

# Méthode Généralisée de Partage des Poids

pour l'amélioration de la précision des estimations de trafic postal en France

Many-to-One indirect sampling with application to the French postal traffic estimation  
(AOAS, 2023)

Estelle Medous

Institut national de l'information géographique et forestière

A. Ruiz-Gazen (TSE-R)

C. Goga (LmB)

A. Dessertaine (La Poste)

P. Puech (La Poste)

J.-F. Beaumont (Statistique Canada)



laboratoire de mathématiques de besançon  
UNIVERSITÉ DE FRANCHE-COMTÉ • CNRS • UMR 5602



- Service historique de distribution postale en France
- Fluctuations du marché de la distribution du courrier  
⇒ La Poste souhaite connaître le trafic postal mensuel.
- Trafic total **inconnu** :
  - information disponible sur le courrier trié automatiquement  
⇒ pas suffisamment fiable
  - certains courriers ne peuvent pas être triés automatiquement⇒ Trafic total estimé par échantillonnage

# L'étude de La Poste

- Échantillons de tournées de facteur depuis 1994
- Tournées de facteur **directement échantillonnées** de 1994 à 2015
- **Sous-couverture** importante depuis 2015 (Lardin-Puech, 2014)  
⇒ **Échantillonnage direct impossible**

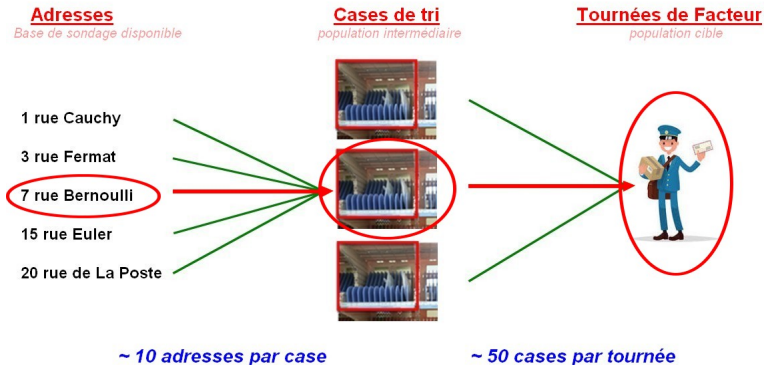
# L'étude de La Poste

- Idée : tirer **indirectement** les tournées à partir des adresses postales  
⇒ durée d'observation trop longue



# L'étude de La Poste

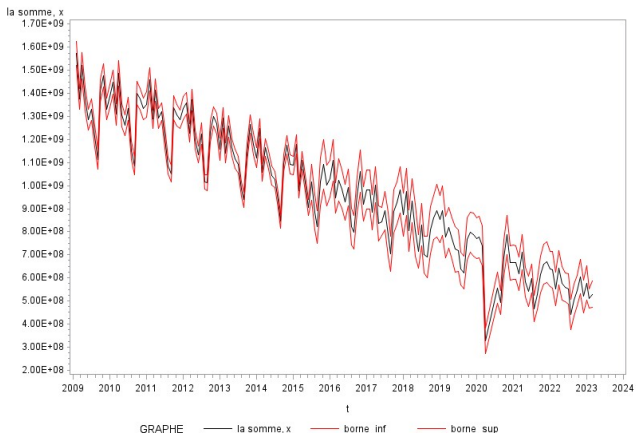
- Idée : **sondage indirect double** avec seconde population intermédiaire



# Perte de précision

- Le sondage indirect double a entraîné une **perte de précision**

Trafic Courrier - de Janvier 2009 à février 2023



# Sondage indirect

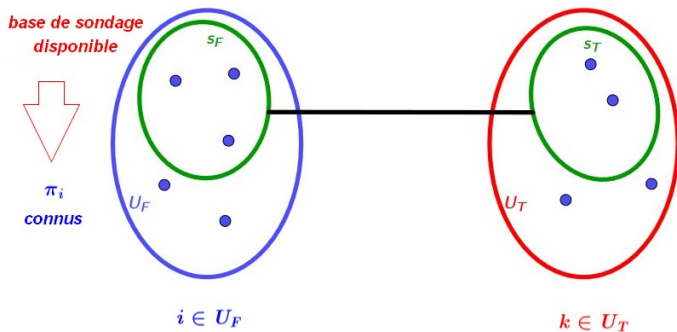
- Sondage indirect double **plus facile à réaliser** que le simple
- Une méthode **plus précise** que l'autre ?

