REVUE DE STATISTIQUE APPLIQUÉE

H. C. HAMAKER
R. VAN STRIK

Efficience de l'échantillonnage double par attributs

Revue de statistique appliquée, tome 5, nº 1 (1957), p. 7-21
<http://www.numdam.org/item?id=RSA_1957__5_1_7_0>

© Société française de statistique, 1957, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

EFFICIENCE DE L'ÉCHANTILLONNAGE DOUBLE PAR ATTRIBUTS [1]

par

H. C. HAMAKER et R. VAN STRIK

Ingénieurs au Service des Recherches de la Société Philips

De nombreux utilisateurs des plans doubles n'ont pas été sans remarquer l'apparente contradiction qui peut, dans certains cas, exister entre les résultats fournis par le premier échantillonnage et ceux qu'aurait donné le plan complet s'il avait été poursuivi, même dans le cas où ce premier échantillon conduit à une décision d'acceptation ou de refus.

Cette contradiction éventuelle résulte, pour partie de fluctuations aléatoires contre l'effet desquelles le jugement sur échantillon reste nécessairement impuissant.

Peut-être cependant est-il possible d'apporter aux plans doubles quelques améliorations susceptibles de limiter l'effet d'un tel inconvénient : c'est le but de l'étude de MM. Hamaker et Van Strick qui se proposent de donner ultérieurement des tables tenant compte des résultats obtenus.

I. - INTRODUCTION

Sur un système d'axes rectangulaires, portons en abscisses le nombre total n d'éléments contrôlés et en ordonnées celui c des défectueux observés; un plan d'échantillonnage double peut être représenté par deux écrans S_1 et S_2. Lorsqu'après le contrôle d'un premier échantillon de n_1 éléments, le point (n_1, c_1) appartient au segment A_1 de S_1, le lot est accepté; s'il vient sur R_1 le lot est rejeté; mais s'il est dans l'ouverture D_1, un deuxième échantillon de taille n_2 est prélevé et le lot est accepté ou rejeté suivant que le point $(n_1 + n_2, c_2)$ se trouve sur les parties A_2 ou R_2 de S_2. Pratiquement un plan d'échantillonnage double est fonction de cinq paramètres : deux pour la taille des échantillons n_1 et n_2 et trois pour les "critères de décision"; si l'on veut remplacer un plan d'échantillonnage simple, fonction seulement de deux paramètres, par un plan d'échantillonnage double, en imposant aux deux plans d'avoir à peu près la même courbe d'efficacité, le problème admet plusieurs solutions parmi lesquelles il s'agit de choisir la meilleure. Ainsi, le "Statistical Research Group" (6) a fourni des tables d'où découlent les "Military Standard 105 A" (4). Bien que de nombreux utilisateurs en soient satisfaits, ces plans d'échantillonnage double ne sont pas les meilleurs dans le sens suivant : Aux fluctuations d'échantillonnage près, durant le contrôle d'éléments provenant d'un lot homogène, le point (n, c) doit se déplacer suivant une droite; et si l'on fait passer une droite par l'origine et par le point P_2; qui sépare en deux

(1) L'étude originale a été publiée dans le Journal of the American Statistical Association. Traduction dûe à Melle Gervaise, publiée avec l'autorisation de l'auteur et des éditeurs.

parties l'écran S_2, il est souhaitable que cette droite coupe S_1 au voisinage du centre de la fenêtre D_1. Envisageons, par exemple, le cas de la fig. 2, le plan correspondant conduit à accepter d'après le premier échantillon des lots qui auraient eu de grandes chances d'être rejetés si l'on avait porté le jugement sur les deux échantillons $n_1 + n_2$ au lieu de n_1 seul; de même, des lots considérés comme douteux en raison du premier échantillon seront très vraisemblablement rejetés après contrôle du deuxième. Il est intuitif qu'un plan comme celui de la fig. 2 ne peut donner toute satisfaction.

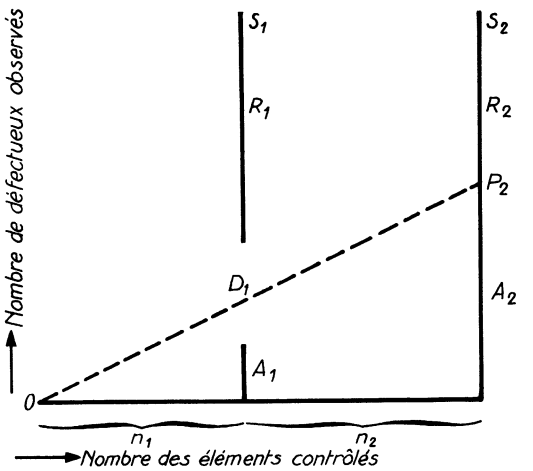

Fig. 1 - Diagramme pour un plan d'échantillonnage double.

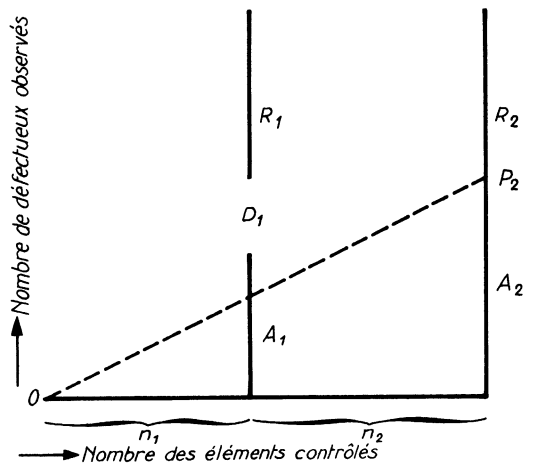

Fig. 2 - Diagramme pour un plan d'échantillonnage double intuitivement non satisfaisant.

Généralement, nous devons pour cette raison imposer à un plan d'échantillonnage double les conditions suivantes :

$$c_1 + \frac{1}{2} < \frac{n_1}{n_1 + n_2} (c_3 + \frac{1}{2}) \qquad (1\ a)$$

$$c_2 + \frac{1}{2} > \frac{n_1}{n_1 + n_2} (c_3 + \frac{1}{2}) \qquad (1\ b)$$

où n_1 et n_2 sont les tailles des échantillons, c_1, c_2, c_3 les trois paramètres de décision, à savoir c_1 (ou c_3) est le nombre maximum des défectueux que l'on peut trouver dans l'échantillon n_1 (ou $n_1 + n_2$) pour accepter le lot; un deuxième échantillon n_2 est contrôlé si dans le premier n_1 on a observé plus de c_1 défectueux et au plus c_2.

