Jean-Claude Deville et le partage des poids

Insee

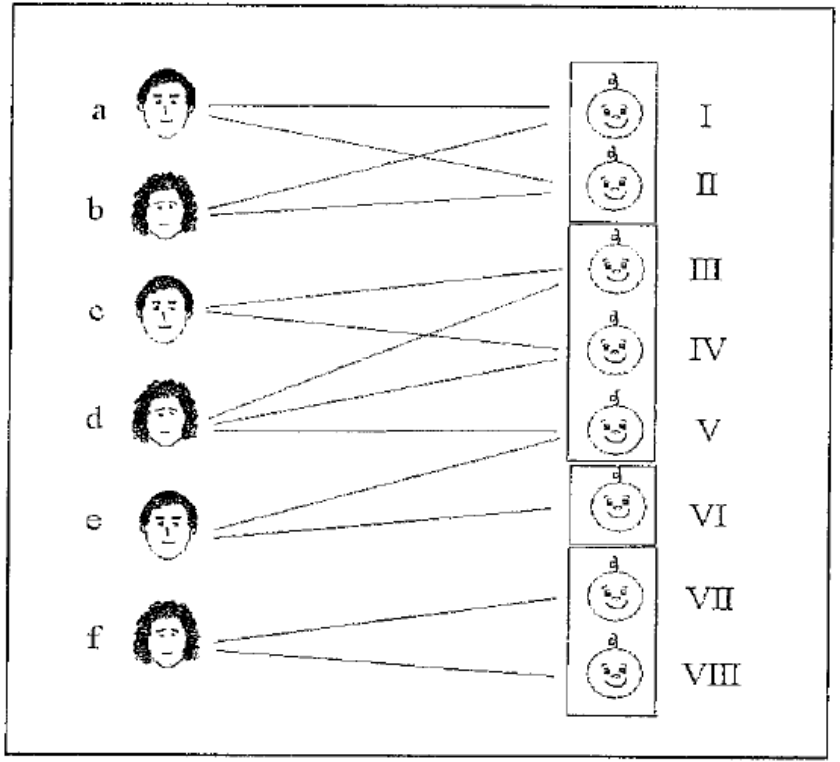
Mesurer pour comprendre

Le partage des poids : motivation

Nous nous intéressons à une population U_B (par exemple, population d'enfants) pour laquelle nous ne disposons pas d'une base de sondage.

Nous disposons en revanche d'une base de sondage pour une population U_A dont les unités sont liées à celles de U_B (par exemple, population de parents).

Comment procéder pour échantillonner des unités de U_B et procéder à l'estimation?



Extrait de Lavallée (2007), *Indirect sampling*.

