

Estimation de la variance pour un plan produit

Jean Rubin Guillaume Chauvet

ENSAI

30 mars 2022

PLAN

- ① Introduction
- ② Propriétés générales
 - Décomposition de Hoeffding-Sobol
 - Résultats théoriques
- ③ Propositions de méthodes d'estimation
 - Constructions générales
 - Estimations classiques
 - Estimations bootstrap
 - Illustrations : plan de sondage simple
 - Estimations classiques
 - Estimations bootstrap
 - Simulations
- ④ Conclusion

INTRODUCTION

INTRODUCTION

Étudier un type de plan de sondage appelé *plan produit* (Cross-Classified Sampling) à D dimensions.

INTRODUCTION

Étudier un type de plan de sondage appelé *plan produit* (Cross-Classified Sampling) à D dimensions.

Population produit

Les individus d'une population produit \mathcal{U} sont de la forme $\mathbf{k} = (k_1, \dots, k_D)$, avec $k_1 \in \mathcal{U}_1, \dots, k_D \in \mathcal{U}_D$.
Ces derniers possèdent une variable d'intérêt réelle $y_{\mathbf{k}}$.

INTRODUCTION

Étudier un type de plan de sondage appelé *plan produit* (Cross-Classified Sampling) à D dimensions.

Population produit

Les individus d'une population produit \mathcal{U} sont de la forme $\mathbf{k} = (k_1, \dots, k_D)$, avec $k_1 \in \mathcal{U}_1, \dots, k_D \in \mathcal{U}_D$.
Ces derniers possèdent une variable d'intérêt réelle $y_{\mathbf{k}}$.

Exemple

Plans produits à deux dimensions :

- Ohlsson $\rightarrow \mathcal{U}_1 = \text{Magasins}, \mathcal{U}_2 = \text{Produits}$.

INTRODUCTION

Étudier un type de plan de sondage appelé *plan produit* (Cross-Classified Sampling) à D dimensions.

Population produit

Les individus d'une population produit \mathcal{U} sont de la forme $\mathbf{k} = (k_1, \dots, k_D)$, avec $k_1 \in \mathcal{U}_1, \dots, k_D \in \mathcal{U}_D$.

Ces derniers possèdent une variable d'intérêt réelle $y_{\mathbf{k}}$.

Exemple

Plans produits à deux dimensions :

- Ohlsson $\rightarrow \mathcal{U}_1 = \text{Magasins}, \mathcal{U}_2 = \text{Produits}$.
- Juillard et al. $\rightarrow \mathcal{U}_1 = \text{Maternités}, \mathcal{U}_2 = \text{Dates}$.

ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \cdots \times \mathcal{U}_D$

ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \cdots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

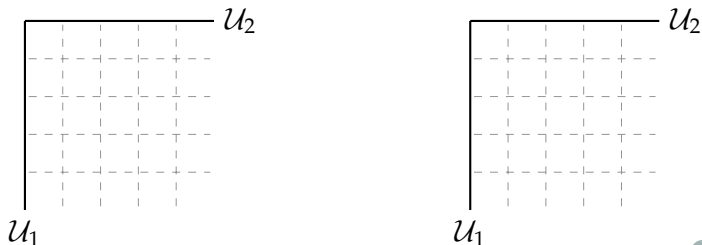
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

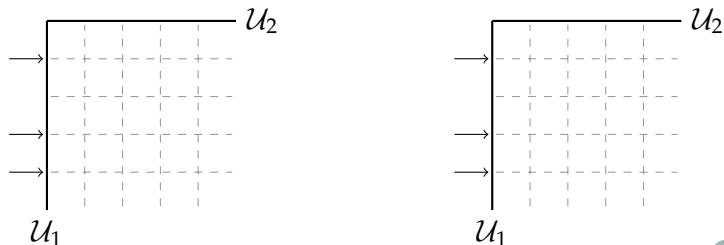
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

