerne la règle de décision retenue pour cette étude, elle est plus générale que la règle initiale de Shewhart. Son usage est très répandu. Elle consiste à décider d'un réglage, non seulement lorsque l'observation courante est hors contrôle, mais aussi lorsque l'observation courante et celle qui la précède sont toutes deux hors surveillance. Cette règle permet une meilleure détection de faibles perturbations. Les limites de surveillance sont établies à $\pm$ LS $\sigma_z$ de la valeur cible, les valeurs usuelles de LS étant 1,96 ou 2. Les limites de contrôles sont établies à $\pm$ LC $\sigma_z$ de la valeur cible, les valeurs usuelles de LC étant 3,09 ou 3. Les trois zones délimitées par ces limites sont désignées dans la suite par :

Zone N : $[-\mathrm{LS}\sigma_z; \mathrm{LS}\sigma_z]$

Zone S : $[-\mathrm{LC}\sigma_z; -\mathrm{LS}\sigma_z[\cup]\mathrm{LS}\sigma_z; \mathrm{LC}\sigma_z]$

Zone C : la zone complémentaire de N $\cup$ S

La règle de Shewhart apparaît comme le cas particulier où LS = LC de sorte que la zone S est alors réduite à l'ensemble vide.

Pour évaluer les performances de ce type de règle de décision, il n'est plus possible de le faire en termes de probabilité de prendre une décision erronée (fonction puissance ou courbe d'efficacité), mais on doit avoir recours à la variable temps nécessaire pour décider d'un premier réglage. On étudie plus précisément son espérance, l'ARL, temps moyen nécessaire pour décider d'un premier réglage. L'ARL, qui est directement lié à la quantité moyenne produite, doit être grand en l'absence de déréglage et d'autant plus faible que le déréglage est important.

Une autre règle de décision est décrite par D.C. Montgomery et C.M. Mastrangelo (1991) dans le même contexte de processus ARMA. La variable mise sous contrôle n'est pas la variable observée comme nous le proposons ici mais l'écart entre la variable observée et la valeur prédite par le modèle. Cette méthode, dont l'inconvénient est de travailler sur une variable qui n'est pas directement interprétable, a par contre l'avantage d'effectuer le suivi d'observations qui se comportent comme étant non corrélées. Les performances de cette méthode ainsi que sa sensibilité aux erreurs éventuelles dues à l'identification du modèle ou à l'estimation des paramètres ne sont pas signalées par les auteurs.

## 3. Détermination de l'ARL

On désigne $T$ la variable premier instant de décision de réglage, à valeurs dans $\{1, 2, ..., n, ...\}$.

Pour une valeur $k = K\sigma_z$, appartenant à $N \cup S$, on pose :

$$\text{ARL}(K) = E_K(T) = \sum_{n \geqslant 1} nP[T = n], \text{ l'espérance de } T \text{ pour le décalage } K$$

On doit donc évaluer les probabilités :

$$\begin{aligned} P[T = n] &= P[Z_n \in C, Z_{n-1} \in N, T > n-1] \\ &\quad + P[Z_n \in C \cup S, Z_{n-1} \in S, T > n-1] \\ &= P[Z_{n-1} \in N, T > n-1] - P[Z_n \in N \cup S, Z_{n-1} \in N, T > n-1] \\ &\quad + P[Z_{n-1} \in S, T > n-1] - P[Z_n \in N, Z_{n-1} \in S, T > n-1] \end{aligned}$$

Cela conduit à déterminer des probabilités du type :

$P[Z_n \in B, Z_{n-1} \in A, T > n-1]$ pour des sous-ensembles $A$ et $B$ de $N \cup S$

Considérons alors la fonction $h^n(z)$ définie pour $z \in N \cup S$ par la récurrence suivante :

$$\begin{aligned} h^1(z) &= f_k(z)[1_N(k) + 1_S(k)1_N(z)] \\ h^n(z) &= 1_N(z) \int_{N \cup S} f_y(z) h^{n-1}(y) dy + 1_S(z) \int_N f_y(z) h^{n-1}(y) dy \end{aligned}$$

pour $n \geqslant 2$, où $1_B$ désigne la fonction indicatrice de l'ensemble $B$ et où $f_y(z)$ désigne la densité de la loi conditionnelle de $Z_n$ sachant $Z_{n-1} = y$, c'est-à-dire la densité de la loi de $\theta y + X_n$, qui est la loi $N(k + \theta(y-k); \sigma^2)$ où $\sigma^2 = (1-\theta^2)\sigma_z^2$, $\sigma_z^2$ étant la variance asymptotique de la variable $Z_n$.

