REVUE DE STATISTIQUE APPLIQUÉE

MARC MOORE

Choix séquentiel des points échantillonnaux pour reconstruire un schéma ordonné

Revue de statistique appliquée, tome 27, n° 3 (1979), p. 15-32

<http://www.numdam.org/item?id=RSA_1979__27_3_15_0>

© Société française de statistique, 1979, 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/

CHOIX SÉQUENTIEL DES POINTS ÉCHANTILLONNAUX POUR RECONSTRUIRE UN SCHÉMA ORDONNÉ

Marc MOORE
Ecole Polytechnique de Montréal, Canada

RESUME

La notion de schéma ordonné sur [0, 1] et le problème de la reconstruction de tels schémas à partir de points échantillonnaux sont introduits et motivés. Le choix des points échantillonnaux optimaux pour les reconstructions bayesiennes est illustré. Devant la difficulté de déterminer ces points optimaux une approche séquentielle pour localiser les points échantillonnaux est suggérée et analysée.

I – INTRODUCTION

Depuis quelques années plusieurs travaux portant sur les schémas (formes) aléatoires sont apparus, GUILBERT [4], MILES [6, 7, 8], MATSCHINSKI [5], CRAIN et MILES [1] pour en citer quelques-uns. Un problème récemment étudié est celui de la reconstruction d'un schéma aléatoire non observable dans son entier, ceci à partir d'observations ponctuelles de ce schéma (SWITZER [10], MOORE [9]). L'exemple simple suivant illustre ce problème et aussi la source de motivation pour les considérations qui suivent.

Considérons une lisière de terrain en bordure d'une rivière et supposons qu'en cet endroit le sous-sol change de nature en s'éloignant de la rive. Plus précisément supposons qu'il y a une bande sur la berge où le sous-sol est trop mou pour supporter une construction, celle-ci étant suivie d'une bande où le sol est propice à la construction et finalement d'une région où le sous-sol est constitué de roc. Ces subdivisions ne sont pas immédiatement apparentes en surface. Etant donné un segment de droite L traversant la bande de terre en question dénotons par θ_1 et θ_2 les points de division entre les parties (figure 1).

Mots-Clés : **Schéma, Reconstruction, Points échantillonaux, Séquentiel, Bayesien.**

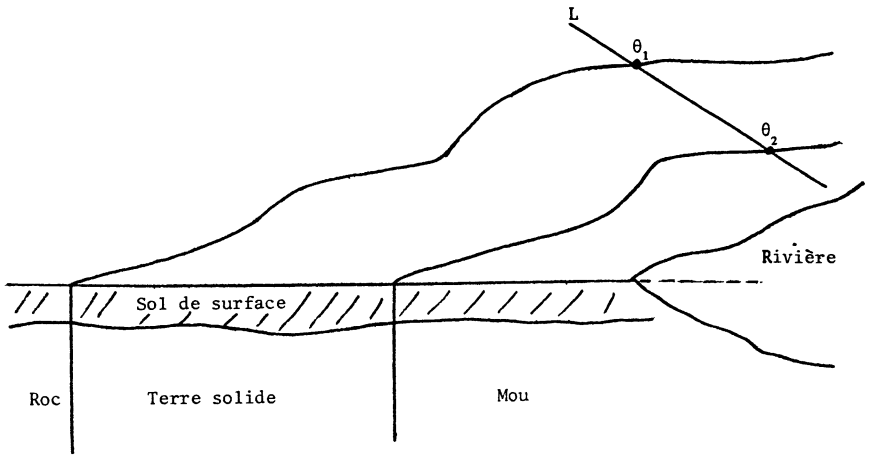

Figure 1

Le problème statistique considéré est le suivant : à partir de certains forages (points échantillonnaux) on désire localiser (faire l'estimation) θ_1 et θ_2. Faisant de même pour plusieurs segments de droite L, il serait possible d'obtenir une reconstruction de la bande de terre solide (figure 2).

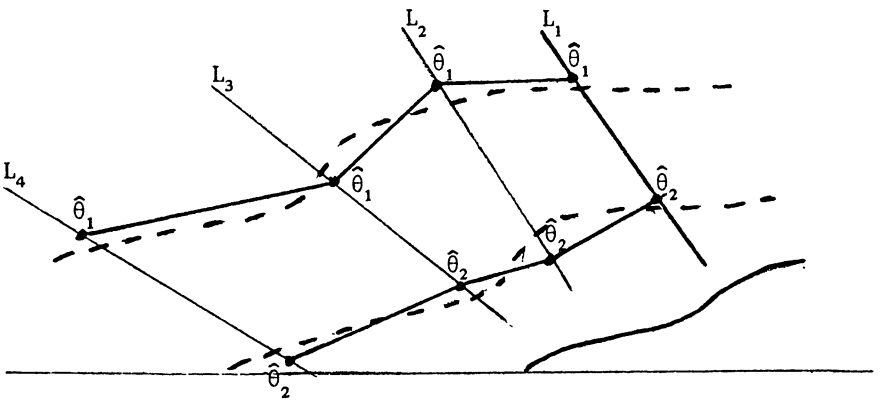

Figure 2

Le même genre de situation peut se rencontrer par exemple quand on veut déterminer le contour d'une nappe de gaz à une certaine profondeur.

Le problème illustré ici peut être abordé à l'aide de résultats obtenus antérieurement par l'auteur [9], dans ce travail nous introduisons ce qui est appelé les schémas m, n sur [0, 1], décrivons ce qu'est le problème de reconstruction d'un tel schéma, définissons les règles bayesiennes de reconstruction, obtenons une classe complète de règles de reconstruction, celle-ci étant incluse dans l'ensemble des procédures bayesiennes étendues ([3], page 50). Les règles bayesiennes de reconstruction jouissent de propriétés très désirables mais souffrent d'un défaut majeur au plan de l'utilisation : elles sont extrêmement difficiles à obtenir expli-

citement. Cette situation, due ici essentiellement à la complexité de l'espace des états de la nature, est assez fréquente en théorie de la décision (voir [2], page 102).

Dans le présent article nous voulons considérer une classe beaucoup plus restreinte de schémas et proposer une approche ne conduisant pas à des résultats théoriques aussi élégants, mais qui a l'avantage de pouvoir facilement être mise en pratique. Plus précisément à la section suivante nous définissons le type de schémas considérés, rappelons ce qu'est une reconstruction bayesienne et examinons la question du choix des points échantillonnaux. Cet examen nous conduit à suggérer un choix séquentiel de ces points et celui-ci est examiné en détail à la section 3. Finalement nous apportons quelques renseignements sur la loi du nombre d'observations requises dans l'application des procédures suggérées.