- 1 Sondage indirect et MGPP
- 2 Sondage indirect double et MGPP double
- 3 Comparaison des deux méthodes
- 4 Simulations
- 5 Conclusion
- 6 References



# Sondage indirect

- Population  $U_F$  liée d'une manière ou d'une autre à  $U_T$
- Échantillon  $s_F$  dans  $U_F \Rightarrow$  échantillon  $s_T$  dans  $U_T$



- Tout individu de  $U_T$  lié à au moins un individu de  $U_F$

# Notations

- **Poids de lien** (poids) entre  $i \in U_F$  et  $k \in U_T$  :

$$\theta_{ik} = \begin{cases} > 0 & \text{si } i \text{ et } k \text{ liés,} \\ 0 & \text{sinon.} \end{cases}$$

- **Poids de lien standardisé**  $\tilde{\theta}_{ik}$  :

$$\sum_{i \in U_F} \tilde{\theta}_{ik} = 1 \text{ pour tout } k \in U_T$$

- Total  $t_Y$  sur  $U_T$  équivalent à un **total sur**  $U_F$  :

$$t_Y = \sum_{k \in U_T} y_k = \sum_{k \in U_T} y_k \sum_{i \in U_F} \tilde{\theta}_{ik} = \sum_{i \in U_F} \tilde{y}_i \text{ avec } \tilde{y}_i = \sum_{k \in U_T} y_k \tilde{\theta}_{ik}.$$

# Méthode Généralisée de Partage des Poids

- Méthode Généralisée de Partage des Poids (MGPP) :

(Kalton and Brick (1995), Deville and Lavallée (2006), Lavallée (2009))

$$\begin{aligned}\hat{t}_{Y, MGPP} &= \sum_{i \in \mathcal{S}_F} \tilde{y}_i / \pi_i \quad \text{avec} \quad \tilde{y}_i = \sum_{k \in U_T} y_k \tilde{\theta}_{ik} \\ &= \sum_{k \in \mathcal{S}_T} y_k \sum_{i \in \mathcal{S}_F} \tilde{\theta}_{ik} / \pi_i\end{aligned}$$

- Sans biais si et seulement si  $\tilde{\theta}_{ik}$  standardisés

# Variance des estimateurs MGPP

- Variance de l'estimateur MGPP :

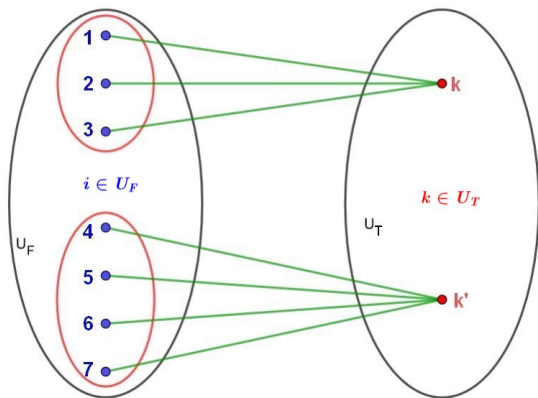
$$\text{Var}(\hat{t}_{Y, MGPP}) = \sum_{i \in U_F} \sum_{i' \in U_F} \Delta_{ii'} \sum_{k \in U_T} \tilde{\theta}_{ik} y_k \sum_{k' \in U_T} \tilde{\theta}_{i'k'} y_{k'}$$

- Poids de liens optimaux :

Poids qui minimisent  $\text{Var}(\hat{t}_{Y, MGPP})$  pour toutes variables d'intérêt

- Existence de poids optimaux ?  $\Rightarrow$  Oui !
- Deux conditions suffisantes (Medous et al., 2023)

# Tous-pour-Un (TpU)



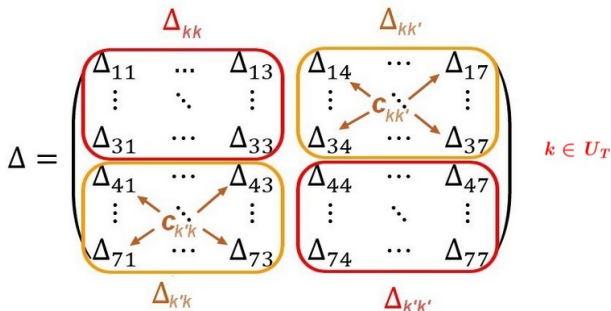
- Liens observés à La Poste

# $\Delta$ -propriété

$$\Delta = (\Delta_{ii'})_{i,i' \in U_F}, \quad \Delta_{ii'} = \frac{\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}}, \quad \Delta_{kk'} = (\Delta_{ii'})_i \text{ lié à } k, i' \text{ lié à } k', k, k' \in U_T$$

