



HAL
open science

SPC Optimisation de la taille des échantillons dans le cas des systèmes multigénérateurs

Maurice Pillet, Emmanuel Duclos

► **To cite this version:**

Maurice Pillet, Emmanuel Duclos. SPC Optimisation de la taille des échantillons dans le cas des systèmes multigénérateurs. Revue pratique de controle industriel, 1997. hal-02314028

HAL Id: hal-02314028

<https://hal.science/hal-02314028v1>

Submitted on 12 Oct 2019

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.

SPC

Optimisation de la taille des échantillons dans le cas des systèmes multigénérateurs

(Cas des presses à injecter, des multibroches, des multipostes...)

Les systèmes multigénérateurs

Dans les entreprises manufacturières, on rencontre souvent des moyens de production qui combinent en fait plusieurs "sous-machines" élémentaires. Par souci de productivité, on cherche à fabriquer plusieurs produits simultanément par la même machine. C'est le cas notamment des presses à injecter pour lesquelles chaque empreinte permet de produire une pièce par cycle d'injection.

Nous appellerons "système multigénérateur" un système dont la production est le résultat de plusieurs productions élémentaires. La production totale est alors la somme de toutes les ré-

partitions élémentaires comme le montre la figure 1.

Nous ferons l'hypothèse dans cet article que les répartitions élémentaires suivent une loi normale. Cependant, comme le montre la figure 1, même si les répartitions élémentaires suivent une loi normale, il n'en est pas de même pour la courbe résultante traduisant la population livrée.

Le suivi SPC de ce type de production

Le but du suivi SPC est de vérifier si le processus de fabrication est sous-contrôle ou hors-contrôle. On cherche par ce suivi à savoir si le processus est toujours centré sur la même moyenne et si sa dispersion reste conforme à la dispersion habituellement constatée sur ce moyen de production.

Le suivi est réalisé la plupart du temps par carte de Shewhart. Pour cela, on prélève régulièrement un échantillon **de taille n** dans la production et on estime les deux paramètres de la courbe de Gauss par la moyenne et l'étendue de l'échantillon.

L'objectif est de déterminer avec la plus petite incertitude possible, et avec le coût de prélèvement le plus faible possible où se trouve le processus de fabrication au moment du prélèvement.

Dans le cas des répartitions unimodales, et dont la répartition suit une loi normale, on montre que la carte de Shewhart donne de bons résultats avec des échantillons de petite taille pour peu que la capacité machine soit bonne. Mais qu'en est-il dans le cas des systèmes multigénérateurs ? Plusieurs questions se posent :

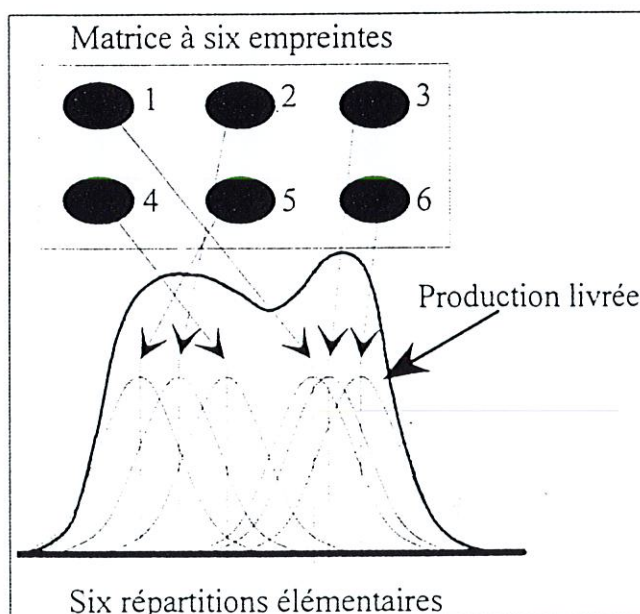


Figure 1 - Système multigénérateur

- Quelle est la taille optimale du prélèvement ?
- Faut-il systématiquement prélever une pièce par générateur ?
- Est-il possible de travailler avec la carte de Shewhart et ses limites traditionnelles ?
- Ne peut-on pas améliorer l'efficacité du contrôle tout en diminuant la taille du prélèvement ?

