

# REVUE DE STATISTIQUE APPLIQUÉE

A. HALD

**Plans d'échantillonnage simple par attributs basés sur des considérations de courbe caractéristique, variations dans la qualité attendue et coûts**

*Revue de statistique appliquée*, tome 17, n° 2 (1969), p. 69-81

[http://www.numdam.org/item?id=RSA\\_1969\\_\\_17\\_2\\_69\\_0](http://www.numdam.org/item?id=RSA_1969__17_2_69_0)

© Société française de statistique, 1969, 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/>

PLANS D'ÉCHANTILLONNAGE SIMPLE  
 PAR ATTRIBUTS BASÉS SUR DES CONSIDÉRATIONS  
 DE COURBE CARACTÉRISTIQUE,  
 VARIATIONS DANS LA QUALITÉ ATTENDUE ET COÛTS

A. HALD

1 - INTRODUCTION

Le but de cet exposé est de présenter, sans démonstration, quelques résultats récents de la théorie des plans d'échantillonnage simple, ainsi que quelques résultats nouveaux.

Soient  $N$ ,  $n$  et  $c$  l'effectif du lot, celui de l'échantillon et le critère d'acceptation. Les variations attendues dans la qualité des lots présentés à l'inspection sont représentées par une distribution binomiale mélangée :

$$f_N(X) = \int_0^1 \binom{N}{x} p^x q^{N-x} dW(p),$$

où  $X$  désigne le nombre de défectueux dans le lot et  $W(p)$  la distribution (a priori) de  $p$ , qualité de la fabrication. On considèrera pour  $p$ , soit une distribution discrète, soit une distribution continue. Le cas discret sera discuté en termes de distribution à deux points, c'est-à-dire que  $p$  prendra les valeurs  $p_1$  et  $p_2$  ( $p_1 < p_2$ ) avec les probabilités  $w_1$  et  $w_2$  ( $w_1 + w_2 = 1$ ), mais les résultats peuvent être généralisés à toute distribution continue, on supposera qu'elle est dérivable, et on écrira  $W'(p) = w(p)$ .

La probabilité moyenne d'acceptation, ou fonction caractéristique, pour des lots produits sous contrôle binomial avec une qualité de fabrication  $p$  est :

$$P(p) = B(c, n, p) = \sum_{x=0}^c \binom{n}{x} p^x q^{n-x}$$

On posera  $Q = 1 - P$ .

Pour décrire les conséquences économiques des décisions erronées, on admettra qu'il existe une qualité limite  $p_r$ , c'est-à-dire que les lots de qualité  $p < p_r$  devraient être acceptés et les lots de qualité  $p > p_r$  refusés ( $0 < p_r < 1$ ), et que la perte entraînée par l'acceptation d'un lot de qualité  $p > p_r$  est  $N(p - p_r)$ ; de même la perte entraînée par le rejet d'un lot de qualité  $p < p_r$  est  $N(p_r - p)$ . Enfin les coûts d'échantillonnage et de contrôle sont  $kn$  ( $k > 0$ ).

Les trois fonctions fondamentales,  $P(p)$ ,  $w(p)$  et  $|p - p_r|$  sont représentées sur la figure 1.

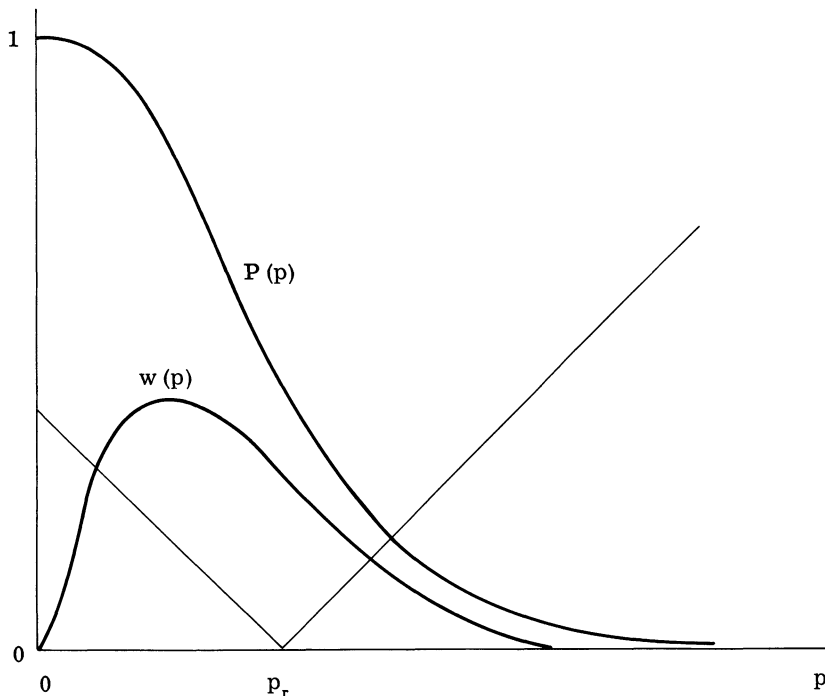


Figure 1

Du point de vue économique, les lots de qualité  $p < p_r$  devraient être acceptés. On définit donc la partie acceptable de la distribution a priori par :

$$w_1 = \int_0^{p_r} dW(p) \quad \text{et} \quad w_2 = 1 - w_1.$$

Le risque du fournisseur  $Q_1$  est, par définition, la probabilité de rejet alors que la décision correcte serait l'acceptation, soit

$$Q_1 = \int_0^{p_r} Q(p) dW(p)/w_1 \tag{1}$$

De même le risque du client est

$$P_2 = \int_{p_r}^1 P(p) dW(p)/w_2 \tag{2}$$

Pour une distribution à deux points, avec  $p_1 < p_r < p_2$ , on a :

$$Q_1 = Q(p_1) = 1 - B(c, n, p_1) \quad \text{et} \quad P_2 = P(p_2) = B(c, n, p_2),$$

C'est la définition habituelle des risques du fournisseur et du client correspondant aux deux niveaux de qualité.