On montre alors que pour tout sous-ensemble $A$ de $N \cup S$ on a :

$$P[Z_n \in A, T > n] \int_A h^n(z) dz$$

En effet, pour $A \subset N \cup S$ et compte tenu de la règle de décision, si $n = 1$ on a :

$$P[Z_1 \in A, T > 1] = \left| \begin{array}{ll} P[Z_1 \in A] = \int_A f_k(z) dz & \text{si } k \in N \\ P[Z_1 \in A \cap N] = \int_{A \cap N} f_k(z) dz & \text{si } k \in S \end{array} \right.$$

d'où le résultat compte tenu de la définition de $h_1(z)$.

si $n \geqslant 2$ on a : $P[Z_n \in A, T > n] = P[Z_n \in A \cap N, Z_{n-1} \in N \cup S, T > n-1]$
$+ P[Z_n \in A \cap S, Z_{n-1} \in N, T > n-1]$

Par suite, du fait que $Z_n = \theta Z_{n-1} + X_n$, que les variables $X_n$ et $Z_{n-1}$ sont indépendantes, utilisant l'hypothèse de récurrence on a :

$$\begin{aligned} P[Z_n \in A, T > n] &= \int_{N \cup S} \left[ \int_{A \cap N} f_y(z) dz \right] h^{n-1}(y) dy \\ &+ \int_N \left[ \int_{A \cap S} f_y(z) dz \right] h^{n-1}(y) dy \\ &= \int_A h^n(z) dz \end{aligned}$$

On déduit alors immédiatement le résultat suivant :

Pour tout sous-ensemble $A$ et $B$ de $N \cup S$ on a :

$$P[Z_n \in B, Z_{n-1} \in A, T > n-1] = \int_A \left[ \int_B f_y(z) h^{n-1}(y) dz \right] dy \tag{1}$$

Cette expression (1) n'étant pas directement calculable, nous proposons de recourir à une discrétisation de l'ensemble continu des valeurs que peut prendre l'observation $z$. Cette méthode, utilisée par Brook et Evans (1972), Fredenucci (1987) et plus récemment Lucas et Saccucci (1990a) pour les EWMA, permet d'approximer le processus markovien continu par une chaîne de Markov discrète. En reprenant les notations de Lucas et Saccucci (1990b), nous effectuons une partition de l'ensemble $N$ en un nombre impair d'intervalles, noté $n$, et de l'ensemble $S$ en un nombre pair d'intervalles, noté $m$ :

$$N = \bigcup_{i \in I} \Delta_i, I = \left\{ \frac{m}{2} + 1, \ldots, \frac{m}{2} + n \right\}$$
$$S = \bigcup_{i \in J} \Delta_i, J = \left\{ 1, \ldots, \frac{m}{2} \right\} \cup \left\{ \frac{m}{2} + n + 1, \ldots, n + m \right\}$$

Nous sommes alors conduits à déterminer les valeurs approchées d'expressions analogues à (1) mais pour des intervalles $\Delta$ :

$$P[Z_n \in \Delta_h, Z_{n-1} \in \Delta_i, T > n-1] = \int_{\Delta_i} \left[ \int_{\Delta_h} f_y(z) h^{n-1}(y) dz \right] dy$$

