

Quelques résultats asymptotiques pour le tirage systématique de Deville

Guillaume Chauvet

Travail joint avec Jean-Claude Deville
(Ensaï)

42èmes Journées de Statistique
Marseille, 27/05/2010



En résumé

Le tirage systématique de Deville est un algorithme permettant de sélectionner des échantillons de taille fixe, avec des probabilités d'inclusion inégales fixées.

En résumé

Le tirage systématique de Deville est un algorithme permettant de sélectionner des échantillons de taille fixe, avec des probabilités d'inclusion inégales fixées.

Les probabilités d'inclusion d'ordre 2 sont calculables exactement, mais leur forme est complexe.

En résumé

Le tirage systématique de Deville est un algorithme permettant de sélectionner des échantillons de taille fixe, avec des probabilités d'inclusion inégales fixées.

Les probabilités d'inclusion d'ordre 2 sont calculables exactement, mais leur forme est complexe.

Nous montrons que, sous certaines conditions, l'algorithme proposé par Deville est proche d'un algorithme de tirage de type stratifié.

En résumé

Le tirage systématique de Deville est un algorithme permettant de sélectionner des échantillons de taille fixe, avec des probabilités d'inclusion inégales fixées.

Les probabilités d'inclusion d'ordre 2 sont calculables exactement, mais leur forme est complexe.

Nous montrons que, sous certaines conditions, l'algorithme proposé par Deville est proche d'un algorithme de tirage de type stratifié. Cette comparaison permet d'obtenir une approximation de variance, et de montrer la normalité asymptotique du π -estimateur.

Le tirage systématique de Deville

Propriétés de l'algorithme

Etude par simulations

Introduction

Cadre

On considère une population finie d'individus

$$U = \{1, \dots, k, \dots, N\},$$

où chaque individu est supposé identifiable par son label k .

Cadre

On considère une population finie d'individus

$$U = \{1, \dots, k, \dots, N\},$$

où chaque individu est supposé identifiable par son label k . Soit y_k la valeur prise par la variable d'intérêt y sur l'individu k de U , et π_k sa probabilité d'inclusion avec

$$\sum_{k \in U} \pi_k = n.$$

Cadre

On considère une population finie d'individus

$$U = \{1, \dots, k, \dots, N\},$$

où chaque individu est supposé identifiable par son label k . Soit y_k la valeur prise par la variable d'intérêt y sur l'individu k de U , et π_k sa probabilité d'inclusion avec

$$\sum_{k \in U} \pi_k = n.$$

Le total $Y = \sum_{k \in U} y_k$ de la variable y est estimé sans biais par le π -estimateur

$$\hat{Y} = \sum_{k \in S} \frac{y_k}{\pi_k}.$$

On note

$$V_k = \sum_{l=1}^k \pi_l \text{ pour tout } k \in U \text{ avec } V_0 = 0.$$

On note

$$V_k = \sum_{l=1}^k \pi_l \text{ pour tout } k \in U \text{ avec } V_0 = 0.$$

Une unité k sera dite *frontalière* si pour un entier i

$$V_{k-1} < i \leq V_k.$$

On note

$$V_k = \sum_{l=1}^k \pi_l \text{ pour tout } k \in U \text{ avec } V_0 = 0.$$

Une unité k sera dite *frontalière* si pour un entier i

$$V_{k-1} < i \leq V_k.$$

On note k_i , $i = 0, \dots, n$ les unités frontalières (avec $k_0 = 1$, $k_n = N$). On note également

$$\begin{aligned} \pi_{k_i} &= V_{k_i} && - && V_{k_{i-1}} \\ &= [i - V_{k_{i-1}}] && + && [V_{k_i} - i] \\ &= a_i && + && b_i. \end{aligned}$$

On note

$$V_k = \sum_{l=1}^k \pi_l \text{ pour tout } k \in U \text{ avec } V_0 = 0.$$

Une unité k sera dite *frontalière* si pour un entier i

$$V_{k-1} < i \leq V_k.$$

On note k_i , $i = 0, \dots, n$ les unités frontalières (avec $k_0 = 1$, $k_n = N$). On note également

$$\begin{aligned} \pi_{k_i} &= V_{k_i} && - && V_{k_{i-1}} \\ &= [i - V_{k_{i-1}}] && + && [V_{k_i} - i] \\ &= a_i && + && b_i. \end{aligned}$$

La *microstrate* U_i est constituée des unités $k_{i-1} \leq k \leq k_i$.