## Définition ( $\Delta$ -propriété)

- $\Delta_{kk} = \Delta_k$  *inversible* pour tout  $k \in U_T$ ,
- Pour  $k \neq k'$ ,  $\Delta_{kk'} = (c_{kk'})_i$  lié à  $k, i'$  lié à  $k'$



## $\Delta$ -propriété

$$\Delta = (\Delta_{ii'})_{i,i' \in U_F}, \Delta_{ii'} = \frac{\pi_{ii'} - \pi_i \pi_{i'}}{\pi_i \pi_{i'}}, \Delta_{kk'} = (\Delta_{ii'})_i \text{ lié à } k, i' \text{ lié à } k', k, k' \in U_T$$

### Définition ( $\Delta$ -propriété)

- $\Delta_{kk} = \Delta_k$  *inversible* pour tout  $k \in U_T$ ,
- Pour  $k \neq k'$ ,  $\Delta_{kk'} = (c_{kk'})_i$  lié à  $k, i'$  lié à  $k'$
- Sondage Aléatoire Simple Sans Remise (SASSR)

$$\Delta_i = \frac{N}{n} - 1, i \in U_F \text{ et } \Delta_{ii'} = \left(1 - \frac{N}{n}\right) \frac{1}{N-1}, i \neq i'$$

## Poids de lien optimaux

Soit  $N_{Fk}$  nombre de  $i \in U_F$  liés à  $k$  et  $\mathbf{1}_k$  vecteur de 1 de taille  $N_{Fk}$ .

Proposition (Medous et al. (2023))

Si les liens sont de type  $TpU$  et la  $\Delta$ -propriété est vérifiée,

$\Rightarrow$  Poids de lien optimaux *uniques* :

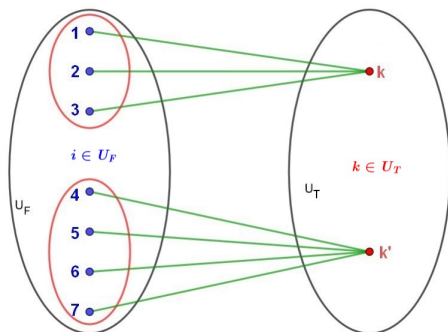
$$(\tilde{\theta}_{ik}^{opt})_{i \text{ lié à } k} = \Delta_k^{-1} \mathbf{1}_k (\mathbf{1}_k^t \Delta_k^{-1} \mathbf{1}_k)^{-1}, \text{ pour tout } k \in U_T.$$

- SASSR :

$$\tilde{\theta}_{ik}^{opt} = \frac{1}{N_{Fk}} \text{ si } i \text{ et } k \text{ liés, } 0 \text{ sinon}$$



# Problème du TpU

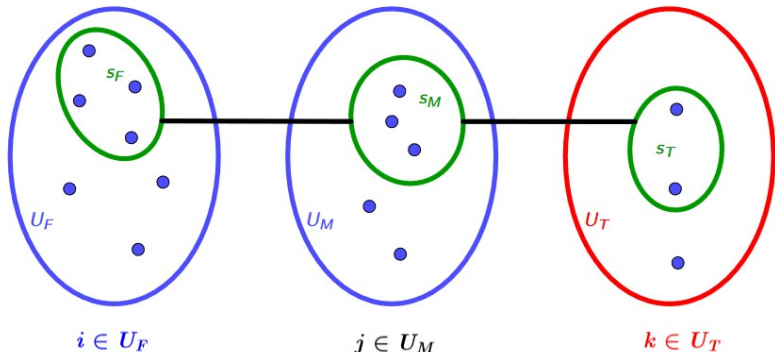


- Nombre élevé d'individus à identifier  
 ⇒ MGPP difficile à mettre en oeuvre avec des liens TpU

- 1 Sondage indirect et MGPP
- 2 Sondage indirect double et MGPP double**
- 3 Comparaison des deux méthodes
- 4 Simulations
- 5 Conclusion
- 6 References

# Sondage Indirect double

- Population  $U_M$  liée d'une manière ou d'une autre à  $U_F$  et  $U_T$
- Sondage Indirect double