Dans le Mil. Standard. 105 A, $c_2 = c_3$ et (l_b) est toujours satisfait. Comme dans ces tables $n_2 = 2\ n_1$, la condition (l_a) devient :

$$c_1 + \frac{1}{2} < \frac{1}{3}\left(c_3 + \frac{1}{2}\right) \qquad (2)$$

qui n'est pas vérifiée par un grand nombre des plans des Mil. Std. 105 A. Ceci conduit à essayer de trouver d'autres plans d'échantillonnage double ayant à peu près la même courbe d'efficacité, mais ayant la propriété ci-dessus. Le but initial de cet article était de déterminer la valeur pratique de cette condition; mais au cours du travail, d'autres aspects de l'échantillonnage double ont été abordés, en particulier la possibilité de tronquer le deuxième échantillon et le choix des paramètres à prendre en considération pour décrire complètement le problème. Dans les § II, III et IV est développée une théorie relative à l'échantillonnage double, déjà décrite dans des articles précédents [2], [3] (1) et qui servira de base pour le reste de notre étude.

II. - PLANS D'ÉCHANTILLONNAGE « ÉQUIVALENTS »

Soit

$$D\left(\frac{n_2}{n_1}\ ;\ c_1,\ c_2,\ c_3\right) \qquad (3)$$

un plan d'échantillonnage double où $\frac{n_2}{n_1}$ est le rapport des tailles des échantillons et c_1, c_2, c_3 sont les "critères de décision". Par exemple :

$$D\ (2\ ;\ 2,\ 6,\ 10) \qquad (4)$$

désigne tous les plans d'échantillonnage double où le deuxième échantillon est deux fois plus grand que le premier; qui accepte un lot si dans un premier échantillon il y a au plus deux défectueux, le rejette s'il y en a au moins 6, demande un deuxième prélèvement pour plus de 2 et moins de 6 et accepte ou rejette suivant que dans $n_1 + n_2$ il y a au plus 10 défectueux ou plus de 10 défectueux.

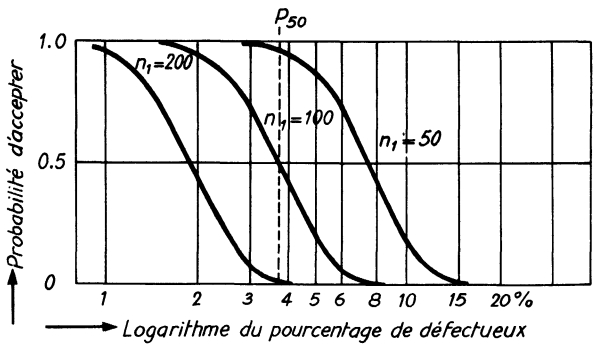

Fig. 3 - Courbe d'efficacité tracée en fonction de log p. Un changement de la taille de l'échantillon produit un déplacement latéral, mais aucune autre modification.

Dans la mesure où les conditions de Poisson sont remplies (ce qui est le cas de la plupart des applications industrielles), la probabilité d'accepter un lot avec un plan tel que (3), est une fonction de n_1 p, où p est la proportion de défectueux contenus dans le lot. La manière la plus commode pour mettre en évidence ce fait consiste à tracer la courbe d'efficacité (O.C) correspondante en portant en abscisse log p comme dans la fig. 3. Alors une variation de n_1 fait glisser la

(1) Les numéros entre crochets renvoient à la bibliographie en fin d'article.

courbe d'efficacité (O.C) parallèlement à elle-même; si l'on veut comparer différents plans d'échantillonnage, il suffit de commencer par ajuster les grandeurs des échantillons de manière à ce que les courbes d'efficacité (O.C) aient un point commun, par exemple, le "point de contrôle" ou qualité d'indifférence p_{50} définie par

$$P(p_{50}) = 0,5 \qquad (5)$$

A cet effet, commençons par modifier les tailles des échantillons de manière que les deux courbes d'efficacité aient le même point de contrôle.

Dans la figure 4, par exemple, la courbe d'efficacité du plan D(2, 5, 10, 10) peut être comparée avec celle des plans simples correspondant à $c_o = 5$ et $c_o = 6$, après un tel ajustement : dans les deux cas, on constate un accord convenable et il est difficile de juger lequel des deux est le meilleur.

Pour rendre notre étude plus précise, il est maintenant nécessaire de réaliser l'ajustement par un procédé mathématique plus précis. Dans ce but nous adoptons le principe que deux courbes d'efficacité peuvent être considérées comme équivalentes lorsque, non seulement elles possèdent le même point de contrôle p_{50}, mais aussi la même pente en ce point.

Il sera commode de faire le graphique en utilisant pour p une échelle logarithmique, car la pente est alors indépendante de la valeur absolue des tailles des échantillons, les conditions de Poisson étant supposées vérifiées.

Nous définissons ainsi un second paramètre, la pente relative :

$$h = -\left(\frac{d \log P}{d \log p}\right)_{p_{50}} = -2\left(p\,\frac{dP}{dp}\right)_{p_{50}} \qquad (6)$$

où P est la probabilité d'accepter.

Deux courbes d'efficacité sont donc appelées équivalentes si elles possèdent les mêmes valeurs de h et de p_{50}.

Ceci est évidemment un point de vue théorique car, en réalité, le critère d'acceptation, et par conséquent h, varient de manière discontinue. Mais on peut surmonter cette difficulté en considérant des valeurs non entières du nombre c_o relatif au plan d'échantillonnage simple.

Ceci peut être conçu de la manière suivante : en prélevant l'échantillon, nous décidons en même temps si la valeur de c_o sera 5 ou 6 par un tirage au sort donnant les probabilités 0,77 pour 5 et 0,23 pour 6 ; nous déterminons ainsi un plan d'échantillonnage qui peut être défini comme caractérisé en moyenne par $c_o = 5,23$. Sa courbe d'efficacité pourra être obtenue par interpolation linéaire entre les courbes $c_o = 5$ et $c_o = 6$.