C'est à toutes ces questions que nous tenterons de répondre dans cet article.

Les différentes techniques de prélèvement

Pour prélever un échantillon d'une production de ce type, plusieurs méthodes d'échantillonnage sont généralement retenues par les industriels. Les deux méthodes les plus couramment retenues sont les suivantes :

- Prélèvement aléatoire de n pièces dans la production totale. Dans ce cas de figure, n est généralement inférieur au nombre de générateur.
- Prélèvement systématique d'une pièce par générateur.

D'autres méthodes se retrouvent de façon moins classique :

- Prélèvement de n pièces provenant toujours des mêmes générateurs. On prélève par exemple systématiquement les générateurs extrêmes (5 et 6 dans le cas de la figure 1) plus une pièce provenant d'un générateur médian.
- Prélèvement d'un échantillon dans un ordre séquentiel. Le premier générateur est tiré au hasard, mais les autres sont systématiquement les générateurs suivants. On aura alors comme type de prélèvement [2, 3, 4] - [5, 6, 1] - [6, 1, 2] - [4, 5, 6] etc. On remarque que le prélèvement systématique de chaque générateur est un cas particulier de ce type de prélèvement.

On a donc trois méthodes d'échantillonnage distinctes :

- prélèvement aléatoire ;
- prélèvement séquentiel ;
- prélèvement systématique d'un sous-ensemble.

Chacune de ces méthodes donne des résultats fort différents en terme d'efficacité du contrôle. Certaines sont excellentes, d'autres sont médiocres. Aussi chaque responsable d'une fabrication de ce type doit connaître les résultats généraux que nous développerons par la suite pour optimiser le coût et l'efficacité de son contrôle.

Étude sur un exemple

Pour illustrer la différence d'efficacité des méthodes d'échantillonnage, nous avons, par simulation, testé les différentes méthodes de prélèvement afin de les comparer. Les démonstrations théoriques sont portées dans

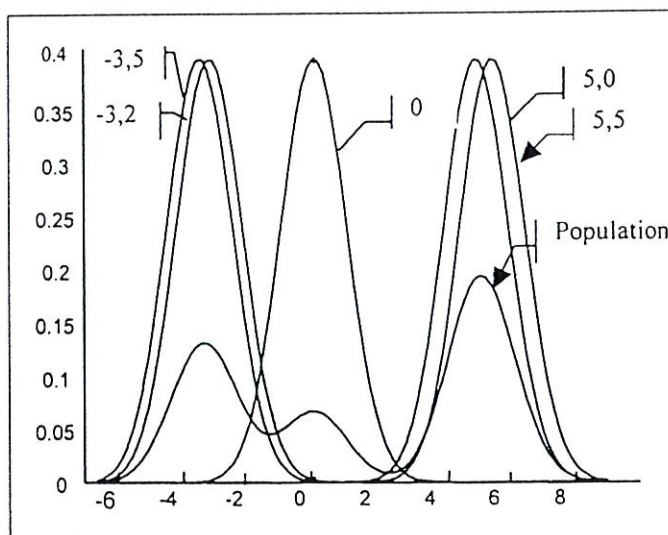


Figure 2 - Répartition des six générateurs

l'encadré 1 afin de ne pas alourdir inutilement l'exposé. Le lecteur intéressé par ces développements pourra utilement s'y reporter.

Nous nous sommes intéressés à une production provenant de six générateurs répartis comme l'indique la figure 2.

Les caractéristiques de chaque générateur sont données par le tableau suivant :

Générateurs	Moyennes	Ecart-type
1	0,0	1
2	5,5	1
3	-3,2	1
4	5,0	1
5	5,0	1
6	-3,5	1

La population globale se caractérise par une moyenne de 1,47 et par un écart-type de $\sigma_{\text{popul}} = 4$.

