LES VARIANCES D'ESTIMATEURS ASYMPTOTIQUEMENT NORMAUX BASES SUR DES ENQUÊTES COMPLEXES

David A. Binder[1]

Nous discutons du problème de la spécification et de l'estimation de la variance de paramètres estimés basés sur des plans d'échantillonnage complexes provenant de populations finies. Les résultats présentés dans cet article sont particulièrement utiles lorsque les estimateurs des paramètres ne sont pas définis explicitement comme étant une fonction des autres statistiques de l'échantillon. Nous montrons comment des résultats peuvent s'appliqer à la régression linéaire, à la régression logistique et aux modèles linéaires logarithmiques de tables de contingence.

I. INTRODUCTION

Nous avons observé ces dernières années une tendance croissante à utiliser des données d'enquête pour estimer les paramètres de modèles classiques, tels que les paramètres de régression, les fonctions discriminantes et les paramètres de logit et de probit. Toutefois, dans nombre d'enquêtes, l'objet premier est l'estimation des valeurs moyennes, des valeurs totales, des tendances et autres caractéristiques d'une population ou d'une

1 D.A. Binder, Division des méthodes d'enquête-institution et agriculture, Statistique Canada.

sous-population. Pour cette raison et pour d'autres considérations d'ordre pratique, le plan d'enquête comprend rarement un échantillon aléatoire simple, mais plutôt un échantillon stratifié, souvent à plusieurs degrés, pouvant présenter des probabilités inégales à certains degrés.

Aussi, on s'est beaucoup interrogé sur le bien-fondé de l'utilisation de coefficients de pondération de l'échantillonnage dans les inférences faites au sujet de ces paramètres de modèles (voir notamment Särndal, 1978). Pour pouvoir répondre à une telle question, il faut d'abord déterminer si un modèle de superpopulation convient à toutes les unités de la population. Dans l'affirmative, les inférences portant sur les paramètres de la superpopulation constituent souvent la première occupation. Il s'agit alors d'inférences basées sur des modèles, c'est-à-dire que, pour un échantillon donné, elles ne dépendent pas des coefficients de pondération de l'échantillonnage.

La question qui nous vient à l'esprit est la suivante, si le modèle de superpopulation ne convient pas, à quoi correspondent les paramètres que nous estimons? Il nous faut admettre que dans de nombreuses études, particulièrement dans le domaine des sciences sociales, le modèle utilisé (par exemple la régression linéaire) n'est qu'une approximation utile de l'univers réel et que les paramètres de ce modèle (par exemple les corrélations et les corrélations partielles) ont plus souvent pour objet de faciliter la compréhension des interdépendances approximatives des variables que de donner lieu à une interprétation scientifique particulière. Par conséquent, les paramètres que nous estimons ne se rapportent pas nécessairement à un modèle de superpopulation réel; ils ont plutôt un caractère descriptif.

Dans cet article, nous partons du principe que nous cherchons à faire des inférences au sujet de ces paramètres "descriptifs" de la population. À titre d'exemple, supposons que $\underset{\sim}{X}$ et $\underset{\sim}{Y}$ désignent respectivement des matrices N x p et N x 1, chaque ligne de $\underset{\sim}{X}$ et $\underset{\sim}{Y}$ correspondant à un individu différent de la population. Nous nous intéressons au paramètre descriptif, $\underset{\sim}{B}$, un vecteur pxl qui satisfait l'équation:

Les paramètres examinés dans cet article ne sont pas définis par une équation explicite comme l'équation (1.3), mais ils sont plutôt définis implicitement sous la forme $\underset{\sim}{U}(\underset{\sim}{Z},\underset{\sim}{\theta}) = 0$. Un exemple simple d'une telle distinction pourrait être le paramètre du rapport :

$$R = \frac{\Sigma Y_k}{\Sigma X_k},$$

qu'on pourrait aussi définir implicitement sous la forme

$$\Sigma Y_k - R\Sigma X_k = 0.$$

Pour certains types de modèles, comme les modèles linéaires logarithmiques indirects ou les modèles de régression logistique, les paramètres peuvent êtres définis seulement en fonction de relation implicites.Bien qu'elle n'apparaise encore sons une forme générale dans la literature l'extension des résultats de Tepping (1968) est relativement simple àppliquer dans ce cas. Il existe toutefois quelques exemples précis de son application,voir notamment Fuller (1975) et Freeman et Koch (1976).

Le cadre général et les principaux résultats de notre étude sont exposés dans la deuxième partie du présent article et des exemples d'un certain nombre de modèles sont présentés dans la troisième partie.

2. CADRE GÉNÉRAL

2.1 Cadre

Les unités de la population sont désignées 1, ..., N. Un vecteur de données q-dimensionnel X_i est associé à la i-ième unité. Nous avons un espace des paramètres $\Theta \subseteq R^p$. Le paramètre $\underset{\sim}{\theta}_o = (\theta_{1o}, \ldots, \theta_{po})$ est défini par les équations p:

$$\underset{\sim}{X}^T \underset{\sim}{X} \underset{\sim}{B} = \underset{\sim}{X}^T \underset{\sim}{Y} \tag{1.1}$$

Cette conception des paramètres descriptifs est la même que celle qui a été considérée par Frankel (1971) et Kish et Frankel (1974)

Pour l'estimation de tels paramètres, on tient compte habituellement des coefficients de pondération de l'échantillonnage. Si nous indiquons par Π_i la probabilité que la i-ième unité de l'échantillon soit prélevée et si $\underset{\sim}{\Pi} = \text{diag}\ (\pi_1, \ldots, \pi_n)$, alors l'estimateur pondéré du paramètre $\underset{\sim}{B}$ satisfait:

$$\underset{\sim}{x}^T \underset{\sim}{\Pi}^{-1} \underset{\sim}{x} \underset{\sim}{B} = \underset{\sim}{x}^T \underset{\sim}{\Pi}^{-1} \underset{\sim}{y} \tag{1.2}$$

où $\underset{\sim}{x}$ et $\underset{\sim}{y}$ représentent respectivement des matrices nxp et nxl dont les lignes correspondent aux lignes échantillonnées de $\underset{\sim}{X}$ et de $\underset{\sim}{Y}$.