II – SCHEMAS ORDONNES ET RECONSTRUCTION BAYESIENNE

Définition : Etant donné un entier $n > 1$ et un ensemble $K = \{k_1, \ldots, k_n\}$ de n couleurs distinctes, un schéma ordonné est une partition de l'intervalle [0 , 1] en n sous-intervalles disjoints de longueur non nulle,

$$C_1 = [0\,,\theta_1]\,,\ C_2 = (\theta_1\,,\theta_2]\,,\ \ldots\,,\ C_{n-1} = (\theta_{n-2}\,,\theta_{n-1}]\,,\ C_n = (\theta_{n-1}\,,1],$$

appelés cellules, la cellule C_i étant de couleur k_i , $i = 1, \ldots, n$.

Ce que nous voulons signifier par couleur est en fait un "état caractéristique", le terme couleur n'est utilisé que pour simplifier. Ce type de schéma est qualifié d'ordonné car les couleurs y apparaissent dans un ordre prédéterminé. Il est à remarquer que chaque couleur est présente une et une seule fois. Ces conditions sont des restrictions par rapport à la notion de schéma m , n introduite en [9].

Dénotons par A la classe des schémas dont il est question ici. Il est clair que A peut être mis en correspondance biunivoque avec

$$\Theta = \{\theta = (\theta_1, \ldots, \theta_{n-1}) : 0 < \theta_1 < \theta_2 < \cdots < \theta_{n-1} < 1\}.$$

Le problème statistique étudié est le suivant : étant donné un élément inconnu de A, nous désirons à partir de r points $y_1, \ldots, y_r$ ($r \geqslant 1$) appelé points échantillonnaux, dont la couleur est observée, faire l'estimation de $\theta_1, \ldots, \theta_{n-1}$, c'est-à-dire reconstruire le schéma inconnu. Désignons par Ω l'espace échantillon correspondant aux observations effectuées. La nature des éléments de Ω dépendra du mode de sélection des points échantillonnaux (aléatoire ou non).

L'approche considérée sera celle de la théorie de la décision. Il nous faut donc spécifier une fonction de perte et une loi a priori sur A. Pour fonction de perte nous pouvons considérer celles qui sont de la forme :

$$L(a\,,b) = \ell\,[(\theta_1, \ldots, \theta_{n-1})\ ;\ (\hat{\theta}_1, \ldots, \hat{\theta}_{n-1})] = \sum_{i=1}^{n-1} \phi_i\,(|\theta_i - \hat{\theta}_i|).$$

où $(\theta_1, \ldots, \theta_{n-1})$ est l'élément de Θ associé à $a \in A$, $(\hat{\theta}_1, \ldots, \hat{\theta}_{n-1})$ est l'élément de Θ associé à $b \in A$ qui est la reconstruction de a et ϕ_i est une fonction de [0 , 1] dans $R^+ \cup \{0\}$, $i = 1, \ldots, n-1$. Dans les exemples qui suivent nous

considérons toujours le cas particulier $\phi_i(x) = x^2$, $i = 1, \ldots, n-1$. Une loi de probabilité a priori λ sur A est complètement déterminée par une mesure de probabilité sur Θ et inversement.

Une règle de reconstruction δ est une fonction de Ω dans A. Le risque correspondant à la règle δ pour reconstruire le schéma a est :

$$R(a, \delta) = \int_{\Omega} L(a, \delta(Y))\, dP_a(Y)$$

où P_a est la mesure de probabilité conditionnelle sur Ω étant donné a. Si λ est une loi a priori sur A alors le risque bayesien correspondant à δ relativement à λ est :

$$r(\lambda, \delta) = \int_{A} R(a, \delta)\, d\lambda(a).$$

Nous disons qu'une règle de reconstruction δ_0 est une règle bayesienne de reconstruction, relativement à la distribution a priori λ, si :

$$r(\lambda, \delta_0) = \inf_{\delta} r(\lambda, \delta).$$

Il est possible de démontrer (en procédant comme dans le théorème 3.1 de [9]) que pour chaque choix aléatoire des points échantillonnaux, il existe un ensemble de points échantillonnaux fixes tel que le risque bayesien correspondant à la règle bayesienne pour ces points, n'est pas plus grand que celui correspondant à la règle bayesienne obtenue dans le cas du choix aléatoire des points échantillonnaux. En conséquence, il est suffisant de considérer les règles bayesiennes de reconstruction basées sur des points échantillonnaux fixes. Dans ce cas l'espace échantillon devient simplement $\Omega = \{k(y_1), \ldots, k(y_r) : k(y_i) \in K, i = 1, \ldots, r\}$. Le problème qui se pose alors est celui du choix des points échantillonnaux optimaux, ceux-ci doivent être tels que le risque bayesien correspondant à la procédure bayesienne relativement à λ soit minimal. Le choix de ces points dépendra en général de λ est ne sera pas facile.

Voyons d'abord comment peuvent être obtenues les règles bayesiennes. La terminologie suivante sera utile : un vecteur observation $Y = (k(y_1), \ldots, k(y_r))$ est compatible avec un schéma a si dans a les points $y_1, \ldots, y_r$ ont respectivement les couleurs $k(y_1), \ldots, k(y_r)$. Nous écrirons l'expression (2.1) en termes des éléments de Θ. Dénotons par G la mesure de probabilité sur Θ correspondant à λ. Alors (2.1) s'écrit :

$$r(G, \delta) = \int_{\Theta} \sum_{Y \in \Omega} \sum_{i=1}^{n-1} \phi_i(|\theta_i - \hat{\theta}_i(Y)|)\, P_\theta(Y)\, dG(\theta)$$

$$= \sum_{Y \in \Omega} \int_{\Theta} \sum_{i=1}^{n-1} \phi_i(|\theta_i - \hat{\theta}_i(Y)|)\, dG(\theta \mid Y)\, P(Y)$$

où : $P_\theta(Y) = 1$ si Y est compatible avec le schéma correspondant à θ et 0 autrement :

$$P(Y) = \int_{\Theta} P_\theta(Y)\, dG(\theta) = \int_{P(Y)} dG(\theta)$$

est la masse de probabilité portée par $Y \in \Omega$, $P(Y) = \{\theta \in \Theta : Y$ est compatible avec le schéma correspondant à $\theta\}$;

$$G(\theta|Y) = P_\theta(Y)\, G(\theta) \Big/ \int_\Theta P_\theta(Y)\, dG(\theta)$$

est la loi a posteriori sur Θ ; $\hat{\theta}_i(Y)$ est l'estimation de θ_i obtenue en utilisant l'information Y. La procédure bayesienne est donc celle qui pour chaque $Y \in \Omega$ rend minimale :

$$\int_{P(Y)} \sum_{i=1}^{n-1} \phi_i\,(|\theta_i - \hat{\theta}_i(Y)|)\, dG(\theta) \Big/ \int_{P(Y)} dG(\theta).$$

Par exemple si $\phi_i(x) = x^2$, $i = 1, \ldots, n-1$, il faut minimiser :

$$\int_{P(Y)} \sum_{i=1}^{n-1} [\theta_i - \hat{\theta}_i(Y)]^2\, dG(\theta) \Big/ \int_{P(Y)} dG(\theta)$$

par rapport aux $\theta_i(Y)$, ceci pour chaque $Y \in \Omega$; il est facile de vérifier que les estimateurs bayesiens, $\hat{\theta}_i^0(Y)$, sont dans ce cas :

$$\hat{\theta}_i^0(Y) = \int_{P(Y)} \theta_i\, dG(\theta) \Big/ \int_{P(Y)} dG(\theta) \qquad i = 1, \ldots, n-1$$

Le problème de la recherche des points échantillonnaux optimaux se ramène donc à trouver les points $y_1, \ldots, y_r$ tels que :

$$\sum_{Y \in \Omega} \int_{P(Y)} \sum_{i=1}^{n-1} \phi_i\,(|\theta_i - \hat{\theta}_i^0(Y)|)\, dG(\theta)$$

soit minimale par rapport à $y_1, \ldots, y_r$.

