REVUE DE STATISTIQUE APPLIQUÉE

R. THEODORESCU

Remarques sur les plans d'échantillonnages optimaux

Revue de statistique appliquée, tome 16, nº 4 (1968), p. 35-40
<http://www.numdam.org/item?id=RSA_1968__16_4_35_0>

REMARQUES SUR LES PLANS D'ÉCHANTILLONNAGES OPTIMAUX

R. THEODORESCU

Université Laval, Faculté des Sciences
Département de mathématiques, Québec

Dans les dernières années, des plans d'échantillonnage ont été déterminés à l'aide de considérations économiques qui conduisent à la perte minimale (voir van der Waerden (1960) et Stange (1964) et (1966)). Il s'agit en fait de faire intervenir le principe minimax.

Dans ce qui suit nous nous proposons de présenter deux façons de construire les jeux qui caractérisent les plans d'échantillonnage. Pour cela nous avons utilisé essentiellement quelques remarques antérieures (voir Penkov et Theodorescu (1967)) ainsi que quelques résultats de Stange (1964) et de Basler (1968). Les constructions indiquées ici s'insèrent dans la théorie générale des jeux, et sont susceptibles d'être utilisées dans la pratique.

1. La théorie des plans d'échantillonnage peut être abordée à l'aide de la théorie des jeux dits statistiques, où l'un des joueurs est "la nature". Il y a plusieurs possibilités ; on peut imaginer un jeu entre la nature et le producteur, un jeu entre la nature et l'acheteur et, enfin, un jeu entre la nature et la coalition producteur-acheteur. Du point de vue mathématique, ces jeux ont la même forme, seuls les paramètres qui interviennent ont des significations différentes. Dans ce qui suit nous examinons un de ces trois jeux, sans préciser lequel ; cela signifie que le plan d'échantillonnage obtenu pourrait être utilisé pour des caractères mesurables ainsi que pour des caractères attributifs, pour le contrôle en cours de fabrication ainsi que pour le contrôle final.

Considérons un lot d'effectif total N. Soit l'intervalle $[0, 1]$ l'ensemble des stratégies pures p de la nature, où p est la proportion de pièces mauvaises dans le lot, et $X = \{x_A, x_R\}$ l'ensemble des stratégies pures du "statisticien" (c'est-à-dire un des joueurs qui intervient dans l'un des trois jeux possibles), où la stratégie x_A revient à accepter le lot et la stratégie x_R à rejeter le lot. Sur le produit cartésien $[0, 1] \times X$ nous définissons une fonction de perte L telle que

$$L(p, x) = \begin{cases} L(p, x_A) & \text{pour} \quad x = x_A , \\ L(p, x_R) & \text{pour} \quad x = x_R . \end{cases}$$

Soit donc

$$L(p, x_A) = Nap + F ,$$

où Np est la partie des pièces mauvaises qui sont retournées au statisticien et qui produisent une perte a par pièce et F le prix fixe du contrôle pour le lot ;

$$L(p, x_R) = Npb + Ne + F ,$$

où b (b < a) est la perte produite par une pièce mauvaise découverte avant la livraison et le prix de contrôle par pièce dans un contrôle 100 %.

Pour ce jeu le principe du minimax est satisfait, c'est-à-dire

$$\max_p \min_x L(p , x) = \min_x \max_p L(p , x),$$

d'où il s'ensuit que ce jeu possède un point-selle. Si nous notons par p_o le point d'intersection des droites

$$l_A : y = Nap + F ,$$

$$l_R : y = Nbp + Ne + F ,$$

nous obtenons

$$p_o = \frac{e}{a - b} = \frac{E}{A - B}$$

avec A = Na, B = Nb et E = eN. Si $p_o \in [0, 1]$, le point-selle est $(1 , x_R)$ et si $p_o > 1$, le point-selle est $(1 , x_A)$. Ces deux solutions représentent des cas extrêmes qui ne sont pas compatibles avec les principes généraux qui dirigent une fabrication. De toute façon, il est naturel de supposer $p_o \in [0, 1]$; dans ce cas, si on connait la stratégie pure de la nature p, la stratégie optimale du statisticien serait x_A si $p \leqslant p_o$ et x_R si $p \geqslant p_o$. Pour $p = p_o$, ce que le statisticien fait est indifférent.

