Sondage déterminantal indirect à deux degrés

[SFdS] Séminaire en ligne sur les sondages

Vincent Loonis
Rapporteur du Comité du label de la statistique publique

11 septembre 2025

Introduction

- Des questions issues de l'enseignement et de la pratique de la théorie des sondages
  - $\Rightarrow$ Les probabilités d'inclusion d'ordre 2 et plus jouent un rôle important, mais sont souvent inconnues

Introduction

- Des questions issues de l'enseignement et de la pratique de la théorie des sondages
  - $\Rightarrow$ Les probabilités d'inclusion d'ordre 2 et plus jouent un rôle important, mais sont souvent inconnues
- Les plans déterminantaux sont une solution possible
  - Les probabilités d'inclusion sont connues de manière explicite à tous les ordres
  - Elles sont paramétrables et donc ajustables

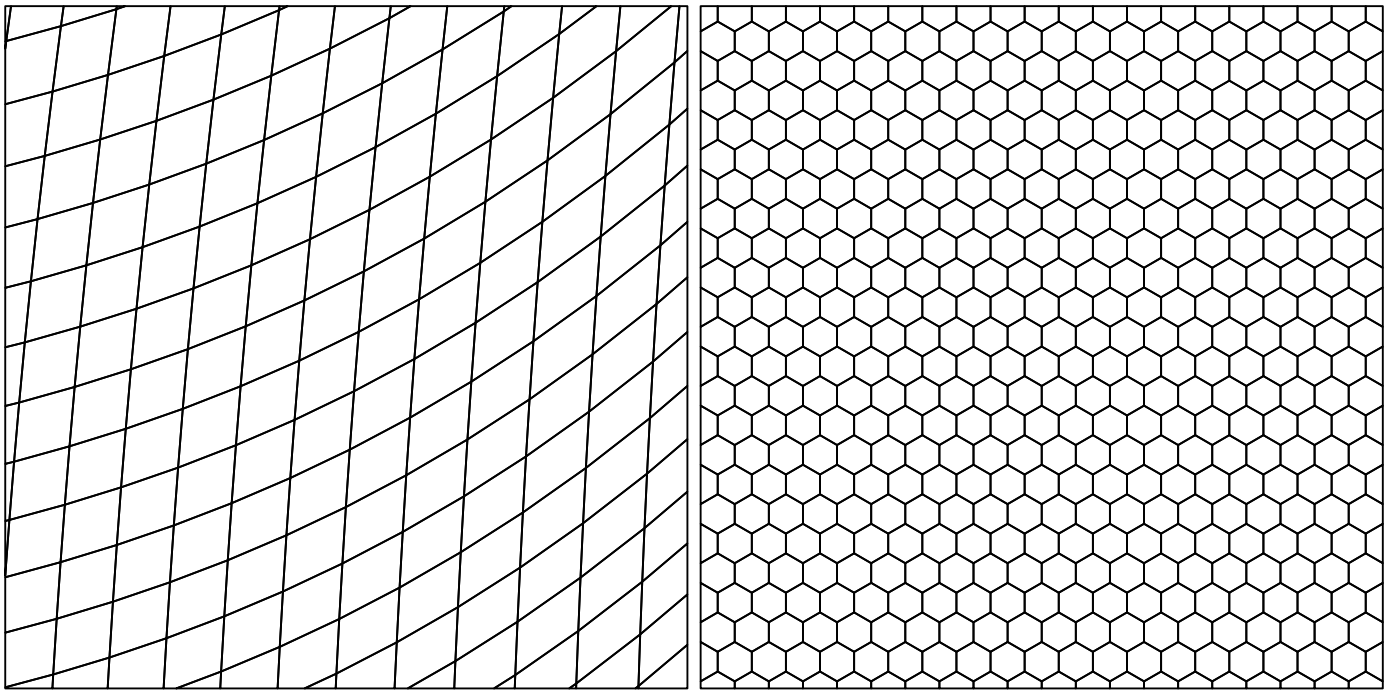
Introduction

- Des questions issues de l'enseignement et de la pratique de la théorie des sondages
  - $\Rightarrow$ Les probabilités d'inclusion d'ordre 2 et plus jouent un rôle important, mais sont souvent inconnues
- Les plans déterminantaux sont une solution possible
  - Les probabilités d'inclusion sont connues de manière explicite à tous les ordres
  - Elles sont paramétrables et donc ajustables
- Deux premiers articles (Loonis et Mary, 2019; Loonis, 2023), centrés sur les propriétés théoriques

Introduction

- Des questions issues de l'enseignement et de la pratique de la théorie des sondages
  - $\Rightarrow$ Les probabilités d'inclusion d'ordre 2 et plus jouent un rôle important, mais sont souvent inconnues
- Les plans déterminantaux sont une solution possible
  - Les probabilités d'inclusion sont connues de manière explicite à tous les ordres
  - Elles sont paramétrables et donc ajustables
- Deux premiers articles (Loonis et Mary, 2019; Loonis, 2023), centrés sur les propriétés théoriques
- Nécessité d'une approche plus pratique, sous la forme
  - de programmes R permettant de s'approprier les différents concepts
  - d'études d'un cas *concret*.

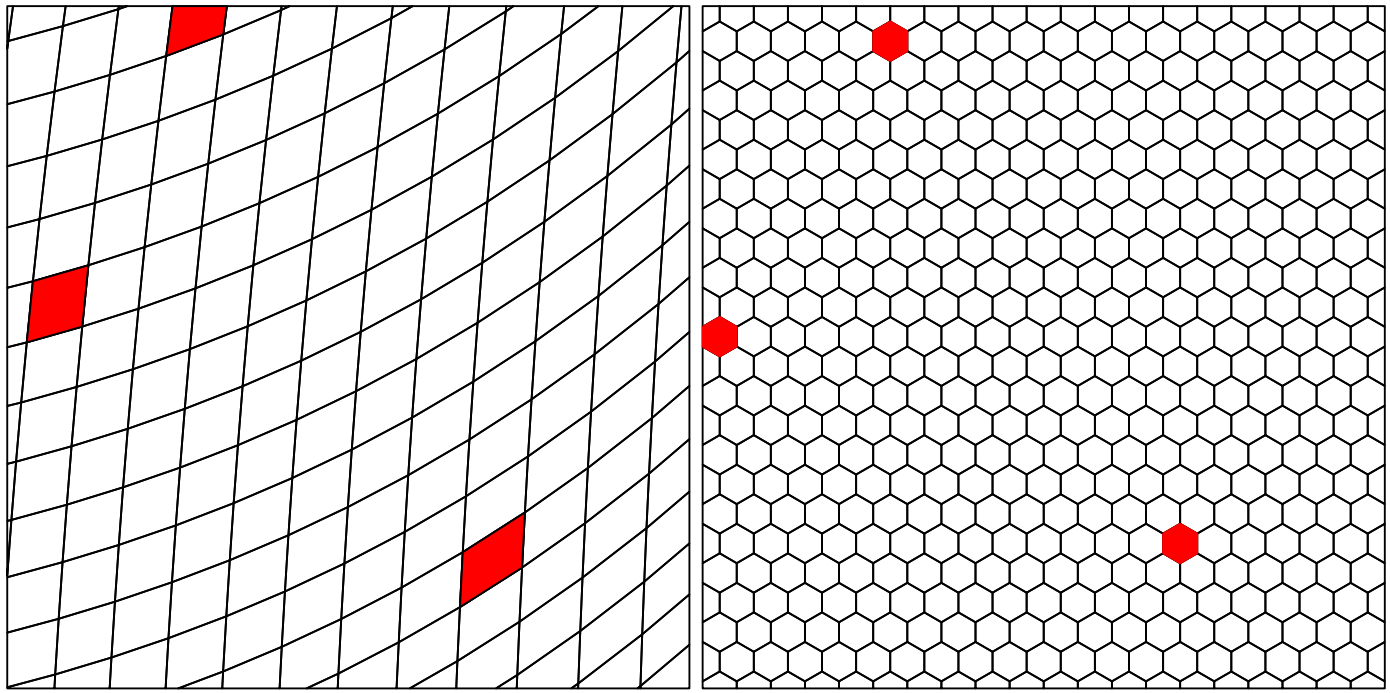
Un problème concret impliquant les probabilités d'inclusion d'ordre 2 et plus



(a) U^A de taille N_A (b) U^B de taille N_B

Figure: Deux populations distinctes (ou non)

Un problème concret impliquant les probabilités d'inclusion d'ordre 2 et plus



(a) $\mathbb{S}^A$ **équilibré sur** X^A (b) $\mathbb{S}^B$ **équilibré sur** X^B

Figure: On souhaite sélectionner deux échantillons $\mathbb{S}^A$ et $\mathbb{S}^B$, **de même taille** (ou presque), **une unité de** $\mathbb{S}^A$ **est proche d'une unité de** $\mathbb{S}^B$ et réciproquement

Un problème concret impliquant les probabilités d'inclusion d'ordre 2 et plus

- Un problème apparu sous une forme proche pour la gestion du réseau enquêteur en face-à-face lors

Un problème concret impliquant les probabilités d'inclusion d'ordre 2 et plus

- Un problème apparu sous une forme proche pour la gestion du réseau enquêteur en face-à-face lors
- des travaux pour l'échantillon maître de l'Insee : (Chevalier *et al.*, 2022)

Un problème concret impliquant les probabilités d'inclusion d'ordre 2 et plus

- Un problème apparu sous une forme proche pour la gestion du réseau enquêteur en face-à-face lors
- des travaux pour l'échantillon maître de l'Insee : (Chevalier *et al.*, 2022)
- Pour le résoudre, ici, on a recours au sondage indirect et aux plans déterminantaux.

Éléments de méthode

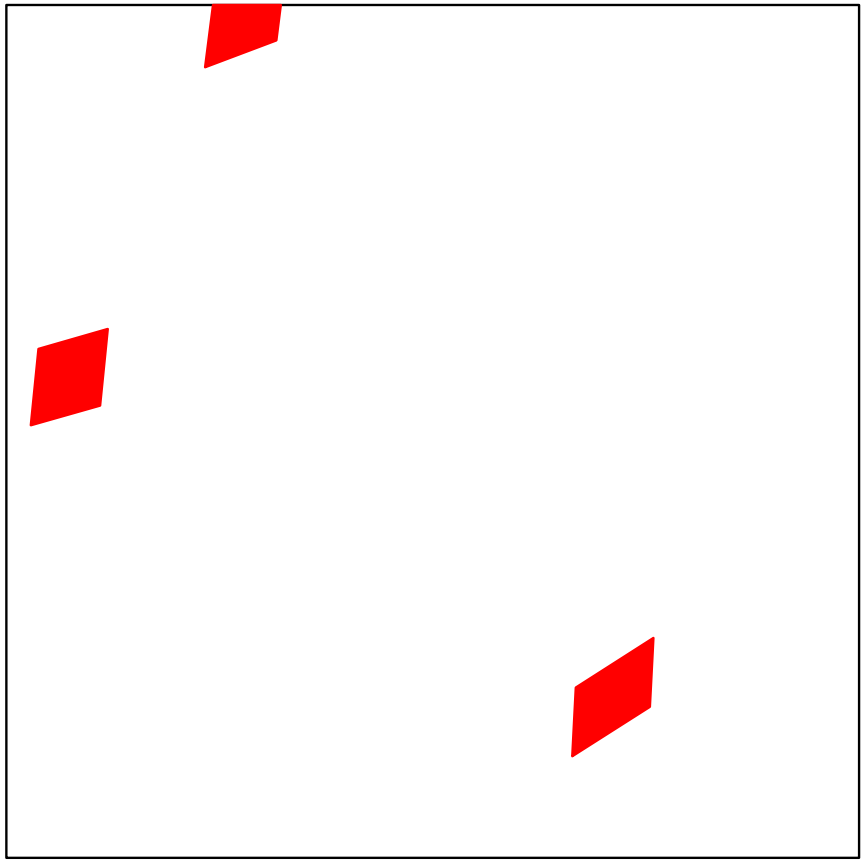


Figure: On sélectionne dans U^A un échantillon aléatoire $\mathbb{S}^A$ selon un plan de sondage donné $\mathcal{P}^A$, avec probabilités d'inclusions fixées Π_i^A. Sur cette population on peut utiliser Horvitz-Thompson.

Éléments de méthode

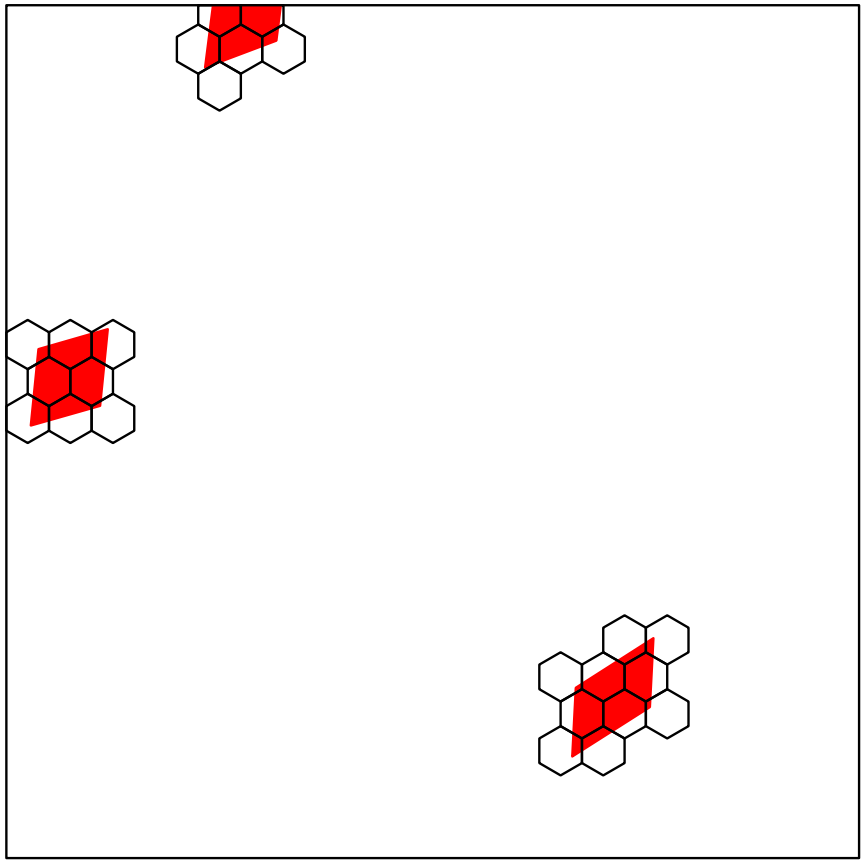


Figure: On considère toutes les unités de U^B qui sont reliés à $\mathbb{S}^A$ selon une relation fixée à l'avance.

Éléments de méthode

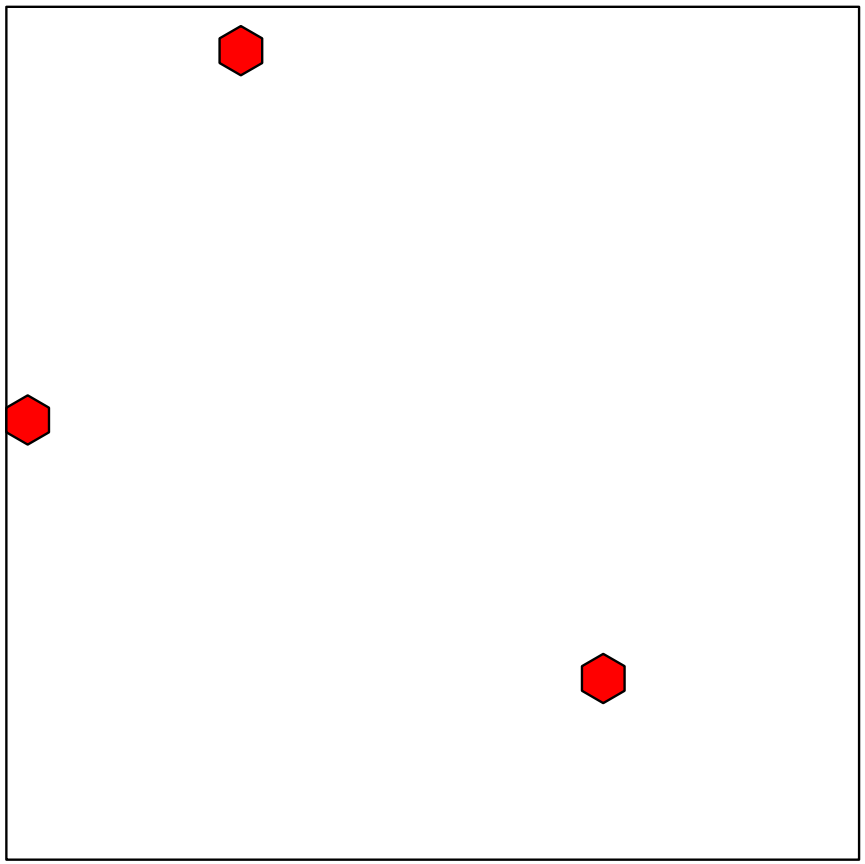


Figure: Dans chaque sous-population de U^B ainsi définie, on sélectionne un sous-échantillon selon un plan prédéfini. On ne peut pas utiliser en général l'estimateur d'Horvitz-Thompson, car les probabilités d'inclusion finales sont inconnues. On utilise la MGPP dans le cadre d'un sondage indirect à deux degrés.

