

Techniques d'enquête

Extension de la méthode de partage des poids lors de l'utilisation d'une base de sondage continue

par Guillaume Chauvet, Olivier Bouriaud et Philippe Brion

Date de diffusion : le 30 juin 2023



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par le ministre de l'Industrie 2023

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Extension de la méthode de partage des poids lors de l'utilisation d'une base de sondage continue

Guillaume Chauvet, Olivier Bouriaud et Philippe Brion¹

Résumé

La définition des unités statistiques est une question récurrente dans le domaine des enquêtes-échantillons. En effet, les populations sondées ne comportent pas toutes une base de sondage déjà disponible. Dans certaines populations, les unités échantillonnées sont différentes des unités d'observation, et la production d'estimations concernant la population d'intérêt soulève des questions complexes qu'il est possible de traiter en utilisant la méthode de partage des poids (Deville et Lavallée, 2006). Les deux populations prises en considération dans cette méthode sont toutefois discrètes. Dans certains champs d'études, la population échantillonnée est continue : c'est, par exemple, le cas des inventaires forestiers dans lesquels, souvent, les arbres sondés sont ceux situés sur des parcelles de terrain dont les centres sont des points tirés aléatoirement dans un secteur donné. La production d'estimations statistiques à partir de l'échantillon d'arbres sondés présente des difficultés d'ordre méthodologique, tout comme les calculs de variance qui y sont associés. Le présent article a pour but d'étendre la méthode de partage des poids au cas de populations continues (population échantillonnée) et de populations discrètes (population sondée), à partir de l'extension proposée par Cordy (1993) de l'estimateur de Horvitz-Thompson pour procéder à un tirage de points dans un univers continu.

Mots-clés : Estimation de la variance; inventaire forestier; plan d'échantillonnage continu; statistique de l'environnement; variable synthétique.

1. Introduction

La définition des unités statistiques est une question récurrente dans le domaine des enquêtes-échantillons. En effet, les populations sondées ne comportent pas toutes une base de sondage déjà disponible. Dans certains cas, les unités échantillonnées (pour lesquelles il existe une base de sondage permettant de sélectionner des unités selon un plan d'échantillonnage donné) sont distinctes des unités d'observation, qui constituent la population d'intérêt.

Cette question est soulevée depuis longtemps dans l'étude des populations difficiles à joindre, comme les sans-domicile fixe (consulter, par exemple, Ardilly et Le Blanc, 2001; De Vitiis, Falorsi et Inglese, 2014; Laporte et coll., 2018), ou les populations nomades ou non localisées (consulter, par exemple, Lohlé-Tart, Clairin, François et Gendreau, 1988; Clairin et Brion, 1996; Himelein, Eckman et Murray, 2014). Ces derniers temps, elle n'a cessé de gagner de l'importance pour la statistique des entreprises en raison de l'utilisation d'une unité « entreprise statistique » n'étant pas nécessairement l'équivalent de l'unité disponible dans les registres des entreprises (Lorenc, Smith et Bavdaž, 2018). Dans ce cas, la production d'estimations sur la population d'intérêt soulève des questions complexes liées au fait que les facteurs de pondération des unités d'observation doivent reposer sur les poids de sondage des unités sélectionnées dans la base de sondage.

Pour gérer ce problème, Deville et Lavallée (2006) ont proposé la méthode dite de partage des poids. Elle est fondée sur un principe de dualité entre la population échantillonnée et la population observée, où

1. Guillaume Chauvet, Ensai (Irmar), Campus de Ker Lann, Bruz, France. Courriel : guillaume.chauvet@ensai.fr; Olivier Bouriaud, Université Stefan cel Mare de Suceava, 13 rue de l'Université, 720229 Suceava, Roumanie et IGN, Laboratoire d'Inventaire Forestier, 14 rue Girardet, 54000 Nancy, France. Courriel : obouriaud.lif@gmail.com; Philippe Brion, Irmar, France. Courriel : philippe.brion55@gmail.com.

une variable d'intérêt définie sur la population observée peut être écrite sous forme de variable synthétique définie sur la base de sondage (voir également Lavallée, 2009). Compte tenu du lien créé entre les unités d'observation et les unités échantillonnées, cette méthode permet d'utiliser les propriétés du plan d'échantillonnage pour définir des estimateurs sans biais des totaux pour les populations observées, et d'en dériver des formules de variance. Les poids d'échantillonnage des unités d'échantillonnage servent plus particulièrement à attribuer des poids d'estimation aux unités d'observation. Le présent article porte sur l'extension de cette méthode au cas où la population échantillonnée est une base de sondage continue. Pour la population échantillonnée et la population observée, nous ferons appel aux notations U^A et U^B dans le cas de populations discrètes, et à \mathcal{U}^A et \mathcal{U}^B dans le cas de populations continues.

Nous nous intéressons particulièrement aux applications utilisées dans les inventaires forestiers, dans lesquels la pratique courante consiste à utiliser un échantillon de points sélectionnés dans un continuum, puis des supports de forme fixe définis à partir de ces points afin de procéder à l'observation d'une population discrète d'arbres. La méthode consistant à « transporter » une variable d'une population discrète dans une population continue n'a rien de nouveau, et a été utilisée par Stevens et Urquhart (2000), Gregoire et Valentine (2007) et Mandallaz (2007), par exemple. Malgré la similitude de ces travaux antérieurs en ce qui concerne leur méthode globale d'échantillonnage indirect, le lien entre les unités de la population échantillonnée et celles de la population cible n'y est qu'implicite.

Les travaux de Stevens et Urquhart (2000) s'apparentent beaucoup aux nôtres. Ils ont étudié le cas où une population finie d'intérêt est liée à un territoire continu, et ont songé à un moyen de transférer une variable d'intérêt dans une base de sondage continue, d'une manière semblable à la variable synthétique que nous présentons dans l'équation (3.15). Ils en ont tiré une condition dite « agrégation-absence de biais » afin d'obtenir des estimateurs totaux sans biais. Ils ont aussi proposé un estimateur de variance de Horvitz-Thompson en faisant appel à la théorie de Cordy (1993). Malgré son importance, cet article n'a pas eu un impact majeur dans la littérature. Le fait que l'article ne soit pas cité dans des ouvrages comme ceux de Gregoire et Valentine (2007) et de Mandallaz (2007) en dit long. Par conséquent, nous pensons qu'il faut procéder à une présentation simple de la méthode, dans laquelle nous explicitons les poids d'estimation, les estimateurs résultant des totaux pour la population finie et les estimateurs de variance de Horvitz-Thompson qui y sont associés. La méthode de partage des poids constitue un outil très utile et simple pour ce faire, comme l'illustrent les applications présentées à la section 3.

Dans les populations naturelles, comme les arbres d'une forêt, les unités sont réparties sur un territoire. L'estimation des totaux de toute caractéristique donnée de cette population requiert la réalisation d'un échantillonnage spatial. À cette fin, Gregoire et Valentine (2007, chapitre 10) ont présenté la méthode dite d'intégration de Monte Carlo, dans laquelle la variable synthétique est appelée « densité de la caractéristique ». Plusieurs exemples sont fournis concernant des méthodes utilisées dans le cadre d'inventaires forestiers (par exemple l'échantillonnage par points à l'aide d'un relascope et l'échantillonnage linéaire). Le lien avec la méthode proposée par Cordy n'est pas indiqué, et l'estimation de la variance se limite au cas d'une sélection des points effectuée selon l'échantillonnage uniforme indépendant. Mandallaz (2007,

section 4.2) a aussi élaboré une méthode connexe où le lien entre la population finie observée et la population continue échantillonnée est établi au moyen de ce qu'il appelle la « densité locale ». La méthode a d'abord été présentée pour l'échantillonnage par parcelles, pour ensuite s'étendre à des situations plus complexes comme l'échantillonnage par grappes (« trakt »), qui est populaire pour les inventaires forestiers (Lawrence et coll., 2010). Elle est toutefois développée de façon ad-hoc, puisqu'il faut calculer la densité locale différemment dans chaque situation. Par ailleurs, la méthode de partage des poids permet de produire des formules générales pour les estimateurs ponctuels et les estimateurs de variance. Dans la section 3.4, nous examinons de plus près le cas de l'échantillonnage par grappes spatiales pour les inventaires forestiers, puisque la méthode de partage des poids fournit une solution générale pour l'estimation et l'estimation de la variance selon un plan d'échantillonnage continu arbitraire.

Dans le texte qui suit, nous commençons par rappeler, dans la section 2, les principes de base de la méthode de partage des poids dans le cas de deux populations discrètes U^A et U^B . Dans la section 3, nous étendons la méthode au cas de l'échantillonnage d'une population continue \mathcal{U}^A , et où nous voulons inférer sur une population discrète U^B . Les résultats de deux études par simulation sont présentés à la section 4 et nos conclusions sont présentées à la section 5.

2. Échantillonnage d'une population discrète

Nous définissons d'abord, dans la section 2.1, les notations que nous employons dans l'échantillonnage d'une population discrète U^A . Puis, dans la section 2.2, nous expliquons la façon dont il est possible d'utiliser la méthode de partage des poids pour produire des estimations dans une autre population discrète U^B couplée à U^A . Nous présentons un exemple simple, à la section 2.3, pour illustrer le tout.

2.1 Notations

Nous portons notre attention sur une population discrète U^A , pour laquelle les unités de la population sont dénombrables et, par conséquent, pour laquelle une base de sondage peut être disponible. Il pourrait, par exemple, s'agir d'une population de ménages ou de particuliers dans des enquêtes sociales, ou d'un registre dans des enquêtes-entreprises. La taille de la population U^A est désignée par N^A . Supposons que nous nous intéressons à une variable d'intérêt y^A prenant la valeur y_i^A pour l'unité $i \in U^A$, et que nous voulons estimer le total de la population

$$\tau_y^A = \sum_{i \in U^A} y_i^A. \quad (2.1)$$

Un échantillon aléatoire S^A est sélectionné dans U^A au moyen d'un plan d'échantillonnage $p^A(\cdot)$. Supposons que s^A désigne une possible réalisation de S^A . L'estimateur de Horvitz-Thompson (HT) est

$$\hat{\tau}_y^A = \sum_{i \in S^A} d_i^A y_i^A, \quad (2.2)$$

où $d_i^A = 1/\pi_i^A$ est le poids de sondage de l'unité i et π_i^A , la probabilité d'inclusion de l'unité i dans l'échantillon. Cet estimateur est sans biais par rapport au plan pour τ_y^A , à condition que tous les π_i^A soient > 0 .