Supposons maintenant qu'un estimateur d'un paramètre d'une population peut être défini de la façon suivante:

$$\hat{\theta} = g(z_1, \ldots, z_k), \tag{1.3}$$

où $E(z_i) = Z_i$. Dans cette équation, $\hat{\theta}$ est un estimateur de $g(Z_i, \ldots Z_k)$. Selon Tepping (1968) et Woodruff (1971), un développement de la série de Taylor pour $\hat{\theta}$ donne:

$$V[\hat{\theta}] \doteq V[\sum_{i=1}^{k} (\frac{\partial g}{\partial Z_i}) (z_i - Z_i)] \, . \tag{1.4}$$

Des exemples de ces formules appliquées à l'estimation des coefficients de régression (1.1) ont été fournis par Tepping (1968). Toutefois, pour le calcul des variances des coefficients de régression, les expressions qui découlent de l'équation (1.4) sont un peu plus complexes que celles qui ont été dérivées par Fuller (1975).

$$U_i(\underset{\sim}{X},\underset{\sim}{\theta}_o) = \sum_{k=1}^{N} u_i(\underset{\sim}{X}_k,\underset{\sim}{\theta}_o) - v_i(\underset{\sim}{\theta}_o) = 0 , \tag{2.1}$$

pour i=1, ..., p. Nous supposons que les équations (2.1) définissent θ_o uniquement dans Θ. Nous supposons également que $\partial u_i(\underset{\sim}{X},\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}$ et $\partial v_i(\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}$ existent dans le voisinage de $\underset{\sim}{\theta}_o$. Un exemple simple de (2.1) est lorsque θ_o représente un total de la population et que nous avons $U(\underset{\sim}{X},\theta_o) = \sum_{k=1}^{N} X_k - \underset{\sim}{\theta}_o$. Dans ce cas, $u(X_k,\theta_o) = X_k$ et $v(\theta_o) = \theta_o$.

Nous tirons un échantillon d'unités selon une distribution de probabilité définie en sur l'ensemble de tous les sous-ensembles non vides de {1, ..., N}. Nous désignons par $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ les valeurs choisies de $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_N$. Nous supposons que pour tout $\underset{\sim}{\theta} \in \Theta$, nous pouvons construire un estimateur convergent asymptotiquement normal de $U_i(X,\underset{\sim}{\theta})$. Nous désignons cet estimateur par $\hat{U}_i(\underset{\sim}{x},\underset{\sim}{\theta})$. Par exemple, pour nombre de plans d'échantillonnage sans remise,

$$\hat{U}_i(\underset{\sim}{x},\underset{\sim}{\theta}) = \sum_{k=1}^{n} (\underset{\sim}{x}_k,\underset{\sim}{\theta})/\pi_k - v_i(\underset{\sim}{\theta}) \tag{2.2}$$

est un estimateur convergent asymptotiquement normal, où π_k représente la probabilité d'inclusion de la k-ième unité.

Posons $\sigma_{ij}(\underset{\sim}{X},\underset{\sim}{\theta}) = \mathrm{Cov}[\hat{U}_i(\underset{\sim}{x},\underset{\sim}{\theta}), \hat{U}_j(\underset{\sim}{x},\underset{\sim}{\theta})]$. Par exemple, pour l'estimateur (2.2), nous constatons que

$$\sigma_{ij}(\underset{\sim}{X},\underset{\sim}{\theta}) = \sum_{k=1}^{N} \sum_{\ell=1}^{N} u_i(\underset{\sim}{X}_k,\underset{\sim}{\theta})\, u_j(\underset{\sim}{X}_\ell,\underset{\sim}{\theta})\, (\pi_{k\ell}-\Pi_k\pi_\ell)/\pi_k\pi_\ell , \tag{2.3}$$

où $\pi_{k\ell}$ représente la probabilité que la k-ième et la ℓ-ième unité soient dans l'échantillon.

Soit $\underset{\sim}{\Sigma}(\underset{\sim}{X},\underset{\sim}{\theta})$ la matrice pxp avec les élements $\sigma_{ij}(\underset{\sim}{X},\underset{\sim}{\theta})$, et $\hat{\underset{\sim}{\Sigma}}(\underset{\sim}{x},\underset{\sim}{\theta})$ un estimateur convergent de $\underset{\sim}{\Sigma}$. Maintenant, pour tout $\underset{\sim}{\theta}$ donné,

$$U_i(\underset{\sim}{X},\underset{\sim}{\theta}) + v_i(\underset{\sim}{\theta}) = \sum_{k=1}^{N} u_i(\underset{\sim}{X}_k,\underset{\sim}{\theta}),$$

de sorte que les estimateurs $\hat{U}_i(\underset{\sim}{X},\underset{\sim}{\theta})$ et $\hat{\underset{\sim}{\Sigma}}(\underset{\sim}{x},\underset{\sim}{\theta})$ peuvent être spécifiés pour tout plan dans lequel nous pouvons dériver des estimateurs convergents asymptotiquement normaux des valeurs totales de la population, de même que des estimateurs convergents des variances des estimateurs des valeurs totales.