La perte minimale inévitable sera

$$L_{min}(p) = \min (Nap + F, Nbp + Ne + F) = \begin{cases} Nap + F & \text{pour } p \leqslant p_o , \\ Nbp + Ne + F & \text{pour } p \geqslant p_o . \end{cases} \quad (1)$$

2. Comme nous l'avons déjà remarqué, il n'est pas économique de rejeter systématiquement le lot si $p_o \in [0, 1]$. Il s'ensuit donc qu'un échantillon d'effectif $n \leqslant N$ pourrait fournir une information supplémentaire, en vue de prendre une décision plus justifiée. Soit donc ξ le caractère à contrôler et $(x_1, \ldots, x_n)$ un échantillon d'effectif n concernant ce caractère. Notons par

$$z = f(x_1, \ldots, x_n)$$

la grandeur qui sert comme variable de contrôle et soient

$$\underline{f} = \inf f(x_1, \ldots, x_n) \qquad \bar{f} = \sup f(x_1, \ldots, x_n) ,$$

la borne inférieure et la borne supérieure, respectivement, prises pour tous les échantillons $(x_1, \ldots, x_n)$ d'effectif n. Ensuite, soit $k \in [\underline{f}, \bar{f}]$ une grandeur, appelée <u>limite d'acceptation</u> ; on remarque tout de suite que k est une fonction de n, soit k = k(n). Si $z \leqslant k$ nous acceptons le lot et si $z > k$ nous rejetons le lot. Un plan d'échantillonnage sera donc défini par le couple (n, k).

Soit maintenant Z_n l'espace formé par toutes les valeurs possibles de z et pour chaque $p \in [0,1]$ définissons une probabilité $P(p, n ; A)$ sur tous les boréliens A de Z_n. Ensuite, soit d une fonction de décision mesurable définie sur Z_n et à valeurs dans X, telle que

$$d(z) = \begin{cases} x_A & \text{pour} \quad z \leqslant k, \\ x_R & \text{pour} \quad z > k, \end{cases}$$

La fonction de risque correspondante sera :

$$\begin{aligned} \rho_{n,k}(p, d) &= \int_{Z_n} L(p, d(z))\, P(p, n ; dz) \\ &= L(p, x_A) \int_{Z_n^A} P(p, n ; dz) + L(p, x_R) \int_{Z_n^R} P(p, n ; dz) \\ &= L(p, x_A)\, P(p, n ; Z_n^A) + L(p, x_R)\, P(p, n ; Z_n^R), \end{aligned}$$

où

$$Z_n^A = \{z : z \leqslant k\}, \qquad Z_n^A \cup Z_n^R = Z_n .$$

Si nous posons

$$W(p ; n, k) = P(p, n ; Z_n^A)$$

nous obtenons <u>l'efficacité</u> du plan d'échantillonnage (n, k), c'est-à-dire la probabilité d'accepter le lot. La fonction de risque devient alors

$$\rho_{n,k}(p, d) = L(p, x_A)\, W(p ; n, k) + L(p, x_R)\,(1 - W(p ; n, k)).$$

Nous avons construit le jeu statistique $([0,1], D, \rho_{n,k})$, où $D = \{d\}$ est l'ensemble des fonctions de décision qui se réduit dans notre cas à une seule fonction d. Si nous tenons compte d'une propriété connue de la théorie des jeux, on peut ajouter à l'expression $\rho_{n,k}$ n'importe quelle constante (par rapport aux stratégies qui y interviennent) sans que la solution du jeu change ; seule la valeur du jeu est affectée par cette opération. Supposons que nous ajoutions à $\rho_{n,k}$ le terme cn qui représente le prix du contrôle des éléments de l'échantillon $(x_1, \ldots, x_n)$ d'effectif n, c'est-à-dire c est le prix de contrôle d'une pièce appartenant à l'échantillon ; nous allons conserver la notation $\rho_{n,k}$ pour cette fonction de risque modifiée, donc

$$\rho_{n,k}(p, d) = L(p, x_A)\, W(p ; n, k) + L(p, x_R)\,(1 - W(p ; n, k)) + cn.$$