La variance de $\hat{\tau}_y^A$ est

$$V_p(\hat{\tau}_y^A) = \sum_{i,j \in U^A} \frac{y_i^A}{\pi_i^A} \frac{y_j^A}{\pi_j^A} (\pi_{ij}^A - \pi_i^A \pi_j^A), \quad (2.3)$$

où π_{ij}^A est la probabilité que les unités i et j soient sélectionnées conjointement dans S^A . Si tous les π_{ij}^A sont positifs, cette variance est estimée sans biais par

$$\hat{V}^A(\hat{\tau}_y^A) = \sum_{i,j \in S^A} \frac{y_i^A}{\pi_i^A} \frac{y_j^A}{\pi_j^A} \left(\frac{\pi_{ij}^A - \pi_i^A \pi_j^A}{\pi_{ij}^A} \right). \quad (2.4)$$

2.2 Méthode de partage des poids

Supposons que nous nous intéressons à une autre population U^B , comportant une variable d'intérêt y^B prenant la valeur y_k^B pour l'unité $k \in U^B$. Nous voulons estimer le total d'une population

$$\tau_y^B = \sum_{k \in U^B} y_k^B. \quad (2.5)$$

Nous supposons qu'aucune base de sondage n'est disponible pour U^B , mais que cette population est couplée à la population U^A . Le lien entre les unités présentes dans U^A et U^B est représenté par les variables indicatrices

$$L^{AB}(i, k) = \begin{cases} 1 & \text{si les unités } i \in U^A \text{ et } k \in U^B \text{ sont liées,} \\ 0 & \text{sinon.} \end{cases} \quad (2.6)$$

L'ensemble d'*ancêtres* d'une certaine unité $k \in U^B$ est $\text{Anc}_k = \{i \in U^A; L^{AB}(i, k) = 1\}$. L'ensemble de *descendants* d'une certaine unité $i \in U^A$ est $\text{Des}_i = \{k \in U^B; L^{AB}(i, k) = 1\}$. Pour toute unité $k \in U^B$,

$$N_{+k}^{AB} = \sum_{i \in U^A} L^{AB}(i, k) \quad (2.7)$$

est le nombre total d'ancêtres. Il faut que toute unité $k \in U^B$ soit couplée à au moins une unité dans U^A ; c'est-à-dire que nous supposons que $N_{+k}^{AB} > 0$ pour toute unité $k \in U^B$.

Un échantillon S^B s'obtient dans U^B en sondant tous les descendants des unités i sélectionnées dans S^A . De manière plus formelle, nous obtenons

$$S^B = \bigcup_{i \in S^A} \text{Des}_i. \quad (2.8)$$

Pour obtenir un estimateur de τ_y^B , la méthode de partage des poids (Deville et Lavallée, 2006) fait appel au principe de dualité entre les populations U^A et U^B fondée sur la fonction de lien donnée dans l'équation (2.6). Le total τ_y^B pourrait s'écrire ainsi :

$$\tau_y^B = \sum_{i \in U^A} y_i^A \quad \text{avec} \quad y_i^A = \sum_{k \in U^B} \frac{L^{AB}(i, k) y_k^B}{N_{+k}^{AB}}, \quad (2.9)$$

consulter Deville et Lavallée (2006, résultat 2). L'équation (2.9) représente le fait que la variable y_k^B pourrait être répartie sur les unités dans U^A afin d'obtenir une variable synthétique y_i^A . Pour ce faire, il faut partager chacune des valeurs y_k^B également entre les ancêtres dans Anc_k .

À partir de l'équation (2.9), il est possible d'estimer sans biais le total τ_y^B en effectuant une estimation de HT de l'échantillon S^A au moyen de la variable synthétique y_i^A . Cet estimateur de HT peut se réécrire comme suit :

$$\hat{\tau}_y^B = \sum_{i \in S^A} d_i^A y_i^A = \sum_{k \in S^B} w_k^B y_k^B, \quad (2.10)$$

où

$$w_k^B = \frac{1}{N_{+k}^{AB}} \sum_{i \in S^A} L^{AB}(i, k) d_i^A,$$

consulter Deville et Lavallée (2006, résultat 3). Chaque unité $k \in S^B$ se voit affecter la somme des poids des unités échantillonnées $i \in S^A$ qui sont couplées à k , divisée par le nombre de liens N_{+k}^{AB} . Les poids d_i^A des unités $i \in S^A$ sont, par conséquent, partagés entre les unités $k \in S^B$, d'où le nom de la méthode. Il est important de prendre note qu'il n'est possible de calculer les poids w_k^B que si le nombre d'ancêtres N_{+k}^{AB} est connu pour chaque unité $k \in S^B$. Il faut donc recueillir ce renseignement durant l'enquête. Dans certaines situations, il sera peut-être difficile, voire impossible d'affirmer qu'une unité dans U^A est ou non liée à une autre unité dans U^B . Xu et Lavallée (2009) appellent cela la non-réponse de lien et proposent des méthodes de traitement pour gérer ce problème.

À partir de l'équation (2.10), la méthode de partage des poids permet d'attribuer à chaque unité $k \in S^B$ un poids w_k^B , qu'il est possible d'utiliser pour toute variable d'intérêt y_k^B et ainsi, l'estimateur $\hat{\tau}_y^B$ est sans biais. Il s'agit d'une très importante propriété. À l'opposé, il est impossible de calculer l'estimateur de HT pour l'échantillon S^B . La probabilité d'inclusion de l'unité k dans l'échantillon S^B est

$$\Pr(k \in S^B) = \sum_{\substack{S^A \subset U^A \\ S^A \cap \text{Anc}_k \neq \emptyset}} p^A(S^A). \quad (2.11)$$

Calculer ces probabilités d'inclusion nécessiterait une spécification complète du plan d'échantillonnage $p^A(\cdot)$ et des liens entre les deux populations, ce qui est habituellement impossible.

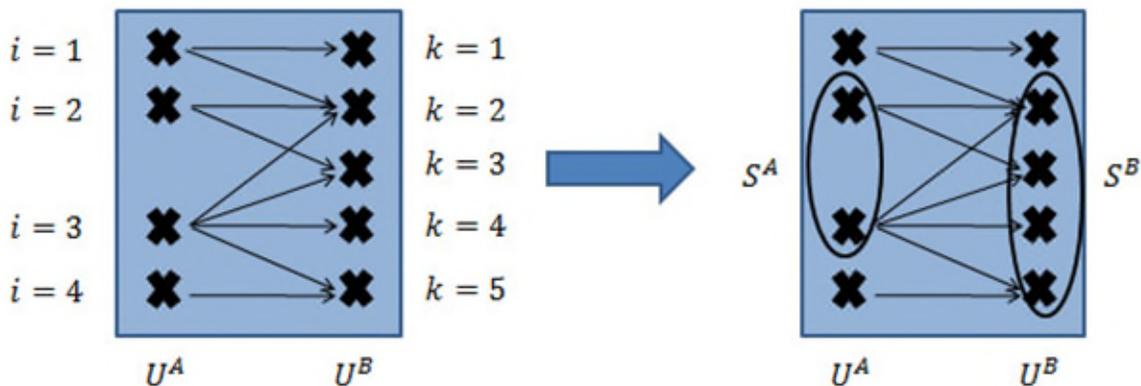
Puisque $\hat{\tau}_y^B$ peut s'écrire sous forme d'estimateur de HT dans l'échantillon S^A , la variance de $\hat{\tau}_y^B$ est obtenue par l'équation (2.3), la variable synthétique y_i^A étant obtenue par l'équation (2.9) et un estimateur de variance est obtenu par l'équation (2.4). Il convient d'indiquer que le calcul exact de la variable y_i^A peut se faire pour toute unité $i \in S^A$, puisqu'on présume que toutes les unités de U^B couplées aux unités de S^A sont sondées, comme l'indique l'équation (2.8).

2.3 Un exemple simple

Pour illustrer cela, nous présentons un exemple simple à la figure 2.1. La population U^A contient $N^A = 4$ unités, et la population U^B contient $N^B = 5$ unités. Les flèches représentent les liens entre les unités. Par exemple, l'unité $i = 3$ dans U^A a quatre descendants, à savoir les unités $k = 2, 3, 4$ et 5 . Par conséquent, nous obtenons $L^{AB}(3,1) = 0$ et $L^{AB}(3,2) = L^{AB}(3,3) = L^{AB}(3,4) = L^{AB}(3,5) = 1$. L'unité $k = 4$ a un seul ancêtre, $i = 3$, tandis que l'unité $k = 5$ en a deux, $i = 3$ et $i = 4$. Supposons que le plan d'échantillonnage mène à la sélection du sous-ensemble $s^A = \{2, 3\}$. Cela entraîne la sélection de tous les descendants des unités de s^A et l'observation du sous-ensemble

$$s^B = \{2, 3, 4, 5\}. \quad (2.12)$$

Figure 2.1 Exemple simple de liens entre deux populations discrètes U^A et U^B .



Supposons maintenant que les valeurs échantillonnées soient

$$y_2^B = 1, \quad y_3^B = 3, \quad y_4^B = 3, \quad y_5^B = 5. \quad (2.13)$$

Nous calculons d'abord la variable synthétique y_i^A pour les unités $i \in s^A$ à l'aide de l'équation (2.9). Nous obtenons

$$y_2^A = \frac{y_2^B}{3} + \frac{y_3^B}{2} \simeq 1,83,$$

$$y_3^A = \frac{y_2^B}{3} + \frac{y_3^B}{2} + \frac{y_4^B}{1} + \frac{y_5^B}{2} \simeq 7,33.$$

Nous calculons ensuite les poids w_k^B pour les unités $k \in s^B$ au moyen de la méthode de partage des poids en utilisant l'équation (2.10). Supposons que l'échantillon s^A est sélectionné dans U^A par un échantillonnage aléatoire simple sans remise, ce qui donne $\pi_i^A = 0,5$ et $d_i^A = 2$ pour tout $i \in U^A$. Nous obtenons

$$w_2^B = \frac{d_2^A + d_3^A}{3} \simeq 1,33,$$

$$w_3^B = \frac{d_2^A + d_3^A}{2} = 2,$$

$$w_4^B = \frac{d_3^A}{1} = 2,$$

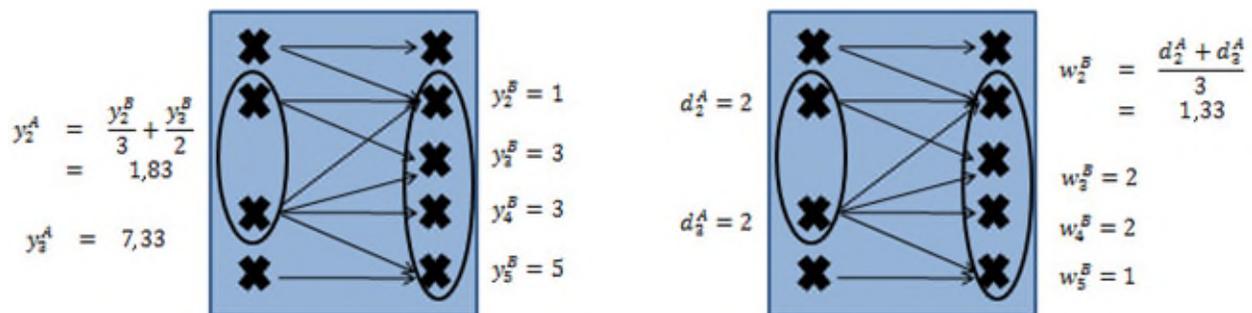
$$w_5^B = \frac{d_3^A}{2} = 1.$$

L'estimation pour le total τ_y^B est donc

$$\hat{\tau}_y^B(s^B) = \sum_{k \in s^B} w_k^B y_k^B \simeq 18,33.$$

La figure 2.2 résume les résultats.