Une telle méthode ne peut évidemment être recommandée en pratique, mais elle constitue un argument théorique nous permettant de trouver un plan simple ayant exactement la même pente h qu'un plan double donné et par conséquent une courbe d'efficacité très voisine, ainsi que le montre la figure 4.

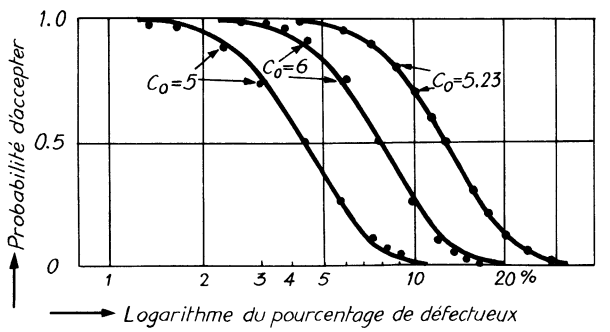

Fig. 4 - La courbe d'efficacité relative au plan d'échantillonnage double D(2;5,10, 10) comparée après un ajustement approprié de la taille de l'échantillon à celles de plans d'échantillonnage simple où $c_o = 5,6$ et $5,23$.

Nous pourrons de même considérer des valeurs non entières de la taille n_o de l'échantillon pour obtenir un parfait ajustement au point de contrôle p_{50}. Pratiquement les valeurs de n_o sont assez grandes pour qu'une telle condition soit négligeable : la considération de valeurs non entières pour c_o à plus d'importance.

L'ajustement est grandement facilité par le fait que les paramètres p_{50} et h sont, pour des plans d'échantillonnage simples, liés aux valeurs de n_o et c_o par les relations ci-après (cf. [2] et [3]).

$$n_o\, p_{50} = c_o + 0,67 \qquad (7)$$

$$\frac{\pi}{2} h^2 = n_o\, p_{50} + 0,06 = c_o + 0,73 \qquad (8)$$

La méthode de comparaison se présente alors comme suit : pour un plan double donné, on détermine h et p_{50}, on en déduit ensuite n_o et c_o à l'aide des équations (7) et (8), on définit ainsi le plan simple équivalent (voir en annexe les détails du calcul).

L'équivalence, ainsi définie, ne signifie pas que les courbes d'efficacité coincident complètement; il subsiste de légères différences dans les extrêmités. Mais ces différences sont sans influence réelle du point de vue pratique.

III. - UNE AUTRE DÉFINITION DE L'ÉQUIVALENCE

Habituellement une courbe d'efficacité n'est pas spécifiée à l'aide de h et p_{50}, mais de p_{95} et p_{10} correspondant à des probabilités d'accepter égales à 0,95 et 0,10 respectivement.

Pour un plan d'échantillonnage simple et dans les conditions de Poisson, le rapport p_{10}/p_{95} est, comme h, seulement fonction de c_o. Par suite un plan d'échantillonnage simple "équivalent" peut être défini comme le plan ayant même rapport p_{10}/p_{95} que le plan double considéré et dont la taille n_o de l'échantillon est calculée de manière que les deux courbes d'efficacité coincident à p_{95} et p_{10}.

Pour un ensemble de 5 plans d'échantillonnage double, les deux définitions de l'équivalence considérées ci-dessus sont comparées dans le tableau 1. Les trois derniers plans sont des plans d'échantillonnage double dont les courbes d'efficacité sont approximativement les mêmes, mais dont les critères de décision sont différents. Le plan D (2 , 5, 13, 13) correspond au diagramme de la fig. 2, c'est-à-dire, a priori, à un choix peu judicieux des critères de décision; le plan D(2 2,9,9) vérifie encore le principe $c_2 = c_3$ mais les critères de décision sont mieux choisis; enfin pour le dernier plan du tableau I on a, $c_2 \neq c_3$.

TABLEAU 1- COMPARAISON DE DEUX DEFINITIONS DE "L'EQUIVALENCE"

Plan d'échantillonnage double	Plan d'échantillonnage simple ajusté au moyen de				
	p_{50} , h		p_{10} , p_{95}		n_o/n_o'
	n_o	c_o	n_o'	c_o'	
D (2;0,1,1), n_1 = 90	139	0.63	132	0.69	1.053
D (2;0,4,4), n_1 = 90	238	3.66	231	3.64	1.030
D (2;5,13,13), n_1 = 150	196	6.96	203	7.64	0.966
D (2;2,9,9), n_1 = 90	202	7.23	194	7.18	1.041
D (2;1,4,8), n_1 = 75	192	6.88	185	6.69	1.038

Les différences entre les deux définitions de l'équivalence sont légères; elles ne dépassent pas 5% de la taille des échantillons. Les courbes d'efficacité des plans du tableau I sont si voisines les unes des autres qu'il est difficile de les représenter distinctement sur une seule figure. Les différences d'efficacité entre le plan D (2 ; 2, 9, 9) et les plans simples correspondants sont montrées dans le

tableau 2; pour les autres plans, elles sont du même ordre de grandeur. Comme on pouvait s'y attendre l'ajustement d'un plan simple à un plan double à l'aide de p_{50} et h est meilleur que celui (p_{10}, p_{95}) pour les valeurs de P voisines de 0,50 et moins bon pour les grandes et petites valeurs de P.

TABLEAU II

Probabilité d'accepter pour D(A;2,9,9) et les plans d'échantillonnage simple équivalents ajustés à l'aide de p_{50}, h et p_{10}, p_{15} respectivement

Pourcentage de défectueux	Plan		
	D(2;2,9,9,) $n_1 = 90$	$c_0 = 7,23$ $n_0 = 202$	$c_0 = 7,18$ $n_0 = 194$
	P		
1%	0.9996	0.9992	0.9993
2	.9636	.9555	.9625
3	.7688	.7623	.7903
4	.4734	.4733	.5155
5	.2431	.2340	.2711
6	.1159	.0963	.1196
7	.0551	.0343	.0458
8	.0266	.0109	.0157

Dans l'ensemble, les deux définitions ci-dessus de l'équivalence sont également acceptables. Dans tout ce qui suit, la définition du § II est adoptée en raison de la facilité des calculs auxquels elle conduit. Si l'autre définition était utilisée, il en résulterait des variations systématiques d'environ 3 à 5% sur les efficiences définies au § IV; ce qui n'altérerait en rien les principales conclusions qui en seront tirées.