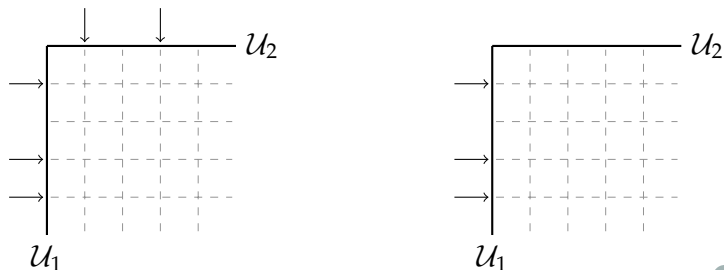
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

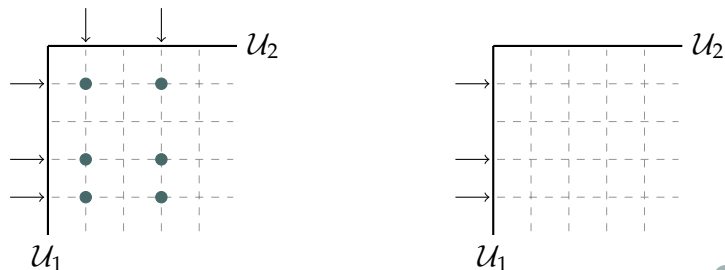
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

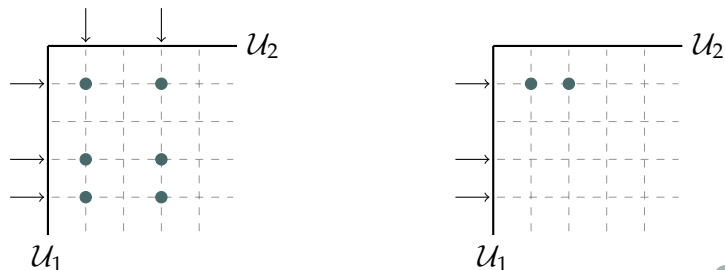
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

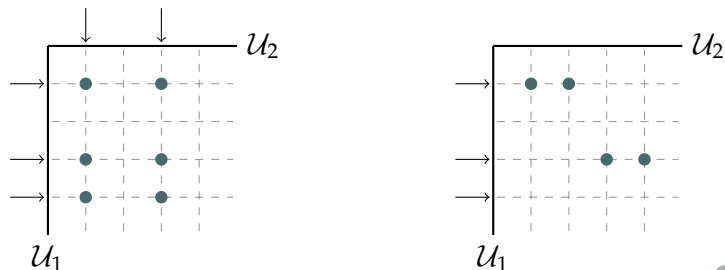
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



ÉCHANTILLONNAGE DE LA POPULATION PRODUIT

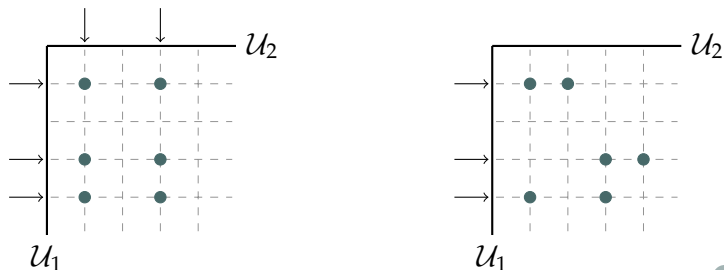
On échantillonne dans la population produit d'une manière naturelle :

- La population est de la forme $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_D$
- On sélectionne des échantillons $S_1 \subseteq \mathcal{U}_1, \dots, S_D \subseteq \mathcal{U}_D$ de manière indépendante.

L'échantillon final est alors $S = S_1 \times S_2 \times \dots \times S_D$.

Ainsi, on a $\pi_{\mathbf{k}} = \prod_{d=1}^D \pi_{k_d}^d$.

Figure: Comparaison entre plan produit (gauche) et plan à deux degrés (droite), voir Juillard et al.