# Notations

- **Poids FM** : poids de lien standardisé  $\tilde{\theta}_{ij}^{FM}$

$$\sum_{i \in U_F} \tilde{\theta}_{ij}^{FM} = 1 \text{ pour tout } j \in U_M.$$

- **Poids MT** : poids de lien standardisé  $\tilde{\theta}_{jk}^{MT}$

$$\sum_{j \in U_M} \tilde{\theta}_{jk}^{MT} = 1 \text{ pour tout } k \in U_T.$$

- Total  $t_Y$  sur  $U_T$  équivalent à un total sur  $U_F$  :

$$t_Y = \sum_{k \in U_T} y_k = \sum_{k \in U_T} y_k \sum_{j \in U_M} \tilde{\theta}_{jk}^{MT} \sum_{i \in U_F} \tilde{\theta}_{ij}^{FM} = \sum_{i \in U_F} \tilde{y}_i^D$$

$$\text{avec } \tilde{y}_i^D = \sum_{j \in U_M} \tilde{\theta}_{ij}^{FM} \sum_{k \in U_T} \tilde{\theta}_{jk}^{MT} y_k$$

# MGPP double

- Méthode Généralisée de Partage des Poids double (MGPP double) : (Medous et al., 2023)

$$\hat{t}_{Y, MGPPd} = \sum_{i \in S_F} \frac{\tilde{y}_i^D}{\pi_i} \quad \text{avec} \quad \tilde{y}_i^D = \sum_{j \in U_M} \tilde{\theta}_{ij}^{FM} \sum_{k \in U_T} \tilde{\theta}_{jk}^{MT} y_k$$

- Sans biais si et seulement si  $\tilde{\theta}_{ij}^{FM}$  and  $\tilde{\theta}_{jk}^{MT}$  standardisés
- Cas particulier de MGPP simple, avec

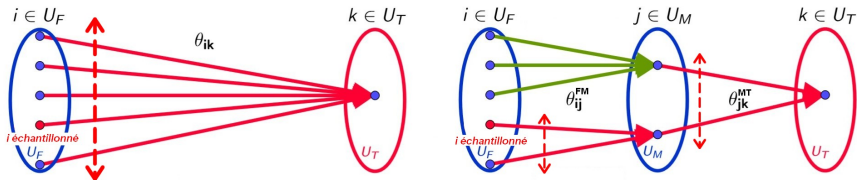
$$\tilde{\theta}_{ik} = \sum_{j \in U_M} \tilde{\theta}_{ij}^{FM} \tilde{\theta}_{jk}^{MT}$$

- 1 Sondage indirect et MGPP
- 2 Sondage indirect double et MGPP double
- 3 Comparaison des deux méthodes**
- 4 Simulations
- 5 Conclusion
- 6 References

# Nombre de liens observés

- Liens TpU-TpU

⇒ Moins d'individus à identifier que la MGPP simple avec liens TpU



- La Poste :

MGPP simple : 500 adresses,

MGPP double : 60 liens = 10 adresses + 50 cases

# Différence de Variance

- Entre MGPP optimale et MGPP double quelconque :

Proposition ((Medous et al., 2023))

Si les liens sont TpU et que la propriété  $\Delta$  est vérifiée :

$$\text{Var}(\hat{t}_{Y, MGPPd}) - \text{Var}(\hat{t}_{Y, MGPP}^{opt}) = \sum_{k \in U_T} y_k^2 \text{Var}(\hat{t}_{\tilde{\theta}_k - \tilde{\theta}_k^{opt}}) \geq 0.$$

avec

$$\hat{t}_{\tilde{\theta}_k - \tilde{\theta}_k^{opt}} = \sum_{i \in S_F} \left( \sum_{j \in U_M} \tilde{\theta}_{ij}^{FM} \tilde{\theta}_{jk}^{MT} - \tilde{\theta}_{ik}^{opt} \right) / \pi_i$$

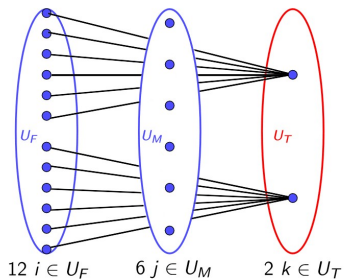
⇒ Impact de la répartition des liens sur la perte de précision ?