Nous nous intéressons dans cet article uniquement au cas de la moyenne. Pour chaque méthode d'échantillonnage, nous avons prélevé plusieurs milliers d'échantillons en calculant à chaque échantillon, trois estimateurs : la moyenne \bar{X} et l'écart-type de variation de cette moyenne.

En calculant la moyenne, on cherche à estimer la valeur de la moyenne de la population totale. Lorsqu'on fait de l'estimation deux erreurs peuvent se produire (figure 3) :

- un biais sur l'estimation ;
- une incertitude dans l'estimation.

Plus la méthode d'échantillonnage est efficace, plus le biais est faible, et plus la dispersion observée sur l'estimateur est faible. En effet, dans ce cas, l'incertitude sur la position moyenne de la production est la plus limitée.

Examinons le comportement des différentes méthodes de prélèvement.

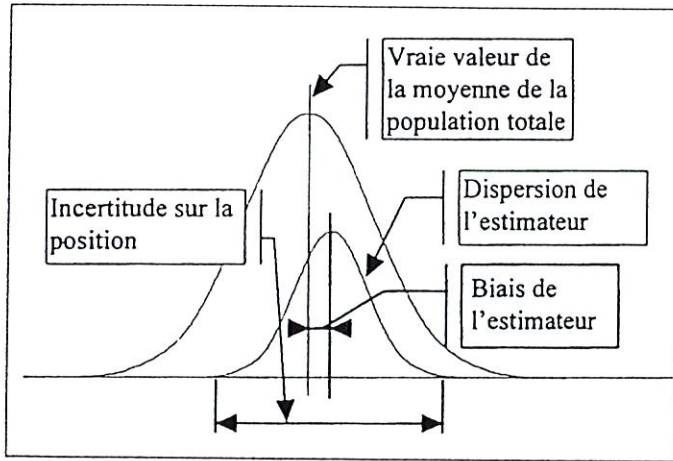


Figure 3 - Efficacité de l'échantillonnage

Prélèvement aléatoire

Ce prélèvement a l'avantage de la simplicité, on prélève au hasard n pièces dans la production qui vient d'être réalisée. Cependant, cette simplicité sera payée par une piètre efficacité du contrôle. On peut résumer les résultats obtenus par le tableau et le graphique de la figure 4.

n	Moyenne		
	Moyenne	Variance	écart-type
3	1,47	5,32	2,31
4	1,47	3,99	2,00
5	1,47	3,19	1,79
6	1,47	2,66	1,63
7	1,47	2,28	1,51
8	1,47	2,00	1,41

Ce résultat évident est conforme à la théorie qui veut que la moyenne d'un échantillon tiré d'une population identique est répartie suivant une loi de Gauss ayant pour moyenne, la moyenne de la population et comme écart-type $\sigma_{\text{popul}}/\sqrt{n}$.

Par exemple, pour n = 4, on retrouve bien :

$$\sigma_{\bar{x}} = \sigma_{\text{popul}}/\sqrt{n} = 4/\sqrt{4} = 2$$

La moyenne de l'estimation est toujours centrée sur la moyenne de la population totale, l'estimation n'est donc pas biaisée.

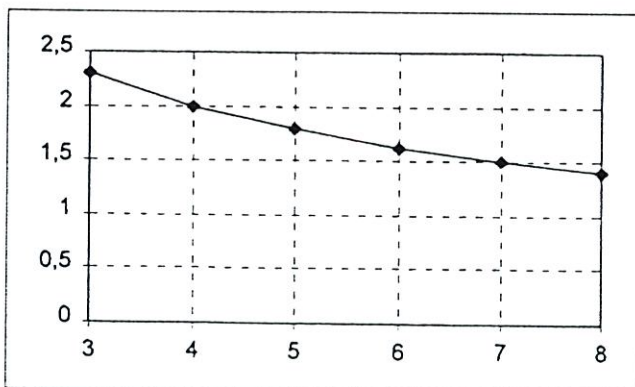


Fig 4 - Cas aléatoire - Écart-type sur la moyenne

La figure 5 donne quelques formes de répartition des moyennes. On remarque que pour n = 3, la distribution des moyennes ne suit pas une loi normale, mais lorsque n augmente (n = 5 ou n = 6) la loi converge vers une population normale.

