MICHEL BATAILLE

Plans d'expériences séquentielles et discrimination entre modèles

Revue de statistique appliquée, tome 21, n° 1 (1973), p. 67-80

<http://www.numdam.org/item?id=RSA_1973__21_1_67_0>

PLANS D'EXPÉRIENCES SÉQUENTIELLES ET DISCRIMINATION ENTRE MODÈLES

Michel BATAILLE

SEMA
O.C.S., Management Consultants

Les problèmes de "modélisation et d'ajustements statistiques" sont en général résolus par l'analyse globale, a posteriori, d'un ensemble de résultats d'expériences.

La théorie des tests séquentiels de Wald et ensuite celle des plans optimaux d'expériences séquentielles de Chernoff ont montré tout l'intérêt des procédures séquentielles qui restent cependant mal connues et peu employées.

Les utilisateurs voient souvent dans le facteur "temps", mis en jeu par ces procédures, un obstacle à leur emploi. Cette affirmation ne semble pas être justifiée puisque d'innombrables études de modélisation statistique sont commencées avant et menées *pendant* la période d'expérimentation.

Dans cet article nous considérons le problème du choix entre deux modèles statistiques dont les paramètres sont fonction des expériences successives choisies.

Nous généraliserons la théorie des plans d'expériences séquentielles, décrirons la procédure utilisée, donnerons des méthodes et des exemples de résolution par la théorie des jeux continus et parlerons enfin des limites de validité du caractère optimal d'une telle approche.

1. INTRODUCTION

Rappelons le problème des tests séquentiels dû à Wald (cf. [11]) où il s'agit de choisir entre deux hypthèses $H_1 : \theta = \theta_1$ et $H_2 : \theta = \theta_2$ au moyen *d'une expérience unique* que l'on répète.

Cette expérience répétée donne naissance à une variable aléatoire Y dont les n premières réalisations $y_1, \ldots y_n$ permettent de calculer le logarithme du rapport de vraisemblance :

$$S_n = \sum_{i=1}^{n} \log \frac{g_1 (y_i)}{g_2 (y_i)}$$

où $g_k (y_i)$ est la densité de Y sous l'hypothèse H_k (k = 1 ou 2).

La procédure séquentielle de Wald est caractérisée par deux nombres A et B et consiste à réagir aux n premières observations $y_1, \ldots\ldots y_n$:

- en acceptant H_1 si $S_n \geqslant A$
- en rejetant H_1 si $S_n \leqslant B$
- en continuant l'expérimentation si $B < S_n < A$

Les constantes A et B sont déterminées par la minimisation d'une fonction de risque adéquate obtenue à partir des coûts suivants :

- c , coût d'une expérience ou d'une observation,
- r_1, coût de rejeter H_1 si H_1 est vraie,
- r_2, coût d'accepter H_1 si H_1 est fausse.

Ainsi la procédure décrite plus haut apparaît comme une règle de Bayes. On montre, de plus, que la procédure de Wald conduit, à risque égal, à un nombre d'observations inférieur à celui obtenu par toute autre procédure (cf. [9] ou [12]).

Chernoff a généralisé ce problème au cas où l'expérimentateur a le choix entre *un nombre fini d'expériences distinctes* $E_1, \ldots E_n$ (Cf. [4]). Il a conservé des règles d'arrêt identiques à celles de Wald et a trouvé une méthode, fondée sur la théorie des jeux, pour choisir après chaque observation l'expérience suivante.

C'est en nous inspirant des résultats de Chernoff, que nous généraliserons encore les hypothèses en supposant cette fois que les valeurs de θ : θ_1 et θ_2 *sont elles-mêmes fonction des expériences* que l'on choisit successivement.

2. DISCRIMINATION ENTRE DEUX MODELES A UNE VARIABLE DE CONTROLE

Nous supposerons toutes les variables unidimensionnelles ; soit Y une variable aléatoire représentant le résultat d'une expérience, fonction d'une variable aléatoire X, donnée (variable non contrôlée) et d'une variable de contrôle à choisir u.

Il s'agit de choisir entre les deux modèles :

$$Y = m(X, \theta_1) \qquad \text{(I)}$$

$$Y = m(X, \theta_2) \qquad \text{(II)}$$

où θ n'intervient qu'en tant que paramètre de la distribution de Y. De façon plus explicite $\theta_1 = f_1(u, \alpha)$, $\theta_2 = f_2(u, \beta)$, c'est-à-dire que la loi de Y est, dans les deux modèles, la même *à un paramètre près fonction de u* et de *constantes* α et β inconnues.