Exemples :

1) Considérons le cas où $n = 2$, la loi de θ_1 est la loi uniforme sur (0,1), $\phi_1(x) = x^2$ et trois points échantillonnaux $y_1 < y_2 < y_3$ sont utilisés. Il est facile de vérifier que :

$\Omega = \{(k(y_1), k(y_2), k(y_3))\}$	δ : reconstruction bayesienne, estimateur de θ_1
(k_1, k_1, k_1)	$(1 + y_3)/2$
(k_1, k_1, k_2)	$(y_2 + y_3)/2$
(k_1, k_2, k_2)	$(y_1 + y_2)/2$
(k_2, k_2, k_2)	$y_1/2$

$$r(G, \delta) = \frac{2}{3}([(1 - y_3)/2]^3 + [(y_3 - y_2)/2]^3 + [(y_2 - y_1)/2]^3 + [y_1/2]^2)$$

de là il découle que le choix optimal des points échantillonnaux est $\left(\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\right)$.

2) Considérons le cas où n = 3, la loi de (θ_1 , θ_2) est celle obtenue en ordonnant deux variables aléatoires indépendantes, chacune suivant une loi uniforme sur (0 , 1), $\phi_1(x) = \phi_2(x) = x^2$ et deux points échantillonnaux $y_1 < y_2$ sont utilisés. Il est possible de vérifier que :

	δ : reconstruction bayesienne	
$\Omega = \{(k(y_1) , k(y_2))\}$	$\hat{\theta}_1$	$\hat{\theta}_2$
(k_1 , k_1)	$y_2 + (1 - y_2)/3$	$y_2 + 2(1 - y_2)/3$
(k_2 , k_2)	$y_1/2$	$(1 + y_2)/2$
(k_3 , k_3)	$y_1/3$	$2y_1/3$
(k_1 , k_2)	$(y_1 + y_2)/2$	$(1 + y_2)/2$
(k_1 , k_3)	$y_1 + (y_2 - y_1)/3$	$y_1 + 2(y_2 - y_1)/3$
(k_2 , k_3)	$y_1/2$	$(y_1 + y_2)/2$

$$r(G , \delta)/2 = \int_{y_2}^{1} \int_{\theta_1}^{1} [\theta_1 - (2y_2 + 1)/3]^2 + [\theta_2 - (y_2 + 2)/3]^2 \, d\theta_2 \, d\theta_1$$

$$+ \int_{0}^{y_1} \int_{y_2}^{1} [\theta_1 - y_1/2]^2 + [\theta_2 - (1 + y_2)/2]^2 \, d\theta_2 \, d\theta_1$$

$$+ \int_{0}^{y_1} \int_{\theta_1}^{y_1} [\theta_1 - y_1/3]^2 + [\theta_2 - 2y_1/3]^2 \, d\theta_2 \, d\theta_1$$

$$+ \int_{y_1}^{y_2} \int_{y_2}^{1} [\theta_1 - (y_1 + y_2)/2]^2 + [\theta_2 - (1 + y_2)/2]^2 \, d\theta_2 \, d\theta_1$$

$$+ \int_{y_1}^{y_2} \int_{\theta_1}^{y_2} [\theta_1 - (2y_1 + y_2)/3]^2 + [\theta_2 - (y_1 + 2y_2)/3]^2 \, d\theta_2 \, d\theta_1$$

$$+ \int_{0}^{y_1} \int_{1}^{y_2} [\theta_1 - y_1/2]^2 + [\theta_2 - (y_1 + y_2)/2]^2 \, d\theta_2 \, d\theta_1 .$$

Afin de trouver la répartition optimale des points échantillonnaux y_1 , y_2 il faut minimiser cette somme d'intégrales par rapport à y_1 et y_2. C'est un travail très long et qui pourrait se compliquer si la loi a priori pour $(\theta_1 . \theta_2)$ n'était pas dérivée de la loi uniforme. Nous pouvons soupçonner ici, à cause de la loi uniforme, que le choix optimal des points échantillonnaux est 1/3 et 2/3.

Les règles bayesiennes obtenues dans les deux exemples précédents semblent indiquer un lien entre celles-ci et la règle du point le plus proche introduite par Switzer [10]. Dans le cas où la loi a priori est obtenue de la loi uniforme comme

dans l'exemple 2, et où $\phi_i(x) = x^2$ pour chaque $i = 1, \ldots, n-1$, il est possible de vérifier les faits suivants : si un point échantillonnal y_t est de couleur k_i et le point échantillonnal suivant y_{t+1} de couleur k_{i+1}, alors l'estimation bayesienne de θ_i est $(y_t + y_{t+1})/2$ (utilisant au besoin que 0 est de couleur k_1 et 1 de couleur k_n), ce qui est la règle du point le plus proche ; par contre si y_t est de couleur k_i et y_{t+1} de couleur k_{i+s} ($s > 1$) cela veut dire que les points de division $\theta_i, \ldots, \theta_{i+s-1}$ sont entre y_t et y_{t+1} ; les estimations bayesiennes de $\theta_i, \ldots, \theta_{i+s-1}$ sont respectivement $y_t + j(y_{t+1} - y_t)/(s+1)$, $j = 1, \ldots, s$, ce qui est une règle de partage égal généralisant d'une certaine façon la règle du point le plus proche. De plus dans ce cas nous pouvons montrer que :

$$R(a, \delta_0) \leqslant \sup_b R(b, \delta)$$

pour chaque schéma a et chaque règle de reconstruction δ, δ_0 étant la règle bayesienne dont il vient d'être question, ainsi cette règle est minimax.

La recherche des points échantillonnaux optimaux bien que théoriquement possible peut devenir très complexe à réaliser. Cette situation nous conduit à envisager une méthode séquentielle pour le choix des points échantillonnaux.