On peut conclure que dans ce cas, la seule solution pour améliorer l'efficacité du prélèvement est d'augmenter la taille de l'échantillonnage, mais cela coûte cher, et il y a sans doute mieux à faire !

Prélèvement séquentiel

Rappelons que pour cette méthode, le premier générateur est tiré au hasard, les suivants sont toujours les successeurs du premier générateur prélevé.

La méthode séquentielle permet de prélever moins de pièces que le nombre de générateurs, tout en balayant tous les générateurs au terme de quelques échantillonnages.

On peut résumer les résultats obtenus par le tableau et le graphique de la figure 6. Là encore, l'estimation n'est évidemment pas biaisée, la moyenne de l'estimateur est centrée sur 1,47.

L'écart-type sur les moyennes dans le cas séquentiel est très inférieur au cas aléatoire. Cela montre bien l'intérêt de la solution séquentielle. Celle-ci donne d'excellents résultats en terme de suivi de procédé par rapport à la solution aléatoire pour un coût très inférieur.

n	Moyenne		
	Moyenne	Variance	écart-type
3	1,47	1,02	1,01
4	1,47	1,31	1,14
5	1,47	0,80	0,89
6	1,47	0,17	0,41
7	1,47	0,45	0,67
8	1,47	0,39	0,62

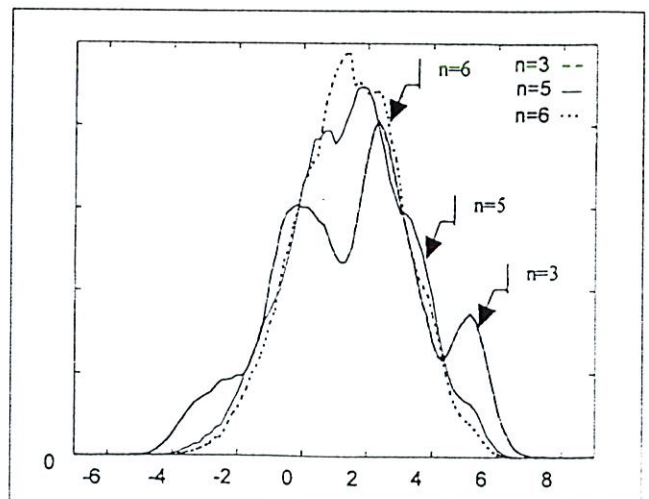


Fig. 5 - Cas aléatoire - Répartition des moyennes

Le graphique de la figure 6 est révélateur. Il nous montre trois résultats fondamentaux qu'il est essentiel de mémoriser.

Dans le cas de figure étudié, il est plus efficace de prélever trois pièces en échantillonnage séquentiel que huit pièces en échantillonnage aléatoire.

L'optimum est obtenu pour un prélèvement dont la taille est identique au nombre de générateurs.

Il est inutile de prélever plus de pièces que le nombre de générateurs à moins de prélever un multiple de celui-ci.

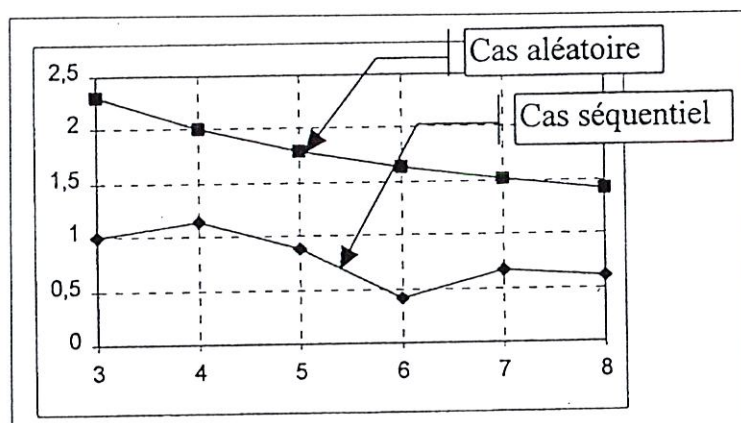


Fig 6 - Cas séquentiel - Écart-type sur la moyenne

est "sous contrôle". Si ce n'est pas le cas, la dérive se verra toujours sur les n générateurs choisis.