Exemple :

$$Y = \alpha + X \qquad \text{(I)}$$

$$X \sim N(0,1)$$

$$Y = \beta u + X \qquad \text{(II)}$$

Nous avons donc ici le choix entre une variable normale de moyenne α fixe et une variable normale de moyenne, fonction linéaire de u.

Le problème est maintenant le suivant : connaissant la loi de X, s'étant donné les fonctions m(.), $f_1(.)$, $f_2(.)$, déterminer de façon optimale les valeurs successives de la commande u en vue de minimiser le nombre d'expériences conduisant au choix entre (I) et (II).

Remarquons ici la différence entre le problème proposé où la commande u intervient dans l'expression de θ_1 et θ_2 et le problème classique du choix entre deux valeurs *fixes* θ_1 et θ_2 du paramètre grâce à un choix judicieux d'expériences séquentielles tels qu'il a été exposé au § I.

3. APPLICATION DE L'ANALYSE SEQUENTIELLE ET DE LA THEORIE DES JEUX CONTINUS

– Soit $g(y, \theta_i)$, c'est-à-dire $g(y, \theta_1) = g(y, f_1(u, \alpha)) = g_1(y, u, \alpha)$ pour le modèle (I)

– Soit $g(y, \theta_i)$, c'est-à-dire $g(y, \theta_2) = g(y, f_2(u, \beta) = g_2(y, u, \beta)$ pour le modèle (II)

la densité des observations (densité de Y) sur le domaine D_Y.

– Soit ξ l'ensemble des valeurs que peut prendre la commande u, ξ^* l'extension de ξ à des valeurs de u non-déterministes correspondant à une commande probabilisée.

– Soient A et B les domaines de variation de α et β, sous-ensembles de R dans le cas unidimensionnel étudié.

Posons :

$$z_1(\alpha, \beta, u) = \log \frac{g_1(y, \alpha, u)}{g_2(y, \beta, u)}, \quad z_{1i}(\alpha, \beta) = z_1(\alpha, \beta, u_i)$$

$$z_2(\alpha, \beta, u) = \log \frac{g_2(y, \beta, u)}{g_1(y, \alpha, u)}, \quad z_{2i}(\alpha, \beta) = z_2(\alpha, \beta, u_i)$$

$$I_1(\alpha, \beta, u) = \int_{D_Y} g_1(y, \alpha, u) \cdot z_1(\alpha, \beta, u)\, dy$$

$$I_2(\alpha, \beta, u) = \int_{D_Y} g_2(y, \beta, u) \cdot z_2(\alpha, \beta, u)\, dy$$

où I_1 et I_2 sont les facteurs d'information de Kullback-Leibler relatifs aux modèles (I) et (II).

Les Y_i sont supposés être mutuellement indépendants et la procédure séquentielle se déroule de la façon suivante (cf. § I) :

On choisit une valeur de commande arbitraire, u_1 ; on observe alors par l'expérience Y_1 ; de (I) et (II) on déduit α_1 et β_1 et le rapport de vraisemblance s'écrit :

$$\lambda_1 = \frac{g_1(Y_1, \alpha_1, u_1)}{g_2(Y_1, \beta_1, u_1)}$$

De manière plus générale, supposons qu'on ait réalisé n expériences, le rapport de vraisemblance est donné par :

$$\lambda_n = \frac{\prod_{i=1}^{n} g_1(Y_i, \bar{\alpha}_i, u_i)}{\prod_{i=1}^{n} g_2(Y_i, \bar{\beta}_i, u_i)}$$

où $\bar{\alpha}_i$ et $\bar{\beta}_i$ sont données par les realtions suivantes :

$$\bar{\alpha}_i = \frac{1}{i} \sum_{k=i}^{n} \alpha_k, \; \bar{\beta}_i = \frac{1}{i} \sum_{k=1}^{n} \beta_k, \; i = 1, 2, \ldots$$

et où α_k et β_k sont tirés de (I) et (II) lors de la k^e expérience.

Les suites (α_i) et (β_i) sont des estimateurs sans biais qui convergent vers α et β.