L'estimateur de Horvitz-Thompson pour (2.3) est

$$\sum_{k=1}^{n}\sum_{\ell=1}^{n} u_i(\underset{\sim}{x}_k,\underset{\sim}{\theta})\ u_j(\underset{\sim}{x}_\ell,\underset{\sim}{\theta})(\pi_{k\ell} - \pi_k\pi_\ell)/\pi_k\pi_\ell\pi_{k\ell}. \tag{2.4}$$

Dans le cas d'un échantillon de taille déterminée, l'estimateur de Yates-Grundy pour (2.3) est:

$$\sum_{k<\ell}\sum{}' \left[\frac{u_i(\underset{\sim}{x}_k,\underset{\sim}{\theta})}{\pi_k} - \frac{u_i(\underset{\sim}{x}_\ell,\underset{\sim}{\theta})}{\pi_\ell}\right]\left[\frac{u_j(\underset{\sim}{x}_k,\underset{\sim}{\theta})}{\pi_k} - \frac{u_j(\underset{\sim}{x}_\ell,\underset{\sim}{\theta})}{\pi_\ell}\right](\pi_k\pi_\ell - \pi_{k\ell}). \tag{2.5}$$

Soit $\underset{\sim}{U}(\underset{\sim}{X},\underset{\sim}{\theta})$ et $\hat{\underset{\sim}{U}}(\underset{\sim}{x},\underset{\sim}{\theta})$ les vecteurs p-dimensionnels dont les composantes sont respectivement $U_i(\underset{\sim}{X},\underset{\sim}{\theta})$ et $\hat{U}_i(\underset{\sim}{x},\underset{\sim}{\theta})$, nous établissons que

$$\underset{\sim}{J}(\underset{\sim}{X},\underset{\sim}{\theta}) = \partial\underset{\sim}{U}(\underset{\sim}{X},\underset{\sim}{\theta})/\partial\underset{\sim}{\theta} \tag{2.6}$$

$$\hat{\underset{\sim}{J}}(\underset{\sim}{x},\underset{\sim}{\theta}) = U(\underset{\sim}{x},\theta)/\partial\underset{\sim}{\theta} \tag{2.7}$$

où $\underset{\sim}{J}$ et $\hat{\underset{\sim}{J}}$ sont des matrices de dérivées partielles pxp. Supposons que les matrices sont des fonctions continues de $\underset{\sim}{\theta}$ et que les dérivées partielles relatives à existent dans le voisinage de $\underset{\sim}{\theta}_o$. Supposons égalemet que

$\hat{J}(\underset{\sim}{x},\underset{\sim}{\theta})$ est un estimateur convergent de $\underset{\sim}{J}(\underset{\sim}{X},\underset{\sim}{\theta})$.

Notre estimateur de $\underset{\sim}{\theta}$ est donné par $\hat{\underset{\sim}{\theta}}$, la solution à:

$$\hat{U}_i(\underset{\sim}{x},\hat{\underset{\sim}{\theta}}) = 0, \text{ puisque } i=1, \ldots, p. \tag{2.8}$$

Nous suppons que la taille de l'échantillon est suffisamment grande pour que la solution de l'équation (2.8) soit unique dans Θ. Nous verrons dans la prochaine section que la matrice des covariances de $\hat{\underset{\sim}{\theta}}$ peut être estimée de façon convergente par:

$$[\hat{\underset{\sim}{J}}^{-1}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})]\ \hat{\underset{\sim}{\Sigma}}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})\ \ \hat{\underset{\sim}{J}}^{-1}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})]^T .$$

2.2 Théorie asymptotique

Conformément aux raisonnements asymptotiques de Madow (1948) et Hajek (1960), nous étudions une séquence de populations avec l'indice t, de tailles $N^{(t)}$ et à partir de données $\underset{\sim}{X}^{(t)}$. Nous supposons que $N^{(t)} \to \infty$ quand $t \to \infty$. Pour la population t, nous tirons un échantillon de taille $n^{(t)}$ et faisons porter nos observations sur les données $\underset{\sim}{x}^{(t)}$. Nous posons que $\nu^{(t)} = E(n^{(t)})$ et nous supposons que $\lim_{t\to\infty} \nu^{(t)} = \infty$ et $\lim_{t\to\infty} (N^{(t)} - \nu^{(t)}) = \infty$.

Pour tout $\underset{\sim}{\theta}$ dans le voisinage de $\underset{\sim}{\theta}_o^{(t)}$, nous supposons que

$$[\nu^{(t)}]^{\frac{1}{2}}\ [\hat{\underset{\sim}{U}}(\underset{\sim}{x}^{(t)}, \underset{\sim}{\theta}) - \underset{\sim}{U}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})]/N^{(t)}$$

suit asymptotiquement une loi $N[0,\underset{\sim}{S}(\underset{\sim}{\theta})]$, où

$$\underset{\sim}{S}(\underset{\sim}{\theta}) = \lim[\nu^{(t)}\ \underset{\sim}{\Sigma}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})/\{N^{(t)}\}^2]$$

existe. Nous supposons également l'existence de

$$\underset{\sim}{K}(\underset{\sim}{\theta}) = \lim \underset{\sim}{J}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})/N^{(t)} \quad \text{et aussi}$$

$$\text{plim } \hat{\underset{\sim}{J}}(\underset{\sim}{x}^{(t)},\underset{\sim}{\theta})/N^{(t)} = \underset{\sim}{K}(\underset{\sim}{\theta}).$$