- 1 Sondage indirect et MGPP
- 2 Sondage indirect double et MGPP double
- 3 Comparaison des deux méthodes
- 4 Simulations**
- 5 Conclusion
- 6 References

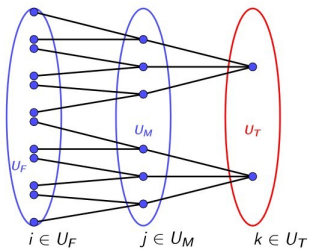
# Populations

- Trois populations de **tailles fixes**
- **500 liens** entre  $k \in U_T$  et  $U_F$

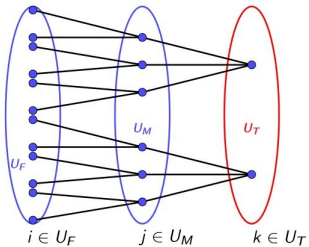


Taille de  $U_T$  : 6 958,  
 Taille de  $U_M$  : 69 580,  
 Taille de  $U_F$  : 3 479 000

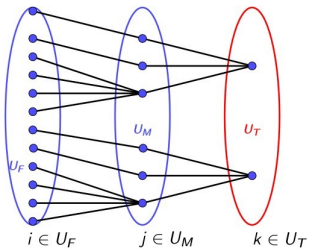
- $N_{Fj}$  nombre d'individus de  $U_F$  liés à  $j \in U_M$ ,  
 $N_{Mk}$  nombre d'individus de  $U_M$  liés à  $k \in U_T$ ,

1.  $N_{Fj}$  &  $N_{Mk}$  uniformes

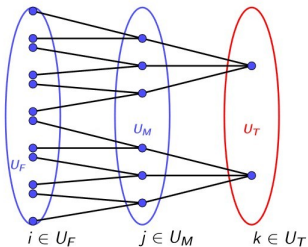
1.  $N_{Fj}$  &  $N_{Mk}$  uniformes



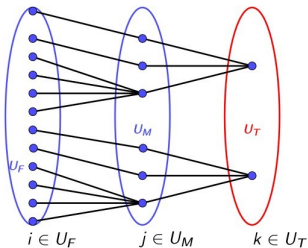
2.  $N_{Fj}$  non uniformes,  $N_{Mk}$  uniformes



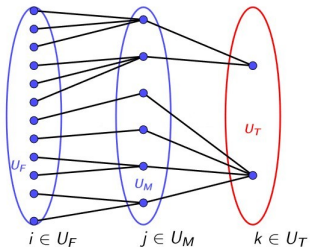
1.  $N_{Fj}$  &  $N_{Mk}$  uniformes

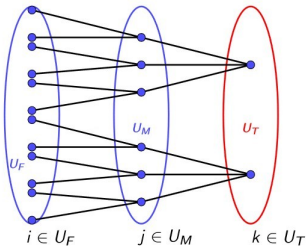
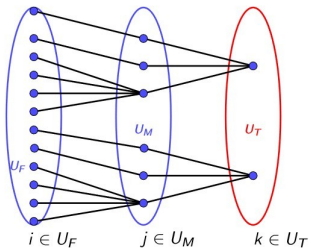
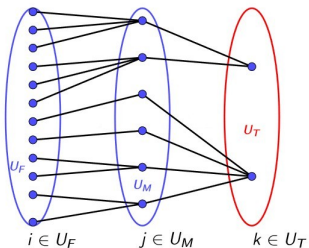
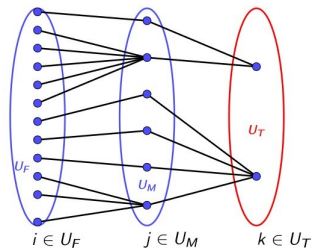


2.  $N_{Fj}$  non uniformes,  $N_{Mk}$  uniformes



3.  $N_{Fj}$  uniformes,  $N_{Mk}$  non uniformes



1.  $N_{Fj}$  &  $N_{Mk}$  uniformes2.  $N_{Fj}$  non uniformes,  $N_{Mk}$  uniformes3.  $N_{Fj}$  uniformes,  $N_{Mk}$  non uniformes4.  $N_{Fj}$  &  $N_{Mk}$  non uniformes

# Étude de cas

- Estimation de  $t_Y$ ,  $Y$  trafic postal
- Monte Carlo avec  $R = 100\,000$  simulations ( $r = 1, \dots, R$ )
- échantillons de taille  $n$  tirés par **SASSR**
- Poids de lien optimaux existant !