IV. - UNE THÉORIE DE L'EFFICIENCE [1]

Après avoir trouvé avec une précision suffisante un plan simple équivalent à un plan double donné, il reste à comparer le nombre moyen des observations nécessitées par ces deux plans.

On appelle "efficience inverse" le rapport du nombre moyen d'éléments à prélever pour l'échantillonnage double à la taille de l'échantillon du plan simple équivalent. Conformément à l'habitude, il est nécessaire d'envisager le cas où le deuxième échantillon du plan double peut être tronqué. (Le premier échantillon du plan double ou l'échantillon du plan simple ne l'étant pas). Dans ce §, nous envisageons seulement le cas où le deuxième échantillon est entièrement contrôlé. Les § V et VI reprennent le problème pour des échantillons tronqués.

Le nombre moyen d'éléments à prélever dans un lot par échantillonnage double et par suite l'efficience inverse, varient avec le pourcentage 100 p des défectueux contenus dans le lot. Pour ce qui nous concerne, il est plus pratique d'étudier l'efficience inverse, non pas en fonction de p mais de P, probabilité d'accepter un lot ayant 100 p% de défectueux et en portant en abscisse les valeurs de P sur une échelle de Henry (a). Les courbes ainsi obtenues seront appelées caractéristiques d'efficiences et serviront à étudier l'efficience de différents plans d'échantillonnage double.

Le plan double D(2;5,10,10) ne satisfait pas le critère (2) du § I et si les arguments de ce § sont valables, il devrait être classé comme peu efficient. La fig. 5 confirme immédiatement cette opinion. Même pour de bons lots, qui ont plus de 99% de chances d'être acceptés, l'efficience inverse est à peine inférieure à 1 et le nombre moyen d'éléments à contrôler est sensiblement le même pour deux plans équivalents simple et double. Pour les lots de mauvaise qualité, le plan double est beaucoup moins économique que le plan simple.

(1) A ne pas confondre avec l'efficacité : il s'agit ici d'une notion de coût devant permettre de choisir entre plusieurs plans de même efficacité au sens de l'O.C.

(2) Echelle fonctionnelle correspondant à la loi normale.

Si l'on passe du plan D(2;5,10,10) au plan D(A; 2,10,10), nous devons, en raison du § I nous attendre à une amélioration de l'efficience et ceci est immédiatement confirmé par la fig. 5. Les caractéristiques d'efficience de la fig. 5 comparent respectivement deux plans doubles aux plans simples équivalents qui leur correspondent. Le problème serait alors d'utiliser des courbes pour représenter l'efficience de deux plans doubles par rapport à n'importe quel autre. Nous pensons que cela est possible pour les raisons suivantes : les deux plans doubles en question ont des courbes d'efficience dont les pentes sont différentes et, par conséquent, les nombres moyens d'éléments à contrôler ne sont pas comparables, mais plus une courbe d'efficacité se rapproche d'un "échelon", plus grand est le nombre des éléments à prélever, que ce soit en échantillonnage simple, double, progressif; il doit donc être possible de mesurer l'influence de la pente et faire la correction correspondante. Numériquement, le calcul donne pour le plan D(2;2,10,10) une pente relative h = 2,427 et pour le plan D(2;5,10,10) h = 1,949 ; en raison de l'équation (8) le rapport des tailles des échantillons simples équivalente est :

$$\frac{n_o}{n'_o} = 1,56$$

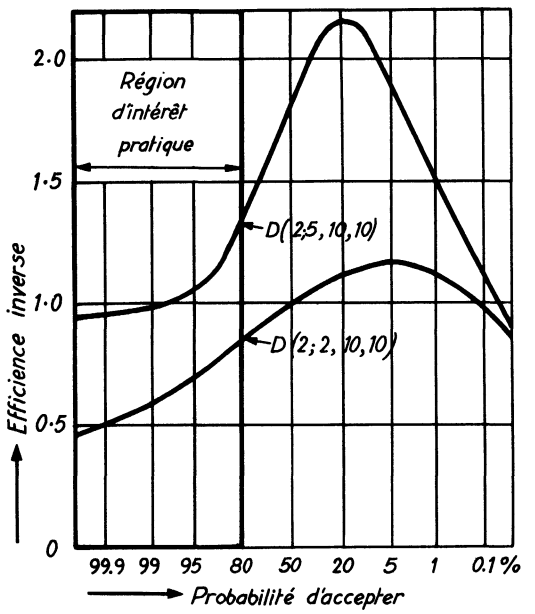

Fig. 5 - Caractéristiques d'efficience pour les plans doubles D(2,5,10,10) et D(2,2,10,10). Une réduction du critère d'acceptation c_1 de 5 à 2 produit une amélioration frappante.

En définitive, le fait de l'accroissement de pente de D(2;5,10,10) à D(2;2,10,10) entraine une multiplication par 1,56 du nombre des éléments à contrôler. Si nous faisons l'hypothèse que ce facteur est encore valable pour l'échantillonnage double, il est alors possible d'éliminer la différence entre les pentes et, par suite de rendre plus directement comparables les courbes qui représentent le nombre moyen des éléments à contrôler pour deux plans doubles. L'efficience inverse réalise immédiatement cette transformation. Nous pouvons exprimer ceci d'une autre manière. Des plans d'échantillonnage simples qui n'utilisent pas les mêmes critères de rejet ont des pentes relatives différentes et, du point de vue efficience ne sont pas comparables. Il n'est donc pas déraisonnable d'accepter les plans d'échantillonnage simple comme un ensemble de référence qui par définition a une efficience égale à 1. De ce point de vue, la fig. 5 peut être considérée comme donnant une esquisse des mérites relatifs des deux plans d'échantillonnage double l'un par rapport à l'autre; et les caractéristiques d'efficience peuvent être utilisées pour juger des mérites relatifs de n'importe quel plan double.

En pratique, la qualité des lots est telle qu'il y a une haute probabilité de les accepter et seul le côté gauche de la figure 5 est surtout utilisé; les maxima des caractéristiques d'efficience sont en réalité de peu d'intérêt puisque les lots qui correspondent à cette région sont rares. Pour cette raison, la partie de la figure 5 où les probabilités d'accepter sont supérieures à 0,80 a été indiquée comme la "<u>région d'intérêt pratique</u>".