Nous allons pour cela procéder en deux étapes. Introduisons la fonction $f_z$, densité de la loi asymptotique de $Z_n$ qui est la loi $N(K\sigma_z; \sigma_z^2)$. Etant donné $z \in R$, le théorème de la moyenne nous permet de dire qu'il existe une valeur $y_i(z) \in \Delta_i$ pour laquelle la fonction $f_y(z)$ calculée en $y = y_i(z)$ est telle que :

$$\int_{\Delta_i} f_y(z) f_z(y) dy = f_{y_i(z)}(z) \int_{\Delta_i} f_z(y) dy$$

Il en résulte que :

$$\int_{\Delta_h} \left[ \int_{\Delta_i} f_y(z) f_z(y) dy \right] dz = \left( \int_{\Delta_h} f_{y_i(z)} dz \right) \left( \int_{\Delta_i} f_z(y) dy \right)$$

d'où le résultat intermédiaire :

$$\int_{\Delta_h} f_{y_i(z)}(z) dz = \frac{\displaystyle\int_{\Delta_h} \int_{\Delta_i} f_y(z) f_z(y) dy dz}{\displaystyle\int_{\Delta_i} f_z(y) dy} = P[Z_n \in \Delta_h | Z_{n-1} \in \Delta_i]$$

Par ailleurs, à condition que l'intervalle $\Delta_i$ soit suffisamment petit pour que l'approximation suivante soit valable :

$$f_y(z) \# f_{y_i(z)}(z) \qquad y \in \Delta_i, z \in \Delta_h$$

on a alors :

$$\int_{\Delta_i} f_y(z) h^{n-1}(y) dy \# f_{y_i(z)} \int_{\Delta_i} h^{n-1}(y) dy \qquad z \in \Delta_h$$

d'où

$$\int_{\Delta_h} \int_{\Delta_i} f_y(z) h^{n-1}(y) dz dy \# \left( \int_{\Delta_h} f_{y_i(z)}(z) dz \right) \left( \int_{\Delta_i} h^{n-1}(y) dy \right)$$

Alors, en utilisant le résultat intermédiaire et du fait que :

$$\int_{\Delta_i} h^{n-1}(y) dy = P[Z_{n-1} \in \Delta_i, T > n-1]$$

on obtient l'approximation suivante :

$$\begin{aligned} &P[Z_n \in \Delta_h, Z_{n-1} \in \Delta_i, T > n-1] \\ &\qquad \# P[Z_n \in \Delta_h | Z_{n-1} \in \Delta_i] \cdot P[Z_{n-1} \in \Delta_i, T > n-1] \qquad (2) \end{aligned}$$

Considérons de nouveau l'expression de $P[T = n]$. Du fait de la discrétisation on a :

$$P[T = n] = \sum_{i \in I} \left( 1 - \sum_{h \in I \cup J} P[Z_n \in \Delta_h, Z_{n-1} \in \Delta_i, T > n-1] \right.$$
$$\left. + \sum_{i \in J} \left( 1 - \sum_{h \in I} P[Z_n \in \Delta_h, Z_{n-1} \in \Delta_i, T > n-1] \right. \right.$$

et du fait de l'approximation (2) on obtient :

$$P[T = n] \# \sum_{i \in I} \left( 1 - \sum_{h \in I \cup J} P[Z_n \in \Delta_h \mid Z_{n-1} \in \Delta_i] \cdot P[Z_{n-1} \in \Delta_i, T > n-1] \right)$$
$$+ \sum_{i \in J} \left( 1 - \sum_{h \in I} P[Z_n \in \Delta_h \mid Z_{n-1} \in \Delta_i] \cdot P[Z_{n-1} \in \Delta_i, T > n-1] \right)$$

Dans cette expression, les termes $P[Z_n \in \Delta_h \mid Z_{n-1} \in \Delta_i]$ dépendent peu de l'instant $n$, au moins pour $n$ suffisamment grand, du fait que le processus $Z_n$ est asymptotiquement stationnaire. Pour les termes $P[Z_n \in \Delta_h, T > n]$ qui dépendent de l'instant $n$, nous pouvons établir les relations de récurrence approchées suivantes :

Pour $\Delta_h \subset N$ on a :

$$P[Z_n \in \Delta_h, T > n] = \sum_{i \in I \cup J} P[Z_n \in \Delta_n, Z_{n-1} \in \Delta_i, T > n-1]$$
$$\# \sum_{i \in I \cup J} P[Z_n \in \Delta_h \mid Z_{n-1} \in \Delta_i] P[Z_{n-1} \in \Delta_i, T > n-1]$$

d'après l'approximation (2).