III – ECHANTILLONNAGE SEQUENTIEL

Par échantillonnage séquentiel, nous entendons la procédure suivante : un premier point échantillonnal est choisi et sa couleur observée, dépendant de celle-ci la position d'un deuxième point échantillonnal est déterminée et la couleur de celui-ci observée, dépendant de la couleur des deux premiers points échantillonnaux la position d'un troisième est déterminée et ainsi de suite. Devant une telle procédure, trois questions se posent immédiatement :

i) Comment ces points échantillonnaux sont-ils choisis ?

ii) Quand l'échantillonnage s'arrête-t-il ?

iii) Quand l'échantillonnage est terminé comment la reconstruction est-elle faite ?

La réponse à la troisième question est immédiate, la règle bayesienne de reconstruction basée sur l'information accumulée doit être utilisée.

Dans ce qui suit nous voulons présenter des exemples de réponses aux deux autres questions. Notre souci premier est de faire des suggestions qui soient facilement transposables en pratique. A travers les justifications données le lecteur pourra voir les modifications à apporter à ces suggestions afin qu'elles s'adaptent mieux à des situations particulières. Afin de simplifier la présentation nous nous restreindrons au cas où la fonction de perte est telle que $\phi_i(x) = x^2$, $i = 1, \ldots, n-1$.

Considérons la deuxième question, il y a ici deux cas distincts à envisager : celui où l'on tient compte d'un coût pour chaque observation et celui où ce coût est négligeable. Nous commençons par traiter ce dernier cas. La réponse à la deuxième question qui est principalement étudiée ici est la suivante.

"Des points échantillonnaux sont prélevés un à un jusqu'à ce que l'on ait au moins un point échantillonnal de chaque couleur".

De cette façon une précision raisonnable peut être obtenue en prenant un nombre limité d'observations. Par contre il peut exister des situations où il est impossible de prendre plus qu'un nombre fixe d'observations dans ce cas notre suggestion est de les prendre séquentiellement suivant la procédure suggérée plus bas. Dans le cas où il n'y a pas d'inconvénient à prendre beaucoup d'observations et avantage à avoir une reconstruction précise, il pourrait être envisagé de prendre des observations une à une jusqu'à ce que pour chaque couleur i, la distance entre le plus grand point échantillonnal de couleur i et le plus petit point échantillonnal de couleur i + 1 soit inférieure à une quantité pré-déterminée $\delta_i > 0$.

Une fois les observations faites les estimateurs bayesiennes sont donnés par :

$$\hat{\theta}_i = \left[\int_{y_{s_1}}^{y_{\ell_2}} \cdots \int_{y_{s_i}}^{y_{\ell_{i+1}}} \theta_i \int_{y_{s_{i+1}}}^{y_{\ell_{i+2}}} \cdots \int_{y_{s_{n-1}}}^{y_{\ell_n}} dG(\theta_1, \ldots, \theta_{n-1}) \right] \Bigg/ \left[\int_{y_{s_1}}^{y_{\ell_2}} \cdots \int_{y_{s_{n-1}}}^{y_{\ell_n}} dG(\theta_1, \ldots, \theta_{n-1}) \right]$$

$i = 1, \ldots, n-1$, y_{s_t} étant le plus grand point échantillonnal de couleur d'indice t ($t = 1, \ldots, n-1$) et y_{ℓ_u} étant le plus petit point échantillonnal de couleur d'indice u ($u = 2, \ldots, n$).

Passons maintenant à la première question. Nous présentons deux suggestions qui ne sont pas très différentes. Certaines comparaisons entre celles-ci seront faites plus loin.

Première suggestion

Le premier point considéré est $E[\theta_1]$; dénotons par Y_1 l'information obtenue en observant ce point. Si le premier point observé n'est pas de couleur k_1 alors le deuxième point considéré est $E[\theta_1 \mid \theta \in P(Y_1)]$. Dénotons par Y_2 l'information obtenue de ce deuxième point, si celui-ci n'est pas de couleur k_1 le troisième point considéré est $E[\theta_1 \mid \theta \in P(Y_1, Y_2)]$. Ce processus continue jusqu'à ce qu'un premier point de couleur k_1 soit obtenu. Rappelons que $P(Y_1, Y_2)$ dénote l'ensemble des $\theta \in \Theta$ correspondant aux schémas $a \in A$ tels que (Y_1, Y_2) est compatible avec a.

Dénotons par $Y_1, \ldots, Y_{N_1}$ les pièces d'information accumulées au moment où pour la première fois un point de couleur k_1 est obtenue, et par i_2 ($i_2 < n$) l'indice de la première couleur après k_1 qui ne soit pas apparue dans les N_1 premières étapes. Alors le point échantillonnal suivant est $E[\theta_{i_2} \mid \theta \in P(Y_1, \ldots, Y_{N_1})]$. Ce processus se continue jusqu'à ce qu'un premier point de couleur k_{i_2} soit obtenu.

En continuant de cette façon au moins un point de chacune des couleurs $k_1, \ldots, k_{n-1}$ est obtenu. A ce moment, si un point de couleur k_n a déjà été obtenu le processus d'échantillonnage est terminé, sinon le point $E[\theta_{n-1} \mid \theta \in P(\bar{Y})]$ est considéré, $\bar{Y}$ représente l'information accumulée au moment où ce point doit être choisi. Le processus est répété jusqu'à ce qu'un point de couleur k_n soit obtenu.

La motivation derrière cette suggestion peut s'énoncer comme suit. D'abord l'estimation bayesienne de θ_1 est faite, ceci dans la situation où l'on a pas encore d'information, cette estimation est $E[\theta_1]$ (avec la fonction de perte considérée). Le point obtenu est observé. Si ce point est de couleur k_1 alors l'estimation bayesienne de θ_2, étant donné l'information déjà disponible, est obtenue, sinon l'estimation bayesienne de θ_1 est refaite en tenant compte cette fois de l'information accumulée. Cela détermine le deuxième point échantillonnal. Ce raisonnement se poursuit jusqu'à ce qu'au moins un point de chaque couleur ait été obtenu et alors la reconstruction est exécutée.

Deuxième suggestion

Suivant l'approche proposée plus haut la recherche d'un point de couleur k_i se fait par l'estimation de l'extrémité droite de la cellule de couleur k_i. En faisant plutôt l'estimation d'un point à l'intérieur de la cellule, le point milieu par exemple, il peut être intuitivement prévu que nous arriverions plus rapidement à obtenir un point de chaque couleur. A cette fin il faut apporter les changements suivants à la première suggestion :

1) Le premier point considéré est $E[\theta_1]/2$, s'il n'est pas de couleur k_1 le point $E[\theta_1 \mid \theta \in P(Y_1)]/2$ est considéré et ainsi de suite jusqu'à ce qu'un premier point de couleur k_1 soit obtenu.