Supposons en outre que

$$\lim\ [\text{rang}\ \{\underset{\sim}{J}(\underset{\sim}{X}^{(t)},\underset{\sim}{\theta})\}] = \text{plim}\ [\text{rang}\{\underset{\sim}{J}\ (\underset{\sim}{x}^{(t)}\underset{\sim}{\theta})\}] = p.$$

Nous définissons $\hat{\underset{\sim}{\theta}}^{(t)}$ pour satisfaire

$$\hat{U}(\underset{\sim}{x}^{(t)},\ \hat{\underset{\sim}{\theta}}^{(t)}) = 0\ .$$

Grâce à un développement de la série de Taylor, nous obtenouns

$$\underset{\sim}{U}(\underset{\sim}{x}^{(t)},\ \underset{\sim}{\theta}_{o}^{(t)} \doteq -\ \hat{J}(\underset{\sim}{x}^{(t)},\ \hat{\underset{\sim}{\theta}}^{(t)})\ (\underset{\sim}{\theta}^{(t)} - \underset{\sim}{\theta}_{o}^{(t)}\ . \tag{2.9}$$

Comme le premier membre de l'équation (2.9) est asymptotiquement normal, il s'ensuit que

$$(n^{(t)})^{\frac{1}{2}}\ (\underset{\sim}{\theta}^{(t)} - \underset{\sim}{\theta}_{o}^{(t)})$$

Soit asymptotiquement une loi $N[\underset{\sim}{0},\ \underset{\sim}{G}(\underset{\sim}{\theta}_{o})]$, où

$$\underset{\sim}{S}\ (\underset{\sim}{\theta}) = \underset{\sim}{K}(\underset{\sim}{\theta}_{o})\ \underset{\sim}{G}(\underset{\sim}{\theta}_{o})[\underset{\sim}{K}(\underset{\sim}{\theta}_{o})]^{T}\ .$$

Par conséquent,

$$\underset{\sim}{G}(\underset{\sim}{\theta}_{o}) = [\underset{\sim}{K}^{-1}(\underset{\sim}{\theta}_{o})]\ \underset{\sim}{S}(\underset{\sim}{\theta}_{o})\ [\underset{\sim}{K}^{-1}(\underset{\sim}{\theta}_{o})]^{T} \tag{2.10}$$

et

$$n^{(t)}[\hat{\underset{\sim}{J}}^{-1}\ (\underset{\sim}{x},\hat{\underset{\sim}{\theta}})]\ \hat{\underset{\sim}{\Sigma}}\ (\underset{\sim}{x},\hat{\underset{\sim}{\theta}})\ [\hat{\underset{\sim}{J}}^{-1}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})]^{T}. \tag{2.11}$$

est un estimateur convergent de $\underset{\sim}{G}(\underset{\sim}{\theta}_{o})$

En conclusion, lorsque la forme fonctionnelle de $\hat{\underset{\sim}{U}}(\underset{\sim}{x},\underset{\sim}{\theta})$ et de $\hat{\underset{\sim}{\Sigma}}(\underset{\sim}{x},\underset{\sim}{\theta})$ est spécifiée, il ne nous reste plus qu'à dériver la matrice $\hat{\underset{\sim}{J}}(\underset{\sim}{X},\underset{\sim}{\theta}_{o})$ et son estimateur $\hat{\underset{\sim}{J}}(\underset{\sim}{x},\hat{\underset{\sim}{\theta}})$ pour pouvoir utiliser ces résulats.

3. EXEMPLES

3.1 Introduction

Dans la présente partie, nous analysons en détail les applications des énoncés généraux formulés dans la deuxième partie au sujet de l'estimation des variances de certains estimateurs des paramètres de la population. Nous examinerons tout particulièrement les rapports, les coefficients de régression et les modèles linéaires logarithmiques se rapportant à des données qualitatives. D'autres modèles comme les modèles de probit pourraient être analysés de façon analogue.

Nous nous fondons généralement sur les notations suivantes. Si $\underset{\sim}{W}_1, \ldots, \underset{\sim}{W}_N$ représentent des valeurs de la population et que $\underset{\sim}{W} = \Sigma \underset{\sim}{W}_k$, lorsque nous tirons un échantillon $\underset{\sim}{w}_1, \ldots, \underset{\sim}{w}_n$, nous avons donc un estimateur sans biais de $\underset{\sim}{W}$ indiqué par $\hat{\underset{\sim}{W}}$. Nous supposons que $\underset{\sim}{V}(\hat{\underset{\sim}{W}})$ représente la matrice des covariances de $\hat{\underset{\sim}{W}}$ et que $\hat{\underset{\sim}{V}}(\hat{\underset{\sim}{W}})$ est un estimateur convergent de $\underset{\sim}{V}(\hat{\underset{\sim}{W}})$. La forme particulière que prendra cet estimateur dépend du plan de d'échantillonage, par exemple, stratifié à plusieurs degrés, ppt avec remise, etc..

3.2 Rapports

Supposons que nous nous intéressions au rapport $R = \Sigma X_{k2}/\Sigma X_{k1}$. Nous définissons

$$U(\underset{\sim}{X},R) = \Sigma X_{k2} - R\Sigma X_{k1}.$$

Par conséquent, dans le cas d'un échantillonnage sans remise, nous observons que

$$\hat{U}(\underset{\sim}{x},R) = \hat{X}_2 - R\hat{X}_1.$$

Si $\hat{U}(\underset{\sim}{x},\underset{\sim}{\hat{R}}) = 0$, nous obtenons

$$\hat{R} = \hat{X}_2/\hat{X}_1. \tag{3.1}$$

Nous posons que $W_k = X_{k2} - R\, X_{k1}$.

