CAHIER DE RECHERCHE

JUIN 1995 ■ N° 73

ALLOCATION INTRA-MÉNAGE DE LA CONSOMMATION ALIMENTAIRE

Patrick BABAYOU

Direction scientifique : Jean-Luc VOLATIER

Département "Prospective de la Consommation"

CRÉDOC

L'ENTREPRISE DE RECHERCHE

Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie

CRÉDOC

Allocation *intra*-ménage de la consommation alimentaire

APPROCHES EMPIRIQUES ÉCONOMÉTRIQUE ET PROBABILISTE D'UN PROBLÈME DE THÉORIE DU CONSOMMATEUR

Patrick BABAYOU

Direction scientifique : Jean-Luc VOLATIER

Département Prospective de la Consommation

JUIN 1995

142, rue du Chevaleret
75013 - PARIS

Sommaire

AVANT PROPOS

Le but de cette recherche est de présenter deux méthodes d'allocation *intra*-ménage de la consommation alimentaire.

La première se situe classiquement dans la lignée de travaux économétriques, en particulier menés au Royaume-Uni. Nous cherchons essentiellement à mettre en évidence les possibilités offertes par de telles méthodes, et surtout à les appliquer à des données d'enquête françaises. A partir des données des enquêtes alimentaires de l'INSEE, nous parvenons donc à explorer différentes composantes des comportements alimentaires des ménages français en matière d'allocation de la consommation. L'angle d'approche réside essentiellement dans l'inférence d'une consommation moyenne des individus qui composent les ménages.

Ces travaux économétriques comportent un certain nombre de limites, surtout du point de vue de leur pouvoir de description des comportements au sein des ménages. Si l'on parvient à établir des échelles de consommation cohérentes au regard des connaissances que l'on peut avoir de la consommation alimentaire par le biais d'autres sources, le résultat n'est valable que dans le cadre d'une estimation des apports de nutriments en moyenne. C'est toutefois une information intéressante en soi et qui peut être directement opérationnelle lors d'études sur les consommations individuelles.

La deuxième approche que nous proposons est nouvelle à notre connaissance. Elle a pour but de se détacher du cadre peut-être trop normatif de l'économétrie. Il s'agit de proposer une formalisation probabiliste du problème de la répartition de la nourriture au sein du ménage. Ainsi, nous cherchons à établir une règle d'allocation par l'inférence d'une loi du poids que prend chaque individu dans les décisions de répartition à l'intérieur du ménage. Ce point de vue mène à la mise en évidence de comportements spécifiques selon les types d'individu qui composent le ménage, illustrables par des courbes de densités de probabilités empiriques.

L'approche consiste donc ici à inférer une information sur les apports nutritionnels <u>en loi</u> et non plus en moyenne.

Parallèlement à une description des comportements, notre approche probabiliste permet aussi d'établir une règle d'allocation. Nous appliquerons cette méthodologie à des données d'enquête de consommation individuelle de l'Institut Scientifique et Technique de la Nutrition et de l'Alimentation (ISTNA) du CNAM.

Nous n'évoquons pas dans ce cahier la question des échelles d'équivalence car l'objectif principal n'est pas de comparer des niveaux de vie entre ménages de compositions différentes mais d'estimer les variations de consommations alimentaires entre membres d'un même ménage.

* * *

Ce cahier de recherche s'articule logiquement selon les orientations qui viennent d'être présentées. Après un exposé des problèmes à la fois micro-économiques, sociologiques et mathématiques soulevés par la question de l'allocation *intra*-ménage de la consommation, et en particulier de la consommation alimentaire, accompagné d'une courte revue bibliographique sur le sujet, nous présentons successivement les formalisations et les résultats préliminaires des approches économétrique et probabiliste.

Ces deux manières d'envisager le problème de l'allocation ne sont pas contradictoires, et il appartiendra à de futures recherches d'en délimiter à la fois les intérêts spécifiques et d'éventuels terrains de convergence qui permettront peut-être d'enrichir une approche par l'autre.

Enfin, l'application à l'alimentation est naturellement à prendre en tant qu'exemple, et les méthodes explorées ici ne semblent en rien liées intrinsèquement à ce type de consommation.

INTRODUCTION

La modélisation d'une allocation *intra*-ménage de la consommation relève à la fois d'un problème général de formalisation du fonctionnement des ménages et d'un problème essentiellement pratique de méthodologie statistique d'enquêtes.

Ce dernier point n'est pas le moins important, et il conditionne même pour une bonne part les méthodes d'évaluations que l'on pourra utiliser dans le cadre d'une telle modélisation. Ainsi, dans le cas des enquêtes de consommation alimentaire, les données de quelque type que ce soit —à visée marketing, ou bien les enquêtes générales de comportements— ne recueillent d'informations qu'au niveau des ménages, par exemple en demandant à la ménagère de recenser précisément la nature des achats qu'elle réalise. Les données disponibles ne donnent donc aucune information sur les consommations individuelles des membres des ménages, ce qui peut être un handicap pour certaines approches. En effet, toujours dans le cas alimentaire, la demande d'information émanant des nutritionnistes ou encore des toxicologues concerne principalement les individus. Il s'agit de déterminer des populations à risques —par exemple des risques liés à l'ingestion d'additifs, de pesticides..., ou encore liés à des déséquilibres nutritionnels—, et s'il peut être intéressant de rechercher des populations de ménages, cela reste d'un intérêt limité.

Cependant, la méthode de recueil au niveau du ménage, sans préoccupation individuelle, reste encore la plus efficace et la plus usitée par les instituts, du fait d'un moindre coût et surtout d'une moindre difficulté de questionnement. Ce type d'enquête de consommation, souvent par remplissage d'un carnet, est déjà fastidieux pour les personnes enquêtées lorsqu'il s'agit de relever les achats du ménage, mais devient franchement inacceptable dès lors que l'on cherche à obtenir une information individuelle précise.

Cette dernière remarque aurait pu être contournée, par le biais du développement de méthodologies d'enquête adéquates, améliorant le confort de saisie d'un carnet de consommation. Toutefois, il n'était pas si crucial d'en venir à de telles recherches dans la mesure où les utilisateurs finaux des données considéraient peu ou prou le ménage comme une boîte noire, dans laquelle tous les individus sont équivalents. Pour passer de la consommation du ménage dans lequel vivent n personnes à la consommation individuelle de chacun des n, l'option par défaut fut de considérer qu'il suffisait de diviser par n la consommation totale du ménage.

Cette méthode abrupte n'est pas absurde, s'il s'agit de calculer une moyenne globale de la consommation individuelle de biens sur l'ensemble de la population. Introduire des disparités individuelles au sein des ménages revient en effet à attribuer un système de pondérations sous contrainte, ce qui ne changera pas la valeur des indicateurs de tendance centrale. Les calculs d'indices de dispersion sont toutefois affectés par un tel système de poids.

Cependant, les moyennes ne sont pas calculables par regroupements d'individus si l'on reste dans le contexte d'une répartition uniforme de la consommation entre les membres des ménages, le postulat étant justement qu'il n'y a pas de différence entre les groupes.

La question que pose la mise en oeuvre d'une démarche d'allocation *intra*-ménage de la consommation, en particulier alimentaire, est donc de déterminer en premier lieu jusqu'à quel point il est tolérable de négliger les disparités de consommations entre les individus.

A la suite de nombreux travaux micro-économiques d'origines très diverses, nous posons que ces disparités existent et ne doivent pas être négligées pour comprendre la consommation des ménages.

TRAVAUX EMPIRIQUES AUTOUR D'UNE MODÉLISATION LINÉAIRE

Travaux de Chesher sur le N.F.S.

Un modèle linéaire d'allocation *intra*-ménage de la consommation alimentaire a été proposé par Andrew Chesher (1991) dans un cadre temporel à partir des données du National Food Survey (N.F.S.) britannique de 1979 à 1989.

Après un modèle AIDS[1] "classique", Chesher propose un modèle linéaire pour l'estimation des prises alimentaires individuelles à partir de la connaissance de la structure démographique des ménages et de leur consommation totale. Le N.F.S. ne permet en effet de connaître que des consommations globales, et non des consommations individuelles. Ce modèle prend en compte 9 types d'individus distincts qui permettent de décrire la structure d'un ménage.

Ces types d'individus sont les suivants :

- enfant de 0 à 4 ans
- enfant de 5 à 11 ans
- adolescent homme de 12 à 17 ans
- adolescente femme de 12 à 17 ans
- adulte homme de 18 à 34 ans
- adulte homme de 35 à 64 ans
- adulte homme de 65 ans et plus
- adulte femme de 18 à 59 ans
- adulte femme de 60 ans et plus

Ces catégories d'individus correspondent à un regroupement des catégories d'âge et de sexe pour lesquelles le ministère de la santé britannique (DHSS) définit des "Apports quotidiens recommandés" d'éléments nutritionnels, et c'est à ce titre que Chesher les utilise.

[1] "Almost Ideal Demand System" (*cf.* Deaton et Muellbauer, 1980). Ce modèle permet d'expliquer les différences d'allocations *intra*-ménage sur des données en prix, mais il n'est pas adapté tel quel à l'estimation de quantités ingérées.

L'hypothèse principale que formule Chesher pour son modèle est que la consommation moyenne d'un ménage est égale à la somme des consommations moyennes de ses membres, ce qui l'amène à écrire le modèle suivant :

$$A_h = \delta_0 + \sum_{i=1}^{9} \delta_i n_{ih} + \varepsilon_h$$

où A_h est le total d'aliments "entrant" dans le ménage h, les δ_i la consommation moyenne d'aliments d'un individu de type i, et n_{ih} le nombre d'individus de type i dans le ménage h. Le terme δ_0 peut s'interpréter comme une mesure de la consommation par des personnes extérieures au ménage (invités...), et aussi comme une mesure des pertes d'aliments (déchets, gâchis...). Il est évidemment difficile de faire la part des choses entre ces deux phénomènes, mais on peut toutefois formuler des hypothèses crédibles à ce sujet à partir de la connaissance acquise par le biais d'enquêtes sur les comportements alimentaires en général, celles-ci permettant de mesurer en particulier les habitudes et pratiques en matière de convivialité et de gâchis.

