

LA METHODOLOGIE DE M.O.R.G.O.A.T. : ENQUETE TOURISME EN BRETAGNE

Jean-Claude Deville, Myriam Maumy¹

RÉSUMÉ

On doit procéder à une enquête portant sur la fréquentation touristique intra et extra-régionale en Bretagne. Pour des raisons matérielles concrètes, les "enquêtes aux frontières" ne peuvent plus s'organiser. Le problème majeur est l'absence de base de sondage permettant d'atteindre directement les touristes. Pour contourner ce problème, on applique la méthode d'échantillonnage indirect ou encore appelée la méthode généralisée du partage des poids développée récemment par Lavallée (2002). Cet article montre comment adapter cette méthode à l'enquête.

1. INTRODUCTION

Une "enquête aux frontières" portant sur la fréquentation touristique d'origine extra-régionale en Bretagne (hormis celle des Bretons) a été réalisée sur la période d'avril à septembre 1997. L'Observatoire Régional du Tourisme de Bretagne (O.R.T.B.), les Comités Départementaux de Tourisme et les Chambres de Commerce et d'Industrie de Bretagne souhaiteraient renouveler ce type d'enquête. Malheureusement ils n'ont plus la possibilité de recueillir une certaine masse d'informations récoltés aux frontières régionales ou intra-régionales, car les forces de la gendarmerie ne collaborent plus à la réalisation d'enquêtes touristiques au bord des routes.

C'est pourquoi l'O.R.T.B. avec l'aide d'un comité technique constitué de méthodologues et d'opérateurs de terrain ont décidé de mettre en place une nouvelle méthodologie d'enquête en remplacement de la méthodologie des "enquêtes aux frontières" (ou enquête cordon).

Un des problèmes majeurs est l'absence d'une base de sondage permettant d'interroger directement les touristes. Pour contourner ce problème, l'idée principale déjà utilisée par la région des Asturies (2001) en Espagne est d'échantillonner des services destinés principalement aux touristes et de les interroger sur les différents lieux de ces nombreuses prestations touristiques. Il est bien évident qu'un touriste peut utiliser une ou plusieurs fois un ou plusieurs services de la base de sondage pendant son séjour. Pour pouvoir estimer des paramètres d'intérêts relatifs aux touristes, il faut relier le jeu de poids des services échantillonnés au jeu de poids des touristes qui ont fréquenté ces services. Le but de cette note est de présenter une méthode qui permet de faire ce calcul. Cette méthode va s'appuyer principalement sur la *Méthode Généralisée de Partage des Poids* (MGPP) mise au point par Lavallée (1995, 2002) et Deville (1999).

2 LA METHODE GENERALISEE DE PARTAGE DES POIDS

On va rappeler très brièvement le principe de la MGPP. Pour de plus amples informations, on renvoie à Lavallée (2002). Notons : U_A une population de N_A individus, où chaque individu est noté j et U_B une population de N_B individus, où chaque individu est noté i . La liaison qui existe entre les 2 populations est

¹ Deville Jean-Claude, Laboratoire de Statistique d'Enquêtes, ENSAI - Campus de Ker Lann - rue Blaise Pascal - BP37203 - 35172 - BRUZ CEDEX, FRANCE, (deville@ensai.fr), Maumy Myriam, Laboratoire de Statistique de Rennes 2, Université Haute Bretagne, Place du Recteur H. Le Moal, CS 24307, 35043 RENNES CEDEX, FRANCE, (myriam.maumy@uhb.fr).

représentée par un *tableau de liens* Θ_{AB} , de taille $N_A \times N_B$, où chaque élément est noté θ_{ji}^{AB} . Autrement dit, l'individu j de U_A est relié à l'individu i de U_B à condition que : $\theta_{ji}^{AB} > 0$; sinon $\theta_{ji}^{AB} = 0$. Le processus du sondage indirect est le suivant :