Plan

1. Rappels
   - La théorie des sondages
   - Les plans déterminantaux
   - Le sondage indirect
2. Sondage indirect et plans déterminantaux

Rappels sur la théorie des sondages

- Plan de sondage, échantillon, échantillon aléatoire
- Probabilités d'inclusion
- Considérations pratiques
- Estimation d'un total, précision

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N
- 2^U est la liste de tous les sous-ensembles de U

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N
- 2^U est la liste de tous les sous-ensembles de U
  - Si $N = 3$ et $U = \{1, 2, 3\}$, 2^U est :
    - l'échantillon vide $\emptyset$
    - les échantillons de taille 1 : $\{1\}, \{2\}, \{3\}$
    - les échantillons de taille 2 : $\{1, 2\}, \{1, 3\}, \{2, 3\}$
    - les échantillons de taille 3 : $\{1, 2, 3\}$

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N
- 2^U est la liste de tous les sous-ensembles de U
  - Si $N = 3$ et $U = \{1, 2, 3\}$, 2^U est :
    - l'échantillon vide $\emptyset$
    - les échantillons de taille 1 : $\{1\}, \{2\}, \{3\}$
    - les échantillons de taille 2 : $\{1, 2\}, \{1, 3\}, \{2, 3\}$
    - les échantillons de taille 3 : $\{1, 2, 3\}$
- Un élément de 2^U est noté s et est appelé échantillon

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N
- 2^U est la liste de tous les sous-ensembles de U
  - Si $N = 3$ et $U = \{1, 2, 3\}$, 2^U est :
    - l'échantillon vide $\emptyset$
    - les échantillons de taille 1 : $\{1\}, \{2\}, \{3\}$
    - les échantillons de taille 2 : $\{1, 2\}, \{1, 3\}, \{2, 3\}$
    - les échantillons de taille 3 : $\{1, 2, 3\}$
- Un élément de 2^U est noté s et est appelé échantillon
- Un plan de sondage $\mathcal{P}$ sur U est une loi de probabilité sur 2^U.
  - $p(\emptyset) = 0.1$,$p(\{1\}) = 0.1, p(\{2\}) = 0.2, p(\{3\}) = 0.1,$
    $p(\{1, 2\}) = 0.2, p(\{1, 3\}) = 0.1, p(\{2, 3\}) = 0.1,$
    $p(\{1, 2, 3\}) = 0.1$

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N
- 2^U est la liste de tous les sous-ensembles de U
  - Si $N = 3$ et $U = \{1, 2, 3\}$, 2^U est :
    - l'échantillon vide $\emptyset$
    - les échantillons de taille 1 : $\{1\}, \{2\}, \{3\}$
    - les échantillons de taille 2 : $\{1, 2\}, \{1, 3\}, \{2, 3\}$
    - les échantillons de taille 3 : $\{1, 2, 3\}$
- Un élément de 2^U est noté s et est appelé échantillon
- Un plan de sondage $\mathcal{P}$ sur U est une loi de probabilité sur 2^U.
  - $p(\emptyset) = 0.1$,$p(\{1\}) = 0.1, p(\{2\}) = 0.2, p(\{3\}) = 0.1,$
    $p(\{1, 2\}) = 0.2, p(\{1, 3\}) = 0.1, p(\{2, 3\}) = 0.1,$
    $p(\{1, 2, 3\}) = 0.1$
- On note $\mathbb{S}$ un échantillon aléatoire dont la loi est $\mathcal{P}$.

Plan de sondages, échantillon, échantillon aléatoire

- Population U de taille N
- 2^U est la liste de tous les sous-ensembles de U
  - Si $N = 3$ et $U = \{1, 2, 3\}$, 2^U est :
    - l'échantillon vide $\emptyset$
    - les échantillons de taille 1 : $\{1\}, \{2\}, \{3\}$
    - les échantillons de taille 2 : $\{1, 2\}, \{1, 3\}, \{2, 3\}$
    - les échantillons de taille 3 : $\{1, 2, 3\}$
- Un élément de 2^U est noté s et est appelé échantillon
- Un plan de sondage $\mathcal{P}$ sur U est une loi de probabilité sur 2^U.
  - $p(\emptyset) = 0.1$,$p(\{1\}) = 0.1, p(\{2\}) = 0.2, p(\{3\}) = 0.1,$
    $p(\{1, 2\}) = 0.2, p(\{1, 3\}) = 0.1, p(\{2, 3\}) = 0.1,$
    $p(\{1, 2, 3\}) = 0.1$
- On note $\mathbb{S}$ un échantillon aléatoire dont la loi est $\mathcal{P}$.
- $\mathbb{S}$ est de taille fixe, n, si taille de $s \neq n \Rightarrow p(s) = 0$

probabilités d'inclusion

- Les probabilités d'inclusion : $pr(s \subseteq \mathbb{S}) \neq p(s) = pr(\mathbb{S} = s)$

probabilités d'inclusion

- Les probabilités d'inclusion : $pr(s \subseteq \mathbb{S}) \neq p(s) = pr(\mathbb{S} = s)$
  - ordre 1 (simple)
    - $\pi_1 = pr(\{1\} \subseteq \mathbb{S}) =$
      $p(\{1\}) + p(\{1,2\}) + p(\{1,3\} + p(\{1,2,3\}) = 0.6$

probabilités d'inclusion

- Les probabilités d'inclusion : $pr(s \subseteq \mathbb{S}) \neq p(s) = pr(\mathbb{S} = s)$
  - ordre 1 (simple)
    - $\pi_1 = pr(\{1\} \subseteq \mathbb{S}) =$
      $p(\{1\}) + p(\{1,2\}) + p(\{1,3\} + p(\{1,2,3\}) = 0.6$
  - ordre 2 (double, jointe)
    - $\pi_{12} = pr(\{1,2\} \subseteq \mathbb{S}) = p(\{1,2\}) + p(\{1,2,3\}) = 0.3$

probabilités d'inclusion

- Les probabilités d'inclusion : $pr(s \subseteq \mathbb{S}) \neq p(s) = pr(\mathbb{S} = s)$
  - ordre 1 (simple)
    - $\pi_1 = pr(\{1\} \subseteq \mathbb{S}) =$
      $p(\{1\}) + p(\{1,2\}) + p(\{1,3\} + p(\{1,2,3\}) = 0.6$
  - ordre 2 (double, jointe)
    - $\pi_{12} = pr(\{1,2\} \subseteq \mathbb{S}) = p(\{1,2\}) + p(\{1,2,3\}) = 0.3$
  - ordre 3
    - $\pi_{123} = pr(\{1,2,3\} \subseteq \mathbb{S}) = p(\{1,2,3\}) = 0.1$

probabilités d'inclusion

- Les probabilités d'inclusion : $pr(s \subseteq \mathbb{S}) \neq p(s) = pr(\mathbb{S} = s)$
  - ordre 1 (simple)
    - $\pi_1 = pr(\{1\} \subseteq \mathbb{S}) =$
      $p(\{1\}) + p(\{1,2\}) + p(\{1,3\} + p(\{1,2,3\}) = 0.6$
  - ordre 2 (double, jointe)
    - $\pi_{12} = pr(\{1,2\} \subseteq \mathbb{S}) = p(\{1,2\}) + p(\{1,2,3\}) = 0.3$
  - ordre 3
    - $\pi_{123} = pr(\{1,2,3\} \subseteq \mathbb{S}) = p(\{1,2,3\}) = 0.1$
- Il y a peu de plans pour lesquels les $p(s)$, et donc toutes les $pr(s \subseteq \mathbb{S})$, sont connues (SAS, Poisson...)

Considérations pratiques

- Dans la pratique, on procède, souvent, comme suit :

Considérations pratiques

- Dans la pratique, on procède, souvent, comme suit :
  - On considère des contraintes d'organisation et de précision et les différentes réponses possibles
    - **Organisation** : plusieurs degrés, deux phases...
    - **Précision** : stratification, probabilités inégales, équilibrage...

Considérations pratiques

- Dans la pratique, on procède, souvent, comme suit :
  - On considère des contraintes d'organisation et de précision et les différentes réponses possibles
    - **Organisation** : plusieurs degrés, deux phases...
    - **Précision** : stratification, probabilités inégales, équilibrage...
  - On fixe les probabilités d'ordre 1 : $\pi_i = \Pi_i$
    - $\pi_i = n/N$ probabilités constantes
    - π_i probabilités proportionnelles à un critère de taille

Considérations pratiques

- Dans la pratique, on procède, souvent, comme suit :
  - On considère des contraintes d'organisation et de précision et les différentes réponses possibles
    - **Organisation** : plusieurs degrés, deux phases...
    - **Précision** : stratification, probabilités inégales, équilibrage...
  - On fixe les probabilités d'ordre 1 : $\pi_i = \Pi_i$
    - $\pi_i = n/N$ probabilités constantes
    - π_i probabilités proportionnelles à un critère de taille
  - On cherche une procédure aléatoire qui respecte les contraintes et $\pi_i = \Pi_i$

Considérations pratiques

- Dans la pratique, on procède, souvent, comme suit :
  - On considère des contraintes d'organisation et de précision et les différentes réponses possibles
    - **Organisation** : plusieurs degrés, deux phases...
    - **Précision** : stratification, probabilités inégales, équilibrage...
  - On fixe les probabilités d'ordre 1 : $\pi_i = \Pi_i$
    - $\pi_i = n/N$ probabilités constantes
    - π_i probabilités proportionnelles à un critère de taille
  - On cherche une procédure aléatoire qui respecte les contraintes et $\pi_i = \Pi_i$

$\Rightarrow$ Les $p(s)$, $pr(s \subseteq \mathbb{S})$, π_{ij} sont souvent inconnues. Le cadre reste valide néanmoins.

Estimation d'un total, précision

- On souhaite estimer $t_y = \sum_{i \in U} y_i$ le total d'une **variable d'intérêt** $y = y_1, \cdots, y_N$

Estimation d'un total, précision

- On souhaite estimer $t_y = \sum_{i \in U} y_i$ le total d'une **variable d'intérêt** $y = y_1, \cdots, y_N$
- On utilise un estimateur de t_y : fonction des valeurs de y pour les seuls individus de $\mathbb{S}$

Estimation d'un total, précision

- On souhaite estimer $t_y = \sum_{i \in U} y_i$ le total d'une **variable d'intérêt** $y = y_1, \cdots, y_N$
- On utilise un estimateur de t_y : fonction des valeurs de y pour les seuls individus de $\mathbb{S}$
- Estimateur de Horvitz-Thompson : $\hat{t}_y^{HT} = \sum_{i \in \mathbb{S}} y_i / \pi_i$

Estimation d'un total, précision

- On souhaite estimer $t_y = \sum_{i \in U} y_i$ le total d'une **variable d'intérêt** $y = y_1, \cdots, y_N$
- On utilise un estimateur de t_y : fonction des valeurs de y pour les seuls individus de $\mathbb{S}$
- Estimateur de Horvitz-Thompson : $\hat{t}_y^{HT} = \sum_{i \in \mathbb{S}} y_i / \pi_i$
- La variance de cet estimateur est : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$
  - où $\tilde{\Delta}^{\mathcal{P}}$ est une matrice $(N \times N)$ telle que $\tilde{\Delta}_{ij} = (\pi_{ij} - \pi_i \pi_j)/(\pi_i \pi_j)$

Estimation d'un total, précision

- On souhaite estimer $t_y = \sum_{i \in U} y_i$ le total d'une **variable d'intérêt** $y = y_1, \cdots, y_N$
- On utilise un estimateur de t_y : fonction des valeurs de y pour les seuls individus de $\mathbb{S}$
- Estimateur de Horvitz-Thompson : $\hat{t}_y^{HT} = \sum_{i \in \mathbb{S}} y_i / \pi_i$
- La variance de cet estimateur est : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$
  - où $\tilde{\Delta}^{\mathcal{P}}$ est une matrice $(N \times N)$ telle que $\tilde{\Delta}_{ij} = (\pi_{ij} - \pi_i \pi_j)/(\pi_i \pi_j)$

$\Rightarrow$ $\tilde{\Delta}^{\mathcal{P}}$ est souvent inconnue car les π_{ij} sont inconnues.

Plans de sondages déterminantaux

- Définition
- Quelques propriétés
- Intérêt principal

Définition

- Soit K une matrice réelle de taille $(N \times N)$, symétrique dont les valeurs propres sont dans $[0, 1]$

Définition

- Soit K une matrice réelle de taille $(N \times N)$, symétrique dont les valeurs propres sont dans $[0, 1]$
- Il existe un plan de sondage déterminantal, DSD(K), associé à K, tel que : $\Pr(s \subseteq \mathbb{S}) = \det(K_{|s})$

Définition

- Soit K une matrice réelle de taille $(N \times N)$, symétrique dont les valeurs propres sont dans $[0, 1]$
- Il existe un plan de sondage déterminantal, DSD(K), associé à K, tel que : $\Pr(s \subseteq \mathbb{S}) = \det(K_{|s})$

Exemple

$$K = \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{10}} & \frac{\sqrt{3}}{2\sqrt{14}} & \frac{\sqrt{3}}{\sqrt{70}} & \frac{1}{\sqrt{35}} & \frac{1}{\sqrt{65}} & \frac{1}{2\sqrt{26}} \\ \frac{1}{\sqrt{10}} & \frac{1}{5} & \frac{\sqrt{3}}{2\sqrt{35}} & \frac{\sqrt{3}}{5\sqrt{7}} & \frac{\sqrt{2}}{5\sqrt{7}} & \frac{\sqrt{2}}{5\sqrt{13}} & \frac{1}{2\sqrt{65}} \\ \frac{\sqrt{3}}{2\sqrt{14}} & \frac{\sqrt{3}}{2\sqrt{35}} & \frac{3}{4} & -\frac{1}{2\sqrt{5}} & -\frac{1}{\sqrt{30}} & -\frac{\sqrt{7}}{\sqrt{390}} & -\frac{\sqrt{7}}{4\sqrt{39}} \\ \frac{\sqrt{3}}{\sqrt{70}} & \frac{\sqrt{3}}{5\sqrt{7}} & -\frac{1}{2\sqrt{5}} & \frac{4}{5} & -\frac{\sqrt{2}}{5\sqrt{3}} & -\frac{\sqrt{14}}{5\sqrt{39}} & -\frac{\sqrt{7}}{2\sqrt{195}} \\ \frac{1}{\sqrt{35}} & \frac{\sqrt{2}}{5\sqrt{7}} & -\frac{1}{\sqrt{30}} & -\frac{\sqrt{2}}{5\sqrt{3}} & \frac{2}{5} & \frac{2\sqrt{7}}{5\sqrt{13}} & \frac{\sqrt{7}}{\sqrt{130}} \\ \frac{1}{\sqrt{65}} & \frac{\sqrt{2}}{5\sqrt{13}} & -\frac{\sqrt{7}}{\sqrt{390}} & -\frac{\sqrt{14}}{5\sqrt{39}} & \frac{2\sqrt{7}}{5\sqrt{13}} & \frac{3}{5} & -\frac{1}{\sqrt{10}} \\ \frac{1}{2\sqrt{26}} & \frac{1}{2\sqrt{65}} & -\frac{\sqrt{7}}{4\sqrt{39}} & -\frac{\sqrt{7}}{2\sqrt{195}} & \frac{\sqrt{7}}{\sqrt{130}} & -\frac{1}{\sqrt{10}} & \frac{3}{4} \end{pmatrix}.$$

$$pr(s = \{1;3;5\} \subseteq \mathbb{S}) = \det(K_{|\{1;3;5\}}) = \det \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2\sqrt{14}} & \frac{1}{\sqrt{35}} \\ \frac{\sqrt{3}}{2\sqrt{14}} & \frac{3}{4} & -\frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{35}} & -\frac{1}{\sqrt{30}} & \frac{2}{5} \end{pmatrix} = \frac{8}{105} = \pi_{135}.$$

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

- $\pi_i = K_{ii} \Rightarrow$ la diagonale de K donne les probabilités d'inclusion d'ordre 1

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

- $\pi_i = K_{ii} \Rightarrow$ la diagonale de K donne les probabilités d'inclusion d'ordre 1
- $\pi_{ij} = K_{ii}K_{jj} - K_{ij}^2 \Rightarrow$ les termes hors diagonale sont directement liés aux probabilités d'inclusion d'ordre 2

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

- $\pi_i = K_{ii} \Rightarrow$ la diagonale de K donne les probabilités d'inclusion d'ordre 1
- $\pi_{ij} = K_{ii}K_{jj} - K_{ij}^2 \Rightarrow$ les termes hors diagonale sont directement liés aux probabilités d'inclusion d'ordre 2
- $\mathbb{S}$ est de taille fixe si et seulement si K est une matrice de projection $\Rightarrow K^2 = K$

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

- $\pi_i = K_{ii} \Rightarrow$ la diagonale de K donne les probabilités d'inclusion d'ordre 1
- $\pi_{ij} = K_{ii}K_{jj} - K_{ij}^2 \Rightarrow$ les termes hors diagonale sont directement liés aux probabilités d'inclusion d'ordre 2
- $\mathbb{S}$ est de taille fixe si et seulement si K est une matrice de projection $\Rightarrow K^2 = K$
- Le complémentaire de $\mathbb{S}$ dans U a pour loi $DSD(I_N - K)$

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

- $\pi_i = K_{ii} \Rightarrow$ la diagonale de K donne les probabilités d'inclusion d'ordre 1
- $\pi_{ij} = K_{ii}K_{jj} - K_{ij}^2 \Rightarrow$ les termes hors diagonale sont directement liés aux probabilités d'inclusion d'ordre 2
- $\mathbb{S}$ est de taille fixe si et seulement si K est une matrice de projection $\Rightarrow K^2 = K$
- Le complémentaire de $\mathbb{S}$ dans U a pour loi $DSD(I_N - K)$
- Pour un jeu de probabilités d'inclusion Π d'ordre 1 donné a priori, on sait construire toutes les matrices K dont la diagonale est Π, y compris, celles de projection orthogonale.