La perte moyenne par unité liée à la décision comprend la perte entraînée par le rejet des lots acceptables.

$$d_1(c, n) = \int_0^{p_r} (p_r - p) Q(p) dW(p)$$

et la perte entraînée par l'acceptation des lots inacceptables.

$$d_2(c, n) = \int_{p_r}^1 (p - p_r) P(p) dW(p), \quad (4)$$

de sorte que :

$$d(c, n) = d_1(c, n) + d_2(c, n).$$

La fonction de regret est :

$$R(c, n, N) = kn + (N - n) d(c, n). \quad (5)$$

Les coûts correspondants pour une acceptation ou un rejet sans contrôle sont :

$$R_a = N \int_{p_r}^1 (p - p_r) dW(p) = N(\bar{p} - p_m)$$

$$R_r = N \int_0^{p_r} (p_r - p) dW(p) = N(p_r - p_m)$$

avec

$$\bar{p} = \int_0^1 p dW(p) \quad \text{et} \quad p_m = \int_0^{p_r} p dW(p) + \int_{p_r}^1 p_r dW(p)$$

Il en résulte que

$$R_a < R_r \quad \text{si} \quad \bar{p} < p_r$$

Une étude plus détaillée de ce modèle a été donnée par Hald (1960, 1967a).

Pour une distribution à deux points, la perte moyenne par unité liée à la décision devient une combinaison linéaire des deux risques, puisque :

$$d(c, n) = w_1 Q(p_1) (p_r - p_1) + w_2 P(p_2) (p_2 - p_r), \quad p_1 < p_r < p_2 \quad (6)$$

Pour une distribution continue on trouve, à partir d'un théorème sur la valeur moyenne :

$$d(c, n) = w_1 Q_1 (p_r - p_1^*) + w_2 P_2 (p_2^* - p_r), \quad p_1^* < p_r < p_2^* \quad (7)$$

On doit toutefois noter que  $(p_1^*, p_2^*)$  sont fonction de  $(c, n)$  tandis que  $(p_1, p_2)$  sont des constantes données. Lorsque  $c/n \rightarrow p_r$ , on trouve que  $p_1 \rightarrow p_r$  et  $p_2 \rightarrow p_r$ .

Si les hypothèses sur lesquelles est basée la fonction de regret  $R$  sont réalistes, le plan d'échantillonnage optimum doit être obtenu en minimisant  $R$ . Le minimum de  $R$  doit ensuite être comparé au plus petit de  $[R_a, R_r]$  pour savoir s'il convient ou non d'effectuer un contrôle sur échantillon. Dans la discussion qui suit, sur les différents systèmes de plans d'échantillonnage, on concentrera l'attention sur la structure de ces systèmes, laissant de côté les détails sur la détermination exacte des plans. On discutera donc surtout les expressions asymptotiques de  $c$  et de  $n$ . Cependant ces expressions sont suffisamment précises pour la plupart des applications pratiques lorsque  $c > 2$ . Pour obtenir les développements asymptotiques, on considèrera les fonctions qui interviennent comme des fonctions dérivables des variables réelles et positives  $(c, n)$ , en exprimant la fonction cumulative de la loi binomiale sous la forme d'une fonction bêta incomplète. On peut ensuite tenir compte du caractère discret de  $c$  en introduisant certaines corrections.

## 2 - PLANS D'ECHANTILLONNAGE DEFINIS PAR LES RISQUES DU FOURNISSEUR ET DU CLIENT

On sait qu'un plan d'échantillonnage peut être déterminé à partir de deux niveaux de qualité  $(p_1, p_2)$  et deux risques  $(\alpha, \beta)$ ,  $\beta < 1 - \alpha$ , en résolvant les équations  $Q(p_1) = \alpha$  et  $P(p_2) = \beta$ . Une étude des solutions exactes et approchées a été donnée par Hald (1967 c).

Cette méthode de détermination d'un plan d'échantillonnage ne donne pas une solution réelle du problème ; elle le transforme seulement en concentrant l'attention sur la fonction caractéristique. Il reste encore à formuler quelques principes pour le choix de  $(p_1, p_2, \alpha, \beta)$  en tenant compte de l'effectif des lots, des variations attendues dans la qualité, et des coûts. On s'intéresse particulièrement aux règles qui spécifient  $(\alpha, \beta)$  comme fonctions décroissantes de  $N$  pour  $(p_1, p_2)$  fixés et de même aux règles qui donnent  $(p_1, p_2)$  en fonction de  $N$  pour  $(\alpha, \beta)$  fixés.