2) Dans la recherche d'un point de couleur k_i $(1 < i < n)$ le point

$$E[(\theta_{i-1} + \theta_i)/2 \mid \theta \in P(\bar{Y})] \quad \text{est examiné.}$$

3) Afin d'obtenir un point de couleur k_n lorsqu'un tel point n'a pas déjà été obtenu le point $\{E[\theta_{n-1} \mid \theta \in P(\bar{Y})] + 1\}/2$ est considéré.

Une propriété minimale d'une procédure séquentielle est qu'elle doit se terminer après un nombre fini d'étapes. Relativement à la première suggestion présentée nous pouvons montrer que si la loi de probabilité a priori G est du type continu alors $\sum_{j=n}^{\infty} P_a[N = j] = 1$ pour chaque $a \in A$, $P_a[N = j]$ étant la probabilité que le nombre N d'observations nécessaires pour terminer l'échantillonnage soit égal à j lorsque le schéma est a. Si G est telle que certaines des composantes de $(\theta_1, \ldots, \theta_{n-1})$ sont discrètes nous pouvons démontrer le même résultat sous les conditions suivantes :

i) dénotons par $j_1, \ldots, j_t$ les indices des variables aléatoires discrètes avec support non fini parmi $\theta_1, \ldots, \theta_{n-1}$, alors pour chaque $1 \leqslant s \leqslant t$ tous les points d'accumulation θ_0 des éléments du support de θ_{j_s} sont tels que $P[\theta_{j_s} = \theta_0] = 0$.

ii) Si θ_{n-1} est une variable aléatoire discrète alors son support n'est pas fini.

Dans le cas de la deuxième suggestion nous avons le même résultat.

Il est à remarquer que la définition donnée pour un schéma est telle que toutes les cellules sont de la forme $(\theta_{i-1}, \theta_i]$ sauf la première qui est $[0, \theta_1]$. Si G n'a pas d'atome cette convention n'a pas d'importance tant qu'à la terminaison du processus d'échantillonnage décrit dans la première suggestion ; dans le cas du deuxiéme processus suggéré la convention n'a pas d'importance quelle que soit la nature de G.

Exemple :

Nous allons considérer en détail le cas où la loi de $(\theta_1, \ldots, \theta_{n-1})$ est celle obtenue en ordonnant $n-1$ variables aléatoires indépendantes, chacune suivant une loi uniforme sur $(0, 1)$. Dans ce cas la fonction de densité de la loi à priori est :

$$g(\theta_1, \ldots, \theta_{n-1}) = \begin{cases} (n-1)! & \text{si } 0 < \theta_1 < \theta_2 < \ldots < \theta_{n-1} < 1 \\ 0 & \text{autrement} \end{cases}$$

Dans le cas du premier plan d'échantillonnage suggéré la procédure s'énonce comme suit :

i) Le premier point considéré est $E[\theta_1] = \frac{1}{n}$.

ii) Supposons que le point $\frac{1}{n}$ est de couleur k_j $j > 1$, cela veut dire que $\theta_1 < \theta_2 < \ldots < \theta_{j-1} < \frac{1}{n} \leqslant \theta_j < \ldots < \theta_{n-1} < 1$ et, suivant le plan d'échantillonnage, pour localiser le point suivant il suffit de calculer

$$E\left[\theta_1 \mid \theta_1 < \ldots < \theta_{j-1} < \frac{1}{n} \leqslant \theta_j < \ldots < \theta_{n-1} < 1\right] :$$

$$= \frac{\displaystyle\int_0^{1/n} \theta_1 \int_{\theta_1}^{1/n} \cdots \int_{\theta_{j-2}}^{1/n} \int_{1/n}^{1} \int_{\theta_j}^{1} \cdots \int_{\theta_{n-2}}^{1} (n-1)!\, d\theta_{n-1} \ldots d\theta_1}{\displaystyle\int_0^{1/n} \int_{\theta_1}^{1/n} \int_{\theta_{j-2}}^{1/n} \int_{1/n}^{1} \int_{\theta_j}^{1} \cdots \int_{\theta_{n-2}}^{1} (n-1)!\, d\theta_{n-1} \ldots d\theta_1}$$

$$= \frac{1/n^j \displaystyle\sum_{i=2}^{j} (-1)^i/[i(j-1)!\,(i-2!]}{1/n^{j-1} \displaystyle\sum_{i=2}^{i} (-1)^i/[(j-i)!\,(i-1)!]} = \frac{1}{jn}$$

Cela se continue jusqu'à ce qu'un point de couleur k_1 ait été obtenu.

iii) Considérons maintenant la recherche d'un point de couleur k_j, $1 < j < n$, lorsque des points de couleur $k_1, \ldots, k_{j-1}$ ont déjà été obtenus. Il se peut que des points dont la couleur a un indice supérieur à j aient aussi déjà été obtenus, dénotons par t le plus petit de ces indices. Ainsi il faudra considérer le point échantillonnal :

$$E[\theta_j \mid y_{s_1} < \theta_1 < y_{\ell_2}, \ldots, y_{s_{j-2}} < \theta_{j-2} \leqslant y_{\ell_{j-1}},$$
$$y_{s_{j-1}} < \theta_{j-1} < \theta_j < \ldots < \theta_{t-1} < y_{\ell_t},$$
$$y_{s_t} < \theta_t < \theta_{t+1} < \ldots < \theta_{n-1} < 1]$$

où comme précédemment y_{s_i} et y_{ℓ_j} désignent respectivement le plus grand point de couleur k_i et le plus petit point de couleur k_j déjà obtenus ; il est sous-entendu que si de l'information supplémentaire sur les cellules de couleur k_u , $u > t$ est disponible celle-ci doit être incorporée dans le bloc $y_{s_t} < \theta_t < \cdots < \theta_{n-1} < 1$. Dans le présent exemple il est facile de voir que l'expression qui précède se ramène à :

$$\frac{\int_{y_{s_{j-1}}}^{y_{\ell_t}} \int_{\theta_{j-1}}^{y_{\ell_t}} \theta_j \int_{\theta_j}^{y_{\ell_t}} \cdots \int_{\theta_{t-2}}^{y_{\ell_t}} d\theta_{t-1} \cdots d\theta_{j-1}}{\int_{y_{s_{j-1}}}^{y_{\ell_t}} \int_{\theta_{j-1}}^{y_{\ell_t}} \int_{\theta_j}^{y_{\ell_t}} \cdots \int_{\theta_{t-2}}^{y_{\ell_t}} d\theta_{t-1} \cdots d\theta_{j-1}} = y_{s_{j-1}} + \frac{2(y_{\ell_t} - y_{s_{j-1}})}{t - j + 2}$$

iv) Finalement pour cherche un point de couleur k_n, si un tel point n'a pas déjà été obtenu, il faut considérer :

$$E\left[\theta_{n-1} \mid y_{s_1} < \theta_1 \leqslant y_{\ell_2}, \ldots, y_{s_{n-2}} \leqslant \theta_{n-2} \leqslant y_{\ell_{n-1}}, \theta_{n-1} \geqslant y_{s_{n-1}}\right]$$

qui dans ce cas-ci est $(1 + y_{s_{n-1}})/2$.