Quelques propriétés

si $\mathbb{S}$ est un échantillon aléatoire de loi $\mathcal{P} = DSD(K)$ alors :

- $\pi_i = K_{ii} \Rightarrow$ la diagonale de K donne les probabilités d'inclusion d'ordre 1
- $\pi_{ij} = K_{ii}K_{jj} - K_{ij}^2 \Rightarrow$ les termes hors diagonale sont directement liés aux probabilités d'inclusion d'ordre 2
- $\mathbb{S}$ est de taille fixe si et seulement si K est une matrice de projection $\Rightarrow K^2 = K$
- Le complémentaire de $\mathbb{S}$ dans U a pour loi $DSD(I_N - K)$
- Pour un jeu de probabilités d'inclusion Π d'ordre 1 donné a priori, on sait construire toutes les matrices K dont la diagonale est Π, y compris, celles de projection orthogonale.
- Pour toute matrice K, on sait sélectionner un échantillon aléatoire selon un plan $DSD(K)$

Intérêt principal

$\Rightarrow$: On sait calculer et optimiser toute fonction des probabilités d'inclusion à n'importe quel ordre

Intérêt principal

$\Rightarrow$: On sait calculer et optimiser toute fonction des probabilités d'inclusion à n'importe quel ordre

- Pour tous les plans : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$

Intérêt principal

$\Rightarrow$: On sait calculer et optimiser toute fonction des probabilités d'inclusion à n'importe quel ordre

- Pour tous les plans : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$
- Pour les plans déterminantaux :

$$\tilde{\Delta} = Diag(K)^{-1} \left[K \odot (I_N - K) \right] Diag(K)^{-1}$$

,

Intérêt principal

$\Rightarrow$: On sait calculer et optimiser toute fonction des probabilités d'inclusion à n'importe quel ordre

- Pour tous les plans : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$
- Pour les plans déterminantaux :

$$\tilde{\Delta} = Diag(K)^{-1} \left[K \odot (I_N - K)\right] Diag(K)^{-1}$$

- Pour un ensemble de Q **variables auxiliaires** $x^1, \cdots, x^Q$ et une matrice K, on peut calculer :

$$\begin{aligned} C_\alpha(X^Q, K) &= \sum_{q=1}^{Q} \alpha_q \, \text{var}(\hat{t}_{x^q}^{HT}) \\ &= \sum_{q=1}^{Q} \alpha_q x^{q\intercal} \text{Diag}(K)^{-1} \left[K \odot (I_N - K)\right] \text{Diag}(K)^{-1} x^q \end{aligned}$$

Intérêt principal

$\Rightarrow$: On sait calculer et optimiser toute fonction des probabilités d'inclusion à n'importe quel ordre

- Pour tous les plans : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$
- Pour les plans déterminantaux :

$$\tilde{\Delta} = Diag(K)^{-1} \left[K \odot (I_N - K)\right] Diag(K)^{-1}$$

- Pour un ensemble de Q **variables auxiliaires** $x^1, \cdots, x^Q$ et une matrice K, on peut calculer :

$$\begin{aligned} C_\alpha(X^Q, K) &= \sum_{q=1}^{Q} \alpha_q \, \text{var}(\hat{t}_{x^q}^{HT}) \\ &= \sum_{q=1}^{Q} \alpha_q x^{q\intercal} \text{Diag}(K)^{-1} \left[K \odot (I_N - K)\right] \text{Diag}(K)^{-1} x^q \end{aligned}$$

- On peut minimiser $C_\alpha(X^Q, K)$ sous les contraintes :
  - K est symétrique
  - K est matrice de projection (Taille fixe)
  - $Diag(K) = \Pi$ (Probabilités d'inclusion fixées).

Intérêt principal

$\Rightarrow$: On sait calculer et optimiser toute fonction des probabilités d'inclusion à n'importe quel ordre

- Pour tous les plans : $var(\hat{t}_y^{HT}) = y^\intercal \tilde{\Delta}^{\mathcal{P}} y$
- Pour les plans déterminantaux :

$$\tilde{\Delta} = Diag(K)^{-1} \left[K \odot (I_N - K)\right] Diag(K)^{-1}$$

- Pour un ensemble de Q **variables auxiliaires** $x^1, \cdots, x^Q$ et une matrice K, on peut calculer :

$$\begin{aligned} C_\alpha(X^Q, K) &= \sum_{q=1}^{Q} \alpha_q \, \text{var}(\hat{t}_{x^q}^{HT}) \\ &= \sum_{q=1}^{Q} \alpha_q x^{q\intercal} \text{Diag}(K)^{-1} \left[K \odot (I_N - K)\right] \text{Diag}(K)^{-1} x^q \end{aligned}$$

- On peut minimiser $C_\alpha(X^Q, K)$ sous les contraintes :
  - K est symétrique
  - K est matrice de projection (Taille fixe)
  - $Diag(K) = \Pi$ (Probabilités d'inclusion fixées).

$\Rightarrow$ D'autres fonctions à calculer, à optimiser, par exemple, avec le sondage indirect

Rappels sur le sondage indirect

Le sondage indirect, pour faire quoi ?

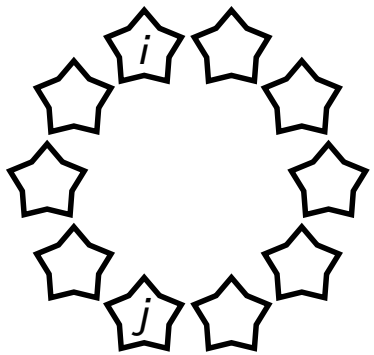
1. Atteindre des populations difficiles à joindre :
   - Des personnes sans-domicile via les services qu'elles consomment
   - Des touristes via les lieux qu'ils visitent
   - Des facteurs via leurs tournées
2. Répondre à des situations plus *théoriques*
   - Pondérations des enquêtes par panel
   - Échantillonnage dans plusieurs bases non disjointes
   - Capture-Recapture
   - Unités légales - Entreprises
   - Coordination spatiale des échantillons

Sondage indirect à deux degrés

- Les populations
- L'échantillon de premier degré
- L'échantillon de second degré
- L'estimation d'un total par la Méthode Généralisée du Partage des Poids (MGPP)
- Des nouveautés ?
  - Une reformulation de la variance de l'estimateur MGPP conduisant à une formulation explicite
    - de la matrice de pondération optimale
    - des probabilités optimales du second degré, dans un cas particulier utile en pratique

Les populations

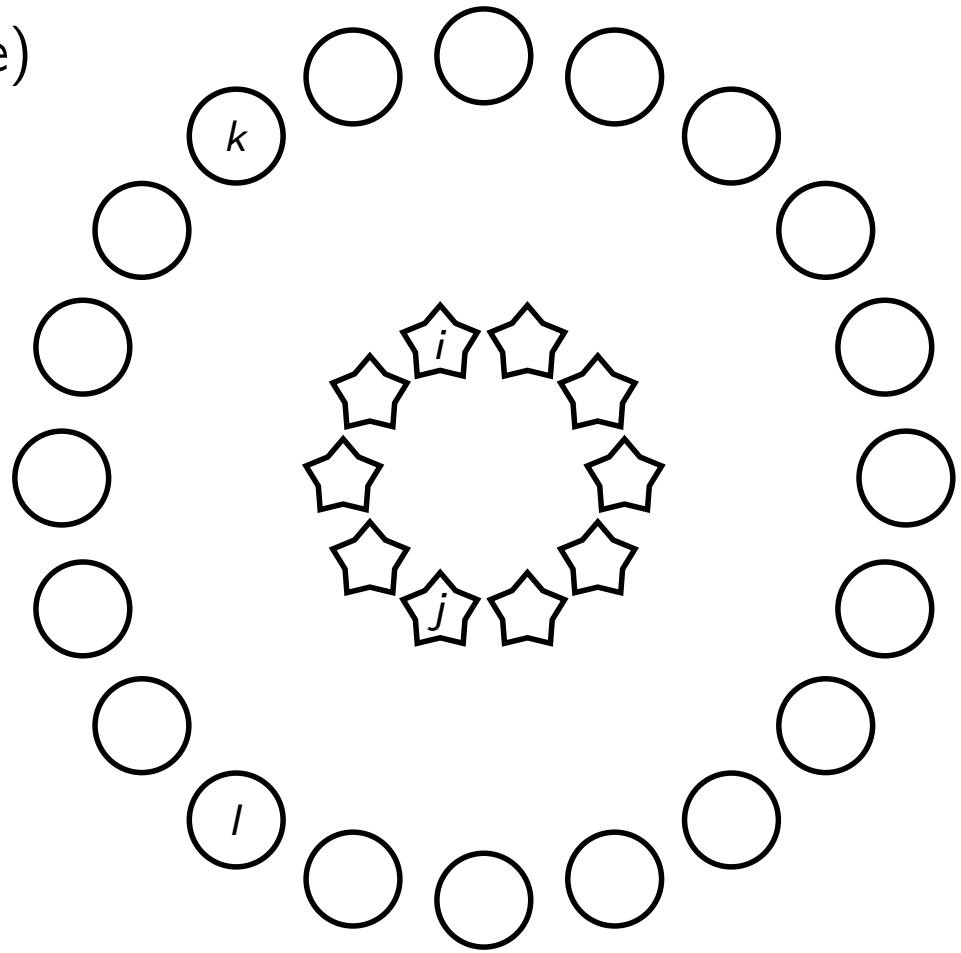
U_A de taille N_A (Population intermédiaire)



Les populations

U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

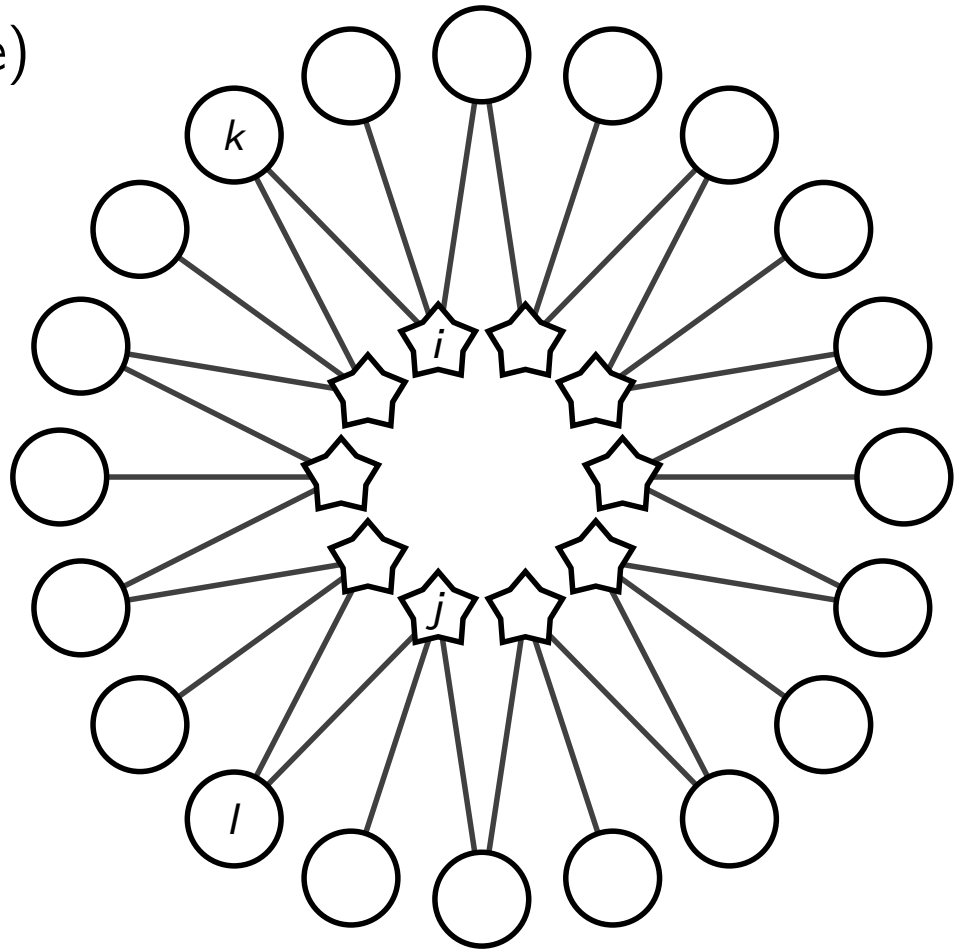


Les populations

U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0



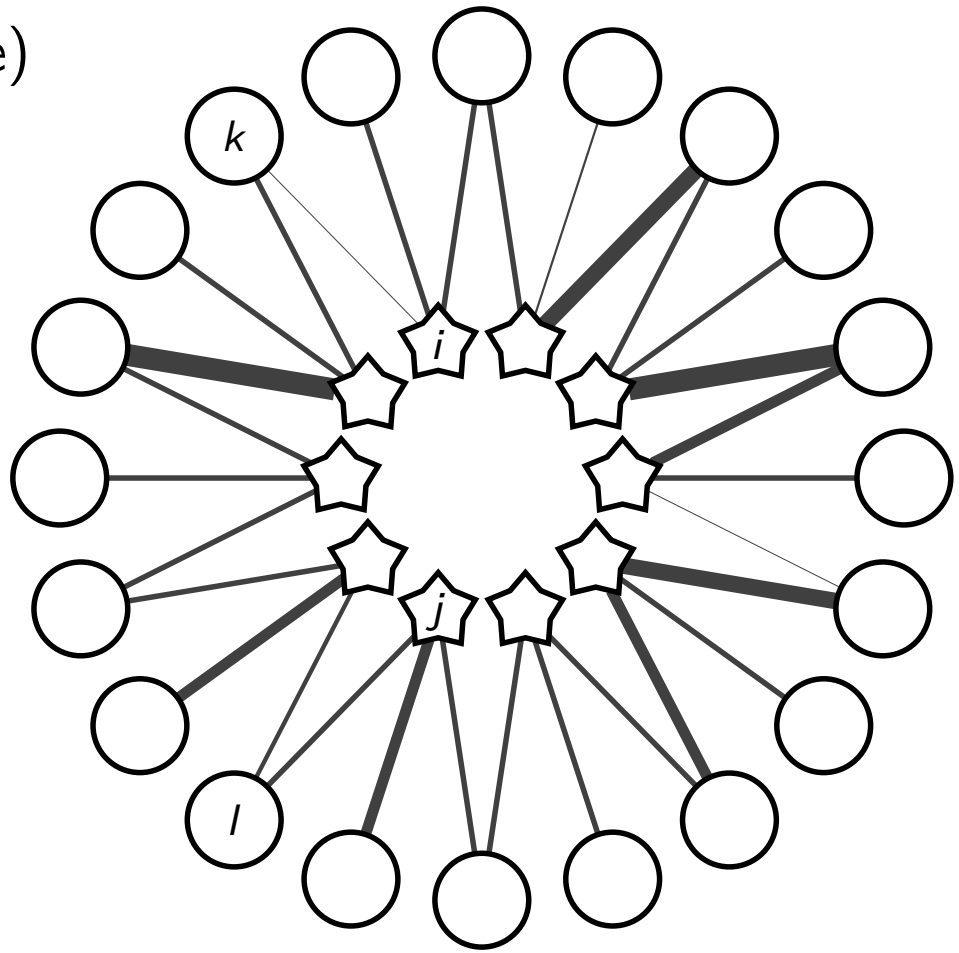
Les populations

U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0

Θ^{AB} de taille $N_A \times N_B$
θ_{ik}^{AB} nombre réel



Les populations

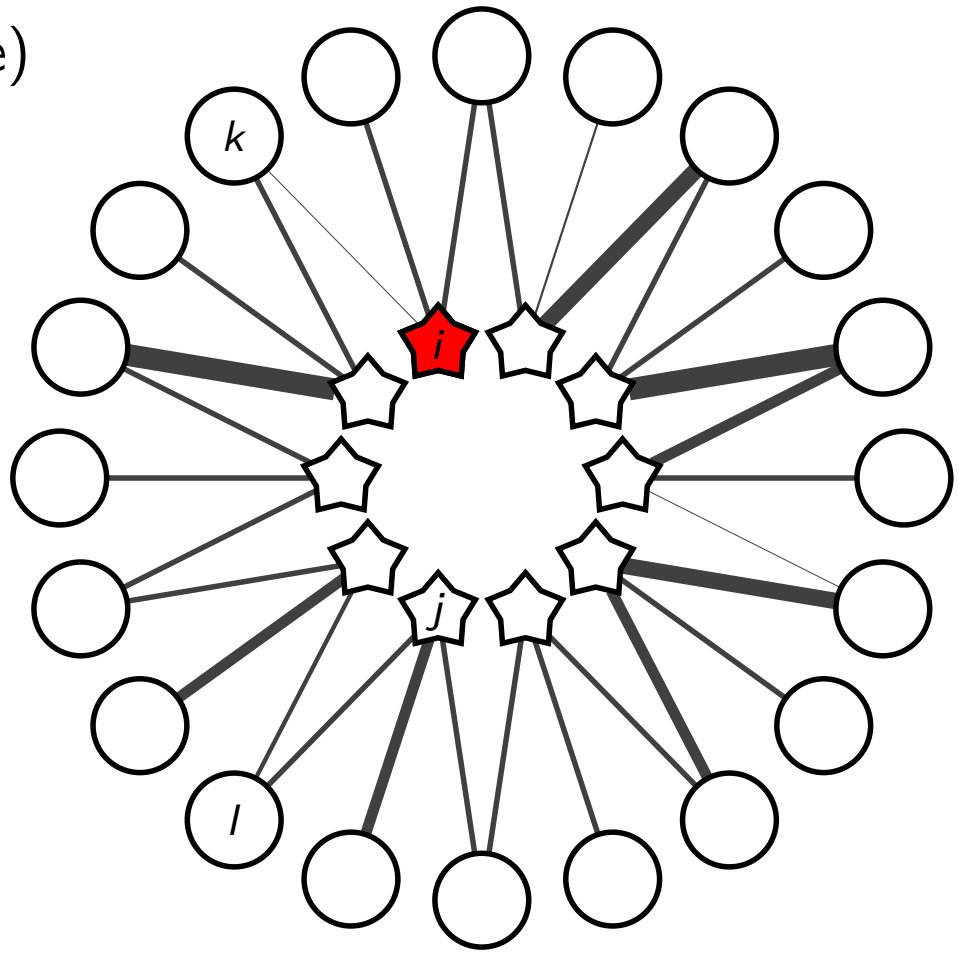
U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0

