UNE GÉNÉRALISATION DE LA MÉTHODE DE PARTAGE DES POIDS DANS LE CAS OÙ LA BASE DE SONDAGE EST CONTINUE

Journées de méthodologie statistique INSEE

Paris – mars 2022

Ph. Brion (Irmar)

O. Bouriaud (IGN / LIF)

G. Chauvet (Ensai / Irmar)

Plan de la présentation

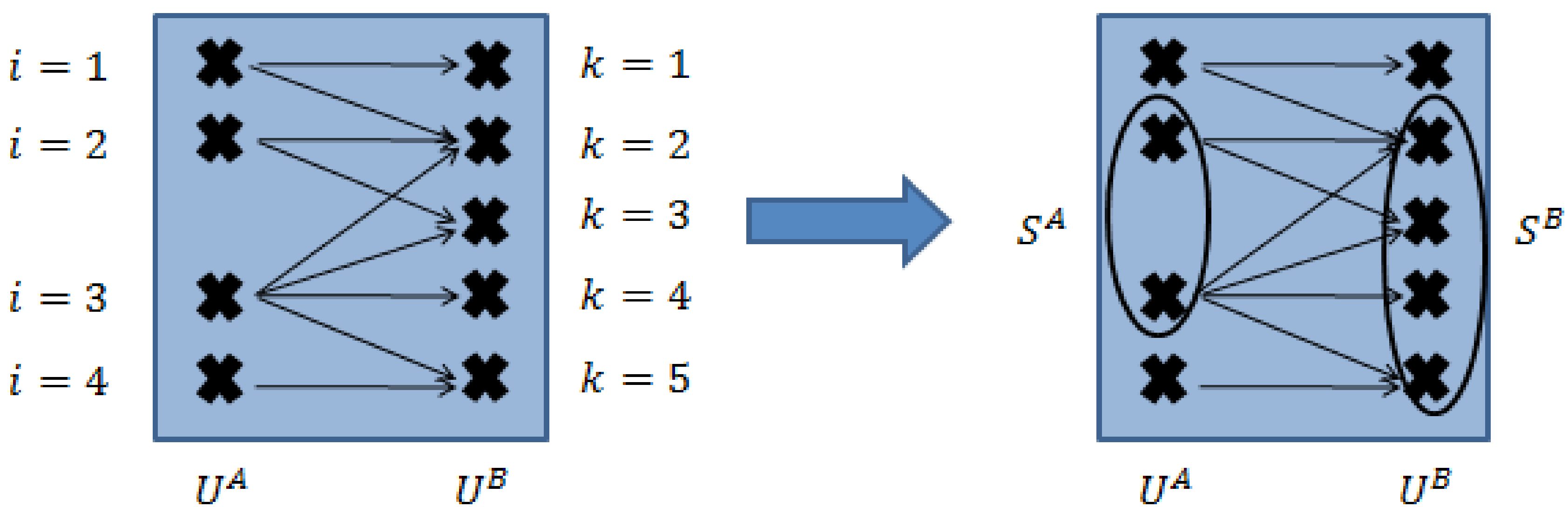
- Rappels sur la méthode de partage des poids
- Exemple de tirage dans une population discrète à partir d'une population continue
- Rappels sur les estimateurs relatifs à un sondage dans une population continue
- Généralisation de la méthode du partage des poids à ce cas
- Conclusion

Rappels sur la méthode de partage des poids (1)

- Une population U^B, discrète, pour laquelle on ne dispose pas de base de sondage
- Une autre population U^A, discrète aussi, avec des liens entre les unités de U^A et U^B
- $L_{ik} = 1 \, si \, i \in U^A \, et \, k \in U^B \, sont \, liées \, , 0 \, sinon$
- On définit *Anc(k)*, l'ensemble des « ancêtres » de l'unité *k* et N_{+k}^{AB} le nombre de ces ancêtres

Rappels sur la méthode de partage des poids (2)

- Pour sélectionner un échantillon s^B dans U^B, on prend l'ensemble des descendants des unités d'un échantillon s^A tiré dans U^A