Les solutions asymptotiques habituelles des deux équations sont :

$$c + 1 = np_1 + u_{1-\alpha} \sqrt{np_1q_1} + a_1 + \underline{0}(n^{-1/2}) \quad (8)$$

et

$$c + 1 = np_2 - u_{1-\beta} \sqrt{np_2q_2} + a_2 + \underline{0}(n^{-1/2}), \quad (9)$$

où  $u_{1-\alpha}$  désigne le fractile  $(1 - \alpha)$  de la loi normale réduite et

$$a_1 = 1/2 + (u^2 - 1) (q_1 - p_1)/6$$

avec  $u = u_{1-\alpha}$  dans  $a_1$  et  $u = u_{1-\beta}$  dans  $a_2$ .

Résolvant par rapport à  $n$  et  $c$ , on obtient :

$$n(p_2 - p_1) - \sqrt{n} (u_{1-\beta} \sqrt{p_2q_2} + u_{1-\alpha} \sqrt{p_1q_1}) + a_2 - a_1 + \underline{0}(n^{-1/2}) = 0$$

pour la détermination de  $n$ , puis

$$c + 1 = n(p_1 + p_2)/2 + \sqrt{n}(u_{1-\alpha} \sqrt{p_1 q_1} - u_{1-\beta} \sqrt{p_2 q_2})/2 + (a_1 + a_2)/2 + \underline{0}(n^{-1/2})$$

Cependant on s'intéresse aussi à la solution pour  $(p_1, p_2)$  fixés et de petites valeurs de  $(\alpha, \beta)$ ; on cherche donc un développement qui soit valable lorsque  $\alpha \rightarrow 0$  et  $\beta \rightarrow 0$ . Supposons que  $c/n = p_0 + \varepsilon$ ,  $\varepsilon \rightarrow 0$  pour  $n \rightarrow \infty$ . Pour  $p_0 < p$ , on a alors :

$$B(c, n, p) = \lambda(p_0, p) n^{-1/2} \exp \{-n\varphi(p_0, p) - n\varepsilon\varphi'(p_0, p) + \underline{0}(\varepsilon) + \underline{0}(n^{-1})\}, \quad (10)$$

où

$$\begin{aligned} \lambda(p_0, p) &= q_0 p / |p - p_0| \sqrt{2\pi p_0 q_0} \\ \varphi(p_0, p) &= p_0 \ln(p/p) + q_0 \ln(q_0/q) \end{aligned} \quad (11)$$

et  $\varphi'(p_0, p) = \ln(p_0 q_0 / q_0 p)$  - voir Hald (1967 a). La même expression est valable pour  $1 - B(c, n, p)$  et  $p_0 > p$ .

Considérons maintenant les équations  $Q(p_1) = \alpha$  et  $P(p_2) = \beta$  lorsque  $\alpha \rightarrow 0$  et  $\beta \rightarrow 0$ . Il est clair que la solution doit satisfaire à la relation  $c/n = p_0 + \varepsilon$ ,  $\varepsilon \rightarrow 0$  et  $p_1 < p_0 < p_2$ , ce qui signifie que les deux risques tendent exponentiellement vers zéro en tant que fonctions de  $n$ . Si  $\varphi(p_0, p_1) \neq \varphi(p_0, p_2)$  alors l'un des deux risques tend exponentiellement vers zéro plus rapidement que l'autre.

On s'intéressera surtout au cas où  $\alpha$  et  $\beta$  tendent vers zéro avec le même taux de convergence. Supposons que  $\alpha/\beta = \rho$ , où  $\rho$  est une constante positive. Ceci conduit à  $\varphi(p_0, p_1) = \varphi(p_0, p_2)$ , soit :

$$p_0 = \left( \ln \frac{q_1}{q_2} \right) / \left( \ln \frac{p_2 q_1}{p_1 q_2} \right) \quad (12)$$

et

$$\varepsilon_n = \left( \ln \frac{p_1(p_2 - p_0)}{p_2(p_0 - p_1)} \right) / \left( \ln \frac{p_2 q_1}{p_1 q_2} \right) = a \quad (13)$$

de sorte que  $c = np_0 + a + \underline{0}(n^{-1})$ . On peut ensuite déterminer  $n$  à partir de l'une des équations originales, par exemple  $P(p_2) = \beta$ , ce qui donne :

$$n\varphi(p_0, p_2) + \frac{1}{2} \ln n - a \ln \frac{p_2 q_0}{p_0 q_2} + \frac{\beta}{\lambda(p_0, p_2)} + \underline{0}(n^{-1}) = 0, \quad (14)$$

Ainsi, en première approximation,  $n$  est proportionnel à  $\ln(1/\beta)$ . Si par exemple on choisit  $\beta$  proportionnel à  $N^{-k}$ ,  $k > 0$ ,  $n$  sera proportionnel à  $\ln N$ .

On voit que la relation entre  $c$  et  $n$  est approximativement linéaire avec des coefficients qui dépendent de  $p_1, p_2$  et  $\alpha/\beta$ , tandis que  $n$  dépend à la fois des 4 paramètres.

Supposons maintenant que la distribution a priori et la qualité limite soient connues. On peut alors trouver les deux risques généralisés,  $Q_1$  et  $P_2$  - voir les relations (1) et (2) - en fonction de  $(c, n)$ . Dans ce qui suit, nous discuterons la solution des équations  $Q_1 = \alpha$  et  $Q_2 = \beta$ .

