REVUE DE STATISTIQUE APPLIQUÉE

J. TIAGO DE OLIVEIRA
SÉBASTIAN B. LITTAUER

Cartes de contrôle à double limites et à séquences

Revue de statistique appliquée, tome 13, nº 2 (1965), p. 61-73
<http://www.numdam.org/item?id=RSA_1965__13_2_61_0>

CARTES DE CONTROLE A DOUBLE LIMITES ET A SEQUENCES (1)

J. TIAGO DE OLIVEIRA
Université de Columbia (N.Y.) et Université de Lisbonne

Sébastian B. LITTAUER
Université de Columbia (N.Y.)

I - INTRODUCTION ET RESUME

Dans le contrôle économique d'un processus de production on peut considérer deux phases dans l'application des concepts de contrôle statistique. Lors de l'introduction d'un processus nouveau (ou de sa reprise après une longue interruption) on applique le contrôle statistique pour amener le processus à un état de stabilité statistique. On utilise donc l'expression "contrôle statistique" dans un sens dynamique en considérant toutes les opérations nécessaires a l'obtention de cet état désiré et celle de "stabilité statistique" quand on se rapporte a cet état qu'on veut maintenir, après son obtention. Une fois réalisée la stabilité statistique, le contrôle statistique est un instrument de vérification que sert à maintenir la stabilité statistique désirée.

Pendant l'application du contrôle statistique (au sens de Shewhart (1939)), on utilise des limites de contrôle a 3 écarts-type (Critère I, Shewhart (1939), p. 29) comme limites d'action ; leur justification a plutôt un caractère empirique-économique qu'une base statistique formelle. Quand, cependant, le processus a atteint la phase de stabilité statistique et que l'on utilise le Critère I (ou un autre critère), pour le contrôler en vue d'agir lorsque l'état de stabilité statistique devient suspect, les propriétés statistiques du critère sont importantes. Il est alors convenable de connaître la probabilité d'une erreur de première espèce (Shewhart (1939), pp. 39, 40) pour pouvoir comparer des critères différents. Contrairement à ce qui se passe lors de la recherche de la stabilité statistique, où l'application du contrôle statistique est faite empiriquement et d'une manière séquentielle pour vérifier si la stabilité statistique a été atteinte, on peut, pendant la stabilité statistique qui s'ensuit, supposer l'existence d'une distribution normale pour X (variable aléatoire du processus) avec $\mu = \mu_o$ et $\sigma = \sigma_o$. Le critère utilisé dans le contrôle statistique peut alors être considéré comme un test de l'hypothèse statistique $H_o : \mu = \mu_o$ contre l'alternative $H_\lambda : \mu = \mu_o + \lambda\sigma_o$, μ_o et σ_o étant connus ($\lambda \neq 0$).

Dans ces conditions, le contrôle se réduit à une succession de tests d'hypothèses (en supposant l'indépendance des moyennes $\overline{X}_i$ successives) où l'erreur de première espèce est exprimée par la fréquence moyenne des échantillons pour lesquels le critère de contrôle a été violé alors que l'hypothèse H_o est vraie. On peut de façon analogue, exprimer l'erreur

---

(1) Cette étude a été subventionnée par l'Office of Naval Research, Contrat ONR 266(04). Elle peut être reproduite totalement ou en partie, à la libre disposition du Gouvernement des Etats-Unis.

du deuxième type, lorsque μ prend la valeur $\mu_0 + \lambda\sigma_0$ ($\lambda \neq 0$), σ_0 demeurant constant. Pendant cette étude on va considérer surtout les variations de μ, car, en général, σ est plus stable que μ. Le critère choisi est le *temps moyen d'action ou d'attente* (T.M.A.), c'est-à-dire le nombre moyen d'échantillons (en incluant celui qui détermine l'action) entre actions successives, le mot action représentant ici la décision d'examiner le processus afin de rechercher des causes possibles de variation(*).

Nous allons déterminer la formule exacte pour le T. M. A. par une technique différente de celle de Kemp (1961). Les résultats analytiques qu'on va obtenir fournissent une base pour comparer les cartes de contrôle de sommes cumulées et les cartes de contrôle à doubles limites et à séquences qui sont utilisées de temps à autre avec un fondement plutôt empirique.

Ces idées peuvent être appliquées à d'autres cartes de contrôle, où les variables aléatoires ont des distributions différentes.

2 - LES CARTES DE CONTROLE A DOUBLE LIMITES ET LA CARTE DE CONTROLE A SEQUENCES

Supposons que X est une variable aléatoire continue représentant les résultats d'un processus avec une même distribution et soit $\overline{X}$ la moyenne d'un échantillon de n valeurs successives de X. Soit $\overline{X}_j$ la moyenne de l'échantillon d'ordre j de dimension n, les échantillons successifs se succedant à des intervalles fixés à priori. Si le processus est statistiquement stable, les $\overline{X}_j$ peuvent être considérés comme étant identiquement et indépendement distribuées. En général, la distribution de X n'est pas normale mais on peut quand même dans un grand nombre de situations supposer que cette distribution est approximativement normale avec une moyenne μ et un écartetype σ. Comme quel que soit le cas, les $\overline{X}$ sont approximativement normalement distribuées, on peut toujours supposer que les $\overline{X}$ ont une distribution normale de moyenne μ et d'écart type $\sigma/\sqrt{n}$.