Il reste le cas où il y a un incident sur un seul générateur. Il peut être prévenu en prélevant régulièrement, mais avec une fréquence plus faible que le SPC tous les générateurs pour une vérification de la dispersion.

Pour étudier ce type de prélèvement, nous avons simulé plusieurs configurations pour $n = 2, 3, 4$ et 6 .

- Pour $n = 2$, on prélève les empreintes 2 et 6 centrées sur 5,5 et -3,5.
- Pour $n = 3$, on prélève les empreintes 1, 3, 4 centrées sur -3,2, 0 et 5.
- Pour $n = 4$, on prélève les empreintes 2, 3, 4 et 6 centrées sur 5 ; 5,5 ; -3,2 ; -3,5.
- Pour $n = 6$, on prélève les 6 empreintes.

D'autres choix auraient pu être faits. Cependant, pour limiter le biais de l'estimation, il est préférable de choisir les générateurs situés aux extrêmes.

On peut résumer les résultats obtenus par le tableau et le graphique de la figure 8.

Les résultats sur les variances sont proches des résultats théoriques $V = \sigma_{\text{gé né}}^2/n$.

Comme on le constate sur la figure 8, la répartition des moyennes est évidemment normale (on revient au cas séquentiel avec $n = 6$).

La figure 7 donne quelques formes de répartition des moyennes.

On remarque immédiatement sur la répartition des moyennes que pour $n = 6$ (égal au nombre de générateurs) la répartition suit une loi normale.

Pour les cas où n est différent du nombre de générateurs, la dispersion peut s'éloigner de manière très significative d'une loi normale dans l'exemple que nous avons choisi.

On montre facilement (voir encadré) que pour $n = 6$ la variance pour les moyennes est égale à $\sigma_{\text{gé né}}^2/n$ lorsque tous les générateurs ont le même écart-type $\sigma_{\text{gé né}}$.

Dans notre cas $\sigma_{\text{gé né}}^2/n = 1/6 = 0,17$.

La conclusion un peu hâtive que l'on pourrait tirer à ce stade est que l'échantillonnage le plus efficace serait le prélèvement systématique de tous les générateurs. Cela reste vrai dans le cas où le nombre de générateur est faible < 10 . Lorsque le nombre de générateurs est important comme c'est souvent le cas sur les carrousels ou sur les presses, il est impossible de prélever systématiquement une valeur par générateur.

Prélèvement systématique d'un sous-ensemble

On cherche par cette méthode à utiliser le résultat précédent pour $n = 6$ mais en diminuant le nombre de générateurs. Pour cela, la méthode consiste à retenir pour l'échantillonnage toujours les mêmes générateurs. L'inconvénient de cette méthode est de ne jamais voir le résultat des générateurs non concernés par l'échantillonnage. Le fait de ne sélectionner que certains générateurs induira un biais dans l'estimation de la position de la production totale. En effet, la moyenne des prélèvements sera forcément centrée sur la position moyenne des générateurs choisis. En fait, ce biais ne pose pas de réel problème pour l'application de la MSP dont le but est de vérifier si le processus