Il y a bien sûr abus de langage de notre part en qualifiant directement les δ_i de "consommation moyenne". En effet, cette assertion est vraie dans le cadre de l'écriture du modèle —c'est bien la manière dont on souhaite modéliser la consommation globale du ménage— mais encore faut-il montrer que leurs estimateurs par les MCO sont égaux, au moins en espérance, aux consommations individuelles moyennes de chacune des neuf catégories retenues.

Convergence des coefficients de la régression vers les consommations moyennes individuelles

Pour montrer l'égalité en espérance des δ_i avec les consommations moyennes par types d'individus correspondants on pose dans un premier temps un modèle d'analyse de la variance à 1 facteur pour décrire la consommation d'un individu donné :

$$A_{ijh} = d_i + u_{ijh}$$

où A_{ijh} est la consommation de l'individu j de type i dans le ménage h, cette consommation s'écrivant comme la somme d'un effet fixe catégoriel et d'une erreur individuelle liée aussi au ménage.

Ce modèle permet d'estimer l'effet moyen de la catégorie :

$$\hat{d}_i = A_{i..}$$

On peut par conséquent écrire la consommation du ménage comme la somme des consommations individuelles. Soit :

$$A_h = \delta_0 + \sum_j A_{ijh} = \delta_0 + \sum_j \left(d_i + u_{ijh}\right) = \delta_0 + \sum_i \hat{d}_i n_{ih} + \sum_j u_{ijh}$$

La dernière égalité se justifie par le fait que la somme des consommations individuelles égale la somme des consommations moyennes par type d'individu pondérées par l'effectif de chaque type d'individu dans le ménage. En identifiant cette dernière expression avec le modèle d'origine, on obtient pour tout h :

$$\sum_i \left(\delta_i - \hat{d}_i\right) n_{ih} = \varepsilon_h - \sum_j u_{ijh}$$

D'où, en passant à l'espérance, pour tout h :

$$\sum_i \left[E(\delta_i) - E\left(\hat{d}_i\right)\right] n_{ih} = 0$$

Le problème posé revient donc à montrer qu'il existe une solution unique d'un système à H équations et $2I$ inconnues, H étant très supérieur à $2I$.

On remarquera que les n_{ih} sont positifs ou nuls, mais, pour chaque valeur possible de i, qu'il existe au moins une équation pour laquelle n_{ih} est non nul, vu la structuration des ménages (il serait absurde de modéliser la consommation selon l'effectif d'individus qui ne se trouvent jamais dans les ménages étudiés).

La présence de ménages composés d'une seule personne impose déjà un grand nombre d'équations h qui se résolvent, pour des valeurs de i précises, à :

$$E(\delta_i) - E\left(\hat{d}_i\right) = 0$$

et ces contraintes s'imposent aux autres équations, d'où la conclusion suivante, pour tout i :

$$E(\delta_i) = E\left(\hat{d}_i\right) = A_{i..}$$

Ce qui montre l'égalité, en espérance, des δ_i avec les consommations moyennes par catégories[2].

Il est dès lors possible d'inférer des consommations individuelles moyennes, ce que fait sans soucis Chesher sur les données du N.F.S. Il met ainsi en évidence des différences sensibles de niveau de consommation entre les catégories d'individus. Son approche étant essentiellement nutritionnelle, il montre en particulier que les estimations d'apports en nutriments obtenues varient de la même manière que les recommandations du DHSS —ce qui est à la fois rassurant quant à la méthode d'estimation et quant aux habitudes de consommation alimentaire des britanniques.

Application du modèle de Chesher à des données françaises

Nous nous proposons dans un premier temps de reprendre la méthode d'estimation de Chesher pour l'appliquer à des données françaises.

Les données des enquêtes alimentaires de l'INSEE sont à ce titre adaptées à cette application, leur structure étant similaire à celle du N.F.S. Nous appliquons par conséquent le modèle linéaire à ces données, pour la dernière année disponible (1991).

[2] La vitesse de convergence des coefficients de la régression vers les moyennes catégorielles en fonction des effectifs considérés est assez rapide, comme nous le montrons rapidement en annexe, en reprenant les résultats d'une note réalisée par le CRÉDOC en 1993 dans le cadre du programme de l'Observatoire des Consommations Alimentaires.

Tableau 1 : Estimation du modèle linéaire sur l'ensemble de la consommation alimentaire (données INSEE 1991)

Catégorie d'individu	Estimateur
Constante	256,6 49,6
Enfant de 0 à 4 ans	503,3 42,9
Enfant de 5 à 11 ans	458,8 33,0
Adolescent homme de 12 à 17 ans	633,8 51,5
Adolescente femme de 12 à 17 ans	628,7 54,6
Adulte homme de 18 à 34 ans	660,2 36,0
Adulte homme de 35 à 64 ans	945,5 40,8
Adulte homme de 65 ans et plus	907,4 56,7
Adulte femme de 18 à 59 ans	834,4 37,3
Adulte femme de 60 ans et plus	842,6 52,2

Note de lecture :

données en kilogramme annuel d'aliments consommés
les nombres indiqués sous les estimations correspondent à leur erreur standard

Les résultats de cette première régression montrent, comme sur les données britanniques du N.F.S., des quantités estimées différentes selon les catégories d'individus dans un ménage. En particulier, les disparités observées évoluent d'une manière cohérente à la prénotion que l'on peut en avoir : consommation supérieure chez les hommes, inférieure chez les enfants, augmentation de la consommation avec l'âge jusqu'à 60 ans, puis baisse de cette consommation.

Ces résultats sont pourtant un peu déroutants dans le sens où ils montrent la part plutôt importante dans la consommation totale du ménage que prennent les individus extérieurs et/ou le gâchis (*cf.* la valeur de la constante du modèle).

Le défaut global d'un tel modèle est aussi de traiter l'ensemble de la consommation alimentaire d'un ménage, sans prendre en compte les différences qui peuvent être induites par la nature des différents aliments. Ainsi, par exemple, il n'est pas irrationnel de supposer que la répartition *intra*-ménage des boissons alcoolisées est différente de celle des légumes. Pour remédier à cette limite du modèle précédent, nous pouvons établir un modèle spécifique pour chaque type d'aliments dont les consommations sont étudiées dans l'enquête de l'INSEE.

Nous cherchons donc à estimer un vecteur de la consommation des différents types d'aliments par les membres des ménages, en empilant autant de régressions qu'il existe de types d'aliments retenus pour décrire la structure de la consommation alimentaire du ménage. D'où la spécification du modèle empilé suivant :

$$A_h^j = \delta_0^j + \sum_{i=1}^{9} \delta_i^j n_{ih} + \varepsilon_h^j$$

où j est l'indice des groupes d'aliments, ayant J modalités.

Le tableau suivant donne les estimations des consommations individuelles pour les aliments agrégés selon la nomenclature de l'INSEE prise à son premier niveau. Dans ce cas, on a J=9 groupes constitués.

Tableau 2 : Estimation du modèle linéaire par groupe d'aliments (données INSEE 1991)

	Produits à base de céréales	Légumes	Fruits	Viande	Volaille	Poisson	Produits laitiers	Produits divers d'épicerie	Boissons
Constante	-2,1	-5,8	26,8	7,2	31,6	1,4	98,8	2,4	96,2
	4,0	10,2	6,5	5,3	16,7	1,1	22,6	4,4	21,0
Enfant de 0 à 4 ans	23,7	34,3	7,1	14,0	61,8	2,5	237,4	81,2	41,3
	3,5	8,8	5,6	4,6	14,4	0,9	19,6	3,8	18,1
Enfant de 5 à 11 ans	38,7	47,2	26,4	27,4	79,6	0,2	200,2	8,7	30,4
	2,7	6,8	4,3	3,5	11,1	0,7	15,0	3,0	13,9
Adolescent homme de 12 à 17 ans	60,7	68,5	36,6	53,1	124,6	3,8	226,8	19,0	40,6
	4,2	10,5	6,7	5,5	17,2	1,1	23,4	4,6	21,7
Adolescente femme de 12 à 17 ans	58,9	77,6	52,0	33,5	152,9	2,8	190,5	22,2	38,3
	4,4	11,2	7,1	5,8	18,3	1,2	24,9	4,9	23,1
Adulte homme de 18 à 34 ans	66,9	101,7	18,0	53,8	96,4	2,8	210,2	25,4	85,1
	2,9	7,4	4,7	3,8	12,1	0,8	16,4	3,2	15,2
Adulte homme de 35 à 64 ans	79,4	152,7	67,7	58,5	143,9	12,1	188,6	27,6	215,1
	3,3	8,4	5,3	4,3	13,7	0,9	18,6	3,7	17,2
Adulte homme de 65 ans et plus	76,0	165,8	85,9	48,5	143,3	8,7	158,8	29,7	190,6
	4,6	11,7	7,4	6,0	19,0	1,2	25,9	5,1	24,0
Adulte femme de 18 à 59 ans	49,6	98,8	61,1	46,7	183,3	5,3	232,8	21,9	135,0
	3,0	7,7	4,9	4,0	12,5	0,8	17,0	3,3	15,8
Adulte femme de 60 ans et plus	71,7	147,8	71,7	41,1	176,5	9,0	177,7	32,1	115,1
	4,2	10,7	6,8	5,6	17,5	1,1	23,9	4,7	22,1

N.B. : en grisé sont indiquées les consommations pour lesquelles un test de significativité du coefficient est rejeté (statistique de Student inférieure à 2)

Les consommations moyennes par types d'individus sont effectivement sensiblement différentes selon les aliments consommés, d'après l'estimation que l'on met en évidence par l'application du modèle de Chesher empilé. Pour la plupart des produits, les hiérarchies observées sont similaires à celle observée pour le modèle global, dans des proportions toutefois variables. Ainsi, les écarts sont plus marqués pour les légumes qu'ils ne le sont pour les produits à base de céréales ou les fruits.

