

La convergence d'estimateurs et d'estimateurs de variance pour l'échantillonnage à deux degrés

Audrey-Anne Vallée

Université de Neuchâtel

et

Guillaume Chauvet

Ensaï (Irmair)

Colloque francophone sur les sondages

26 Octobre 2018

Lyon

Contexte

Échantillonnage à deux degrés

Estimateurs pour les plans à deux degrés

Estimateurs et hypothèses

Estimateurs

Hypothèses

Résultats principaux

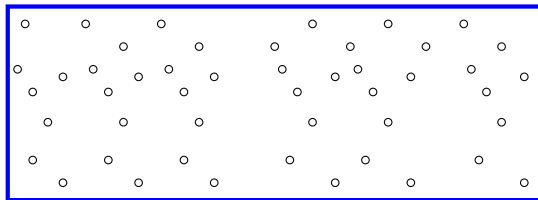
Convergence des estimateurs du total et de la variance

Convergence d'un estimateur de variance simplifié

Cas des plans de sondage à grande entropie

Étude par simulations

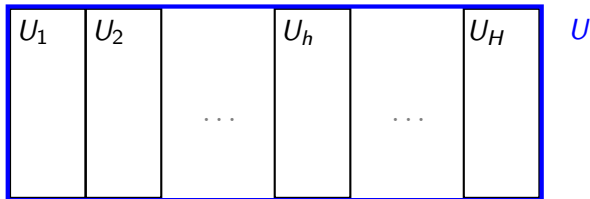
Échantillonnage à deux degrés

 U

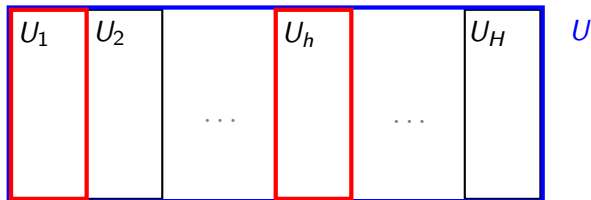
Échantillonnage à deux degrés

 U

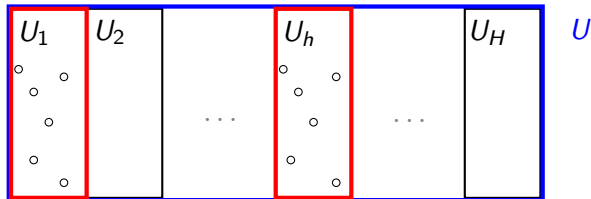
Échantillonnage à deux degrés



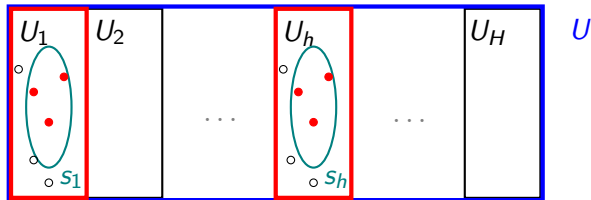
Échantillonnage à deux degrés



Échantillonnage à deux degrés



Échantillonnage à deux degrés



Estimateurs pour les plans à deux degrés

Pour estimer un total:

- ▶ Estimateur du total Horvitz-Thompson;
- ▶ Estimateurs de variance habituels (Horvitz-Thompson, Yates-Grundy).

Certaines propriétés des estimateurs sont nécessaires:

- ▶ Convergence de l'estimateur;
- ▶ Convergence de l'estimateur de variance;
- ▶ Validité d'un théorème central limite.

Objectif: Établir ces propriétés pour les plans à deux degrés, précisément pour un estimateur de variance simplifié.

Contexte

Échantillonnage à deux degrés

Estimateurs pour les plans à deux degrés

Estimateurs et hypothèses

Estimateurs

Hypothèses

Résultats principaux

Convergence des estimateurs du total et de la variance

Convergence d'un estimateur de variance simplifié

Cas des plans de sondage à grande entropie

Étude par simulations

Estimateurs

- ▶ Population U de N Unités Secondaires d'Échantillonnage (USEs),
- ▶ Partitionnée en une population U_I de N_I Unités Primaires d'Échantillonnage (UPEs).
- ▶ Premier degré:
 - ▶ Échantillon S_I de n_I UPEs est sélectionné dans U_I .
 - ▶ π_{Ii} et π_{Iij} : probabilités d'inclusion d'ordre un et deux.
 - ▶ Définissons $\Delta_{ij} = \pi_{Iij} - \pi_{Ii}\pi_{Ij}$.
- ▶ Deuxième degré: échantillon S_i de n_i USEs est sélectionné dans chaque UPE $u_i \in S_I$.