Exemple

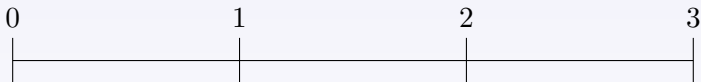
Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$

Exemple

Population U de taille $N = 8$, avec $n = 3$ et

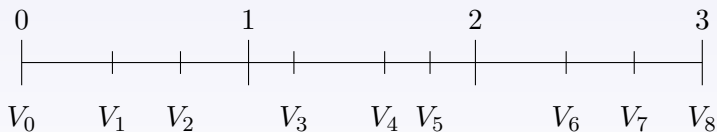
$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



Exemple

Population U de taille $N = 8$, avec $n = 3$ et

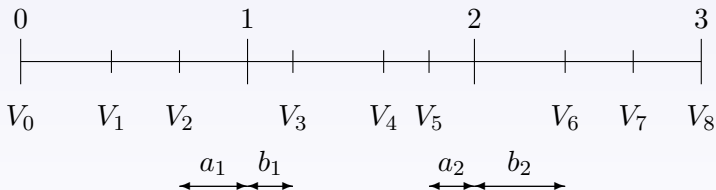
$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



Exemple

Population U de taille $N = 8$, avec $n = 3$ et

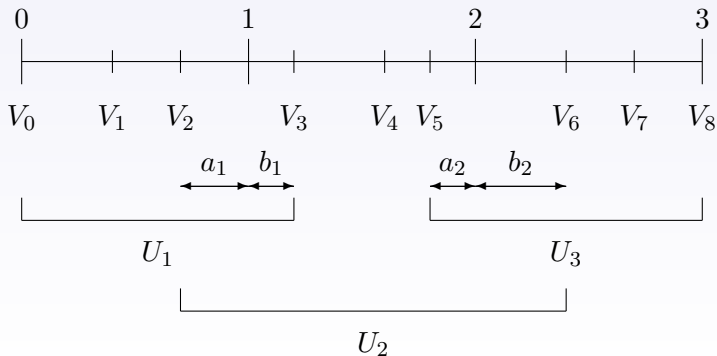
$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



Exemple

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Etape 1:

Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Etape 1:

- ▶ On génère $u_1 \sim \mathcal{U}[0, 1]$,

Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Etape 1:

- ▶ On génère $u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Etape 1:

- ▶ On génère $u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Etape 1:

- ▶ On génère $u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

- ▶ On génère

$$u_i \sim \begin{cases} \mathcal{U}[b_{i-1}, 1] & \text{si } k_{i-1} \text{ a été tirée à l'étape } i-1, \\ \mathcal{U}[0, b_{i-1}] \wedge \mathcal{U}[0, 1] & \text{sinon.} \end{cases}$$

Méthode de Deville

On réalise n tirages, un dans chaque microstrate. Les tirages sont coordonnés de façon à ce qu'aucune unité frontalière ne soit sélectionnée deux fois.

Etape 1:

- ▶ On génère $u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

- ▶ On génère

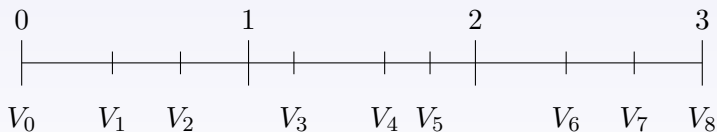
$$u_i \sim \begin{cases} \mathcal{U}[b_{i-1}, 1] & \text{si } k_{i-1} \text{ a été tirée à l'étape } i-1, \\ \mathcal{U}[0, b_{i-1}] \wedge \mathcal{U}[0, 1] & \text{sinon.} \end{cases}$$