- sélection de l'échantillon s_A de taille n_A à partir de U_A selon un plan d'échantillonnage;
- $\pi_j^A > 0$ désigne la probabilité de sélection de l'individu j de la population U_A ;
- pour chaque j sélectionné dans s_A , identification des individus i de U_B pour lesquels $\theta_{ji}^{AB} > 0$;
- obtention de l'échantillon s_B de taille n_B où n_B identifiées au moyen des individus j qui appartiennent à s_A , c'est-à-dire $s_B = \{i \in U_B ; \exists j \in s_A \text{ et } \theta_{ji}^{AB} > 0\}$.
- pour chaque i de s_B , mesure d'une variable d'intérêt y_i auprès de l'individu i .

Remarque importante : Dans cette méthode, on ne doit connaître les θ_{ji}^{AB} du tableau de liens que pour j appartenant à s_A et i appartenant à s_B .

Exemple : On veut estimer un total T_B de la population cible U_B qui se définit par :

$$T^B = \sum_{i=1}^{N^B} y_i, \quad (2.1)$$

avec y_i mesurées auprès de l'ensemble des individus de la population U_B . On utilise un estimateur :

$$\hat{T}^B = \sum_{i=1}^{N^B} w_i y_i, \quad (2.2)$$

où w_i est le poids de i de s_B , avec $w_i = 0$ pour $i \notin s_B$. Pour obtenir une estimation sans biais de T_B , il suffit d'utiliser comme poids w_i l'inverse de la probabilité de sélection π_i^B de i . Il est généralement impossible d'obtenir ces probabilités. On a alors recours à la MGPP. L'intérêt de cette méthode est de calculer des poids w_i qui fournissent un estimateur de sans biais de T_B . Les poids sont donnés par :

$$w_i = \sum_{j \in s^A} \frac{\tilde{\theta}_{ji}^{AB}}{\pi_j^A}, \quad (2.3), \text{ avec} \quad \tilde{\theta}_{ji}^{AB} = \frac{\theta_{ji}^{AB}}{\sum_{j=1}^{N^A} \theta_{ji}^{AB}}, \quad (2.4)$$

3 L'ENQUETE TOURISME EN MILIEU OUVERT

3.1 Principe de l'enquête et hypothèse principale à la réalisation de celle-ci

Le principe de l'enquête est le suivant : "*atteindre les touristes (étrangers ou français résidant ou non en Bretagne) par le biais de services destinés à satisfaire leurs besoins élémentaires*" comme l'hébergement, la nourriture, les activités de loisirs, les transports. L'hypothèse principale à la réalisation de l'enquête est la suivante : *on admet que tout ménage touristique consomme au moins un des services (achats en boulangeries, visites de sites emblématiques de la Bretagne, passage au péage autoroutier de La Gravelle), ou tout du moins, que très peu de ménages touristiques ne consomment aucun d'entre eux.*"

3.2 La population d'intérêt

Soit G un **champ géographique**. Ici dans l'enquête G représentera les quatre départements bretons. Soit P une **période de référence**. Pour nous, la période de référence est celle qui s'étend du mois de février 2005 au mois de décembre 2005. Un **touriste** t est une personne ayant passé au moins une nuit dans le champ géographique G hors de sa résidence principale (nuitée). Pour un touriste t , un **séjour** est un intervalle s de la période P de durée le cardinal du séjour s noté $|s|$, au cours duquel le touriste passe toutes ses nuits dans