Des produits induisant un comportement spécifique à l'intérieur du ménage apparaissent clairement : ce sont les boissons, pour lesquelles la consommation des adultes est très significativement supérieure à celle des enfants, les produits laitiers, que les enfants consomment au contraire en plus grande quantité que les adultes, ou encore les produits divers d'épicerie, pour lesquels le même phénomène est observé —probablement du fait de la présence de catégories de produits alimentaires spécialisés tels que les aliments pour bébés...

On notera enfin la forte variabilité de la valeur que prend la constante du modèle : pour trois groupes d'aliments, les fruits, les produits laitiers et les boissons, elle garde une valeur significativement non nulle. On peut supposer que ces produits sont particulièrement associés soit à un cadre festif —aliments surconsommés lorsqu'il y a des invités— ou qu'ils ont un taux de déchet supérieur à la normale.

Dans le cas des boissons, on peut penser que le poids de la convivialité est plus fort pour expliquer la valeur de la constante. Dans le cas des fruits et des produits laitiers, l'interprétation est plus délicate dans un premier temps, puisque ces produits sont certes plus souvent jetés par les ménages, mais certains peuvent aussi être associés à la convivialité — par exemple le fromage.

Pour tenter d'aller plus loin dans l'interprétation de la valeur de la constante du modèle, nous pouvons estimer des consommations individuelles à un niveau de détail plus fin dans la nomenclature de l'INSEE, par exemple en passant au deuxième niveau pour les produits qui semblent *a priori* intéressants.

Tableau 3 : Estimation du modèle pour des produits laitiers et pour les boissons

	Laits	Fromages	Boissons alcoolisées	Boissons non alcoolisées	Café, thé, infusions
Constante	-1,9	11,2	25,8	85,0	2,9
	8,7	1,7	8,4	18,7	0,6
Enfant de 0 à 4 ans	93,3	6,3	9,1	33,1	0,3
	6,9	1,3	6,6	14,7	0,5
Enfant de 5 à 11 ans	83,9	5,3	2,1	23,6	0,4
	5,2	1,0	5,0	11,1	0,4
Adolescent homme de 12 à 17 ans	84,9	8,7	-13,8	50,0	1,7
	8,4	1,6	8,1	17,9	0,6
Adolescente femme de 12 à 17 ans	72,0	7,0	3,7	38,2	0,3
	8,7	1,6	8,3	18,6	0,6
Adulte homme de 18 à 34 ans	69,0	10,1	28,3	55,9	1,7
	6,2	1,2	5,9	13,2	0,5
Adulte homme de 35 à 64 ans	63,7	16,9	90,4	120,4	4,2
	7,0	1,3	6,7	15,0	0,5
Adulte homme de 65 ans et plus	59,7	14,5	89,7	85,2	3,1
	9,9	1,9	9,5	21,2	0,7
Adulte femme de 18 à 59 ans	52,9	12,9	17,2	101,5	2,9
	6,4	1,2	6,1	13,7	0,5
Adulte femme de 60 ans et plus	76,0	11,1	12,1	83,7	3,4
	9,3	1,7	8,9	19,7	0,7

Le premier résultat de cette estimation à un niveau plus fin de la nomenclature réside dans la valeur toujours importante et significative de la constante du modèle, à la seule exception de celle du modèle sur le lait en général. On voit aussi que la non significativité des consommations globales de boissons estimées précédemment pour les individus de 17 ans et moins est due pour l'essentiel aux boissons alcoolisées et aux cafés, thés et infusions. Des comportements spécifiques apparaissent de plus : non significativité de la consommation de boissons alcoolisées chez les femmes de plus de 60 ans, consommation de cafés, thés et infusions chez les adolescents.

Ces estimations ne permettent pourtant pas de statuer définitivement sur l'interprétation à donner à la valeur de la constante du modèle. Toutefois, les aliments concernés par une valeur forte de la constante sont tous associés, à des degrés divers, à la préparation des "repas soignés" tels qu'ils sont décrits en particulier dans les enquêtes sur les comportements alimentaires du CRÉDOC. Leur consommation par des individus extérieurs au ménage est par conséquent à envisager.

Problèmes liés à l'inférence d'une règle d'allocation

La modélisation linéaire que nous venons de mettre en oeuvre permet donc d'estimer des consommations moyennes par types d'individus appartenant à des ménages dans l'ensemble de la population, en ne connaissant *a priori* que les consommations globales des ménages et leur structure démographique.

A l'inverse il est difficile d'estimer à partir de cette modélisation la répartition statistique des consommations individuelles au sein des ménages eux-mêmes.

Pour chercher à estimer non plus une quantité moyenne d'aliments ingérés mais une part de la consommation individuelle dans la consommation du ménage, il faudrait pouvoir spécifier un modèle estimant les parts individuelles correspondantes —c'est-à-dire normer le modèle de Chesher par la consommation totale du ménage—, à taille de ménage constante. Cette dernière contrainte est décisive, la notion de part n'ayant de sens que dans ce contexte. Toutefois, un tel modèle n'est pas facilement identifiable, puisque la variable expliquée est constante (et égale à 100%).

Une autre façon d'inférer une valeur moyenne du point moyen d'un individu de type donné dans un ménage consistera à examiner pour chaque taille de ménage possible, la valeur de la consommation totale prédite par le modèle, et de la diviser par les coefficients du modèle. On obtiendra alors une estimation des parts prises par les individus des ménages, mais la somme de ces parts n'égalera pas 100%, puisqu'il faudrait alors prendre en compte l'erreur du modèle.

Limites de la modélisation linéaire

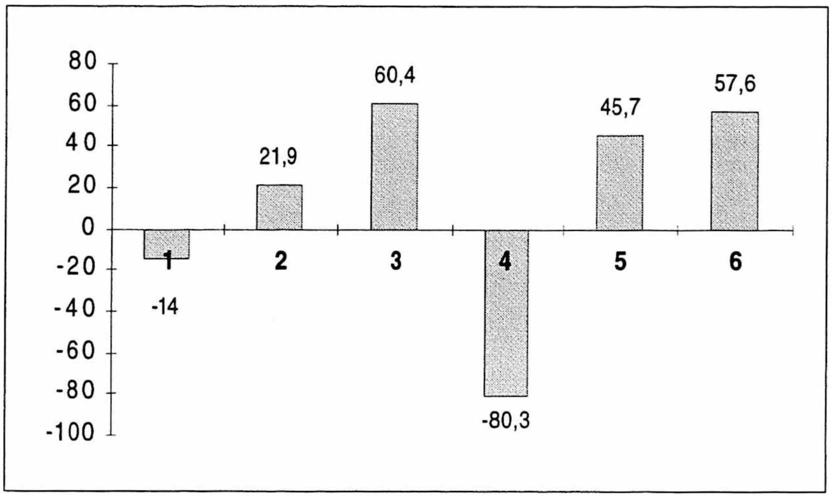
Le modèle que nous avons estimé par les MCO est de facture classique, et dépend donc des hypothèses habituelles faites dans la modélisation linéaire. Les deux principales sont d'une part la normalité des résidus, et d'autre part l'indépendance des variables explicatives, ou d'une combinaison linéaire des variables explicatives, avec le terme résiduel (homoscédasticité).

La mise en oeuvre d'un test de Shapiro-Wilk montre que l'hypothèse de normalité des résidus du modèle est tenable.

En revanche, l'hypothèse d'homoscédasticité pose problème, d'abord sur un plan intuitif. En effet, nous pouvons nous interroger au sujet des implications d'une telle hypothèse sur la modélisation. Celle-ci revenant à présupposer l'existence d'une indépendance relationnelle entre les variables explicatives elles-mêmes, cela signifierait par exemple qu'il n'y a pas de lien entre la présence des différents types d'individus.

En particulier, l'effectif du ménage a une influence notable sur la valeur résiduelle du modèle, comme le montre le graphique suivant :

Figure 1 : Valeurs moyennes du résidu du modèle global en fonction de l'effectif du ménage



Ces résultats tendent à prouver qu'il existe une corrélation des résidus avec des variables sous-jacentes au modèle. Ce fait rend très difficile une tentative de construction d'une règle d'allocation à partir des coefficients du modèle, l'hétéroscédasticité de celui-ci demandant encore à être maîtrisée.

Recherche des variables corrélées avec le terme résiduel d'un modèle estimé par les MCO

L'estimation par les MCO de notre modèle a été faite en formulant l'hypothèse d'homoscédasticité, c'est-à-dire :

$$A_h = \delta' n_h + \varepsilon_h$$
$$\text{s.h. } E(\varepsilon_h / n_h) = 0$$

où l'on désigne par n_h le vecteur des n_{ih}.

Pour mettre en évidence l'existence d'une corrélation d'une variable avec le terme résiduel, on peut postuler que le modèle s'écrit en fait de la manière suivante :

$$A_h = dn_h + bx_h + u_h$$
$$\begin{cases} E(u_h / n_h, x_h) = 0 \\ E(x_h / n_h) = \delta' n_h \neq 0 \end{cases}$$

c'est-à-dire que l'on suppose l'existence de l'effet d'une variable x.

De ce qui précède, on déduit :

$$E(A_h / n_h) = dn_h + bE(x_h / n_h) = n_h(d + b\delta')$$

Ceci signifie que les MCO estiment dans notre modèle de départ la quantité :

$$\eta = d + b\delta'$$

On en déduit alors la forme des résidus estimés par le modèle :

$$\varepsilon_h = A_h - n_h\eta = A_h - n_h(d + b\delta') \Leftrightarrow \varepsilon_h = bx_h + u_h - n_h b\delta'$$

Cette dernière relation implique que :

$$E(\varepsilon_h / x_h) = f(x_h) \neq 0$$

C'est cette fonction *f* que nous avons représentée sur le graphique qui précède, en prenant pour variable x la taille du ménage, montrant ainsi qu'il existe bien un effet de cette variable sur la valeur du terme résiduel.