OBJECTIF

À partir d'un échantillon S , un estimateur $f(\hat{Y})$ de $f(Y)$ est construit, où

$$Y = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}}, \quad \hat{Y} = \sum_{\mathbf{k} \in S} \frac{y_{\mathbf{k}}}{\pi_{\mathbf{k}}}$$

Le but est de produire des estimations de $V_p(f(\hat{Y}))$.

OBJECTIF

À partir d'un échantillon S , un estimateur $f(\hat{Y})$ de $f(Y)$ est construit, où

$$Y = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}}, \quad \hat{Y} = \sum_{\mathbf{k} \in S} \frac{y_{\mathbf{k}}}{\pi_{\mathbf{k}}}$$

Le but est de produire des estimations de $V_p(f(\hat{Y}))$.

Deux approches possibles:

- Produire un estimateur classique $\hat{V}_p(\hat{Y})$ puis utiliser une méthode de linéarisation de f .

OBJECTIF

À partir d'un échantillon S , un estimateur $f(\hat{Y})$ de $f(Y)$ est construit, où

$$Y = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}}, \quad \hat{Y} = \sum_{\mathbf{k} \in S} \frac{y_{\mathbf{k}}}{\pi_{\mathbf{k}}}$$

Le but est de produire des estimations de $V_p(f(\hat{Y}))$.

Deux approches possibles:

- Produire un estimateur classique $\hat{V}_p(\hat{Y})$ puis utiliser une méthode de linéarisation de f .
- Construire une méthode bootstrap, de sorte que $V_*(\hat{Y}^*) \approx \hat{V}_p(\hat{Y})$, puis calculer $V_*(f(\hat{Y}^*))$ par une méthode de Monte-Carlo.

PROPRIÉTÉS GÉNÉRALES

- 1 Introduction
- 2 Propriétés générales
Décomposition de Hoeffding-Sobol
Résultats théoriques
- 3 Propositions de méthodes d'estimation
- 4 Conclusion

DÉCOMPOSITION DE Hoeffding-SOBOL

Théorème (Décomposition de Hoeffding-Sobol)

Il est possible de décomposer \hat{Y} sous la forme

$$\hat{Y} = \sum_{I \subseteq \{1, \dots, D\}} \hat{Y}^I$$

où les termes \hat{Y}^I ne dépendent que des S_d avec $d \in I$. De plus, les termes de cette décomposition sont décorrélés, d'où

$$V_p(\hat{Y}) = \sum_{I \subseteq \{1, \dots, D\}} V_p(\hat{Y}^I).$$

► Détails

- 1 Introduction
- 2 Propriétés générales**
Décomposition de Hoeffding-Sobol
Résultats théoriques
- 3 Propositions de méthodes d'estimation
- 4 Conclusion

RÉSULTATS THÉORIQUES

On a le résultat principal suivant :

Théorème (Ordre de grandeur des termes de variance)

Sous des hypothèses classiques (voir Juillard et al.), on peut établir que

$$V_p(\hat{Y}^I) = O\left(\frac{N^2}{\prod_{d \in I} n_d}\right)$$

► Détails

RÉSULTATS THÉORIQUES

On a le résultat principal suivant :

Théorème (Ordre de grandeur des termes de variance)

Sous des hypothèses classiques (voir Juillard et al.), on peut établir que

$$V_p(\hat{Y}^I) = O\left(\frac{N^2}{\prod_{d \in I} n_d}\right)$$

► Détails

⇒ Les termes dominants sont les $V_p(\hat{Y}^{\{d\}}) = O(N^2/n_d)$.