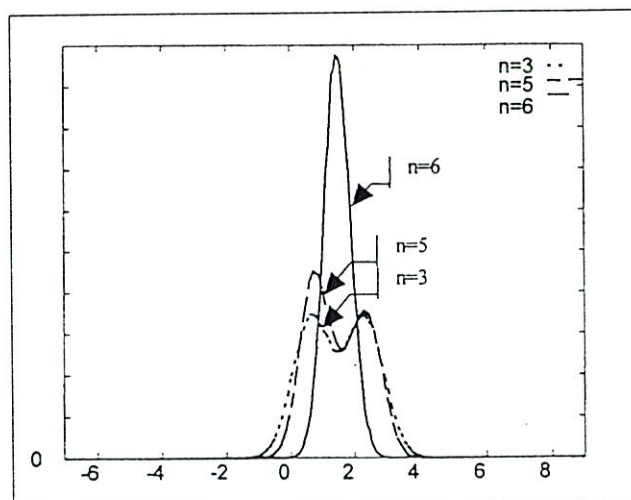


Fig 7 - Cas séquentiel - Répartition des moyennes

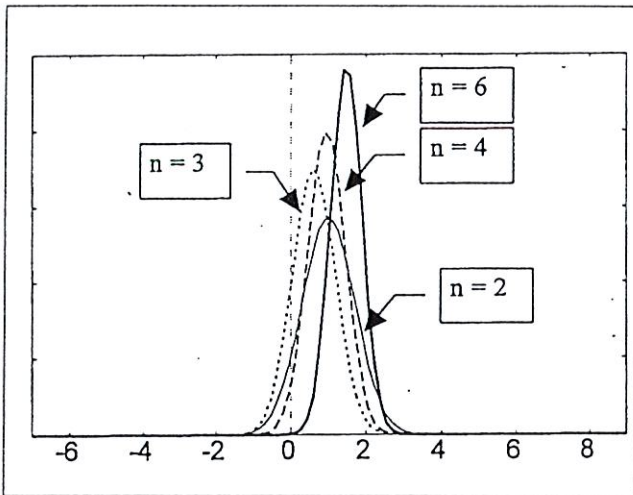


Fig. 8 - Cas systématique - Répartition des moyennes

La variance de la répartition est minimale par rapport au cas aléatoire ou séquentiel pour une taille d'échantillon donnée.

On remarque cependant que la moyenne de l'échantillon systématique n'est pas centrée sur la moyenne de la production mais sur la moyenne des positions des générateurs prélevés. L'estimation de la position de la population est donc biaisée. Cela n'est pas gênant dans le cas d'un suivi par carte de contrôle car il suffit alors de centrer la carte sur la moyenne historique des générateurs choisis pour l'échantillonnage.

Moyenne			
n	Moyenne	Variance	écart-type
2	1,0	0,49	0,70
3	0,61	0,32	0,57
4	0,95	0,25	0,50
6	1,47	0,17	0,41

La figure 9, nous indique qu'il est préférable - dans l'exemple choisis pour cet article - de prendre un échantillonnage systématique avec n = 2 qu'un échantillonnage aléatoire avec n = 8 pour minimiser l'écart-type sur la moyenne. Or d'un point de vue coût du contrôle, il est clair que les deux solutions ne sont pas compa-

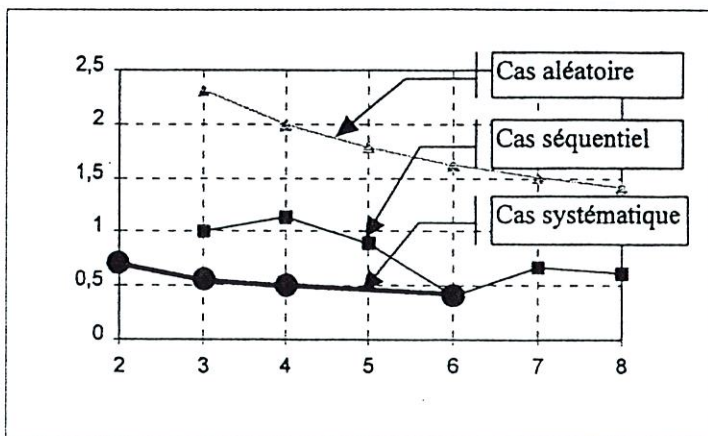


Figure 9 - Comparaison des différentes solutions

rables. C'est donc le cas le plus favorable pour surveiller la position d'un processus par carte de contrôle.