APPROCHE PROBABILISTE

Introduction à la méthode

Le contexte d'étude de cette partie est radicalement différent de celui de la précédente. En changeant de modélisation, nous changeons de point de vue sur le problème posé par la désagrégation de la consommation des ménages. En effet, le but n'est plus ici de donner une estimation de la quantité moyenne d'aliments ingérés par les individus d'une catégorie donnée, mais d'estimer la loi de la part que prend chaque membre du ménage dans sa consommation alimentaire totale.

En faisant un minimum d'hypothèses micro-économiques, nous appliquons une méthode non paramétrique —estimation de la densité par un noyau— à des données d'enquête individuelles pour étudier la distribution de la part que prennent les individus de chaque type dans la consommation totale des ménages.

Nous tentons par conséquent d'estimer des *coefficients individuels* associés à chaque type d'individu composant un ménage et, plus précisément, la nature de la *loi de probabilité* de ces coefficients. L'allocation *intra*-ménage de la consommation alimentaire revient donc, au moins sur un plan théorique, à l'attribution d'un poids aléatoire aux membres des ménages étudiés.

Cette démarche globale s'avérant complexe, nous présentons ci-après les résultats d'une recherche préliminaire sur cette question. Dans un premier temps, le cadre d'analyse étant précisé, une première application de la méthode non paramétrique conduira à la mise en évidence des disparités de comportements par grands regroupements d'aliments, d'où un regard neuf sur l'influence de la composition du ménage sur les consommations individuelles. Après cette partie descriptive, on proposera une règle générale d'allocation de la consommation, basée sur l'édification d'une échelle d'équivalence.

Ces deux approches, si elles sont très loin de couvrir l'ensemble des possibilités offertes par la méthode non paramétrique proposée ici, permettent d'en présenter deux expressions décisives, descriptive et inférentielle.

Cadre d'analyse

La méthode d'estimation que nous nous proposons de formaliser consiste à estimer, pour un nombre donné de types d'individus dans un ménage, les lois de probabilité de la part que prend chacun dans la consommation alimentaire totale du ménage. Les données utilisées pour estimer un tel modèle d'allocation doivent permettre de connaître les consommations de chacun des membres des ménages.

Cette estimation se fait ici sans contrainte, dans un cadre statique. On n'impose pas en particulier une contrainte visant à répartir toute la consommation du ménage entre ses membres, et on pourra même obtenir des coefficients dont la somme sera supérieure à 1. Toutefois, on ne cherchera pas à estimer une part de "perte" (déchets, consommation de personnes étrangères au ménage,...). Ce choix impose en particulier que l'on n'étudiera pas les ménages constitués d'une seule personne, celle-ci étant alors supposée consommer l'ensemble des aliments qui entrent dans ce type de ménage.

Il y a autant de lois à estimer —par regroupement d'aliments— qu'il y a de types d'individus à envisager, pris dans des ménages de taille fixée.

Les visées d'une telle démarche sont multiples : on peut espérer dans un premier temps parvenir à décrire le comportement des individus par une loi de probabilité. Il sera possible d'examiner la forme de la loi estimée et d'en interpréter les conséquences sur le plan des comportements individuels à l'intérieur des ménages. On peut même chercher à contrôler l'appartenance des lois estimées à des familles de lois de probabilités connues par des tests d'adéquation. Deux cas sont toutefois à distinguer, selon que la loi estimée semble "pure" ou non. Là réside toute la difficulté de l'exercice. En effet, dans le cas d'une loi non "pure" — ou encore de lois "contaminées"—, c'est-à-dire d'une combinaison linéaire de plusieurs lois, il serait intéressant d'en arriver à une typologie des ménages décrits par ces différents comportements.

Dans un second temps, la méthode doit avoir une portée inférentielle pratique, et doit donc rendre possible, dans le cas de la connaissance de la consommation alimentaire totale d'un ménage et de sa constitution, l'estimation des consommations individuelles.

Nous retenons seulement trois types d'individus dans les ménages :

- adulte homme
- adulte femme
- enfant

Ainsi, les estimations de coefficients sont valables dans le cadre d'une formalisation minimale des positions individuelles dans le ménage. Travaillant sur une part, et non une grandeur absolue, nous formulons l'hypothèse que les répartitions *intra*-ménage sont indépendantes de l'âge et des quantités d'aliments ingérées.

Ce sont à la fois les caractéristiques des individus et celles de leur environnement qui sont étudiées pour l'influence qu'elles peuvent avoir sur les comportements d'allocation de la nourriture au sein du ménage.

Cette hypothèse est peut-être trop forte et il appartient pour une part à cette recherche d'en délimiter les conséquences.

Formalisation de l'estimation

Le problème posé par notre méthode consiste à inférer la loi de la part que prend dans l'alimentation totale du ménage un individu de type donné, dans un ménage de type donné, à partir de l'observation d'un nombre même minime de réalisations de cette part dans une population d'individus regroupés en ménages.

Soit la variable aléatoire continue X^{jh} décrivant la part de consommation de l'individu de type j (j=1 à 3 dans notre cas) dans le ménage de type h (h correspond à l'effectif du ménage). Nous observons les réalisations x_i^{jh} de cette variable aléatoire. C'est à partir de ces observations que nous cherchons à donner une fonction de densité approchant la densité $f(x)$ de X^{jh}. Nous nous basons par conséquent sur une étape d'étude de la fonction de répartition empirique de X^{jh}.

Cette densité est estimée par un noyau de convolution. Il s'agit d'approximer $f(x)$ en écrivant, pour tout x, l'expression suivante :

$$\hat{f}(x)=\frac{1}{n.h_n}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h_n}\right)$$

où :

les x_i sont les réalisations de X^{jh}

n est le nombre de réalisations de X^{jh} qui sont observées

h_n est un paramètre de lissage, dépendant de n

$K(u)$ est une fonction continue, vérifiant certaines propriétés (*cf.* ci-dessous), appelée noyau de convolution.

Les noyaux que nous avons envisagés appartiennent à la famille des noyaux de Parzen-Rosenblatt, qui permettent de se situer dans de bonnes conditions de convergence vers la vraie densité. Ces noyaux sont parmi les plus utilisés en statistique non paramétrique :

- noyau gaussien : $K(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$
- noyau d'Epanechnikov : $K(u)=\frac{3}{4\sqrt{5}}\left(1-\frac{u^2}{5}\right)$, pour $-\sqrt{5}\leq u\leq\sqrt{5}$
- noyau polynômial[3] : $K(u)=\frac{105}{64}\left(1-u^2\right)^2\left(1-3u^2\right)$, pour $-1\leq u\leq 1$

On peut ajouter à cette liste de noyaux celui dit "de la fenêtre mobile", consistant en une fonction indicatrice, ce qui permet de constituer un histogramme à pas fixe.

Les estimations de densités qui sont présentées par la suite ont été déterminées grâce à l'utilisation d'un noyau gaussien, qui a l'avantage d'être moins sensible aux effets de bords,

[3] Les noyaux polynômiaux sont d'usage peut-être moins fréquents, mais permettent d'obtenir un bon lissage de la courbe de densité estimée. Celui que nous avons utilisé a été proposé par Lejeune (1985).

puisqu'il ne fait pas intervenir une indicatrice dans son expression et qui est, de plus, toujours positif. Il est toutefois plus gourmand en calculs, mais ce n'est pas une limitation réelle dans notre contexte.

Définitions et propriétés du noyau de convolution
(d'après Bosq et Lecoutre, 1987)

On appelle "noyau" une fonction continue K définie sur les réels, intégrable par rapport à la mesure de Lebesgue et d'intégrale 1.

Un noyau réel est de Parzen-Rosenblatt si :

$$\lim_{\|x\| \to \infty} \|x\| K(x) = 0$$

L'estimateur $f_n(x)$ associé au noyau K —la densité de probabilité empirique— est défini par :

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right) = \left(K_{h_n} * \mu_n\right)(x)$$

Les estimateurs associés aux noyaux de Parzen-Rosenblatt convergent (en moyenne quadratique) vers la vraie densité de probabilité.

Pour obtenir un estimateur de f qui soit une densité de probabilité, il est de plus conseillé de prendre un noyau positif.

Résultats descriptifs sur des données d'enquête

L'application de la méthode non paramétrique d'estimation de la loi des coefficients individuels nécessite des données d'enquête pour lesquelles sont connues les consommations des individus à l'intérieur de chaque ménage.

Nous exploitons dans cette partie les résultats de l'enquête alimentaire individuelle réalisée en 1988 dans le département du Val de Marne[4] par l'Institut Scientifique et Technique de l'Alimentation, sous la direction du docteur Serge Hercberg.

Cette enquête consistait en un relevé sur une semaine des consommations des ménages pour des aliments regroupables selon la nomenclature suivante :

- viande, charcuterie et oeufs
- poissons et crustacés
- produits laitiers (lait, yaourts et desserts lactés, fromage)
- produits à base de céréales (riz, pâtes, pain et farine)
- sucre et produits sucrés (biscuits, pâtisseries, bonbons, chocolat, viennoiseries)
- fruits et légumes
- matières grasses (beurre, huile et autres matières grasses)
- boissons sans alcool (boissons aux fruits, eau, café et thé)
- boissons alcoolisées

Nous pouvons par conséquent étudier les lois de probabilité des coefficients individuels associés à la consommation de ces différents types d'aliments, selon les trois types d'individus que nous avons retenus.

Ces résultats sont donnés sous forme graphique dans les pages qui suivent. On représente, pour des effectifs de ménages égaux à 2, 3 ou 4, les estimations des lois que prennent la part d'un adulte homme, d'une adulte femme et d'un enfant dans la consommation alimentaire totale des ménages, par groupes d'aliments. C'est donc une part qui est donnée en abscisse.