Figure 2.2 Calcul de la variable synthétique y_i^A et des poids d_k^B obtenus par la méthode de partage des poids dans un exemple simple.



3. Échantillonnage d'une population continue

Nous définissons d'abord, dans la section 3.1, les notations que nous employons lors de l'échantillonnage d'un univers continu \mathcal{U}^A , d'après Cordy (1993). Dans la section 3.2, nous expliquons la façon dont il est possible d'étendre la méthode de partage des poids pour produire des estimations dans une certaine population discrète U^B couplée à \mathcal{U}^A . Nous examinons les applications des plans d'échantillonnage utilisés dans les inventaires forestiers, pour lesquels des méthodes analogues ont été proposées dans la littérature. Le cas d'un échantillonnage direct par parcelles est d'abord étudié dans la section 3.3. L'application de l'échantillonnage par grappes est abordée ensuite dans la section 3.4.

3.1 Notations

Nous évaluons, dans un premier temps, la méthode d'échantillonnage et d'estimation dans un univers continu \mathcal{U}^A présenté par Cordy (1993). Ce cadre, fortement motivé par les applications environnementales,

a mené à une extension de l'estimateur de HT existant pour une population discrète, que nous avons vue à la section 2.1.

Supposons que nous nous intéressions à une population continue ou à un univers continu, c'est-à-dire, selon la définition de Gregoire et Valentine (2007, page 93), « qui ne se divise pas naturellement en unités discrètes plus petites ». Il pourrait s'agir d'un paysage ou d'un lac, par exemple. Nous supposons que l'univers \mathcal{U}^A est compris dans \mathbf{R}^q et que $q \geq 1$. Nous nous intéressons à une certaine fonction Lebesgue intégrable $y^A : \mathcal{U}^A \rightarrow \mathbf{R}$ et voulons estimer le total (intégrale)

$$\tau_y^A = \int_{\mathcal{U}^A} y^A(x) dx \quad (3.1)$$

de cette fonction pour \mathcal{U}^A .

Un échantillon aléatoire $S^A = \{S_1^A, \dots, S_{n^A}^A\}$ de n^A localisations est sélectionné dans \mathcal{U}^A , et nous supposons que $s^A = \{s_1^A, \dots, s_{n^A}^A\}$ désigne une réalisation possible de S^A . Nous supposons l'existence de la fonction de densité de probabilité jointe

$$f(s_1^A, \dots, s_{n^A}^A) \quad (3.2)$$

des localisations, de même que l'existence de la fonction de densité de probabilité marginale et de la fonction de densité de probabilité jointe

$$f_i(s_i^A) \quad \text{et} \quad f_{ij}(s_i^A, s_j^A) \quad (3.3)$$

de s_i^A et de s_j^A pour $i \neq j$. Si, par exemple, l'échantillon S^A est obtenu par n^A sélections indépendantes d'un point, de manière uniforme dans \mathcal{U}^A , nous obtenons

$$f_i(s_i^A) = \frac{1(s_i^A \in \mathcal{U}^A)}{M^A} \quad (3.4)$$

et $f(s_1^A, \dots, s_{n^A}^A) = \prod_{i=1}^{n^A} f_i(s_i^A)$, où $M^A = \int_{\mathcal{U}^A} dx$ est la mesure globale de l'univers et $1(\cdot)$, une fonction indicatrice.

Supposons que la fonction de densité de probabilité soit absolument continue au regard de la mesure de Lebesgue. Comme l'ont indiqué Cordy (1993) et Stevens (1997), pour certains plans de sondage utilisés en pratique, comme l'échantillonnage systématique, cette hypothèse n'est pas valide. Pour tout point $x \in \mathcal{U}^A$, la fonction de densité d'inclusion se définit par :

$$\pi^A(x) = \sum_{i=1}^{n^A} f_i(x). \quad (3.5)$$

On pourrait voir cela comme une mesure locale du nombre de points échantillonnés selon l'unité de mesure. Nous obtenons plus précisément $\int_{\mathcal{U}^A} \pi^A(x) dx = n^A$, une propriété usuelle des plans d'échantillonnage de taille fixe n^A . De même, la fonction de densité d'inclusion jointe se définit par :

$$\pi^A(x, x') = \sum_{i=1}^{n^A} \sum_{\substack{j=1 \\ j \neq i}}^{n^A} f_{ij}(x, x') \quad (3.6)$$

pour $x, x' \in \mathcal{U}^A$.

L'estimateur de HT de τ_y^A est

$$\hat{\tau}_y^A = \sum_{s \in S^A} d^A(s) y^A(s), \quad (3.7)$$

et $d^A(x) = 1/\pi^A(x)$ est le poids de sondage du point x . Cet estimateur est sans biais pour τ_y , à condition que $\pi^A(x) > 0$ presque partout, comme l'a précisé Cordy (1993, théorème 1).

Si la fonction $y^A(\cdot)$ est bornée et que $\int_{\mathcal{U}^A} \{1/\pi^A(x)\} dx < \infty$, la variance de $\hat{\tau}_y^A$ sera alors obtenue à l'aide de la formule de Horvitz-Thompson

$$\begin{aligned} V(\hat{\tau}_y^A) &= \int_{x \in \mathcal{U}^A} \frac{\{y^A(x)\}^2}{\pi^A(x)} dx \\ &+ \int_{x \in \mathcal{U}^A} \int_{x' \in \mathcal{U}^A} \left\{ \pi^A(x, x') - \pi^A(x) \pi^A(x') \right\} \frac{y^A(x)}{\pi^A(x)} \frac{y^A(x')}{\pi^A(x')} dx dx', \end{aligned} \quad (3.8)$$

ou de la même façon à l'aide de la formule de Sen-Yates-Grundy

$$V(\hat{\tau}_y^A) = \frac{1}{2} \int_{x \in \mathcal{U}^A} \int_{x' \in \mathcal{U}^A} \left\{ \pi^A(x) \pi^A(x') - \pi^A(x, x') \right\} \left\{ \frac{y^A(x)}{\pi^A(x)} - \frac{y^A(x')}{\pi^A(x')} \right\}^2 dx dx', \quad (3.9)$$

selon Cordy (1993, section 2). Les estimateurs de variance correspondants sont, respectivement,

$$\begin{aligned} \hat{V}_{\text{HT}}(\hat{\tau}_y^A) &= \sum_{s \in S^A} \left\{ \frac{y^A(s)}{\pi^A(s)} \right\}^2 \\ &+ \sum_{\substack{s \in S^A \\ s' \in S^A \\ s' \neq s}} \left\{ \frac{\pi^A(s, s') - \pi^A(s) \pi^A(s')}{\pi^A(s, s')} \right\} \frac{y^A(s)}{\pi^A(s)} \frac{y^A(s')}{\pi^A(s')} \end{aligned} \quad (3.10)$$

et

$$\hat{V}_{\text{YG}}(\hat{\tau}_y^A) = \frac{1}{2} \sum_{\substack{s \in S^A \\ s' \in S^A \\ s' \neq s}} \left\{ \frac{\pi^A(s) \pi^A(s') - \pi^A(s, s')}{\pi^A(s, s')} \right\} \left\{ \frac{y^A(s)}{\pi^A(s)} - \frac{y^A(s')}{\pi^A(s')} \right\}^2. \quad (3.11)$$

De plus, si $\pi^A(x) > 0$ et $\pi^A(x, x') > 0$ presque partout dans \mathcal{U}^A , ces deux estimateurs de variance seront alors sans biais, comme l'a indiqué Cordy (1993, théorème 2). Il convient de noter que la condition de $\pi^A(x, x')$ peut ne pas être valide, par exemple lors de l'utilisation de plans d'échantillonnage systématique.

3.2 Méthode de partage des poids : le cas continu-discret

Supposons que nous nous intéressions encore à la population U^B et à l'estimation du total τ_y^B obtenu dans l'équation (2.5). Les unités dans U^B ne sont pas échantillonnées directement; l'échantillonnage utilise

un univers continu \mathcal{U}^A couplé à U^B . Les liens entre les unités présentes dans les populations \mathcal{U}^A et U^B sont représentés par la fonction indicatrice

$$L^{AB}(x, k) = \begin{cases} 1 & \text{si } x \in \mathcal{U}^A \text{ et } k \in U^B \text{ sont liés,} \\ 0 & \text{sinon.} \end{cases} \quad (3.12)$$

Nous conservons la terminologie utilisée à la section 2.2 et écrivons Anc_k pour le *sous-ensemble d'ancêtres* de $k \in U^B$, et $\text{Des}(x)$ pour le *sous-ensemble de descendants* de $x \in \mathcal{U}^A$. Pour tout $k \in U^B$,

$$M_{+k}^{AB} = \int_{x \in \mathcal{U}^A} L^{AB}(x, k) dx \quad (3.13)$$

est la mesure du sous-ensemble d'ancêtres de l'unité k . Comme dans le cas d'une population discrète, nous supposons que $M_{+k}^{AB} > 0$ pour tout $k \in U^B$.

Un échantillon S^B s'obtient dans U^B en sondant tous les descendants des points sélectionnés dans S^A . Nous obtenons donc

$$S^B = \bigcup_{s \in S^A} \text{Des}(s). \quad (3.14)$$

Pour obtenir un estimateur de τ_y^B , nous établissons un principe de dualité entre les populations \mathcal{U}^A et U^B . La proposition 1 résume ce processus. Le principe de dualité est semblable à celui obtenu au moyen d'une population discrète : chaque valeur y_k^B est partagée également entre les points du sous-ensemble d'ancêtres de k . Il est par conséquent possible d'interpréter la fonction synthétique $y^A(x)$ en tant que mesure locale de densité de la variable y^B selon l'unité de surface. Cette méthode a déjà été envisagée dans le domaine des inventaires forestiers par Mandallaz (2007, section 4.2) et par Gregoire et Valentine (2007, chapitre 10), par exemple, comme nous l'avons mentionné en introduction.