Le critère I de Shewhart nous conduit aux limites de contrôle (supérieure et inférieure) suivantes :

$$\text{L. C. } : \overline{\overline{X}} \pm 3\,\sigma'/\sqrt{n}$$

($\overline{\overline{X}} = \sum_1^N \overline{X}_j/N$; σ' estimation de σ), écart-type de l'échantillon) la justification du multiplicateur 3 est connue. On trouve cependant utile d'avoir aussi des limites de surveillance données par l'expression :

$$\text{L. S. } : \overline{\overline{X}} \pm 2\,\sigma'/\sqrt{n}$$

On peut utiliser le critère suivant : quand une valeur $\overline{X}_j$ est plus grande que L. C. Sup. ou plus petite que L. C. Inf. où alors quand deux

----

(*) Le concept du T. M. A. à été essentiellement introduit pour Aroian et Levene (1950) dans la notion de "decision points". L'expression "average run length" à été introduite par Kemp (1961) pour les cartes de contrôle de sommes cumulées. Quoiqu'il ne paraisse point désirable d'augmenter une terminologie déjà longue, il nous semble que l'expression "temps moyen d'action (ou d'attente)", non seulement évite toute confusion avec la notion de séquence (run) et de limites de contrôle, mais aussi exprime d'une manière convenable son idée même, c'est-à-dire le temps moyen entre deux actions , de plus les cartes de contrôle ont aussi pour but la recherche de causes de variation.

valeurs successives de $\overline{X}_j$ se trouvent entre $\overline{\overline{X}} - 3\sigma'/\sqrt{n}$ et $\overline{\overline{X}} - 2\sigma'/\sqrt{n}$ ou entre $\overline{\overline{X}} + 2\sigma'/\sqrt{n}$ et $\overline{\overline{X}} + 3\sigma'/\sqrt{n}$ on décidera l'action.

Pour des raisons en même temps pratiques et analytiques, il est convenable de généraliser le critère en choisissant deux paramètres, w et a, tels que $0 < w \leqslant a$, la carte de contrôle menant à la décision d'action lorsque $\overline{X}_j \leq \overline{\overline{X}} - a\sigma'/\sqrt{n}$ ou $\overline{X}_j \geqslant \overline{\overline{X}} + a\sigma'/\sqrt{n}$ ou encore, si l'on a soit $\overline{\overline{X}} - a\sigma'/\sqrt{n} < \overline{X}_{j-1}$, $\overline{X}_j < \overline{\overline{X}} - w\sigma'/\sqrt{n}$, soit $\overline{\overline{X}} + w\sigma'/\sqrt{n} < \overline{\overline{X}}_{j-1}$, $\overline{\overline{X}}_j < X + a\sigma'/\sqrt{n}$.

Nous appelons ceci un *contrôle statistique à doubles limites*, ou mieux, quand nous parlerons de la carte de contrôle, nous l'appellerons *carte de contrôle à doubles limites*. On voit maintenant que le critère classique I est un cas particulier pour lequel on a w=a=3.

Il est fréquent d'utiliser un critère a séquences sans avoir reconnu un Critère I (Wolfowitz) (1943)), ou alors simultanément avec ce même critère. Dans le deuxième cas, la longueur maximum des séquences a été fixée d'une façon empirique et les tentatives de quantification faites sont restées plutôt vagues et imprécises. Nous nous réferrons à *la carte de contrôle à séquences* qui utilise le critère suivant : on décidera d'agir si $\overline{X}_j \leqslant \overline{\overline{X}} - c\sigma'/\sqrt{n}$ ou si $\overline{X}_j \geqslant \overline{\overline{X}} + c\sigma'/\sqrt{n}$ ou encore si $\overline{\overline{X}} - c\sigma'/\sqrt{n} < \overline{X}_{j-R+1}, \ldots, \overline{X}_j < \overline{\overline{X}}$ ou si $\overline{\overline{X}} < \overline{X}_{j-R+1}, \ldots, \overline{X}_j < \overline{\overline{X}} + c\sigma'/\sqrt{n}$, où c,R sont deux paramètres, $c > 0$, $R > 1$ (R entier). Ceci est un critère utile et simple dont on va établir les propriétés analytiques en utilisant le concept de T. M. A. Ce critère se réduit à celui de Shewhart si $R = \infty$.

Les propriétés qu'on va établir permettent non seulement une grande flexibilité dans le choix du critère statistique de décision mais aussi un choix convenable (du point de vu économique) des limites de contrôle.

3 - LE CONCEPT DE TEMPS MOYEN D'ACTION

Les critères de contrôle statistique ayant des applications opérationnelles définies, il nous faut des mesures qui permettent de comparer ces critères, mesures qui doivent avoir une signification intuitive. Lorsque, l'on contrôle un processus qui a atteint la stabilité statistique, le nombre d'échantillons entre deux échantillons successifs violant le critère de décision est une indication de plus haut intérêt. En particulier la fréquence des "violations" pour H_0 peut être considérée comme l'analogue de l'erreur de première espèce. Une difficulté inhérente à la nature statistique des critères présentés ici nous interdit d'utiliser comme mesure de leur efficacité la fonction de puissance classique. En effet, bien que les échantillons successifs soient indépendants, une violation du critère statistique entraînant l'action peut dépendre des moyennes antérieures. Au lieu de la fonction de puissance nous allons utiliser le temps moyen d'action où le temps d'action est mesuré au moyen du nombre d'échantillons entre deux actions successives, en considérant également dans ce nombre l'échantillon qui correspond à la dernière action, c'est-à-dire lorsque le critère choisi est violé.