On peut alors décider d'arrêter l'expérimentation et d'opter pour le modèle (I) ou (II) ou décider de continuer celle-ci. Dans ce dernier cas, on adopte provisoirement comme étant "le plus vraisemblable" le modèle (I) si $\lambda_n > 1$, le modèle (II) si $\lambda_n < 1$.

Avant de passer au choix optimal de la commande suivante, rappelons l'expression des règles d'arrêt dans le cas où l'on décide de ne pas continuer l'expérimentation (cf. §. I.).

Ces règles sont une généralisation à notre problème de celles données par Chernoff en [4] pour les plans d'expériences séquentielles :

si (1) $\begin{cases} \lambda_n > 1 \\ \inf\limits_{\beta \in B} \sum\limits_{i=1}^{n} z_{1i}(\bar{\alpha}_n, \beta) > -\log c \end{cases}$ arrêter et opter pour le modèle (I)

si (2) $\begin{cases} \lambda_n < 1 \\ \inf\limits_{\alpha \in A} \sum\limits_{i=1}^{n} z_{2i}(\alpha, \bar{\beta}_n) > -\log c \end{cases}$ arrêter et opter pour le modèle (II)

où c est le coût unitaire d'une expérience.

La n^e expérience ayant été réalisée et ni (1) ni (2) n'étant vérifié, il convient de continuer à expérimenter, donc de choisir u_{n+1}.

Pour ce faire, il faut en appliquant la théorie des plans d'expériences séquentielles à notre problème (cf. [4], [1], [2]), résoudre le jeu :

$$\max_{u \in \xi} . \min_{\beta \in B} . \ I_1(\bar{\alpha}_n, \beta, u) \text{ si } \lambda_n > 1 \tag{3}$$

ou

$$\max_{u \in \xi} . \min_{\alpha \in A} . \ I_2(\alpha, \bar{\beta}_n, u) \text{ si } \lambda_n < 1 \tag{4}$$

où l'un des joueurs est l'expérimentateur (u), l'autre la nature (β avec $\bar{\alpha}_n$ fixé ou α avec $\bar{\beta}_n$ fixé).

La résolution des jeux (3) ou (4) correspond au désir de maximiser l'"information" – représentée par I_1 ou I_2 - provenant de l'expérience suivante dans le cas le plus défavorable, c'est-à-dire lorsque β rend I_1 minimum dans (3) ou bien lorsque α rend I_2 minimum dans (4).

4. MODES DE RESOLUTION ET POLITIQUES STATIONNAIRES

On peut s'étonner que le choix de u_{n+1} se ramène à résoudre un jeu a deux personnes et à somme nulle entre d'une part l'expérimentateur et d'autre part "la nature" considérée comme un joueur "prudent" et non pas comme un phénomène connu statistiquement.

En effet on a pris un critère de choix "pessimiste" (le maximin) justifié par le fait que d'un point de vue statistique, on ne sait rien sur α et β ; ainsi l'expérimentateur en est réduit à obtenir le maximum "d'informations" de la nature à chaque expérience comme si elle était un joueur malveillant.

L'inconvénient majeur d'autre part de cette procédure est *qu'à chaque* expérience, il faut théoriquement résoudre le jeu (3) ou (4) ce qui peut conduire à des calculs prohibitifs dans le cas de plusieurs dizaines ou centaines d'expériences ! .

Il faudrait donc essayer de trouver des politiques optimales qui soient *stationnaires* vis-à-vis de $\overline{\alpha}_n$ et $\overline{\beta}_n$ et réduisent le problème à la résolution de deux seuls jeux tout en fournissant pour chaque expérience des règles de choix simples pour la prochaine valeur de la commande.

Nous reviendrons sur ce point important.

Il s'agit à ce stade de résoudre les jeux (3) et (4) obtenus à la fin du § III, jeux à deux personnes, de somme nulle et le noyau(1) $I_1(\overline{\alpha}_n, \beta, u)$ ou $I_2(\alpha, \overline{\beta}_n, u)$

Fixons notre attention sur l'un des deux jeux, *par exemple* celui ayant pour noyau $I_1(\overline{\alpha}_n, \beta, u) = I_1(\beta, u)$ – la résolution de l'autre se fera selon le même principe.