Θ^{AB} de taille $N_A \times N_B$
θ_{ik}^{AB} nombre réel

$U_i^B = \{k \in U^B | \ell_{ik} = 1\}$



Les populations

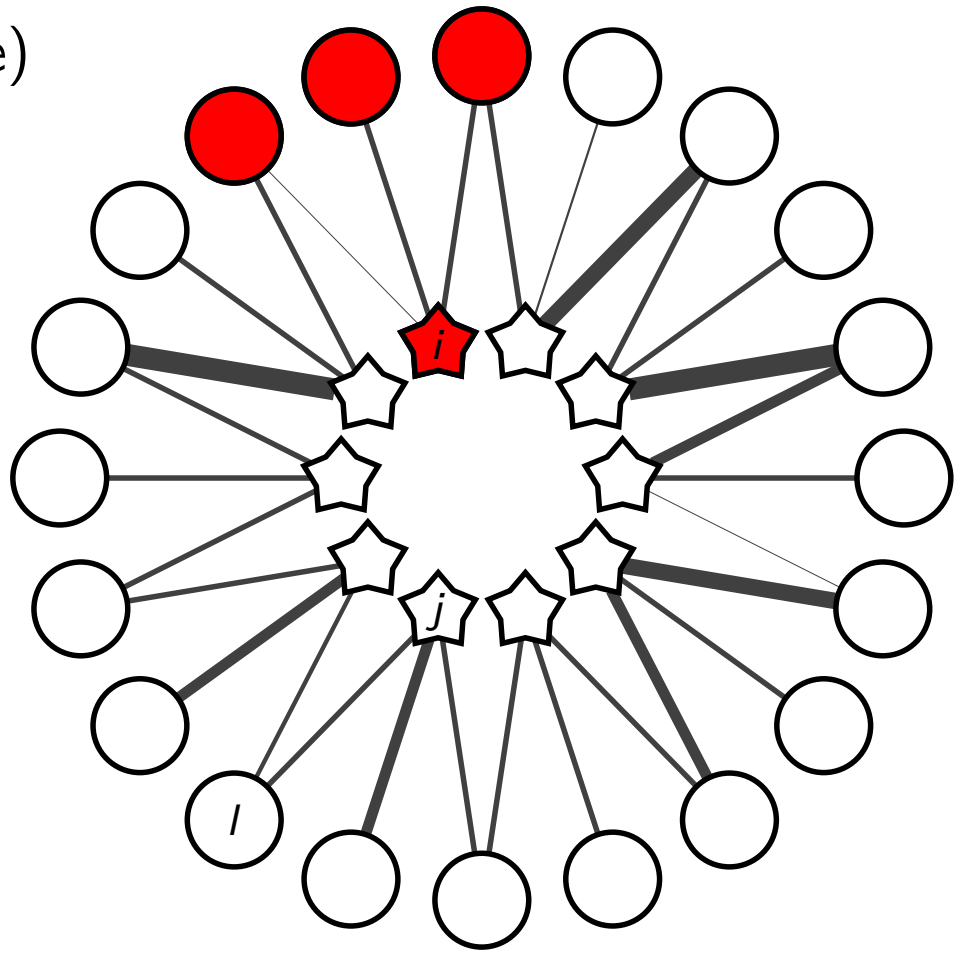
U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0

Θ^{AB} de taille $N_A \times N_B$
θ_{ik}^{AB} nombre réel

$U_i^B = \{k \in U^B | \ell_{ik} = 1\}$



Les populations

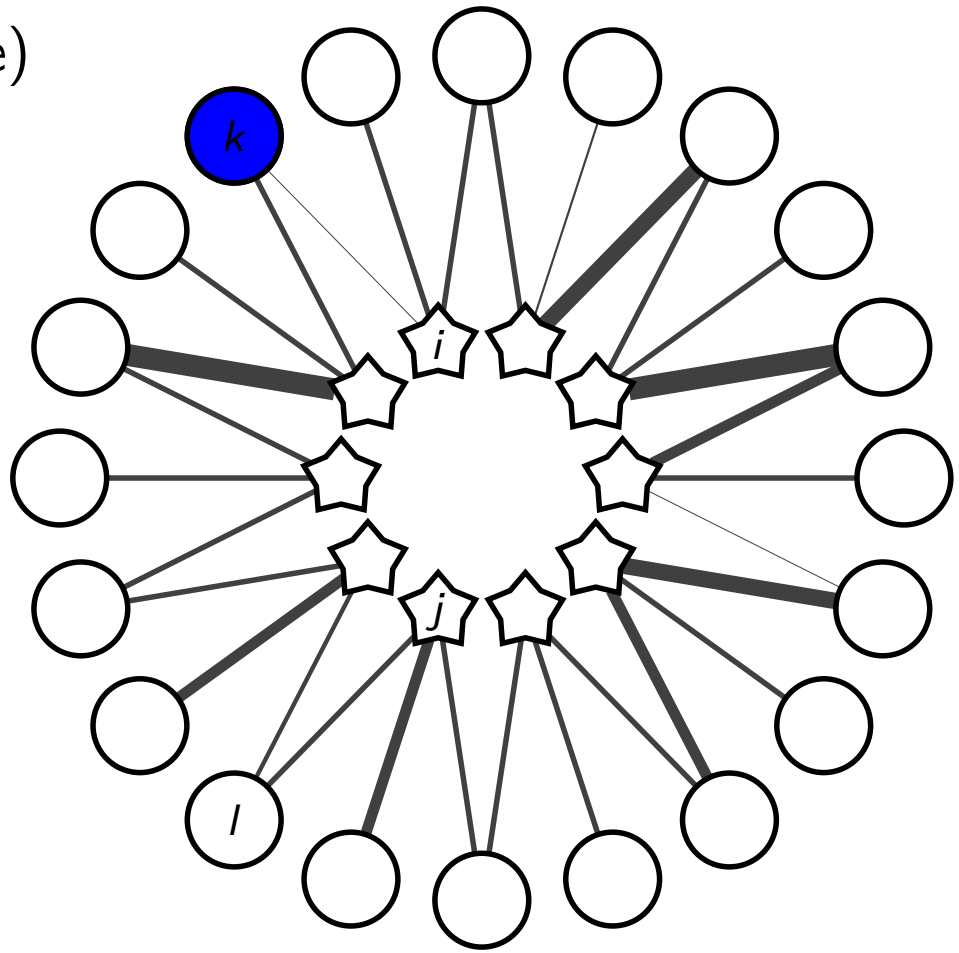
U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0

Θ^{AB} de taille $N_A \times N_B$
θ_{ik}^{AB} nombre réel

$U_i^B = \{k \in U^B | \ell_{ik} = 1\}$
$U_k^A = \{i \in U^A | \ell_{ik} = 1\}$



Les populations

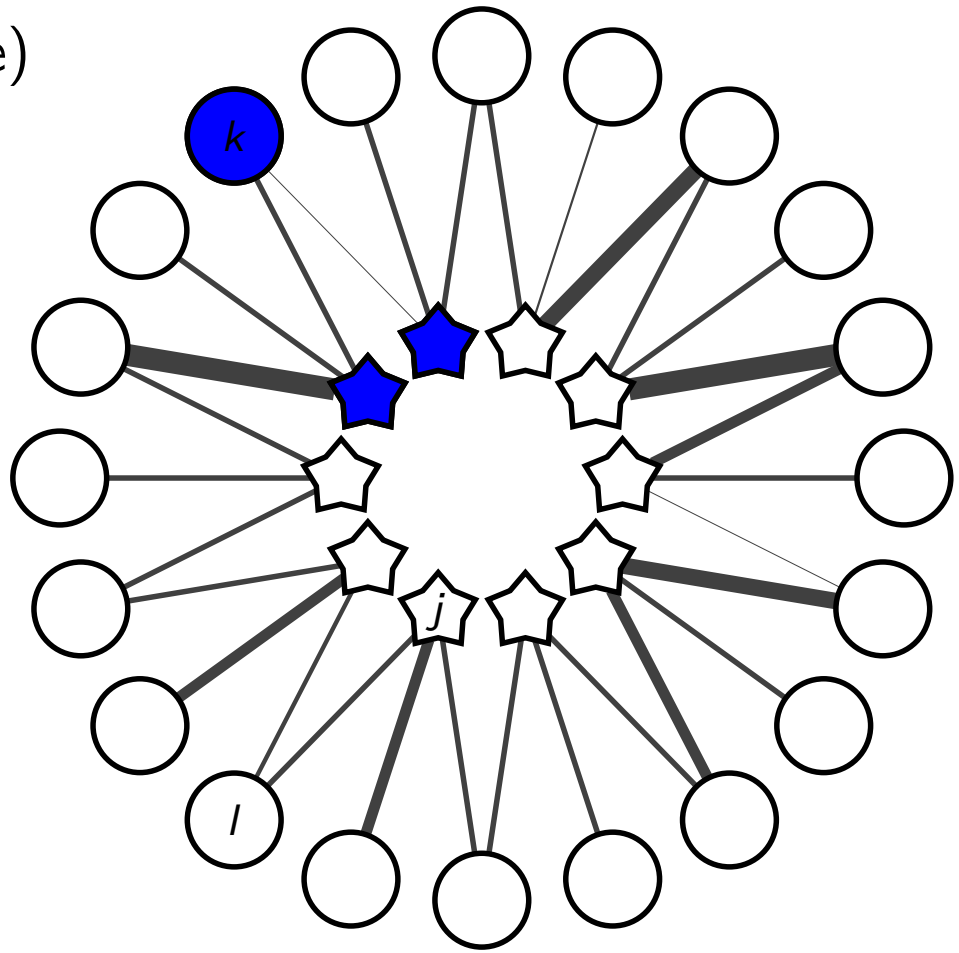
U_A de taille N_A (Population intermédiaire)

U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0

Θ^{AB} de taille $N_A \times N_B$
θ_{ik}^{AB} nombre réel

$U_i^B = \{k \in U^B | \ell_{ik} = 1\}$
$U_k^A = \{i \in U^A | \ell_{ik} = 1\}$



Les populations

U_A de taille N_A (Population intermédiaire)

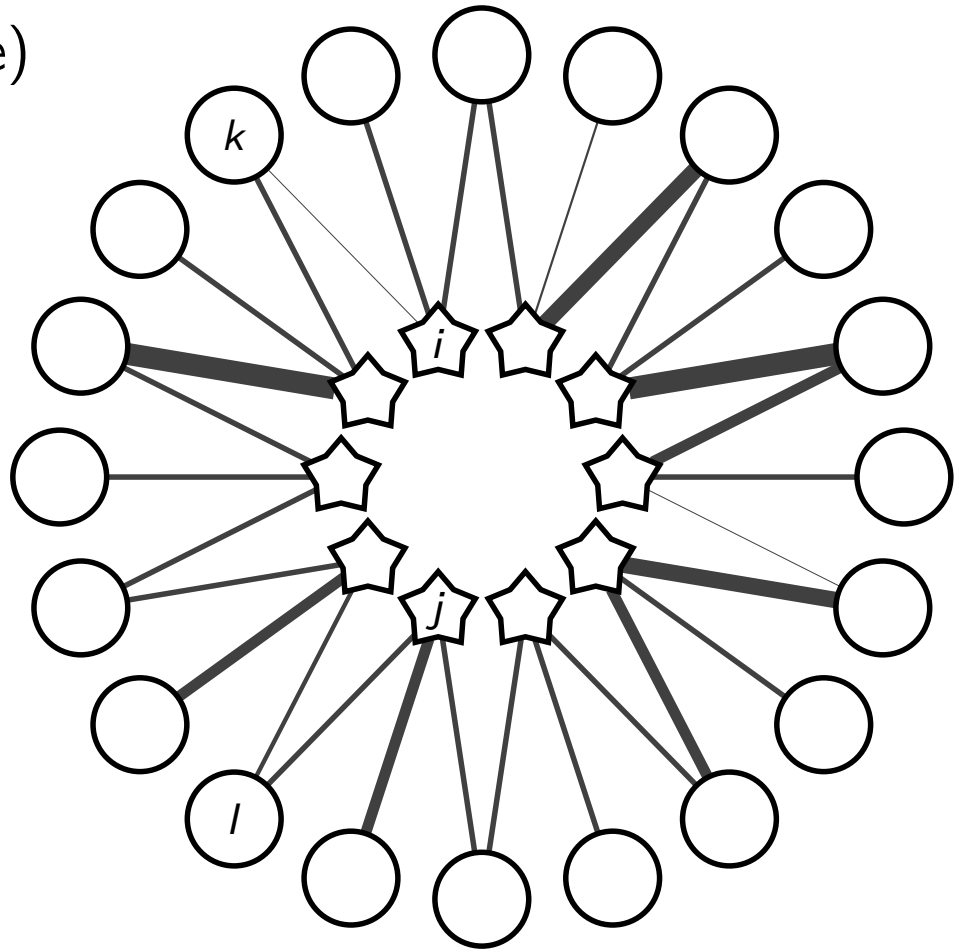
U_B de taille N_B (Population cible)

L^{AB} de taille $N_A \times N_B$
$\ell_{ik}^{AB} = 1$ ou 0

Θ^{AB} de taille $N_A \times N_B$
θ_{ik}^{AB} nombre réel

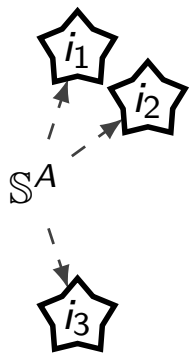
$U_i^B = \{k \in U^B | \ell_{ik} = 1\}$
$U_k^A = \{i \in U^A | \ell_{ik} = 1\}$

L^{AB} est une donnée du problème
Θ^{AB} est un paramètre



Échantillon de premier degré

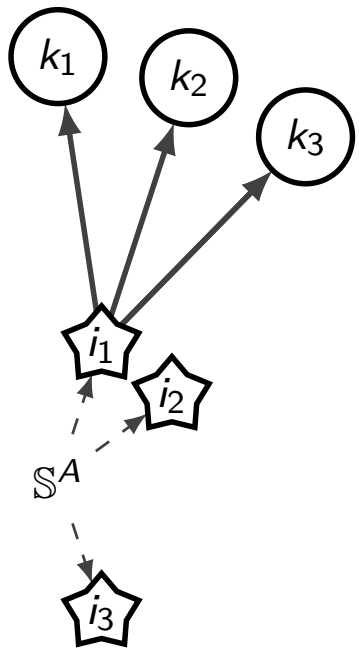
On sélectionne un échantillon $\mathbb{S}^A$ dans U^A selon un plan $\mathcal{P}^A$ **donné** $\Rightarrow \tilde{\Delta}^A$.



Échantillon de premier degré

On sélectionne un échantillon $\mathbb{S}^A$ dans U^A selon un plan $\mathcal{P}^A$ **donné** $\Rightarrow \tilde{\Delta}^A$.

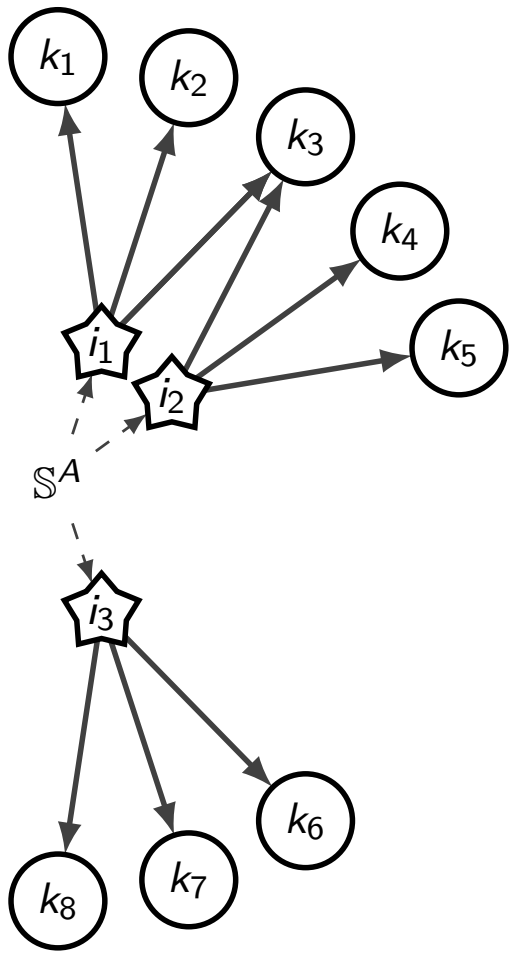
Pour une unité i de $\mathbb{S}^A$, on considère toutes les unités de U^B en lien avec elle (U_i^B).



Échantillon de premier degré

On sélectionne un échantillon $\mathbb{S}^A$ dans U^A selon un plan $\mathcal{P}^A$ **donné** $\Rightarrow \tilde{\Delta}^A$.