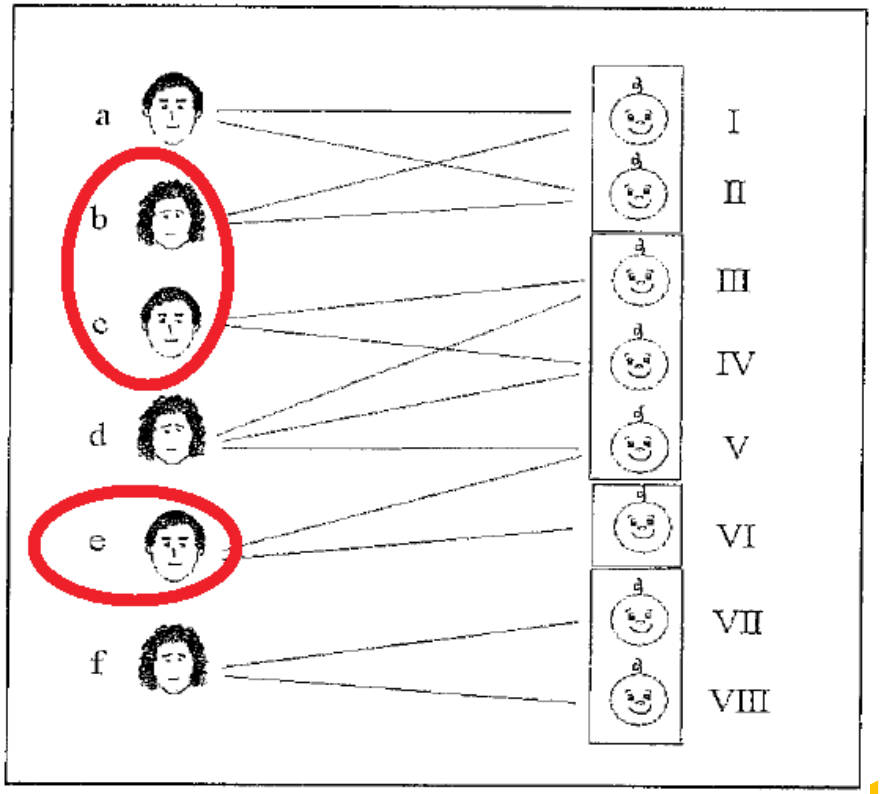
Le partage des poids : principe

Des solutions proposées dans la littérature pour estimer la prévalence d'une maladie rare (Birnbaum et Sirken, 1965) ou pour les enquêtes répétées dans le temps (Ernst, 1989).

Généralisation et traitement unifié dans Deville, J-C et Lavallée, P. (2006) :
Sondage indirect – les fondements de la méthode généralisée du partage des poids

Principe :

- Nous tirons un échantillon S_A dans U_A selon un plan de sondage choisi.



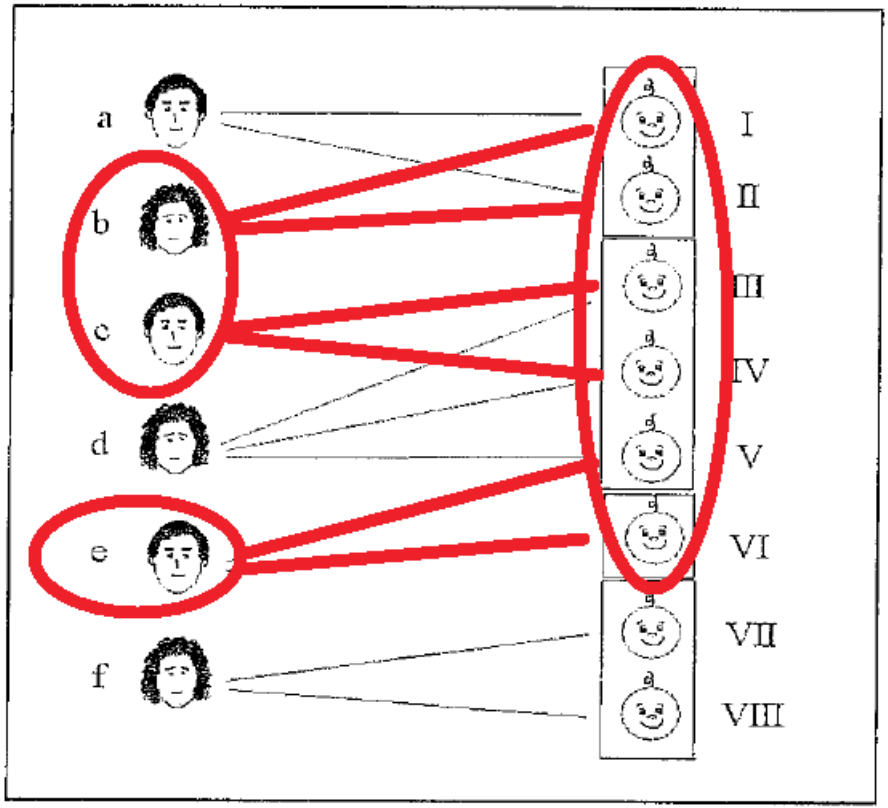
Le partage des poids : principe

Des solutions proposées dans la littérature pour estimer la prévalence d'une maladie rare (Birnbaum et Sirken, 1965) ou pour les enquêtes répétées dans le temps (Ernst, 1989).

Généralisation et traitement unifié dans Deville, J-C et Lavallée, P. (2006) :
Sondage indirect – les fondements de la méthode généralisée du partage des poids

Principe :

- Nous tirons un échantillon S_A dans U_A selon un plan de sondage choisi.
- Nous enquêtons dans U_B l'échantillon S_B des unités liées à au moins une unité de S_A.