Il est facile de vérifier que si le plan échantillonnal présenté dans la deuxième suggestion est utilisé, les changements à apporter sont de remplacer en i) 1/n par 1/2n, en ii) 1/jn par 1/4jn, en iii)

$$y_{s_{j-1}} + 2(y_{\ell_t} - y_{s_{j-1}})/(t-j+2) \text{ par } y_{s_{j-1}} + 3(y_{\ell_t} - y_{s_{j-1}})/2\,(t-j+2)$$

et en iv) $(1 + y_{s_{n-1}})/2$ par $3/4 + y_{s_{n-1}}/4$.

Nous voulons terminer cette section en présentant une suggestion dans le cas où le coût d'une observation doit être pris en considération. Nous supposons que le coût d'une observation est constant et indépendant des observations déjà accumulées ; il peut être commode ici de considérer une fonction de perte où $\phi_i(x) = kx^2$, $i = 1, \ldots, n-1$, k une constante.

L'idée à la base de la procédure envisagée est que lorsque j observations ont été prises, une observation supplémentaire sera prise seulement si le risque global encouru en la prenant, étant donné les j observations déjà obtenues, est inférieur au risque global encouru en arrêtant après les j observations disponibles étant donné ces observations. Plus simplement, il s'agit d'arrêter l'échantillonnage dès que le coût d'une observation supplémentaire est plus grand que le gain moyen enregistré en prenant une telle observation. Il s'agit là en fait d'une modification aux règles d'arrêt déjà proposées. Les points échantillonnaux continuent à être localisés suivant une des suggestions faites plus haut. Nous allons maintenant expliciter la procédure envisagée. Dénotons par c le coût d'une observation.

Il doit d'abord être décidé si une première observation sera prise. Dans le cas où aucune observation ne serait prise les estimations de $\theta_1, \ldots, \theta_{n-1}$ seraient respectivement $E[\theta_1], \ldots, E[\theta_{n-1}]$ et le risque serait :

$$r_0 = \int_\Theta k \sum_{i=1}^{n-1} [\theta_i - E(\theta_i)]^2 \, dG(\theta_1, \ldots, \theta_{n-1})$$

$$= k \sum_{i=1}^{n-1} \text{Var}(\theta_i).$$

Il faut comparer r_0 avec le risque global encouru en prenant une observation, celui-ci est donné par :

$$r_1 = \sum_{Y \in \Omega_1} \int_{P(Y)} k \sum_{i=1}^{n-1} [\theta_i - \hat{\theta}_i(Y)]^2 \, dG(\theta_1, \ldots, \theta_{n-1}) + c$$

où Ω_1 est l'espace échantillon dans le cas où une observation est prise. Ainsi cette première observation sera prise si $r_0 - r_1 > c$.

Supposons que cette première observation fut prise en y_1 (par exemple en $E[\theta_1]$ suivant le premier plan d'échantillonnage suggéré) et que la couleur de ce point est k_{i_1}. Alors avant de prendre une deuxième observation il faut calculer :

$$\bar{r}_1(k(y_1) = k_{i_1}) = \int_{P(k(y_1)=k_{i_1})} k \sum_{i=1}^{n-1} [\theta_i - \hat{\theta}_i(k(y_1) = k_{i_1})]^2 \, dG(\theta_1, \ldots, \theta_{n-1})$$

$$\bar{r}_2(k(y_1) = k_{i_1}) = \sum_{Y \in \Omega_2(k(y_1)=k_{i_1})} \int_{P(Y)} k \sum_{i=1}^{n-1} [\theta_i - \hat{\theta}_i(Y)]^2 \, dG(\theta_1, \ldots, \theta_{n-1})$$

où $\Omega_2(k(y_1) = k_{i_1})$ est l'espace échantillon correspondant à deux points échantillonnaux le premier étant y_1 et de couleur k_{i_1}, $\bar{r}_1(k(y_1) = k_{i_1})$ est le risque conditionnel encouru en prenant une observation, celle-ci étant en y_1 et de couleur k_{i_1}, $\bar{r}_2(k(y_1) = k_{i_1})$ est le risque conditionnel encouru en prenant deux observations, la première étant en y_1 et de couleur k_{i_1}. Une deuxième observation sera prise si :

$$\bar{r}_1(k(y_1) = k_{i_1}) - \bar{r}_2(k(y_1) = k_{i_1}) > c.$$

Ce raisonnement se continue, supposons qu'ayant pris t observations en $y_1, \ldots, y_t$ et ayant observé les couleurs $k_{i_1}, \ldots, k_{i_t}$ l'on doive décider si une $(t+1)^{\text{ième}}$ observation doit être prise. Il faut alors calculer :

$$\bar{r}_t(k(y_1) = k_{i_1}, \ldots, k(y_t) = k_{i_t})$$

$$= \int_{P(k(y_1)=k_{i_1}, \ldots, k(y_t)=k_{i_t})} k \sum_{i=1}^{n-1} [\theta_i - \hat{\theta}_i(k(y_1) = k_{i_1}, \ldots, k(y_t) = k_{i_t})]^2 \, dG(\theta_1, \ldots, \theta_{n-1})$$

$$\bar{r}_{t+1}(k(y_1) = k_{i_1}, \ldots, k(y_t) = k_{i_t})$$

$$= \sum_{Y \in \Omega_{t+1}(k(y_1)=k_{i_1}, \ldots, k(y_t)=k_{i_t})} \int_{P(Y)} k \sum_{i=1}^{n-1} [\theta_i - \hat{\theta}_i(Y)]^2 \, dG(\theta_1, \ldots, \theta_{n-1})$$

et une $(t+1)^{\text{ième}}$ observation sera prise si :

$$\bar{r}_t(k(y_1) = k_{i_1}, \ldots, k(y_t) = k_{i_t}) - \bar{r}_{t+1}(k(y_1) = k_{i_1}, \ldots, k(y_t) = k_{i_t}) > c.$$

L'exemple suivant montre que cette procédure peut facilement être mise en pratique.