Comme $J\,(\underset{\sim}{X},R) = -\Sigma X_{k1}$, nous pouvons affirmer que $(V(\hat{R})$ peut s'approximer par $V(\hat{W})/(\Sigma X_{k1})^2$, estimé par $\hat{V}(\hat{W})/\hat{X}_1^2$. Dans le cas d'un é'chantillonage stratifé, le résultat obtenu serait le même que dans l'étude de Woodruff (1971).

3.3 Coefficients de régression et R

Supposons que notre matrice de données $\underset{\sim}{X}$ est répartie en $[\underset{\sim}{Z}|\underset{\sim}{Y}]$, la première colonne de $\underset{\sim}{Z}$ étant le vecteur des 1, $\underset{\sim}{Y}$ étant un vecteur par Nx1. Les paramètres d'intérêt θ, $\underset{\sim}{B}$, et R^2 sont définis par:

$$U_1 = \theta - \underset{\sim}{Y}^T \underset{\sim}{1} = 0, \tag{3.2a}$$

$$\underset{\sim}{U}_2 = \underset{\sim}{Z}^T \underset{\sim}{Z}\, \underset{\sim}{B} - \underset{\sim}{Z}^T \underset{\sim}{Y} = 0, \tag{3.2b}$$

$$U_3 = (\underset{\sim}{Y}^T\underset{\sim}{Y}-N^{-1}\theta^2)(R^2-1) + \underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{Y}^T \underset{\sim}{Z}\, \underset{\sim}{B} = 0. \tag{3.2c}$$

Dans cet exemple, $\underset{\sim}{B}$ représente le vecteur des coefficients de régression, R^2 le coefficient de détermination multiple et θ, le total de tous les Y. Prenons d'abord le cas où N est connu. Nous supposons que $SSY = \underset{\sim}{Y}^T\underset{\sim}{T} - N^{-1}\theta^2$. Définissons également, $\underset{\sim}{S}_{ZZ}$ l'estimateur de $\underset{\sim}{Z}^T\underset{\sim}{Z}$, S_{YY}, l'estimateur de $\underset{\sim}{Y}^T\underset{\sim}{Y}$ et $\underset{\sim}{S}_{ZY}$, l'estimateur de $\underset{\sim}{Z}^T\underset{\sim}{Y}$. Par conséquent, nous avons:

$$\hat{\theta} = \hat{Y}\ , \tag{3.3a}$$

$$\underset{\sim}{\hat{B}} = \underset{\sim}{S}_{ZZ}^{-1}\, \underset{\sim}{S}_{ZY}\ , \tag{3.3b}$$

$$\hat{R}^2 = 1 - \frac{S_{YY} - \hat{\underset{\sim}{B}}^T \underset{\sim}{S}_{ZY}}{S_{YY} - N^{-1}\hat{Y}^2} , \tag{3.3c}$$

et

$$\underset{\sim}{J} = \partial\underset{\sim}{U}(\underset{\sim}{Z},\underset{\sim}{Y},\underset{\sim}{B},R^2,\theta)/\partial(\underset{\sim}{B},\hat{R},\theta) = \begin{bmatrix} \underset{\sim}{0}^T & 0 & 1 \\ \underset{\sim}{Z}^T\underset{\sim}{Z} & \underset{\sim}{0} & \underset{\sim}{0} \\ -\underset{\sim}{Y}^T\underset{\sim}{Z} & SSY & 2\bar{Y}(1-R^2) \end{bmatrix},$$

où $\bar{Y} = \theta/N$.

Il résulte donc que

$$\underset{\sim}{J}^{-1} = \begin{bmatrix} \underset{\sim}{0} & (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1} & \underset{\sim}{0} \\ -2\bar{Y}(1-R^2)/SSY & B^T/SSY & 1/SSY \\ 1 & \underset{\sim}{0}^T & 0 \end{bmatrix} .$$

Si nous supposons maintenant que $\underset{\sim}{W}_k^T(B) = (Z_{k1}\, e_k, \ldots, Z_{kp}\, e_k)$, où $e_k = Y_k - \sum_j Z_{kj}\, B_j$, nous obtenons:

$$\underset{\sim}{V}[\hat{\underset{\sim}{B}}] \doteq (\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1}\, \underset{\sim}{V}[\hat{\underset{\sim}{W}}(\underset{\sim}{B})](\underset{\sim}{Z}^T\underset{\sim}{Z})^{-1} . \tag{3.4}$$

Il s'agit là d'une conséquence directe de l'équation (2.10). Notons que l'ensemble des vecteurs de $\underset{\sim}{W}_k(\underset{\sim}{B})$ correspond à $\underset{\sim}{U}_2$ dans l'équation (3.2b). Fuller (1975) obtient le même résultat avec un échantillon stratifié ou un échantillon stratifié à deux degrés.

