Techniques d'enquête, décembre 1987
Vol. 13, n° 2, pp. 193–209
Statistique Canada

Une classe de méthodes utilisant des chiffres de population dans la pondération des ménages

CHARLES H. ALEXANDER[1]

RÉSUMÉ

Nous analysons une série de méthodes fondées sur la notion de "distance minimum conditionnelle" et qui assurent la concordance des poids des ménages avec les données supplémentaires concernant l'effectif de diverses cellules d'âge × race × sexe. Les poids conditionnels se rapprochent autant que possible des poids initiaux, qui sont fondés sur l'inverse de la probabilité de sélection. Cette série de méthodes comprend la méthode itérative du quotient et la méthode des moindres carrés généralisés de même que la méthode multinomiale du maximum de vraisemblance, où les cellules de la distribution représentent des catégories de ménages. À l'aide de modèles d'observation simples, nous étudions les propriétés de ces méthodes dans une situation de sous-dénombrement systématique des catégories de ménages. Après une comparaison avec la méthode de la personne principale, nous concluons à la nécessité de mieux connaître la nature du sous-dénombrement avant de décider de la méthode la plus appropriée (méthodes fondées sur la distance minimum conditionnelle ou méthode de la personne principale).

MOTS CLÉS: Pondération; information supplémentaire; méthode itérative du quotient; méthode de la personne principale; champ d'enquête.

1. INTRODUCTION

La post-stratification est souvent utilisée pour rajuster les poids de sondage en fonction de données supplémentaires sur le nombre d'unités de certaines catégories dans la population. Par exemple, on peut obtenir des estimations indépendantes de la population pour diverses cellules de post-stratification portant sur l'âge, l'origine raciale et le sexe en rajustant les chiffres du recensement en fonction des variations démographiques observées depuis le recensement. Ces estimations indépendantes sont souvent appelées "chiffres de population". Avant la stratification a posteriori, chaque personne (ou ménage) de l'échantillon a déjà un poids qui correspond habituellement à l'inverse de sa probabilité de sélection. Dans la post-stratification, chaque poids est multiplié par un facteur de correction sous forme de ratio qui s'applique à chaque cellule de sorte que la somme des poids corrigés pour une cellule est égale à son chiffre de population. Cette correction est particulièrement importante lorsqu'il y a un sous-dénombrement systématique des ménages ou des personnes à l'intérieur des ménages.

Dans la plupart des enquêtes démographiques du U.S. Census Bureau, la post-stratification sert à attribuer des poids aux personnes de l'échantillon mais non aux ménages, car il est plus difficile d'obtenir des estimations indépendantes pour les ménages. Pour attribuer des poids aux ménages dans ces enquêtes, on utilise plutôt une version de la méthode de la personne principale. Selon la version fondamentale de cette méthode, on attribue aux ménages un poids équivalent à celui attribué par post-stratification à la personne principale du ménage. Dans la section 2, nous exposons les règles qui permettent de définir la personne principale d'un ménage. En utilisant les poids attribués aux personnes après la post-stratification, la méthode de la personne principale permet d'incorporer les estimations indépendantes du nombre de personnes dans les poids attribués aux ménages.

[1] Charles H. Alexander, Statistical Methods Division, Bureau of the Census, Washington, D.C. 20233, U.S.A.

Toutefois, cette méthode a un inconvénient majeur. En effet, lorsqu'on se sert des poids de ménage obtenus par cette méthode pour calculer des estimations pondérées du nombre de personnes dans chaque cellule de post-stratification, chaque personne étant affectée du poids du ménage auquel elle appartient, ces estimations ne correspondent pas aux chiffres de population utilisés dans la post-stratification. C'est pourquoi on s'est intéressé à des méthodes de pondération des ménages qui produisent des poids à partir desquels on obtient nécessairement des estimations de population conformes aux chiffres de population.

Le présent article vise à exposer une série de méthodes de pondération des ménages qui produisent des poids à partir desquels on obtient nécessairement des estimations conformes aux chiffres de population des diverses cellules. Le principe de base de ces méthodes est de déterminer des poids de ménage qui satisfont aux conditions et qui se rapprochent le plus possible du vecteur de poids attribués initialement aux ménages. Les méthodes analysées dans ce document sont des façons de mesurer la distance entre le vecteur de poids initial et le vecteur de poids corrigé.

Dans la section 2, nous décrivons six méthodes de pondération fondées sur la notion de "distance minimum conditionnelle" et une version de la méthode de la personne principale. Trois des six méthodes précitées ont déjà fait l'objet d'analyses; les autres sont présentées pour donner une image plus complète de la question. La section 3 porte sur le calcul des poids. Dans la section 4, nous voyons comment la composition du ménage influe sur la correction des poids. Dans la section 5, nous analysons les résultats et présentons des exemples qui peuvent aider à comprendre le mécanisme de ces méthodes. Enfin dans la section 6, nous proposons des sujets de recherche.

Cet article n'est pas le premier du genre. Luery (1986) propose d'appliquer la série de méthodes fondées sur la notion de distance minimum conditionnelle à la pondération des ménages. Poursuivant le travail de Luery, Zieschang (1986a) propose d'appliquer l'une de ces méthodes, la méthode des moindres carrés généralisés, à la pondération dans les enquêtes sur les dépenses des consommateurs aux États-Unis. Une autre méthode de la série est la "méthode de l'information minimum discriminante", connue aussi sous le nom d'"estimation itérative par la méthode du quotient" ou simplement "méthode itérative du quotient". Oh et Scheuren (1978a) traitent particulièrement de l'application de cette méthode à la pondération des ménages et enrichissent la bibliographie déjà très abondante sur cette méthode d'estimation et les méthodes connexes. Deming et Stephan (1940) comptent parmi les premiers qui ont associé la méthode itérative du quotient à la notion de distance minimum conditionnelle. Les principes fondamentaux de cette approche sont analysés dans Ireland et Kullback (1968). Brackstone et Rao (1979) ont bien décrit l'application de cette méthode à la correction des poids de sondage. La série de méthodes comprend aussi deux fonctions de critère qui se rattachent à la méthode multinomiale du maximum de vraisemblance. La relation entre cette méthode et la méthode itérative du quotient a fait l'objet de nombreuses études; voir, par exemple, Bishop, Fienberg et Holland (1976). Fienberg (1986) souligne que les critères de distance analysés dans le présent article peuvent être considérés comme des cas spéciaux d'une famille de fonctions paramétriques analysées dans Cressie et Read (1984).

2. MÉTHODES FONDÉES SUR LA NOTION DE DISTANCE MINIMUM CONDITIONNELLE

2.1 Méthodes fondées sur les poids des ménages

Considérons un échantillon de K ménages, dont les poids initiaux sont définis par le vecteur $\underline{S} = (S_1, \ldots, S_K)$. Dans cet article, S_k désigne l'inverse de la probabilité de sélection du k-ième ménage; dans certaines applications, le poids initial peut inclure d'autres facteurs de correction, par exemple des facteurs de non-réponse.