Pour une unité i de $\mathbb{S}^A$, on considère toutes les unités de U^B en lien avec elle (U_i^B).

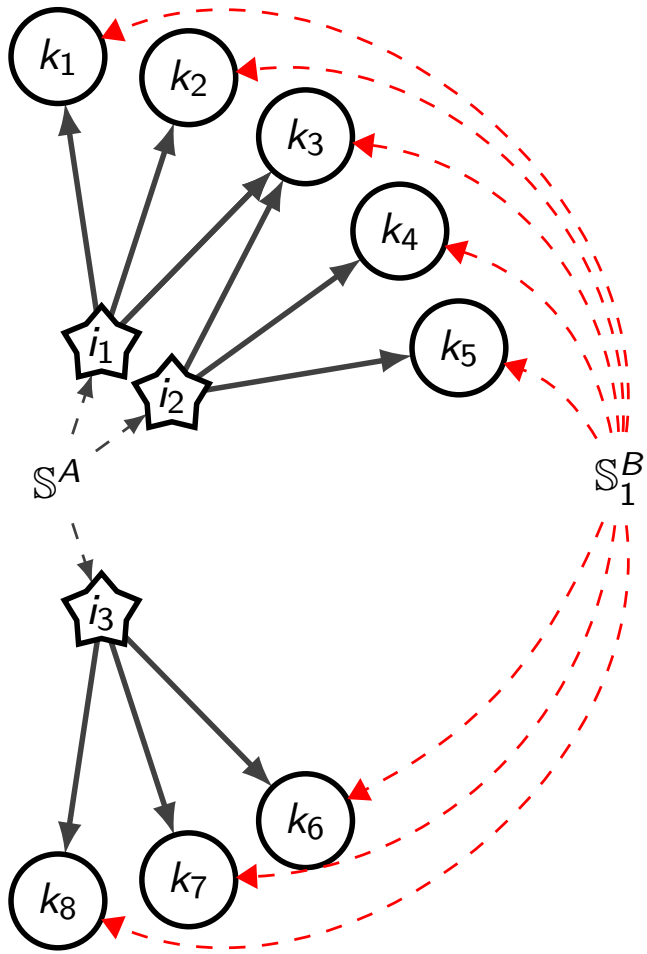


Échantillon de premier degré

On sélectionne un échantillon $\mathbb{S}^A$ dans U^A selon un plan $\mathcal{P}^A$ **donné** $\Rightarrow \tilde{\Delta}^A$.

Pour une unité i de $\mathbb{S}^A$, on considère toutes les unités de U^B en lien avec elle (U_i^B).

L'échantillon cible de premièr degré $(\mathbb{S}^{1B})$ est la réunion de toutes les unités cibles ainsi repérées: $\mathbb{S}^{1B} = \bigcup_{i \in \mathbb{S}^A} U_i^B$



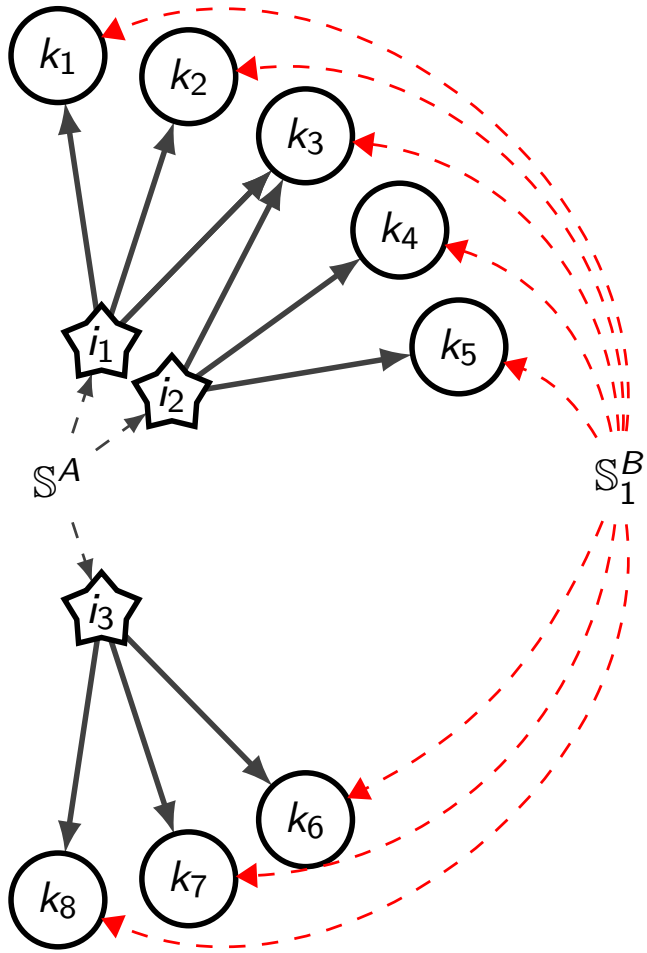
Échantillon de premier degré

On sélectionne un échantillon $\mathbb{S}^A$ dans U^A selon un plan $\mathcal{P}^A$ **donné** $\Rightarrow \tilde{\Delta}^A$.

Pour une unité i de $\mathbb{S}^A$, on considère toutes les unités de U^B en lien avec elle (U_i^B).

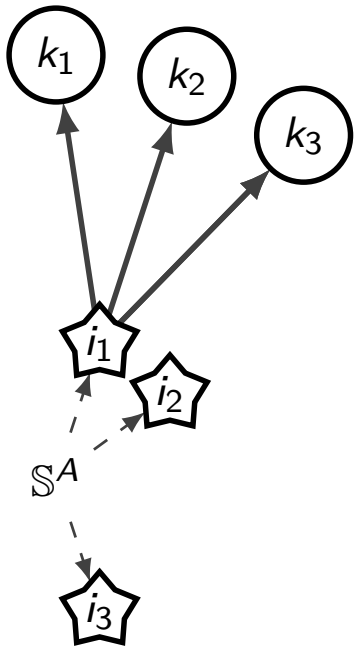
L'échantillon cible de premièr degré $(\mathbb{S}^{1B})$ est la réunion de toutes les unités cibles ainsi repérées: $\mathbb{S}^{1B} = \bigcup_{i \in \mathbb{S}^A} U_i^B$

C'est le cadre étudié par Deville et Lavallée (2006) : ⚠La taille de $\mathbb{S}^{1B}$ peut être grande et variable



Échantillon au second degré (Préparation)

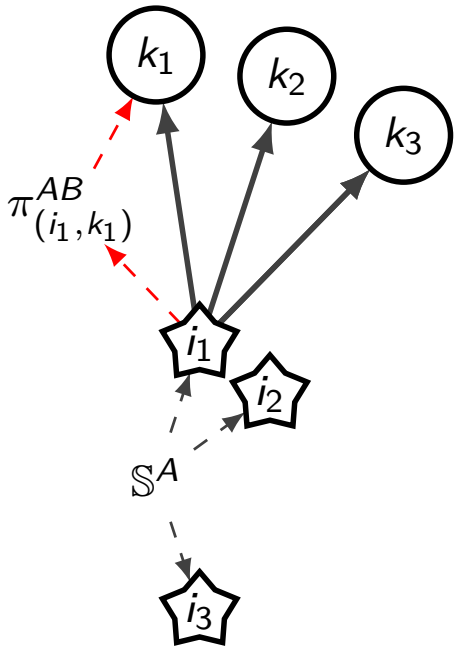
Pour préparer le deuxième degré, on définit (comme paramètre) des probabilités d'inclusion.



Échantillon au second degré (Préparation)

Pour préparer le deuxième degré, on définit (comme paramètre) des probabilités d'inclusion.

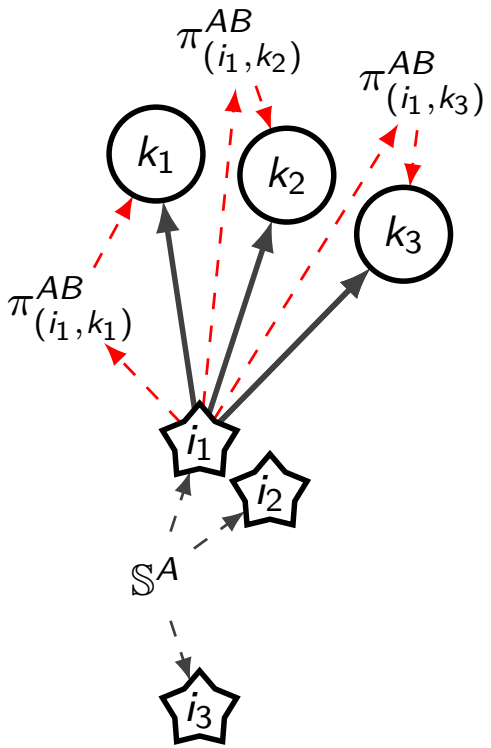
d'ordre 1... Probabilité que k soit sélectionné depuis i : $\pi^{AB}_{(i,k)}$



Échantillon au second degré (Préparation)

Pour préparer le deuxième degré, on définit (comme paramètre) des probabilités d'inclusion.

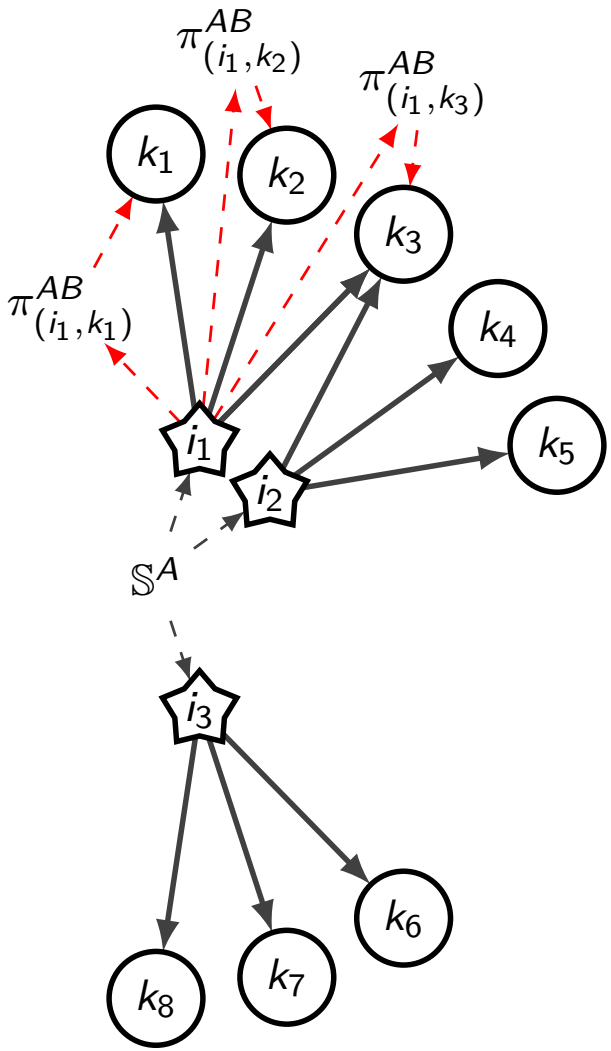
d'ordre 1... Probabilité que k soit sélectionné depuis i : $\pi^{AB}_{(i,k)}$



Échantillon au second degré (Préparation)

Pour préparer le deuxième degré, on définit (comme paramètre) des probabilités d'inclusion.

d'ordre 1... Probabilité que k soit sélectionné depuis i : $\pi^{AB}_{(i,k)}$

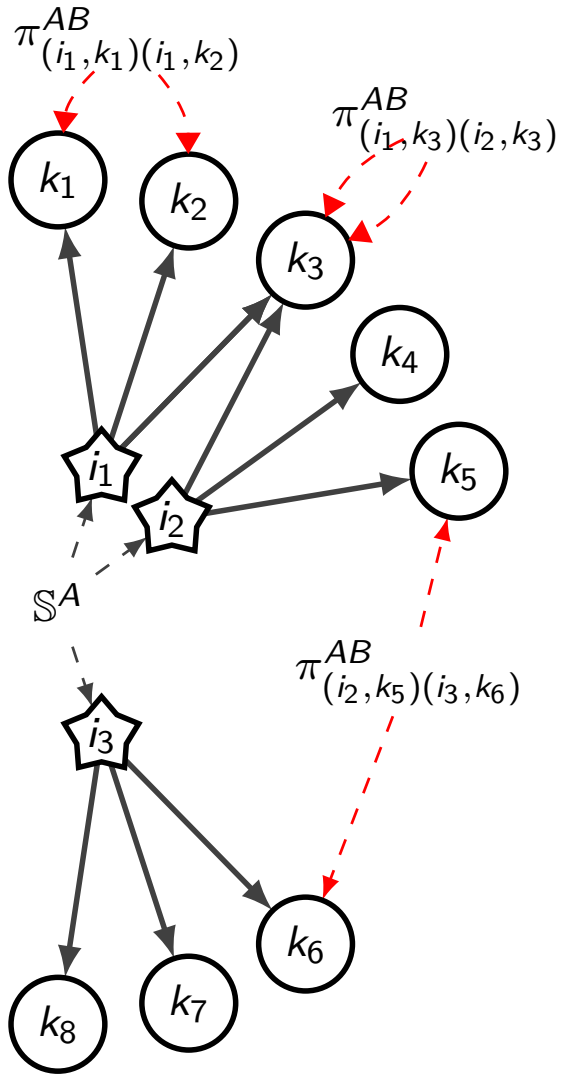


Échantillon au second degré (Préparation)

Pour préparer le deuxième degré, on définit (comme paramètre) des probabilités d'inclusion.

d'ordre 1... Probabilité que k soit sélectionné depuis i : $\pi^{AB}_{(i,k)}$

d'ordre 2... Probabilité que k soit tiré depuis i et l depuis j : $\pi^{AB}_{(i,k)(j,l)}$.



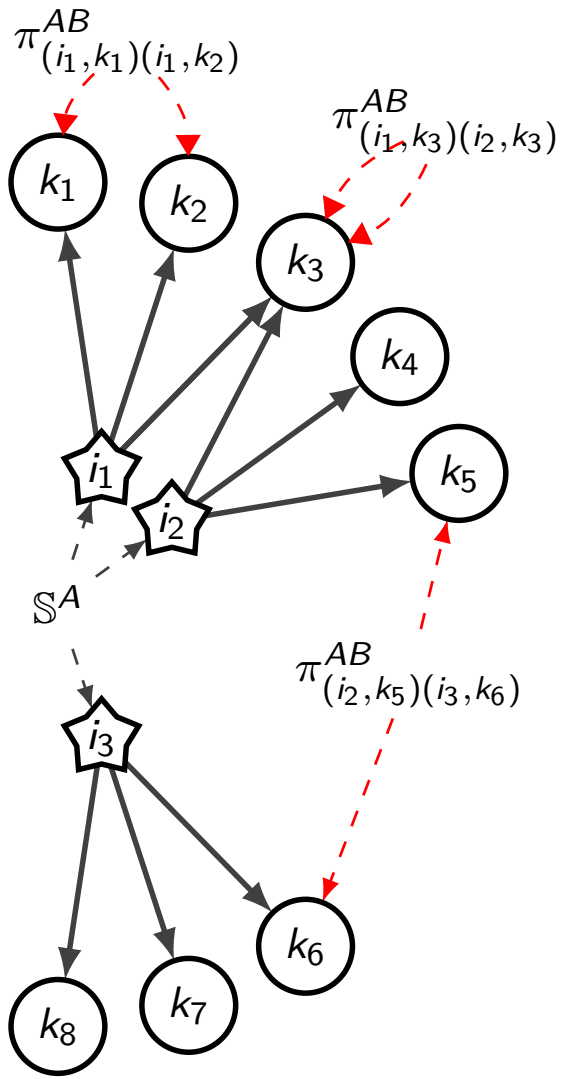
Échantillon au second degré (Préparation)

Pour préparer le deuxième degré, on définit (comme paramètre) des probabilités d'inclusion.

d'ordre 1... Probabilité que k soit sélectionné depuis i : $\pi^{AB}_{(i,k)}$

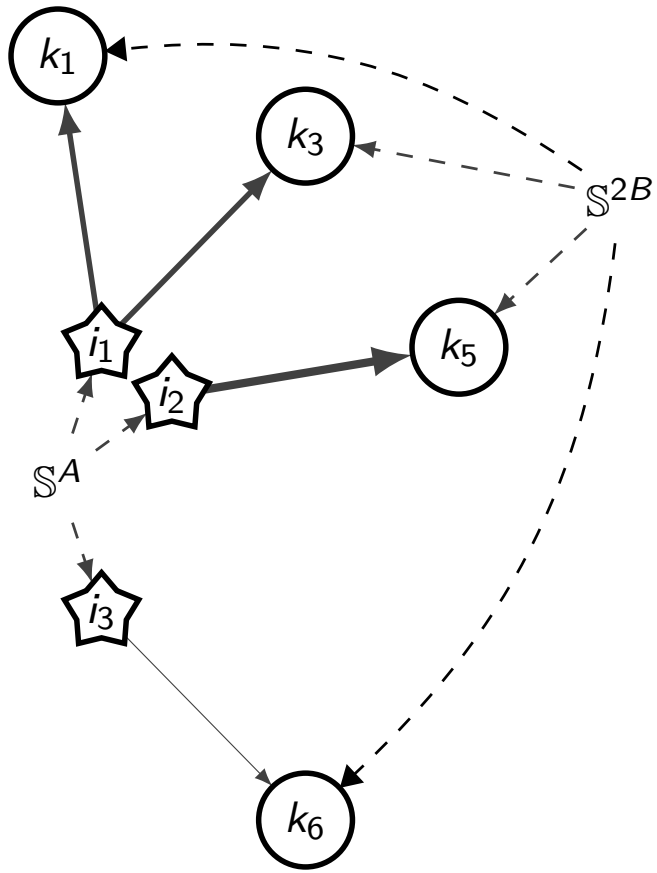
d'ordre 2... Probabilité que k soit tiré depuis i et l depuis j : $\pi^{AB}_{(i,k)(j,l)}$.