Pour une distribution à deux points, on obtient les mêmes équations que précédemment, mais  $p_1$  et  $p_2$  sont maintenant donnés a priori.

Pour une distribution continue on trouve, à partir de  $w_1 Q_2 - w_2 P_1 = w_1 \alpha - w_2 \beta$  :

$$\int_0^1 B(c, n, p) w(p) dp = w_1 (1 - \alpha) + w_2 \beta.$$

Par un développement asymptotique de cette intégrale - voir Hald (1968 a)-on obtient :

$$c + 1 = (n + 1) p_0 - \frac{1}{2} p_0 q_0 w'(p_0)/w(p_0) + \underline{O}(n^{-1/2}), \quad (15)$$

où  $p_0$  est défini comme le  $w_1 (1 - \alpha) + w_2 \beta$  fractile de la distribution a priori, soit :  $W(p_0) = w_1 (1 - \alpha) + w_2 \beta$ .

A partir de cette équation et de  $W(p_r) = w$  on trouve :

$$(p_0 - p_r) w(p_r) + \frac{1}{2} (p_0 - p_r)^2 w'(p_r) + \dots = -\alpha w_1 + \beta w_2,$$

ce qui détermine  $\varepsilon = p_0 - p_r$ . Posant  $p_0 = p_r + \varepsilon$  dans (15) on obtient :

$$c = np_r + n\varepsilon - q_r - \frac{1}{2} p_r q_r w'(p_r)/w(p_r) + \underline{O}(\varepsilon) + \underline{O}(n^{-1/2}). \quad (16)$$

Comme pour la distribution discrète, on ne considèrera que le cas où  $\alpha$  et  $\beta$  tendent vers zéro avec le même taux de convergence, de sorte que  $\alpha/\beta \rightarrow \rho$ , où  $\rho$  est une constante positive. On peut démontrer que cette hypothèse conduit à une relation entre  $c$  et  $n$  de la forme

$$c = np_r + a_1 \sqrt{n} + a_2 + \underline{O}(n^{-1/2}) \quad (17)$$

et de plus que

$$w_2 p_2 = \sqrt{\frac{p_r q_r}{n}} w(p_r) m_1(-b_1) + \frac{e_2}{n} + \underline{O}(n^{-3/2}) \quad (18)$$

$$w_1 Q_1 = \sqrt{\frac{p_r q_r}{n}} w(p_r) m_1(b_1) + \frac{e_1}{n} + \underline{O}(n^{-3/2}) \quad (19)$$

où

$$b_1 = a_1/\sqrt{p_r q_r}, \quad m_1(-b_1) = \varphi(b_1) + b_1 \Phi(b_1), \quad \text{et } (e_1, e_2)$$

sont des constantes dépendant de  $(a_1, a_2)$  et  $[w(p_r), w'(p_r)]$

En comparant (16) et (17) on obtient :

$$a_1 = \varepsilon \sqrt{n} \quad \text{et} \quad a_2 = -q_r - \frac{1}{2} p_r q_r w'(p_r)/w(p_r)$$

Posant  $P_2 = \beta$  dans (18) on aboutit à une équation du 2e degré pour la détermination de  $n$ . En première approximation, on trouve :

$$n = p_r q_r [w(p_r) m_1(-b_1)]^2 / (w_2 \beta)^2 \quad (20)$$

Si  $\beta$  est proportionnel à  $N^{-k}$ , alors  $n$  est proportionnel à  $N^{2k}$ . On trouve aussi :

$$w_1 Q_1 / w_2 P_2 = m_1(b_1) / m_1(-b_1) \quad \text{de sorte que} \quad \rho = [w_2 m_1(b_1)] / [w_1 m_1(-b_1)]$$

En utilisant l'approximation normale de la loi binomiale, on trouve, à partir de (17) que  $P(p_r) \simeq \Phi(b_1)$  de sorte que toutes les courbes caractéristiques des plans satisfaisant à (17) ont en commun le point  $[p_r, \Phi(b_1)]$ .

Dans le cas particulier où  $\alpha/\beta = w_2/w_1$ , on a  $p_0 = p_r$  de sorte que (15) conduit à :

$$c = np_r - q_r - \frac{1}{2} p_r q_r w'(p_r) / w(p_r) + \underline{0}(n^{-1}) \quad (21)$$

Si  $p$  obéit à une distribution bêta de moyenne  $\bar{p}$  et de variance  $\sigma_p^2$  ce résultat peut s'exprimer comme suit :

$$c = np_r - \frac{1}{2} + \frac{1}{2} (p_r - \bar{p}) (\bar{p} - \sigma_p^2) / \sigma_p^2 \quad (22)$$

On notera que la solution des deux équations  $Q_1 = \alpha$  et  $Q_2 = \beta$  a des propriétés tout à fait différentes suivant que la distribution a priori est discrète ou continue.

Pour déterminer un système de plans à partir de ces équations, il est nécessaire de choisir  $\alpha$  et  $\beta$  en fonction de  $N$ . Lorsque ce choix est fait, les formules précédentes permettent de déterminer  $c$  et  $n$ .