Supposons qu'il y a J cellules de post-stratification et que l'on connaît le nombre de personnes (N_j) dans chacune. Par exemple, dans le cas de l'enquête sur les dépenses des consommateurs aux États-Unis, il y a $J = 48$ cellules formées par la combinaison des deux sexes, de deux origines raciales (noir, non noir) et de douze catégories d'âge. Les personnes de moins de 14 ans ne sont pas visées par cette enquête. Les chiffres de population pour ces cellules sont définis en l'occurrence par un vecteur $\underline{N} = (N_1, \ldots, N_J)$.

La composition des ménages de l'échantillon est définie par une matrice $A = (a_{kj})$, où a_{kj} est égal au nombre de personnes du k-ième ménage qui se trouvent dans la j-ième cellule de post-stratification. La sommation des cellules de post-stratification pour le k-ième ménage donne $a_{k.}$, soit le nombre total de personnes que compte le ménage. Le vecteur $(a_{k1}, \ldots, a_{kJ})$ décrit la composition du ménage k. Par exemple, un vecteur $(2, 1, 0, 0, \ldots, 0)$ signifie que deux personnes du ménage se trouvent dans la première cellule et une troisième dans la seconde cellule.

Étant donné les poids initiaux $\underline{S}$, l'estimation pondérée du nombre de personnes dans la cellule j sera $\hat{N}_j = \Sigma_k a_{kj} S_k$ ou, de façon générale, $\hat{\underline{N}} = A'\underline{S}$.

En règle générale, $\hat{\underline{N}} \neq \underline{N}$, autrement dit, l'estimation pondérée initiale du nombre de personnes dans les cellules de post-stratification peut différer de l'effectif réel de la cellule.

Nous cherchons ici à définir un nouveau vecteur de poids $\underline{W} = (W_1, \ldots, W_K)$ pour les ménages de l'échantillon de sorte que $\underline{N} = A'\underline{W}$ ou

$$\sum_k a_{kj} W_k = N_j \text{ pour } j = 1, \ldots, J. \tag{1}$$

L'équation (1) n'a pas nécessairement une solution unique. Les méthodes fondées sur la notion de distance minimum conditionnelle consistent essentiellement à choisir un vecteur $\underline{W}$ de manière à minimiser la distance $D(\underline{W},\underline{S})$ entre les vecteurs $\underline{W}$ et $\underline{S}$ à condition que l'équation (1) soit respectée. Ainsi, en respectant la condition d'équivalence entre les poids corrigés et les totaux de contrôle, nous nous trouvons à modifier le moins possible les poids initiaux $\underline{S}$. Il convient de souligner que pour certaines valeurs $N_1, \ldots, N_J$, il peut être impossible de calculer un vecteur de poids $\underline{W}$ qui satisfasse à la condition (1). En pratique toutefois, cette éventualité ne semble pas poser de problème si l'échantillon est suffisamment grand pour inclure un nombre raisonnable d'unités des diverses catégories de ménages, puisque les totaux de contrôle $\underline{N}$ sont tirés de la population réelle et que par conséquent, on peut les considérer comme "possibles".

Il y a plusieurs façons de mesurer la différence entre deux vecteurs. Nous allons étudier trois critères de distance $D(\underline{W},\underline{S})$ qui correspondent respectivement, au niveau des ménages, à une fonction objective des moindres carrés généralisés (MCG-M), à une fonction d'information minimum discriminante (IMD-M) et à une fonction estimatrice du maximum de vraisemblance (EMV-M). Les critères sont:

$$\text{MCG - M:} \quad \sum_k (W_k - S_k)^2 / S_k, \tag{2a}$$

$$\text{MDI - H:} \quad (S_. - W_.) + \sum_k W_k \ln(W_k / S_k), \tag{2b}$$

$$\text{MLE - H:} \quad (W_. - S_.) - \sum_k S_k \ln(W_k / S_k). \tag{2c}$$

Dans cet article, les points au bas des lettres désignent la sommation par rapport à un indice inférieur.

Dans chaque cas, $D(\underline{W},\underline{S})$ est non négatif et est égal à zéro si et seulement si $\underline{W} = \underline{S}$. On peut le vérifier, de la façon habituelle, en examinant les dérivées partielles du premier et du second ordre de chaque expression par rapport à W_k.

Dans la section 3, nous étudierons des algorithmes permettant de calculer $\underline{W}$ de manière à minimiser ces trois critères tout en respectant la condition (1) au niveau d'approximation voulu.

2.2 Méthodes fondées sur les poids attribués aux personnes

La question peut être envisagée sous un autre angle; cette approche amène toutefois une modification légère mais importante des trois critères de distance. Les critères modifiés sont définis par les expressions (5a), (5b) et (5c) ci-dessous. Bien que ces critères servent à déterminer des poids pour les ménages, ils sont le produit d'une méthode qui, au départ, cherche à définir des poids pour les personnes. En conséquence, nous devons tout d'abord déterminer, pour les personnes, les poids qui se rapprochent le plus possible des poids initiaux des ménages auxquels appartiennent ces personnes, à condition que l'estimation pondérée du nombre de personnes dans chaque cellule de post-stratification égale le chiffre de population. Désignons les membres du k-ième ménage par $i = 1, \ldots, a_{k.}$ et définissons S_{ki} comme le poids initial de la i-ème personne dans le k-ième ménage; remarquez que $S_{ki} = S_k$.

Soit b_{kij} une variable indicatrice (0-1) qui montre si la i-ème personne du k-ième ménage se trouve dans la j-ième cellule de post-stratification. Alors, la condition pour qu'il y ait correspondance avec les chiffres de population est

$$\sum_k \sum_i b_{kij} W_{ki} = N_j. \tag{3}$$

Les trois critères pour la pondération des personnes seraient

$$\sum_k \sum_i (W_{ki} - S_{ki})^2 / S_{ki}, \tag{4a}$$

$$S_{..} - W_{..} + \sum_k \sum_i W_{ki} \ln(W_{ki}/S_{ki}), \tag{4b}$$

$$W_{..} - S_{..} - \sum_k \sum_i S_{ki} \ln(W_{ki}/S_{ki}). \tag{4c}$$

Ces critères peuvent servir à déterminer des poids pour les personnées. De fait, le critère (4c) produit des poids de post-stratification qui servent à la pondération des personnes dans l'enquête sur les dépenses des consommateurs (voir Alexander 1986). Cependant, nous cherchons ici à déterminer des poids pour les ménages. Nous pouvons déterminer de tels poids à l'aide des fonctions de critère ci-dessus si nous posons comme condition additionnelle que tous les membres d'un même ménage aient la même pondération. Par conséquent, posons $W_{ki} = W_k$ pour $i = 1, \ldots, a_k$. Sous cette condition, (3) devient

$$N_j = \sum_k \left(\sum_i b_{kij} \right) W_{ki} = \sum_k a_{kj} W_k,$$

qui est identique à l'équation (1) de la section 2.1. Les critères de distance (4a), (4b) et (4c) s'écrivent maintenant:

$$\text{MCG:} \qquad \sum_k a_{k.} \, (W_k - S_k)^2 / S_k, \tag{5a}$$

$$\text{IMD-P:} \qquad \sum_k a_{k.} \, S_k - \sum_k a_{k.} \, W_k + \sum_k a_{k.} W_k \ln \, (W_k / S_k), \tag{5b}$$

$$\text{EMV-P:} \qquad \sum_k a_{k.} \, W_k - \sum_k a_{k.} \, S_k - \sum_k a_{k.} \, S_k \ln \, (W_k / S_k). \tag{5c}$$

Les critères sont devenus des sommations en fonction des ménages mais la taille du ménage $a_{k.}$ y a été incluse dans le but de mesurer la distance entre le vecteur de poids initial et le vecteur de poids corrigé. Nous verrons que ces critères présentent des avantages par rapport à la méthode plus directe qui a servi à établir (2a), (2b) et (2c).