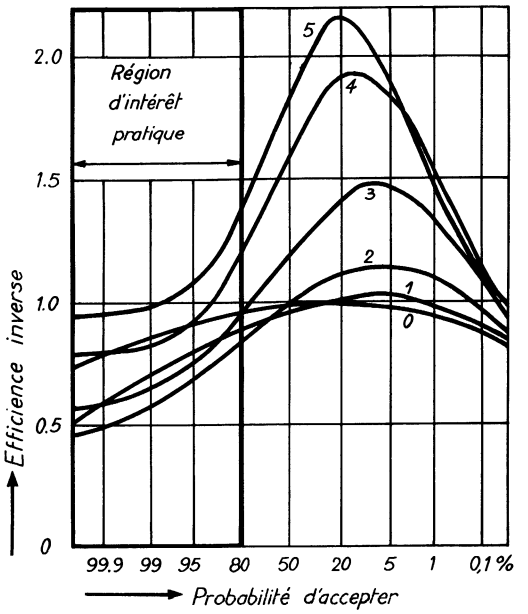

Fig. 6 - Caractéristiques d'efficience pour la suite des plans D(2; c_1, 10,10) quand c_1 = 5,4, ... 0.

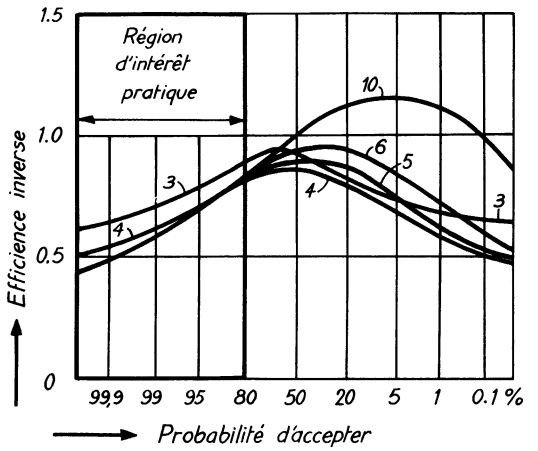

Fig. 7 - Caractéristiques d'efficience pour les plans D(2;2,c_2,10) avec c_2 = 10, 6, 5, 4 et 3.

Une application plus étendue des caractéristiques d'efficience est considérée fig. 6 pour les plans D(2;5,10,10) ; D(2;4,10,10) ; D(2;3,10,10) etc ... On sent immédiatement que dans la région d'intérêt pratique, le plan D(2;2,10,10) est le meilleur.

En prenant $c_1 = 2$, nous étudions, fig. 7, l'amélioration que l'on pourrait obtenir si l'on abandonnait le principe $c_2 = c_3$. Comme on s'y attendait, un changement de D(2;2,10,10) à D(2;2,5,10) améliore l'efficience principalement du côté des mauvais lots. Dans la région d'intérêt pratique, l'effet est presque nul et les deux plans sont également satisfaisants.

V. - ÉCHANTILLONNAGE TRONQUÉ

Les caractéristiques d'efficience considérées jusqu'ici reposent sur l'hypothèse de l'inspection complète des échantillons. Pour réduire le nombre des observations, il est quelquefois recommandé de tronquer le deuxième échantillon; ainsi, le contrôle du deuxième échantillon est arrêté dès que le nombre total des défectueux constaté dépasse le critère de rejet c_3. Evidemment, le procédé n'altère pas la courbe (O.C) mais améliore l'efficience définie ci-dessus.

Pour étudier l'influence d'un examen tronqué, nous avons utilisé les trois plans d'échantillonnage double considérés précédemment dans la table 1, c'est-à-dire D(2;5,13,13) ; D(2;2,9,9,) et D(2;1,4,8). Ils ont presque les mêmes pentes relatives h = 2,21 ; 2,25, 2,20 respectivement et si les tailles des échantillons sont correctement ajustées comme dans la table 1, ils ont des courbes d'efficacité (O.C) presque identiques.

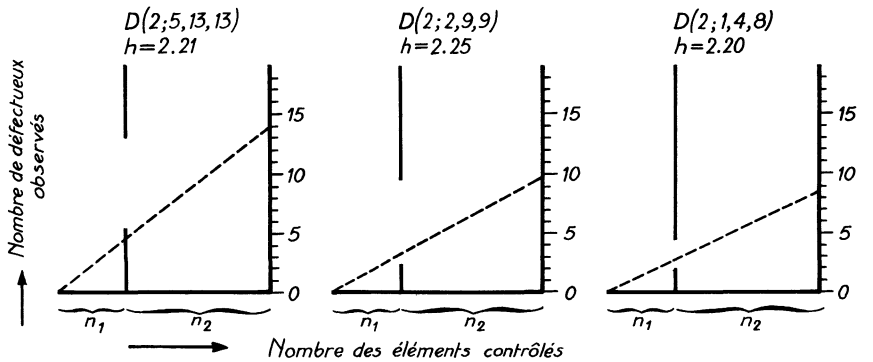

Fig. 8 - Diagrammes relatifs à trois plans doubles ayant presque la même pente h. L'efficience de ces plans, tronqués ou non, est montrée fig. 9.

Les diagrammes correspondants sont donnés fig. 8 ; le plan D(2;5,13,13) est du type de la fig. 2 et n'est pas réellement efficient. Avec un plan de ce type, les lots considérés comme douteux par le contrôle du premier échantillon ont beaucoup de chances d'être rejetés lors du contrôle du deuxième échantillon; on peut par conséquent s'attendre à réaliser une économie importante en tronquant le deuxième échantillon. Pour le plan D(2;2,9,9) l'amélioration sera moindre, bien qu'appréciable puisque la moitié supérieure de l'ouverture pratiquée dans le premier écran est encore trop grande. Pour le plan D(2;1,4,8) le fait de tronquer l'échantillon est sans conséquence appréciable.

Les caractéristiques d'efficience (fig. 9) sont en accord avec ces résultats (les efficiences relatives aux plans tronqués ont été calculées à l'aide des formules données dans "Sampling Inspection", [6, page 209]. Il sera noté que dans la région d'intérêt pratique l'échantillonnage tronqué a relativement peu d'effet. Ceci provient du fait que la coupure est pratiquée uniquement dans le deuxième échantillon dont le prélèvement n'est fréquent que pour les lots de mauvaise qualité. Comme la région pratique correspond à des lots de bonne qualité, on utilise rarement cet avantage dans cette région.