Exemple :

Nous envisageons à nouveau le deuxième exemple mentionné à la section 2 et supposons que les points échantillonnaux sont localisés suivant la première suggestion faite. Il est facile de vérifier qu'alors $r_0 = k/9$. Dans le cas où une première observation serait prise celle-ci serait située en 1/3 et en arrêtant l'échantillonnage après cette observation les estimations seraient :

$k\left(\frac{1}{3}\right)$	$\hat{\theta}_1$	$\hat{\theta}_2$
k_1	$\frac{5}{9}$	$\frac{7}{9}$
k_2	$\frac{1}{6}$	$\frac{2}{3}$
k_3	$\frac{1}{9}$	$\frac{2}{9}$

et

$$r_1 = 2k\left\{\int_{1/3}^{1}\int_{\theta_1}^{1}\left[\left(\theta_1-\frac{5}{9}\right)^2+\left(\theta_2-\frac{7}{9}\right)^2\right]d\theta_2\,d\theta_1\right.$$

$$+\int_{0}^{1/3}\int_{1/3}^{1}\left[\left(\theta_1-\frac{1}{6}\right)^2+\left(\theta_2-\frac{2}{3}\right)^2\right]d\theta_2\,d\theta_1$$

$$\left.+\int_{0}^{1/3}\int_{\theta_1}^{1/3}\left[\left(\theta_1-\frac{1}{9}\right)^2+\left(\theta_2-\frac{2}{9}\right)^2\right]d\theta_2\,d\theta_1\right\}$$

$$= (0{,}043896)\,k$$

Ainsi une première observation sera prise si $k(0{,}067215) > c$. Supposons que cette première observation soit prise et s'avère être de couleur k_1. Alors :

$$r_1\left(k\left(\frac{1}{3}\right)=k_1\right) = 2k\int_{1/3}^{1}\int_{\theta_1}^{1}\left[\left(\theta_1-\frac{5}{9}\right)^2+\left(\theta_2-\frac{7}{9}\right)^2\right]d\theta_2\,d\theta_1 =$$

$$= (0{,}021948)\,k.$$

Si une deuxième observation est prise elle doit être en :

$$E\left[\theta_2 \mid \frac{1}{3} < \theta_1 < \theta_2 < 1\right] = \frac{7}{9}.$$

Dans le cas où cette observation serait prise et conditionnellement à la première observation, les estimations seraient après la deuxième observation :

$\left(k\left(\frac{1}{3}\right), k\left(\frac{7}{9}\right)\right)$	$\hat{\theta}_1$	$\hat{\theta}_2$
(k_1, k_1)	$\frac{23}{27}$	$\frac{25}{27}$
(k_1, k_2)	$\frac{5}{9}$	$\frac{8}{9}$
(k_1, k_3)	$\frac{13}{27}$	$\frac{17}{27}$

et il est facile de vérifier que $\bar{r}_2\left(k\left(\frac{1}{3}\right) = k_1\right) = (0{,}008670)\,k$. Une deuxième observation serait donc prise si $(0{,}013278)\,k > c$.

Le traitement serait analogue dans les cas où le point 1/3 serait de couleur k_2 et k_3. La procédure se poursuit de la même façon.

IV – LE NOMBRE D'OBSERVATIONS

Bien que nous ayons donné précédemment des conditions sous lesquelles les procédures d'échantillonnage proposées se terminent avec certitude, il serait souhaitable, comme c'est le cas dans toute procédure séquentielle, d'obtenir de l'information sur la loi du nombre d'observations requises, N, pour que la procédure d'échantillonnage se termine. En général l'obtention de la loi de N est très difficile. Cependant dans le cas particulier où $n = 2$ et θ_1 suit une loi uniforme sur $(0, 1)$ le problème peut être complètement résolu et nous présenterons d'abord cette solution.

Désignons par $N_1(N_2)$ le nombre de points à prélever pour obtenir un point de couleur $k_1(k_2)$. Les valeurs de ces variables sont les suivantes (rappelons que l'échantillonnage se termine dès qu'au moins un point de chaque couleur a été obtenu) :

N_1	1	2, 3, 4, . . .
N_2	1, 2, 3, . . .	0

Alors $P[N = k] = P[N_1 = 1,\ N_2 = k - 1] + P[N_1 = k,\ N_2 = 0]$, $k \geqslant 2$ et dans le cas où les points échantillonnaux sont localisés suivant la première suggestion :

$$P[N_1 = 1,\ N_2 = k - 1] = P[(2^{k-1} - 1)/2^{k-1} < \theta < (2^k - 1)/2^k] = 1/2^k$$

$$P[N_1 = k,\ N_2 = 0] = P[1/2^k < \theta < 1/2^{k-1}] = 1/2^k$$

$$P[N = k] = 1/2^{k-1} \quad \text{et} \quad E[N] = 3.$$

Ainsi dans ce cas il faudra prélever en moyenne trois observations. Le risque bayesien correspondant à cette procédure peut s'écrire :

$$\sum_{j=1}^{\infty} \int_{(2^j-1)/2^j}^{(2^{j+1}-1)/2^{j+1}} \left(\theta - \frac{2^{j+2}-3}{2^{j+2}}\right)^2 d\theta + \sum_{j=1}^{\infty} \int_{1/2^{j+1}}^{1/2^j} + \left(\theta - \frac{3}{2^{j+2}}\right)^2 d\theta = \frac{1}{(16)(21)}$$

Comparons cette procédure avec une procédure non séquentielle où trois observations seraient prises. Il a été mentionné dans le premier exemple de la section 2 que le choix optimal de ces points serait 1/4, 1/2, 3/4. Dans ce cas le risque bayesien est :

$$\int_{3/4}^{1} \left(\theta - \frac{7}{8}\right)^2 d\theta + \int_{1/2}^{3/4} \left(\theta - \frac{5}{8}\right)^2 d\theta + \int_{1/4}^{1/2} \left(\theta - \frac{3}{8}\right)^2 d\theta +$$

$$+ \int_{0}^{1/4} \left(\theta - \frac{1}{8}\right)^2 d\theta = \frac{1}{(64)(3)}$$

et le rapport entre les risques bayesiens (non séquentiel/séquentiel) est 1.75. La procédure séquentielle est donc plus précise mais peut requérir plus d'observations, en fait $P[N > 3] = 1/4$. Considérons maintenant le cas où les points échantillonnaux seraient localisés suivant la deuxième suggestion faite. Il est possible de vérifier que $P[N = k] = 12/2^{2k}$, $k \geqslant 2$, $E[N] = 7/3$ et le risque bayesien est 1/64. Tel que prévu le nombre moyen d'observations requises est plus petit en suivant la deuxième suggestion, par contre on constate que la précision de la reconstruction est moins bonne. En fait, il peut facilement être vérifié que si les deux points fixes 1/3, 2/3 sont utilisés (choix optimal) pour effectuer la reconstruction, alors le risque bayesien est 1/108, ce qui est inférieur à celui obtenu en suivant la deuxième procédure séquentielle suggérée.

Passons maintenant à la loi de N dans le cas d'une valeur quelconque de n. Si $n > 2$, il n'est pas simple de dresser un tableau, comme celui donné pour $n = 2$, présentant les différentes valeurs de N en fonction du nombre d'essais requis pour atteindre chaque couleur. En général il est donc difficile de procéder comme dans l'exemple plus haut pour étudier la loi de N. Nous avons donc effectué une simulation.