2.3 Méthode de la personne principale

Dans la version de base de la méthode de la personne principale, le poids de post-stratification attribué à la ''personne principale'' du ménage sert de poids pour le ménage. Pour déterminer la personne principale, il faut tout d'abord déterminer la personne repère du ménage. Pour l'intervieweur, la personne repère sera celle dont le nom figure à côté de la question suivante: ''Donnez-moi tout d'abord le nom du propriétaire ou du locataire du logement''.

On définit les rapports entre les membres du ménage en fonction de leur rapport avec la personne repère. Pour les besoins de la cause, la notion de ''personne repère'' est substituée à celle de ''chef de ménage''.

Si la personne repère est un homme marié qui vit avec sa femme, celle-ci est reconnue comme la personne principale. Autrement, la personne repère devient la personne principale. Ces critères reposent sur l'idée que la personne principale doit être une personne qui est peu susceptible d'échapper à l'enquête s'il y avait sous-dénombrement à l'intérieur du ménage. En règle générale, les femmes échappent moins souvent aux enquêtes que les hommes. En outre, le propriétaire ou le locataire d'un logement peut difficilement échapper à l'enquête.

Le principe de base de la méthode de la personne principale veut qu'il n'y ait qu'une personne principale par ménage. Ainsi, l'estimation du nombre de personnes principales suffit pour estimer le nombre de ménages. Cette méthode est utilisée dans la U.S. National Crime Survey. D'autres enquêtes, comme la U.S. Consumer Expenditure Survey ou la Current Population Survey, comportent des mesures de redressement additionnelles fondées sur des hypothèses concernant le sous-dénombrement de personnes principales à l'intérieur des ménages par comparaison au sous-dénombrement d'autres personnes dans la même cellule de post-stratification (Alexander 1986).

Il est difficile de construire un modèle théorique de la méthode de la personne principale parce que la définition de la personne repère est quelque peu arbitraire. Dans les exemples fictifs de la section 5, nous allons utiliser une version simplifiée de la méthode de la personne principale, selon laquelle la personne principale est le membre du ménage dont la cellule de post-stratification présente le meilleur taux de dénombrement, c'est-à-dire celle dont le facteur de post-stratification est le plus près de un. Scheuren (1981) applique une version semblable.

Cette version simplifiée de la méthode de la personne principale sera représentée de la façon suivante. Pour le k-ième ménage de l'échantillon, soit $j(k)$ la cellule de post-stratification à laquelle appartient la personne principale de ce ménage. Alors, le poids attribué à cette personne est

$$W_k = S_k (N_{j(k)} / \hat{N}_{j(k)}).$$

3. CALCUL DES POIDS

Les deux méthodes des moindres carrés, MCG-M et MCG-P, ont des expressions en forme analytique pour $\underline{W}$, à condition qu'il existe une solution pour l'équation (1). En ce qui concerne la méthode MCG-M, l'équation des poids corrigés est

$$\underline{W} = \underline{S} + MA(A'MA)^{-1}(\underline{N} - A'\underline{S}) \tag{6}$$

où $\underline{S} = (S_1, \ldots, S_K)$, $\underline{N} = (N_1, \ldots, N_J)$, A est la matrice (a_{kj}) et M est la matrice diagonale $K \times K$, dont la diagonale principale est constituée des éléments de $\underline{S}$. En ce qui a trait à la méthode MCG-P, les poids $\underline{W}$ sont aussi définis par l'équation (6) sauf que M est une matrice diagonale $K \times K$ dont la diagonale principale est constituée des valeurs $S_1 / a_{1.}, \ldots, S_K / a_{K.}$.

L'inconvénient de la solution (6) pour l'une ou l'autre des méthodes précitées est qu'elle peut produire des poids négatifs, ce qui est déroutant au point de vue théorique et inacceptable au point de vue pratique. Il est habituellement possible d'introduire de nouvelles conditions pour faire en sorte que les poids soient positifs. Zieschang (1986a) et Huang et Fuller (1978) montrent comment. Toutefois, l'introduction de nouvelles conditions élimine l'avantage que présente une expression en forme analytique simple.

La méthode itérative du quotient (IMD-P) a déjà servi à la pondération des ménages (voir, par exemple, Oh et Scheuren (1978a)). Une méthode connexe, qui a été largement expérimentée, est décrite dans Pugh, Tyler et George (1976), ceux-ci s'étant fondés sur l'approche de Stephan (1942). S'inspirant de Darroch et Ratcliff (1972), Luery (1986) définit un algorithme d'itération qui s'avère convergent lorsqu'il existe une solution à l'équation (1). Nous décrivons ici cette méthode étant donné que le processus itératif est facile à interpréter. Le processus débute par les poids à l'étape 0.

$$W_k(0) = S_k(N_. / \hat{N}_.)$$

Autrement dit, le poids initial S_k est multiplié par un indice de correction global qui correspond au quotient de la population connue $N_.$ par la population totale initiale pondérée. Pour les itérations suivantes, la correction est définie par

$$W_k(i) = W_k(i-1) \prod_j \left(N_j / \sum_s a_{sj} W_s(i-1) \right)^{a_{kj}/a_{k.}}$$

Vous remarquerez que $W_k(i-1)$ est multiplié par la moyenne géométrique des facteurs de stratification a posteriori qui s'appliquent aux membres du k-ième ménage, ces facteurs étant calculés à l'aide des poids de l'itération $i-1$.

Les trois autres méthodes (IMD-M, EMV-M et EMV-P) ont fait l'objet de peu d'analyses. Les algorithmes d'itération décrits ci-dessous se sont avérés efficaces dans des exemples fictifs simples comme ceux présentés dans la section 5. Dans chaque cas, on peut déterminer à l'aide des multiplicateurs de Lagrange un système d'équations que les poids doivent satisfaire pour minimiser le critère de distance assujetti aux conditions. Il n'est pas possible de résoudre directement ces équations mais s'il se trouve une méthode itérative qui produit des solutions de la forme voulue, ces solutions minimisent le critère de distance. Si les algorithmes sont convergents, les solutions satisferont les équations. Toutefois, on n'a pu vérifier de façon générale la convergence de ces algorithmes. L'approche de Haber et Brown (1986) pourrait être une solution alternative dans le cas des critères du "maximum de vraisemblance". Fagan et Greenberg (1985) proposent également des solutions à cet égard.