Nous conservons aussi la notation pour le jeu modifié correspondant, soit $([0,1], D, \rho_{n,k})$. Pour ce jeu le principe minimax est trivialement vérifié, parce que

$$\max_p \min_d \rho_{n,k}(p, d) = \min_d \max_p \rho_{n,k}(p, d) = \max_p \rho_{n,k}(p, d),$$

d'où il s'ensuit que la solution de ce jeu sera fournie par le point-selle (p_{max}, d) ; ici nous avons posé p_{max} pour la valeur de p qui réalise ce maximum. La valeur de ce jeu sera

$$v_{n,k} = \max_p \rho_{n,k}(p, d). \tag{2}$$

3. Regardons maintenant $([0,1], D, \rho_{n,k})$ comme une famille de jeux, dépendant de deux paramètres n et k. Il est donc raisonnable de choisir n et k de façon que la valeur $v_{n,k}$ soit minimale, c'est-à-dire que nous cherchons

$$v = \min_{n,k} v_{n,k} = \min_{n,k} \max_{p} \rho_{n,k}(p, d).$$

Il s'agit ici d'un simple problème de minimisation qui n'a rien à faire avec le principe minimax, au moins apparemment. Le plan d'échantillonnage optimal sera donc (n_o, k_o), où

$$v = v_{n_o,k_o} = \min_{n,k} v_{n,k} . \tag{3}$$

D'autre part, on peut imaginer à ce stade un autre jeu. Supposons que l'ensemble des stratégies pures de la nature reste égal à $[0,1]$, mais l'ensemble des stratégies pures du statisticien devient $Y = \{(n, k)\}$, c'est-à-dire l'ensemble des plans d'échantillonnage (n, k) avec la fonction de perte $\rho_{n,k}(p, d)$, qui sera notée par $L(p ; n, k)$. Donc, nous avons construit un nouveau jeu contre la nature $([0,1], Y, L)$. Nous savons d'après la théorie des jeux qu'en général

$$\max_{p} \min_{n,k} L(p ; n, k) \leqslant \min_{n,k} \max_{p} L(p ; n, k) ;$$

sous une forme équivalente, nous savons que le principe minimax n'est pas en général vérifié. Il s'ensuit que le plan d'échantillonnage (n_o, k_o), déterminé par (3), donne une solution acceptable, mais pas une solution minimax. Elle sera une solution minimax, donc optimale de ce point de vue, si

$$\max_{p} \min_{n,k} L(p ; n, k) = \min_{n,k} \max_{p} L(p ; n, k). \tag{4}$$

Sous des conditions assez générales, Basler (1968) a démontré que le principe minimax a lieu pour le <u>regret</u> $R(p ; n, k)$, défini par

$$R(p ; n, k) = L(p ; n, k) - L_{min}(p)$$

$$= \begin{cases} (A - B)(p_o - p)(1 - W(p ; n, k)) + cn & \text{pour } p \leqslant p_o , \\ (A - B)(p_o - p) W(p ; n, k) + cn & \text{pour } p \geqslant p_o , \end{cases}$$

où $L_{min}(p)$ est donné par (1), ce qui est équivalent à la relation (4). C'est seulement dans ce cas que la solution (n_o, k_o) du jeu $([0,1], Y, L)$ coïncide avec le couple qui résulte de la minimisation de la valeur $v_{n,k}$, donnée par (2).

Mais en général, quand le principe minimax n'est pas vérifié pour le jeu $([0,1], Y, L)$, ou bien nous acceptons le couple (n_o, k_o) comme le résultat d'une opération de minimisation, ou bien nous sommes obligés de passer aux stratégies mixtes. Dans ce dernier cas, nous envisageons le jeu $(\Psi, \Lambda, \bar{L})$, où Ψ représente l'ensemble les stratégies mixtes ψ de la nature, Λ l'ensemble des stratégies mixtes λ du statisticien et

$$\bar{L}(\psi, \lambda) = \int_{[0,1] \times Y} L(p, m) \psi(dp) \lambda(dm) ,$$

où $m = (n, k)$ et $L(p, m) = L(p ; n, k)$.