PROPOSITIONS DE MÉTHODES D'ESTIMATION

- 1 Introduction
- 2 Propriétés générales
- 3 Propositions de méthodes d'estimation**
 - Constructions générales
 - Estimations classiques
 - Estimations bootstrap
 - Illustrations : plan de sondage simple
 - Estimations classiques
 - Estimations bootstrap
 - Simulations
- 4 Conclusion

CONSTRUCTION D'ESTIMATEURS SIMPLIFIÉS

Puisque $V_p(\hat{Y}) = \sum_{I \subseteq \{1, \dots, D\}} V_p(\hat{Y}^I) \approx \sum_{d=1}^D V_p(\hat{Y}^{\{d\}})$.
On peut utiliser

$$\hat{V}^{SIMP1}(\hat{Y}) = \sum_{d=1}^D \hat{V}_p(\hat{Y}^{\{d\}})$$

en utilisant des estimateurs sans biais de $\hat{V}_p(\hat{Y}^{\{d\}})$.

CONSTRUCTION D'ESTIMATEURS SIMPLIFIÉS

Puisque $V_p(\hat{Y}) = \sum_{I \subseteq \{1, \dots, D\}} V_p(\hat{Y}^I) \approx \sum_{d=1}^D V_p(\hat{Y}^{\{d\}})$.
On peut utiliser

$$\hat{V}^{SIMP1}(\hat{Y}) = \sum_{d=1}^D \hat{V}_p(\hat{Y}^{\{d\}})$$

en utilisant des estimateurs sans biais de $\hat{V}_p(\hat{Y}^{\{d\}})$.

Une simplification supplémentaire est possible, à l'aide d'une "estimation plug-in".

$$\hat{V}^{SIMP2}(\hat{Y}) = \sum_{d=1}^D \hat{V}^{PLUG}(\hat{Y}^{\{d\}})$$

CONTRAINTES DE MOMENTS POUR UN ESTIMATEUR BOOTSTRAP

On souhaite que " \hat{Y}^* se comporte vis-à-vis de \hat{Y} de la même manière que \hat{Y} se comporte vis-à-vis de Y ".

CONTRAINTES DE MOMENTS POUR UN ESTIMATEUR BOOTSTRAP

On souhaite que " \hat{Y}^* se comporte vis-à-vis de \hat{Y} de la même manière que \hat{Y} se comporte vis-à-vis de Y ".

Contraintes de moments

L'estimateur bootstrap \hat{Y}^* vérifie les contraintes de moments d'ordres 1 et 2 lorsque :

$$E_*(\hat{Y}^*) = \hat{Y}$$

$$V_*(\hat{Y}^*) = \hat{V}_p(\hat{Y})$$

MÉTHODE BOOTSTRAP À POIDS

Imaginons que l'on dispose déjà pour chaque dimension d d'une méthode bootstrap à poids adaptée générant les poids

$(w_{k_d}^{d*})_{k_d \in S_d}$ [▶ Détails](#)

MÉTHODE BOOTSTRAP À POIDS

Imaginons que l'on dispose déjà pour chaque dimension d d'une méthode bootstrap à poids adaptée générant les poids

$(w_{k_d}^{d*})_{k_d \in S_d}$. [▶ Détails](#)

On peut alors utiliser l'estimateur bootstrap naturel :

$$\hat{Y}^* = \sum_{\mathbf{k} \in S} y_{\mathbf{k}} \prod_{d=1}^D w_{k_d}^{d*} = \sum_{\mathbf{k} \in S} y_{\mathbf{k}} w_{\mathbf{k}}^*.$$

Cet estimateur vérifie également asymptotiquement les contraintes de moments.

- 1 Introduction
- 2 Propriétés générales
- 3 Propositions de méthodes d'estimation**
 - Constructions générales
 - Estimations classiques
 - Estimations bootstrap
 - Illustrations : plan de sondage simple
 - Estimations classiques
 - Estimations bootstrap
 - Simulations
- 4 Conclusion

PLAN DE SONDAGE SIMPLE À DEUX DIMENSIONS

On tire n_1 unités dans \mathcal{U}_1 et n_2 unités dans \mathcal{U}_2 selon un tirage simple sans remise. L'estimateur de Horvitz-Thompson associé s'écrit sous la forme

$$\hat{Y} = \frac{N_1 N_2}{n_1 n_2} \sum_{k_1 \in \mathcal{S}_1} \sum_{k_2 \in \mathcal{S}_2} y_{k_1 k_2}.$$