# Estimateurs

- MGPP simple avec **poids optimaux**  $1/N_{Fk}$
- MGPP double avec **poids usuels**  
⇒ **poids usuels** :  $1/N_{Fj} \times 1/N_{Mk}, j \in U_M$   
⇒ poids obtenus quand :

$$\theta_{ij}^{FM} = \theta_{jk}^{MT} = 1$$

⇒ poids **utilisés à La Poste**



# Étude de cas

- Biais relatif :

$$BR(\hat{t}) = (1/R) \sum_{r=1}^R (\hat{t}^{(r)} - t_Y) / t_Y$$

non rapportés car négligeables

- Erreur quadratique moyenne

$$EQM(\hat{t}) = \sum_{r=1}^R \left( \hat{t}^{(r)} - (1/R) \sum_{r=1}^R \hat{t}^{(r)} \right)^2$$

- Racine de l'Erreur quadratique moyenne relative :

$$EQMR_{opt}(\hat{t}) = \left( \frac{EQM(\hat{t})}{EQM(\hat{t}_{Y, MGPP}^{opt})} \right)^{1/2}$$

1.  $N_{Fj}$  &  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 1.00         |
| 1000 | 1.00         |

60 liens observés en moyenne  
par  $k \in U_T$ .

$$\sum_{j \in U^M} \tilde{\theta}_{ij} \tilde{\theta}_{jk} = \tilde{\theta}_{ik} \Rightarrow \text{MGPPd} = \text{MGPP}.$$

MGPP simple : 500 liens observés en moyenne par  $k \in U_T$ .

1.  $N_{Fj}$  &  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 1.00         |
| 1000 | 1.00         |

60 liens observés en moyenne  
par  $k \in U_T$ .

2.  $N_{Fj}$  non uniformes,  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 7.19         |
| 1000 | 7.14         |

60 liens observés en moyenne  
par  $k \in U_T$ .

$$\sum_{j \in U^M} \tilde{\theta}_{ij} \tilde{\theta}_{jk} = \tilde{\theta}_{ik} \Rightarrow \text{MGPPd} = \text{MGPP}.$$

MGPP simple : 500 liens observés en moyenne par  $k \in U_T$ .

1.  $N_{Fj}$  &  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 1.00         |
| 1000 | 1.00         |

60 liens observés en moyenne  
par  $k \in U_T$ .

2.  $N_{Fj}$  non uniformes,  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 7.19         |
| 1000 | 7.14         |

60 liens observés en moyenne  
par  $k \in U_T$ .

$$\sum_{j \in U^M} \tilde{\theta}_{ij} \tilde{\theta}_{jk} = \tilde{\theta}_{ik} \Rightarrow \text{MGPPd} = \text{MGPP}.$$

3.  $N_{Fj}$  uniformes,  $N_{Mk}$  non uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 1.00         |
| 1000 | 1.00         |

314 liens observés en moyenne  
par  $k \in U_T$ .

MGPP simple : 500 liens observés en moyenne par  $k \in U_T$ .

1.  $N_{Fj}$  &  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 1.00         |
| 1000 | 1.00         |

60 liens observés en moyenne  
par  $k \in U_T$ .

2.  $N_{Fj}$  non uniformes,  $N_{Mk}$  uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 7.19         |
| 1000 | 7.14         |

60 liens observés en moyenne  
par  $k \in U_T$ .

$$\sum_{j \in U^M} \tilde{\theta}_{ij} \tilde{\theta}_{jk} = \tilde{\theta}_{ik} \Rightarrow \text{MGPPd} = \text{MGPP}.$$

3.  $N_{Fj}$  uniformes,  $N_{Mk}$  non uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 1.00         |
| 1000 | 1.00         |

314 liens observés en moyenne  
par  $k \in U_T$ .

MGPP simple : 500 liens observés en moyenne par  $k \in U_T$ .

4.  $N_{Fj}$  &  $N_{Mk}$  non uniformes

| $n$  | $EQMR_{opt}$ |
|------|--------------|
| 500  | 11.78        |
| 1000 | 11.55        |

314 liens observés en moyenne  
par  $k \in U_T$ .