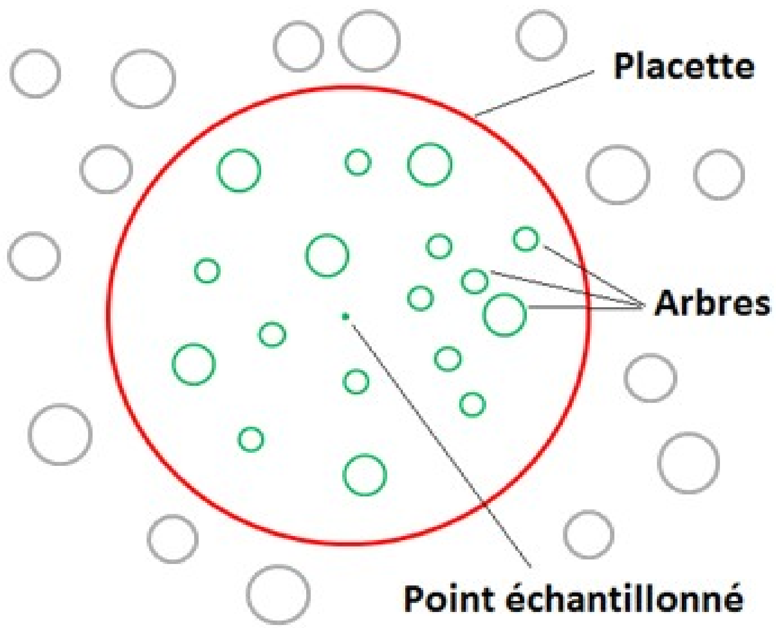
Rappels sur la méthode de partage des poids (3)

Le total d'une variable *y* sur la population U^B peut s'écrire

- $$T_y^B = \sum_{i \in U^A} y_i^A$$
- Avec : $$y_i^A = \sum_{k \in U^B} \frac{L_{ik}^{AB}}{N_{+k}^{AB}} y_k^B$$
- On peut donc estimer ce total à partir de l'estimateur HT sur l'échantillon s^A
- L'estimateur s'écrit également $\sum_{k \in s^B} w_k^B y_k^B$

$$w_k^B = \frac{1}{N_{+k}^{AB}} \sum_{i \in s^A} L_{ik}^{AB} d_i^A$$

Un exemple de tirage dans une population discrète à partir d'une population continue



- Inventaire forestier : un échantillon de points est tiré, et l'enquête de terrain consiste à observer les arbres situés sur des placettes centrées sur ces points

Rappels sur les estimateurs relatifs à une population continue (1)

- Référence : *An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe, Cordy (1993), Statistics and Probability Letters*
- Total qu'on cherche à estimer : $T_y^A = \int_{U^A} y^A(x)\,dx$
- Exemple : estimation du volume de neige sur un lac gelé
- Ce total est l'intégrale d'une densité sur le domaine U^A considéré

Rappels sur les estimateurs relatifs à une population continue (2)

- On va sélectionner un échantillon de points, en nombre fini, à partir d'une densité de probabilité $f(s_1,\ldots,s_n)$, avec des densités marginales f_i
- Par exemple, si on tire chaque point *i* de manière uniforme dans U^A : $f_i(x)=\frac{1(x\in U^A)}{aire(U^A)}$
- On calcule pour chaque point *x* $\Pi(x)=\sum_{i=1,n} f_i(x)$
- $\sum_s \frac{1}{\Pi(x)} y(x)$ est un estimateur sans biais de T_y^A

Généralisation du partage des poids à une base de sondage continue (1)

- Deux populations : U^A population continue, U^B population discrète
- On définit le lien entre un « point » x de U^A et une unité k de U^B comme $L_k^{AB}(x)=1 \text{ si } x \in U^A \text{ et } k \text{ sont liés}, 0 \text{ sinon}$
- On définit $M_{+k}^{AB}=\int_{x \in U^A} L_k^{AB}(x)\,dx$
- Le total de Y sur U^B s'écrit $\sum_{k \in U^B} y_k^B=\int_{x \in U^A} y^A(x)\,dx$
- Avec

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

Généralisation du partage des poids à une base de sondage continue (2)

- Si on tire un échantillon s^A dans U^A et que l'on enquête l'ensemble des unités de U^B « liées » aux unités de l'échantillon s^A (exemple des arbres localisés dans les placettes tirées), on peut proposer un estimateur du total d'une variable y sur U^B grâce à la formalisation de Cordy appliquée à la variable y^A sur U^A

Conclusion

- Travail réalisé dans le cadre de l'inventaire forestier national français, mais applicable à d'autres contextes
- Intérêt : donner une formalisation unifiée à des développements pratiques déjà existants (en particulier pour les inventaires forestiers)
  - formalisation utile en particulier pour des situations « complexes »
- La généralisation du partage des poids proposée concerne l'estimateur HT utilisé dans le cas continu (tel que présenté par Cordy), mais d'autres estimateurs (utilisés par exemple dans le cas de sondages en deux phases) peuvent également donner lieu à généralisation du partage des poids au cas continu