PLAN DE SONDAGE SIMPLE À DEUX DIMENSIONS

On tire n_1 unités dans \mathcal{U}_1 et n_2 unités dans \mathcal{U}_2 selon un tirage simple sans remise. L'estimateur de Horvitz-Thompson associé s'écrit sous la forme

$$\hat{Y} = \frac{N_1 N_2}{n_1 n_2} \sum_{k_1 \in \mathcal{S}_1} \sum_{k_2 \in \mathcal{S}_2} y_{k_1 k_2}.$$

On introduit pour la suite les moyennes

$$\forall k_1 \in \mathcal{U}_1, \bar{Y}_{k_1 \bullet} = \frac{1}{N_1} \sum_{k_2 \in \mathcal{U}_2} y_{k_1 k_2} \quad \text{moyenne sur la ligne } k_1$$

$$\forall k_2 \in \mathcal{U}_2, \bar{Y}_{\bullet k_2} = \frac{1}{N_2} \sum_{k_1 \in \mathcal{U}_1} y_{k_1 k_2} \quad \text{moyenne sur la colonne } k_2$$

$$\bar{Y}_{\bullet \bullet} = \frac{1}{N_1 N_2} \sum_{k_1 \in \mathcal{U}_1} \sum_{k_2 \in \mathcal{U}_2} y_{k_1 k_2} \quad \text{moyenne totale}$$

EXPRESSION DE LA VARIANCE

La variance de \hat{Y} peut s'exprimer sous la forme suivante

$$V_p(\hat{Y}) = N_1^2 N_2^2 \left[(1 - f_1) \frac{S_1^2}{n_1} + (1 - f_2) \frac{S_2^2}{n_2} + (1 - f_1)(1 - f_2) \frac{S_{12}^2}{n_1 n_2} \right] \quad (\star)$$

EXPRESSION DE LA VARIANCE

La variance de \hat{Y} peut s'exprimer sous la forme suivante

$$V_p(\hat{Y}) = N_1^2 N_2^2 \left[(1 - f_1) \frac{S_1^2}{n_1} + (1 - f_2) \frac{S_2^2}{n_2} + (1 - f_1)(1 - f_2) \frac{S_{12}^2}{n_1 n_2} \right] \quad (*)$$

avec

$$S_1^2 = \frac{1}{N_1 - 1} \sum_{k_1 \in \mathcal{U}_1} (\bar{Y}_{k_1 \bullet} - \bar{Y}_{\bullet \bullet})^2 \quad \text{dispersion des lignes}$$

$$S_2^2 = \frac{1}{N_2 - 1} \sum_{k_2 \in \mathcal{U}_2} (\bar{Y}_{\bullet k_2} - \bar{Y}_{\bullet \bullet})^2 \quad \text{dispersion des colonnes}$$

$$S_{12}^2 = \frac{1}{N_1 - 1} \frac{1}{N_2 - 1} \sum_{k_1 \in \mathcal{U}_1} \sum_{k_2 \in \mathcal{U}_2} (y_{k_1 k_2} - \bar{Y}_{\bullet k_2} - \bar{Y}_{k_1 \bullet} + \bar{Y}_{\bullet \bullet})^2$$

ESTIMATION SANS BIAIS DE LA VARIANCE

On peut estimer sans biais chacun des termes de variances par

$$\hat{S}_1^2 = \left[\frac{1}{n_1 - 1} \sum_{k_1 \in S_1} (\bar{y}_{k_1 \bullet} - \bar{y}_{\bullet \bullet})^2 \right] - (1 - f_2) \frac{\hat{S}_{12}^2}{n_2}$$

$$\hat{S}_2^2 = \left[\frac{1}{n_2 - 1} \sum_{k_2 \in S_2} (\bar{y}_{\bullet k_2} - \bar{y}_{\bullet \bullet})^2 \right] - (1 - f_1) \frac{\hat{S}_{12}^2}{n_1}$$