Pour estimer (3.4), nous utilisons:

$$\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{B}}] = \underset{\sim}{S}_{ZZ}^{-1}\, \hat{\underset{\sim}{V}}[\hat{\underset{\sim}{W}}(\hat{\underset{\sim}{B}})]\, \underset{\sim}{S}_{ZZ}^{-1} .$$

En supposant que $\underset{\sim}{B}$ est l'estimateur du maximum de vraisemblance de $\underset{\sim}{\beta}$, alors $\underset{\sim}{B}$ satisfait

$$\underset{\sim}{U} = \underset{\sim}{Z}^T \underset{\sim}{P}(\underset{\sim}{B}) - \underset{\sim}{Z}^T \underset{\sim}{Y} = 0 , \tag{3.8}$$

où $\underset{\sim}{P}(\underset{\sim}{B})^T = [p_1(\underset{\sim}{B}), \ldots, p_N(\underset{\sim}{B})]$.

Pour une population finie donnée, nous définissons $\underset{\sim}{B}$ comme notre paramètre d'intérêt.

Nous choisissons $\underset{\sim}{C}(\underset{\sim}{B})$ comme notre estimateur de $\underset{\sim}{Z}^T\underset{\sim}{P}(\underset{\sim}{B})$ et $\underset{\sim}{S}_{ZY}$ comme notre estimateur de $\underset{\sim}{Z}^T\underset{\sim}{Y}$. Par conséquent, $\hat{\underset{\sim}{B}}$ satisfait $\underset{\sim}{C}(\hat{\underset{\sim}{B}}) = \underset{\sim}{S}_{ZY}$. Ces équations doivent habituellement être résolues de façon itérative. Posons également

$$\underset{\sim}{J} = \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{B}} .$$

La (i, j)-ième composante de $\underset{\sim}{J}$ est $\sum_k Z_{ki}\, Z_{kj}\, p_k\, (\underset{\sim}{B})\, [1-p_k(\underset{\sim}{B})]$. Nous notons l'estimateur de $\underset{\sim}{J}$ par $\hat{\underset{\sim}{J}}$.
Pour estimer la variance de $\hat{\underset{\sim}{B}}$, nous posons:

$$\underset{\sim}{W}_k^T = (Z_{k1}\, \hat{e}_k, \ldots, Z_{kr}\, \hat{e}_k)$$

où $\hat{e}_k = p_k\, (\hat{\underset{\sim}{B}}) - Y_k$. L'estimateur de $\underset{\sim}{V}[\hat{\underset{\sim}{B}}]$ est donné par:

$$\hat{\underset{\sim}{J}}^{-1}\, \hat{\underset{\sim}{V}}(\hat{\underset{\sim}{W}})\, \hat{\underset{\sim}{J}}^{-1}.$$

3.5 Modèles linéaires logarithmiques pour des données qualitatives

Supposons que chaque membre de la population appartienne à une seule catégorie d'un nombre q de catégories distinctes. Un vecteur a_i de rxl est associé à la catégorie i de sorte que la proportion d'individus faisant partie de la i-ième catégorie est approximativement:

$$p_i(\underset{\sim}{\beta}) = \frac{\exp(\underset{\sim}{a}_i^T\, \underset{\sim}{\beta})}{\sum_j \exp(\underset{\sim}{a}_j^T\, \underset{\sim}{\beta})} .$$

Nous pouvons également estimer la variance de $\hat{R}^2$. Si $\underset{\sim}{W}_k^T(B,R^2) = [Y_k, Z_{k1} e_k, \ldots, Z_{kp} e_k, Y_k (\sum_j Z_{kj} B_j - R^2 Y_k)]$ et $\underset{\sim}{c}^T = [-2\hat{Y}(1-\hat{R}^2)/N, \underset{\sim}{\hat{B}}^T, 1]/(S_{YY}-1^{-1}\hat{Y}^2)$, nous obtenons:

$$\hat{V}[\hat{R}^2] \doteq \underset{\sim}{c}^T \, \underset{\sim}{\hat{V}}[\underset{\sim}{\hat{W}}(\underset{\sim}{\hat{B}},\hat{R}^2)] \, \underset{\sim}{c}. \tag{3.5}$$

Dans le cas où N est inconnu, par exemple lorsque les unités primaires d'échantillonnage sont des régions géographiques, nous obtenons l'équation supplémentaire suivante:

$$U_4 = N - \Sigma 1, \tag{3.6}$$

Après avoir ajouté les unités de la ligne et de la colonne appropriées à $\underset{\sim}{J}$ et après inversion, nous pouvons estimer $V[\hat{R}^2]$ de la façon suivante

Nous posons: $\underset{\sim}{W}_k^T(\underset{\sim}{B},R^2) = [Y_k, Z_{k1} e_k, \ldots, Z_{kp} e_k, Y_k(\sum_j Z_{kj} B_j - R^2 Y_k), 1]$

et que $\underset{\sim}{c}^T = [-2\hat{Y}(1-\hat{R}^2)/\hat{N}, \hat{B}^T, 1, \hat{Y}^2(1-\hat{R}^2)/\hat{N}^2]/(S_{YY} - \hat{N}^{-1}\hat{Y}^2)$.

Nous pouvons alors déterminer $\hat{V}[\hat{R}^2]$ à l'aide de l'équation (3.5) pour ces nouvelles valeurs de $\underset{\sim}{W}_k(\underset{\sim}{B},R^2)$ et $\underset{\sim}{c}$.