3.1 Solution pour IMD-M

L'équation des poids est

$$W_k = S_k \prod_j \gamma_j a_{kj} \tag{7}$$

à condition que (1) soit respectée. S'il est possible de déterminer des valeurs, $\gamma_1, \ldots, \gamma_J$ de telle sorte que les poids calculés selon (7) satisfassent l'équation (1), alors ces poids minimisent (2b) à condition que (1) soit respectée. Nous décrivons ci-dessous un algorithme d'itération qui produit un vecteur de poids $\underline{W}$.

Posons $W_k(0) = S_k$ et $\gamma_{(0)} = 1$. Puis, à la i-ème itération, posons

$$\gamma_j(i) = \gamma_j(i-1) \left[1 - (\hat{N}_j(i-1) - N_j) / \sum_s a_{sj}^2 W_s(i-1) \right],$$

où $\hat{N}_j(i-1) = \sum_s a_{kj} W_s(i-1)$. Enfin, posons $W_k(i) = S_k \prod_j (\gamma_j(i))^{a_{kj}}$.

3.2 Solution pour EMV-M

La solution a la forme

$$W_k = S_k / \left(1 + \sum_j \gamma_j a_{kj}\right),$$

à condition que (1) soit respectée.

Une solution itérative est

$$W_k(0) = S_k \qquad \text{et} \qquad \gamma_j(0) = 0,$$

$$\gamma_j(i) = \gamma_j(i-1) + (\hat{N}_j(i-1) - N_j) / \left(\sum_s (a_{sj} W_s(i-1))^2 / S_k \right),$$

$$W_k(i) = S_k / \left(1 + \sum_j \gamma_j(i) a_{kj}\right).$$

3.3 Solution pour EMV-P

La solution a la forme

$$W_k = S_k / \left(\sum_j \gamma_{kj} a_{kj} / a_{k.} \right),$$

à condition que (1) soit respectée.

Une solution itérative est

$$W_k(0) = S_k \qquad \text{et} \qquad \gamma_j(0) = 1,$$

$$\gamma_j(i) = \gamma_j(i-1) \hat{N}_j(i-1) / N_j,$$

$$W_k(i) = S_k / \left(\sum_j \gamma_j(i) a_{kj} / a_{k.} \right).$$

4. RÔLE DU "MODE DE COMPOSITION" D'UN MÉNAGE

En ce qui concerne les six méthodes fondées sur la notion de distance minimum conditionnelle, le rapport du poids initial d'un ménage à son poids corrigé dépend de la répartition des membres de ce ménage dans les diverses cellules de post-stratification. Avant de poursuivre cette analyse, nous devons définir ici le "mode de composition" d'un ménage. Deux ménages d'un échantillon (par exemple k et m) auront le même mode de composition s'ils comptent exactement le même nombre de personnes dans chacune des cellules de post-stratification, c'est-à-dire si

$$a_{kj} = a_{mj} \text{ pour } j = 1, \ldots, J. \tag{8}$$

On pourrait avoir, par exemple, un ménage constitué d'un homme de race blanche âgé de 35 à 39 ans et d'une femme de race blanche âgée de 30 à 34 ans. Il convient de souligner que le mode de composition ne dépend pas des liens de parenté.

Le rapport du poids corrigé au poids initial, W_k / S_k, est identique pour tous les ménages qui ont le même mode de composition. Autrement dit, si k et m satisfont l'équation (8), alors $W_k / S_k = W_m / S_m$. Ireland et Scheuren (1975) ont appliqué cette relation. Alexander et Roebuck (1986) en font la démonstration.

Une conséquence avantageuse de cette relation est que la catégorie de ménages peut remplacer le ménage comme unité d'analyse dans le calcul des poids pour les méthodes fondées sur la distance minimum conditionnelle. Un exemple simple nous permettra de mieux saisir les conséquences de ces observations. Supposons qu'il y a deux cellules de post-stratification: $j = 1$ pour le femmes et $j = 2$ pour les hommes. L'échantillon est formé de K ménages. Pour le ménage k, le vecteur (a_{k1}, a_{k2}) définit le nombre de femmes et d'hommes qui constituent le ménage. Un vecteur (2,1) signifie que le ménage compte deux personnes de sexe féminin et une de sexe masculin.

En pratique, un ménage ne peut dépasser une certaine taille et le nombre de catégories de ménages est aussi limité. Pour les besoins de la cause, supposons qu'aucun ménage ne compte plus de trois personnes. Nous avons donc $T = 9$ modes de composition de ménage qui correspondent aux vecteurs (1,0), (0,1), (2,0), (1,1), (0,2), (2,1), (1,2), (3,0), (0,3). Ces modes seront désignés dans l'ordre par $t = 1, \ldots, 9$. On leur attribuera aussi, dans l'ordre, les codes mnémotechniques F, H, FF, FH, HH, FFH, FHH, FFF, HHH. Le tableau 1 contient des données d'échantillon et des totaux de contrôle hypothétiques. Précisons que S est le poids total initial attribué aux ménages de la catégorie t.

Les corrections de poids selon les méthodes fondées sur la distance minimum conditionnelle peuvent être effectuées à l'aide des poids attribués aux modes de composition des ménages $S_1, \ldots, S_9$ sans tenir compte des poids des ménages proprements dits. Nous pouvons calculer les poids corrigés $W_1, \ldots, W_9$ au moyen des algorithmes présentés dans la section 3 en effectuant la sommation en fonction de t plutôt qu'en fonction de k. Ainsi, le poids corrigé d'un ménage de catégorie t est défini comme le produit de W_t / S_t par le poids initial de ce ménage. (Malgré le risque de confusion, nous avons choisi de désigner le poids des ménages par S_k et le poids total d'une catégorie de ménages par S_t pour souligner le fait que les formules définies dans les sections 2 et 3 s'appliquent aussi bien aux catégories de ménages qu'aux ménages proprement dits. Au moment des calculs, le contexte permet de faire la distinction entre S_k et S_t.)

Le fait de pouvoir utiliser des catégories de ménages comme unités d'analyse au lieu des ménages proprement dits est très avantageux pour la présentation d'exemples simples. Même lorsque l'analyse s'étend aux 48 cellules de post-stratification, il peut être encore avantageux d'utiliser les catégories de ménages. Malgré qu'il existe en théorie un nombre indéfini de catégories de ménages, le nombre de catégories incluses dans un échantillon ne peut jamais dépasser la taille de cet échantillon et est, le plus souvent, beaucoup moindre. C'est ce qu'Ireland et Scheuren (1975) ont constaté pour des cellules de ménages connexes. Le simple fait de réduire le volume des calculs en combinant les poids des ménages à personne unique d'une même catégorie peut avoir des avantages; c'est ce qu'a fait le U.S. Bureau of Labor Statistics en appliquant la méthode des moindres carrés généralisés à l'enquête sur les dépenses des consommateurs.