Proposition 1. *Le total τ_y^B peut s'écrire ainsi :*

$$\tau_y^B = \int_{x \in \mathcal{U}^A} y^A(x) dx \quad (3.15)$$

où

$$y^A(x) = \sum_{k \in U^B} \frac{L^{AB}(x, k) y_k^B}{M_{+k}^{AB}}.$$

Preuve. *Nous obtenons :*

$$\begin{aligned} \tau_y^B &= \sum_{k \in U^B} y_k^B = \sum_{k \in U^B} y_k^B \times \frac{1}{M_{+k}^{AB}} \int_{x \in \mathcal{U}^A} L^{AB}(x, k) dx \\ &= \int_{x \in \mathcal{U}^A} \sum_{k \in U^B} \frac{L^{AB}(x, k) y_k^B}{M_{+k}^{AB}} dx = \int_{x \in \mathcal{U}^A} y^A(x) dx. \end{aligned}$$

La proposition 1 permet de réécrire τ_y^B en tant qu'intégrale sur l'univers \mathcal{U}^A et d'ainsi utiliser l'estimateur de HT étendu donné dans l'équation (3.7). Cet estimateur peut à son tour s'écrire sous forme de somme pondérée sur l'échantillon S^B . La proposition 2 résume ce raisonnement.

Proposition 2. Le total τ_y^B peut être estimé sans biais par

$$\hat{\tau}_y^B = \sum_{k \in S^B} w_k^B y_k^B = \sum_{s \in S^A} d^A(s) y^A(s) \quad (3.16)$$

où

$$w_k^B = \frac{1}{M_{+k}^{AB}} \sum_{s \in S^A} L^{AB}(s, k) d^A(s).$$

Preuve. Nous pouvons réécrire

$$\begin{aligned} \hat{\tau}_y^B &= \sum_{s \in S^A} \sum_{k \in S^B} \frac{L^{AB}(s, k) y_k^B d^A(s)}{M_{+k}^{AB}} \\ &= \sum_{s \in S^A} \sum_{k \in U^B} \frac{L^{AB}(s, k) y_k^B d^A(s)}{M_{+k}^{AB}} \end{aligned}$$

où la dernière égalité vient du fait que si $s \in S^A$, toutes les unités k de son sous-ensemble de descendants $\text{Des}(x)$ sont sélectionnées dans S^B . Il s'ensuit que

$$\hat{\tau}_y^B = \sum_{s \in S^A} d^A(s) \sum_{k \in U^B} \frac{L^{AB}(s, k) y_k^B}{M_{+k}^{AB}} = \sum_{s \in S^A} d^A(s) y^A(s),$$

qui est simplement l'estimateur de HT de l'intégrale $\int_{x \in \mathcal{U}^A} y^A(x) dx$.

La méthode de partage des poids apporte ainsi une solution à l'estimation de τ_y^B , en s'appuyant sur un estimateur pondéré calculé dans l'échantillon S^B où les poids w_k^B sont obtenus à l'aide de l'équation (3.16). Chaque unité $k \in S^B$ est obtenue par la somme des poids des points échantillonnés $s \in S^A$ qui sont liés à k , divisée par M_{+k}^{AB} , la mesure du sous-ensemble d'ancêtres de l'unité k . Le principe est donc le même que celui de la méthode de partage des poids habituelle appliquée aux populations discrètes. Il est important de noter que pour chaque unité $k \in S^B$, nous devons connaître la mesure M_{+k}^{AB} de son sous-ensemble d'ancêtres.

Puisque $\hat{\tau}_y^B$ peut s'écrire sous forme d'estimateur de HT dans l'échantillon S^A , la variance est obtenue par l'équation (3.8) ou (3.9), en s'appuyant sur la variable synthétique $y^A(x)$ obtenue dans l'équation (3.15). Il est possible d'obtenir un estimateur de variance en appliquant l'équation (3.10) ou (3.11). Par conséquent, l'estimation de la variance provient directement de l'estimateur du partage des poids.

3.3 Application à l'échantillonnage par parcelles pour les inventaires forestiers

Nous examinons d'abord une application de la méthode de partage des poids à un cas d'inventaire forestier réalisé par échantillonnage direct de parcelles. Nous nous intéressons à une population U^B d'arbres situés dans un territoire \mathcal{U}^A . Un échantillon de points S^A est d'abord sélectionné dans \mathcal{U}^A à l'aide d'un plan d'échantillonnage continu. Pour chaque point $s \in S^A$, on dessine le cercle $C_r(s)$ centré sur s selon

un rayon r prédéterminé. Tous les arbres $k \in U^B$ dont le centre x_k se trouve à l'intérieur de ces cercles sont sélectionnés dans l'échantillon S^B et enquêtés.

La fonction de lien est donc

$$L^{AB}(x, k) = \begin{cases} 1 & \text{si } x_k \in C_r(x), \\ 0 & \text{sinon,} \end{cases} = \begin{cases} 1 & \text{si } x \in C_r(x_k), \\ 0 & \text{sinon.} \end{cases} \quad (3.17)$$

Pour tout arbre $k \in U^B$, la quantité M_{+k}^{AB} est

$$M_{+k}^{AB} = \int_{x \in \mathcal{U}^A} L^{AB}(x, k) dx = \int_{x \in \mathcal{U}^A} 1\{x \in C_r(x_k)\} dx, \quad (3.18)$$

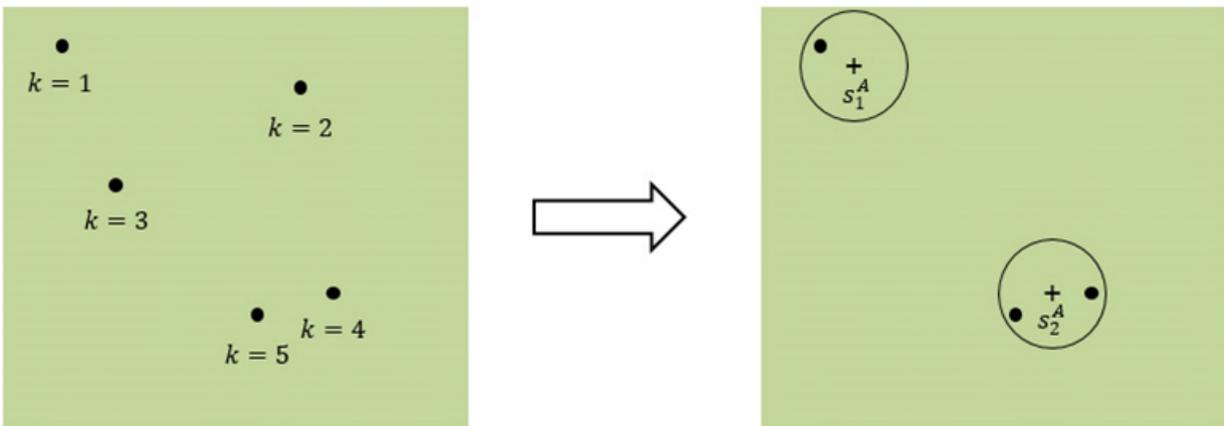
qui correspond à l'aire de l'intersection entre \mathcal{U}^A et le cercle centré sur x_k . Pour tout point $x \in \mathcal{U}^A$, la variable synthétique est

$$y^A(x) = \sum_{k \in U^B} \frac{L^{AB}(x, k) y_k^B}{M_{+k}^{AB}} = \sum_{x_k \in C_r(x)} \frac{y_k^B}{M_{+k}^{AB}}. \quad (3.19)$$

La solution obtenue dans ce cas équivaut à la fonction d'agrégation de Stevens et Urquhart (2000, section 4.2), à la densité de caractéristique de Gregoire et Valentine (2007, section 10.2) ou à la densité locale de Mandallaz (2007, équation 4.5).

Afin d'illustrer cela, nous présentons un exemple simple à la figure 3.1. Nous nous intéressons à un territoire rectangulaire \mathcal{U}^A ayant une mesure globale $M^A = 7 \times 8 = 56 \text{ m}^2$. À l'intérieur de cette aire, nous avons une population U^B de $N^B = 5$ arbres. Un échantillon de $n^A = 2$ points est sélectionné au moyen de tirages indépendants à l'aide de la fonction de densité de probabilité marginale donnée dans l'équation (3.4), qui mène à l'observation de $s^A = \{s_1^A, s_2^A\}$.

Figure 3.1 Exemple simple de liens entre une population continue \mathcal{U}^A et une population discrète U^B .



Pour chaque point $s \in s^A$, on dessine le cercle $C_r(s)$ centré sur s selon un rayon $r = 1$, et tous les arbres k dont le centre x_k se trouve à l'intérieur de ces cercles sont enquêtés. Dans l'exemple présenté à la figure 3.1, nous obtenons $s^B = \{1, 4, 5\}$. Les valeurs de la variable d'intérêt pour les unités dans s^B sont :

$$y_1^B = 1, y_4^B = 4, y_5^B = 3.$$

Nous calculons les quantités M_{+k}^{AB} pour les arbres $k \in s^B$ en appliquant l'équation (3.18). Pour les arbres $k = 4$ et 5 , le cercle $C_r(x_k)$ est compris dans \mathcal{U}^A , ce qui donne $M_{+k}^{AB} \simeq 3,14$. Pour l'arbre $k = 1$, nous obtenons $M_{+1}^{AB} \simeq 3,00$. Nous calculons la variable synthétique $y^A(x)$ pour les points dans s^A à l'aide de l'équation (3.19). Nous obtenons

$$y^A(s_1) = \frac{y_1^B}{3,00} \simeq 0,33,$$

$$y^A(s_2) = \frac{y_4^B}{3,14} + \frac{y_5^B}{3,14} \simeq 2,23.$$

Pour terminer, nous calculons les poids w_k^B pour les arbres $k \in s^B$ au moyen de la méthode de partage des poids en utilisant l'équation (3.16). Nous obtenons

$$w_1^B = \frac{d^A(s_1)}{M_{+1}^{AB}} = \frac{28}{3,00} \simeq 9,33,$$

$$w_4^B = \frac{d^A(s_2)}{M_{+4}^{AB}} = \frac{28}{3,14} \simeq 8,91,$$

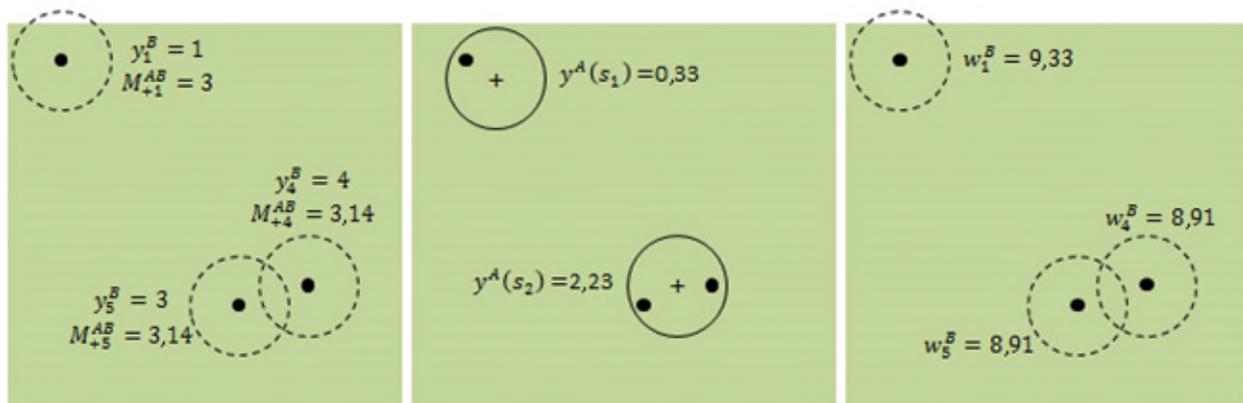
$$w_5^B = \frac{d^A(s_2)}{M_{+5}^{AB}} = \frac{28}{3,14} \simeq 8,91.$$

L'estimation du total τ_y^B est donc

$$\hat{\tau}_y^B(s^B) = \sum_{k \in s^B} w_k^B y_k^B \simeq 71,72.$$

La figure 3.2 résume les résultats.