Ainsi, que l'on tronque ou ne tronque pas le deuxième échantillon, le plan double D(2;2,9,9) est toujours considérablement meilleur que le plan D(2;5,13,13). Ce dernier figure dans les Mil Std. 105 A, et la figure 9 illustre l'amélioration que l'on peut obtenir en choisissant mieux les critères d'acceptation.

VI. - ÉCHANTILLONNAGE TRONQUÉ ET CHOIX D'UN PLAN D'ÉCHANTILLONNAGE DOUBLE

Les tables de Dodge et Romig [1] contiennent des plans d'échantillonnage double pour lesquels $c_2 = c_3$ et depuis, tous les plans d'échantillonnage double proposés à l'usage courant ont été de ce même type. Dodge et Romig ne mentionnaient pas quel choix spécial ils avaient fait; par suite, peu ou aucune attention n'était faite à ce problème. Celui-ci avait été discuté par Stem et Shaw [7] dans un rapport qui n'est pas facilement accessible et n'est pas cité dans les livres relatifs à ce sujet. C'est pourquoi nous allons lui consacrer quelques lignes.

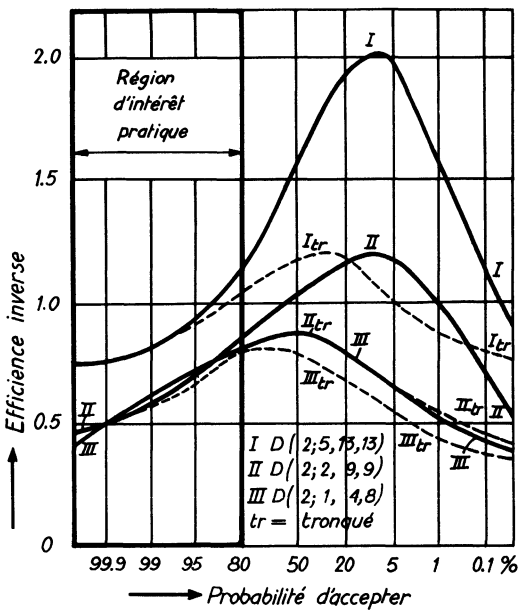

Fig. 9 - Effet de l'échantillonnage tronqué sur l'efficience (au sens de l'article). Caractéristiques d'efficience pour le plan D(2,5,13,13); D(2,2,9,9) et D(2,1,4,8) échantillonnage complet et tronqué.

Le choix $c_2 = c_3$ peut être justifié pour simplifier la spécification d'un plan d'échantillonnage double quand la perte d'efficience se produit dans une région de petite importance pratique (fig. 9). D'un autre côté, si nous prenons $c_2 \neq c_3$ et tronquons les échantillons quand cela est possible, nous simplifions d'un côté et compliquons de l'autre; l'on peut se demander si cela est correct. En fait, nous voyons sur la figure 9 que le plan D(2;2,9,9) avec échantillonnage tronqué a presque la même caractéristique d'efficience que D(2;1,4,8) avec contrôle total du deuxième échantillon. Lequel de ces deux plans doit-on préférer en pratique?

Dans les Mil. Std. 105 A, les entrées correspondant aux deux plans d'échantillonnage double précédemment utilisés D(2;2,9,9) et D(2;1,4,8) sont respectivement :

	Accepter	Rejeter
1er échantillon	2	10
2e "	9	10
1er "	1	5
2e "	8	9

L'avantage présenté par $c_2 = c_3$ est que le critère de rejet est le même pour les deux échantillons. Nous ne croyons pas toutefois que ceci soit d'un intérêt réel en pratique; habituellement, le contrôleur doit de toute façon consulter la table à

la fin du contrôle du premier échantillon pour prendre une décision; il y a beaucoup de chances pour qu'il soit obligé de le faire à la fin du contrôle du deuxième, même lorsque $c_2 = c_3$.

Lorsque l'on peut tronquer le contrôle du deuxième échantillon, il n'en est plus ainsi. Le contrôleur doit avoir constamment sous les yeux le nombre des défectueux trouvés chaque fois qu'il inspecte un deuxième échantillon et la routine du contrôle n'est plus la même.

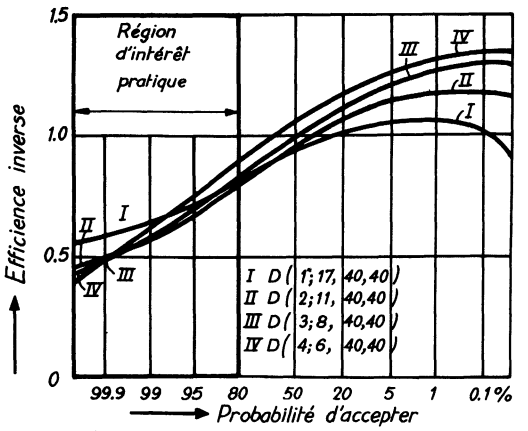

Fig. 10 - Effet du rapport des tailles des 1er et 2ème prélèvements, n_2/n_1, quand $c_2 = c_3$.

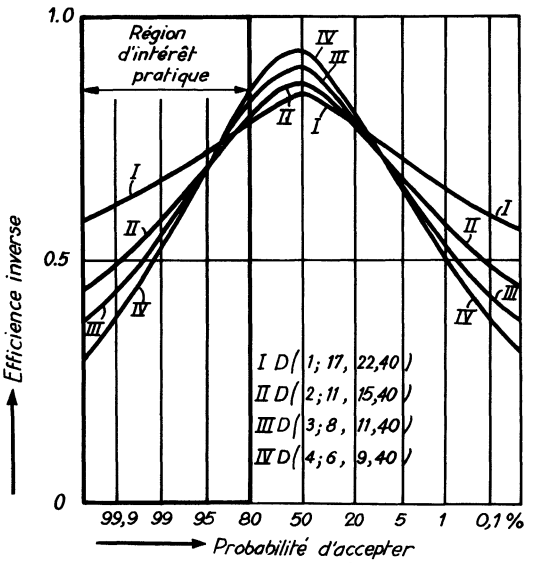

Fig. 11 - Effet du rapport des tailles des 1er et 2ème prélèvements, n_2/n_1, quand $c_2 \neq c_3$.