De même, pour $\Delta_h \subset S$ on a :

$$P[Z_n \in \Delta_h, T > n] \# \sum_{i \in I} P[Z_n \in \Delta_h \mid Z_{n-1} \in \Delta_i] P[Z_{n-1} \in \Delta_i, T > n-1]$$

Pour unifier et simplifier la formulation, nous adoptons les notations suivantes :

$$p_n(\Delta_h) = P[Z_n \in \Delta_h, T > n]$$
$$p(h \mid i) = \left| \begin{array}{l} P[Z_n \in \Delta_h | Z_{n-1} \in \Delta_i] \text{ si } h \in I \text{ et } i \in I \cup J \\ 0 \text{ si } h \in J \text{ et } i \in J \end{array} \right.$$

Soit alors $P$ la matrice carrée d'ordre $n + m$ dont le terme général est la probabilité de transition $P_{ih} = p(h \mid i)$. D'après ce qui précède on obtient :

$$[p_n(\Delta_1) \dots p_n(\Delta_{n+m})] \# [p_{n-1}(\Delta_1) \dots p_{n-1}(\Delta_{n+m})] P$$

d'où

$$[p_n(\Delta_1)\dots p_n(\Delta_{n+m})]\#[p_0(\Delta_1)\dots p_0(\Delta_{n+m})]\,P^n$$

d'autre part :

$$P[T=n]\#[p_{n-1}(\Delta_1)\dots p_{n-1}(\Delta_{n+m})]\;[I-P]\,J$$

où $J$ est le vecteur colonne de terme général 1.

Il résulte de ces deux dernières approximations que :

$$P[T=n]\#[(p_0(\Delta_1)\dots p_0(\Delta_{n+m}]\,P^{n-1}[I-P]\,J$$

L'expression approchée suivante de l'ARL en découle :

$$\mathrm{ARL}(K)\#[p_0(\Delta_1)\dots p_0(\Delta_{n+m})]\sum_{n\geqslant 1} nP^{n-1}[I-P]\,J$$

c'est-à-dire

$$\mathrm{ARL}(K)\#[p_0(\Delta_1)\dots p_0(\Delta_{n+m})](I-P)^{-1}J$$

avec $p_0(\Delta_i)=1_{\Delta i}(k)$

L'évaluation approchée de l'ARL ne nécessite plus que la détermination des $p(h\mid i)$.

## 4. Méthode de calcul de la matrice $P$

La détermination de la matrice $P$ de probabilités de transition nécessite le calcul des probabilités conditionnelles du type :

$$p(h/i)=P[Z_n\in\Delta_h/Z_{n-1}\in\Delta_i]$$

Or, du fait que $\Delta_i$ est un intervalle non réduit à un point, la loi conditionnelle de $Z_n$ sachant $Z_{n-1}\in\Delta_i$ n'est pas une loi normale. Le calcul de ces probabilités conditionnelles nécessite d'avoir recours à un calcul approché. L'approximation la plus simple consiste à se ramener à une loi normale en considérant la loi conditionnelle de $Z_n$ sachant $\delta_i$ où $\delta_i$ est le milieu de l'intervalle $\Delta_i$ ; c'est, dans un autre contexte, le choix fait par Lucas et Saccucci (1990b) qui complétaient cette méthode par une extrapolation. Une autre solution serait d'adapter l'approximation normale préconisée par Mee et Owen (1983) pour la loi conditionnelle de $Z_n$ sachant $Z_{n-1}\in]-\infty,h]$.