3.4 Régression logistique

Comme dans la section précédente, nous posons comme hypothèse que la matrice de données $\underset{\sim}{X}$ peut être répartie en $[\underset{\sim}{Z}|\underset{\sim}{Y}]$, mais ici $\underset{\sim}{Y}$ est un vecteur de 0 et des 1. Dans le cadre de l'analyse statistique classique, le modèle de régression logistique de $\underset{\sim}{Y}$ conditionnel à $\underset{\sim}{Z}$ implique que les valeurs de $Y_1, \ldots, Y_N$ sont indépendantes et que $Pr(Y_k = 1) = p_k(\underset{\sim}{\beta})$, où:
où

$$p_k(\beta) = \frac{\exp(\underset{\sim}{\beta}^T \underset{\sim}{z}_k)}{1 + \exp(\underset{\sim}{\beta}^T \underset{\sim}{z}_k)} \quad . \tag{3.7}$$

Supposons que $\underset{\sim}{p}(\underset{\sim}{\beta})^T = [p_1(\underset{\sim}{\beta}), \ldots, p_q(\underset{\sim}{\beta})]$ et que $\underset{\sim}{N}^T = (N_1, \ldots, N_q)$, où N_i représente le nombre d'individus dans la i-ième catégorie. Si la population est générée à partir d'une distribution multinomiale de probabilité $\underset{\sim}{p}(\underset{\sim}{\beta})$, l'estimateur à maximum du vraisemblance de $\underset{\sim}{\beta}$, donné par $\underset{\sim}{B}$, satisfait

$$\underset{\sim}{U} = \underset{\sim}{A}^T \underset{\sim}{N} - [\underset{\sim}{A}^T \underset{\sim}{p}(\underset{\sim}{B})] \underset{\sim}{1}^T \underset{\sim}{N} = 0,$$

où $\underset{\sim}{A}$ est une matrice qxr dont la i-ième ligne correspond à $\underset{\sim}{a}_i^T$. Nous *considérons* $\underset{\sim}{B}$ *comme notre paramètre d'intérêt de toute population finie* donnée.

Soit $\hat{\underset{\sim}{N}}$ un estimateur convergent asymptotiquement normal de $\underset{\sim}{N}$, $\underset{\sim}{V}[\hat{\underset{\sim}{N}}]$ la matrice des variances-covariances et $\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{N}}]$ la matrice estimée. Notre estimateur, $\hat{\underset{\sim}{B}}$, satisfait:

$$\underset{\sim}{A}^T \hat{\underset{\sim}{N}} - [\underset{\sim}{A}^T \underset{\sim}{p}(\hat{B})] \underset{\sim}{1}^T \hat{\underset{\sim}{N}} = 0. \tag{3.9}$$

Cette méthode d'estimation a été proposée par Freeman et Koch (1976). Elle est peut-être moins efficace que la méthode de régression asymptotique fonctionnelle de Imrey, Koch et Stokes (1981, 1982), mais il n'est pas nécessaire de calculer toutes les composantes de $\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{N}}]$ pour appliquer la formule (3.9).

Nous supposons que $\underset{\sim}{D}(\underset{\sim}{B})$ correspond à $\text{diag}[\underset{\sim}{p}(\underset{\sim}{B})]$ et $\underset{\sim}{H}(\underset{\sim}{B}) = \underset{\sim}{D}(\underset{\sim}{B}) - \underset{\sim}{p}(\underset{\sim}{B}) \underset{\sim}{p}(\underset{\sim}{B})^T$. Alors

$$\underset{\sim}{J} = \frac{\partial \underset{\sim}{U}}{\partial \underset{\sim}{B}} = - (\underset{\sim}{1}^T \underset{\sim}{N}) \underset{\sim}{A}^T \underset{\sim}{H}(\underset{\sim}{B}) \underset{\sim}{A} .$$

Par conséquent, la matrice de variance asymptotique pour B est donnée par:

$$\underset{\sim}{V}[\hat{\underset{\sim}{B}}] = (\underset{\sim}{N}^T \underset{\sim}{1})^{-2} (\underset{\sim}{A}^T \underset{\sim}{H}(\underset{\sim}{B}) \underset{\sim}{A})^{-1}$$

$$\underset{\sim}{A}^T(\underset{\sim}{I}-\underset{\sim}{p}(\underset{\sim}{B})\underset{\sim}{1}^T) V[\hat{N}] (\underset{\sim}{I}-\underset{\sim}{1} \underset{\sim}{p}(\underset{\sim}{B})^T) A(\underset{\sim}{A}^T \underset{\sim}{H}(\underset{\sim}{B}) \underset{\sim}{A})^{-1} . \tag{3.10}$$

Cette expression peut quelquefois être simplifiée de la manière suivante. Si on peut supposer que $\underset{\sim}{N}/\underset{\sim}{N}^T\underset{\sim}{1} \doteq \underset{\sim}{p}(\underset{\sim}{B})$, alors pour $\hat{\underset{\sim}{\pi}} = \underset{\sim}{N}/n^T\underset{\sim}{1}$ nous avons:

$$\underset{\sim}{V}[\hat{\underset{\sim}{\pi}}] \doteq (\underset{\sim}{N}^T\underset{\sim}{1})^{-2}\ (\underset{\sim}{I}-\underset{\sim}{p}(\underset{\sim}{B})\underset{\sim}{1}^T)\ \underset{\sim}{V}[\hat{\underset{\sim}{N}}](\underset{\sim}{I}-\underset{\sim}{1}\ \underset{\sim}{p}(\underset{\sim}{B})^T),$$

de sorte que

$$\underset{\sim}{V}[\underset{\sim}{B}] \doteq (\underset{\sim}{A}^T\ \underset{\sim}{H}(\underset{\sim}{B})\ \underset{\sim}{A})^{-1}\ \underset{\sim}{A}^T\ \underset{\sim}{V}[\hat{\underset{\sim}{\pi}}]\ \underset{\sim}{A}(\underset{\sim}{A}^T\ \underset{\sim}{H}\ (\underset{\sim}{B})\ \underset{\sim}{A})^{-1}. \qquad (3.11)$$