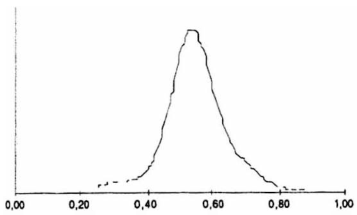
[4] L'objectif de cette enquête alimentaire était de mesurer les apports en micronutriments (minéraux et vitamines) au sein d'un échantillon individuel, représentatif de la population du Val de Marne selon les critères d'âge et de sexe. L'échantillon analysé a été sélectionné aléatoirement dans 12 des 47 communes du département à partir de l'annuaire téléphonique ; 527 familles sur les 849 familles contactées après tirage au sort ont accepté de participer à l'étude, soit au total 1064 individus âgés de 6 mois à 97 ans.

Pour des raisons de représentativité dans l'échantillon, les ménages mono-parentaux ne sont pas étudiés.

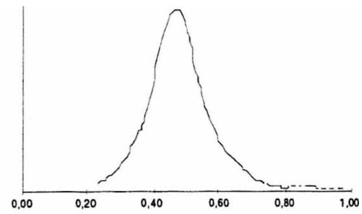
Viande, charcuterie et oeufs

Ménage de 2 personnes

homme

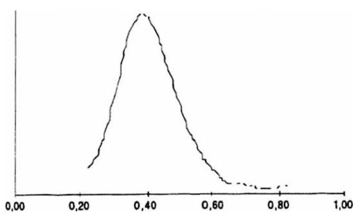


femme

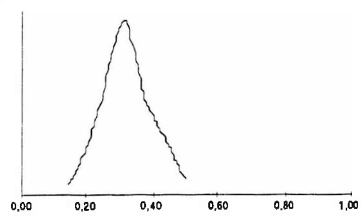


Ménage de 3 personnes

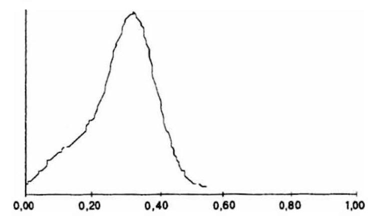
homme



femme

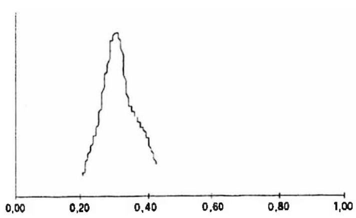


enfant

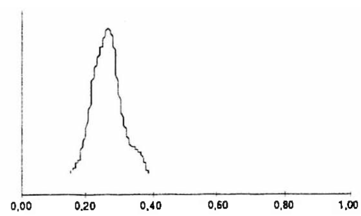


Ménage de 4 personnes

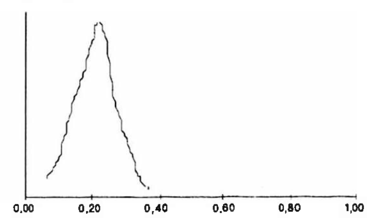
homme



femme



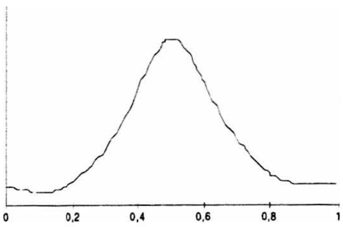
enfant



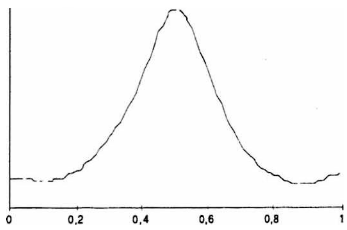
Poissons et crustacés

Ménage de 2 personnes

homme

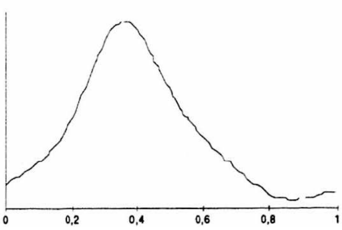


femme

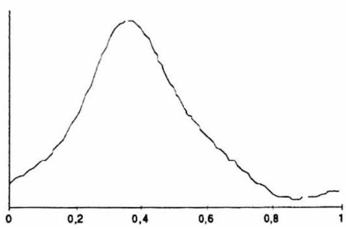


Ménage de 3 personnes

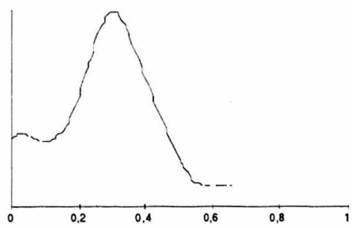
homme



femme

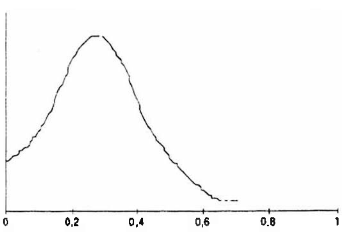


enfant

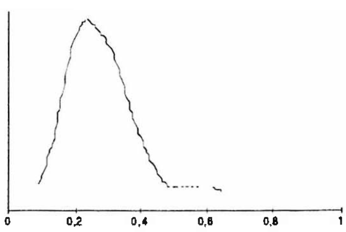


Ménage de 4 personnes

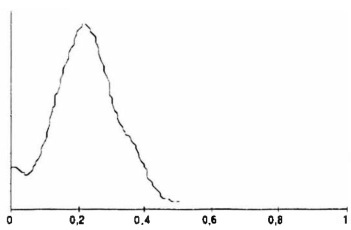
homme



femme



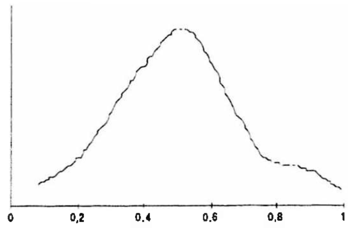
enfant



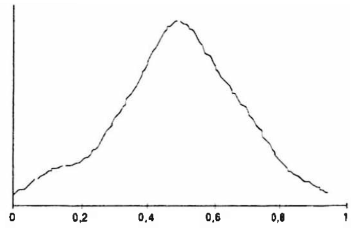
Produits laitiers

Ménage de 2 personnes

homme

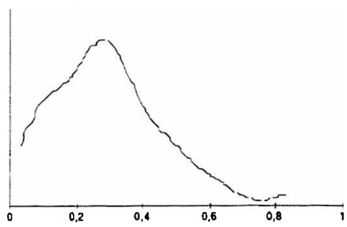


femme

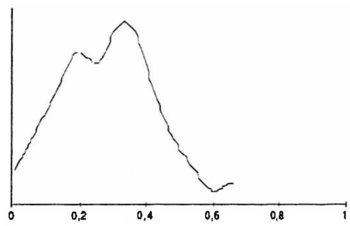


Ménage de 3 personnes

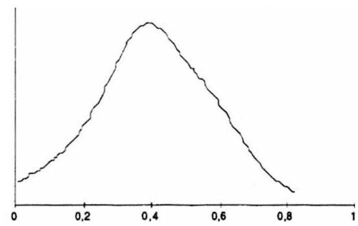
homme



femme

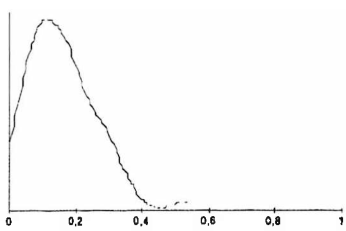


enfant

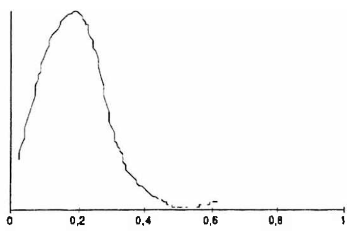


Ménage de 4 personnes

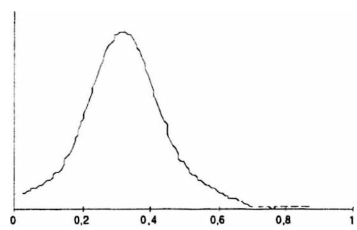
homme



femme



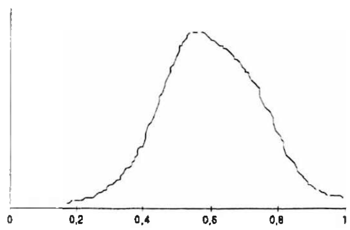
enfant



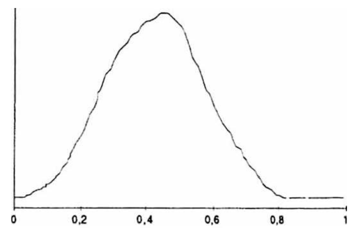
Produits à base de céréales

Ménage de 2 personnes

homme

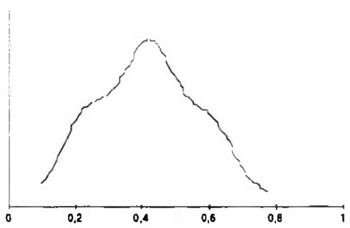


femme

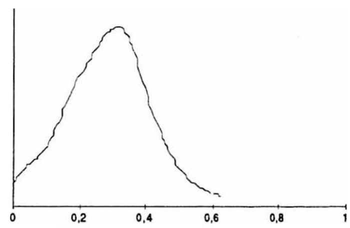


Ménage de 3 personnes

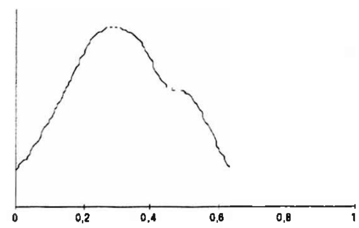
homme



femme

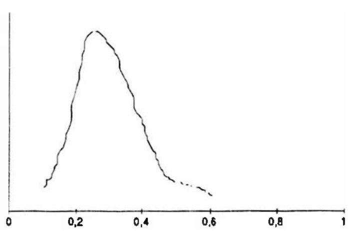


enfant

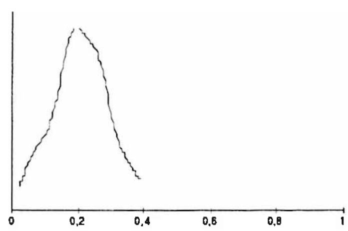


Ménage de 4 personnes

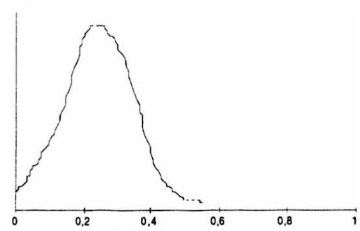
homme



femme



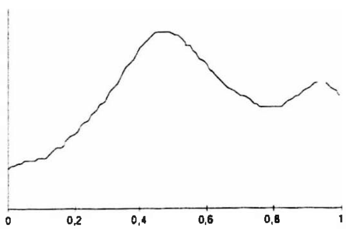
enfant



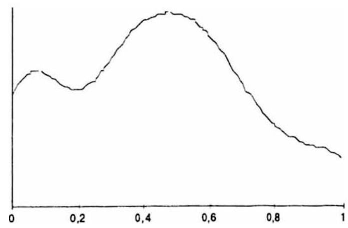
Sucre et produits sucrés

Ménage de 2 personnes

homme

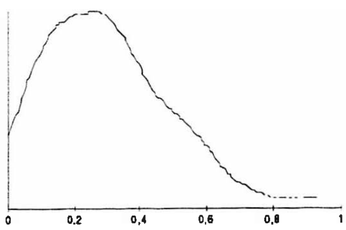


femme

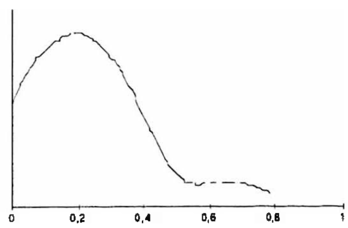


Ménage de 3 personnes

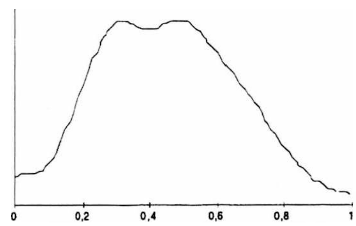
homme



femme

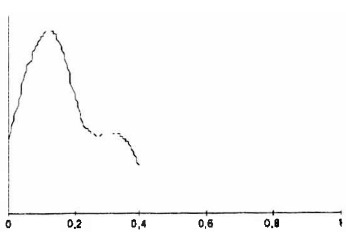


enfant

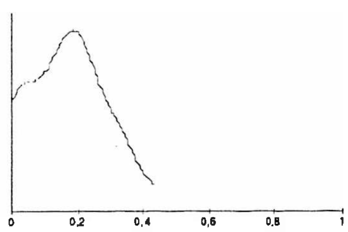


Ménage de 4 personnes

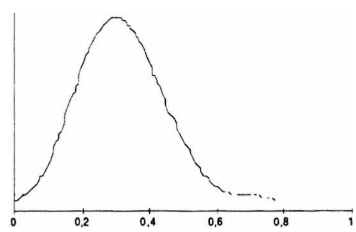
homme



femme



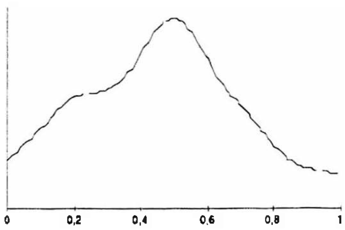
enfant



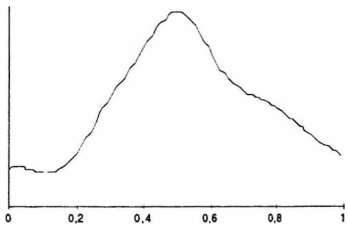
Fruits

Ménage de 2 personnes

homme

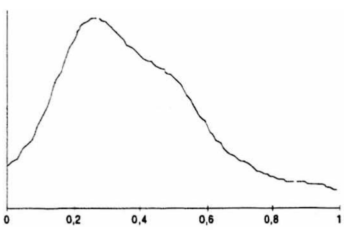


femme

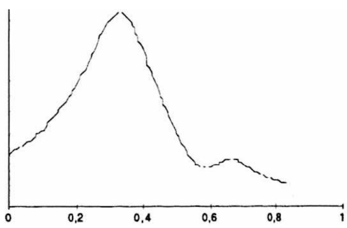


Ménage de 3 personnes