### 3 - PLANS D'ECHANTILLONNAGE MINIMISANT LES COÛTS

Soit  $R_0(N) = \min_{c, n} R(c, n, N)$ . On se limite à l'échantillonnage simple, à l'acceptation sans inspection et au rejet sans inspection ; la solution optimale est celle qui correspond à  $\min [R_0(N), R_a(N), R_r(N)]$ . Nous appellerons cette solution la "solution bayésienne". Pour des lots importants, la solution sera toujours l'inspection par l'échantillonnage, pourvu que  $0 < W(p_r) < 1$ . On discutera les propriétés asymptotiques des plans d'échantillonnage bayésiens.

Pour une distribution a priori à deux points, le plan d'échantillonnage minimisant  $R(c, n, N)$  a les propriétés suivantes (pour la démonstration, voir l'article de Hald - 1967 a).

$$c = np_0 + a_0 + \underline{0}(n^{-1}),$$

où  $p_0$  est défini dans (12) et  $a_0$  dépend des  $(p_1, p_2, w_1, p_r)$ .



$$n \varphi_0 + \frac{1}{2} \ln n + b + \underline{Q}(n^{-1}) = \ln N, \quad (23)$$

où  $\varphi_0 = p_0 \ln(p_0/p_1) + q_0 \ln(q_0/q_1)$  et  $b$  dépendent des cinq paramètres.

$$\min_{(c, n)} R = k \{ n + 1/\varphi_0 + \underline{Q}(n^{-1}) \},$$

où  $n$  est la fonction de  $N$  donnée par (23).

$$d_1(c, n) = w_1 Q(p_1) (p_r - p_1) = k(1 - \delta)/N\varphi_0 (1 + \underline{Q}(n^{-1})), \quad (24)$$

$$d_2(c, n) = w_2 P(p_2) (p_2 - p_r) = k\delta/N\varphi_0 (1 + \underline{Q}(n^{-1})), \quad (25)$$

où

$$\delta = \left( \ln \frac{p_0 q_1}{p_1 q_0} \right) \left( \ln \frac{p_2 q_1}{p_1 q_2} \right)$$

Ce résultat ne dépend pas d'une structure linéaire des coûts ; les deux pertes  $(p_r - p_1, p_2 - p_r)$  peuvent être remplacées par deux nombres positifs quelconques.

Le théorème montre que  $n$  est asymptotiquement proportionnel à  $\ln N$ , et que la perte moyenne par lot liée à la décision,  $(N - n) d(c, n)$  tend vers  $k/\varphi_0$  tandis que les coûts d'échantillonnage sont  $(k/\varphi_0) \ln N$ .

Les deux risques  $Q(p_1)$  et  $P(p_2)$  sont proportionnels à  $N^{-1}$  (c'est-à-dire à  $e^{-n\varphi_0}$ ) et :

$$\frac{Q(p_1)}{P(p_2)} \longrightarrow \frac{w_2(p_2 - p_r) (1 - \delta)}{w_1(p_r - p_1) \delta} = \rho_0. \quad (26)$$

L'étendue des valeurs de  $(p_1, p_2)$  que l'on rencontre en pratique donne normalement des valeurs de  $\delta$  comprises entre 0,5 et 0,6.

La structure de la solution est la même que celle obtenue en résolvant les deux équations  $Q(p_1) = \alpha$  et  $P(p_2) = \beta$  pour  $\alpha = \alpha_0/N$  et  $\beta = \beta_0/N$ . Les deux solutions se confondent asymptotiquement si  $\alpha_0$  et  $\beta_0$  sont choisis comme il est indiqué dans (24) et (25).

Pour une distribution a priori continue, le plan d'échantillonnage minimisant  $R(c, n, N)$  a les propriétés suivantes (pour la démonstration voir l'article de Hald - 1968 b)

$$c = ap_r - \frac{1}{2} + p_r - q_r - p_r q_r w'(p_r)/w(p_r) + \underline{Q}(n^{-1}) \quad (27)$$

$$n = \lambda_1 \sqrt{N} + \lambda_2 + \underline{Q}(N^{-1/2}), \quad (28)$$

où  $\lambda_1^2 = p_r q_r w(p_r)/2k$  ;  $\lambda_2$  dépend de  $w(p_r)$ ,  $w'(p_r)$  et  $w''(p_r)$ .

$$\min_{(c, n)} R = k(2n - \lambda_1^2 - \lambda_2 + \underline{Q}(n^{-1})),$$

où  $n$  est la fonction de  $N$  donnée par (28).

$$d_1(c, n) = \frac{k\lambda_1}{2\sqrt{N}} (1 + \underline{O}(N^{-1/2})) = d_2(c, n) (1 + \underline{O}(N^{-1/2})) \quad (29)$$

$$w_1 Q_1 = (k\lambda_1 w(p_r)/\pi)^{1/2} N^{1/4} (1 + \underline{O}(N^{-1/4})) = w_2 P_2 (1 + \underline{O}(N^{-1/4})) \quad (30)$$

Ce résultat dépend étroitement de la structure linéaire et symétrique des pertes ; si les pertes sont  $\underline{l}_1(p_r - p)$  et  $\underline{l}_2(p - p_r)$  respectivement, avec  $\underline{l}_1 \neq \underline{l}_2$ , alors la relation entre  $c$  et  $n$  n'est plus linéaire mais contient un terme d'ordre  $\sqrt{N}$ . Les résultats pour des fonctions de pertes non linéaires et asymétriques ont été donnés par Hald (1967 b).