Figure 3.2 Calcul de la variable synthétique $y^A(x)$ et des poids w_k^B obtenus par la méthode de partage des poids dans un exemple simple.

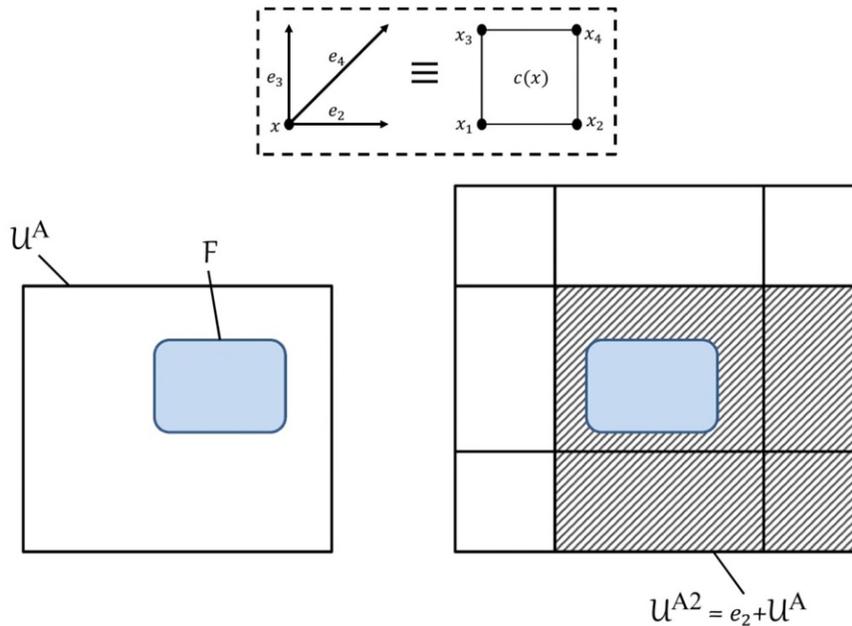


3.4 Application de l'échantillonnage par grappes spatiales pour les inventaires forestiers

Lorsque l'accès au terrain est difficile, la pratique courante pour les inventaires forestiers consiste à utiliser des grappes de parcelles (Köhl et Magnussen, 2015). Ces grappes ont une forme géométrique fixe, déterminée avant l'enquête. Les parcelles peuvent, par exemple, se situer dans les coins d'un carré de 50 mètres (consulter Mandallaz, 2007, section 4.3).

Supposons que nous nous intéressions encore à une population U^B d'arbres situés dans une forêt \mathcal{F} . Supposons que \mathcal{U}^A désigne un ensemble tel que $\mathcal{F} \subset \mathcal{U}^A$, et que e_1, \dots, e_L désigne un ensemble de L vecteurs dans \mathbf{R}^2 . Pour tout point $x \in \mathcal{U}^A$, la grappe $c(x)$ correspond à un ensemble de points $\{x_l \equiv x + e_l; l = 1, \dots, L\}$. Selon la notation de Mandallaz (2007), nous considérons e_1 comme le vecteur nul et $x_1 = x$, comme l'origine de la grappe. Notons $\mathcal{U}^{Al} = e_l + \mathcal{U}^A$ pour $l = 1, \dots, L$, et $\mathcal{V}^C = \bigcup_{l=1}^L \mathcal{U}^{Al}$ pour leur réunion. On suppose aussi que l'ensemble \mathcal{U}^A est suffisamment grand pour faire en sorte que $\mathcal{F} \subset \mathcal{U}^{Al}$ pour tout $l = 1, \dots, L$ (consulter Mandallaz, 2007, section 4.3). À des fins d'illustration, la figure 3.3 présente un exemple du cas où $L = 4$.

Figure 3.3 Exemple d'échantillonnage par grappes dans des populations continues.



Note de présentation. Dans le panneau supérieur se trouve un exemple de grappe de taille $L = 4$ dont l'origine est le point x . Dans le panneau de gauche, deux populations continues correspondant à $\mathcal{F} \subset \mathcal{U}^A$ sont données. Dans le panneau de droite, les ensembles $\mathcal{U}^{Al}, l = 1, \dots, 4$ associés à \mathcal{U}^A sont présentés, où l'ensemble hachuré correspond à \mathcal{U}_{l2} .

L'échantillonnage par grappes s'effectue en sélectionnant d'abord un échantillon S^A de n^A points dans \mathcal{U}^A , selon un plan d'échantillonnage continu. Pour chaque point $x \in S^A$, nous obtenons la grappe associée $c(x) = \{x_1, \dots, x_L\}$ dont l'origine est $x_1 = x$, de même que les cercles associés $C_r(x_l)$ dessinés selon un

rayon prédéterminé r . Tous les arbres $k \in U^B$ dont le centre x_k se trouve à l'intérieur d'un de ces cercles $C_r(x_l), l = 1, \dots, L$ pour $x \in S^A$ sont sondés. La figure 3.4 en montre un exemple.

Pour tout point $x \in U^A$, la variable synthétique $y^A(x)$ est obtenue en deux étapes, \mathcal{V}^C servant de population pivot. Nous définissons d'abord une fonction de lien entre \mathcal{V}^C et U^B sous la forme

$$L^{CB}(z, k) = \begin{cases} 1 & \text{si } x_k \in C_r(z), \\ 0 & \text{sinon,} \end{cases} \tag{3.20}$$

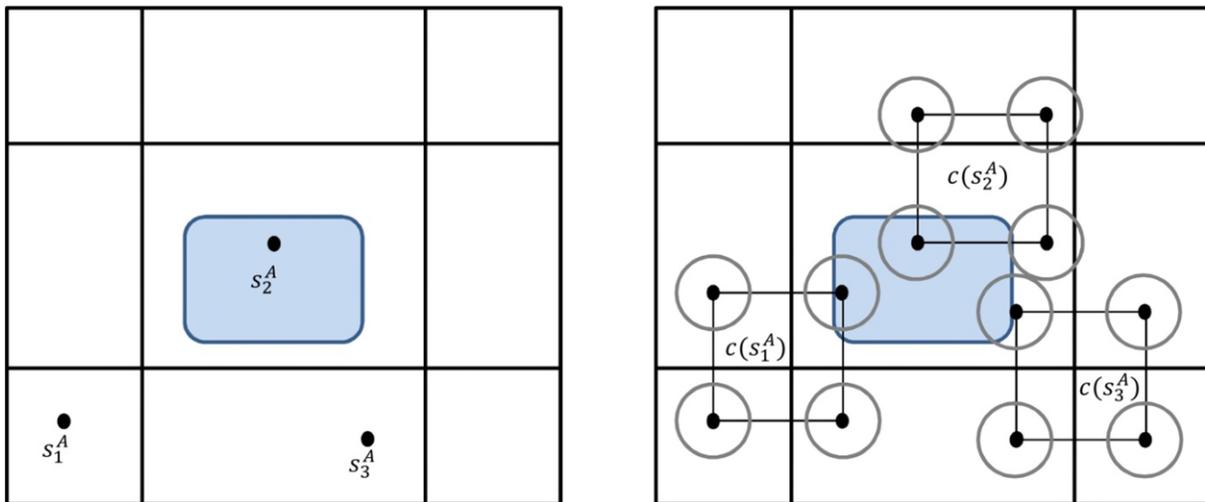
où x_k correspond au centre de l'arbre k et z , à un point dans l'ensemble unifié \mathcal{V}^C , ce qui est semblable à la fonction de lien définie à l'équation (3.17). En suivant le même raisonnement, nous obtenons la variable synthétique intermédiaire

$$y^C(z) = \sum_{x_k \in C_r(z)} \frac{y_k^B}{M_{+k}^{CB}} \text{ pour tout } z \in \mathcal{V}^C, \tag{3.21}$$

où $M_{+k}^{CB} = \int_{z \in \mathcal{V}^C} 1\{z \in C_r(x_k)\} dz$ correspond à l'aire de l'intersection entre \mathcal{V}^C et le cercle centré sur x_k . Nous définissons ensuite une fonction de lien entre U^A et \mathcal{V}^C sous la forme

$$L^{AC}(x, z) = \begin{cases} 1 & \text{si } z \in c(x) = \{x_1, \dots, x_L\}, \\ 0 & \text{sinon.} \end{cases} \tag{3.22}$$

Figure 3.4 Exemple d'échantillonnage par grappes où $n^A = 3$ points sont d'abord sélectionnés dans U^A .