Dans ce cas :

$$\max_p \min_m L(p\,,\,m) \leqslant \min_\psi \max_\lambda \overline{L}(\psi\,,\,\lambda) = \overline{L}(\psi_o,\,\lambda_o) =$$
$$= \max_\lambda \min_\psi \overline{\overline{L}}(\psi\,,\,\lambda) \leqslant \min_m \max_p L(p\,,\,m)\,,$$

d'où il s'ensuit que le "plan d'échantillonnage" $(\psi_o\,,\,\lambda_o)$ sera optimal, donc meilleur que le plan d'échantillonnage (n , k) au moins du point de vue théorique. Cela revient à prendre des plans d'échantillonnage (n, k) avec des probabilités données, problème qui reste ouvert du point de vue de l'utilité pratique.

La conclusion qui s'impose, au moins pour le moment, c'est qu'on doit chercher à trouver des conditions qui assurent la vérification du principe minimax pour le jeu ([0, 1], Y, L), ou bien se restreindre à utiliser le plan d'échantillonnage qui résulte de l'opération de minimisation de la valeur $v_{n,k}$, donnée par (2).

4. Nous remarquons que nous aurions pu arriver (voir Basler (1968)) au jeu ([0, 1], Y, L) directement, en partant de la fonction de perte L (p ; n , k). Pour cela, soit

$$L_1(p\,,\,n) = Nap + F + cn$$

la perte du statisticien si le lot est accepté et

$$L_2(p\,,\,n) = Nbp + Ne + F + cn^{(1)}$$

la perte du statisticien si le lot est rejeté. La perte moyenne sera alors

$$L(p\,;\,n\,,\,k) = L_1(p\,,\,n)\,W(p\,;\,n\,,\,k) + L_2(p\,,\,n)\,(1 - W(p\,;\,n\,,\,k))\,,$$

donc la même expression que plus haut.

La première présentation du jeu ([0, 1], Ψ, L) a l'avantage d'être entièrement constructive, en mettant mieux en évidence les différents aspects du jeu qui caractérisent les plans d'échantillonnage. D'autre part, la deuxième construction de ce jeu présente l'avantage d'être plus directe et se prête à des généralisations possibles ; de même, on peut admettre aussi le cas n = 0, qui implique que l'efficacité du plan d'échantillonnage (0 , k) est

$$W(p\,;\,0\,,\,k) \equiv k\,.$$

5. Le plan d'échantillonnage (n , k) est valable pour le contrôle par mesures ainsi que pour le contrôle par attributs.

Pour le contrôle par mesures, l'efficacité s'exprime toujours à l'aide de la fonction de répartition du caractère ξ controlé. Pour le contrôle par attributs, l'espace Z_n devient $\{0, 1, \ldots, n\}$, $z = f(x_1, \ldots, x_n)$ est égale au nombre ν des pièces mauvaises dans l'échantillon $(x_1, \ldots, x_n)$ d'effectif n, la limite d'acceptation $k \in Z_n$ et

$$W(p\,;\,n\,,\,k) = \sum_{\nu=0}^{k} \binom{n}{\nu} p^\nu (1-p)^{n-\nu}$$

(voir Penkov et Theodorescu (1967)).

----

(1) On peut prendre aussi la variante (voir Basler (1968))

$$L_2(p\,,\,n) = Nbp + (N-n)e + F + cn$$

si n n'est suffisamment faible par rapport à N.

REFERENCES

BASLER H. (1968) - Bestimmung kostenoptimaler Prüfpläne mittels des Mini-Max-Prinzips. Metrika, 12, 115-154.

PENKOV B., THEODORESCU R. (1967) - Uber die spieltheoretische Behandlung von Prüfplänen. Wiss. Z. Humboldt-Univ. Berlin, Math. - Natur. Reihe, 16, 70-72.

STANGE K. (1964) - Die Berechnung wirtschaftlicher Pläne für messende Prüfung. Metrika, 8, 48-82.

STANGE K. (1966) - Ein Näherungsverfahren zur Berechnung optimaler Pläne fur messende Prüfung bei bekannten Kosten und bekannter Verteilung der Schlecht-anteile in den vorgelegten Liefermengen. Metrika, 10, 92-136.

van der WAERDEN B.L. (1960) - Sampling inspection as a minimum loss problem. Ann. Math. Statist., 31, 369-384.