La version simplifiée de la méthode de la personne principale dépend aussi uniquement du mode de composition du ménage. Si deux ménages ont le même mode de composition, leurs personnes principales se trouveront dans la même cellules de post-stratification, celle dont le facteur de stratification a posteriori est le plus près de 1. On utiliserait donc le même facteur de correction pour les deux ménages. Selon la version fondamentale de la méthode de la personne principale, la personne principale est en partie définie en fonction de celui ou celle qui se trouve être la personne repère. Dans ce cas, le facteur de correction n'est donc pas entièrement déterminé par le mode de composition du ménage.

Il convient de souligner que la méthode EMV-M équivaut à calculer des estimations multinomiales du maximum de vraisemblance (moyennant le respect de la condition (1)) de p_t, $t = 1, \ldots, T$, où p_t est la proportion de ménages de la catégorie t dans la population. La méthode EMV-P peut être interprétée de la même manière. Aucun de ces modèles, qui se rattachent aussi aux méthodes MCG et IMD correspondantes, n'admet le sous-dénombrement systématique.

5. ANALYSE DES MÉTHODES

Nous allons tout d'abord formuler des hypothèses sur les propriétés des méthodes fondées sur la distance minimum conditionnelle en nous servant des résultats de la section 4 et nous allons poursuivre avec des exemples fictifs simples qui, de façon générale, semblent appuyer les hypothèses.

La première hypothèse est que les méthodes EMV-M, MCG-M et IMD-M tendront à produire des résultats similaires et que les méthodes EMV-P, MCG-P et IMD-P tendront à se rapprocher l'une de l'autre, du moins pour de grands échantillons. Cette hypothèse repose sur le fait que les estimateurs de ces méthodes sont tous de meilleurs estimateurs asymptotiques normaux selon le modèle d'échantillonnage multinomial pertinent, où les cellules représentent les catégories de ménages. En ce qui concerne les échantillons de taille faible ou moyenne, on peut s'attendre à des écarts plus marqués entre les résultats des diverses méthodes, surtout s'il y a un grand nombre de modes de composition des ménages, ce qui crée des échantillons de faible taille dans les "cellules" du modèle multinomial.

Les exemples présentés plus loin tendent à confirmer cette hypothèse; les méthodes pour les ménages, tout comme celles pour les personnes, produisent toutes des résultats très comparables, même lorsque les données hypothétiques s'ajustent mal au modèle. Toutefois, comme ces exemples ne concernent qu'un petit nombre de catégories de ménages et de cellules de post-stratification, ils servent plus à illustrer l'hypothèse qu'à la vérifier.

La seconde hypothèse repose sur une analyse qui vise à déterminer le genre de modèles d'échantillonnage où l'on peut appliquer plus particulièrement les fonctions estimatrices du maximum de vraisemblance ou des approximations asymptotiques de celles-ci. Dans ce genre de modèles, on suppose que le dénombrement est complet. On suppose aussi une distribution correspondant à des probabilités qui représentent les proportions réelles dans la population et qui sont conformes aux totaux de contrôle "réels" utilisés dans l'équation (1). Selon ce genre de modèles, les estimations initiales de l'échantillon se rapprocheraient des totaux de contrôle pour des échantillons suffisamment grands. Cela ne pourrait être le cas s'il y avait un fort taux de sous-dénombrement dans la base de sondage. D'ailleurs, le sous-dénombrement est une justification importante de la post-stratification. La question du sous-dénombrement peut prendre une importance particulière dans les enquêtes par téléphone qui ne prévoient pas de base de sondage supplémentaire pour les ménages qui n'ont pas le téléphone. Si aucune formule particulière n'est prévue pour compenser la non-réponse attribuable à la non-interview (par exemple, refus ou incapacité de répondre), la non-réponse peut être alors une autre source de sous-dénombrement.

Compte tenu de ces observations, nous pouvons maintenant formuler la seconde hypothèse: en l'absence de toute forme de compensation, les méthodes fondées sur la distance minimum conditionnelle peuvent s'avérer inefficace devant le sous-dénombrement systématique même pour de grands échantillons. Dans les modèles où on suppose un dénombrement complet, les méthodes produisent des résultats optimaux; il serait raisonnable de penser qu'elles produisent des résultats inférieurs lorsque le dénombrement n'est plus complet.

Les exemples que nous produisons ci-dessous confirment en partie cette hypothèse. Selon certaines hypothèses concernant le sous-dénombrement, les méthodes fondées sur la distance minimum conditionnelle ne sont pas aussi efficaces que la version simplifiée de la méthode de la personne principale. Selon d'autres hypothèses, certaines de ces méthodes peuvent être assez efficaces. Nous devons en conclure qu'il faut mieux connaître la nature du sous-dénombrement avant de juger si l'une ou l'autre de ces méthodes est supérieure à la méthode de la personne principale. Oh et Scheuren (1978b) soulèvent le même genre de question à propos de l'erreur quadratique moyenne de l'estimateur de la méthode itérative du quotient lorsqu'il y a sous-dénombrement.

Tableau 1

Sous-dénombrement des ménages:
Description de la population et de l'échantillon

Catégorie et description	Population réelle	Poids total initial	Poids total (W_t) pour les méthodes: MCG-M	IMD-M	EMV-M	MCG-P IMD-P EMV-P Personne principale
1: F	25,000	22,500	23,785	23,745	23,704	25,000
2: H	15,000	13,500	14,120	14,097	14,075	15,000
3: FF	7,000	6,300	7,020	7,016	7,013	7,000
4: FH	40,000	36,000	39,708	39,672	39,632	40,000
5: HH	5,000	4,500	4,913	4,906	4,900	5,000
6: FFH	12,000	10,800	12,529	12,506	12,594	12,000
7: FHH	12,000	10,800	12,408	12,428	12,449	12,000
8: FFF	0	0	0	0	0	0
9: HHH	0	0	0	0	0	0
Total	116,000	104,400	114,483	114,370	114,367	116,000

Chiffres de population:	Nombre de femmes	=	115,000
	Nombre d'hommes	=	101,000
Comptes initiaux pondérés:	Femmes	=	103,500
	Hommes	=	90,900

Nous présenterons deux exemples qui reproduisent deux formes de sous-dénombrement contraires. Dans le premier exemple (sous-dénombrement de ménages), nous supposons qu'il y a un taux uniforme de sous-dénombrement des ménages de 10% mais qu'il n'y a pas de sous-dénombrement à l'intérieur des ménages. Dans le second exemple (sous-dénombrement à l'intérieur des ménages), nous supposons qu'il y a un taux de sous-dénombrement des hommes de 10% attribuable au sous-dénombrement à l'intérieur des ménages qui comptent des personnes des deux sexes et au sous-dénombrement des ménages qui ne comptent que des personnes de sexe masculin. Pour les ménages formés d'une seule personne, le "sous-dénombrement à l'intérieur des ménages" signifie que le ménage au complet échappe à l'enquête.