En utilisant le T. M. A., nous représenterons la moyenne et la variance de chaque variable aléatoire par μ et σ et les écarts de ces valeurs par : $\lambda = (\mu - \mu_0)/\sigma_0$ et $\delta = \sigma/\sigma_0$ respectivement, où μ_0 et σ_0 sont les paramètres du processus. Si $P_j(\lambda, \delta)$ représente la probabilité de décider l'action après l'échantillon d'ordre "j" (compté à partir de l'action précédente), le T. M. A. est défini par :

$$T(\lambda, \delta) = P_1 + \sum_{2}^{\infty} j(P_j - P_{j-1}) \qquad (1)$$

On sait que la condition nécéssaire et suffisante pour l'existence de $T(\lambda, \delta)$ est que $\lim_{j\to\infty} j(1-P_j) = 0$. Il est facile de voir, en se servant de la définition, que P_j est toujours plus grande que la probabilité d'action P_j' dans la carte de contrôle usuelle de Shewhart, (P_j' est la somme des premiers termes d'une série géométrique), dont les limites coincident avec les limites d'action de la carte de contrôle à double limite ; comme $j(1 - P_j') \longrightarrow o$ et $P_j \longrightarrow P_j'$ nous avons $j(1 - P_j) \longrightarrow o$; utilisant ce fait nous obtenons immédiatement pour le T. M. A. l'expression

$$T(\lambda, \delta) = 1 + \sum_{1}^{\infty} (1 - P_j) \qquad (1')$$

qui est préférable pour les calculs ; nous voyons également que $P_j \longrightarrow 1$ ce qui nous montre, comme on pouvait s'y attendre, qu'il nous faut toujours se décider pour l'action à un moment ou un autre. Il est évident que $T(\lambda, \delta)$ est une fonction de μ_o, σ_o, μ et σ par l'intermédiaire de λ et δ, que l'on étudie pour des valeurs données de λ et δ. Il est raisonable d'étudier principalement T pour $\delta = 1$ car, on stabilise premièrement σ ce qui permet d'étudier objectivement les variations de "μ" ; en plus, la variance du processus est, en général, plus stable que sa valeur moyenne (Western Electric (1958)). De toutes manières on étudiera $T(\lambda, \delta)$ pour des combinaisons différentes de λ et de δ, celles du type $(\lambda, 1)$ ayant un intérêt particulier. Il est facile de voir que la fonction $T(\lambda, 1)$ est symétrique en "λ". Nous pouvons établir une relation entre le T. M. A. et la notion bien connue de *période de retour* (Gumbel (1958)) dont l'usage s'est répandu dans les applications de la statistique aux problèmes des ingénieurs :

Considérons une succession d'expériences où l'on a une probabilité constante "p" d'obtenir un résultat donné (violation d'une borne, besoin de contrôle, etc.). La probabilité que l'événement se produise lors de l'expérience d'ordre "j" est : $(1 - p)^{j-1}p$; on a donc pour le T. M. A. $T = \sum_{1}^{\infty} j(1 - p)^{j-1}p = 1/p$. Si $p = 0,05$ on aura $T = 20$ c'est-à-dire qu'en "moyenne", le délai entre deux échantillons successifs où se produit ce résultat est de 20 réalisations. L'interprétation d'une probabilité constante a donc été donnée en termes de T. M. A.

Comme la probabilité d'action lors de l'échantillon d'ordre "j" dépend de la moyenne des échantillons précédents, la fonction de puissance classique n'est plus applicable dans ce cas. A sa place les valeurs $T(0,1)$ et $T(\lambda, 1)$ peuvent être utilisées au lieu des erreurs de première et de deuxième espèces. On obtiendra des expressions exactes pour $T(\lambda, \delta)$ pour les cartes à doubles limites et à séquences aussi bien que quelques propriétés utiles. Des tables pour des valeurs choisies de (λ, δ) sont en cours de calcul, au moyen de ces expressions afin de permettre de comparer l'efficacité du critère que nous avons choisi à celles du critère I de Shewhart et du critère de la carte des sommes cumulées. A la différence de celles de Kemp (1961) et de Page (1962) les formules obtenues ici sont exactes. La règle pour la combinaison de valeurs de T. M. A. suivie par Kemp (1961) et Page (1962) est la suivante : décomposons l'action A en deux actions indépendantes et disjointes A_1 et A_2 et soient T_1 et T_2 les T. M. A. correspondents. Le T. M. A. de l'action composée est donné par :

$$1/T = 1/T_1 + 1/T_2$$

Pour des valeurs grandes du T. M. A. cette formule est, en général une bonne approximation, mais l'on peut en donner un contre-exemple simple qui nous montre qu'elle n'est pas exacte.