Pour une valeur donnée de n, $(n - 1)$ copies indépendantes d'une variable aléatoire suivant une loi uniforme sur $(0, 1)$ sont générées, celles-ci sont ordonnées pour obtenir $\theta_1, \ldots, \theta_{n-1}$. Suivant l'une ou l'autre des suggestions faites plus haut, des points échantillonnaux sont localisés ceci jusqu'à ce qu'au moins un point de chaque couleur soit obtenu. Alors le nombre d'observations effectuées est enregistré, les estimations $\hat{\theta}_1, \ldots, \hat{\theta}_{n-1}$ sont calculées ainsi que la quantité $\sum_{i=1}^{n-1} (\theta_i - \hat{\theta}_i)^2$. En répétant ce processus un grand nombre de fois il est possible d'obtenir une approximation de la loi de N et du risque bayesien. Les résultats pour 10,000 répétitions du processus avec $n = 3, 4, 5, 6$ sont présentés dans le tableau 1.

TABLEAU 1

Loi a priori uniforme et application de la procédure correspondant à cette loi

	Première suggestion				Deuxième suggestion			
n	3	4	5	6	3	4	5	6
E [N]	4,598	6,116	7,644	9,165	3,827	5,347	6,839	8,417
Mode (probabilité)	3 (0,343)	4 (0,246)	6 (0,216)	7 (0,186)	3 (0,521)	4 (0,358)	6 (0,272)	7 (0,234)
Risque bayesien	0,0076	0,0100	0,0097	0,0088	0,0152	0,0125	0,0104	0,0088

Comme dans l'exemple plus haut il est constaté que la deuxième suggestion est plus économique mais moins précise, bien que cet inconvénient semble s'atténuer lorsque n augmente.

La simulation peut aussi être utilisée pour évaluer la robustesse de la méthodologie suggérée. Plus précisément, supposons que la procédure obtenue pour une loi a priori correspondant à la loi uniforme sur (0 , 1) est utilisée, mais qu'en réalité le schéma est engendré par une autre loi, quel sera alors l'effet sur N ? Dans le but d'illustrer cette question nous avons substitué à la loi uniforme deux types de lois triangulaires pour générer $\theta_1, \ldots, \theta_{n-1}$, mais nous avons conservé la procédure obtenue de la loi uniforme pour localiser les points échantillonnaux. Les lois triangulaires considérées sont :

$$f(x) = \begin{cases} 2x & \text{si } 0 < x < 1 \\ 0 & \text{autrement} \end{cases} \tag{4.1}$$

$$f(x) = \begin{cases} 4x & \text{si } 0 < x < 1/2 \\ 4(1-x) & \text{si } 1/2 \leqslant x < 1 \\ 0 & \text{autrement} \end{cases} \tag{4.2}$$

Les tableaux 2 et 3 présentent les résultats.

TABLEAU 2

Loi a priori triangulaire (4.1) et application de la procédure correspondant à la loi uniforme.

	Première suggestion				Deuxième suggestion			
n	3	4	5	6	3	4	5	6
E [N]	4,396	5,942	7,526	9,116	4,167	5,992	7,604	9,352
Mode (probabilité)	3 (0,366)	5 (0,273)	6 (0,226)	8 (0,194)	3 (0,364)	5 (0,291)	7 (0,224)	8 (0,207)
Risque bayesien	0,0091	0,0095	0,0084	0,0073	0,0111	0,0084	0,0070	0,0056

TABLEAU 3

Loi a priori triangulaire (4.2) et application de la procédure correspondant à la loi uniforme

	Première suggestion				Deuxième suggestion			
n	3	4	5	6	3	4	5	6
E [N]	4,036	5,514	7,065	8,654	3,619	5,179	6,709	8,274
Mode (Probabilité)	3 (0,477)	4 (0,348)	5 (0,249)	7 (0,215)	3 (0,641)	4 (0,428)	6 (0,293)	7 (0,249)
Risque bayesien	0,0070	0,0076	0,0074	0,0068	0,0131	0,0116	0,0094	0,0083

A l'examen de ces deux derniers tableaux nous constatons que l'utilisation de la procédure correspondant à la loi uniforme dans les situations envisagées où la loi a priori n'est pas obtenue de la loi uniforme ne pose aucun inconvénient. Pour la loi triangulaire (4.2) nous remarquons encore que l'échantillonnage suivant la deuxième suggestion est plus économique mais engendre une reconstruction moins précise. La situation n'est pas aussi claire dans le cas de la loi (4.1).

En conclusion nous pouvons dire que l'échantillonnage suivant une des deux suggestions faites est facile à implanter, économique et les reconstructions obtenues sont assez précises. De plus la procédure correspondant à la loi a priori obtenue de la loi uniforme semble robuste contre les autres lois a priori.

L'auteur désire exprimer sa gratitude envers le Centre de Recherches Mathématiques de l'Université de Montréal où cette recherche fut effectuée durant l'année académique 1977-1978. Des remerciements sont adressés à M. Michel Girardeau pour son aide dans la préparation de la simulation et au lecteur de la Revue de Statistique Appliquée pour ses judicieux conseils. Cette recherche a pu être effectuée grâce à une subvention du Conseil de recherche en sciences naturelles et en génie Canada (Subvention A-8211).

BIBLIOGRAPHIE

[1] CRAIN R., MILES R. – Monte Carlo estimates of the distributions of the random polygons determined by random lines in a plane. *J. Statist. Comput. Simul.*, Vol. 4, 1976, pages 293-325.

[2] FREIBERGER W., GRENANDER U . – A short course in computational probability and statistics. *Springer-Verlag,* New-York, 1975.

[3] FERGUSON T. – Mathematical statistics – A decision theoretic approach. *Academic Press,* New-York, 1967.

[4] GUILBERT E. – Random subdivision of space into crystals. *Ann. Math. Stat.*, Vol. 33, 1962, pages 958-972.

[5] MATSCHINSKI M. – Détermination expérimentale des probabilités (W_n) du nombre des côtés des polygônes couvrant chaotiquement un plan sans lacune. *Publication de l'Institut de Statistique de l'Université de Paris,* Vol. 18, fasc. 3, 1969, pages 1-20.

[6] MILES R.E. – Random polygons determined by random lines in a plane. *Proc. Nat. Acad. Sci.* (U.S.A.), Vol. 52, 1964, pages 901-907 et pages 1157-1160.

[7] MILES R.E. – On the homogeneous planar Poisson point process. *Math. Biosc.*, Vol. 6, 1970, pages 85-127.

[8] MILES R.E. – The various aggregates of random polygons determined by random lines in a plane. *Adv. Math.*, Vol. 10, 1973, pages 256-290.

[9] MOORE M. – Bayesian reconstruction of m, n-patterns. *Ann. Stat.*, Vol. 2, 1974, pages 1226-1237.

[10] SWITZER P. – Reconstructing patterns from sample data. *Ann. Math. Stat.*, Vol. 38, (1967), pages 138-154.