En outre, la matrice des covariances de $\underset{\sim}{p}(\hat{\underset{\sim}{B}})$, soient les probabilités estimées par case, nous est donnée par:

$$\underset{\sim}{V}[\underset{\sim}{p}(\hat{\underset{\sim}{B}})] = \underset{\sim}{H}(\underset{\sim}{B})\ \underset{\sim}{A}\ \underset{\sim}{V}[\hat{\underset{\sim}{B}}]\ \underset{\sim}{A}^T\ \underset{\sim}{H}(\underset{\sim}{B}).$$

Les estimateurs de $\underset{\sim}{V}[\hat{\underset{\sim}{B}}]$ et de $\underset{\sim}{V}[\underset{\sim}{p}(\underset{\sim}{B})]$ sont des expressions semblables dans lesquelles $\underset{\sim}{N}$ et $\underset{\sim}{B}$ sont remplacés respectivement par $\hat{\underset{\sim}{N}}$ et $\hat{\underset{\sim}{B}}$. On pourrait alors déterminer directement la valeur de $\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{N}}]$. Dans le cas où q **est** relativement plus quand que r, il serait plus efficace de procéder de la façon suivante. Supposons que

Y_{ki} = 1 si la k-ième unité appartient à la i-ième catégorie
= 0 dans tout autre cas,

par k=1, ..., N; i-1, ..., q. Soient $\underset{\sim}{Y}^T = (Y_{k1}, \ldots, Y_{kq})$, et
$\underset{\sim}{W}_k = \underset{\sim}{A}^T\ [\underset{\sim}{I}-\underset{\sim}{p}(\hat{\underset{\sim}{B}})\ \underset{\sim}{1}^T]\ \underset{\sim}{Y}_k$.

Alors
$$\hat{\underset{\sim}{V}}[\hat{\underset{\sim}{B}}] \doteq (\hat{\underset{\sim}{N}}^T\underset{\sim}{1})^2\ (\underset{\sim}{A}^T\ \underset{\sim}{H}(\hat{\underset{\sim}{B}})\ \underset{\sim}{A})^{-1}\ \hat{\underset{\sim}{V}}(\hat{\underset{\sim}{W}})\ (\underset{\sim}{A}^T\ \underset{\sim}{H}(\underset{\sim}{B})\ \underset{\sim}{A})^{-1}.$$

Soulignons que les méthodes décrites dans la présente section peuvent facilement être appliquées aux modèles de distribution multinomiale du produit, dans lesquels il existe un modèle linéaire logarithmique pour $\{N_{ij}\}$, mais où les marges $\{\sum_j N_{ij}\}$ sont connues.

4. DISCUSSION

Les méthodes décrites dans le présent article s'appliquent à certains modèles précis, à titre d'exemple, voir Fuller (1975) et Freeman et Koch (1976). Les résultats généraux n'ont toutefois pas été présentés de façon explicite. Il est possible d'utiliser de nombreux progammes statistiques types pour estimer les paramètres des modèles décrits, mais les variances et les tests d'hypothèses compris dans ces programmes ne seront pas valides.

Les résultats présentés dans cet article sont fondés sur l'hypothèse que les estimateurs sont asymptotiquement normaux. Il importe d'effectuer des études empiriques sur la validité de ces approximations.

D'autres méthodes d'estimation de bon nombre de paramètres décrits dans cet article sont présentées par Imrey, Koch et Stokes (1981, 1982). Leur méthode de régression asymptotique fonctionnelle s'inscrit également dans le cadre général de la présente analyse, quant à l'estimation et la dérivation des variances.

BIBLIOGRAPHIE

[1] Frankel, M.R. (1971), Inference from Survey Samples , University of Michigan, Ann Arbor.

[2] Freeman, D.H. Jr., and Koch, G.G. (1976), "An Asymptotic Covariance Structure for Testing Hypotheses on Raked Contingency Tables from Complex Sample Surveys", Proc. Amer. Statist. Ass. (Social Statistics Section), Part 1, 330-335.

[3] Fuller, W.A. (1975), "Regression Analysis for Sample Survey", Sankya, Series C, 37, 117-132.

[4] Hajek, J. (1960), "Limiting Distributions in Simple Random Sampling from a Finite Population", Publ. Math. Inst. Hung. Acad. Sci., 5, 361-374.

[5] Imrey, P.B., Koch, G.G., Stokes, M.E. (1981-1982), "Categorical Data Analysis: Some Reflections on the Log Linear Model and Logistic Regression; Part I: Historical and Methodological Overview. Part II: Data Analysis", International Statistical Review. À paraître.

[6] Kish, L., and Frankel, M.R. (1974), "Inference from Complex Samples", J. Roy. Statistic. Soc. B, 36, 1-22.

[7] Madow, W.G. (1948), "On the Limiting Distribution of Estimates Based on Samples from Finite Universes", Ann. Math. Statist., 19, 535-545.

[8] Särndal, C.E. (1978), "Design-based and Model-based Inference in Survey Sampling", Scand. J. Statist. 5, 27-52.

[9] Tepping, B.J. (1968), "The Estimation of Variance in Complex Surveys", Proc. Amer. Statist. Assoc. (Social Statistics Section), 11-18.

[10] Woodruff, R.S. (1977), "A Simple Method for Approximating the Variance of a Complicated Estimate", J. Amer. Statist. Assoc. 66, 411-414.