Supposons que nous utilisons la carte de contrôle usuelle de Shewhart avec un décalage d'un échantillon, c'est-à-dire, que quel que soit le résultat du premier échantillon nous accepterons que l'état de stabilité se trouve réalisé et n'agirons point. Représentant par p_1 et p_2 les probabilités de violation des limites supérieure et inférieure, nous obtenons immédiatement : $T = 1 + 1/(p_1 + p_2)$, $T_1 = 1 + 1/p_1$ et $T_2 = 1 + 1/p_2$ et nous avons alors : $1/(T - 1) = 1/(T_1 - 1) + 1/(T_2 - 1)$ et non la règle de Kemp pour la combinaison des T. M. A.. Montrons maintenant que cette règle-ci est une bonne approximation. Nous allons maintenir la signification de p_1 et p_2 pour le cas de la carte de contrôle à double limite ou à séquences. Comme : $T = \frac{1 - \varepsilon}{p_1 + p_2}$, $T_1 = \frac{1 - \varepsilon_1}{p_1}$ et $T_2 = \frac{1 - \varepsilon_2}{p_2}$ avec $0 \leqslant \varepsilon, \varepsilon_1, \varepsilon_2 < 1$, en éliminant p_1 et p_2 nous obtenons l'expression :

$$\frac{1 - \varepsilon}{T} = \frac{1 - \varepsilon_1}{T_1} + \frac{1 - \varepsilon_2}{T_2}$$

Si l'on représente par T_o la solution de l'équation : $\frac{1}{T_o} = \frac{1}{T_1} + \frac{1}{T_2}$ nous obtenons :

$$\frac{T}{T_o} = \frac{(1 - \varepsilon)(T_1 + T_2)}{(1 - \varepsilon_1)T_2 + (1 - \varepsilon_2)T_1}$$

En posant $\bar{\varepsilon} = \max(\varepsilon, \varepsilon_1, \varepsilon_2)$ nous obtenons immédiatement :

$$1 - \bar{\varepsilon} < T/T_o \leqslant 1/(1 - \bar{\varepsilon})$$

Pour $\bar{\varepsilon} = 1/10$ on obtient $\frac{9}{10} < \frac{T}{T_o} < \frac{10}{9}$ ce qui montre que T_o est une approximation raisonable de T.

4 - PROPRIETES DE LA CARTE DE CONTROLE A DOUBLES LIMITES

En accord avec les notations antérieures, soit P_j la probabilité d'un temps pour l'action de longueur au plus égale a "j", c'est-à-dire la probabilité que après une action, on aura une séquence de k-1 échantillons sans prendre aucune action suivie d'une action à l'échantillon d'ordre "$k \leqslant j$". On peut écrire :

$$P_j = P_j(\lambda, \delta \mid w, a) \tag{2}$$

dès que "P_j" dépend de la dimension "n" de l'échantillon on prendra en considération cette dimension lorsque on en éprouvera le besoin.

Définissons les probabilités suivantes :

$$A = \text{Prob}\{-a < Z < -w\} = \Phi\left(-w - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(-a - \sqrt{n}\frac{\lambda}{\delta}\right)$$

$$W = \text{Prob}\{-w \leqslant Z \leqslant -w\} = \Phi\left(w - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(-w - \sqrt{n}\frac{\lambda}{\delta}\right)$$

$$A = \text{Prob}\{w < Z < a\} = \Phi\left(a - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(w - \sqrt{n}\frac{\lambda}{\delta}\right)$$

où

$$Z = \sqrt{n}\,\frac{\overline{X} - \mu_o}{\sigma_o}$$ est une variable aléatoire normale et :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\,(-\,t^2/2)\,dt \tag{4}$$

A_1 représente la probabilité de la moyenne d'un échantillon prenne une valeur située entre les limites de surveillance et d'action inférieures, A_2 est l'analogue pour les limites supérieures et W la probabilité que la moyenne prenne une valeur située entre les limites de surveillance ; ces fonctions dépendent de "n" et de "λ" par $\sqrt{n}\,\lambda$. Nous allons maintenant poursuivre le calcul de T. M. A. en termes de A_1, A_2 et de W au moyen d'un système d'équations aux différences finies.

Soient B'_{1j}, et B'_{2j} les probabilités pour l'échantillon d'ordre "j" d'avoir une moyenne entre deux limites inférieures ou supérieures respectivement et B_{1j} et B_{2j} les probabilités que les échantillons d'ordre "j-1" et "j" tombent entre les mêmes limites.

La probabilité d'action lors de l'échantillon d'ordre "j", $(P_j - P\ )$, est donnée par l'équation :

$$P_j - P_{j-1} = (1 - P_{j-1})\,(1 - A_1 - A_2 - W) + B_{1j} + B_{2j} \tag{5}$$

où le premier terme du second membre représente la probabilité que les "j-1" premiers termes ne conduisent point à l'action et que la moyenne de l'échantillon d'ordre "j" tombe au dehors de l'intervalle défini par les limites pour l'action. Les deuxième et troisième termes représentent les probabilités d'action en conséquence de ce que les moyennes des échantillons d'ordre "j-1" et "j" tombent entre les deux limites inférieures ou supérieures.