$$\hat{S}_{12}^2 = \frac{1}{n_1 - 1} \frac{1}{n_2 - 1} \sum_{k_1 \in S_1} \sum_{k_2 \in S_2} (y_{k_1 k_2} - \bar{y}_{\bullet k_2} - \bar{y}_{k_1 \bullet} + \bar{y}_{\bullet \bullet})^2$$

où les \bar{y} sont des estimateurs de \bar{Y} , par exemple pour $k_1 \in S_1$,

$$\bar{y}_{k_1 \bullet} = \frac{1}{n_2} \sum_{k_2 \in S_2} y_{k_1 k_2}$$

PREMIÈRE SIMPLIFICATION

On peut estimer **asymptotiquement** sans biais chacun des termes de variances par

$$\hat{S}_1^2 = \left[\frac{1}{n_1 - 1} \sum_{k_1 \in S_1} (\bar{y}_{k_1 \bullet} - \bar{y}_{\bullet \bullet})^2 \right] - (1 - f_2) \frac{\hat{S}_{12}^2}{n_2}$$

$$\hat{S}_2^2 = \left[\frac{1}{n_2 - 1} \sum_{k_2 \in S_2} (\bar{y}_{\bullet k_2} - \bar{y}_{\bullet \bullet})^2 \right] - (1 - f_1) \frac{\hat{S}_{12}^2}{n_1}$$

~~$$\hat{S}_{12}^2 = \frac{1}{n_1 - 1} \frac{1}{n_2 - 1} \sum_{k_1 \in S_1} \sum_{k_2 \in S_2} (y_{k_1 k_2} - \bar{y}_{\bullet k_2} - \bar{y}_{k_1 \bullet} + \bar{y}_{\bullet \bullet})^2$$~~

où les \bar{y} sont des estimateurs de \bar{Y} , par exemple pour $k_1 \in S_1$,

$$\bar{y}_{k_1 \bullet} = \frac{1}{n_2} \sum_{k_2 \in S_2} y_{k_1 k_2}$$

DEUXIÈME SIMPLIFICATION

On peut estimer **asymptotiquement** sans biais chacun des termes de variances par

$$\hat{S}_1^2 = \left[\frac{1}{n_1 - 1} \sum_{k_1 \in S_1} (\bar{y}_{k_1 \bullet} - \bar{y}_{\bullet \bullet})^2 \right] \cancel{-(1 - f_2) \frac{\hat{S}_{12}^2}{n_2}}$$

$$\hat{S}_2^2 = \left[\frac{1}{n_2 - 1} \sum_{k_2 \in S_2} (\bar{y}_{\bullet k_2} - \bar{y}_{\bullet \bullet})^2 \right] \cancel{-(1 - f_1) \frac{\hat{S}_{12}^2}{n_1}}$$

$$\hat{S}_{12}^2 = \cancel{\frac{1}{n_1 - 1} \frac{1}{n_2 - 1} \sum_{k_1 \in S_1} \sum_{k_2 \in S_2} (y_{k_1 k_2} - \bar{y}_{\bullet k_2} - \bar{y}_{k_1 \bullet} + \bar{y}_{\bullet \bullet})^2}$$

où les \bar{y} sont des estimateurs de \bar{Y} , par exemple pour $k_1 \in S_1$,

$$\bar{y}_{k_1 \bullet} = \frac{1}{n_2} \sum_{k_2 \in S_2} y_{k_1 k_2}$$

MÉTHODE BOOTSTRAP

On considère la méthode de Rao et Wu. Il s'agit d'une méthode bootstrap pour le tirage simple sans remise de taille n . On génère des poids bootstrap $(w_k^*)_{k \in S}$ de la forme

$$w_k^* = \left[1 + \sqrt{\frac{n^*(1-f)}{n-1}} \left(\frac{nm_k^*}{n^*} - 1 \right) \right] \frac{N}{n}$$

où les variables $(m_k^*)_{k \in S}$ ont été générées suivant une loi multinomiale $\mathcal{M}(n^*; \frac{1}{n}, \dots, \frac{1}{n})$ avec n^* un paramètre choisi.