- 1 Sondage indirect et MGPP
- 2 Sondage indirect double et MGPP double
- 3 Comparaison des deux méthodes
- 4 Simulations
- 5 Conclusion**
- 6 References

# Conclusion

- Poids de lien optimaux pour MGPP simple avec liens  $TpU$
- $TpU$  + MGPP simple : problématique !
- Moins d'individus à identifier avec MGPP double avec liens  $TpU-TpU$
- Perte de précision entre MGPP simple optimale et MGPP double  
⇒ due à la variabilité des poids MGPP double.
- La Poste : variabilité des poids MGPP double causé par la variabilité des liens.
- Bouriaud et al. (2024)

- Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *The Annals of Mathematical Statistics*, pages 105–110.
- Bouriaud, O. B., Brion, P., Chauvet, G., Duong, T. H. K., and Pulkinen, M. (2024). The weight share method in forest inventories : refining the relation between points and trees. *Canadian Journal of Forest Research*, (ja).
- Deville, J.-C. and Lavallée, P. (2006). Sondage indirect : les fondements de la méthode généralisée du partage des poids. *Techniques d'enquête*, 32(2) :185.
- Kalton, G. and Brick, J. M. (1995). Weighting schemes for household panel surveys. *Survey Methodology*, 21(2) :33–34.
- Kiesl, H. (2016). Indirect sampling : a review of theory and recent applications. *AStA Wirtschafts-und Sozialstatistisches Archiv*, 10(4) :289–303.



- Lardin-Puech, P. (2014). Estimation du trafic de courrier distribué en France métropolitaine par sondage indirect. In *the proceedings of the 8ème colloque francophone sur les sondages, Dijon*.  
[http://paperssondages14.sfds.asso.fr/submission\\_100.pdf](http://paperssondages14.sfds.asso.fr/submission_100.pdf).
- Lavallée, P. (2007). *Indirect sampling*. Springer Science & Business Media.
- Lavallée, P. (2009). *Indirect sampling*. Springer Science & Business Media.
- Medous, E., Goga, C., Ruiz-Gazen, A., Beaumont, J.-F., Dessertaine, A., and Puech, P. (2023). Many-to-one indirect sampling with application to the French postal traffic estimation. *The Annals of Applied Statistics*, 17(1) :838–859.
- Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in Statistics : Foundations and basic theory*, pages 235–247. Springer.

# tous-pour-Un généralisé

- Clusters d'individus de  $U_T$  liés aux mêmes individus de  $U_F$
- Population de clusters  $U'$

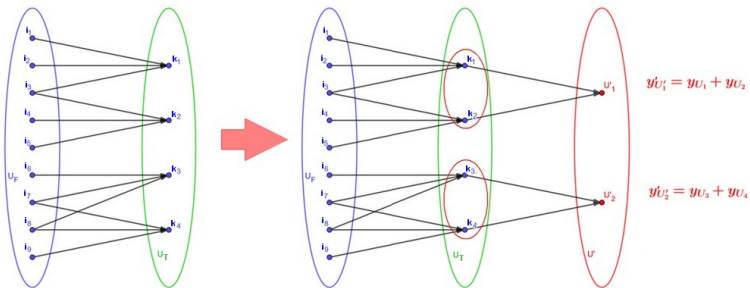


Figure – Création d'une population de clusters

# Tous-pour-Un généralisé

- TpU entre  $U_F$  et  $U$
- $\Rightarrow$  Poids de lien optimaux existent

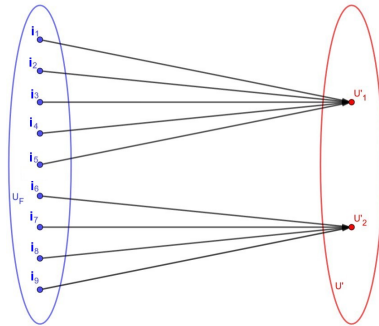


Figure – Liens entre  $U_F$  et  $U$ .

## Poids de lien optimaux pour non-TpU

- $T_Y$  comme total sur  $U'$  de  $y'_q = \sum_{k \in q} y_k$  :

$$T_Y = \sum_{k \in U_T} y_k = \sum_{q \in U'} \sum_{k \in q} y_k = \sum_{q \in U'} y'_q = T_{Y'}.$$