Les relations

$$\begin{aligned} (1 - P_{j-1})\,A_1 &= B_{1j} + B'_{1j} \\ (1 - P_{j-1})\,A_2 &= B_{2j} + B'_{2j} \end{aligned} \tag{6}$$

sont également immédiates car si les "j-1" premiers échantillons ne conduisent point a l'action et si l'échantillon d'ordre "j" tombe entre les limites supérieures ou inférieures nous avons nécessairement un ou deux échantillons entre les deux limites inférieures ou supérieures. Enfin les relations :

$$B_{1j} = A_1\,B'_{1,j-1}\ , \quad B_{2j} = A_2\,B'_{2,j-1}$$

sont évidentes. En éliminant les B' nous obtenons les 3 équations aux différences finies suivantes :

$$\begin{aligned} P_j - P_{j-1} &= (1 - A_1 - A_2 - W)\,(1 - P_{j-1}) + B_{1j} + B_{2j} \\ (1 - P_{j-1})\,A_1 &= B_{1j} + B_{1j+1}/A_1 \\ (1 - P_{j+1})\,A_2 &= B_{2j} + B_{2j+1}/A_2 \end{aligned} \tag{7}$$

avec les conditions initiales :

$$\begin{aligned} B_{11} = B_{21} = 0 \quad ; \quad B_{12} = A_1^2 \quad ; \quad B_{22} = A_2^2 \\ P_1 = 1 - A_1 - A_2 - W \end{aligned} \tag{8}$$

Nous allons maintenant calculer T. Utilisant la formule (1'), soient B_1 et B_2 les sommes des séries $\sum_1^\infty B_{1j}$ et $\sum_1^\infty B_{2j}$ lesquelles convergent parce qu'elles sont majorées par la série $\sum_2^\infty (P_j - P_{j-1})$. En sommant membre à membre les trois successions d'équations pour j variant de 2 à l'infini on obtient :

$$\begin{aligned} 1 - P_1 &= (1 - A_1 - A_2 - W)(T - 1) + B_1 + B_2 \\ (T - 1) A_1 &= B_1 + (B_1 - A_1^2)/A_1 \\ (T - 1) A_2 &= B_2 + (B_2 - A_2^2)/A_2 \end{aligned} \qquad (9)$$

car

$$T - 1 = \sum_1^\infty (1 - P_j), \quad \sum_2^\infty B_{1j} = B_1,$$
$$\sum_2^\infty B_{2j} = B_2, \quad \sum_3^\infty B_{1j} = B_1 - A_1^2, \quad \sum_3^\infty B_{2j} = B_2 - A_2^2$$

compte tenu des conditions initiales. En substituant les valeurs de B_1 et B_2 on obtient finalement

$$T = \left[\frac{1 - A_1 A_2}{(1 + A_1)(1 + A_2)} - W\right]^{-1} \qquad (10)$$

De l'expression obtenue pour $T(\lambda, \delta \,|\, w, a)$ on vérifie que

$$\left(\frac{\partial T}{\partial \lambda}\right)_{\lambda = 0,\, \delta = 1} = 0 \qquad (11)$$

et que $T(\lambda, \delta \,|\, w, a)$ a un maximum pour $\lambda = 0$. Des bornes inférieures et supérieures pour T sont obtenues facilement comme suit, des formules (7) et (8) et de la définition de P_j. Compte tenu du fait que la probabilité de ne pas s'arrêter avant l'échantillon d'ordre "j" est $(1-P_j)$, moindre que la probabilité que les moyennes de tous les échantillons tombent entre les limites d'action et plus grande que la probabilité que toutes les moyennes soient entre les limites de surveillance on a :

$$W^j \leqslant 1 - P_j \leqslant (A_1 + A_2 + W)^j \qquad (12)$$

ce qui donne, en utilisant la formule (1')

$$\frac{1}{1 - W} \leqslant T \leqslant \frac{1}{1 - (A_2 + A_2 + W)} \qquad (13)$$

inégalités dont la vérification est facile.

5 - LA CARTE DE CONTROLE A SEQUENCES

Conformément à la description donnée ces cartes de contrôle ont deux paramètres (c, R) tels que lorsqu'un échantillon tombe en dehors de l'intervalle $\left[\mu_o - c\frac{\sigma_o}{\sqrt{n}},\ \mu_o + c\frac{\sigma_o}{\sqrt{n}}\right]$ où lorsqu'il y a une séquence, avec une longueur $R > 1$, de moyennes d'échantillons au-dessous (ou au-dessus) de la moyenne générale on arrête le processus pour rechercher les causes possibles de variation. Pour calculer le T. M. A.', $T(\lambda, \delta \,|\, c, R)$ nous nous servirons a nouveau de l'expression :

$$P_j \ (\lambda, \delta \mid c, R) \tag{14}$$

définie comme au début du paragraphe 4.

La probabilité $1-P_j$ de continuer après l'échantillon d'ordre "j" est égale à

Prob. (tous les échantillons avec des moyennes dans l'intervalle

$$\left[\mu_o - c\frac{\sigma_o}{\sqrt{n}},\ \mu_o + c\frac{\sigma_o}{\sqrt{n}}\right]) \times \tag{15}$$

x Prob. (séquences < R| les échantillons ont leurs moyennes dans l'intervalle.

$$\left[\mu_o - c\frac{\sigma_o}{\sqrt{n}},\ \mu_o + c\frac{\sigma_o}{\sqrt{n}}\right])$$

Un calcul assez compliqué (voir annexe 1) donne l'expression du facteur D_j de la série (16). La valeur

$$T(\lambda, \delta \mid c, R) = 1 + \sum_1^{\infty}\left[\Phi\left(c - \sqrt{n}\,\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\,\frac{\lambda}{\delta}\right)\right]^j D_j\ (\lambda, \delta \mid c, R) \tag{16}$$

est telle que :

$$T(\lambda, \delta \mid c, R) \leqslant 1/ \quad 1 - \left[\Phi\left(c - \sqrt{n}\,\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\,\frac{\lambda}{\delta}\right)\right] \tag{17}$$

6 - L'ERREUR DANS T (λ, δ) RESULTANT DE L'USAGE D'ESTIMATIONS DE LA MOYENNE ET DE L'ECART TYPE.