La solution que nous proposons ici est l'intégration numérique de la densité de la loi conditionnelle de $Z_n$ sachant $Z_{n-1}\in D_i$. Cette méthode requiert l'utilisation d'un logiciel. En effet, pour un décalage $K\sigma_Z$ donné

$$\begin{aligned} P\big[Z_n \in [a\sigma_Z; b\sigma_Z]/Z_{n-1} \in [c\sigma_Z; d\sigma_Z]\big] \\ = \int_{a-K}^{b-K} f_{Z/Y\in[c-K;d-K]}(z)dz \end{aligned}$$

où $Z = \dfrac{Z_n - K}{\sigma_Z}$ et $Y = \dfrac{Z_{n-1} - K}{\sigma_Z}$ sont les variables centrées réduites.

On montre que,

d'une part

$$f_{Z/Y\in[c-K;d-K]}(x) = \frac{F(d-K)f_{Z/Y\leqslant d-K}(x) - F(c-K)f_{Z/Y\leqslant c-K}(x)}{F(d-K) - F(c-K)}$$

d'autre part

$$f_{Z/Y\leqslant c-K}(x) = \frac{F\left(\dfrac{c-K-\theta z}{\sqrt{1-\theta^2}}\right)}{F(c-K)} f(z)$$

où $F$ et $f$ sont respectivement la fonction de répartition et la densité de la loi normale centrée réduite.

Il en résulte que $p(h/i)$ vaut :

$$\frac{1}{F(d-K) - F(c-K)} \int_{a-K}^{b-K} \left[F\left(\frac{d-K-\theta z}{\sqrt{1-\theta^2}}\right) - F\left(\frac{c-K-\theta z}{\sqrt{1-\theta^2}}\right)\right] f(z)dz$$

où $\theta$ est le coefficient d'autocorrélation

$a\sigma_Z$ est la borne inférieure de $\Delta_h, b\sigma_Z$ la borne supérieure

$c\sigma_Z$ est la borne inférieure de $\Delta_i, d\sigma_Z$ la borne supérieure

Un logiciel approprié permet d'effectuer cette intégration numérique. Nous avons choisi le logiciel NAG.

A titre d'exemple, on obtient la matrice $P$ suivante lorsque $\theta = 0,25$ et $K = 0,5$ avec le choix $n = 3$ et $m = 4$ (cf. Tableau 1).

Les résultats sont exprimés en millième ($10^{-3}$).

## 5. Application

En guise d'illustration, on considère le cas où les limites de contrôle et de surveillance sont établies respectivement à $\pm 3,09\sigma_Z$ et à $\pm 1,96\sigma_Z$. Les valeurs de l'ARL, en fonction des décalages $K$ de la moyenne exprimés en unité de $\sigma_Z$, figurent dans le tableau 2 ci-dessous. Ces valeurs arrondies résultent du choix de la

TABLEAU 1. – Matrice $P$ des probabilités conditionnelles

| | S | | N | | | S | |
|---|---|---|---|---|---|---|---|
| S | 0 | 0 | 209 | 499 | 248 | 0 | 0 |
| | 0 | 0 | 174 | 490 | 288 | 0 | 0 |
| N | 1 | 5 | 113 | 447 | 369 | 47 | 14 |
| | 0,3 | 2 | 67 | 377 | 440 | 76 | 28 |
| | 0,1 | 1 | 38 | 297 | 485 | 112 | 48 |
| S | 0 | 0 | 21 | 230 | 500 | 0 | 0 |
| | 0 | 0 | 15 | 193 | 496 | 0 | 0 |

partition $n = 41$ $m = 50$. Les valeurs obtenues pour la partition $n = 11$ $m = 20$ sont parfois inférieures à 0,1 unité près.

TABLEAU 2. – Valeurs de l'ARL, exprimées en nombre d'unités consécutives produites, en fonction de $\theta$ coefficient d'autocorrélation et $K$ décalage de la moyenne exprimée en unité de $\sigma_Z$

| $K$ \ $\theta$ | 0 | 0,20 | 0,25 | 0,30 | 0,40 | 0,50 | 0,60 |
|---|---|---|---|---|---|---|---|
| 0 | 238,1 | 202,2 | 187,0 | 172,0 | 144,2 | 122,3 | 106,6 |
| 0,5 | 99,2 | 81,2 | 77,1 | 73,3 | 66,9 | 62,2 | 59,5 |
| 1 | 26,1 | 23,9 | 23,6 | 23,4 | 23,3 | 23,8 | 25,0 |
| 1,5 | 8,9 | 9,4 | 9,6 | 9,9 | 10,5 | 11,4 | 12,7 |
| 2 | 4,1 | 4,9 | 5,1 | 5,4 | 5,9 | 6,7 | 7,8 |
| 2,5 | 2,5 | 3,1 | 3,3 | 3,5 | 4,0 | 4,6 | 5,5 |