MÉTHODE BOOTSTRAP

Pour le plan produit, $S = S_1 \times S_2$. On génère des poids $(w_{k_1}^{(1*)})_{k_1 \in S_1}$ et $(w_{k_2}^{(2*)})_{k_2 \in S_2}$ selon la méthode de Rao et Wu et on construit les poids

$$\forall k_1 \in S_1, k_2 \in S_2, w_{k_1 k_2}^* = w_{k_1}^{(1*)} w_{k_2}^{(2*)}$$

pour construire l'estimateur bootstrap

$$\hat{Y}^* = \sum_{k_1 \in S_1} \sum_{k_2 \in S_2} y_{k_1 k_2} w_{k_1 k_2}^*$$

CONTRAINTES DE MOMENTS

Cet estimateur bootstrap vérifie d'une part la contrainte de moment d'ordre 1 :

$$E_*(\hat{Y}^*) = \hat{Y}$$

CONTRAINTES DE MOMENTS

D'autres part, sa variance $V_*(\hat{Y}^*)$ s'exprime sous la forme (\star).

CONTRAINTES DE MOMENTS

D'autres part, sa variance $V_*(\hat{Y}^*)$ s'exprime sous la forme (*).
On peut montrer que

$$S_1^{2*} = \hat{S}_1^2 + (1 - f_2) \frac{\hat{S}_{12}^2}{n_2},$$

$$S_2^{2*} = \hat{S}_2^2 + (1 - f_1) \frac{\hat{S}_{12}^2}{n_1},$$

$$S_{12}^{2*} = \hat{S}_{12}^2.$$

ainsi, lorsque n_1, n_2 sont suffisamment grands, la variance bootstrap est asymptotiquement sans biais.

- ① Introduction
- ② Propriétés générales
- ③ Propositions de méthodes d'estimation
 - Constructions générales
 - Estimations classiques
 - Estimations bootstrap
 - Illustrations : plan de sondage simple
 - Estimations classiques
 - Estimations bootstrap
 - Simulations
- ④ Conclusion

SIMULATIONS

Modèle de génération

Pour $k_1 \in \mathcal{U}_1, k_2 \in \mathcal{U}_2$, on génère une population avec $N_1 \times N_2 = 1000 \times 1000$ variables d'intérêts :

$$y_{k_1 k_2} = \mu + \sigma_1 u_{k_1} + \sigma_2 v_{k_2} + \sigma_{12} w_{k_1 k_2}$$

avec $u_{k_1}, v_{k_2}, w_{k_1 k_2} \sim \mathcal{N}(0, 1), \mu = 200, \sigma_1 = \sigma_2 = \sigma_{12} = 5$.

Table: Biais relatifs (en pourcentage) d'estimateurs de $V_p(\hat{Y})$

| n_1 | n_2 | RaoWu | Skinner | Simp1 | Simp2 |
|-------|-------|-------|---------|-------|-------|
| 5 | 5 | 16 | 21 | -12 | 10 |
| 10 | 10 | 8 | 11 | -4 | 6 |
| 10 | 100 | 2 | 5 | 0.1 | 2 |
| 100 | 100 | 0.6 | 12 | -0.8 | 0.2 |
| 500 | 500 | -0.6 | 99 | -0.8 | 0.7 |

CONCLUSION

CONCLUSION

Le plan produit est composé de plusieurs sous-plans.

CONCLUSION

Le plan produit est composé de plusieurs sous-plans.

On peut produire des estimations asymptotiquement sans biais en se concentrant sur une dimension à la fois.