Le théorème montre que  $n$  est asymptotiquement proportionnel à  $\sqrt{N}$ , et que la perte moyenne par lot due à la décision est de même ordre de grandeur que les coûts d'échantillonnage, soit  $k\lambda_1 \sqrt{N}$ .

Les deux risques,  $Q_1$  et  $P_2$  sont proportionnels à  $N^{-1/4}$  (c'est-à-dire à  $n^{-1/2}$ ) et  $Q_1/P_2 \rightarrow w_2/w_1$ .

La structure de la solution est la même que celle que l'on obtient en résolvant les deux équations  $Q_1 = \alpha$  et  $P_2 = \beta$  dans le cas particulier où  $\alpha/\beta = w_2/w_1$ . Les deux solutions se confondent asymptotiquement si  $\beta$  est choisi égal à  $\beta_0 N^{1/4}$ , avec  $\beta_0$  déterminé par (30).

La décroissance des deux risques dans le cas discret et dans le cas continu est illustrée dans la table ci-après, où le risque pour un lot de 100 unités a été choisi égal à 10 %.

N	$\alpha_1 N^{-1}$	$\alpha_2 N^{-1/4}$
$10^2$	10	10
$10^3$	1	5.6
$10^4$	0.1	3.2
$10^5$	0.01	1.8
$10^6$	0.001	1.0

L'efficacité d'un plan d'échantillonnage non optimum  $(c_1, n_1)$  peut être exprimée par le rapport :

$$e(c_1, n_1, N) = R(c_0, n_0, N)/R(c_1, n_1, N), \quad 0 \leq e \leq 1, \quad (31)$$

où  $R_0(N) = R(c_0, n_0, N) = \min R(c, n, N)$ . Il a été démontré (voir Hald 1964) qu'il suffit que la relation correcte  $(c, n)$  entre  $c$  et  $n$  soit satisfaite, soit  $c = np_0 + \underline{O}(1)$  et  $c = np_r + \underline{O}(1)$ , pour que l'efficacité soit élevée même si  $n_1$  diffère considérablement de  $n_0$ . Dans le cas continu, on a asymptotiquement  $e \sim 2/(n_1/n_0 + n_0/n_1)$ , de sorte que  $e > 0,8$  si  $0,5 < n_1/n_0 < 2$ .

La distribution a priori est toujours inconnue, mais en pratique, il est souvent possible de choisir une distribution a priori hypothétique basée sur l'expérience antérieure sur les variations de qualité. Cette hypothèse peut ensuite être améliorée de proche en proche, sur la base des résultats du contrôle et autres informations sur le processus de fabrication. Les résultats ci-dessus concernant l'efficacité de plans non optimaux montrent que

la fonction de regret est peu sensible à des changements d'importance modérée de la distribution a priori. Des résultats numériques plus détaillés et des exemples ont été donnés par Hald (1967 a, 1968 b).

#### 4 - PLANS D'ECHANTILLONNAGE DE STRUCTURE BAYESIENNE DEFINIS PAR LA FONCTION CARACTERISTIQUE

La solution bayésienne qui vient d'être présentée repose sur les hypothèses suivantes : (1)  $W(p)$  est connu, (2)  $W(p)$  ne change pas avec le temps, (3)  $W(p)$  est indépendant de  $N$ , (4) les valeurs successives de  $p$  sont indépendantes (pas de corrélation sériale), (5) les fonctions de coûts sont linéaires, et (6)  $k$  et  $p_r$  sont connus.

De plus, la solution bayésienne présente l'inconvénient, du point de vue pratique, de ne pas conduire à un plan d'échantillonnage pour les petits lots, et il en résulte qu'une vérification progressive des hypothèses sur la distribution a priori peut faire défaut. Si la qualité se détériore et qu'il n'y a pas de contrôle, de lourdes pertes peuvent être subies avant que le changement soit mis en évidence.

C'est pour l'ensemble de ces raisons que la solution bayésienne n'a pas été beaucoup utilisée dans la pratique. On a préféré construire des systèmes de plans d'échantillonnage, comme par exemple les MIL STD 105, basés sur des règles empiriques telles que (1) l'effectif de l'échantillon est une certaine fonction (arbitraire) de l'effectif de lot, et (2) le risque de rejeter des lots de qualité acceptable doit être petit, et il doit décroître en fonction de l'effectif du lot. Pour la construction de ces systèmes, et l'appréciation de leur valeur, les courbes d'efficacité ont une importance fondamentale.

Nous allons maintenant essayer d'unifier les deux approches. Nous le ferons en transférant les propriétés asymptotiques des plans d'échantillonnage bayésiens aux conditions imposées aux fonctions caractéristiques. Le système de plans d'échantillonnage sera ainsi défini en spécifiant certaines propriétés des courbes caractéristiques. Si, par ailleurs, les hypothèses qui sont à la base de la solution bayésienne sont remplies, alors le système devrait avoir une efficacité asymptotiquement égale à 1 pour un choix convenable des paramètres.

Pour une distribution a priori à deux points, on définit le système au moyen des équations  $Q(p_1) = \alpha_0/N$  et  $P(p_2) = \beta_0/N$ . On déduit de (24) et (25) comment  $\alpha_0$  et  $\beta_0$  devraient être choisis sous conditions bayésiennes.