- ▶ On sélectionne l'unité k telle que $V_{k-1} \leq u_i + (i-1) < V_k$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$

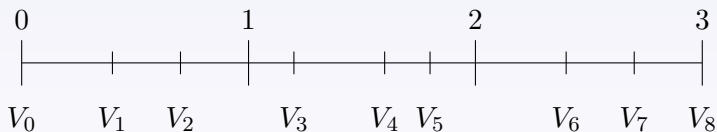


Etape 1:

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



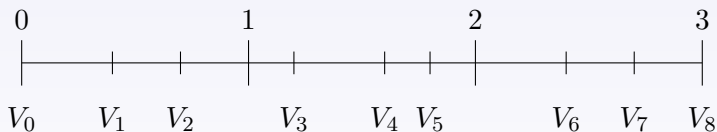
Etape 1:

- Tirage dans la 1ère microstrate,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



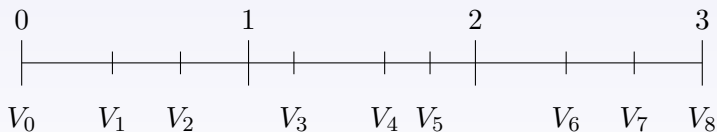
Etape 1:

- ▶ Tirage dans la 1ère microstrate,
- ▶ On génère $u_1 \sim \mathcal{U}[0, 1] \Rightarrow u_1 = 0.65$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



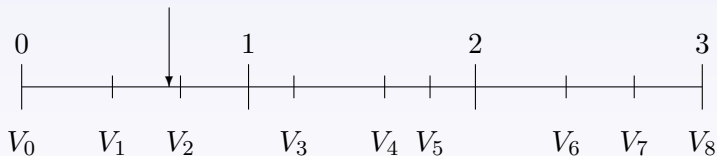
Etape 1:

- ▶ Tirage dans la 1ère microstrate,
- ▶ On génère $u_1 \sim \mathcal{U}[0, 1] \Rightarrow u_1 = 0.65$,
- ▶ On sélectionne l'unité 2.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



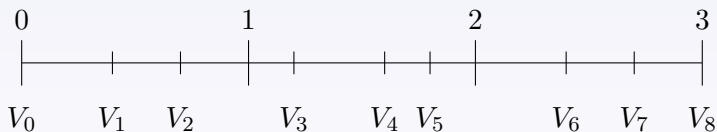
Etape 1:

- ▶ Tirage dans la 1ère microstrate,
- ▶ On génère $u_1 \sim \mathcal{U}[0, 1] \Rightarrow u_1 = 0.65$,
- ▶ On sélectionne l'unité 2.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$

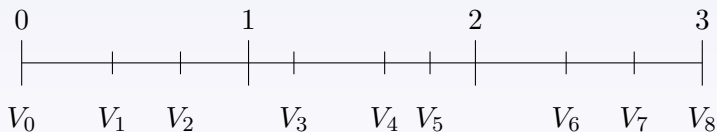


Etape 2:

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



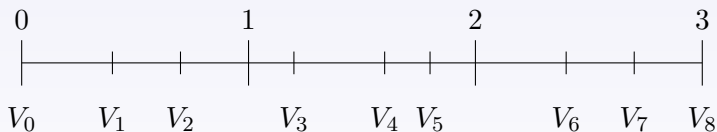
Etape 2:

- ▶ L'unité $k_1 = 3$ n'a pas été retenue,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



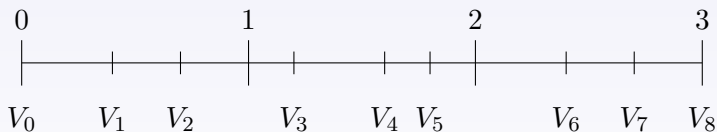
Etape 2:

- ▶ L'unité $k_1 = 3$ n'a pas été retenue,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge \mathcal{U}[0, 1] \Rightarrow u_2 = 0.92$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