Enfin, pour surveiller si un des générateurs ne dérive pas seul, alors que les autres restent sous contrôle, il est possible de réaliser un prélèvement double :

- 2 ou 4 générateurs pour suivre la position du procédé en notant sur une carte la moyenne ;
- tous les générateurs pour suivre la dispersion du procédé en notant l'étendue relevée sur l'ensemble des générateurs. la fréquence d'échantillonnage pouvant être plus faible que pour la position.

3. Choix des limites de contrôle dans le cas d'un système multigénérateurs

Echantillonnage aléatoire

Si, malgré les remarques que nous venons de faire quant à l'efficacité de cette méthode, certains persistent à utiliser l'échantillonnage aléatoire ils auront au moins la consolation de la simplicité des calculs.

En effet, en procédant à un échantillonnage aléatoire, on se trouve dans le cas traditionnel de Shewhart, et les limites de contrôle sont calculées de façon classique par les célèbres formules :

$$LIC_{\bar{x}} = \text{Cible} - 3 \cdot \frac{\sigma_{\text{popul}}}{\sqrt{n}} \quad (1)$$

$$LSC_{\bar{x}} = \text{Cible} + 3 \cdot \frac{\sigma_{\text{popul}}}{\sqrt{n}}$$

On remarquera cependant que la répartition des moyennes peut être multimodale (figure 5) pour n petit. Le risque alpha de fausse alarme seront donc différents de 0,27% mais restent du même ordre de grandeur.

Echantillonnage séquentiel

On a vu précédemment que dans ce cas, la variance sur la moyenne des échantillons, n'était pas distribuée selon la loi écart-type σ_i/\sqrt{n} . Il ne faut donc pas, théoriquement utiliser les équations (1).

En pratique, on peut toujours utiliser les limites classiques, cela diminue considérablement le risque alpha de fausse alarme. En revanche, on détecte plus tardivement des dérives.

Pour calculer les limites de contrôle, il faut connaître l'écart-type de la dispersion des moyennes. Or comme le montrent les calculs de l'encadré 1, on a quelques difficultés à le calculer car il dépend de la moyenne et de la dispersion de chaque générateur.

Le plus simple consiste à observer le procédé sur une trentaine d'échantillons, et à calculer à

partir des observations l'écart-type de la distribution des moyennes $\bar{\sigma}_x$.

Les limites de contrôle pour les moyennes se calculent alors à partir des formules

$$\begin{aligned} \text{LIC}_{\bar{x}} &= \text{Cible} - 3\bar{\sigma}_x \\ \text{LSC}_{\bar{x}} &= \text{Cible} + 3\bar{\sigma}_x \end{aligned} \quad (2)$$

On a vu (figure 7) que les moyennes pouvaient être réparties suivant une loi multimodale. Le risque alpha sera donc dans ce cas différent de 0,27% trouvé dans le cas d'une loi normale. Mais cela n'empêche pas de placer les limites à $\pm 3\bar{\sigma}_x$. Le risque alpha restera dans le même ordre de grandeur même en cas de multimodalité.

Echantillonnage systématique d'un sous-ensemble

Ce cas est le plus délicat à traiter. Considérons l'exemple où nous avons prélevé systématiquement les deux générateurs extrêmes de la population décrite au paragraphe 2.

Dans ce cas, la moyenne des trois générateurs ne serait pas égale à la moyenne de la population (1,47), mais à 1. Pour calculer les limites, il ne faut donc pas viser la cible, mais la moyenne des moyennes () de la carte d'observation.