4.1. Existence des solutions

4.1.1. ξ n'est pas un sous-ensemble compact(2) de R : on ne sait à l'heure actuelle rien conclure sur l'existence d'une politique optimale ou même ϵ – optimale(3) pour l'expérimentateur et, sauf cas très particulier, le problème est insoluble.

4.1.2. ξ est un compact de R : on peut alors affirmer qu'il existe une politique au moins ϵ-optimale pour l'expérimentateur sous certaines conditions :

– *B est un compact de R :* alors, moyennant un changement d'échelle, on a un jeu sur le carré unité et des solutions optimales pour les deux joueurs si :

. soit le noyau $I_1(\beta, u)$ est continu par rapport à (β, u),
. soit la discontinuité n'existe que pour $\beta = u$.

– *B n'est pas un compact de R mais $I_1(\beta, u)$ est continu par rapport à (β, u):* il existe alors une politique ϵ-optimale pour l'expérimentateur sans qu'on puisse rien conclure pour son adversaire (la nature) mais ce dernier point n'est guère gênant.

4.2. Calcul des solutions

Ce calcul est légèrement simplifié par rapport à la résolution complète du jeu car on ne cherche ni la valeur du jeu(4), ni la stratégie de l'adversaire de l'expérimentateur.

Nous allons examiner les méthodes de résolution dans le cas où il existe une solution au moins ϵ-optimale pour l'expérimentateur (§.4.1.2.).

---

(1) On appelle noyau d'un jeu, la fonction de gain des joueurs.

(2) Un ensemble est dit compact s'il est fermé et borné (par exemple un intervalle fermé et borné de R).

(3) Une politique est dite ϵ-optimale si elle conduit aussi près que l'on veut de l'espérance théorique maximum de gain d'un des joueurs.

(4) La valeur du jeu est l'espérance maximum de gain d'un des joueurs s'il joue de façon optimale.

4.2.1. ξ, B compacts, $I_1(\beta, u)$ continu ; on obtiendra pour certains jeux un "point selle" ou une stratégie pure non probabilisée, c'est le cas des jeux "concave-convexes" (voir référence [6] et [11]).

4.2.2. ξ, B compacts, $I_1(\beta, u)$ continu : on obtiendra pour les jeux dits "polynomiaux", "séparables" ou "en cloche" des stratégie mixtes (donc probabilisées) sur un nombre fini de valeurs (voir référence [8]).

4.2.3. ξ, B conpacts, $I_1(\beta, u)$ discontinu pour $u = \beta$: il s'agit d'un jeu "de duel" qui conduit à une solution analytique qui est une stratégie mixte sur un nombre infini de valeurs (voir références [6] [8]).

4.2.4. ξ compact, $I_1(\beta, u)$ continu : il s'agit ici des jeux pour lesquels ou bien B est compact mais qui n'ont pas été mentionnés précédemment, ou bien B n'est pas compact.

Il n'existe pas de méthode générale et il faut exploiter chaque structure particulière pour avoir une chance d'aboutir. On considère l'ensemble ξ^*, c'est-à-dire qu'on probabilise la commande u.

Soit donc une variable aléatoire U dont les réalisations appartiennent à ξ, possédant une fonction de répartition H(u) qu'il s'agit de trouver en résolvant :

$$\boxed{\max_{H \in \xi^*} \inf_{\beta \in B} E\{I_1(\beta, u)\}}$$

où E{ } est l'espérance mathématique par rapport à H(u). Nous cherchons donc à déterminer la stratégie mixte de l'expérimentateur, optimale vis-à-vis de toutes les stratégies pures de la nature (paramètre β).

Dans le cas favorable, l'opération $\inf_{\beta \in B}$ pourra se faire directement par différentiation et donnera $\beta = \beta_0(H)$, ce qui formellement veut dire qu'on aura obtenu β en fonction de certaines caractéristiques de H(u), en particulier ses moment : $E\{U\}$, $E\{U^2\}$, . . .

Il restera à résoudre $\max_{H \in \xi^*} E\{I_1(\beta_0, u)\}$ sachant que H(u) est une fonction de répartition sur ξ. Ce problème relève de l'analyse fonctionnelle et peut s'avérer fort ardu à résoudre. On notera cependant que le lemme de Neyman-Pearson peut se révéler utile (voir référence [2] [8]).