Les valeurs de μ_o et σ_o n'étant pas connues on utilisera a leur place des estimations $\bar{\bar{x}}$ et σ' provenant d'un grand échantillon de $N > n$ observations pour le calcul du temps moyen d'action. L'erreur a craindre sera obtenue suivant une technique analogue a celle utilisés déjà par Wald et Wolfowitz (1946).

Soit donc T (λ, σ) le temps moyen d'action dans un des deux critères en étude, où l'on use les valeurs estimés $\bar{\bar{x}}$ et σ' à la place de μ_o et σ_o et supposons que les paramètres du processus ont changé pour devenir μ et σ. En rappelant la définition d'un T. M. A., on a donc, dans ce cas, la valeur

$$T\left(\frac{\mu - \bar{\bar{x}}}{\sigma'},\ \frac{\sigma}{\sigma'}\right)$$

Le calcul de la valeur moyenne de T (λ, δ) conduit (voir annexe 2) au résultat suivant

$$T\ (\lambda, \delta) = T\ (\lambda, \delta) + 0\ (1/N) \tag{18}$$

$0(1/N)$ désignant des termes d'ordre $1/N$, qui tendent vers zéro lorsque N augmente. Dans la pratique, on prend en général $N > 100$ (Shewhart (1931), Wilks (1941, 1942)) ; de cette façon T (λ, δ) peut être estimé avec une erreur de l'ordre de 0, 01 a partir de $\bar{\bar{x}}$ et σ', ce qui est négligeable pour les valeurs significatives du T. M. A..

7 - OBSERVATIONS FINALES

Les résultats exacts que l'on vient d'obtenir pour les deux cartes de contrôle dans les conditions de stabilité statistique peuvent être utilisés pour les différents objectifs ci-après :

1/ Faire un choix rationnel des paramètres qui déterminent le T. M. A., ce qui permet un usage plus flexible de ces cartes.

2/ Analyser l'efficacité des critères au moyen du T. M. A. d'une manière semblable à l'utilisation usuelle des erreurs de première et de deuxième espèces.

3/ Comparer des cartes de contrôle au moyen des courbes du T. M. A. ; en particulier on a maintenant une base rationnelle pour l'utilisation des séquences dans le contrôle statistique.

Dans un autre travail nous allons nous servir de ces résultats-ci et des tables partiellement calculées déjà de façon à donner une base économique pour le choix de cartes de contrôle et de leurs paramètres.

ANNEXE 1 (CARTE DE CONTROLE A SEQUENCES) :

La probabilité définie par la relation (15) peut s'écrire

$$1 - P_j = \left[\Phi\left(c - \sqrt{n}\,\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\,\frac{\lambda}{\delta}\right)\right]^j D_j\,(\lambda, \delta\,|c, R)$$

L'expression pour D_j est :

$$D = \sum_{\substack{0 \leqslant i_1 \leqslant j_1 \\ 0 \leqslant i_2 \leqslant j_2 \\ j_1 + j_2 = j}} \Delta_R(i_1, j_1)\,\Delta_R(i_2, j_2)\,G(i_1, i_2)\,p^{j_1}(1-p)^{j_2} \tag{1.0}$$

où :

$$p = \frac{\Phi\left(c - \sqrt{n}\,\frac{\lambda}{\delta}\right) - \Phi\left(- \sqrt{n}\,\frac{\lambda}{\delta}\right)}{\Phi\left(c - \sqrt{n}\,\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\,\frac{\lambda}{\delta}\right)}$$

et :

$$\Delta_R\,(i, j) = \sum \frac{i\,!}{i_1\,!\ldots\,i_{R-1}!}\,(i \leqslant j)$$

où la sommation est faite pour toutes les nombres tels que

$$i_1 + i_2 + \ldots + i_{R-1} = i$$
$$i_1 + 2i_2 + \ldots + (R - 1)\,i_{R-1} = j$$

et $G(i_1, i_2)$ est définie par

$$\begin{aligned} G(i_1, i_2) &= 0 \quad \text{si} \quad |i_1 - i_2| > 1 \\ &= 1 \quad \text{si} \quad |i_1 - i_2| = 1 \\ &= 2 \quad \text{si} \quad |i_1 - i_2| = 0 \end{aligned}$$

Pour obtenir ces résultats nous avons utilisé l'expression (11. 3. 3) de Fisz (1963) en prenant :

$$n_1 = j_1 \ ; \ n_2 = j_2 \ ; \ k_1 = i_1 \ ; \ k_2 = i_2$$
$$k_{1,R} = \ldots k_{1,n_1} = 0$$
$$k_{2,R} = \ldots k_{2,n_2} = 0$$

et remplaçant les autres k par les i correspondants.