Par exemple, pour un procédé tel que $Z_n$ évolue dans le temps selon un processus autorégressif d'ordre 1 dont le coefficient d'autocorrélation est $\theta = 0,25$, les valeurs de l'ARL figurent dans la troisième colonne du tableau 2. Par comparaison au cas où $\theta = 0$, on constate une modification des performances de la règle de décision. Il y a une détérioration sensible en l'absence de déréglage, puisque la valeur de l'ARL est de 187 unités produites au lieu de 238, et une légère amélioration lorsque $K = 0,5$.

Pour retrouver une règle de décision dont l'ARL est de 238 en l'absence de déréglage, la seule solution est de choisir pour LC et LS des valeurs supérieures aux valeurs initiales. Le programme proposé en annexe permet d'évaluer l'ARL pour ces nouveaux choix. A titre d'exemple, le choix des valeurs LC = 3,03 et LS = 2,05 permet d'obtenir la valeur de 238,6 pour l'ARL en l'absence de déréglage; mais ce choix conduit à une valeur de l'ARL de 95 lorsque $K$ = 0,5.

## 6. Conclusions

Les techniques usuelles appliquées pour la mise en place de la maîtrise statistique des procédés sont encore valables lorsque les observations successives sont issues d'un processus autorégressif d'ordre 1. Cependant, dans cette situation, les performances des règles de décision usuelles se trouvent modifiées. L'analyse théorique qui est faite ici et l'outil pratique qui est proposé permettent d'appréhender ces modifications de performance en terme d'ARL, temps moyen de première décision de réglage. Le programme joint autorise le choix des valeurs des limites de contrôle et de surveillance pour atteindre un objectif de performance. Par ailleurs, les valeurs obtenues pour l'ARL dépendent, pour une faible part, du choix opéré pour discrétiser les intervalles de décision. Ainsi, on constate une légère sous-estimation de l'ARL en présence de fortes valeurs de $K$ lorsqu'on adopte une discrétisation comportant peu d'intervalles.

## Remerciements

L'auteur tient à remercier Pham Dinh Tuan et A. Le Breton pour l'intérêt qu'ils ont manifesté pour ce travail et pour leurs suggestions, ainsi que le rapporteur dont les remarques ont permis d'améliorer une première version de l'article.

## Bibliographie

BROOK D., EVANS D.A. (1972), "An approach to the probability distribution of CUSUM run length". *Biometrika*, 59, 539-549.

FALTIN F.W., TUCKER W.T. et HAHN G.J. (1990), "Integrating SPC with automated controls : concepts and experiences". Communication, journées Renault de la qualité, Paris (1990).

FREDENUCCI B. (1987), «Assurance qualité en fabrication dans le cas d'un processus autorégressif». Communication, Congrès ASU, Lausanne.

LUCAS J.M., SACCUCCI M.S. (1990a), "Exponentially weighted moving average control schemes : properties and enhancements". *Technometrics* 32, 1, 1-12.

LUCAS J.M., SACCUCCI M.S. (1990b), "Average run lengths for exponentially weighted moving average control schemes using the Markov chain approach". Computer programs, *Journal of quality technology*, 22 2, 154-162.

Mac GREGOR J.F.K. (1990), "A different view of the funnel experiment". *Journal of quality technology*, 22 4, 255-259.

MEE R., OWEN D. (1983), "A simple approximation for bivariate normal probabilities", *Journal of quality technology*, 15 2, 72-77.

MONTGOMERY D.C., MASTRANGELO C.M. (1991), "Some statistical process control methods for autocorrelated data", *Journal of quality technology*, 23 2, 179-193.