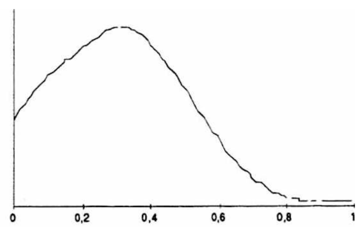
homme



femme

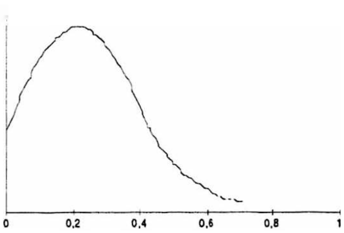


enfant

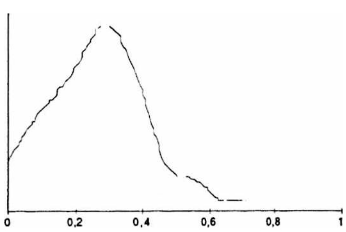


Ménage de 4 personnes

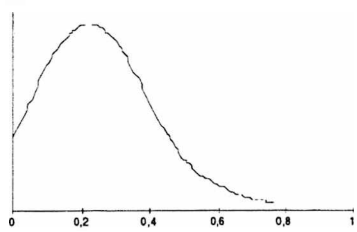
homme



femme



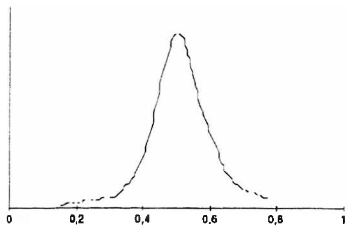
enfant



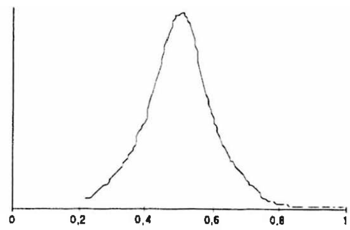
Légumes

Ménage de 2 personnes

homme

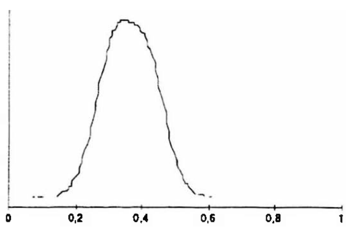


femme

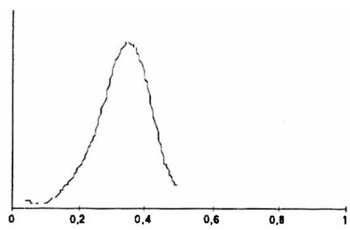


Ménage de 3 personnes

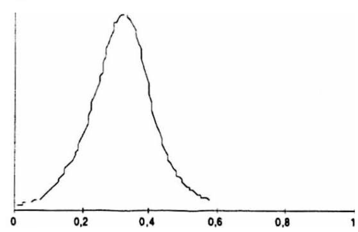
homme



femme

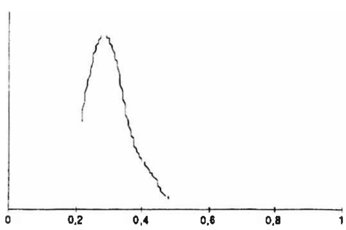


enfant

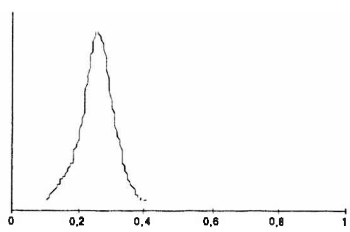


Ménage de 4 personnes

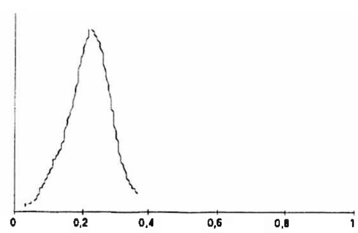
homme



femme



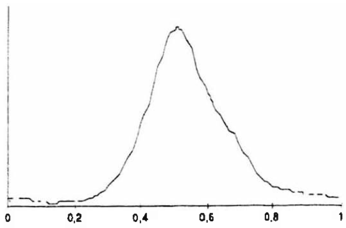
enfant



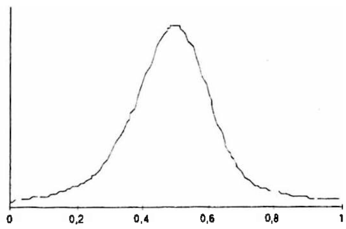
Matières grasses

Ménage de 2 personnes

homme

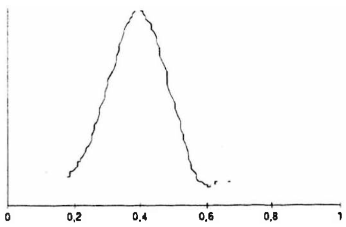


femme

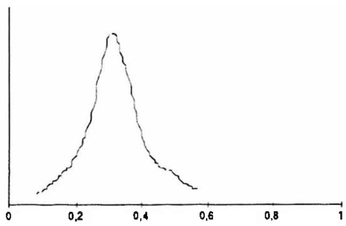


Ménage de 3 personnes

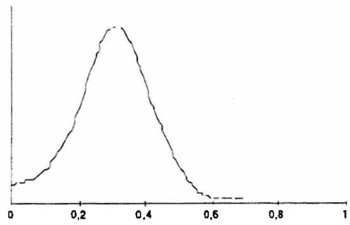
homme



femme

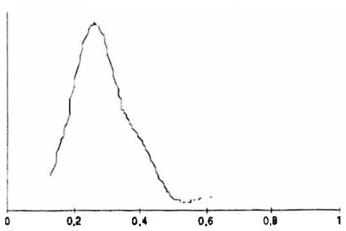


enfant

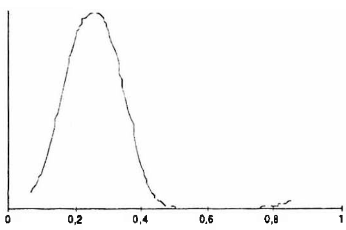


Ménage de 4 personnes

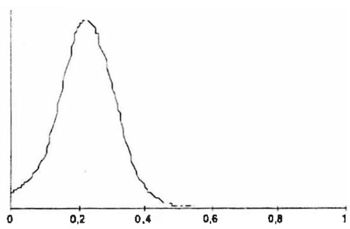
homme



femme



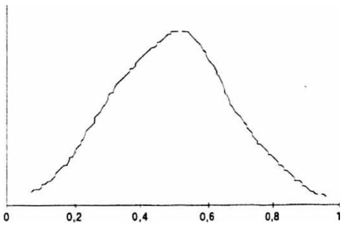
enfant



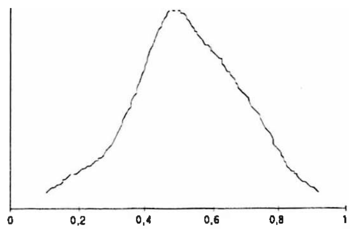
Boissons non alcoolisées

Ménage de 2 personnes

homme

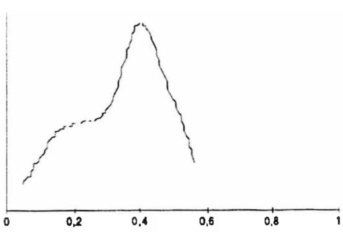


femme

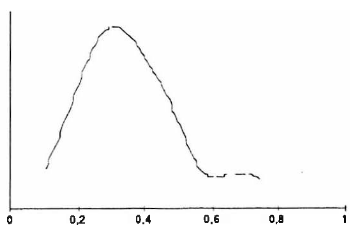


Ménage de 3 personnes

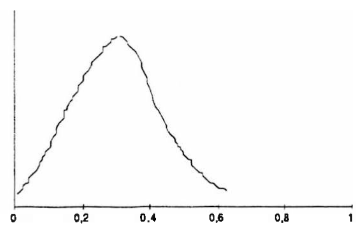
homme



femme

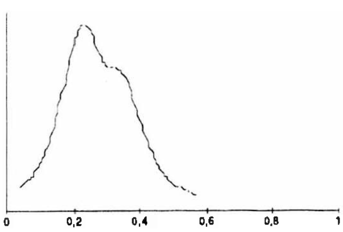


enfant

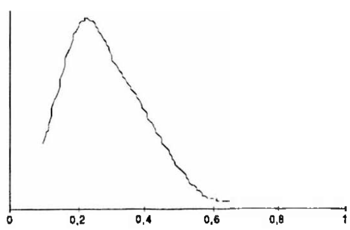


Ménage de 4 personnes

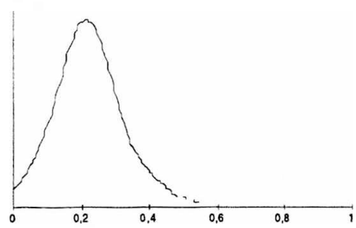
homme



femme



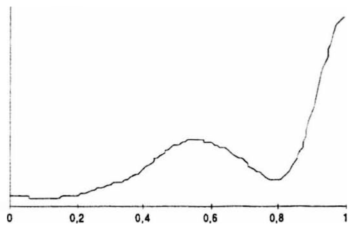
enfant



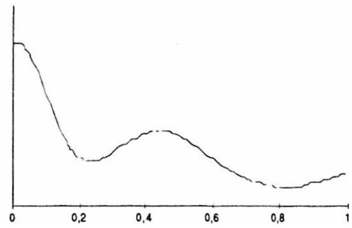
Boisson alcoolisées

Ménage de 2 personnes

homme

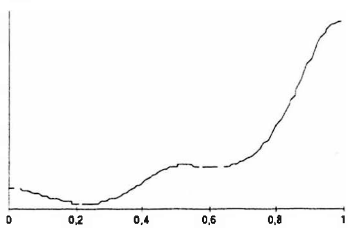


femme

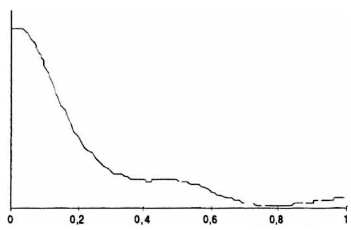


Ménage de 3 personnes

homme



femme

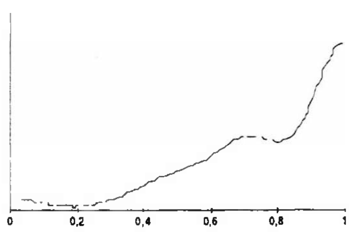


enfant

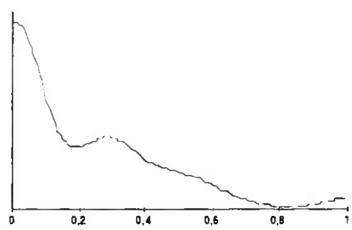
non significatif

Ménage de 4 personnes

homme



femme



enfant

non significatif

Tests d'adéquation

La partie très descriptive qui précède peut aboutir à une recherche de spécification des lois de probabilités obtenues par notre modélisation non paramétrique. La connaissance d'une loi de la part individuelle dans le ménage qui appartiendrait à une famille de lois de probabilité précise rendrait possible une meilleure description des phénomènes, et pourrait rendre envisageable l'estimation de consommations individuelles par l'application de méthodes de *bootstrap*. Il s'agirait en effet de dupliquer les individus des ménages avec un poids tiré aléatoirement dans les réalisations d'une variable aléatoire continue précise.

Une spécification des lois de probabilités des coefficients serait bien sûr possible dans notre contexte, surtout pour les aliments qui font apparaître des lois de forme familière (courbes quasi gaussiennes pour les légumes, la viande...). Toutefois, l'étude descriptive qui précède permet de mettre en évidence un grand nombre de lois contaminées (le cas des boissons alcoolisées est exemplaire), et une spécification devient dès lors délicate.

Inférence d'une règle d'allocation