Ce faisant on définit un plan de sondage $\mathcal{P}^{AB}$ sur $U^A \times U^B \Rightarrow \mathbb{S}^{AB}, \tilde{\Delta}^{AB}$ avec $\mathcal{P}^A$ indépendant de $\mathcal{P}^{AB}$.



Échantillon au second degré (sélection)

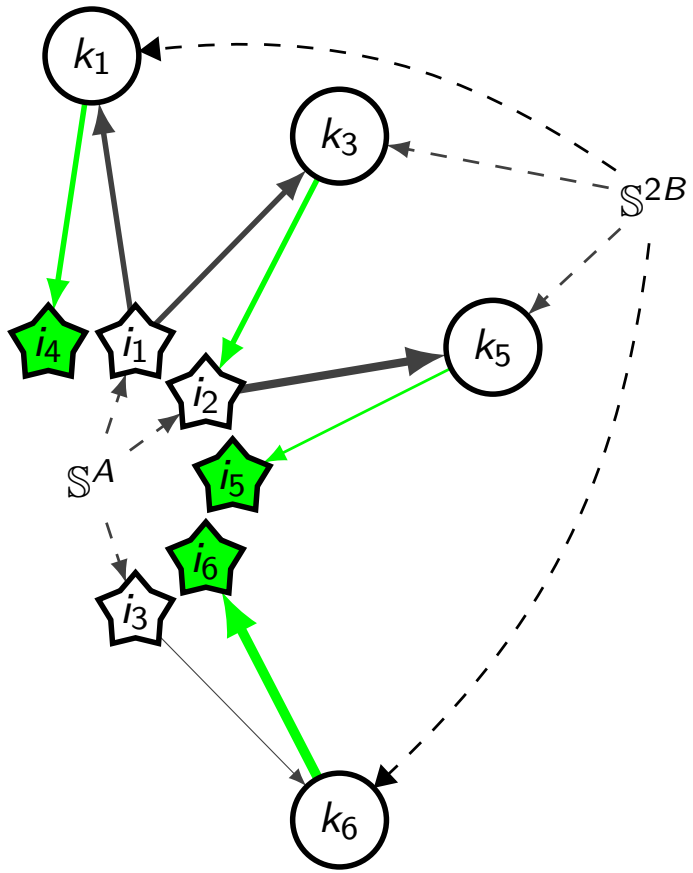
L'échantillon cible final du second degré est : $\mathbb{S}^{2B} = \{k_1, k_3, k_5, k_6\}$



Échantillon au second degré (sélection)

L'échantillon cible final du second degré est : $\mathbb{S}^{2B} = \{k_1, k_3, k_5, k_6\}$

Pour la suite des opérations, pour chaque unité cible sélectionnée, il faut connaître toutes les unités intermédiaires qui conduisent vers elle : $\{U_k^A\}$ pour $k \in \mathbb{S}^{2B}$



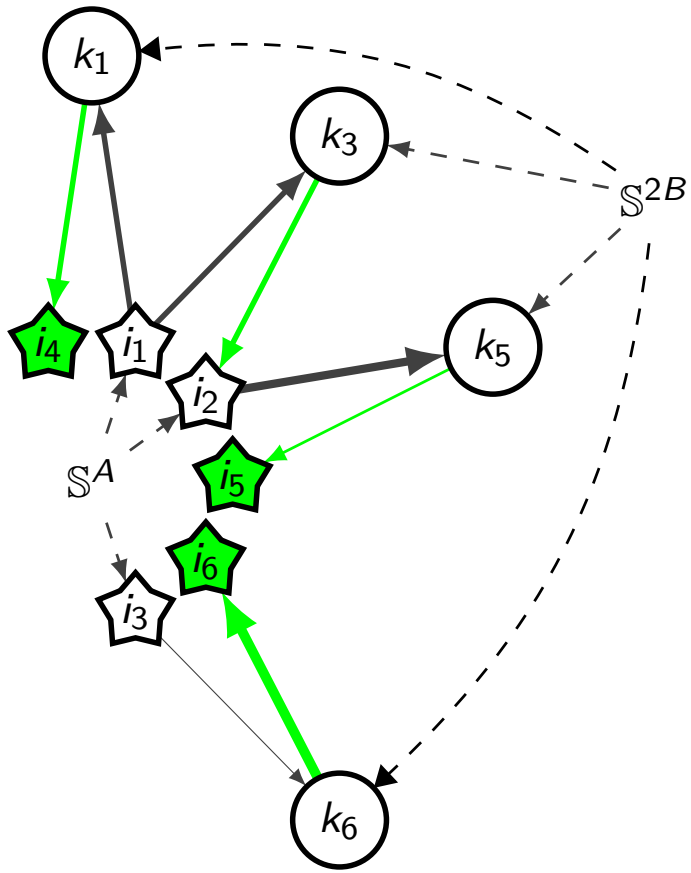
Échantillon au second degré (sélection)

L'échantillon cible final du second degré est : $\mathbb{S}^{2B} = \{k_1, k_3, k_5, k_6\}$

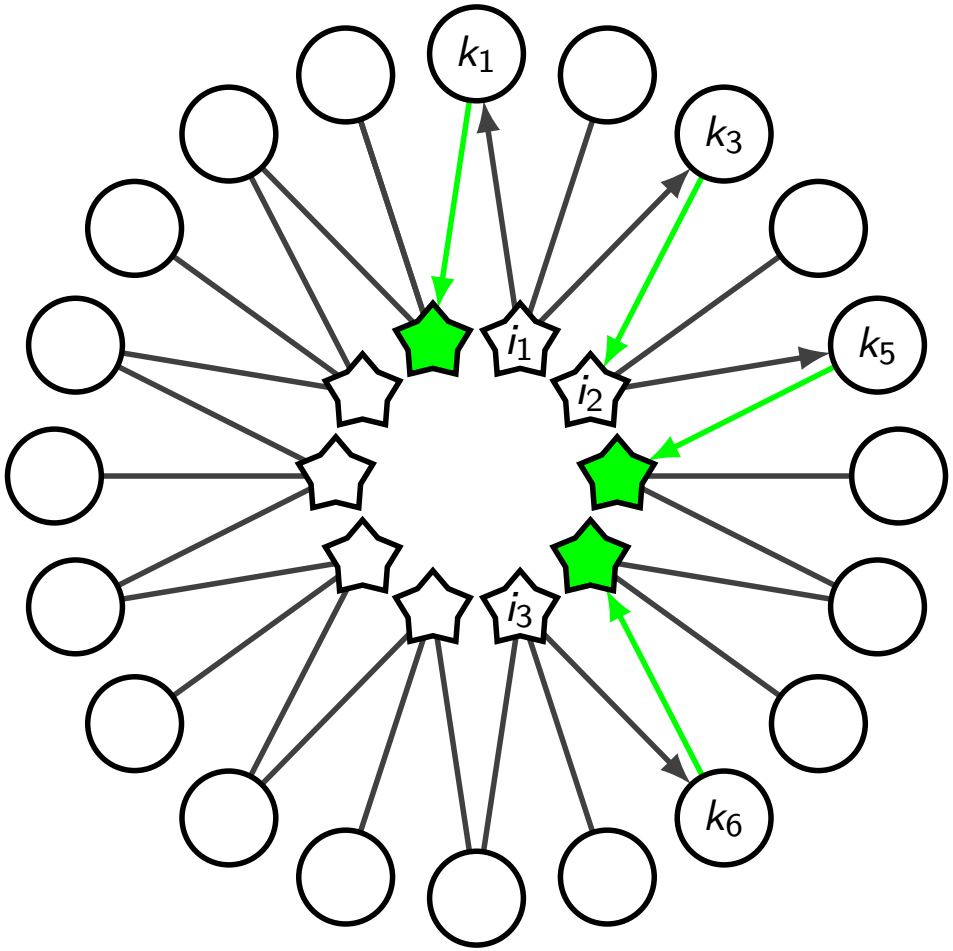
Pour la suite des opérations, pour chaque unité cible sélectionnée, il faut connaître toutes les unités intermédiaires qui conduisent vers elle : $\{U_k^A\}$ pour $k \in \mathbb{S}^{2B}$

Cette information minimale sera utile pour garantir l'absence de biais.

Dans la pratique, elle provient souvent de questions supplémentaires posées à cette attention.



Échantillon au second degré (sélection)



Dans certaines situations, toute l'information : L^{AB} est connue dès le début : coordination spatiale

.

Estimateur MGPP d'un total

- On souhaite estimer le total t_{y^B} d'une variable y^B sur U^B

Estimateur MGPP d'un total

- On souhaite estimer le total t_{y^B} d'une variable y^B sur U^B
- On ne peut pas utiliser $\hat{t}_{y^B}^{HT}$ car les π_k^B sont généralement inconnues

Estimateur MGPP d'un total

- On souhaite estimer le total t_{y^B} d'une variable y^B sur U^B
- On ne peut pas utiliser $\hat{t}_{y^B}^{HT}$ car les π_k^B sont généralement inconnues
- On utilise l'estimateur MGPP défini par

$$\hat{t}_{y^B}(L^{AB}, \Theta^{AB}, \mathbb{S}^A, \mathbb{S}^{AB}) = \sum_{k \in \mathbb{S}^{2B}} \left[\sum_{i \in U_k^A} \frac{\theta_{ik}^{AB} 1(i \in \mathbb{S}^A) 1((i,k) \in \mathbb{S}^{AB})}{\pi_i^A \pi_{ik}^{AB}} \right] y_k^B$$

Estimateur MGPP d'un total

- On souhaite estimer le total t_{y^B} d'une variable y^B sur U^B
- On ne peut pas utiliser $\hat{t}_{y^B}^{HT}$ car les π_k^B sont généralement inconnues
- On utilise l'estimateur MGPP défini par
$\hat{t}_{y^B}(L^{AB}, \Theta^{AB}, \mathbb{S}^A, \mathbb{S}^{AB}) = \sum_{k \in \mathbb{S}^{2B}} \left[\sum_{i \in U_k^A} \frac{\theta_{ik}^{AB} 1(i \in \mathbb{S}^A) 1((i,k) \in \mathbb{S}^{AB})}{\pi_i^A \pi_{ik}^{AB}} \right] y_k^B$
- Cet estimateur est sans biais si la somme des termes de chaque colonne de $L^{AB} \odot \Theta^{AB}$ vaut 1.

Précision de l'estimateur MGPP

On définit la vectorisation d'une matrice comme suit :

$$M = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \Rightarrow vect(M) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Précision de l'estimateur MGPP

On définit la vectorisation d'une matrice comme suit :

$$M = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \Rightarrow vect(M) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

On montre que

$$var(\hat{t}_{y^B}) = vect(\Theta^{AB})^{\intercal} \mathcal{Q}(L^{AB}, y^B y^{B\intercal}, \tilde{\Delta}^A, \tilde{\Delta}^{AB}) vect(\Theta^{AB})$$

où $\mathcal{Q}$ est une matrice de taille $N_A N_B \times N_A N_B$.

Précision de l'estimateur MGPP

On définit la vectorisation d'une matrice comme suit :

$$M = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \Rightarrow vect(M) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

On montre que

$$var(\hat{t}_{yB}) = vect(\Theta^{AB})^{\intercal} \mathcal{Q}(L^{AB}, y^{B} y^{B\intercal}, \tilde{\Delta}^{A}, \tilde{\Delta}^{AB}) vect(\Theta^{AB})$$

où $\mathcal{Q}$ est une matrice de taille $N_A N_B \times N_A N_B$.
$\Rightarrow$ On cherche à minimiser $var(\hat{t}_{yB})$ en Θ^{AB}, $\tilde{\Delta}^{AB}$ et $\tilde{\Delta}^{A}$

Précision : Minimiser la variance MGPP en Θ^{AB}

Quelques propriétés de $\mathcal{Q}$

- $\mathcal{Q}$ est symétrique.

Précision : Minimiser la variance MGPP en Θ^{AB}

Quelques propriétés de $\mathcal{Q}$

- $\mathcal{Q}$ est symétrique.
- Pour Q variables auxiliaires cibles $X^B = x_1^B, \cdots, x_q^B, \cdots, x_Q^B$

$$\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B}) = vect(\Theta^{AB})^\intercal \mathcal{Q}(L^{AB}, \sum_{q=1}^{Q} \alpha_q^B x_q^B x_q^{B\intercal}, \tilde{\Delta}^A, \tilde{\Delta}^{AB}) vect(\Theta^{AB})$$

Précision : Minimiser la variance MGPP en Θ^{AB}

Quelques propriétés de $\mathcal{Q}$

- $\mathcal{Q}$ est symétrique.
- Pour Q variables auxiliaires cibles $X^B = x_1^B, \cdots, x_q^B, \cdots, x_Q^B$

$$\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B}) = vect(\Theta^{AB})^\intercal \mathcal{Q}(L^{AB}, \sum_{q=1}^{Q} \alpha_q^B x_q^B x_q^{B\intercal}, \tilde{\Delta}^A, \tilde{\Delta}^{AB}) vect(\Theta^{AB})$$

- Θ^{AB} est à l'extérieur...

Précision : Minimiser la variance MGPP en Θ^{AB}

Quelques propriétés de $\mathcal{Q}$

- $\mathcal{Q}$ est symétrique.
- Pour Q variables auxiliaires cibles $X^B = x_1^B, \cdots, x_q^B, \cdots, x_Q^B$

$$\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B}) = vect(\Theta^{AB})^\intercal \mathcal{Q}(L^{AB}, \sum_{q=1}^{Q} \alpha_q^B x_q^B x_q^{B\intercal}, \tilde{\Delta}^A, \tilde{\Delta}^{AB}) vect(\Theta^{AB})$$

- Θ^{AB} est à l'extérieur...

$\Rightarrow$ Minimiser $\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_Q^B})$ en Θ^{AB}, à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés, sous contrainte d'absence de biais est un problème classique d'optimisation

Précision : Minimiser la variance MGPP en Θ^{AB}

Quelques propriétés de $\mathcal{Q}$

- $\mathcal{Q}$ est symétrique.
- Pour Q variables auxiliaires cibles $X^B = x_1^B, \cdots, x_q^B, \cdots, x_Q^B$

$$\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B}) = vect(\Theta^{AB})^\intercal \mathcal{Q}(L^{AB}, \sum_{q=1}^{Q} \alpha_q^B x_q^B x_q^{B\intercal}, \tilde{\Delta}^A, \tilde{\Delta}^{AB}) vect(\Theta^{AB})$$

- Θ^{AB} est à l'extérieur...

$\Rightarrow$ Minimiser $\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_Q^B})$ en Θ^{AB}, à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés, sous contrainte d'absence de biais est un problème classique d'optimisation

$$vect(\Theta_{opt}^{AB}(X^B)) = \begin{pmatrix} \mathcal{Q}(L^{AB}, \sum_{q=1}^{Q} \alpha_q^B x_q^B x_q^{B\intercal}, \tilde{\Delta}^A, \tilde{\Delta}^{AB}) & E^{AB\intercal} \\ E^{AB} & 0_{(N_B \times N_B)} \end{pmatrix}^{-1} \begin{pmatrix} 0_{N_A N_B} \\ 1_{N_B} \end{pmatrix}$$

Minimiser la variance MGPP en $\tilde{\Delta}^{AB}$

- On définit $\mathcal{P}_0^{AB}$ comme suit :
  - Pour chaque i, on sélectionne une et une seule unité cible dans U_i^B
  - La sélection pour i est indépendante de celle pour j

  $\Rightarrow$ $\tilde{\Delta}_0^{AB}$ est fonction des seuls $\pi_{(i,k)}^{AB}$

Minimiser la variance MGPP en $\tilde{\Delta}^{AB}$

- On définit $\mathcal{P}_0^{AB}$ comme suit :
  - Pour chaque i, on sélectionne une et une seule unité cible dans U_i^B
  - La sélection pour i est indépendante de celle pour j

  $\Rightarrow$ $\tilde{\Delta}_0^{AB}$ est fonction des seuls $\pi_{(i,k)}^{AB}$

$\Rightarrow$ Les $\pi_{(i,k)}^{AB}$ qui minimisent $\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ sont telles que :

$$\pi_{(i,k)}^{AB,opt} = \frac{|\theta_{ik}^{AB}| \sqrt{\sum_{q=1}^{Q} \alpha_q^B \left(x_{q,k}^B\right)^2}}{\sum_{l \in U_i^B} |\theta_{il}^{AB}| \sqrt{\sum_{q=1}^{Q} \alpha_q^B \left(x_{q,l}^B\right)^2}} 1(k \in U_i^B).$$

Minimiser la variance MGPP en $\tilde{\Delta}^{AB}$

- On définit $\mathcal{P}_0^{AB}$ comme suit :
  - Pour chaque i, on sélectionne une et une seule unité cible dans U_i^B
  - La sélection pour i est indépendante de celle pour j

  $\Rightarrow$ $\tilde{\Delta}_0^{AB}$ est fonction des seuls $\pi_{(i,k)}^{AB}$

$\Rightarrow$ Les $\pi_{(i,k)}^{AB}$ qui minimisent $\sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ sont telles que :

$$\pi_{(i,k)}^{AB,opt} = \frac{|\theta_{ik}^{AB}| \sqrt{\sum_{q=1}^{Q} \alpha_q^B \left(x_{q,k}^B\right)^2}}{\sum_{l \in U_i^B} |\theta_{il}^{AB}| \sqrt{\sum_{q=1}^{Q} \alpha_q^B \left(x_{q,l}^B\right)^2}} 1(k \in U_i^B).$$

$\Rightarrow$ On notera que $\mathcal{P}_0^{AB}$ est déterminantal, de noyau K_0^{AB}...