En conséquence, notre opinion est que, si les critères de rejet et d'acceptation sont présentés explicitement comme dans les Mil Std 105 A, les plans d'échantillonnage double où $c_2 = c_3$ seront plus satisfaisants en pratique que ceux où $c_2 = c_3$, mais avec possibilité de tronquer le deuxième échantillon. Nous pouvons d'ailleurs ajouter que nous ne savons pas dans quelle mesure la possibilité de tronquer est réellement utilisée. Ce procédé est recommandé dans la plupart des livres, mais n'est pas mentionné dans les Mil Std 105A. Nous sommes amenés à conclure qu'il a été trouvé peu pratique, peut être pour les raisons ci-dessus. Quelque information à ce sujet serait désirable.

En conclusion, il pourrait être souligné une fois de plus que dans la région d'intérêt pratique le gain apporté soit par l'échantillonnage tronqué, soit par $c_2 \neq c_3$ n'est pas important.

VII. - RAPPORT DES TAILLES DES ÉCHANTILLONS

Dans les tables originales de Dodge et Romig [1] le rapport $\frac{n_2}{n_1}$ des tailles des échantillons varie quelque peu, tandis que dans celles ultérieures il est constant et $\frac{n_2}{n_1} = 2$. A l'aide des méthodes considérées ci-dessus, il est facile de voir dans quelle mesure le choix est satisfaisant.

La fig. 10 et 11 représentent les caractéristiques d'efficience de plans d'échantillonnage où $\frac{n_2}{n_1} = 1, 2, 3$ et 4. Pour le calcul de ces courbes les critères d'acceptation c_1 et c_2 sont grands et par suite ajustés avec assez de précision pour donner une efficience (au sens de cet article) optimum. Au seuil de probabilité d'accepter égal à 95%, la valeur n_2/n_1 a relativement peu d'importance, tandis que pour les probabilités inférieures à celle-ci, l'augmentation de n_2/n_1 conduit à une diminution de l'efficience. Les différences dans cette région sont plus grandes pour $c_2 = c_3$ que pour $c_2 \neq c_3$ (voir fig. 10-11).

Dans l'ensemble, le choix $n_2/n_1 = 2$ semble satisfaisant. Quand la qualité est très bonne et le rejet des lots rare, il peut y avoir avantage à augmenter le rapport n_2/n_1, mais il est peut être encore mieux d'arrêter l'échantillonnage.

Nous avons donc discuté l'efficience des plans d'échantillonnage double sous différents aspects. Dans des articles futurs, nous avons l'intention de voir comment choisir pratiquement un plan d'échantillonnage et fournir une suite de tables qui permettent la détermination rapide des plans simples et doubles appropriés.

APPENDICE

Formules relatives au calcul de p_{50} et h pour des plans d'échantillonnage double.

Nous écrirons
$$q(k;m) = e^{-m} \frac{m^k}{k!} \tag{9}$$

$$R(c;m) = \sum_{0}^{c} q(k;m) \tag{10 a}$$

$$Q(c;m) = 1 - R(c;m) = \sum_{k=c+1}^{\infty} q(k;m). \tag{10 b}$$

et nous aurons alors

$$\dot{q}(k;m) = \frac{dq(k;m)}{dm} = q(k-1;m) - q(k;m) \tag{11}$$

$$\dot{Q}(c;m) = \frac{dQ(c;m)}{dm} = q(c,m) = Q(c-1\ ;m) - Q(c;m) \tag{12}$$

formules qui peuvent être utilisées ainsi pour $k = 0$ ou $c = 0$ si nous posons

$$q(-1;m) = 0 \qquad Q(-1;m) = 1 \tag{13}$$

La probabilité d'accepter pour un plan d'échantillonnage double avec $c_2 = c_3$ est :

$$P = R(c_1;m_1) + \sum_{k=c_1+1}^{c_2} q(k;m_1)\, R(c_2 - k;m_2) \tag{14}$$

où
$$m_1 = n_1 p \quad , \quad m_2 = n_2 p$$

et p est la proportion des défectueux contenus dans le lot.

Ensuite,

$$R\left\{ c_2 \,;\, (m_1 + m_2) \right\} = \sum_{k=0}^{c_2} q(k;m_1)\ R(c_2 - k;m_2) \tag{15}$$

dont l'introduction dans (14) donne :

$$P = R(c_1;m_1) + R\left[c_2;(m_1 + m_2)\right] - \sum_{k=0}^{c} q(k;m_1)R(c_2-k;m_2) \qquad (16)$$

Remplaçant Q par 1 - R :

$$P = 1 - Q(c_2 ; m_1 + m_2) + \sum_{k=0}^{c_1} q(k;m_1)\, Q\,(c_2-k ; m_2) \qquad (17)$$

Puisque, pour un plan donné, les tailles n_1 et n_2 des échantillons sont fixées nous avons, lorsque p varie :

$$h = -2\,p\,\frac{dP}{dp} = -2\,m_1\,\frac{dP}{dm_1} \qquad (18)$$

et

$$\frac{dm_2}{dm_1} = \frac{n_2}{n_1} \qquad (19)$$

La différentiation de (17) donne :

$$-\dot{P} = -\frac{dP}{dm_1} = q\,(c_1 ; m_1)\,Q(c_2 - c_1 - 1 ; m_2)$$

$$+ \left(1 + \frac{n_2}{n_1}\right)\left[q(c_2 ; m_1 + m_2) + \sum_{k=0}^{c_1} q(k;m_1)Q(c_2-k;m_2)\right.$$

$$\left. - \sum_{k=0}^{c_1} q(k;m_1)\,Q(c_2-k-1 ; m_2)\right] \qquad (20)$$

Comme les fonctions $q(k;m)$, $Q(c,m)$ ont été complètement tabulées par Molina [5], les formules (17) et (20) sont d'utilisation pratique. Le calcul d'une somme de produits de 0 à c_1 nécessite moins de travail que celui de $(c_1 + 1)$ à c_2, quand $c_2 = c_3$.