- $\hat{t}_{Y, GWSM}$  comme estimateur MGPP de  $T_{Y'}$  avec  $\tilde{\theta}_{iq} = \sum_{k \in q} \tilde{\theta}_{ik} y_k / y'_q$  :

$$\begin{aligned} \hat{t}_{Y, MGPP} &= \sum_{k \in U} \left( \sum_{i \in S_F} \frac{\tilde{\theta}_{ik}}{\pi_i} \right) y_k = \sum_{q \in U'} \sum_{k \in q} \left( \sum_{i \in S_F} \frac{\tilde{\theta}_{ik}}{\pi_i} \right) y_k \\ &= \sum_{q \in U'} \left( \sum_{i \in S_F} \frac{1}{\pi_i} \left( \sum_{k \in q} \frac{\tilde{\theta}_{ik}}{y'_q} y_k \right) \right) y'_q = \sum_{q \in U'} \left( \sum_{i \in S_F} \frac{\tilde{\theta}_{iq}}{\pi_i} \right) y'_q \\ &= \hat{t}_{Y', MGPP} \end{aligned}$$

## Comparaison avec HT direct

- $D_i = \{(k, y_k), k \in U_T, l_{ik} \neq 0\}$ ,  $D_F = \{(i, D_i), i \in s_F\}$ ,  
 $D_T = \{(k, y_k), k \in s_T\}$ .
- Lavallée (2007) :  $D_T$  statistique suffisante pour  $D_F$ .
- $\hat{t}_{Y, MGPP} = \hat{t}_{Y, MGPP}(D_F)$
- théorème de Rao-Blackwell (Rao (1992), Blackwell (1947)) :  
indépendamment des liens et des poids,

$$\text{Var}(\hat{t}_{Y, MGPP}) \leq \text{Var}(E[\hat{t}_{Y, MGPP} | D_T]).$$

- $E[\hat{t}_{Y, MGPP} | D_T]$  estimateur HT direct de  $T_Y$ .

## Comparaison avec HT direct

- Preuve par Kalton and Brick (1995) (pour les enquêtes longitudinales)
- Espérance et variance prises **conditionnellement** à  $Y$  et

$$\begin{aligned}
 E[\hat{t}_{Y, MGPP} | D_T] &= E[\hat{t}_{Y, MGPP} | D_T, Y] \\
 &= \sum_{k \in s_T} y_k E \left[ \sum_{i \in s_F} d_i \tilde{\theta}_{ik} | D_T, Y \right] \\
 &= \sum_{k \in s_T} y_k \sum_{i \in U_F} d_i \tilde{\theta}_{ik} E[l_i | k \in s_T] \\
 &= \sum_{k \in s_T} y_k \sum_{i \in U_F} d_i \tilde{\theta}_{ik} \frac{\pi_i}{\pi_k^d} = \sum_{k \in s_T} \frac{y_k}{\pi_k^d}
 \end{aligned}$$

avec  $\pi_k^d = P(k \in s_T)$ .

# Comparaison avec HT direct

- Kiesel (2016) : pourquoi  $E[l_i | s_T] = E[l_i | k \in s_T]$  ?
- $k(i) = \{k \in U_T, k \text{ lié à } i\}$ ,  $\bar{s}_T^k := \{k' \in s_T, k' \neq k\}$

$$\begin{aligned}
 E[l_i | s_T] &= \frac{\pi_i}{\pi_k^d} \frac{P(i \in s_F \cap k(i) \in s_T \cap \bar{s}_T^{k(i)})}{\pi_i} \frac{\pi_k^d}{P(k(i) \in s_T \cap \bar{s}_T^{k(i)})} \\
 &= \frac{\pi_i}{\pi_k^d} \frac{P(i \in s_F \cap \bar{s}_T^{k(i)})}{\pi_i} \frac{\pi_k^d}{P(k \in s_T \cap \bar{s}_T^k)} = \frac{\pi_i}{\pi_k^d} \frac{P(\bar{s}_T^{k(i)} | i)}{P(\bar{s}_T^k | k)}
 \end{aligned}$$

- $P(\bar{s}_T^{k(i)} | i)$  conditionnellement à tout échantillon  $s_F$  contenant  $i$
- $P(\bar{s}_T^k | k)$  conditionnellement à tout échantillon  $s_F$  contenant un individu lié à  $k$

## Comparaison avec HT direct

- Plan de Poisson : sélection de  $\bar{s}_T^{k(i)}$  indépendante de  $i$  :

$$P(\bar{s}_T^{k(i)} | i) = P(\bar{s}_T^{k(i)})$$

- Si TpU, alors  $k(i) = k$  et :

$$P(\bar{s}_T^k | k) = P(\bar{s}_T^k) = P(\bar{s}_T^{k(i)})$$

⇒ Alors le HT direct est plus précis que la MGPP indirecte (optimale ou non)