Sondage indirect et plans déterminantaux

$\mathcal{P}^A$ est déterminantal de noyau K^A de diagonale fixée Π^A et $\mathcal{P}^{AB}$ est déterminantal de noyau K^{AB}

- Améliorer la variance MGPP en K^A
- Résolution du problème initial
- Calcul des probabilités d'inclusion cible (simples et doubles)

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$
- Quand L^{AB} **est connue**, on sait optimiser formellement en
  - Θ^{AB} à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$
- Quand $\underline{L^{AB}}$ **est connue**, on sait optimiser formellement en
  - Θ^{AB} à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés
    - $\text{⚠}\Theta^{AB}_{opt}$ dépend de $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$
- Quand L^{AB} **est connue**, on sait optimiser formellement en
  - Θ^{AB} à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés
    - ⚠Θ^{AB}_{opt} dépend de $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$
  - $\tilde{\Delta}^{AB}$, dans un cas particulier, à $\tilde{\Delta}^A$ et Θ^{AB} donnés

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$
- Quand L^{AB} **est connue**, on sait optimiser formellement en
  - Θ^{AB} à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés
    - $\triangle$ Θ^{AB}_{opt} dépend de $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$
  - $\tilde{\Delta}^{AB}$, dans un cas particulier, à $\tilde{\Delta}^A$ et Θ^{AB} donnés
- Si $\mathcal{P}^A$ et $\mathcal{P}^{AB}$ sont déterminantaux

  $\Rightarrow$ $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ sont calculables et donc Θ^{AB}_{opt}

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$
- Quand $\underline{L^{AB}}$ **est connue**, on sait optimiser formellement en
  - Θ^{AB} à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés
    - $\triangle\!\!!\,\Theta^{AB}_{opt}$ dépend de $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$
  - $\tilde{\Delta}^{AB}$, dans un cas particulier, à $\tilde{\Delta}^A$ et Θ^{AB} donnés
- Si $\mathcal{P}^A$ et $\mathcal{P}^{AB}$ sont déterminantaux
  - $\Rightarrow$ $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ sont calculables et donc Θ^{AB}_{opt}
  - $\Rightarrow$ On peut optimiser en K^A à $\tilde{\Delta}^{AB}$ et Θ^{AB} données

Minimiser la variance MGPP en K^A

- La variance MGPP dépend de Θ^{AB}, $\tilde{\Delta}^A$, $\tilde{\Delta}^{AB}$
- Quand $\underline{L^{AB}}$ **est connue**, on sait optimiser formellement en
  - Θ^{AB} à $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ donnés
    - $\triangle\!\!!\,\Theta^{AB}_{opt}$ dépend de $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$
  - $\tilde{\Delta}^{AB}$, dans un cas particulier, à $\tilde{\Delta}^A$ et Θ^{AB} donnés
- Si $\mathcal{P}^A$ et $\mathcal{P}^{AB}$ sont déterminantaux
  - $\Rightarrow$ $\tilde{\Delta}^A$ et $\tilde{\Delta}^{AB}$ sont calculables et donc Θ^{AB}_{opt}
  - $\Rightarrow$ On peut optimiser en K^A à $\tilde{\Delta}^{AB}$ et Θ^{AB} données

Minimiser la variance MGPP en K^A

- $C_{\alpha^B}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ est fonction de K^A

Minimiser la variance MGPP en K^A

- $C_{\alpha^B}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ est fonction de K^A

  $\Rightarrow$ On peut optimiser $C_{\alpha^B}(X^B)$ en K^A à Θ^{AB} et $\tilde{\Delta}^{AB}$ donnés,

Minimiser la variance MGPP en K^A

- $C_{\alpha^B}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ est fonction de K^A

  $\Rightarrow$ On peut optimiser $C_{\alpha^B}(X^B)$ en K^A à Θ^{AB} et $\tilde{\Delta}^{AB}$ donnés,

- Sur la population intermédiaire, on dispose de P variables auxiliaires $x_1^A, \cdots, x_p^A, \cdots, x_P^A$

Minimiser la variance MGPP en K^A

- $C_{\alpha^B}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ est fonction de K^A

  $\Rightarrow$ On peut optimiser $C_{\alpha^B}(X^B)$ en K^A à Θ^{AB} et $\tilde{\Delta}^{AB}$ donnés,

- Sur la population intermédiaire, on dispose de P variables auxiliaires $x_1^A, \cdots, x_p^A, \cdots, x_P^A$

- $C_{\alpha^A}(X^A) = \sum_{q=1}^{Q} \alpha_q^A var(\hat{t}_{x_p^A}^{HT})$ est aussi fonction de K^A

Minimiser la variance MGPP en K^A

- $C_{\alpha^B}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ est fonction de K^A

  $\Rightarrow$ On peut optimiser $C_{\alpha^B}(X^B)$ en K^A à Θ^{AB} et $\tilde{\Delta}^{AB}$ donnés,

- Sur la population intermédiaire, on dispose de P variables auxiliaires $x_1^A, \cdots, x_p^A, \cdots, x_P^A$

- $C_{\alpha^A}(X^A) = \sum_{q=1}^{Q} \alpha_q^A var(\hat{t}_{x_p^A}^{HT})$ est aussi fonction de K^A

$\Rightarrow$ On peut optimiser $C_{\alpha^A}(X^A)$ en K_A

Minimiser la variance MGPP en K^A

- $C_{\alpha^B}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B})$ est fonction de K^A

  $\Rightarrow$ On peut optimiser $C_{\alpha^B}(X^B)$ en K^A à Θ^{AB} et $\tilde{\Delta}^{AB}$ donnés,

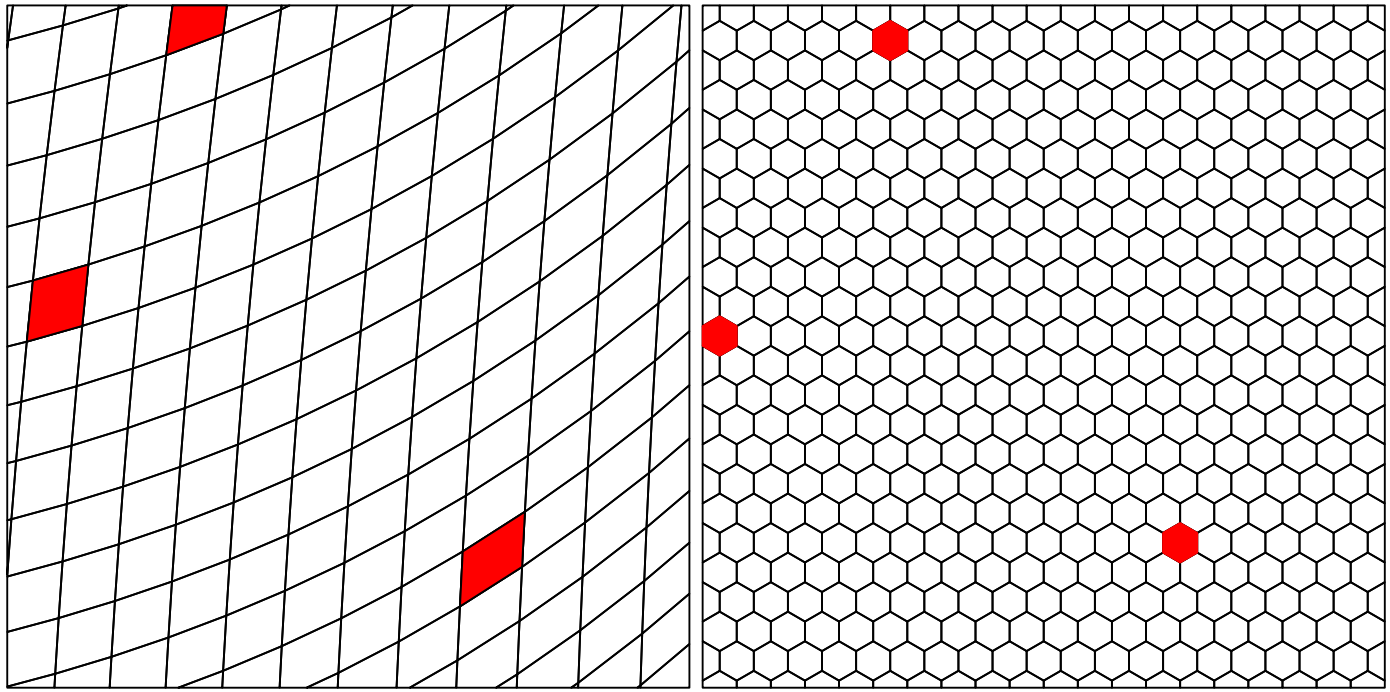
- Sur la population intermédiaire, on dispose de P variables auxiliaires $x_1^A, \cdots, x_p^A, \cdots, x_P^A$

- $C_{\alpha^A}(X^A) = \sum_{q=1}^{Q} \alpha_q^A var(\hat{t}_{x_p^A}^{HT})$ est aussi fonction de K^A

$\Rightarrow$ On peut optimiser $C_{\alpha^A}(X^A)$ en K_A

$\Rightarrow$ On peut optimiser $C_{\alpha^A}(X^A) + C_{\alpha^B}(X^B)$ en K^A à Θ^{AB} et $\tilde{\Delta}^{AB}$ données

Résolution du problème initial (Rappel)



(a) $\mathbb{S}^A$ **équilibré sur** X^A (b) $\mathbb{S}^B$ **équilibré sur** X^B

Figure: On souhaite sélectionner deux échantillons $\mathbb{S}^A$ et $\mathbb{S}^B$, **de même taille** (ou presque), **une unité de** $\mathbb{S}^A$ **est proche d'une unité de** $\mathbb{S}^B$ et réciproquement

Résolution du problème initial

On se donne une matrice de lien L^{AB}

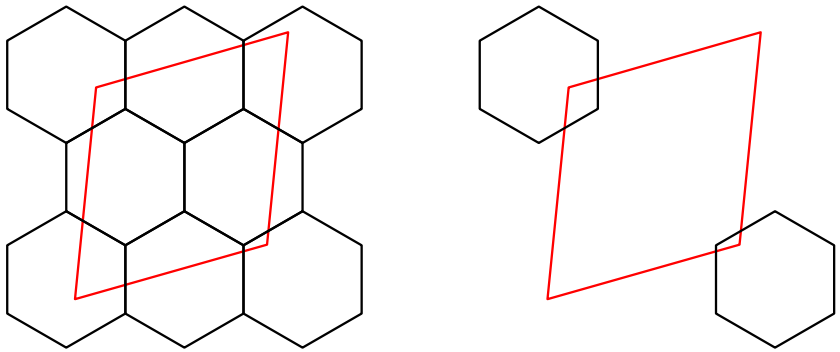
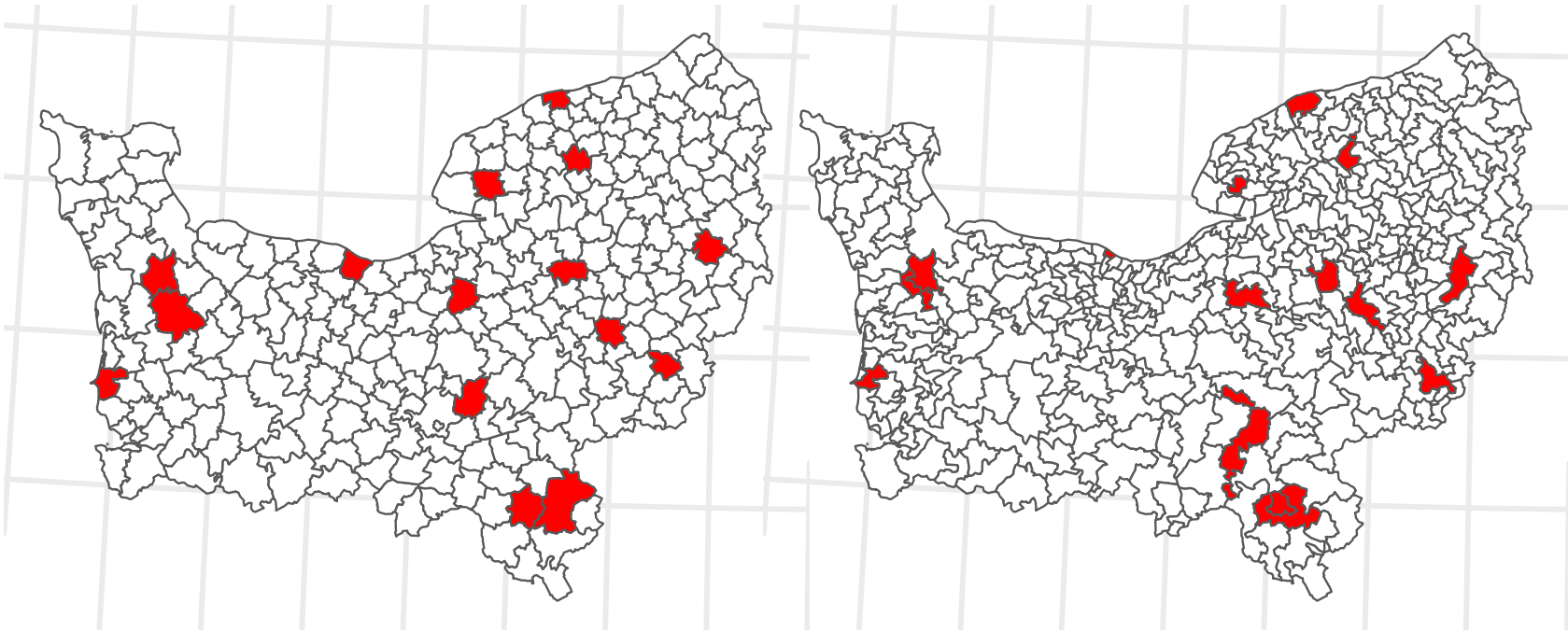


Figure: Construction de L^{AB}

Résolution du problème initial sur la Normandie



(a) U^A : $N^A = 250$, $n^A = 15$, $\Pi^A\sharp$ Nb hab

(b) U^B : $N^B = 337$, $n^B \simeq 15$

Figure: $P = Q = 3$, X^A, X^B contiennent le nombre de 55 ans et plus, le montant total des pensions et celui des salaires

Résolution du problème initial sur la Normandie

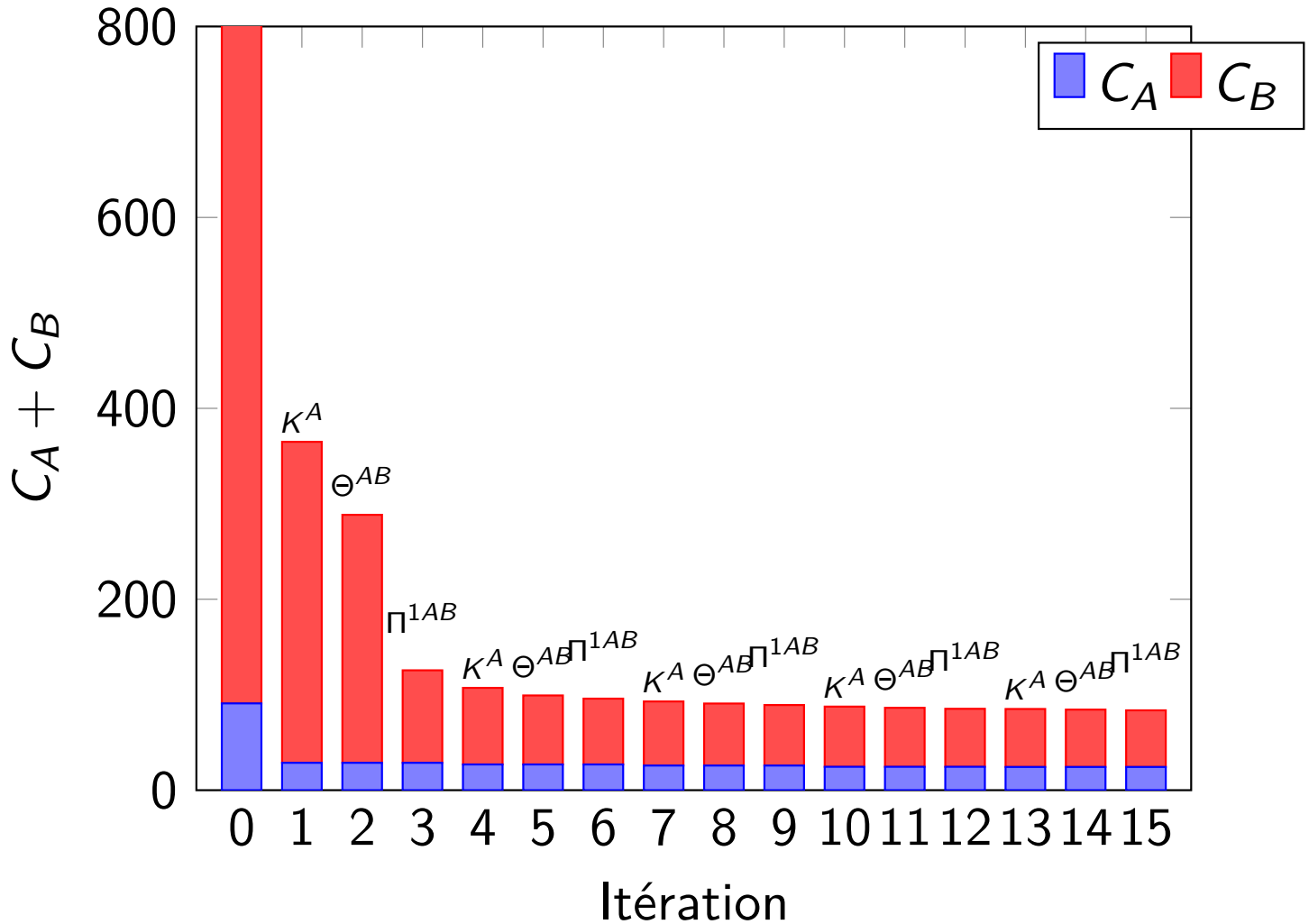


Figure: Optimisation séquentielle de $C_{\alpha^A}(X^A) + C_{\alpha^B}(X^B)$ en K^A, Θ^{AB}, Π^{1AB} à partir d'une situation initiale K_0^A, Θ_0^{AB}, Π_0^{1AB}