le champ géographique G hors de sa résidence principale et, les nuits immédiatement avant ou après le séjour s étant passées hors du champ géographique G . Par exemple il peut avoir passé ces nuits-là à sa résidence principale ou dans tout autre hébergement qui ne se situe pas dans le champ géographique G . **Un voyage** ou encore **un ménage touristique** est un ensemble de touristes partageant le même séjour et avec le même hébergement au cours du séjour. **L'unité statistique i** de l'enquête est le ménage touristique. **Les sous unités d'enquête** sont les séjours, les touristes et les nuitées. Un voyage v comporte n_v touristes pendant le séjour de durée $|s|$ et donc $|s|n_v$ nuitées. Ici la population U_B est donc l'ensemble des voyages dans le champ géographique G au cours de la période P . **Le champ de l'enquête** : la méthodologie permet de décrire principalement le tourisme d'agrément puisque les principaux points d'enquête sont des sites de loisirs et des boulangeries. En effet, il y aura peu de touristes d'affaires attrapés car nous n'enquêterons pas dans les modes d'hébergement marchand, en particulier les hôtels qui nous permettraient d'atteindre cette cible.

3.3 Le plan de sondage de l'enquête

Pour utiliser la MGPP, la population U_A est constituée par un ensemble de « services ». Dans cette enquête, ceux-ci sont :

- les achats en boulangerie, constituant une première sous-population de la population U_A ;
- les visites de 16 sites emblématiques de la Bretagne qui se divisent en trois familles : les sites naturels, les sites patrimoniaux, les sites familiaux. En pratique, pour chacun d'eux, un « point de passage obligé » a été défini. C'est l'ensemble des passages par ce point qui est la seconde sous-population de la population U_A ;
- les passages sortant de Bretagne au péage autoroutier de La Gravelle. La voiture caractérise 80% des séjours de non-résidents bretons. Ces passages constituent la troisième sous-population de la population U_A .

Dans *la première sous-population*, on réalise un échantillon à 3 degrés :

1. un échantillon de boulangeries : 19 boulangeries;
2. un échantillon de jours d'enquête ;
3. un échantillon de clients dans les 19 boulangeries sélectionnées à un jour donné.

Dans *la deuxième sous-population*, on réalise un échantillon à 2 degrés :

1. un échantillon de jours d'enquête ;
2. un échantillon de personnes qui passent sur un des 16 sites référés à un jour donné.

Enfin dans *la troisième sous-population*, on réalise un échantillon à 2 degrés :

1. un échantillon de jours d'enquête ;
2. un échantillon de personnes qui passent au péage autoroutier de La Gravelle à un jour donné.

3.4 Les spécificités de M.O.R.G.O.A.T.

La première spécificité : le bâtonnage. Le système de bâtonnage est une mini-enquête mise en œuvre préalablement à la passation du questionnaire. Ce système a essentiellement **deux objectifs** :

1. déterminer les proportions des touristes étrangers, des touristes français non bretons et des touristes intra-régionaux. C'est l'objectif fondamental du bâtonnage puisque ces proportions seront utilisées lors du calcul des poids. Il est donc impossible de supprimer le bâtonnage du dispositif de l'enquête. Ces proportions sont inconnues quel que soit le lieu enquête. Par exemple, on peut obtenir le nombre de visiteurs journaliers d'un zoo, d'un château ou d'un musée mais la part de visiteurs étrangers est inconnue. De même, on peut obtenir le nombre de clients journaliers venus effectuer un achat en boulangerie et même le nombre d'acheteurs pendant les heures d'enquête dans la boulangerie sélectionnée ; mais la proportion d'achats des locaux dans la boulangerie reste inconnue pour les artisans boulangers.
2. Filtrer les populations hors champs, c'est-à-dire les excursionnistes, les autocaristes et les locaux. C'est un objectif d'ordre pratique.