WETHERILL G.B. (1969), "Sampling inspection and quality control". Science paperbacks, Chapman and Hall.

## ANNEXE
## Programme et exemple numérique de son utilisation

PROGRAMME SOURCE

```
c       programme ARLAR (fortran)
c       Etude de l'ARL d'un processus AR(1)
c       Z(n)=PHI*Z(n-1)+X(n)
c       a l'exeption de PHI les autres parametres du modele
c       sont exprimes en unite de SIGMAZ l'ecart-type de Z ce qui
c       permet de prendre SIGMAZ=1. Ces parametres sont :
c       LSZ limite de surveillance, LCZ de controle, SHIFTZ decalage.
c       Les parametres de precision de calcul sont :
c       n le nombre impair d'intervalles constituant la zone centrale,N,
c       de la carte de controle et m le nombre pair d'intervalles
c       constituant la zone de surveillance,S. On calcule alors
c       la largeur des intervalles WIDTHN=2*LSZ/n et
c       WIDTHS=2*(LCZ-LSZ)/m exprimes en unites SIGMAZ.
c
c       Initilisation des parametres definis ensuite :

        integer nmax,ia,iunit
        parameter (nmax=201,ia=nmax,iunit=nmax)
        real phi
        real lsz,lcz ,li ,lk,x,y,pi
        real vshift(6)
        real p(201,201),q(201),u(201,201) ,arl(201),wkspce(201)
        real s15aef ,f01aaf,d01ahf
c       ces trois fonctions proviennent de la bibliotheque NAG
        integer npts,nlimit ,ifail,jfail
        external d01ahf, fun ,f01aaf
        common /cc/ phi, pi, a, b
        data vshift /0.,0.5,1.0,1.5,2.,2.5/
c       vshift constitue le vecteur des 6 decalages de la moyenne proposes,
c       en unite sigmaz. ces valeurs peuvent etre modifiees mais en restant
c       inferieures a lcz.

        write (*,1001)
1001    format (/'Coefficient PHI d''autocorrelation')
        read (*,*) phi
        write (*,1002)
1002    format (/'LSZ et LCZ a donner en unite SIGMAZ')
        read (*,*) lsz
        read (*,*) lcz
        write (*,1004)
1004    format(/'Nombre impair d''intervalles de la zone N n= ')
        read (*,*) n
        write (*,1006)
1006    format(/'Nombre pair ou nul d''intervalles de la zone S m= ')
```

```
      read (*,*) m
      if (lsz.eq.lcz) then
      m=0
      end if

c     Fin de l'entree des parametres du modele AR(1)

c     Initialisation de la matrice P ;creation des largeurs d intervalles

      write(*,2001)
2001  format(/' DECALAGE ARL ')

      do 1 ns=1,6
      do 20 i=1,max
      do 30 k=1,max
      p(i,k)=0
30    continue
20    continue
      shiftz= vshift(ns)
      wn=2*lsz/n
      if (m.ne.0) then
      ws =2*(lcz-lsz)/m
      end if

c     Determination de Q(i) = P(delta i)

      max=n+m
      m1=m/2
      n1 =m1+n
      li=lcz
      wi=ws

      do 10 i=1,max
      ifail=0
      j=i
      if (i.gt.m1 .and. i.le.n1) then
      j=i-m1
      li=lsz
      wi=wn
      end if
      if (i.gt.n1) then
      j=i-n1
      li=-lsz
      wi=ws
      end if

      a =-li+(j-1)*wi-shiftz
      b =-li+ j *wi-shiftz
```

```
c       la fonction S15AEF permet de calculer la fonction de repartition
c       de la loi Normale centree reduite F(t) :
c       F(t)=(1+s15aef(t/sqrt(2.)))/2
        a=a/sqrt(2.)
        b=b/sqrt(2.)
        ga=s15aef(a, ifail)
        gb=s15aef(b, ifail)
        ga=(1+ga)/2
        gb=(1+gb)/2