Lorsque $c_2 \neq c_3$, nous avons

$$P = R(c_1 ; m_1) + \sum_{k=c_1+1}^{c_2} q(k;m_1)R(c_3 - k - m_2)$$

ou

$$P = 1-Q(c_2;m_1) - \sum_{k=c_1+1}^{c_2} q(k;m_1)\, q(c_3-k ; m_2) \qquad (21)$$

et par différentiation :

$$-\dot{P} = -\frac{dP}{dm_1} = q(c_2 ; m_1) - q(c_2 ; m_1)Q(c_3 - c_2 - 1 ; m_1)$$

$$+ q(c_1 ; m_1)\, Q\,(c_3 - c_1 - 1 ; m_2)$$

$$+ \left(1 + \frac{n_2}{n_1}\right) \sum_{k=c_1+1}^{c_2} q(k;m_1)Q(c_3 - k - 1;m_2)$$

$$- \sum_{k=c_1+1}^{c_2} q(k ; m_1)\, Q\,(c_3 - k ; m_2) \qquad (22)$$

Il n'est guère possible d'aller plus loin pour transformer les données, car lorsque $c_2 \neq c_3$, c_1 et c_2 sont assez voisins l'un de l'autre et une sommation des produits de $c_1 + 1$ à c_2 est rapide.

La méthode générale consiste à calculer P et $-\overset{*}{P}$ pour deux valeurs de m_1 qui donnent des valeurs de P voisines de P = 0,50 et trouver ensuite la valeur correcte par interpolation inverse. La table 3 fournit quelques détails sur les calculs numériques.

Pour obtenir la dérivée $\overset{*}{P}_{50}$ une méthode d'interpolation un peu plus précise est nécessaire. Pour deux valeurs de m_1, $m_{10} = 2,9$ et $m_{12} = 3,0$ nous avons cal-

culé les valeurs correspondantes P_0, P_2 et P_0, $\dot{P}_2$. Il reste à ajuster une équation du 3ème degré dont le graphique passe par les points m_{10}, P_0 et m_{12}, P_2 avec des pentes correctes. De cette équation, nous pouvons déduire la pente au point m_{11}, $P_1 = 1/2$. Ce principe donne la formule d'interpolation suivante :

$$\overset{*}{P} = \frac{P_2 - P_0}{m_{12} - m_{10}} \, 6\theta\Phi + \dot{P}_0 \theta \, (1-3\Phi) + \dot{P}_2 \, \Phi \, (1 - 3\theta),$$

ou

$$\theta = \frac{m_{11} - m_{10}}{m_{12} - m_{10}} \qquad \Phi = 1 - \theta$$

C'est la formule qui a été utilisée dans cet article; initialement une méthode moins raffinée avait été adoptée mais avait conduit à quelques imprécisions. Utilisant les tables de Molina, le calcul peut être fait directement à partir des tables. Le calcul de h pour un plan double particulier ne prend pas plus d'une demi-heure.

Il doit être noté que

$$P_M(c;m) = Q(c - 1 \, m) \tag{23}$$

où P_M désigne la probabilité telle qu'elle a été tabulée par Molina. Pour une valeur quelconque de m_1, la probabilité d'avoir à contrôler un deuxième échantillon est :

$$\sum_{k=c_1+1}^{c_2} q(k;m_1) = Q(c_1 + 1 \, ; m_1 - Q(c_2; m_1) \tag{24}$$

et l'efficience inverse est :

$$I.E. = \frac{n_1}{n_0} \left[1 + \frac{n_2}{n_1} \left\{ Q(c_1 + 1 \, ; m_1) - Q(c_2 \, ; m) \right\} \right] \tag{25}$$

tandis que la probabilité correspondante d'accepter est fournie par (17) ou (21).

TABLE III Calcul numérique de $n_1 p_{50}$ et h pour le plan d'échantillonnage D (2 1 4 8)

m_1 m_2	2.9 5.8	3.0 6.0
$a = 1 - Q(c_2; m_1)$ $b = q(c_2; m_1)$ $c = q(c_2; m_1) Q(c_3 - c_2 - 1; m_2)$ $d = q(c_1; m_1) Q(c_3 - c_1 - 1; m_2)$ $e = \sum_{k=c_1+1}^{c_2} q(k; m_1) \, Q(c_3 - k; m_2)$ $f = \sum_{k=c_1+1}^{c_2} q(k; m_1) \, Q(c_3 - k - 1; m_2)$	0.831777 0.162154 0.134594 0.057701 0.311792 0.409015	0.815263 0.168031 0.142624 0.058803 0.332528 0.426992
$P = a - e$ $-\dot{P} = b - c + d + 3(f - e) =$	0.519985 0.376930	0.482735 0.367602
Par interpolation inverse		
$m_{11} = n_1 p_{50}$ $-\dot{P}_{50}$ $h = -2 \, m_{11} \dot{P}_{50}$	2.954 0.3726 2.201	
c_0 $n_0 p_{50}$ n_1/n_0	6.88 7.551 0.392	

La valeur de $m_{11} = n_1 p_{50}$ était obtenue par interpolation inverse linéaire : la courbe O.C dans la région entre P = 0,52 et 0,48 est suffisamment rectiligne pour arriver à une valeur précise de cette manière.

BIBLIOGRAPHIE

(1) DODGE, HAROLD F., and ROMIG, HARRY G., Sampling inspection tables, New-York; John Wiley and sons, Inc. 1944 : or Bell System Technical journal 8 (1928) 613-31.

(2) HAMAKER H.C., "Some notes on lot-by-lot inspection by attributes". Review of the International Statistical Institute 18 (1950), 179-96.

(3) HAMAKER H.C., "The theory of Sampling inspection plans" Philips Technical Review, 11 (1950), 260-70.

(4) MILITARY STANDARD 105 A : Sampling procedures and tables for inspection by attributes. U.S. Government Printing Office. The standard is reproduced in full in "Quality Control and Industrial Statistics" by A.J. Duncan, Chicago R.D. Irwin Inc., 1952.

(5) MOLINA Edward C, Poisson's Exponential Binomial limit, New-York; Van Nostrand, 1947.

(6) STATISTICAL RESEARCH GROUP, Columbia University, Sampling inspection, New-York; Mc. Graw-Hill Book Co, 1948.

(7) STEIN, ARTHUR, and SHAW, LAWRENCE W., "Some methods of reducing the amount of inspection in the application of double sampling inspection procedure". Ballistic Research Laboratory Report n° 248, Aberdeen Proving Ground, 1943.