Estimateurs

- ▶ Total de la variable y :

$$Y = \sum_{k \in U} y_k = \sum_{u_i \in U_I} Y_i,$$

où Y_i est le total dans l'UPE u_i .

- ▶ L'estimateur Horvitz-Thompson de Y :

$$\hat{Y}_\pi = \sum_{u_i \in S_I} \frac{\hat{Y}_i}{\pi_{Ii}},$$

où \hat{Y}_i estime le total Y_i .

Estimateurs

- ▶ La variance de \hat{Y}_π est

$$V(\hat{Y}_\pi) = \underbrace{\sum_{u_i, u_j \in U_I} \Delta_{ij} \frac{Y_i}{\pi_{li}} \frac{Y_j}{\pi_{lj}}}_{V_1(\hat{Y}_\pi)} + \underbrace{\sum_{u_i \in U_I} \left(\frac{1 - \pi_{li}}{\pi_{li}} \right) V_i}_{V_2(\hat{Y}_\pi)} + \underbrace{\sum_{u_i \in U_I} V_i}_{V_3(\hat{Y}_\pi)},$$

où $V_i = V(\hat{Y}_i)$ est la variance de \hat{Y}_i dans l'UPE u_i .

- ▶ L'estimateur Horvitz-Thompson de la variance est

$$\hat{V}_{HT}(\hat{Y}_\pi) = \underbrace{\sum_{u_i, u_j \in S_I} \frac{\Delta_{ij}}{\pi_{lij}} \frac{\hat{Y}_i}{\pi_{li}} \frac{\hat{Y}_j}{\pi_{lj}}}_{\hat{V}_{HT,A}(\hat{Y}_\pi)} + \underbrace{\sum_{u_i \in S_I} \frac{\hat{V}_{HT,i}}{\pi_{li}}}_{\hat{V}_{HT,B}(\hat{Y}_\pi)},$$

où $\hat{V}_{HT,i}$ est un estimateur de V_i .

Hypothèses

Variable y

- ▶ Moment d'ordre quatre fini $\frac{1}{N} \sum_{k \in U} y_k^4 \leq \infty$.

Premier degré du plan de sondage

- ▶ Fraction de sondage plus petite que 1,
- ▶ Probabilités d'inclusion d'ordre de grandeur de n_I/N_I ,
- ▶ Bornes pour les probabilités d'inclusion d'ordre deux et trois.

Deuxième degré du plan de sondage

- ▶ Tailles N_i et n_i comparables,
- ▶ Probabilités d'inclusion d'ordre de grandeur de n_i/N_i ,
- ▶ Bornes pour les probabilités d'inclusion d'ordre deux et trois.

Contexte

Échantillonnage à deux degrés

Estimateurs pour les plans à deux degrés

Estimateurs et hypothèses

Estimateurs

Hypothèses

Résultats principaux

Convergence des estimateurs du total et de la variance

Convergence d'un estimateur de variance simplifié

Cas des plans de sondage à grande entropie

Étude par simulations

Convergence des estimateurs du total et de la variance

- ▶ Convergence de \hat{Y}_π vers le total Y .
 - ▶ Convergence de $\hat{V}_{HT,A}(\hat{Y}_\pi)$ vers $V_1(\hat{Y}_\pi) + V_2(\hat{Y}_\pi)$.
 - ▶ Convergence de $\hat{V}_{HT,B}(\hat{Y}_\pi)$ vers $V_3(\hat{Y}_\pi)$.
- ⇒ $\hat{V}_{HT}(\hat{Y}_\pi) = \hat{V}_{HT,A}(\hat{Y}_\pi) + \hat{V}_{HT,B}(\hat{Y}_\pi)$ est un estimateur convergent de $V(\hat{Y}_\pi)$.

Convergence d'un estimateur de variance simplifié

► Rappel: $\hat{V}_{HT}(\hat{Y}_\pi) = \underbrace{\sum_{u_i, u_j \in S_I} \frac{\Delta_{ij}}{\pi_{lij}} \frac{\hat{Y}_i}{\pi_{li}} \frac{\hat{Y}_j}{\pi_{lj}}}_{\hat{V}_{HT,A}(\hat{Y}_\pi)} + \underbrace{\sum_{u_i \in S_I} \frac{\hat{V}_{HT,i}}{\pi_{li}}}_{\hat{V}_{HT,B}(\hat{Y}_\pi)}$.