Le partage des poids : formules

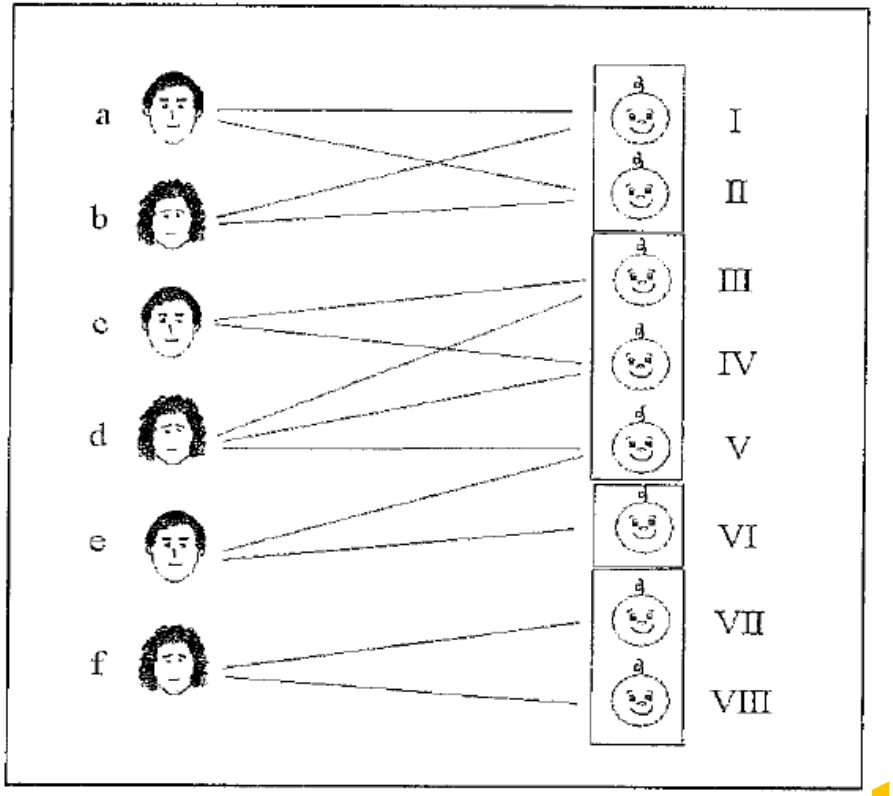
La méthode de partage des poids repose sur un principe de dualité :

- Une variable y_{kB} dans U_B peut être transportée comme une variable y_{iA} dans U_A de même total :

$$y_{iA} = \sum_{\text{Unités } k \text{ de } U_B \text{ liées à } i} \frac{y_{kB}}{\text{Nombre de liens entre } k \text{ et } U_A}$$

- Les poids des unités de S_B s'obtiennent à partir des poids d_{iA} des unités de S_A :

$$d_{kB} = \frac{\sum_{\text{Unités } i \text{ de } S_A \text{ liées à } k} d_{iA}}{\text{Nombre de liens entre } k \text{ et } U_A}$$



Le partage des poids : applications

L'idée est simple et géniale. Elle permet de se tirer de situations où le calcul d'un estimateur direct serait presque impossible. Cette plasticité a permis de très nombreuses applications.

- **Enquêtes répétées dans le temps**
  U_B = ménages au temps t+1 U_A = individus au temps t

- Enquête MORGOAT sur le tourisme en Bretagne
  U_B = touristes U_A = voitures OU lieux touristiques OU boulangeries

- Echantillonnage forestier (« local density », Mandallaz 2007)
  U_B = population d'arbres $\mathcal{U}_A$ = continuum de points (Brion et al, 2022)

- Echantillonnage de réseaux (« ancestral BIG sampling », Zhang, 2022)

- Redécouverte en biostatistique (« time-location sampling », Leon et al., 2015)
  U_B = individus U_A = lieux * moments

Que retenir de Jean-Claude Deville ?

Une certaine idée de la statistique d'enquête. Des travaux (calage, échantillonnage équilibré, estimation de variance, partage des poids, …) qui ont profondément infusé et ont durablement modifié la pratique des enquêtes.

Une recherche très largement motivée par les applications, et tournée vers les applications. Un sportif complet, avec à son actif des articles de recherche mathématiquement brillants et d'autres très appliqués. Jean-Claude était également un très bon programmeur (avec un goût prononcé pour MATLAB).

Une méfiance pour les modèles (et des blagues récurrentes sur les économètres) et probablement pour une trop grande mathématisation de la statistique. Il est révélateur qu'il n'ait jamais travaillé en estimation sur petits domaines.

Kenavo, Jean-Claude.