Dans le paragraphe suivant nous donnons un exemple de résolution dans le cas polynomial avec B non-compact.

5. EXEMPLE DE RESOLUTION (noyau polynomial, B non-compact)

Reprenons l'exemple du § .I. :

soit à choisir entre les modèles (I) $Y = \alpha + X$ $\qquad X \sim N(0,1)$

(II) $Y = \beta u + X$ $\qquad 0 \leqslant u \leqslant a$

Par changement de variable, on peut toujours se ramener au cas a = 1 qu'on considèrera désormais :

Nous avons :

$$\begin{cases} \theta_1 = f_1(u, \alpha) = \alpha \\ \theta_2 = f_2(u, \beta) = \beta u \\ m(X, \theta_1) = \theta_1 + X \\ m(X, \theta_2) = \theta_2 + X \\ g(y, \theta) = \dfrac{1}{\sqrt{2\Pi}} e^{-\frac{(y-\theta)^2}{2}} \\ Y \sim N(\theta, 1) \end{cases}$$

5.1.

Si après n expériences, le modèle (I) nous paraît le plus vraisemblable avec $\alpha = \bar{\alpha}_n$, pour choisir u_{n+1} de façon optimale, nous avons à résoudre le jeu :

$$\max_{H \in \xi^*} \inf_{\beta \in B} E\{I_1(\beta, u)\} \qquad \text{avec} \qquad \begin{aligned} \xi &= [0,1] \\ B &=]-\infty, +\infty[\end{aligned} \tag{5}$$

Nous avons successivement les résultats :

$$I_1(\beta, u) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-\bar{\alpha}_n)^2}{2}} \log \frac{e^{-\frac{(y-\bar{\alpha}_n)^2}{2}}}{e^{-\frac{(y-\beta u)^2}{2}}} dy = \frac{1}{2}(\beta^2 u^2 - 2\beta u \bar{\alpha}_n + \bar{\alpha}_n^2)$$

$$E\{I_1(\beta, u)\} = \frac{1}{2}(\beta^2 E(U^2) - 2\beta \bar{\alpha}_n E(U) + \bar{\alpha}_n^2)$$

$$\inf_{\beta \in B} E\{I_1(\beta, u)\} = \frac{1}{2} \bar{\alpha}_n^2 \frac{\text{var}(U)}{E(U^2)} \quad \text{pour} \quad \beta = \frac{E(U)}{E(U^2)} \bar{\alpha}_n = \beta_o$$

en supposant :

$$E(U^2) \neq 0. \tag{6}$$

Il reste à résoudre :

$$\max_{H \in \xi^*} \frac{1}{2} \bar{\alpha}_n^2 \frac{\text{var}(U)}{E(U^2)} \tag{7}$$

L'expression (7) est maximum pour var (U) = $E(U^2)$, donc pour $E(U) = 0$ et (5) vaut alors $\frac{1}{2} \bar{\alpha}_n^2$.

Comme d'autre part U est une variable aléatoire positive, $E(U) = 0 \Rightarrow E(U^2) = 0$ et U est presque certainement nulle, ce qui est contraire à l'hypothèse (6).

Nous voyons alors qu'il n'y a pas de solution optimale en U mais seulement une solution ϵ-optimale, ce qui est normal puisque B n'est pas compact.

Nous pouvons approcher la valeur du jeu $\frac{1}{2}\bar{\alpha}_n^2$ à ϵ près en choisissant E(U) très proche de 0, donc $E(U^2)$ très proche de zéro mais différent de 0.

La stratégie optimale consiste donc à jouer u = 0 "presque toujours" et, par exemple, u = 1 le reste du temps.

Remarques :

– La solution ϵ-optimale obtenue est stationnaire car elle ne dépend pas de $\bar{\alpha}_n$; elle est donc très simple à mettre en oeuvre : à chaque fois que le modèle (I) semble l'emporter et qu'on décide de continuer à expérimenter, on choisit u = 0 avec une probabilité égale à 99 % par exemple et u = 1 avec une probabilité de 1 %.

– L'interprétation physique de la solution est la suivante : si $\alpha \neq 0$, u = 0 est évidemment une bonne valeur pour différencier (I) de (II) par contre, dans le cas où $\alpha = 0$, u = 0 ne permet plus de conclure d'où la nécessité de jouer u = 1 de temps en temps.

5.2.