- Convergence de $\hat{V}_{HT,A}(\hat{Y}_\pi)$ vers $V(\hat{Y}_\pi)$ si

$$\frac{V_3(\hat{Y}_\pi)}{V_1(\hat{Y}_\pi) + V_2(\hat{Y}_\pi)} \rightarrow 0.$$

(... si la fraction de sondage au premier degré $n_I/N_I \rightarrow 0$.)

Échantillonnage réjectif

- ▶ Échantillonnage réjectif au premier degré (Hájek, 1964);
- ▶ Approximation du premier terme de l'estimateur de variance:

$$\hat{V}_{HAJ,A}(\hat{Y}_{r\pi}) = \sum_{u_i \in S_{rI}} (1 - \pi_{li}) \left(\frac{\hat{Y}_i}{\pi_{li}} - \hat{R}_\pi \right)^2,$$

où \hat{R}_π est une fonction de π_{li} et \hat{Y}_i .

- ▶ Convergence de $\hat{V}_{HAJ,A}(\hat{Y}_{r\pi})$ vers $V_1(\hat{Y}_\pi) + V_2(\hat{Y}_\pi)$.
- ▶ Version simplifiée.

- ▶ $\frac{\hat{Y}_{r\pi} - Y}{\sqrt{V(\hat{Y}_{r\pi})}} \rightarrow_{\mathcal{L}} \mathcal{N}(0, 1)$.

Contexte

Échantillonnage à deux degrés

Estimateurs pour les plans à deux degrés

Estimateurs et hypothèses

Estimateurs

Hypothèses

Résultats principaux

Convergence des estimateurs du total et de la variance

Convergence d'un estimateur de variance simplifié

Cas des plans de sondage à grande entropie

Étude par simulations

Simulations

- ▶ Population de $N_l = 2000$ UPEs;
- ▶ Nombre moyen de 40 USEs par UPEs, coefficient de variation 0.03;
- ▶ $y_{i,k} = \lambda + \sigma \nu_i + [\rho^{-1}(1 - \rho)]^{0.5} \sigma \varepsilon_k$,
où $\lambda = 20$, $\sigma = 2$,
 ρ_h t.q. le coefficient de corrélation intra-classe est 0.3,
 $\varepsilon_k \sim N(0, 1)$, $\nu_i \sim N(0, 1)$ dans l'UPE u_i ;
- ▶ Premier degré: $n_l = 20, 40, 100, 200$ UPEs; échantillonnage réjectif;
- ▶ Deuxième degré: $n_i = 5$ USEs; échantillonnage aléatoire simple sans remise;
- ▶ $R = 1000$ répétitions: \hat{Y}_π , $\hat{V}_{HAJ,A}(\hat{Y}_\pi)$ et $\hat{V}_{HAJ}(\hat{Y}_\pi)$;
- ▶ Biais relatif Monte-Carlo $BR_{MC}(\hat{V})$; Stabilité relative Monte-Carlo $SR_{MC}(\hat{V})$;
Probabilité de couverture de l'intervalle de confiance normal 95% (IC_{MC}).

Simulations et remarques

Table: Biais relatif, stabilité relative, taux de couverture de $\widehat{V}_{HAJ,A}(\widehat{Y}_\pi)$ et $\widehat{V}_{HAJ}(\widehat{Y}_\pi)$.

n_I	BR_{MC}		SR_{MC}		IC_{MC}	
	$\widehat{V}_{HAJ,A}$	\widehat{V}_{HAJ}	$\widehat{V}_{HAJ,A}$	\widehat{V}_{HAJ}	$\widehat{V}_{HAJ,A}$	\widehat{V}_{HAJ}
20	-0.72	-0.43	32.89	32.88	0.94	0.94
40	-0.77	-0.19	22.58	22.56	0.94	0.94
100	-1.63	-0.14	14.09	14.00	0.95	0.95
200	-3.26	-0.16	9.80	9.25	0.95	0.95

- ▶ Scénarios différents: variation de N_i , ICC et n_i .
- ▶ Autres plans de sondage à grande entropie.
- ▶ Autres paramètres et applications.

Référence

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, **35**, 1491–1523.