Le tableau 1 décrit le premier exemple. On constatera que chaque catégorie de ménages dans l'échantillon est sous-représentée dans une proportion de 10%. Dans le cas d'un échantillon suffisamment grand, cette sous-représentation serait sûrement attribuable au sous-dénombrement systématique plutôt qu'à l'erreur d'échantillonnage. En appliquant les méthodes fondées sur la distance minimum conditionnelle et la méthode de la personne principale, nous obtenons pour chaque catégorie de ménages les poids totaux corrigés qui figurent dans les quatre dernières colonnes du tableau.

On remarquera que les méthodes MCG-P, IMD-P et EMV-P produisent toutes des poids équivalents à la population réelle. Il s'agit donc de poids "non biaisés".Comme toutes les personnes ont un facteur du second degré de 1/.9, la méthode de la personne principale produit aussi des poids équivalents à la population réelle. Les autres méthodes, MCG-M, IMD-M et EMV-M, produisent toutes des poids beaucoup trop faibles pour les ménages formés d'une seule personne et des poids beaucoup trop élevés pour les ménages formés de trois personnes. Cela est compréhensible intuitivement. Puisque ces trois méthodes n'admettent pas le sous-dénombrement systématique et qu'elles doivent expliquer la sous-représentation de personnes par l'erreur d'échantillonnage, l'explication la plus plausible est que l'échantillon comportait par hasard un nombre de ménages de trois personnes inférieur à la moyenne. Le rendement supérieur de EMV-P est compréhensible puisque cette méthode repose sur un modèle d'échantillonnage multinomial, selon lequel les personnes sont échantillonnées sans égard aux ménages.

Au point de vue pratique, cet exemple déprécie largement les méthodes MCG-M, IMD-M et EMV-M. Un taux de sous-dénombrement uniforme n'empêchera pas ces méthodes de fausser la distribution de la taille des ménages. Chose encore plus grave toutefois, la distorsion créée par l'application de ces méthodes dans l'exemple est contraire au fait généralement admis en ce qui concerne les différences de taux de dénombrement des ménages, c'est-à-dire que les ménages de faible taille sont plus susceptibles d'échapper à l'enquête que les ménages de grande taille et que, par conséquent, ils devraient avoir des poids relativement plus élevés et non des poids relativement moins élevés.

Le deuxième exemple met l'accent sur le sous-dénombrement des hommes à l'intérieur des ménages. Le cas est plus complexe que dans l'exemple précédent parce que la composition apparente d'un ménage peut être différente de sa composition réelle. Par exemple, un ménage composé d'un homme et d'une femme peut sembler être un ménage d'une seule personne. Nous allons différencier la composition réelle de la composition apparente en modifiant la notation utilisée jusqu'à maintenant. Par exemple, un ménage de type FH où l'homme n'est pas dénombré sera désigné par F[H]. Un ménage désigné par [H] ou [HH] aura tout simplement été omis. Le tableau 2 contient les données hypothétiques. La population réelle est la même que celle de l'exemple précédent.

On observe un taux de sous-dénombrement des hommes de 10% à cause du sous-dénombrement à l'intérieur des ménages et du sous-dénombrement des ménages composés uniquement de personnes de sexe masculin. Chaque homme a 10% de chances d'échapper à l'enquête.

Tableau 2

Sous-dénombrement à l'intérieur des ménages:
Description de la population et de l'échantillon

Catégorie de ménages (composition réelle)	Catégorie de ménages (composition apparente)	Population réelle	Poids totaux initiaux
1: F	F	25,000	25,000
2: H	H	13,500	13,500
	[H]	1,500	0
3: FF	FF	7,000	7,000
4: FH	FH	36,000	36,000
	F[H]	4,000	4,000
5: HH	HH	4,500	4,500
	[HH]	500	0
6: FHH	FFH	10,800	10,800
	FF[H]	1,200	1,200
7: FHH	FHH	10,800	10,800
	FH[H]	1,200	1,200
8: FFF	FFF	0	0
9: HHH	HHH	0	0
		116,000	114,000

Chiffres de population:	Nombre de femmes	115,000
	Nombre d'hommes	101,000
Comptes initiaux pondérés:	Femmes	115,000
	Hommes	90,900

Aucune des colonnes de chiffres du tableau 2 n'est observée puisqu'il n'y a pas de comptage de contrôle de ménages. En outre, on ne connaît pas le mode de composition réel des ménages auxquels appartiennent les unités de l'échantillon. Ainsi, les ménages de type F[H] (catégorie 4) passent pour des ménages de type F (catégorie 1). On observe plutôt les poids totaux initiaux des ménages qui semblent avoir un mode de composition donné. Le tableau 3 donne les poids corrigés calculés selon trois méthodes: EMV-M, EMV-P et personne principale. Comme les méthodes MCG-M et IMD-M produisent des résultats comparables à ceux de EMV-M et que les méthodes MCG-P et IMD-P produisent les mêmes résultats que EMV-P, nous n'avons pas cru nécessaire de reproduire tous ces résultats.

Les trois dernières colonnes du tableau 3 contiennent les poids totaux corrigés qui ont été calculés pour chaque catégorie de ménages (composition réelle) à l'aide des méthodes EMV-M et EMV-P et de la méthode de la personne principale. On constate que les poids obtenus par la méthode de la personne principale pour chaque catégorie "réelle" correspondent aux chiffres de la population réelle qui figurent dans la troisième colonne du tableau 1. C'est pourquoi on peut dire que la méthode de la personne principale produit des poids non biaisés.

Cet exemple reprend les hypothèses qui sous-tendent la version simplifiée de la méthode de la personne principale. Il est donc normal, en l'occurrence, que cette méthode produise des résultats très satisfaisants. Il va de soi que les ménages de type [H] ou [HH] qui échappent entièrement à l'enquête n'ont pas de poids; en contrepartie, le poids des ménages de

Tableau 3
Sous-dénombrement à l'intérieur des ménages:
Catégories et poids observés,
avec les poids corrigés obtenus selon les trois méthodes

		Poids attribué à la catégorie de ménages (composition apparente)			Poids attribué à la catégorie de ménages (composition réelle)		
Catégorie de ménages	Poids total initial	EMV-M	EMV-P	Personne principale	EMV-M	EMV-P	Personne principale
F	29,000	27,450	26,973	29,000	23,664	23,253	25,000
H	13,500	14,997	16,338	15,000	14,997	16,338	15,000
FF	8,200	7,368	7,626	8,200	6,290	6,510	7,000
FH	37,200	38,887	39,128	37,200	41,419	41,586	40,000
HH	4,500	5,623	5,446	5,000	5,623	5,446	5,000
FFH	10,800	10,661	10,885	10,800	11,739	12,001	12,000
FHH	10,800	12,605	11,878	10,800	13,859	13,140	12,000
FFF	0	0	0	0	0	0	0
HHH	0	0	0	0	0	0	0
Total	114,000	117,591	118,274	116,000	117,591	118,274	116,000

type H ou HH qui sont dénombrés est augmenté en conséquence. Le nombre total pondéré de ménages obtenu par la méthode de la personne principale est égal au nombre de ménages dans la population.