L'expression de D_j peut être écrite :

$$D_j = 2 \sum_{\substack{0 < i < \min(j_1, j_2) \\ j_1 + j_2 = j}} \Delta_R(i, j_1)\ \Delta_R(i, j_2)\ p^{j_1}(1-p)^{j_2} +$$
$$+ \sum_{0 \leqslant i \leqslant \min(j_1, j_2 - 1)} \Delta_R(i, j_1)\ \Delta_R(i+1, j_2)\ p^{j_1}(1-p)^{j_2} + \sum_{0 \leqslant i \leqslant \min(j_1 - 1, j_2)} \Delta_R(i+1, j_2)\ \Delta_R(i, j_2)\ p^{j_1}(1-p)^{j_2}$$

En nous servant de ce développement, que nous venons de présenter, nous pouvons obtenir pour D_j la formule plus compacte qui suit :

$$D_j = \sum_{j_1 + j_2 = j} L(j_1, j_2)\ p^{j_1}(1-p)^{j_2}$$

avec

$$L(j_1, j_2) = 2 \sum_0^{\min(j_1, j_2)} \Delta_R(i, j_1)\ \Delta_R(i, j_2) + \sum_0^{\min(j_1, j_2 - 1)} \Delta_R(i, j_1)\ \Delta_R(i+1, j_2) + \sum_0^{\min(j_1 - 1, j_2)} \Delta_R(i+1, j_1)\ \Delta_R(i, j_2)$$

Lorsque $j_1 - 1 \leqslant 0$ ou $j_2 - 1 \leqslant 0$ le terme correspondant s'annulle, car la somme se fait pour un ensemble vide d'indices.

En partant de cette dernière formule, l'on remarque immédiatement la symétrie $L(p, q) = L(q, p)$. En conséquence il nous suffit de calculer $L(p, q)$ pour $p \leqslant q$. On voit immédiatement de l'expression des Δ_R que :

$$\Delta_R(0, 0) = 1 \quad ; \quad \Delta_R(0, j) = 0 \quad \text{pour} \quad j > 0$$

En utilisant tous ces résultats on obtient finalement :

$$L(p, p) = 2 \left(\sum_1^p \Delta_R^2(i, p) + \sum_1^{p-1} \Delta_R(i, p)\ \Delta_R(i+1, p) \right)$$

et pour $p < q$

$$L(p, q) = 2 \sum_1^p \Delta_R(i, p)\ \Delta_R(i, q) + \sum_1^p \Delta_R(i, p)\ \Delta_R(i+1, q) + \sum_1^{p-1} \Delta_R(i+1, p)\ \Delta_R(i, q)$$

Le calcul de $\Delta_R(i, j)$ peut se faire comme suit. Dans l'identité :

$$(x_1 + x_2 + \ldots + x_{R-1})^i = \sum \frac{i!}{i_1!\ i_2! \ldots i_{R-1}!}\ x_1^{i_1} x_2^{i_2} \ldots x_{R-1}^{i_{R-1}}$$

on peut prendre $x_1 = x, x_2 = x^2, \ldots, x_{R-1} = x^{R-1}$ ce qui donne

$$(x + x^2 + \ldots + x^{R-1})^i = \sum \frac{i!}{i_1!\ i_2! \ldots i_{R-1}!}\ x^{i_1 + 2i_2 + \ldots + (R-1)i_{R-1}} = \frac{(x - x_R)^i}{(1 - x)^i}$$

$\Delta_R(i, j)$ est le coefficient de x dans le développement de

$$\frac{(x - x^R)^i}{(1 - x)^i}$$

suivant les puissances de x.

L'expression pour le T. M. A. est évidement :

$$T(\lambda, \delta) = 1 + \sum_1^{\infty} \left[\Phi\left(c - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\frac{\lambda}{\delta}\right) \right]^j D_j \qquad (1.b)$$

On peut remarquer de l'expression (16) que T dépend de "n" et de "λ" par $\sqrt{n}\,\lambda$. Comme auparavant en tenant compte la symétrie de L, on a :

$$\left(\frac{\partial T}{\partial \lambda}\right)_{\lambda = 0, \delta = 1} = 0 \qquad (1.c)$$

et $T(\lambda, \delta \mid c, R)$ a une borne supérieure semblable à celle qu'on a obtenue dans la section 4. Comme $D_j \leqslant 1$ nous obtenons :

$$T(\lambda, \delta) < 1 + \sum_1^{\infty} \left[\Phi\left(c - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\frac{\lambda}{\delta}\right) \right]^j = \frac{1}{1 - \Phi\left(c - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\frac{\lambda}{\delta}\right)} \qquad (1.d)$$

On peut faire, encore, une remarque concernant le calcul de la série donnant T. Si on prend k termes de la série, le reste de la série est inférieur à :

$$\sum_{k+1}^{\infty} \left[\Phi\left(c - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\frac{\lambda}{\delta}\right) \right]^j \leqslant \sum_{k+1}^{\infty} [\Phi(c) - \Phi(-c)]^j = \frac{(2\,\Phi(c) - 1)^{k+1}}{(2\,(1 - \Phi(c))}$$

car

$$D_j \leqslant 1 \quad \text{et} \quad \Phi\left(c - \sqrt{n}\frac{\lambda}{\delta}\right) - \Phi\left(- c - \sqrt{n}\frac{\lambda}{\delta}\right) \leqslant \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

en conséquence de la symétrie de la fonction de distribution normale ; pour chaque valeur de c on peut calculer k de façon que la borne supérieure du reste soit inférieure a 1/2 ce qui donnera l'approximation de la série ; pour le calcul des tables on prendra $\delta = 1$.