Si après n expériences, le modèle (II) nous paraît le plus vraisemblable avec $\beta = \bar{\beta}_n$, pour choisir u_{n+1} de façon optimale nous avons à résoudre le jeu (en posant $I_2(\alpha, \bar{\beta}_n, u) = I_2(\alpha, u)$) :

$$\max_{H \in \xi^*} \inf_{\alpha \in A} E\{I_2(\alpha, u)\} \qquad \text{avec} \qquad \begin{aligned} \xi &= [0,1] \\ A &=]-\infty + \infty[\end{aligned} \tag{8}$$

Nous avons successivement les résultats :

$$I_2(\alpha, u) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{+\infty} e^{-\frac{(y-\bar{\beta}_n u)^2}{2}} \log \frac{e^{-\frac{(y-\bar{\beta}_n)^2}{2}}}{e^{-\frac{(y-\alpha)^2}{2}}} dy = \frac{1}{2}(\alpha^2 - 2\alpha u\bar{\beta}_n + \bar{\beta}_n^2 u^2)$$

$$E\{I_2(\alpha, u)\} = \frac{1}{2}(\alpha^2 - 2\alpha\bar{\beta}_n E(U) + \bar{\beta}_n^2 E(U^2))$$

$$\inf_{\alpha \in A} E\{I_2(\alpha, u)\} = \frac{1}{2}\bar{\beta}_n^2 \text{ var } (U) \text{ pour } \alpha = E(U)\bar{\beta}_n = \alpha_o$$

Il reste à résoudre :

$$\max_{H \in \xi^*} \frac{1}{2}\bar{\beta}_n^2 \text{ var } (U) \tag{9}$$

Comme U a ses réalisations sur [0,1], (9) conduit immédiatement à la solution :

$$\begin{cases} u = 0 \text{ avec probabilité } 1/2 \\ u = 1 \text{ avec probabilité } 1/2 \end{cases}$$

Remarque :

– la solution est encore *stationnaire* et l'application du résultat est particulièrement simple puisqu'il suffit de jouer u = 0 et u = 1 à pile ou face dès que le modèle (II) semble l'emporter et qu'on décide de continuer l'expérimentation.

Applications :

Nous allons décrire un exemple fictif d'application d'ordre médical qui illustre le calcul précédent.

Supposons qu'un médecin (l'expérimentateur) veuille tester l'influence d'un médicament sur le taux de glucose du sang d'un patient et admettons que, *sans médicament,* ce taux, ramené à sa moyenne, soit une variable normale de moyenne 0 et de variance 1 (variable X).

La variable d'expérimentation dont dispose le médecin est la quantité du médicament qu'il décide d'administrer : u, comprise entre 0 et a.

Après chaque prise du médicament, un test sanguin met en évidence un taux de glucose Y (variable aléatoire). Le médecin veut déterminer aussi rapidement que possible si le médicament augmente le taux moyen de glucose d'une quantité constante quelle que soit la dose prise ou bien si ce taux est proportionnel à la dose administrée.

Il a donc le choix entre les deux modèles décrits précédemment :

$$\text{(I)} \quad Y = \alpha + X$$

$$X \sim N(0,1)$$

$$\text{(II)} \quad Y = \beta u + X$$

et il expérimentera de la façon suivante :

– si le calcul du rapport de vraisemblance montre que (I) est plus vraisemblable, il n'administrera aucun médicament à l'expérience suivante avec une probabilité de 90 %, par exemple,

– si au contraire (II) semble plus vraisemblable, il décidera à pile ou face s'il doit donner la quantité maximale (a) ou rien du tout,

– lorsqu'une des conditions d'arrêt citées au § III. s'appliquera, il arrêtera l'expérimentation et optera pour le modèle le plus vraisemblable.

On peut également, par exemple, imaginer qu'on pratique le "rattrapage de jeu" sur un tour et qu'on cherche à savoir si ce rattrapage influe de façon constante ou linéaire sur une dimension des pièces produites par ce tour.