résolution du problème initial sur la Normandie

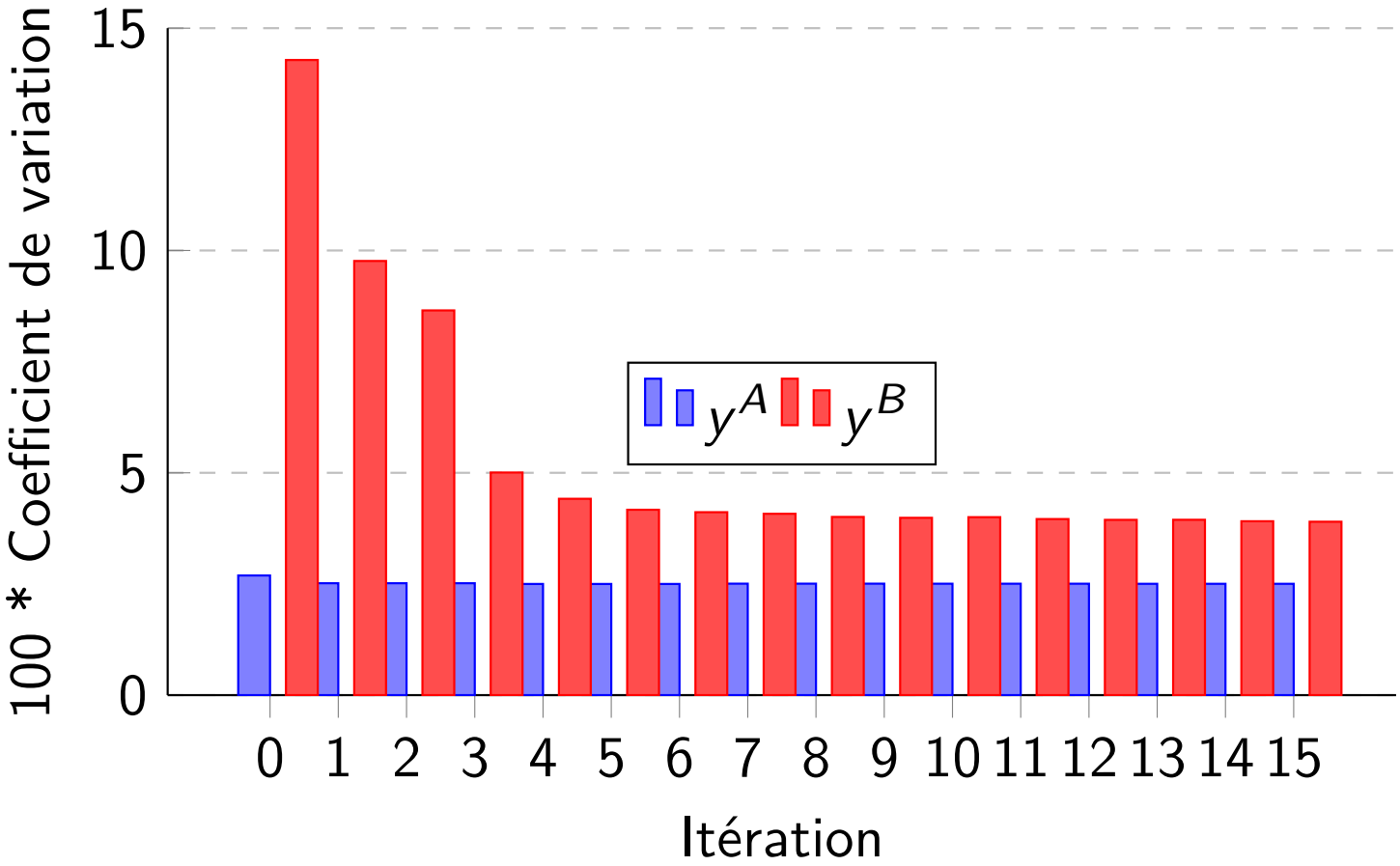


Figure: Coefficient de variation estimateurs pour deux variables d'intérêt y^A (HT), y^B (MGPP) : total des allocations chômage

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.
- Cela revient à : $\Pr\left(\exists\, i \in U_k^A,\ i \in \mathbb{S}^A\right)$.

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.
- Cela revient à : $\Pr\left(\exists\, i \in U_k^A,\ i \in \mathbb{S}^A\right)$.
- Ce qui équivaut à : $1 - \Pr\left(\forall\, i \in U_k^A,\ i \in \mathbb{S}^{A^c}\right)$.

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.
- Cela revient à : $\Pr\left(\exists\, i \in U_k^A,\ i \in \mathbb{S}^A\right)$.
- Ce qui équivaut à : $1 - \Pr\left(\forall\, i \in U_k^A,\ i \in \mathbb{S}^{A^c}\right)$.
- Donc : $1 - \Pr\left(U_k^A \subseteq \mathbb{S}^{A^c}\right)$.

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.
- Cela revient à : $\Pr\left(\exists\, i \in U_k^A,\ i \in \mathbb{S}^A\right)$.
- Ce qui équivaut à : $1 - \Pr\left(\forall\, i \in U_k^A,\ i \in \mathbb{S}^{A^c}\right)$.
- Donc : $1 - \Pr\left(U_k^A \subseteq \mathbb{S}^{A^c}\right)$.

$\Rightarrow$ Il faut donc connaître les probabilités d'inclusion **à tous les ordres** pour l'échantillon complémentaire $\mathbb{S}^{A^c}$.

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.
- Cela revient à : $\Pr\left(\exists\, i \in U_k^A,\ i \in \mathbb{S}^A\right)$.
- Ce qui équivaut à : $1 - \Pr\left(\forall\, i \in U_k^A,\ i \in \mathbb{S}^{A^c}\right)$.
- Donc : $1 - \Pr\left(U_k^A \subseteq \mathbb{S}^{A^c}\right)$.

$\Rightarrow$ Il faut donc connaître les probabilités d'inclusion **à tous les ordres** pour l'échantillon complémentaire $\mathbb{S}^{A^c}$.

- C'est faisable, par exemple, si le plan $\mathcal{P}^A$ est stratifié (Fizzala et Bodet, 2022),

Probabilités d'inclusion cibles

Dans ce qui suit, on se limite au cas d'un tirage en une seule phase.

- ⚠L'échantillon indirect $\mathbb{S}^{1B}$ n'est pas déterminantal, en général !!
- Pour une unité k, on cherche à calculer : $\Pr\left(\{k\} \subseteq \mathbb{S}^{1B}\right)$.
- Cela revient à : $\Pr\left(\exists\, i \in U_k^A,\ i \in \mathbb{S}^A\right)$.
- Ce qui équivaut à : $1 - \Pr\left(\forall\, i \in U_k^A,\ i \in \mathbb{S}^{A^c}\right)$.
- Donc : $1 - \Pr\left(U_k^A \subseteq \mathbb{S}^{A^c}\right)$.

$\Rightarrow$ Il faut donc connaître les probabilités d'inclusion **à tous les ordres** pour l'échantillon complémentaire $\mathbb{S}^{A^c}$.

- C'est faisable, par exemple, si le plan $\mathcal{P}^A$ est stratifié (Fizzala et Bodet, 2022),
- ou si $\mathcal{P}^A$ est un plan déterminantal.

Probabilités d'inclusion cibles

Dans le cas d'une phase, on montre que :

$$\pi_k^{1B} = \Pr(\{k\} \subseteq \mathbb{S}^{1B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right)$$

Probabilités d'inclusion cibles

Dans le cas d'une phase, on montre que :

$$\pi_k^{1B} = \Pr(\{k\} \subseteq \mathbb{S}^{1B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right)$$

$$\begin{aligned}\pi_{k\ell}^{1B} &= \Pr(\{k,\ell\} \subseteq \mathbb{S}^{1B}) \\ &= 1 + \det\left(I_{N_{k\ell}^A} - K^A_{|U_k^A \cup U_\ell^A}\right) \\ &\quad - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right) - \det\left(I_{N_\ell^A} - K^A_{|U_\ell^A}\right).\end{aligned}$$

Probabilités d'inclusion cibles

Dans le cas d'une phase, on montre que :

$$\pi_k^{1B} = \Pr(\{k\} \subseteq \mathbb{S}^{1B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right)$$

$$\begin{aligned}\pi_{k\ell}^{1B} &= \Pr(\{k,\ell\} \subseteq \mathbb{S}^{1B}) \\ &= 1 + \det\left(I_{N_{k\ell}^A} - K^A_{|U_k^A \cup U_\ell^A}\right) \\ &\quad - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right) - \det\left(I_{N_\ell^A} - K^A_{|U_\ell^A}\right).\end{aligned}$$

Dans le cas de deux phases on montre que :

$$\pi_k^{2B} = \Pr(\{k\} \subseteq \mathbb{S}^{2B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A} K^{AB}_{|(U_k^A,k)}\right),$$

Probabilités d'inclusion cibles

Dans le cas d'une phase, on montre que :

$$\pi_k^{1B} = \Pr(\{k\} \subseteq \mathbb{S}^{1B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right)$$

$$\begin{aligned}\pi_{k\ell}^{1B} &= \Pr(\{k,\ell\} \subseteq \mathbb{S}^{1B}) \\ &= 1 + \det\left(I_{N_{k\ell}^A} - K^A_{|U_k^A \cup U_\ell^A}\right) \\ &\quad - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right) - \det\left(I_{N_\ell^A} - K^A_{|U_\ell^A}\right).\end{aligned}$$

Dans le cas de deux phases on montre que :

$$\pi_k^{2B} = \Pr(\{k\} \subseteq \mathbb{S}^{2B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A} K^{AB}_{|(U_k^A,k)}\right),$$

Si $\mathcal{P}^{AB} = \mathcal{P}_0^{AB}$

Probabilités d'inclusion cibles

Dans le cas d'une phase, on montre que :

$$\pi_k^{1B} = \Pr(\{k\} \subseteq \mathbb{S}^{1B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right)$$

$$\begin{aligned}\pi_{k\ell}^{1B} &= \Pr(\{k,\ell\} \subseteq \mathbb{S}^{1B}) \\ &= 1 + \det\left(I_{N_{k\ell}^A} - K^A_{|U_k^A \cup U_\ell^A}\right) \\ &\quad - \det\left(I_{N_k^A} - K^A_{|U_k^A}\right) - \det\left(I_{N_\ell^A} - K^A_{|U_\ell^A}\right).\end{aligned}$$

Dans le cas de deux phases on montre que :

$$\pi_k^{2B} = \Pr(\{k\} \subseteq \mathbb{S}^{2B}) = 1 - \det\left(I_{N_k^A} - K^A_{|U_k^A} K^{AB}_{|(U_k^A,k)}\right),$$

Si $\mathcal{P}^{AB} = \mathcal{P}_0^{AB}$

$$\begin{aligned}\pi_{k\ell}^{2B} &= \Pr(\{k,\ell\} \subseteq \mathbb{S}^{2B}) \\ &= 1 + \det\left(I_{N_{k\ell}^A} - K^A_{|U_k^A \cup U_\ell^A} \cdot \left(K^{AB}_{0|(U_k^A \cup U_\ell^A,k)} + K^{AB}_{0|(U_k^A \cup U_\ell^A,\ell)}\right)\right) \\ &\quad - \det\left(I_{N_k^A} - K^A_{|U_k^A} K^{AB}_{0|(U_k^A,k)}\right) - \det\left(I_{N_\ell^A} - K^A_{|U_\ell^A} K^{AB}_{0|(U_\ell^A,\ell)}\right).\end{aligned}$$

Probabilités d'inclusion cibles

- On peut calculer π_k^B et π_{kl}^B dans tous les cas, pas seulement quand L^{AB} est entièrement connue

Probabilités d'inclusion cibles

- On peut calculer π_k^B et π_{kl}^B dans tous les cas, pas seulement quand L^{AB} est entièrement connue
- On peut utiliser Horvitz-Thompson à la place de la MGPP

Probabilités d'inclusion cibles

- On peut calculer π_k^B et π_{kl}^B dans tous les cas, pas seulement quand L^{AB} est entièrement connue
- On peut utiliser Horvitz-Thompson à la place de la MGPP
- π_k^B et π_{kl}^B dépendent de K^A et K_0^{AB}, si L^{AB} est connue

$\Rightarrow$ On peut chercher K^A et K_0^{AB} telles que $\pi_k^B \simeq \Pi_k^B$ fixé a priori (Non fait, encore ?)

Probabilités d'inclusion cibles

- On peut calculer π_k^B et π_{kl}^B dans tous les cas, pas seulement quand L^{AB} est entièrement connue
- On peut utiliser Horvitz-Thompson à la place de la MGPP
- π_k^B et π_{kl}^B dépendent de K^A et K_0^{AB}, si L^{AB} est connue

$\Rightarrow$ On peut chercher K^A et K_0^{AB} telles que $\pi_k^B \simeq \Pi_k^B$ fixé a priori (Non fait, encore ?)

$\Rightarrow$ On peut chercher K^A et K_0^{AB} qui minimisent $C_{\alpha^B}^{HT}(X^B) = \sum_{q=1}^{Q} \alpha_q^B var(\hat{t}_{x_q^B}^{HT})$ (Non fait, encore ?)

Conclusions

- Nous avons fourni des formules explicites pour Θ^{AB}_{opt} et $\pi^{AB,\text{opt}}_{(i,k)}$ (dans un cas particulier)

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A
  - obtenir les probabilités cibles d'inclusion simples et doubles

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A
  - obtenir les probabilités cibles d'inclusion simples et doubles
- **Perspectives** :
  - Optimiser conjointement K^A et K^{AB} pour approcher $\pi_k^B \simeq \Pi_k^B$

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A
  - obtenir les probabilités cibles d'inclusion simples et doubles
- **Perspectives** :
  - Optimiser conjointement K^A et K^{AB} pour approcher $\pi_k^B \simeq \Pi_k^B$
  - Minimiser, via K^A et K^{AB}, la variance cible de l'estimateur de Horvitz–Thompson

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A
  - obtenir les probabilités cibles d'inclusion simples et doubles
- **Perspectives** :
  - Optimiser conjointement K^A et K^{AB} pour approcher $\pi_k^B \simeq \Pi_k^B$
  - Minimiser, via K^A et K^{AB}, la variance cible de l'estimateur de Horvitz–Thompson
  - Trouver K^A et K^{AB} tels que $n^A = n^B$ (plutôt que $n^A \simeq n^B$)

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A
  - obtenir les probabilités cibles d'inclusion simples et doubles
- **Perspectives** :
  - Optimiser conjointement K^A et K^{AB} pour approcher $\pi_k^B \simeq \Pi_k^B$
  - Minimiser, via K^A et K^{AB}, la variance cible de l'estimateur de Horvitz–Thompson
  - Trouver K^A et K^{AB} tels que $n^A = n^B$ (plutôt que $n^A \simeq n^B$)
  - Explorer des noyaux non symétriques...

Conclusions

- Nous avons fourni des formules explicites pour Θ_{opt}^{AB} et $\pi_{(i,k)}^{AB,\text{opt}}$ (dans un cas particulier)
- Les plans déterminantaux ont permis de :
  - calculer Θ_{opt}^{AB} et l'optimiser par rapport à K^A
  - obtenir les probabilités cibles d'inclusion simples et doubles
- **Perspectives** :
  - Optimiser conjointement K^A et K^{AB} pour approcher $\pi_k^B \simeq \Pi_k^B$
  - Minimiser, via K^A et K^{AB}, la variance cible de l'estimateur de Horvitz–Thompson
  - Trouver K^A et K^{AB} tels que $n^A = n^B$ (plutôt que $n^A \simeq n^B$)
  - Explorer des noyaux non symétriques...
  - d'autres problèmes...

Merci pour votre attention

Bibliographie I

Chevalier, M., Costa, L., Delta, L., Deroyon, T., Favre-Martinoz, C., Givois, S., Guillo, C., Merly-Alpa, T., Paliod, N., Pendoli, P.-A., Sauvaget, T. et Vincent, L. (2022). *Le renouvellement de l'échantillon-maître des enquêtes auprès des ménages et de l'échantillon de l'enquête Emploi de l'Insee*. Insee Méthodes. Insee. N°141, mai 2022.

Deville, J. et Lavallée, P. (2006). Indirect sampling: The foundations of the generalized weight share method. *Survey Methodology*, 32(2):165.

Fizzala, A. et Bodet, H. (2022). Utilisation des probabilités d'inclusion exactes pour le sondage indirect en population asymétrique. *Journées de Méthologie Statistique de l'Insee*.

Loonis, V. (2023). Constructing all determinantal sampling designs. *Survey Methodology*, 49(2).

Loonis, V. et Mary, X. (2019). Determinantal sampling designs. *Journal of Statistical Planning and Inference*, 199:60–88.