ANNEXE 2 (ERREUR DANS T(λ, δ) RESULTANT DE L'USAGE D'ESTIMATIONS DE LA MOYENNE ET DE L'ECART-TYPE)

La valeur moyenne de $T(\lambda, \delta)$ est donnée par l'expression

$$\tilde{T}(\lambda, \delta) = \int_{-\infty}^{+\infty} \int_{0}^{-\infty} T\left(\frac{\mu - \bar{\bar{x}}}{\sigma'}, \frac{\sigma}{\sigma'}\right) \frac{\sqrt{N}}{\sigma_0^2} \Phi'\left(\sqrt{n}\frac{\bar{\bar{x}} - \mu_0}{\sigma_0^2}\right) h\left(\frac{\sigma'}{\sigma_0}\right) d\sigma'\, d\bar{\bar{x}} \qquad (2.a)$$

où le noyau de (2.a) est la densité de probabilité de $(\bar{\bar{x}}, \sigma)$.

En posant $\sqrt{N} \cdot \frac{\bar{\bar{x}} - \mu_0}{\sigma_0} = \xi$ et $\frac{\sigma'}{\sigma_0} = \eta$

(2.a) devient

$$\tilde{T}(\lambda, \delta) = \int_{-\infty}^{+\infty}\int_{0}^{+\infty} T\left(\frac{\lambda - \xi}{\eta}, \frac{\delta}{\eta}\right)\Phi'(\xi)\ h\ (\eta)\, d\eta\, d\xi \qquad (2.b)$$

on a de nouveau :

$$\left(\frac{\delta T}{\partial \lambda}\right)_{\lambda = 0, \delta = 1} = 0 \qquad (2.c)$$

Nous pouvons maintenant comparer $\tilde{T}(\lambda, \delta)$ avec $T(\lambda, \delta)$.

Avant tout on a :

$$T\left(\frac{\lambda - \xi}{\eta}, \frac{\delta}{\eta}\right) = T(\lambda, \delta) + A(\lambda,\delta)\ \xi + B(\lambda, \delta)\ (\eta - 1) +$$
$$+ \frac{1}{2}\ [\tilde{C}(\lambda, \delta)\ \xi^2 + 2\tilde{D}(\lambda, \delta)\ \xi(\eta - 1) + \tilde{E}(\lambda, \delta)\ (\eta - 1)^2]$$

où $\tilde{C}$, $\tilde{D}$ et $\tilde{E}$ sont les trois dérivées partielles de deuxième ordre de $T(\lambda, \delta)$ pour des points de $[0, \xi] \times [1, \eta]$ avec $\xi > 0$ et $\eta > 1$ et dans des régions analogues pour les autres cas et A et B sont les dérivées partielles de premier ordre au point $(\xi, \eta) = (0, 1)$. D'autre part, on voit que les dérivées en question existent pour les deux critères.

En substituant (2.d) dans (2;b) la valeur moyenne $\tilde{T}(\lambda, \delta)$ devient (M notant la valeur moyenne)

$$\tilde{T}(\lambda, \delta) = T(\lambda, \delta) + B\ M\ [\eta - 1] + \qquad (2.e)$$
$$+ \frac{1}{2} \int_{-\infty}^{+\infty}\int_{0}^{+\infty} (|\tilde{C}|\ \xi^2 + 2\ |\tilde{D}|\ |\xi(\eta - 1)| + |\tilde{E}|\ (\eta - 1)^2)\ \Phi'(\xi)\ h_N(\eta)\, d\eta\, d\xi$$

En conséquence on a

$$|T(\lambda, \delta) - \tilde{T}(\lambda, \delta) - BM\ [\eta - 1] \leqslant \qquad (2.f)$$
$$\leqslant \frac{1}{2} \int_{-\infty}^{+\infty}\int_{0}^{+\infty} (|\tilde{C}|\ \xi^2 + 2\ |\tilde{D}|\ |\xi|\ (\eta - 1)| + |\tilde{E}|\ (\eta - 1)^2)\ \Phi'(\xi)\ h_N(\eta)\, d\eta\, d\xi$$

Les termes sous le signe d'intégration étant $0(1/N)$ aussi bien que la valeur moyenne $M\ [\eta - 1] = 0(1/N)$, on a :

$$\tilde{T}(\lambda, \delta) = T(\lambda, \delta) + 0(1/N) \qquad (2.g)$$

REFERENCES

ARROIAN, Leo A., LEVENE, H. (1950) - The effectiveness of quality control charts, *Journal of the American Statistical Association,* XLV

FISZ, M. (1963) - *Probability Theory and Mathematical Statistics,* J. Wiley and Sons, New York.

GUMBEL, E.J. (1958) - *Statistics of Extremes.* Columbia University Press, New York

KEMP, K. W. (1961) - The Average Run Length of Cumulative sum chart when a V - mask is used, *J.R. Statist. Soc.* B. 23

PAGE, E. S. (1962) - A modified control chart with warning limits, *Biom.*, 49

SHEWHART, W. A. (1931) - *Economic control of Manufactured Product*, D. Van Nostrand, New York

(1939) - *Statistical Method from the Viewpoint of Quality Control*, Grad, School of Agriculture, Washington

WALD, A. and WOLFOWITZ, J. (1946) - Tolerance Limits for a normal distribution, *Ann. Math. Statist.*, 17

WESTERN ELECTRIC (1958) - *Statistical Quality Control Hankbook* Mack Printing Company, Easton, Pennsylvania

WILKS, S. S. (1941) - On the determination of sample sizes for setting tolerance limits, *Ann. Math. Statist.*, 12

(1942) - Statistical prediction with special reference tothe problem of tolerance limits. *Ann. Math. Statist.* 13

WOLFOWITZ, J. (1943) - On the theory of runs with some applications to quality control, *Ann. Math. Statist.*, 14