6. GENERALISATION

On pourrait par exemple généraliser le modèle $Y = m(X,\theta)$ dans le sens de la multiplication des variables aléatoires non contrôlées en faisant de X un vecteur de R^k avec une distribution jointe à k dimensions et même en faisant du résultat de l'expérience Y un vecteur de R^l, la distribution $g(y, \theta)$ devenant alors l-dimensionnelle : $g(y_1, y_2, \ldots . y_l, \theta)$. Il nous semble qu'un tel développement, fort utile dans la pratique où il peut y avoir de nombreux facteurs non contrôlés et plusieurs résultats d'une même expérience, n'apporterait ici rien d'un point de vue méthodologique et ne ferait que compliquer inutilement les notations.

Il nous semble par contre beaucoup plus intéressant d'introduire un *vecteur de contrôle* $u \in R^n$ et les paramètres correspondants $\alpha, \beta \in R^n$. La fonction m(.) ne change pas et $f_1(.)$, $f_2(.)$ deviennent des fonctions de $(u_1, \ldots . u_n, \alpha_1, \ldots, \alpha_n)$ et $(u_1, \ldots u_n, \beta_1, \ldots, \beta_n)$ respectivement.

Que devient la procédure du plan d'expérience ?

Pour déterminer un ensemble de valeurs des vecteurs α et β, il convient de réaliser n expériences avec la même valeur u_1 de commande. On observe alors les $Y_1, \ldots, Y_n$ successifs et on obtient un système de 2n équations aux inconnues vectorielles α et β. Nous appelerons "expérience groupée" un ensemble de n expériences consécutives. Une telle "expérience groupée" est équivalente à une expérience ordinaire dans le cas de la commande unidimensionnelle (cf. §.III) puisqu'elle permet de déterminer les premières valeurs des vecteurs α, β qu'on appelera α_1, β_1.

La méthode de test exposée en III reste valable à condition de remplacer :

$$\lambda_1 \text{ par } \lambda'_1 = \frac{\prod_{j=1}^{n} g_1(y_j, \overline{\alpha}_1 u_1)}{\prod_{j=1}^{n} g_2(y_j, \overline{\beta}_1, u_1)} \text{ et après m expèriences :}$$

$$\lambda_m \text{ par } \lambda'_m = \prod_{i=1}^{n} G(u_i) \text{ où } G(u_i) = \frac{\prod_{j=(i-1)n+1}^{ni} g_1(y_j, \overline{\alpha}_i, u_i)}{\prod_{j=(i-1)n+1}^{ni} g_2(y_j, \overline{\beta}_i, u_i)}$$

Remarquons qu'on a supposé que les Y_j successifs sont mutuellement independants et donc que la densité des n résultats de n expériences consécutives (ou d'une "expérience groupée) est le produit des n densités $g_i(y_1, \overline{\alpha}_k, u_k) \times \times g_i(y_n, \overline{\alpha}_k, u_k)$ pour les mêmes valeurs du vecteur de commande et des vecteurs paramètres $\overline{\alpha}_k$, $\overline{\beta}_k$.

Une telle hypothèse paraît raisonnable à moins que les expériences aient une certaine "hystérésis" ou mémoire auquel cas il conviendrait de se donner ou de trouver la loi jointe de n Y_j consécutifs à commande fixée.

De toutes façons, toute l'analyse séquentielle repose sur l'indépendance des résultats des expériences donc, même dans le cas où les Y_j d'une "expèrience groupée" ne sont pas indépendants, il faut au moins supposer que les "expériences groupées" sont, elles mutuellement indépendances.

Comme précédemment, nous sommes conduits à la résolution des jeux :

$$\max_{u \in \xi} \ \min_{\beta \in B} \ I_1(\overline{\alpha}_n, \beta, u)$$

et

$$\max_{u \in \xi} \ \min_{\alpha \in A} \ I_2(\alpha, \overline{\beta}_n, u)$$

où α, β, $\overline{\alpha}_n$, $\overline{\beta}_n$, $u \in R^n$.

et ξ, A, B sont des sous-ensembles de R^n.

On voit alors apparaître les grandes difficultés que va poser la résolution de ce jeu à deux personnes et somme nulle où les stratégies sont n-dimensionnelles.

L'existence des solutions a été discutée en 3.1. et reste valable ainsi que le calcul théorique mais, dans ce cas, il faut introduire une fonction jointe de répartition H(u) définie sur $E \subset R^n$ ce qui complique en général singulièrement les calculs.

7. LIMITES DE VALIDITE ET CONCLUSIONS

7.1. Optimalité

Nous avons vu au § I que la procédure de Wald est une règle de Bayes et qu'en conséquence elle est optimale *vis-à-vis de la fonction de coût* que nous avons considérée (c, r_1, r_2).