Pour une distribution a priori continue, on pourrait de même poser  $P(p_2) = \beta_0 \exp[-\beta_1 \sqrt{N}]$ , mais il pourrait être assez difficile de choisir  $p_2$ . Nous prendrons plutôt des risques fixés et nous considérerons  $(p_1, p_2)$  comme fonction de  $N$ .

A partir de (27) et (28) on obtient la structure de la solution bayésienne, soit  $c = np_r + \underline{0}(1)$  and  $n = \lambda_1 \sqrt{N} + \underline{0}(1)$ .

Les deux équations  $Q(p_1) = \alpha$  et  $P(p_2) = \beta$ , où  $\alpha$  et  $\beta$  sont des constantes (arbitrairement choisies) conduisent à :

$$c = np_1 + u_{1-\alpha} \sqrt{np_1 q_1} + \underline{0}(1) = np_2 - u_{1-\beta} \sqrt{np_2 q_2} + \underline{0}(1) = np_r + \underline{0}(1),$$

voir (8) et (9). Résolvant en  $(p_1, p_2)$  on obtient :

$$p_1 = p_r - u_{1-\alpha} \sqrt{p_r q_r / n} + \underline{0}(n^{-1}) \text{ and } p_2 = p_r + u_{1-\beta} \sqrt{p_r q_r / n} + \underline{0}(n^{-1}),$$

et en substituant  $n = \lambda_1 \sqrt{N} + \underline{0}(1)$  on a :

$$p_1 = p_r - u_{1-\alpha} \sqrt{p_r q_r / \lambda_1} N^{-1/4} + \underline{0}(N^{-1/2}) \quad (32)$$

et de même pour  $p_2$ . Ce résultat montre que si  $(p_1, p_2)$  sont choisis comme

$$p_{1N} = p_r - k_1 N^{-1/4} \text{ et } p_{2N} = p_r + k_2 N^{-1/4}$$

alors le système de plans d'échantillonnage satisfaisant aux équations  $Q(p_{1N}) = \alpha$  et  $P(p_{2N}) = \beta$ , aura la même structure que le système bayésien. Ainsi, les points correspondant au risque du fournisseur et au risque du client, convergent vers  $p_r$  lorsque l'effectif des lots augmente.

Pour déterminer le système, on doit d'abord choisir un effectif de lot,  $N_0$ , et fixer deux points de la courbe caractéristique pour cet effectif, soit

$$(p_{1N_0}, 1 - \alpha) \text{ et } (p_{2N_0}, \beta)$$

à partir de  $(p_{1N_0}, p_{2N_0})$  on trouve ensuite  $(k_1, k_2)$  ce qui conduit à :

$$p_{1N} = p_r - (p_r - p_{1N_0}) (N_0/N)^{1/4} \quad (33)$$

$$p_{2N} = p_r + (p_{2N_0} - p_r) (N_0/N)^{1/4} \quad (34)$$

Un système particulièrement simple s'obtient à partir des équations

$$p(p_r) = 1/2 \text{ soit } c = np_r - (2 - p_r)/3 + \underline{0}(n^{-1}) \text{ et } P(p_{2N}) = \beta$$

En introduisant d'autres taux de convergence, on peut généraliser le système. En remplaçant  $N^{-1/4}$  par  $N^{-1/3}$ , on obtient un système qui a les mêmes propriétés que la solution minimax.

Nous allons donner un exemple. Supposons que pour  $p_r = 0,05$  et  $N_0 = 1\,000$  on considère le système  $(c, n) = (4, 90)$  comme raisonnable. Résolvant les équations  $B(c, n, p_1) = 0,95$  et  $B(c, n, p_2) = 0,1$  on obtient  $p_{1N_0} = 0,02214$  et  $p_{2N_0} = 0,08687$ . Pour construire un système cohérent de plans d'échantillonnage pour d'autres effectifs de lot, on calcule

$$k_1 = (p_r - p_{1N_0}) N_0^{1/4} = 0.1567$$

$$k_2 = (p_{2N_0} - p_r) N_0^{1/4} = 0.2073$$

et on définit

$$p_{1N} = 0.05 - 0.1567 N^{-1/4}$$

$$p_{2N} = 0.05 + 0.2073 N^{-1/4}$$

Le plan d'échantillonnage pour un lot d'effectif  $N$  est ensuite obtenu en résolvant les équations  $B(c, n, p_{1N}) = 0,95$  et  $B(c, n, p_{2N}) = 0,10$ . (En raison du caractère discret de  $c$ , on obtient un intervalle pour  $n$ , corres-