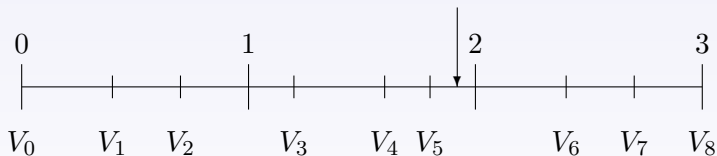
Etape 2:

- ▶ L'unité $k_1 = 3$ n'a pas été retenue,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge \mathcal{U}[0, 1] \Rightarrow u_2 = 0.92$,
- ▶ On sélectionne l'unité 6.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



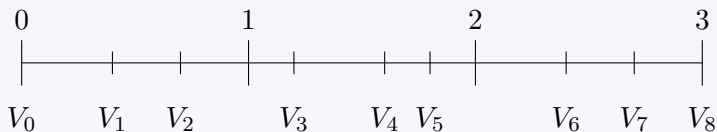
Etape 2:

- ▶ L'unité $k_1 = 3$ n'a pas été retenue,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge \mathcal{U}[0, 1] \Rightarrow u_2 = 0.92$,
- ▶ On sélectionne l'unité 6.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$

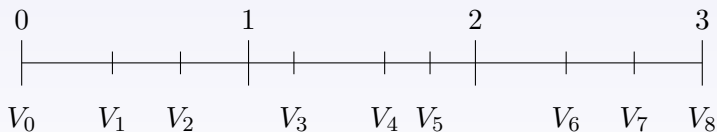


Etape 3:

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



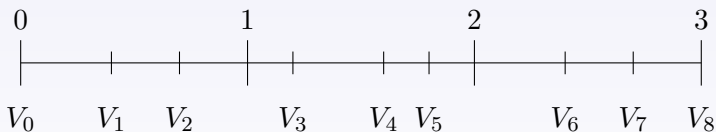
Etape 3:

- ▶ L'unité $k_2 = 6$ a été retenue,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



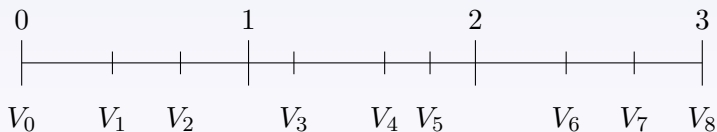
Etape 3:

- ▶ L'unité $k_2 = 6$ a été retenue,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



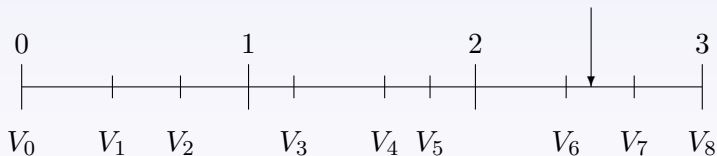
Etape 3:

- ▶ L'unité $k_2 = 6$ a été retenue,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,
- ▶ On sélectionne l'unité 7.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



Etape 3:

- ▶ L'unité $k_2 = 6$ a été retenue,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,
- ▶ On sélectionne l'unité 7.

Propriétés de l'algorithme

L'algorithme est de taille fixe par construction, et permet de respecter exactement les probabilités d'inclusion voulues.

L'algorithme est de taille fixe par construction, et permet de respecter exactement les probabilités d'inclusion voulues. Deville (1998) obtient également une forme explicite pour les probabilités d'inclusion d'ordre deux, et montre que les conditions de Yates-Grundy sont satisfaites.

L'algorithme est de taille fixe par construction, et permet de respecter exactement les probabilités d'inclusion voulues. Deville (1998) obtient également une forme explicite pour les probabilités d'inclusion d'ordre deux, et montre que les conditions de Yates-Grundy sont satisfaites.

Avantages/inconvénients de la méthode :

- Beaucoup de probabilités π_{kl} sont nulles \Rightarrow pas d'estimateur sans biais de variance,

L'algorithme est de taille fixe par construction, et permet de respecter exactement les probabilités d'inclusion voulues. Deville (1998) obtient également une forme explicite pour les probabilités d'inclusion d'ordre deux, et montre que les conditions de Yates-Grundy sont satisfaites.