ANNEXE

SOMMES PARTIELLES

Pour $I \subseteq \{1, \dots, D\}$, on introduit les populations partielles $\mathcal{U}_I = \prod_{d \in I} \mathcal{U}_d$, et les sommes partielles pour $\mathbf{k}' \in \mathcal{U}_I$,

$$Y_{\mathbf{k}'} = \sum_{\substack{\mathbf{l} \in \mathcal{U} \\ \forall d \in I, l_d = k'_d}} y_{\mathbf{l}}$$

DÉCOMPOSITION DE Hoeffding-SOBOL

Formules explicites

Pour $I \subseteq \{1, \dots, D\}$, \hat{Y}^I est donné par

$$\hat{Y}^I = \sum_{\mathbf{k} \in \mathcal{U}} y_{\mathbf{k}} \prod_{d \in I} \left(\frac{\delta_{k_d}^d}{\pi_{k_d}^d} - 1 \right).$$

Ainsi, on a également

$$V_p(\hat{Y}^I) = \sum_{\mathbf{k}', I' \in \mathcal{U}_I} \frac{Y_{\mathbf{k}'} Y_{I'}}{\pi_{\mathbf{k}'} \pi_{I'}} \Delta_{\mathbf{k}' I'}$$

avec $\Delta_{\mathbf{k}' I'} = \prod_{d \in I} \Delta_{k'_d I'_d}^d$.

ORDRE DE GRANDEUR DES TERMES DE VARIANCES

Hypothèses classiques

Pour chaque $d = 1, \dots, D$,

1. $n_d \rightarrow \infty$ et $N_d \rightarrow \infty$, et il existe une constante $f_d \in [0, 1]$ telle que

$$\frac{n_d}{N_d} \rightarrow f_d.$$

2. Il existe une constante $\lambda_d > 0$ telle que

$$\forall k \in \mathcal{U}_d, \pi_k^d \geq \lambda_d \frac{n_d}{N_d}.$$

3. Il existe une constante γ_d telle que

$$\forall k \neq l \in \mathcal{U}_d, |\Delta_{k,l}^d| \leq \gamma_d \frac{n_d}{N_d^2}.$$

ESTIMATEUR PLUG-IN

L'estimateur plug-in de $V_p(\hat{Y}^I)$ est donné par l'expression

$$\hat{V}^{PLUG}(\hat{Y}^I) = \sum_{\mathbf{k}', I' \in \mathcal{S}_I} \frac{\hat{Y}_{\mathbf{k}'} \hat{Y}_{I'}}{\pi_{\mathbf{k}'} \pi_{I'}} \frac{\Delta_{\mathbf{k}' I'}}{\pi_{\mathbf{k}' I'}}$$

avec $\hat{Y}_{\mathbf{k}'}$ un estimateur sans biais de $Y_{\mathbf{k}'}$:

$$\hat{Y}_{\mathbf{k}'} = \sum_{\substack{I \in \mathcal{S} \\ \forall d \in I, l_d = k'_d}} \frac{y_I}{\prod_{d \notin I} \pi_{l_d}^d}.$$

On peut vérifier qu'il est bien asymptotiquement sans biais :

$$E_p \left[\hat{V}^{PLUG}(\hat{Y}^I) \right] = \sum_{I' \supseteq I} V_p(\hat{Y}^{I'})$$

HYPOTHÈSES POUR LA MÉTHODE BOOTSTRAP À POIDS

Hypothèses

On dispose pour $d = 1, \dots, D$ de poids bootstrap $(w_{k_d}^{d*})$ vérifiant les contraintes de moments.

$$\forall k_d \in S_d, E_*(w_{k_d}^{d*}) = \frac{\delta_{k_d}^d}{\pi_{k_d}^d}$$

$$\forall k_d, l_d \in S_d, Cov_*(w_{k_d}^{d*}, w_{l_d}^{d*}) = \frac{\Delta_{k_d, l_d}^d}{\pi_{k_d, l_d}^d} \frac{\delta_{k_d}^d \delta_{l_d}^d}{\pi_{k_d}^d \pi_{l_d}^d}$$