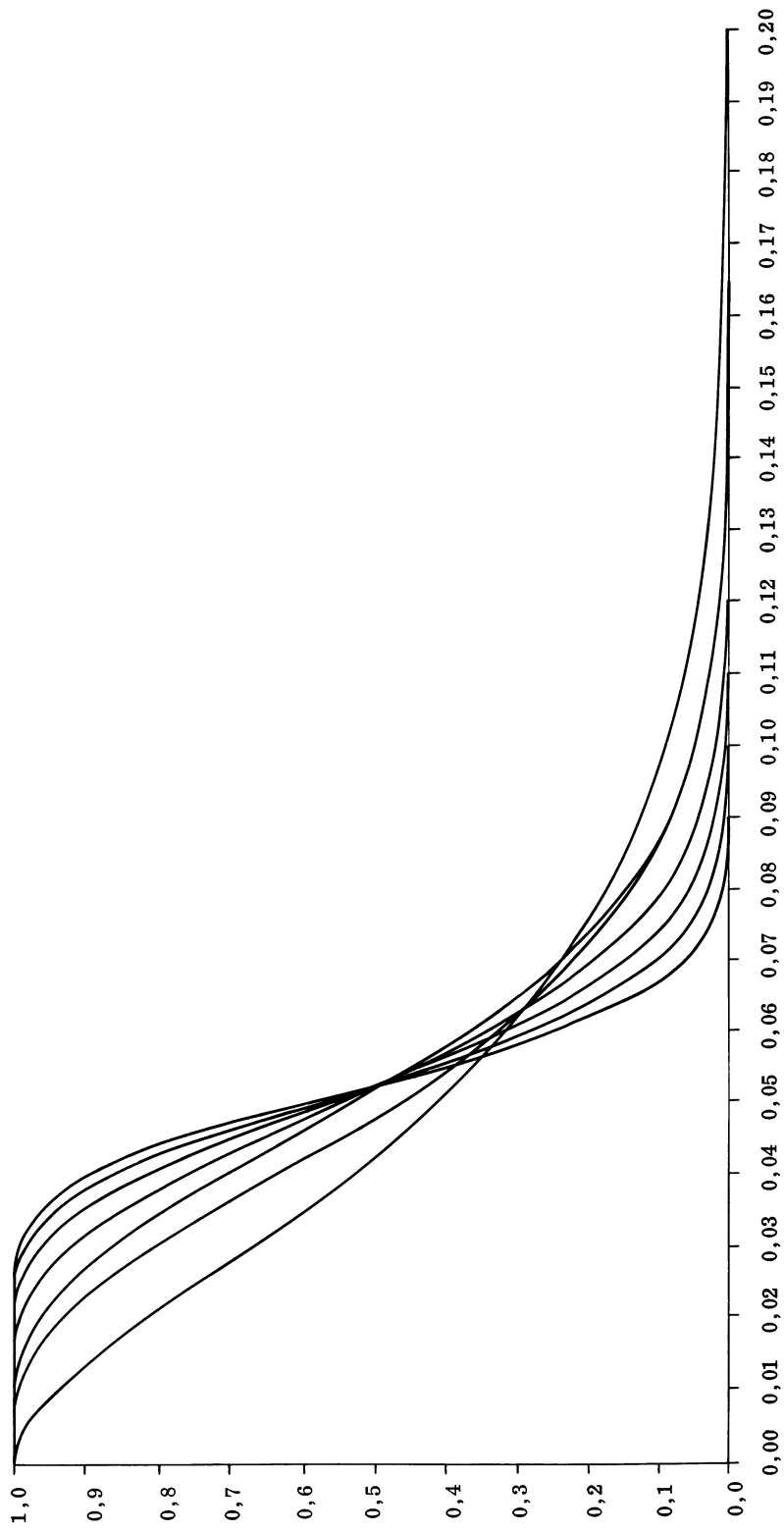


Figure 2

pendant à chaque valeur entière  $c$ , voir Hald (1967 c). On a ici choisi pour  $n$  le milieu de cet intervalle). La solution est donnée dans la table ci-après pour 7 valeurs de  $N$ . Les courbes caractéristiques sont données sur la figure 2. La solution bayésienne correspondante pour  $w(p) = 19q^{18}$  a été discutée par Hald (1964).

Plans d'échantillonnage satisfaisant aux équations

$$Q(p_{1N}) = 0.05 \text{ and } P(p_{2N}) = 0.10$$

No.	N	$100p_{1N}$	$100p_{2N}$	n	c	c
1	200	0.8333	10.51	39	1	1
2	500	1.686	9.384	76	3	3
3	1.000	2.214	8.687	90	4	4
4	2.000	2.657	8.099	147	7	7
5	5.000	3.137	7.465	221	11	11
6	10.000	3.433	7.073	316	16	16
7	20.000	3.682	6.743	450	23	23

ACKNOWLEDGEMENTS

My thanks are due to Mr. P. Thyregod for discussions during the preparation of the paper.

REFERENCES

- A. HALD (1960) - The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs. *Technometrics*, 2, 275 - 340, Discussion, 370 - 372.
- A. HALD (1964) - Efficiency of sampling inspection plans for attributes. *Bull. Int. Statist. Inst.*, 40, 681 - 697.
- A. HALD (1967a) - On the theory of single sampling inspection by attributes based on two quality levels. *Rev. Inst. Internat. Statist.*, 35, 1 - 29.
- A. HALD (1967b) - Asymptotic properties of Bayesian single sampling plans. *J. Roy. Statist. Soc. Ser. B*, 29, 169 - 173 and 586.
- A. HALD (1967c) - The determination of single sampling attribute plans with given producer's and consumer's risk. *Technometrics*, 9, 401-415.
- A. HALD (1968a) - The mixed binomial distribution and the posterior distribution of  $p$  for a continuous prior distribution. *J. Roy. Statist. Soc. Ser. B*, 30.
- A. HALD (1968b) - Bayesian single sampling attribute plans for continuous prior distributions. *Technometrics*, 10.