        q(i)=gb-ga
10      continue

c       a ce stade on a constitue le vecteur des denominateurs P(delta i)

c       Phase principale de la constitution de P,matrice de proba de transition
c       calcul de P(delta i * delta k) puis de P(k/i)=P(aller delta k/delta i)

        li=lcz
        wi=ws
        do 100 i=1,max
        lk=lcz
        wk=ws
        j=i
        if (i.gt.m1 .and. i.le.n1) then
        j=i-m1
        li=lsz
        wi=wn
        end if
        if (i.gt.n1) then
        j=i-n1
        li=-lsz
        wi=ws
        end if
        do 101 k=1,max
        l=k
c       detection du cas : delta i et delta k inclus en S.alors p(k/i) reste nul
        if (i.le.m1.or.i.gt.n1 ) then
        if (k.le.m1.or.k.gt.n1 ) then
        goto 101
        end if
        end if

        if (k.gt.m1 .and. k.le.n1) then
        l=k-m1
        lk=lsz
        wk=wn
        end if
        if (k.gt.n1) then
        ...
```

```
      l=k-n1
      lk=-lsz
      wk=ws
      end if
c
      a=-li+(j-1)*wi
      b=-li+ j *wi
      c=-lk+(l-1)*wk
      d=-lk+ l *wk
c
c     evaluation des proba p(k/i) par integration numerique
      a=a-shiftz
      b=b-shiftz
      c=c-shiftz
      d=d-shiftz

      nlimit=0
      epsr=1.0e-5
      ifail=0
      crelerr=0
c     la fonction D01AHF calcule numeriquement l integrale de la
c     fonction FUN entre c et d.
      ans=d01ahf(c, d, epsr,npts,relerr,fun, nlimit, ifail)
      p(i, k)=ans
      p(i, k)=p(i, k)/q(i)
      jfail=jfail+ifail
      crelerr=crelerr+relerr

101   continue
100   continue

c     constitution de la matrice (I-P), notee P et de son inverse Q

      data one/1.d0/
      do 14 i=1, max
      do 13 k=1, max
      p(i, k)=-p(i,k)
13    continue
      p(i, i)=one+p(i,i)
14    continue

c     inversion de la matrice et calcul de l ARL.

      ifail=1
c     la fonction F01AAF fournit l inverse de la matrice P et la nomme U
      call f01aaf(p, ia , max,u,iunit,wkspce,ifail)
      if (ifail.ne.0) then
      write (*,*) ifail
```

```
        end if
        do 220 i=1,max
        do 221 j=1,max
        arl(i)=arl(i)+u(i,j)
221     continue
220     continue
        nn=(m+n-1)/2+1
        write(*,*)
        write(*,2002) shiftz,arl(nn)
2002    format (5x,f3.1,12x,f9.4)
        do 3 i=1,max
        arl(i)=0
3       continue
1       continue
        end

c       la fonction FUN calcule la fonction a integrer numeriquement
c
        real function fun(x)
        common /cc/ phi, pi, a ,b
        real x, y, z, g, h
        real s15aef
        data pi /3.141592654/
        y=(a-phi*x)/sqrt(1-phi**2)
        z=(b-phi*x)/sqrt(1-phi**2)
        y=y/sqrt(2.)
        z=z/sqrt(2.)
        g=s15aef(y, ifail)
        h=s15aef(z, ifail)
        g=(1+g)/2.
        h=(1+h)/2.
        fun=(h-g)*exp(-(x**2)/2.)/sqrt(2.*pi)
        return
        end
```

## EXEMPLE D UTILISATION DU PROGRAMME ARLAR

Coefficient PHI d'autocorrelation

? .25

LSZ et LCZ a donner en unite SIGMAZ

? 1.96

? 3.09

Nombre impair d'intervalles de la zone N n=

? 41

Nombre pair ou nul d'intervalles de la zone S m=
? 50

| DECALAGE | ARL |
|---|---|
| 0.0 | 186.9869 |
| .5 | 77.0778 |
| 1.0 | 23.5793 |
| 1.5 | 9.6172 |
| 2.0 | 5.1007 |
| 2.5 | 3.2834 |