Dans cet exemple, les méthodes fondées sur la distance minimum conditionnelle surestiment le nombre total de ménages mais pondèrent trop faiblement les ménages qui ne comptent pas de personne de sexe masculin. De façon générale, ces méthodes produisent des poids trop élevés pour les ménages qui comptent des personnes de sexe masculin.

Il ne faudrait pas en conclure que la méthode de la personne principale est toujours supérieure aux méthodes fondées sur la distance minimum conditionnelle lorsqu'il y a sous-dénombrement à l'intérieur des ménages. Suivant d'autres hypothèses sur le dénombrement, elle pourrait ne pas être aussi efficace. De fait, il existe différentes versions de la méthode de la personne principale selon les enquêtes où elle est appliquée, chaque enquête ayant ses propres hypothèses sur le dénombrement. Soulignons aussi qu'il est possible de combiner la méthode de la personne principale avec la méthode itérative du quotient; à ce sujet, voir Scheuren (1981).

Même dans cet exemple, les poids biaisés qui sont produits par les méthodes fondées sur la distance minimum conditionnelle pourraient être utiles pour l'estimation de certaines caractéristiques. Si les ménages où des hommes ne sont pas dénombrés tendent à minimiser la variable d'intérêt, un poids trop élevé pour ces ménages pourrait néanmoins compenser le biais de réponse lié au sous-dénombrement à l'intérieur des ménages.

L'exemple le plus frappant de cette utilité est l'estimation du nombre total d'hommes; dans ce cas, les poids calculés selon EMV-M et EMV-P correspondent aux totaux de contrôle, ce qui n'est pas le cas des poids calculés selon la méthode de la personne principale. Toutefois, les poids biaisés ne sont pas souhaitables lorsqu'il s'agit de caractéristiques qui font peu souvent l'objet d'erreurs de déclaration attribuables au sous-dénombrement de personnes de sexe masculin (par exemple, mode d'occupation: locataire/propriétaire). L'efficacité des méthodes de pondération dans des cas comme ceux exposés ici dépend incontestablement de la nature du sous-dénombrement et de sa relation avec la variable estimée. Alexander et Roebuck (1986) traitent plus en détail le sujet et fournissent d'autres exemples.

En attendant d'autres recherches sur la couverture des enquêtes et ses effets sur la pondération, quelles recommandations pouvons-nous faire? En ce qui concerne les méthodes fondées sur la distance minimum conditionnelle, MCG-M, IMD-M et EMV-M semblent présenter peu d'intérêt à cause de l'irrégularité de leurs résultats dans le contexte d'un sous-dénombrement uniforme des ménages, ceci en dépit du fait que dans un contexte de dénombrement complet, EMV-M semble reposer sur un modèle plus sensible que EMV-P puisqu'en l'occurrence, la dernière unité d'échantillonnage est le ménage et non la personne.

La possibilité de poids négatifs nous amène à nous interroger sur l'utilité de la méthode MCG-P, malgré que certaines applications (voir, par exemple, Zieschang 1986b) comportent très peu de poids négatifs de sorte qu'on pourrait leur substituer des poids positifs sans que cela ait d'effet notable sur les estimations. Restent les méthodes IMD-P et EMV-P. Les résultats observés ne permettent pas vraiment de déterminer laquelle de ces méthodes est préférable. Si l'on s'en tient uniquement au calcul, on est porté à choisir la méthode itérative du quotient (IMD-P). Selon les quelques résultats obtenus avec les algorithmes de la section 3, la convergence est plus lente pour les méthodes EMV que pour les méthodes IMD. En outre, de nombreuses recherches ont été faites dans le but de trouver des moyens d'accroître l'efficacité de la méthode itérative du quotient dans des applications à grande échelle (voir, par exemple, Ireland et Scheuren 1975). Compte tenu des remarques précédentes, la méthode itérative du quotient (IMD-P) est celle qui semble offrir les meilleures perspectives parmi les méthodes fondées sur la distance minimum conditionnelle.

Contrairement à la méthode de la personne principale, les méthodes fondées sur la distance minimum conditionnelle produisent des poids de ménages qui correspondent aux totaux de contrôle pour les personnes. Cependant, il n'est pas du tout démontré que les méthodes fondées sur la distance minimum conditionnelle sont plus efficaces que la méthode de la personne principale en ce qui concerne le redressement en fonction du sous-dénombrement. Le sous-dénombrement est un aspect fondamental du problème de la pondération dans les enquêtes. La méthode de la personne principale est un palliatif du sous-dénombrement, qui repose sur des hypothèses très simplistes concernant la couverture des enquêtes. En revanche, nous avons vu dans la section 4 que les méthodes fondées sur la distance minimum conditionnelle pouvaient être considérées comme des estimateurs "optimaux" (c'est-à-dire, estimateurs du maximum de vraisemblance ou l'équivalent asymptotique) dans les modèles où l'on pose comme hypothèse le dénombrement complet. Il s'agit donc de choisir entre une solution optimale appliquée à un faux problème et un palliatif de ce qui peut être ou non le vrai problème. Ces observations indiquent qu'il faut manifestement pousser la recherche.

6. QUELQUES SUJETS DE RECHERCHE

6.1 Totaux de contrôle pour les ménages

S'il existait des estimations supplémentaires du nombre de ménages des diverses catégories, on pourrait appliquer les méthodes courantes de post-stratification aux estimations des ménages. On étudie actuellement la possibilité d'établir des chiffres de population de ménages selon la taille dans le cadre d'un projet qui vise à mettre à jour les données du recensement de 1980 (Das Gupta et coll. 1986). Les chiffres de population de ménages nous permettraient d'envisager d'une toute autre façon le problème de la pondération des ménages.

Les chiffres de population de ménages ne devraient pas exclure pour autant l'utilisation des chiffres de population de personnes. Les chiffres de population de ménages ne sont pas susceptibles de fournir des renseignements détaillés sur l'âge, l'origine raciale et le sexe des membres des ménages. Se servant d'une estimation du nombre total de ménages, Scheuren (1981) applique la méthode itérative du quotient pour comparer simultanément les estimations à des totaux de contrôle indépendants pour les personnes et les ménages. Zieschang

(1986b) montre comment apporter des corrections similaires à l'aide de la méthode des moindres carrés généralisés.

Les chiffres de population de ménages réunissent sans aucun doute toutes les caractéristiques nécessaires pour tenir compte des variations de taux de couverture des diverses catégories de ménages. Ils peuvent néanmoins poser des problèmes en ce qui concerne le sous-dénombrement à l'intérieur des ménages puisque leur application peut fausser le calcul de la taille réelle des ménages, ce qui conduirait à une mauvaise classification des ménages de l'échantillon à l'étape de la post-stratification.