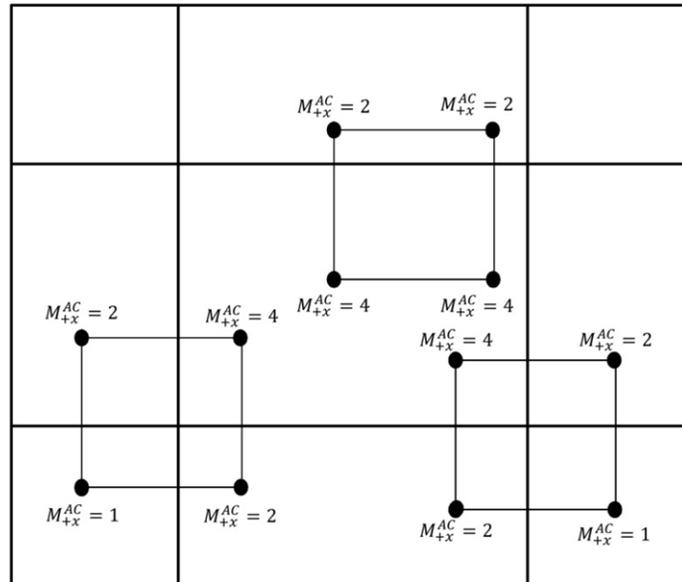
Autrement dit, x et z sont couplés si z est un des sommets de la grappe qui prend naissance dans x . La quantité

$$M_{+z}^{AC} = \int_{x \in U^A} L^{AC}(x, z) dx \tag{3.23}$$

est le nombre de grappes ayant z pour sommet. À la figure 3.5, nous donnons les valeurs obtenues pour les sommets des grappes dans l'exemple initié à la figure 3.4. Nous obtenons finalement la variable synthétique

$$y^A(x) = \sum_{z \in c(x)} \frac{y^C(z)}{M_{+z}^{AC}} = \sum_{z \in c(x)} \frac{1}{M_{+z}^{AC}} \sum_{x_k \in C_r(z)} \frac{y_k^B}{M_{+k}^{CB}} \quad \text{pour tout } x \in \mathcal{U}^A. \quad (3.24)$$

Figure 3.5 Valeurs des quantités M_{+x}^{AC} pour les sommets d'un échantillon par grappes.



Comme le prouve la proposition 2, il est possible d'estimer τ_y^B sans biais au moyen de $\hat{\tau}_y^B$, et un estimateur de la variance sans biais s'obtient directement en appliquant l'équation (3.10) ou (3.11). Par conséquent, la méthode de partage des poids fournit une solution générale d'estimation et d'estimation de la variance pour l'échantillonnage par grappes, selon un plan d'échantillonnage arbitraire réalisé dans \mathcal{U}^A . En revanche, la solution décrite dans Mandallaz (2007, équation 4.17) est valable uniquement dans le cas où on a obtenu S^A à partir de sélections uniformes indépendantes tirées de \mathcal{U}^A .

4. Étude par simulation

Dans la présente étude par simulation, nous examinons la question de l'estimation et de l'estimation de la variance pour un échantillonnage par grappes spatiales. Nous voulons comparer la méthode de partage des poids avec la solution décrite dans Mandallaz (2007, équation 4.17), dans le cas où l'échantillon S^A est obtenu à partir de sélections uniformes indépendantes dans \mathcal{U}^A . C'est l'objet de la première étude par simulation décrite à la section 4.1. Nous voulons aussi évaluer la méthode de partage des poids dans le cas où l'échantillon S^A des origines de la grappe n'est pas sélectionné à l'aide d'un échantillonnage uniforme

indépendant et où la technique de Mandallaz ne serait par conséquent pas applicable. C'est l'objet de la deuxième étude par simulation décrite à la section 4.2.

La population continue \mathcal{U}^A que nous examinons est un carré de 1 000 mètres. À l'intérieur de \mathcal{U}^A , la forêt est située dans un territoire carré \mathcal{F} de 600 mètres, mais l'emplacement de la forêt et sa superficie sont considérés comme inconnus avant l'enquête. Une population de $N^B = 30\,942$ *Pinus nigra var. austriaca* (Hoess) est générée dans la forêt \mathcal{F} . Les principales caractéristiques de la population de pins, en ce qui a trait au volume (y_{1k}^B , mètres cubes) et au diamètre à hauteur de poitrine (y_{2k}^B , centimètres), sont résumées dans le tableau 4.1.

Tableau 4.1

Moyenne, écart-type, minimum et maximum du diamètre à hauteur de poitrine et du volume de la population de pins tordus latifoliés

	Moyenne	Écart-type	Minimum	Maximum
Diamètre à hauteur de poitrine (centimètres)	17,28	6,38	3,57	41,78
Volume (mètres cubes)	0,17	0,16	0,00	0,99

Nous voulons estimer les paramètres suivants : le nombre total d'arbres $\tau_{y_0}^B = \sum_{k \in \mathcal{U}^B} y_{0k}^B$ selon $y_{0k}^B = 1$, le volume total de bois $\tau_{y_1}^B = \sum_{k \in \mathcal{U}^B} y_{1k}^B$, le volume moyen des arbres $\mu_{y_1}^B = \tau_{y_1}^B / N^B$, le diamètre moyen à hauteur de poitrine $\mu_{y_2}^B = \tau_{y_2}^B / N^B$, le volume moyen de bois par mètre carré de forêt $\bar{Y}_1^B = \tau_{y_1}^B / M^{\mathcal{F}}$, où $M^{\mathcal{F}}$ correspond à la superficie de la forêt \mathcal{F} .

4.1 Comparaison entre la méthode de partage des poids et l'estimateur de Mandallaz pour l'échantillonnage par grappes

D'abord, nous comparons l'efficacité des estimateurs proposés à ceux proposés par Mandallaz (2007, section 4.3), dans le cas d'un échantillonnage par grappes (consulter la section 3.4) où l'échantillon S^A des origines de la grappe est sélectionné à partir d'un échantillonnage uniforme indépendant du territoire \mathcal{U}^A . Nous sélectionnons un échantillon S^A de taille $n^A = 100, 200$ ou 400. Un échantillon d'arbres est ensuite sélectionné et enquêté grâce à la technique d'échantillonnage par grappes décrite à la section 3.4, dans laquelle nous utilisons des grappes carrées de taille $L = 4$ et de longueur 60 mètres, et des parcelles avec un rayon r de 25 mètres.

Pour un échantillon donné, les estimateurs sont obtenus comme suit par la méthode de partage des poids. Les estimateurs des totaux $\tau_{y_0}^B$ et $\tau_{y_1}^B$ sont obtenus à l'aide de l'équation (3.16) sous la forme

$$\hat{\tau}_{y_0}^B = \frac{M^A}{n^A} \sum_{x \in S^A} y_0^A(x), \quad (4.1)$$

$$\hat{\tau}_{y_1}^B = \frac{M^A}{n^A} \sum_{x \in S^A} y_1^A(x),$$

où y_0^A et y_1^A s'obtiennent en insérant dans l'équation (3.24) les variables y_{0k}^B et y_{1k}^B , respectivement, et où M^A est la superficie de \mathcal{U}^A . Les estimateurs des moyennes de population $\mu_{y_1}^B$ et $\mu_{y_2}^B$ sont

$$\hat{\mu}_{y_1}^B = \frac{\hat{\tau}_{y_1}^B}{\hat{\tau}_{y_0}^B} \quad \text{et} \quad \hat{\mu}_{y_2}^B = \frac{\hat{\tau}_{y_2}^B}{\hat{\tau}_{y_0}^B},$$

respectivement, et sont obtenus à l'aide du principe de substitution. $\hat{\tau}_{y_2}^B$ est obtenu selon l'équation (4.1). L'estimateur de \bar{Y}_1^B est

$$\hat{\bar{Y}}_1^B = \frac{\hat{\tau}_{y_1}^B}{\hat{M}^{\mathcal{F}}},$$

où

$$\hat{M}^{\mathcal{F}} = \frac{M^A}{n^A} \sum_{x \in S^A} y_3^A(x) \quad \text{avec} \quad y_3^A(x) = \sum_{z \in c(x)} \frac{1\{z \in \mathcal{F}\}}{M_{+z}^{AC}} \quad (4.2)$$

est un estimateur sans biais de $M^{\mathcal{F}}$. Les estimateurs proposés par Mandallaz s'obtiennent comme suit. L'estimateur de \bar{Y}_1^B est

$$\hat{\bar{Y}}_{1,\text{mand}}^B = \frac{\sum_{x \in S^A} \sum_{z \in c(x)} 1\{z \in \mathcal{F}\} \sum_{k \in C_r(z)} \frac{y_{1k}^B}{M_{+k}^{\mathcal{F}}}}{\sum_{x \in S^A} \sum_{z \in c(x)} 1\{z \in \mathcal{F}\}}$$

où $M_{+k}^{\mathcal{F}}$ correspond à l'aire de l'intersection entre \mathcal{F} et le cercle centré sur x_k , comme l'indique l'équation (4.17) dans Mandallaz (2007). Les estimateurs des totaux $\tau_{y_0}^B$ et $\tau_{y_1}^B$ sont

$$\hat{\tau}_{y_0,\text{mand}}^B = \frac{1}{L} \frac{M^A}{n^A} \sum_{x \in S^A} \sum_{z \in c(x)} 1\{z \in \mathcal{F}\} \sum_{k \in C_r(z)} \frac{y_{0k}^B}{M_{+k}^{\mathcal{F}}},$$

$$\hat{\tau}_{y_1,\text{mand}}^B = \frac{1}{L} \frac{M^A}{n^A} \sum_{x \in S^A} \sum_{z \in c(x)} 1\{z \in \mathcal{F}\} \sum_{k \in C_r(z)} \frac{y_{1k}^B}{M_{+k}^{\mathcal{F}}},$$

selon l'équation (4.16) de Mandallaz (2007). Les estimateurs des moyennes de population $\mu_{y_1}^B$ et $\mu_{y_2}^B$ sont

$$\hat{\mu}_{y_1,\text{mand}}^B = \frac{\hat{\tau}_{y_1,\text{mand}}^B}{\hat{\tau}_{y_0,\text{mand}}^B} \quad \text{et} \quad \hat{\mu}_{y_2,\text{mand}}^B = \frac{\hat{\tau}_{y_2,\text{mand}}^B}{\hat{\tau}_{y_0,\text{mand}}^B},$$

respectivement, et sont obtenus à l'aide du principe de substitution.

Les étapes d'échantillonnage et d'estimation sont répétées $D = 10\,000$ fois. Pour un estimateur $\hat{\theta}$ d'un paramètre θ , nous calculons le biais relatif de Monte Carlo en pourcentage

$$\text{BR}\{\hat{\theta}\} = 100 \times \frac{D^{-1} \sum_{d=1}^D \hat{\theta}_d - \theta}{\theta}, \quad (4.3)$$

et l'erreur quadratique moyenne de Monte Carlo

$$\text{EQM}\{\hat{\theta}\} = \frac{1}{D} \sum_{d=1}^D (\hat{\theta}_d - \theta)^2. \quad (4.4)$$

Les résultats sont présentés dans le tableau 4.2. Pour l'un ou l'autre des cinq paramètres, les deux méthodes d'estimation mènent à des estimateurs pratiquement sans biais, et les erreurs quadratiques moyennes sont très semblables. L'estimateur de Mandallaz fonctionne un peu mieux pour \bar{Y}_1^B , tandis que la méthode de partage des poids fonctionne un peu mieux pour les autres paramètres.