Moyennant certaines hypothèses, on peut montrer que la procédure décrite au § III qui est caractérisée par les règles d'arrêt et par la méthode de choix de l'expérience suivante, possède un certain caractère d'optimalité – dite "asymptotique" – *vis-à-vis d'une procédure quelconque.*

En particulier, Chernoff a montré (cf. [4]) que si l'on suppose que le coût de chaque observation c est très petit (par exemple par rapport à 1), les risques (α ou β) de rejeter la bonne hypothèse lorsqu'on décide de s'arrêter deviennent eux-mêmes très petits si on utilise la procédure proposée ici.

Autrement dit la procédure envisagée réalise une sorte d'équilibre optimum entre le coût d'expérimentation et les coûts d'erreur (r_1, r_2).

En effet si par une autre procédure, on s'arrête après un nombre d'observations plus grand, le coût d'expérimentation croît plus vite que les risques d'erreur ne diminuent ; si par contre on décide de prendre moins d'observations par une autre procédure, le coût d'expérimentation décroît d'une quantité très petite alors que les risques d'erreur deviennent finis.

La procédure proposée est donc "asymptotiquement optimale" en ce sens qu'elle conduit au nombre d'observations minimum pour que les risques d'erreur deviennent des infiniments petits du même ordre que c.

La méthode possède deux inconvénients : d'une part elle implique un nombre d'observations élevé (c très petit), d'autre part le critère du rapport de vraisemblance peut, *au début de l'expérimentation* conduire à de mauvaises valeurs de la commande. On a essayé de pallier ces défauts de deux façons :

– en se limitant à un échantillon de taille moyenne ("medium sample size" ou MSS), voir par exemple les références [5] et [10],

– en choisissant mieux les expériences *au début* (voir référence [3]) pour rendre la procédure plus performante.

Ces tentatives dans deux directions différentes n'ont pas été couronnées de succès.

7.2. Conclusion

Si les conditions suivantes sont réunies :

- faible coût unitaire de l'expérience,
- indépendance des résultats successifs,
- possibilité de résoudre les jeux,
- obtention d'une solution *stationnaire,*

alors la procédure séquentielle exposée dans cet article pour le choix des commandes successives se révèlera, à risque égal, beaucoup plus rapide qu'une procédure classique. Rappelons à cet égard qu'une procédure séquentielle, dans le cas du test de la moyenne d'une loi normale par exemple, demande, pour le même risque, au *moins* 47 % de moins d'observations qu'avec une procédure classique.

Il serait donc souhaitable de voir se développer l'emploi des procédures séquentielles qui sont, dans les cas favorables, très efficaces et simples à mettre en oeuvre mais, malheureusement, trop mal connues.

BIBLIOGRAPHIE

[1] ARMITAGE P. – Sequential medical trials. Blackwell Scientific Publications, Oxford, 1962.

[2] BLACKWELL D., GIRSHICK M.A. – Theory of games and statistical decisions. John Wiley, New-York, 1954.

[3] BOX G.E.P., HILL W.J. – Discrimination among mechanistic models. Technometrics Vol. 9, n° 1, February 1967, p. 57-71.

[4] CHERNOFF H. – Sequential design of experiments. Annals of Mathematical Statistics, Vol. 30, n° 3, September 1959.

[5] CHERNOFF H. – Technical Report n° 35, Stanford University, 1968.

[6] DRESHER M. – Jeux de stratégie. Dunod, Paris, 1965.

[7] FERGUSON T.S. – Mathematical statistics. Academic Press, New-York, 1967.

[8] KARLIN S. – Mathematical methods and theory in games, programming and economics. Pergamon Press, London, 1959.

[9] MATTHES T.K. – On the optimality of sequential probability ratio tests. Annals of Mathematical Statistics, Vol. 34, n° 1, March 1963.

[10] MEETER, PIRIE, BLOT. – Internal technical report, Florida State University, 1968.

[11] OWEN D.B. – Game theory. Saunders, Philadelphia, 1968.

[12] WALD A. – Sequential analysis. John Wiley, New-York, 1947.

[13] WETHERILL G.B. – Sequential methods in statistics. John Wiley, New-York, 1966.