6.2 Études relatives à la couverture

On peut évaluer assez bien le taux de couverture des personnes en comparant les estimations initiales de l'enquête (N_j) aux totaux de contrôle N_j. Il est difficile de déterminer dans quelle proportion le sous-dénombrement est attribuable à un sous-dénombrement des ménages ou à un sous-dénombrement des personnes à l'intérieur des ménages. On pourrait obtenir des renseignements supplémentaires en comparant les estimations pondérées initiales des ménages aux chiffres de population de ménages, lorsque ces chiffres seraient disponibles. Pour le moment, on pourrait comparer, pour chaque catégorie de ménages, les estimations d'enquête de 1980 et les chiffres correspondants du recensement de la même année.

Cependant, les renseignements supplémentaires ne nous permettent pas de faire parfaitement la distinction entre le sous-dénombrement de ménages et le sous-dénombrement à l'intérieur des ménages; pour cela, nous devons absolument poser de nouvelles hypothèses. Alexander et Roebuck (1986) proposent certaines solutions préliminaires pour ajuster des données d'enquête et de recensement à une série de modèles de dénombrement. Wolter (1986) décrit d'autres façons de construire un modèle de dénombrement.

6.3 Estimation de variances

En ce qui concerne la plupart des méthodes fondées sur la distance minimum conditionnelle, on n'a pas cherché à mettre au point des méthodes d'estimation de la variance des estimateurs pondérés. Seule la méthode itérative du quotient a fait l'objet de telles études; voir Arora et Brackstone (1977), Bankier (1978) et Fan et coll. (1981).

Dans tous les cas, des méthodes d'itération pourraient servir à estimer la variance des estimateurs pondérés. Ce genre de méthodes s'est avéré raisonnablement efficace dans des conditions assez générales; voir, par exemple, Krewski et Rao (1981). Il reste à déterminer si ces conditions peuvent s'appliquer aux méthodes fondées sur la distance minimum conditionnelle.

6.4 Applications

Zieschang (1986b) a appliqué les méthodes des moindres carrés généralisés à l'enquête sur les dépenses des consommateurs des États-Unis. Scheuren (1981) décrit une application à grande échelle de la méthode itérative du quotient à la pondération des ménages. Les algorithmes du maximum de vraisemblance (EMV-M et EMV-P) n'ont pas été soumis à des applications de ce genre. S'ils devaient servir à la pondération dans les enquêtes, on pourrait devoir se livrer à des recherches en vue d'accroître leur applicabilité.

REMERCIEMENTS

L'auteur tient à exprimer sa reconnaissance à Michael J. Roebuck pour l'aide qu'il lui a apportée dans cette étude, de même qu'au rédacteur associé et aux arbitres pour leurs commentaires utiles.

L'auteur tient également à exprimer sa gratitude à Brenda Kelly pour son assiduité dans la préparation de la copie dactylographiée.

BIBLIOGRAPHIE

ALEXANDER, C.H. (1986). The present Consumer Expenditure Surveys weighting method. In *Population Controls in Weighting Sample Units,* Section 7, Washington D.C.: U.S. Bureau of Labor Statistics, 1-32.

ALEXANDER, C.H., et ROEBUCK, M.J. (1986). Comparison of alternative methods for household estimation. *Proceedings of the Section on Survey Research, American Statistical Association,* 54-64.

ARORA, H.R., et BRACKSTONE, G.J. (1977). An investigation of the properties of raking ratio estimates: II. With cluster sampling. *Survey Methodology,* 4, 232-252.

BANKIER, M.D. (1978). An estimate of the efficiency of raking ratio estimators under sample random sampling. *Survey Methodology,* 4, 115-124.

BRACKSTONE, G.J., et RAO, J.N.K. (1979). An investigation of raking ratio estimators. *Sankhyā,* Série C, 41, 97-114.

BISHOP, Y.M.M., FIENBERG, S.W., et HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Massachusetts: MIT Press.

CRESSIE, N., et READ, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society,* Série B, 46, 440-464.

DARROCH, J.N., et RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics,* 63, 1470-1480.

DAS GUPTA, P., GIBSON, C., HERRIOT, R.A., LAMAS, E., et ZITTER M. (1986). New approaches to estimating households and their characteristics for states and counties. Article présenté au congrès annuel du Population Association of America de 1986.

DEMING, W.E., et STEPHAN, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected margins are known, *Annals of Mathematical Statistics,* 11, 427-444.

FAGAN, J.T., et GREENBERG, B. (1985). Algorithms for making tables additive: raking, maximum likelihood, and minimum chi-square. Statistical Research Division Report Series No. Census/SRD/RR-85/12, U.S. Bureau of the Census,

FAN, M.C., WOLTMAN, H.F., MISKURA, S.M., et THOMPSON, J.H. (1981). 1980 census variance estimation procedure. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 176-181.

FIENBERG, S.E. (1986). Comments on some estimation problems in the Consumer Expenditure Surveys. Dans *Population Controls in Weighting Sample Units.* Section 5, Washington, D.C.: U.S. Bureau of Labour Statistics, 1-12.

HABER, M., et BROWN, M.B. (1986). Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints, *Journal of the American Statistical Association,* 81, 477-482.

HUANG, E.T., et FULLER, W. (1978). Nonnegative regression estimation for sample survey data. *American Statistical Association Proceedings of Social Statistics Section,* 300-305.

IRELAND, C.T., et SCHEUREN, F.J. (1975). The rake's progress, Dans *Computer Programs for Contingency Table Analysis,* Washington, D.C.: The George Washington University, 155-216.

KREWSKI, D., et RAO, J.N.K. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Annals of Statistics* 9, 1010-1019.

LUERY, D.M. (1986). Weighting sample survey data under linear constraints on the weights. *Proceedings of the Social Statistics Section, American Statistical Association,* 325-330.

OH, H.L., et SCHEUREN, F.J. (1978a). Multivariate raking ratio estimation in the 1973 Exact Match Study. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 716-722.

OH, H.L., et SCHEUREN, F.J. (1978b). Some unresolved application issues in raking ratio estimation. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 723-725.

PUGH, R.E., TYLER, B.S., et GEORGE, S. (1976). Computer-based procedure for N-dimensional adjustment of data - NJUST. Staff Paper No. 24, U.S. Social Security Administration.

SCHEUREN, F.J. (1981). Methods of estimation for the 1973 exact match study. Dans *Studies from Interagency Data Linkages, Report No. 10,* U.S. Department of Health and Human Services, U.S. Social Security Administration, 9-122.

STEPHAN, F.F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *Annals of Mathematical Statistics,* 13, 166-178.

WOLTER, K.M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association,* 81, 338-346.

ZIESCHANG, K.D. (1986a). Generalized least squares: an alternative to principal person weighting. *Population Controls in Weighting Sample Units,* Section 2, Washington, D.C.: U.S. Bureau of Labor Statistics, 1-41.

ZIESCHANG, K.D. (1986b). A generalized least squares weighting system for the Consumer Expenditure Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 64-71.