Tableau 4.2

Biais relatif en pourcentage et erreur quadratique moyenne pour cinq paramètres estimés au moyen de la méthode de partage des poids et au moyen des estimateurs proposés par Mandallaz

			$\tau_{y_0}^B$	$\tau_{y_1}^B$	$\mu_{y_1}^B$	$\mu_{y_2}^B$	\bar{Y}_1^B
$n^A = 100$	Partage des poids	BR (%)	-0,01	-0,01	0,00	0,00	-0,15
		EQM	$2,04 \cdot 10^7$	$6,24 \cdot 10^5$	$1,22 \cdot 10^{-6}$	$1,81 \cdot 10^{-3}$	$8,76 \cdot 10^{-7}$
	Mandallaz	BR (%)	-0,02	-0,02	0,00	0,00	-0,17
		EQM	$2,06 \cdot 10^7$	$6,29 \cdot 10^5$	$1,25 \cdot 10^{-6}$	$1,85 \cdot 10^{-3}$	$8,65 \cdot 10^{-7}$
$n^A = 200$	Partage des poids	BR (%)	-0,01	-0,01	0,00	0,00	-0,15
		EQM	$9,91 \cdot 10^6$	$3,03 \cdot 10^5$	$6,07 \cdot 10^{-7}$	$9,09 \cdot 10^{-4}$	$4,15 \cdot 10^{-7}$
	Mandallaz	BR (%)	-0,01	-0,01	0,00	0,00	-0,15
		EQM	$9,99 \cdot 10^6$	$3,05 \cdot 10^5$	$6,20 \cdot 10^{-7}$	$9,27 \cdot 10^{-4}$	$4,08 \cdot 10^{-7}$
$n^A = 400$	Partage des poids	BR (%)	-0,02	-0,02	0,00	0,00	-0,03
		EQM	$5,03 \cdot 10^6$	$1,54 \cdot 10^5$	$2,99 \cdot 10^{-7}$	$4,44 \cdot 10^{-4}$	$2,15 \cdot 10^{-7}$
	Mandallaz	BR (%)	-0,02	-0,03	0,00	0,00	-0,04
		EQM	$5,07 \cdot 10^6$	$1,55 \cdot 10^5$	$3,05 \cdot 10^{-7}$	$4,52 \cdot 10^{-4}$	$2,12 \cdot 10^{-7}$

Note : BR signifie Biais relatif; EQM signifie Erreur quadratique moyenne.

Pour les estimateurs obtenus à l'aide de la méthode de partage des poids, nous avons également examiné l'estimation de la variance. Nous n'avons pas procédé à une estimation de la variance pour les estimateurs de Mandallaz, puisque dans son ouvrage (2007, section 4.3), des estimations de la variance sont proposées seulement pour des moyennes spatiales comme \bar{Y}_1^B , et non pour des totaux ou des moyennes de population. À partir d'un échantillonnage uniforme indépendant de \mathcal{U}^A , l'application de l'équation (3.10) permet d'obtenir l'estimateur de variance sans biais

$$\hat{V}_{\text{HT}}(\hat{\tau}_{y_0}^B) = \frac{(M^A)^2}{n^A} \times \frac{1}{n^A - 1} \sum_{x \in S^A} \left\{ y_0^A(x) - \frac{1}{n^A} \sum_{s \in S^A} y_0^A(s) \right\}^2, \quad (4.5)$$

pour $\hat{\tau}_{y_0}^B$, et une expression semblable pour $\hat{\tau}_{y_1}^B$. Un estimateur de la variance approximativement sans biais $\hat{\mu}_{y_1}^B$ est obtenu en remplaçant la variable $y_0^A(x)$ dans l'équation (4.5) par la variable linéarisée

$$u_1^A(x) = \frac{1}{\hat{\tau}_{y_0}^B} \{ y_1^A(x) - \hat{\mu}_{y_1}^B y_0^A(x) \}, \quad (4.6)$$

qu'il est aussi possible d'obtenir pour $\hat{\mu}_{y_2}^B$. Un estimateur de variance approximativement sans biais pour \hat{Y}_1^B est obtenu en remplaçant la variable $y_0^A(x)$ dans l'équation (4.5) par la variable linéarisée

$$u_2^A(x) = \frac{1}{\hat{\tau}_{y_3}^B} \{y_1^A(x) - \hat{Y}_1^B y_3^A(x)\}. \quad (4.7)$$

Les étapes d'échantillonnage et d'estimation sont répétées $E = 1000$ fois. Pour mesurer le biais d'un estimateur de variance $\hat{V}_{HT}(\hat{\theta})$, nous utilisons le biais relatif de Monte Carlo en pourcentage

$$\text{BR}\{\hat{V}_{HT}(\hat{\theta})\} = 100 \times \frac{E^{-1} \sum_{e=1}^E \hat{V}_{HT}(\hat{\theta}_e) - \text{EQM}(\hat{\theta})}{\text{EQM}(\hat{\theta})}, \quad (4.8)$$

où $\text{EQM}(\hat{\theta})$ est obtenu indépendamment du premier cycle de $D = 10\,000$ simulations, voir l'équation (4.4). Les résultats sont présentés dans le tableau 4.3. Tous les estimateurs de variance sont approximativement sans biais.

Tableau 4.3

Biais relatif en pourcentage de l'estimateur de variance pour cinq paramètres estimés au moyen de la méthode de partage des poids

n^A	$\hat{V}(\hat{\tau}_{y_0}^B)$	$\hat{V}(\hat{\tau}_{y_1}^B)$	$\hat{V}(\hat{\mu}_{y_1}^B)$	$\hat{V}(\hat{\mu}_{y_2}^B)$	$\hat{V}(\hat{Y}_1^B)$
100	-1,21	-1,17	-0,48	-0,04	-0,17
200	1,75	1,83	-0,59	-0,76	3,35
400	0,49	0,62	0,40	1,09	-0,29

4.2 Évaluation de la méthode de partage des poids pour un échantillonnage par grappes avec un échantillonnage non uniforme des grappes

Nous envisageons un second cas d'échantillonnage par grappes où la sélection de l'échantillon S^A ne découle pas d'un échantillonnage uniforme indépendant de \mathcal{U}^A , de sorte que l'on ne peut pas utiliser les estimateurs proposés par Mandallaz (2007). La population \mathcal{U}^A est d'abord divisée en une sous-population \mathcal{U}_1^A d'une longueur de 300 mètres et d'une hauteur de 1 000 mètres (partie occidentale) et en une sous-population \mathcal{U}_2^A d'une longueur de 700 mètres et d'une hauteur de 1 000 mètres (partie orientale). La superficie de \mathcal{U}_1^A et de \mathcal{U}_2^A est désignée par M_1^A et M_2^A , respectivement. Un échantillon de $n_1 = 75, 150$ ou 300 points est d'abord sélectionné de façon générale dans \mathcal{U}^A , et un deuxième échantillon de taille $n_2 = 25, 50$ ou 100 est ensuite sélectionné dans \mathcal{U}_2^A . Ce cas pourrait, par exemple, survenir lorsqu'il est particulièrement intéressant de procéder à une estimation de la sous-population \mathcal{U}_2^A et qu'il est possible de financer une extension de l'échantillon pour que l'échantillon global sélectionné dans \mathcal{U}_2^A soit assez important. L'union de ces deux échantillons est désignée par S^A . Supposons que n_1^A désigne la taille (aléatoire) de $S_1^A = S^A \cap \mathcal{U}_1^A$, et que n_2^A désigne la taille (aléatoire) de $S_2^A = S^A \cap \mathcal{U}_2^A$.

Conditionnellement aux valeurs de n_1^A et n_2^A , les estimateurs sans biais des totaux $\tau_{y_0}^B$ et $\tau_{y_1}^B$ sont

$$\hat{\tau}_{y_0, \text{cond}}^B = \frac{M_1^A}{n_1^A} \sum_{x \in S_1^A} y_0^A(x) + \frac{M_2^A}{n_2^A} \sum_{x \in S_2^A} y_0^A(x),$$

$$\hat{\tau}_{y_1, \text{cond}}^B = \frac{M_1^A}{n_1^A} \sum_{x \in S_1^A} y_1^A(x) + \frac{M_2^A}{n_2^A} \sum_{x \in S_2^A} y_1^A(x),$$

où y_0^A et y_1^A s'obtiennent en insérant dans l'équation (3.24) les variables y_{0k}^B et y_{1k}^B , respectivement. Les estimateurs des moyennes de population $\mu_{y_1}^B$ et $\mu_{y_2}^B$ sont

$$\hat{\mu}_{y_1, \text{cond}}^B = \frac{\hat{\tau}_{y_1, \text{cond}}^B}{\hat{\tau}_{y_0, \text{cond}}^B} \quad \text{et} \quad \hat{\mu}_{y_2, \text{cond}}^B = \frac{\hat{\tau}_{y_2, \text{cond}}^B}{\hat{\tau}_{y_0, \text{cond}}^B},$$

respectivement, et sont obtenus à l'aide du principe de substitution. L'estimateur de \bar{Y}_1^B est

$$\hat{Y}_{1, \text{cond}}^B = \frac{\hat{\tau}_{y_1, \text{cond}}^B}{\hat{M}_{\text{cond}}^{\mathcal{F}}},$$

où

$$\hat{M}_{\text{cond}}^{\mathcal{F}} = \frac{M_1^A}{n_1^A} \sum_{x \in S_1^A} y_3^A(x) + \frac{M_2^A}{n_2^A} \sum_{x \in S_2^A} y_3^A(x)$$

est un estimateur sans biais de $M^{\mathcal{F}}$, où $y_3^A(x)$ est défini dans l'équation (4.2).

Les étapes d'échantillonnage et d'estimation sont répétées $D = 10\,000$ fois. Pour un estimateur $\hat{\theta}$ d'un paramètre θ , nous calculons le biais relatif de Monte Carlo en pourcentage obtenu à l'équation (4.3) et l'erreur quadratique moyenne de Monte Carlo obtenue à l'équation (4.4). Les résultats sont présentés dans le tableau 4.4. Pour l'un ou l'autre des cinq paramètres, les estimateurs sont quasi sans biais. L'erreur quadratique moyenne diminue à mesure que la taille de l'échantillon augmente, comme on peut s'y attendre.

Tableau 4.4
Biais relatif en pourcentage et erreur quadratique moyenne pour cinq paramètres estimés au moyen de la méthode de partage des poids

		$\tau_{y_0}^B$	$\tau_{y_1}^B$	$\mu_{y_1}^B$	$\mu_{y_2}^B$	\bar{Y}_1^B
$n^A = 100$	BR (%)	0,03	0,03	0,00	0,00	-0,02
	EQM	$1,85 \cdot 10^7$	$5,67 \cdot 10^5$	$1,19 \cdot 10^{-6}$	$1,78 \cdot 10^{-3}$	$8,5 \cdot 10^{-7}$
$n^A = 200$	BR (%)	-0,02	-0,02	0,01	0,00	-0,06
	EQM	$9,40 \cdot 10^6$	$2,87 \cdot 10^5$	$5,83 \cdot 10^{-7}$	$8,70 \cdot 10^{-4}$	$4,34 \cdot 10^{-7}$
$n^A = 400$	BR (%)	0,09	0,09	0,00	0,00	0,05
	EQM	$4,67 \cdot 10^6$	$1,43 \cdot 10^5$	$2,90 \cdot 10^{-7}$	$4,35 \cdot 10^{-4}$	$2,18 \cdot 10^{-7}$

Note : BR signifie Biais relatif; EQM signifie Erreur quadratique moyenne.