A partir des données de l'enquête "Val-de-Marne", on a pu appliquer la méthode non paramétrique d'estimation de la densité. L'inférence d'une règle d'allocation est donc elle aussi possible. Pour cela, nous avons appliqué une méthode de détermination d'un coefficient moyen par type d'individu, calculé à partir d'une moyenne pondérée par la valeur du noyau de convolution ayant permis de spécifier la densité.

Moyenne pondérée par un noyau de convolution
(d'après Lejeune, 1985)

Le calcul des coefficients individuels présentés ci-après consiste à inférer un indicateur de tendance centrale des valeurs des coefficients qui sont observées dans la population des individus. Pour cela, nous tirons profit de l'estimation de la densité qui précède en tenant compte de la déformation de la courbe de répartition des coefficients. Il s'agit donc d'établir une valeur moyenne robuste $\overline{X}^{i}_{K,n}$ du coefficient individuel des individus de type i en pondérant les réalisations de la variable aléatoire "coefficient individuel" par la valeur du noyau K prise en chaque point.

D'où la valeur du coefficient individuel estimé :

$$\overline{X}^i_{K,n} = \frac{\sum_j K\left(\frac{x-x_j}{h_n}\right).x_j}{\sum_j K\left(\frac{x-x_j}{h_n}\right)}$$

Cet estimateur du coefficient individuel peut être considérée comme un indicateur fiable du moment d'ordre 1 de la loi de ce moment.

Dans l'état actuel de la recherche de spécification d'un tel modèle, les problèmes de contamination éventuelle des densités obtenues n'ont pas été pris en compte. De plus, aucune technique visant à appliquer une contrainte sur la valeur des coefficients n'étant pour l'instant appliquée, la somme des coefficients individuels pour chaque type de ménage n'égale pas 100%.

Il est enfin à noter qu'en l'absence de données individuelles sur un grand échantillon, il n'est pas aisé d'éprouver la robustesse des estimations présentées ci-après.

Dans ce contexte quelque peu restrictif, le tableau suivant donne, pour chaque type d'individu envisagé, l'estimation de la part (en pourcentage) qu'il prend dans la consommation du ménage, pour chaque groupe d'aliments considérés, en fonction de l'effectif du ménage. On exclut ici les familles monoparentales, trop peu nombreuses dans l'échantillon dont nous disposons. On envisage par conséquent trois cas : les ménages composés de deux adultes, avec 0, 1 ou 2 enfants.

Tableau 4 : Part de chaque type d'individu dans la consommation du ménage (en pourcentage)

Effectif du ménage / Groupe	Homme 2	Homme 3	Homme 4	Femme 2	Femme 3	Femme 4	Enfant 3	Enfant 4
Viande, charcuterie et oeufs	54,1	39,9	31,3	47,2	31,7	26,3	30,7	21,4
Poissons et crustacés	51,1	37,6	28,8	50,0	34,5	26,4	30,0	22,3
Produits laitiers	49,9	26,9	14,0	50,1	28,4	17,9	41,2	32,0
Produits à base de céréales	59,7	41,5	28,8	42,9	29,0	21,1	31,7	24,9
Sucre et produits sucrés	55,7	26,4	14,6	43,7	20,7	17,2	41,2	32,0
Fruits	45,8	34,6	22,5	53,0	32,5	26,5	30,0	22,7
Légumes	50,8	36,2	29,1	50,2	34,5	26,0	32,1	22,6
Matières grasses	52,6	39,3	27,4	49,3	31,6	25,5	31,2	22,8
Boissons non alcoolisées	49,3	37,1	26,6	52,8	33,2	25,7	29,5	21,5
Boissons alcoolisées	63,8	83,8	77,8	33,1	13,0	20,7	ns	ns

Ces résultats montrent que la répartition *intra*-ménage de la consommation alimentaire n'est pas réellement uniforme. A ce titre, ils s'accordent avec ceux déjà mis en évidence par un grand nombre d'économistes, et ils justifient *a posteriori* la démarche entreprise.