La seconde spécificité : les sites en rase campagne. Sur certains sites de loisirs, en particulier ceux qui ne sont pas dotés de systèmes de billetteries, nous ne connaissons pas le nombre de visiteurs journaliers. Ce

nombre est nécessaire pour appliquer la MGPP. Comme ce total est inconnu, il va falloir obtenir une estimation de ce nombre de visiteurs journaliers. Nous avons donc mis en place des compteurs, généralement placés à l'entrée des parkings et qui comptent le nombre de voitures qui franchissent l'entrée du parking. Ces compteurs nous donnent un nombre de véhicules et non un nombre d'individus. Or sur les autres sites, notre référentiel est le nombre d'individus visitant le site. Pour disposer du même référentiel, un enquêteur placé sur le parking ou sur le site (nous discuterons par la suite de quel dispositif nous devons choisir afin d'obtenir le dispositif le plus précis en terme de variance et le plus économique) demande aux visiteurs si la voiture qui les a conduit est garée sur le parking du site et dans le cas échéant combien de voyageurs étaient présents dans ce véhicule. Il ne reste plus qu'à relier le nombre de voitures noté T_V à l'estimateur du nombre total de visiteurs \hat{T}_P .

Remarque : Le nombre de voitures stationnées sur le parking du site en rase campagne T_V est parfaitement connu aux erreurs de mesure près puisqu'il est fourni par un compteur installé par l'O.R.T.B. spécialement pour les besoins de la méthodologie de l'enquête.

3.5 Les paramètres d'intérêt

Les paramètres d'intérêt de l'enquête sont des totaux, des effectifs et des ratios. Supposons par exemple, que l'on s'intéresse à l'estimation d'un total relatif à une variable y définie sur la population U_B

$$T^B = \sum_{i \in U^B} y_i, \quad (3.1)$$

où y_i est une variable mesurée auprès de l'individu i . Dans le cadre de l'enquête, ici, l'individu i est un des membres du ménage touristique i . Un cas particulier de ces totaux est le nombre total de touristes venus visiter la Bretagne. Le total T_B peut-être aussi : le nombre de personnes ayant pratiqué une certaine activité, le budget total dépensé par le ménage touristique à l'intérieur de la Bretagne, la provenance géographique des ménages touristiques, le nombre de jours que le ménage touristique passe en Bretagne...

Remarque importante : Pour beaucoup de variables, le total dépend de la taille du ménage touristique et de la longueur du séjour (uniquement les jours passés en Bretagne).

Abordons maintenant la question qui semble la plus importante : *Comment allons nous définir le total T_B défini en (3.1) dans le cadre de la MGPP ?* Pour répondre à cette interrogation, nous allons définir l'application F , qui à tout service j durant la période de référence P dans les 3 types d'établissements du champ de l'enquête, associe le ménage touristique i utilisateur de ce service.

$$\begin{array}{lcl} F : \text{services} & \rightarrow & \text{ménage touristique} \\ j & \rightarrow & F(j) = i. \end{array} \quad (3.2)$$

Soit U_B , la population des ménages touristiques i de la période P . Cette population d'intérêt U_B est l'image par F de l'ensemble des services de U_A utilisés durant la période P dans les 3 types d'établissements de l'enquête. La population U_A (ici l'ensemble des services utilisés par les touristes en Bretagne) est l'image par la fonction réciproque de F de l'ensemble des ménages touristiques i durant la période P . Pour tout individu i de la population des voyages U_B , on définit :

$$N_{ij} = \text{card}(F^{-1}(i)), \quad (3.3)$$

autrement dit, le nombre d'antécédents de i au cours de la période P , ou encore, le nombre de services j utilisés par le ménage touristique i donné pendant le séjour s . Désormais, T_B s'écrit :

$$T^B = \sum_{l=1}^3 \sum_{a_l \in A_l} \sum_{d_l \in D_l} \sum_{j \in C_{d_l}} z_j, \quad (3.4)$$

$$z_j = \frac{y_i}{N_{ij}}, \quad \text{pour } j \in F^{-1}(i). \quad (3.5)$$

A_1 désigne l'ensemble des boulangeries de l'enquête repéré par l'indice a_1 , A_2 désigne les 16 sites de l'enquête repérés par l'indice a_2 , A_3 désigne le péage de La Gravelle repéré par l'indice a_3 , D_1 désigne l'ensemble des jours d'enquête, repérés par l'indice d_1 dans un établissement a_1 de A_1 , pour l variant de 1 à 3,

C_{dl} désigne l'ensemble des services dans un établissement a_l de l'ensemble des établissements A_l de la journée d_l de l'ensemble des jours D_l repérés par l'indice j .