Nous examinons maintenant l'estimation de la variance. Conditionnellement à n_1^A et n_2^A , l'estimateur de variance sans biais pour $\hat{\tau}_{y_0, \text{cond}}^B$ est

$$\hat{V}_2(\hat{\tau}_{y_0}^B) = \frac{(M_1^A)^2}{n_1^A} \times \frac{1}{n_1^A - 1} \sum_{x \in S_1^A} \left\{ y_0^A(x) - \frac{1}{n_1^A} \sum_{s \in S_1^A} y_0^A(s) \right\}^2 + \frac{(M_2^A)^2}{n_2^A} \times \frac{1}{n_2^A - 1} \sum_{x \in S_2^A} \left\{ y_0^A(x) - \frac{1}{n_2^A} \sum_{s \in S_2^A} y_0^A(s) \right\}^2, \quad (4.9)$$

qu'il est possible d'exprimer de façon similaire pour $\hat{\tau}_{y_1, \text{cond}}^B$. Un estimateur de variance approximativement sans biais pour $\hat{\mu}_{y_1, \text{cond}}^B$ est obtenu en remplaçant la variable $y_0^A(x)$ dans l'équation (4.9) par la variable linéarisée obtenue dans l'équation (4.6), qu'il est possible d'obtenir de la même façon pour $\hat{\mu}_{y_2}^B$. Un estimateur de variance approximativement sans biais pour $\hat{Y}_{1, \text{cond}}^B$ est obtenu en remplaçant la variable $y_0^A(x)$ dans l'équation (4.9) par la variable linéarisée obtenue dans l'équation (4.7).

Les étapes d'échantillonnage et d'estimation sont répétées $E = 1000$ fois. Pour mesurer le biais d'un estimateur de variance $\hat{V}_2(\hat{\theta})$, nous utilisons le biais relatif de Monte Carlo en pourcentage défini à l'équation (4.3). Les résultats sont présentés dans le tableau 4.5. Tous les estimateurs de variance sont approximativement sans biais.

Tableau 4.5

Biais relatif en pourcentage de l'estimateur de variance pour cinq paramètres estimés au moyen de la méthode de partage des poids

n^A	$\hat{V}_2(\hat{\tau}_{y_0}^B)$	$\hat{V}_2(\hat{\tau}_{y_1}^B)$	$\hat{V}_2(\hat{\mu}_{y_1}^B)$	$\hat{V}_2(\hat{\mu}_{y_2}^B)$	$\hat{V}_2(\hat{Y}_1^B)$
100	3,18	3,16	-3,02	-2,96	-0,28
200	0,97	0,92	-1,59	-1,01	-1,88
400	1,12	1,03	0,05	-0,36	-1,46

5. Discussion

En pratique, plusieurs raisons peuvent expliquer l'absence, pour une population, d'une base de sondage manipulable. Lorsqu'il est possible de coupler une population d'intérêt à une population discrète pour laquelle il existe une base de sondage, la méthode de partage des poids habituelle (Deville et Lavallée, 2006) permet d'obtenir des échantillons de probabilité pour la population d'intérêt, de même que des estimateurs et des estimateurs de variance sans biais pour cette population. Nous avons montré que cette méthode pourrait être généralisée de manière naturelle lorsque la population d'intérêt est couplée à une population continue, au moyen d'une fonction synthétique pour cette population continue qui pourrait être interprétée comme une mesure locale de densité. Dans le cas d'un échantillonnage par grappes spatiales comportant une sélection uniforme indépendante des origines de grappe, les résultats de notre simulation montrent que la méthode de partage des poids et l'estimateur de Mandallaz procurent des résultats similaires. L'estimateur de Mandallaz n'est pas applicable quand la sélection des origines de grappe ne se fait pas au moyen d'un échantillonnage uniforme indépendant. Dans ce cas, les résultats de notre simulation confirment que la méthode de partage des poids mène à des estimateurs sans biais.

De notre point de vue, l'utilisation de la méthode de partage des poids présente plusieurs avantages. Tout d'abord, elle permet de faciliter la gestion des effets de bord, c'est-à-dire le fait que certaines unités ont un sous-ensemble d'ancêtres qui intersecte la superficie d'intérêt. L'utilisation de la superficie d'un sous-ensemble d'ancêtres M_{+k}^{AB} dans les poids d'estimation (consulter l'équation [3.16]) procure des estimateurs sans aucun biais, tandis que les méthodes alternatives de corrections de limites peuvent se révéler quelque peu fastidieuses, comme nous l'observons dans Gregoire et Valentine (2007, section 10.7) ou Roesch, Green et Scott (1993), par exemple. De plus, notre méthode nous permet de retourner à la population qui a effectivement été échantillonnée. Cela s'avère nécessaire pour utiliser l'estimateur de Horvitz-Thompson et calculer un estimateur de variance sans biais. Il est possible de le faire en toute généralité, c'est-à-dire selon un plan d'échantillonnage continu arbitraire, en faisant appel à la théorie élaborée par Cordy (1993).

Cette méthode ne se limite évidemment pas aux inventaires forestiers. On en voit un exemple dans l'enquête sur les pratiques culturelles réalisée en France par le service statistique ministériel de l'agriculture et de l'alimentation jusqu'en 2006. L'échantillon de cette enquête se composait de parcelles de terrain sélectionnées à partir de points de l'enquête sur l'utilisation du territoire (Teruti) (Chapelle-Barry, 2008). L'enquête Teruti est axée sur la production de statistiques relatives à l'utilisation du sol, et les points où l'utilisation du sol était associée à des pratiques culturelles constituait la base sur laquelle reposait la sélection des parcelles de terrain. Le calcul des poids pour les parcelles de terrain s'est fait en considérant que chaque parcelle avait une probabilité d'être tirée proportionnelle à sa superficie. De cette manière, la méthode a permis de produire les mêmes poids que la méthode de partage des poids, exception faite des effets de bord. L'utilisation de la méthode de partage des poids mènerait toutefois à l'obtention d'un cadre méthodologique général permettant la dérivation d'un estimateur de variance. D'autres sujets, liés à des questions environnementales, pourraient également tirer avantage de l'application de cette méthode. Il serait également judicieux d'envisager l'extension de la méthode de partage des poids à des estimateurs qui ne sont pas directement des estimateurs de Horvitz-Thompson, comme ceux relatifs à l'échantillonnage à deux phases, souvent utilisés pour les questions environnementales.

Remerciements

Les auteurs tiennent à remercier Pierre Lavallée et Minna Pulkkinen pour leurs échanges constructifs. Ils veulent également remercier l'examineur, le rédacteur associé et le rédacteur adjoint pour leurs commentaires utiles sur la version précédente du texte.

Bibliographie

Ardilly, P., et Le Blanc, D. (2001). [Échantillonnage et pondération d'une enquête auprès de personnes sans domicile : un exemple français](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2001001/article/5859-fra.pdf). *Techniques d'enquête*, 27, 1, 117-127. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2001001/article/5859-fra.pdf>.

- Chapelle-Barry, C. (2008). Enquête sur les pratiques culturelles en 2006. *Agreste Chiffres et Données*.
- Clairin, R., et Brion, P. (1996). *Manuel de Sondages : Applications aux Pays en Développement*. Documents et manuels du CEPED, 3.
- Cordy, C.B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters*, 18(5), 353-362.
- De Vitiis, C., Falorsi, S. et Inglese, F. (2014). Implementing the first ISTAT survey of homeless population by indirect sampling and weight sharing method, Springer. *Contributions to Sampling Statistics*, 119-138.
- Deville, J.-C., et Lavallée, P. (2006). [Sondage indirect : Les fondements de la méthode généralisée du partage des poids](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2006002/article/9551-fra.pdf). *Techniques d'enquête*, 32, 2, 185-196. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2006002/article/9551-fra.pdf>.
- Gregoire, T.G., et Valentine, H.T. (2007). *Sampling Strategies for Natural Resources and the Environment*. CRC Press.
- Himelein, K., Eckman, S. et Murray, S. (2014). Sampling nomads: A new technique for remote, hard-to-reach, and mobile populations. *Journal of Official Statistics*, 30(2), 191-213.
- Köhl, M., et Magnussen, S. (2015). Sampling in forest inventories, Springer. *Tropical Forestry Handbook*, 1-50.
- Laporte, A., Vandentorren, S., Détrez, M.-A., Douay, C., Le Strat, Y., Le Méner, E., Chauvin, P. et The Samenta Research Group (2018). Prevalence of mental disorders and addictions among homeless people in the greater Paris area, France. *International Journal of Environmental Research and Public Health*, 15(2), 241.
- Lavallée, P. (2009). *Indirect Sampling*. Springer Science & Business Media.
- Lawrence, M., McRoberts, R.E., Tomppo, E., Gschwantner, T. et Gabler, K. (2010). Comparisons of national forest inventories, Springer. *National Forest Inventories*, 19-32.
- Lohlé-Tart, L., Clairin, R., François, M. et Gendreau, F. (1988). *De l'Homme au Chiffre. Réflexions sur l'Observation Démographique en Afrique*.

Lorenc, B., Smith, P.A. et Bavdaž, M. (2018). *The Unit Problem and Other Current Topics in Business Survey Methodology*. Cambridge Scholars Publishing.

Mandallaz, D. (2007). *Sampling Techniques for Forest Inventories*. CRC Press.

Roesch, F.A., Green, E.J. et Scott, C.T. (1993). [Un nouveau concept pour l'échantillonnage des forêts](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1993002/article/14453-fra.pdf). *Techniques d'enquête*, 19, 2, 211-216. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/1993002/article/14453-fra.pdf>.

Stevens, D. (1997). Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics: The Official Journal of the International Environmetrics Society*, 8(3), 167-195.

Stevens, D.L., et Urquhart, N.S. (2000). Response designs and support regions in sampling continuous domains. *Environmetrics: The Official Journal of the International Environmetrics Society*, 11(1), 13-41.

Xu, X., et Lavallée, P. (2009). [Traitements de la non-réponse de lien dans l'échantillonnage indirect](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2009002/article/11038-fra.pdf). *Techniques d'enquête*, 35, 2, 165-177. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2009002/article/11038-fra.pdf>.