Pour l'estimation de l'écart-type, on procédera de manière identique au cas séquentiel. Les limites deviennent :

$$\begin{aligned} \text{LIC}_{\bar{x}} &= \bar{\bar{X}} - 3\bar{\sigma}_x \\ \text{LSC}_{\bar{x}} &= \bar{\bar{X}} + 3\bar{\sigma}_x \end{aligned} \quad (3)$$

Dans le cas systématique, il n'y a pas de problème de normalité sur la répartition des moyennes comme le montre la figure 8

Conclusions

Nous avons montré dans cet article, l'intérêt d'étudier de façon approfondie le choix de l'échantillonnage dans le cas des procédés multigénérateurs. Les résultats avancés dans cet article nous permettent de hiérarchiser dans un ordre décroissant les méthodes d'échantillonnage en terme de coût de prélèvement et de performance :

1. prélèvement systématique de tous les générateurs ;
2. prélèvement systématique d'une sélection de n générateurs ;
3. prélèvement séquentiel de n générateurs avec choix aléatoire du premier ;
4. prélèvement aléatoire dans la production.

Le premier devant être retenu dans les cas où le nombre de générateurs est faible, le second et le troisième dans les autres cas. Le dernier est à proscrire pour faire une production de qualité à moindre coût.

Variances des moyennes dans le cas séquentiel

Prenons l'exemple des séquences de 4 générateurs. Les six séquences possibles sont : 1/2/3/4 - 2/3/4/5 - 3/4/5/6 - 4/5/6/1 - 5/6/1/2 - 6/1/2/3. Ces six séquences sont réparties en proportions égales car le premier générateur est choisi aléatoirement. Il est facile de calculer la moyenne de chaque séquence ($E(\text{seq}_i)$)

On rappelle les relations suivantes

$$V(X) = E(X^2) - E(X)^2$$

$$E(X^s) = \sum \pi_i m_i^s$$

On calcule la variance des moyennes

$$V(E(\text{seq})) = E(E(\text{seq})^2) - E(E(\text{seq}))^2$$

$$E(E(\text{seq})) = \sum \pi_i E(\text{seq}_i) = \frac{1}{6} \sum E(\text{seq}_i)$$

$$E(E(\text{seq})^2) = \sum \pi_i m_i^2 = \frac{1}{6} \sum (\sigma_{\text{seq}_i}^2 + E(\text{seq}_i)^2)$$

avec

$$\sigma_{\text{seq}_i}^2 = V\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2} V(\sum x_i) = \frac{1}{n^2} \sum V(x_i) = \frac{\sigma^2}{n}$$

car dans notre cas, chaque générateur a le même écart-type d'où :

$$V(E(\text{seq})) = \frac{1}{6} \sum \sigma_{\text{seq}_i}^2 + \frac{1}{6} \sum E(\text{seq}_i)^2 - \left[\frac{1}{6} \sum E(\text{seq}_i) \right]^2$$

Variance intra-séquence

Variance inter-séquence, nulle si on prend tous les générateurs (une seule séquence)

Application numérique pour n = 4

$$\begin{aligned} E(\text{Seq}_1) &= 1,83 & E(\text{Seq}_2) &= 3,08 & E(\text{Seq}_3) &= 0,83 \\ E(\text{Seq}_4) &= 1,63 & E(\text{Seq}_5) &= 1,75 & E(\text{Seq}_6) &= -0,30 \end{aligned}$$

$$V(E(\text{seq})) = 0,25 + \frac{19,33}{6} - (1,47)^2 = 1,31$$

Bibliographie

1. Emmanuel Duclos - Maurice Pillet - Alain Courtois - *Optimisation de la maîtrise statistique des procédés par une méthode de filtrage d'ordre* - Revue de statistique appliquée - A paraître Juin 1996
2. Maurice Pillet - *Appliquer la Maîtrise Statistique des Procédés MSP/SPC* - Ouvrage 340 p - Ed organisation - 1994
3. Maurice Pillet - *Les mesures de capabilité* - Revue Contrôle industriel & Qualité n°171- 58:64 - Nov 1991
4. Philippe Tassi - *Méthodes statistiques* - ouvrage 440 p Economica
5. M. Pillet - S. Rochon - E. Duclos - *Case of unilateral tolerances. SPC- Generalization of Capability index Cpm*. Quality engineering - Avril 96. □

Maurice Pillet
(Université de Savoie (74) - Laboratoire LLP)
Emmanuel Duclos
(Université de Savoie (74) - Laboratoire LLP)