Pour la majeure partie des groupes d'aliments, les hommes s'avèrent consommer systématiquement une plus grande part que leur part "théorique" —au sens d'une équi-répartition de la consommation— à l'exception des fruits et des boissons non alcoolisées. En revanche, les boissons alcoolisées sont exclusivement masculines. La présence d'un enfant induit même en moyenne une chute de la part de la conjointe dans la consommation d'alcool.

Les enfants se voient quant à eux accorder une part plus grande de la consommation de produits laitiers et de produits sucrés.

CONCLUSIONS ET PERSPECTIVES DE RECHERCHE

Les champs de recherche statistique qui ont été partiellement couverts au cours de cette recherche mériteront dans l'avenir de nouveaux développements. En particulier, la recherche de points de convergences entre les différentes approches représente un sujet important car il n'y a pas de dichotomie radicale entre les deux formalisations économétrique et probabiliste que nous avons explorées.

Les développements à venir au CRÉDOC de nouvelles recherches de spécifications de modèles d'allocation *intra*-ménage de la consommation alimentaire devront en particulier tenir compte des demandes toujours plus précises d'informations individuelles sur les ingestions de nutriments. A ce titre, les domaines d'applications de ces méthodes restent plus que jamais de types à la fois socio-économiques et marketing —influence de la composition du ménage sur la consommation globale, interactions entre les consommations individuelles...

De plus, dans le cadre d'une surveillance des risques liés à l'hygiène alimentaire, l'individualisation des consommations devra apporter un nouveau souffle aux recherches et études actuellement en cours. Plus spécifiquement, les programmes de mesure des populations à risques au regard de certains aspects toxicologiques —par exemple la consommation de nitrates ou de pesticides dans les aliments— déjà engagés dans le cadre de l'Observatoire des Consommations Alimentaires, auquel le CRÉDOC apporte une contribution importante, devraient tirer le plus grand profit des avancées statistiques présentées dans cette recherche.

BIBLIOGRAPHIE

[1.] BOSQ Denis, LECOUTRE Jean-Pierre, (1987).- *Théorie de l'estimation fonctionnelle*, Economica.

[2.] CHESHER Andrew, (1991).- "Household Composition and Household Food Purchases", in *Fifty years of the National Food Survey, 1940-1990*, MAFF.

[3.] DEATON Angus, MUELLBAUER J., (1980).- "An almost Ideal Demand System", *American Economic Review,* n°70, juin.

[4.] DEATON Angus, (1992).- *Understanding Consumption*, Clarendon Press, Oxford.

[5.] LEJEUNE Michel, (1985).- "Estimation non-paramétrique par noyaux : régression polynomiale mobile", *Revue de Statistique appliquée*, vol. XXXIII, n°3.

[6.] RENAULT Chantal, (1993).- *Répartition des consommations des ménages entre ses différents membres*, CRÉDOC, Observatoire des Consommations Alimentaires, document n°11.

ANNEXE 1 : ÉTUDE EMPIRIQUE DE LA VITESSE DE CONVERGENCE DES ESTIMATEURS DU MODÈLE DE CHESHER VERS LES CONSOMMATIONS INDIVIDUELLES MOYENNES

En 1993, le CRÉDOC, dans le cadre du programme de l'Observatoire des Consommations Alimentaires, a publié des résultats préliminaires sur l'application du modèle de Chesher, dans le cas d'un échantillon de faible taille (Renault, 1993). Ce travail exploitait les données de l'enquête "Val de Marne", que nous avons nous-mêmes utilisées pour l'illustration empirique de notre approche probabiliste. L'avantage de ces données était de permettre une comparaison directe des estimateurs obtenus avec les vraies moyennes, du fait de la nature des données de l'enquête.

Notre propos n'est pas ici de reprendre cette étude, mais simplement d'en utiliser les résultats à titre d'illustration de la rapidité de convergence des coefficients de la régression de Chesher vers les moyennes individuelles globales. Nous somme en effet dans le cadre d'un échantillon de faible taille, et cette étude permet donc de se rendre compte des possibilités offertes par les MCO pour modéliser les consommations individuelles.

Le modèle estimé sur l'ensemble des aliments donne des estimations très proches des moyennes vraies, ce qui est déjà un résultat important.

Tableau 5 : Modèle de Chesher comparé aux vraies valeurs des moyennes

	Coefficient du modèle	Moyenne
Enfant de moins de 5 ans	1853 11,9	1908
Enfant de 5 à 11 ans, sexe masculin	2498 13,5	2568
Enfant de 5 à 11 ans, sexe féminin	1990 9,4	2221
Adolescent de 12 à 17 ans, sexe masculin	3117 15,0	3176
Adolescent de 12 à 17 ans, sexe féminin	2592 13,5	2682
Adulte de 18 à 65 ans, sexe masculin	3334 39,2	3266
Adulte de 18 à 65 ans, sexe féminin	2768 36,7	2835
Personne âgée de 65 ans et +, sexe masculin	2798 12,1	2869
Personne âgée de 65 ans et +, sexe féminin	2644 15,6	2559

Le modèle de Chesher est en revanche beaucoup moins efficace pour estimer, sur petit échantillon, les consommations individuelles dans des cas atypiques. C'est en particulier le cas du modèle estimé sur la consommation de boissons alcoolisées, dont nous avons vu le caractère fortement asymétrique dans la partie consacrée à l'approche probabiliste.

Le tableau suivant donne à ce titre, selon le même schéma que précédemment, les coefficients du modèle comparés aux vraies moyennes, dont ils diffèrent sensiblement.

Tableau 6 : Modèle de Chesher comparé aux vraies valeurs des moyennes pour les boissons alcoolisées

	Coefficient du modèle	Moyenne
Adulte de 18 à 65 ans, sexe masculin	292,0 9,6	343,6
Adulte de 18 à 65 ans, sexe féminin	66,3 2,3	112,3
Personne âgée de 65 ans et +, sexe masculin	394,9 4,6	386,9
Personne âgée de 65 ans et +, sexe féminin	123,9 1,7	128,6

C'est probablement ici moins l'effectif que la forte asymétrie du phénomène de répartition *intra*-ménage de la consommation qui "trouble" le modèle, ce qui donne une idée du type de problèmes que l'on pourra rencontrer dans une telle application.

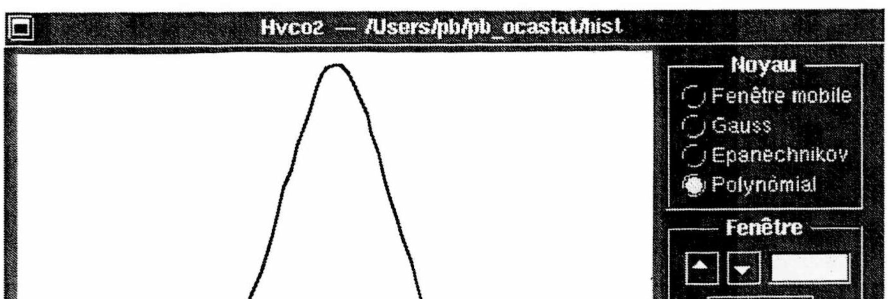
ANNEXE 2 : OUTIL INFORMATIQUE POUR L'ESTIMATION DE LA DENSITÉ PAR LA MÉTHODE DU NOYAU

Le choix d'une méthode non paramétrique d'estimation de la densité a dû s'accompagner d'un développement informatique adapté, les outils disponibles traditionnellement au CRÉDOC ne permettant pas de la mettre en oeuvre.

Tous les travaux statistiques ayant été effectués sous environnement UNIX avec le logiciel SAS, nous sommes restés dans le même environnement pour développer une petite application permettant d'appliquer directement la méthode non paramétrique d'estimation à des "sorties" au format ASCII de SAS. Ce développement a été fait sur une station de travail NeXT, mise en réseau avec les stations de travail sur lesquelles est installé SAS. Le langage de programmation utilisé est une variante orientée objet du langage C.

Cette application permet d'appliquer les quatre types de noyaux indiqués dans le corps de cette recherche : fenêtre mobile, gaussien, Epanechnikov et polynômial. On obtient une estimation de la densité en faisant varier la taille de la fenêtre (paramètre h), ainsi que la moyenne pondérée par le noyau.

La copie d'écran suivante donne une représentation de l'interface de l'application.



L'application peut donner en sortie un nouveau fichier au format ASCII, intégrable dans un logiciel bureautique afin de produire une courbe de densité.

Dépôt légal : Juin 1995

ISSN : 1257-9807

ISBN : 2-84104-035-6

CAHIER DE RECHERCHE

Récemment parus :

Parcours d'insertion de jeunes en difficulté
Anne-Lise AUCOUTURIER, Valérie BEAUDOUIN - **n°66 (1994)**

Le sentiment de restrictions
Franck BERTHUIT, Aude COLLERIE de BORELY et Anne-Delphine KOWALSKI - **n°67 (1995)**

Les spécificités des enquêtes quantitatives auprès de populations socialement marginales
Marie-Odile GILLES - **n°68 (1995)**

Articles de sociologie
Michel MESSU, Christine LE CLAINCHE - **n°69 (1995)**

Méthode d'étude sectorielle
Philippe MOATI - **n°70 (1995)**

Le consommateur de 1995 face à la reprise économique
Aude COLLERIE DE BORELY, Jean-Luc VOLATIER - **n°71 (1995)**

Design et forme naturelle de l'objet : la mise au point d'un outil de design et d'ingénierie de l'immatériel
Patrick BABAYOU, Valérie BAUDOUIN, Aude COLLERIE DE BORELY, Chantal RENAULT, Cécile STEPANCZAK et Jean-Luc VOLATIER - **n°72 (1995)**

Président : Bernard SCHAEFER   Directeur : Robert ROCHEFORT
142, rue du Chevaleret, 75013 PARIS - Tél. : (1) 40.77.85.00

ISBN : 2-84104-035-6

CRÉDOC

Centre de recherche pour l'Étude et l'Observation des Conditions de Vie