3.6 Calcul des estimateurs des paramètres d'intérêt

Dans le paragraphe précédent, nous avons montré que le total s'écrit comme un total sur l'ensemble des services du champ de l'enquête. Supposons que l'on dispose d'un échantillon de services répondants j , auxquels on peut associer des poids de sondage δ_j . Ces poids seront supposés sans biais. Disposant d'un jeu de poids de sondage δ_j pour les services répondants, et si on connaît le nombre de services N_{ij} utilisés par le ménage touristique i , on estime alors le total T_B sans biais par :

$$\hat{T}^B = \sum_{i \in s^B} w_i \frac{y_i}{N_{ij}} \quad (3.6), \text{ où} \quad w_i = \sum_{l=1}^3 \sum_{s_{A_l}} \sum_{s_{D_l}} \sum_{s_{d_l}} \delta_j, \quad (3.7)$$

et où : s^B est l'ensemble des ménages touristiques i correspondant à l'ensemble des services échantillonnés au cours de la période d'enquête, s_{A_l} est l'ensemble des établissements échantillonnés, s_{D_l} est l'ensemble des jours échantillonnés dans l'établissement a_l , et $s_{C_{dl}}$ est le sous-échantillon de services j correspondant au jour de l'établissement a_l . On est ramené à une estimation sur la population des ménages touristiques. Cette formule n'est autre que celle donnée par la MGPP. Notons que $\theta_{ji}^{AB} = 1$ si le service j a été utilisé par le ménage touristique i et que $\delta_j = 1 / \pi_j^A$. L'estimation de la variance de \hat{T}^B est possible selon les mêmes principes (cf. Lavallé (2002)). Elle ne sera pas détaillée ici car elle n'est qu'une application assez lourde en calcul des principes généraux.

3.7 Cas particulier de certains sites : les points de visite en rase campagne

Dans certains sites, on ne connaît pas le nombre total de personnes venant visiter le site ou se promener sur le site. En effet, dans l'ensemble des 16 sites emblématiques de la Bretagne noté A_2 on ne connaît pas tous les services (ici le nombre de visites) de la population car comme mentionné au paragraphe 3.4, certains sites parmi les 16 sélectionnés ne disposent pas de systèmes de billetterie ou de système équivalent capable de fournir le nombre exact de visiteurs. C'est ce nombre qui sert de dénominateur dans le calcul de la probabilité. On ne peut donc pas avoir directement la probabilité $\pi_j^{A_2}$ qui est un rapport de cardinaux (nombre de visiteurs interrogés par nombre total de visiteurs du site) et donc δ_j pour $j \in A_2$. Pour contourner ce problème, on estime alors le nombre de visiteurs journaliers afin de déduire une estimation de la probabilité $\pi_j^{A_2}$, rapport de la taille de l'échantillon des visiteurs interrogés sur l'un des 16 sites et de l'estimation du nombre de visiteurs de ce même site.

RÉFÉRENCES

Deville, J-C., (1999), « Les enquêtes par panel : en quoi différent-elles des autres enquêtes ? », suivi de « Comment attraper une population en se servant d'une autre », *Actes des journées de méthodologie statistiques*, INSEE Méthodes n°84-85-86.

Lavallée, P. (1995) « Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode du partage des poids », *Techniques d'enquête*, 21, p. 27-35.

Lavallée, P. (2002) : *Le Sondage Indirect, ou la méthode généralisée du partage des poids*, Editions de l'Université de Bruxelles, Bruxelles.

Valdés Pelaàez, L. et al. (2001) : « A methodology to measure tourism expenditure and total tourism production at the region level », Lennon, J. (Editor): *Tourism Statistics*. Continuum, London.