Avantages/inconvénients de la méthode :

- Beaucoup de probabilités π_{kl} sont nulles \Rightarrow pas d'estimateur sans biais de variance,
- + Effet de stratification \Rightarrow variance réduite si la variable y est liée à l'ordre du fichier.

L'algorithme est de taille fixe par construction, et permet de respecter exactement les probabilités d'inclusion voulues. Deville (1998) obtient également une forme explicite pour les probabilités d'inclusion d'ordre deux, et montre que les conditions de Yates-Grundy sont satisfaites.

Avantages/inconvénients de la méthode :

- Beaucoup de probabilités π_{kl} sont nulles \Rightarrow pas d'estimateur sans biais de variance,
- + Effet de stratification \Rightarrow variance réduite si la variable y est liée à l'ordre du fichier.

Si les micro-strates ne se chevauchent pas, l'algorithme est équivalent au tirage stratifié de taille 1 dans chaque microstrate.

Approximation de l'algorithme

Intuitivement, l'algorithme est proche d'un tirage stratifié de taille 1 dans chaque microstrate si les probabilités d'inclusion des unités frontalières sont faibles.

Approximation de l'algorithme

Intuitivement, l'algorithme est proche d'un tirage stratifié de taille 1 dans chaque microstrate si les probabilités d'inclusion des unités frontalières sont faibles.

Pour le montrer, on définit une procédure conjointe de tirage d'un échantillon systématique de Deville ($S^{(sd)}$) et d'un échantillon "stratifié" ($S^{(st)}$).

Approximation de l'algorithme

Intuitivement, l'algorithme est proche d'un tirage stratifié de taille 1 dans chaque microstrate si les probabilités d'inclusion des unités frontalières sont faibles.

Pour le montrer, on définit une procédure conjointe de tirage d'un échantillon systématique de Deville ($S^{(sd)}$) et d'un échantillon "stratifié" ($S^{(st)}$).

On note

$$\hat{Y}^{(sd)} = \sum_{k \in S^{(sd)}} \frac{y_k}{\pi_k} \text{ et } \hat{Y}^{(st)} = \sum_{k \in S^{(st)}} \frac{y_k}{\pi_k}.$$

Procédure de tirage conjointe

Etape 1:

Procédure de tirage conjointe

Etape 1:

- ▶ On génère $v_1 = u_1 \sim \mathcal{U}[0, 1]$,

Procédure de tirage conjointe

Etape 1:

- ▶ On génère $v_1 = u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne dans $S^{(sd)}$ et $S^{(st)}$ l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Procédure de tirage conjointe

Etape 1:

- ▶ On génère $v_1 = u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne dans $S^{(sd)}$ et $S^{(st)}$ l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

Procédure de tirage conjointe

Etape 1:

- ▶ On génère $v_1 = u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne dans $S^{(sd)}$ et $S^{(st)}$ l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

- ▶ On génère $v_i \sim \mathcal{U}[0, 1]$, et on sélectionne dans $S^{(st)}$ l'unité k telle que $V_{k-1} \leq v_i + (i - 1) < V_k$.

Procédure de tirage conjointe

Etape 1:

- ▶ On génère $v_1 = u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne dans $S^{(sd)}$ et $S^{(st)}$ l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

- ▶ On génère $v_i \sim \mathcal{U}[0, 1]$, et on sélectionne dans $S^{(st)}$ l'unité k telle que $V_{k-1} \leq v_i + (i - 1) < V_k$.
- ▶ On génère

$$u_i \sim \begin{cases} \mathcal{U}[b_{i-1}, 1] & \text{si } k_{i-1} \text{ a été tirée à l'étape } i - 1, \\ \mathcal{U}[0, b_{i-1}] \wedge v_i & \text{sinon .} \end{cases}$$

Procédure de tirage conjointe

Etape 1:

- ▶ On génère $v_1 = u_1 \sim \mathcal{U}[0, 1]$,
- ▶ On sélectionne dans $S^{(sd)}$ et $S^{(st)}$ l'unité k telle que $V_{k-1} \leq u_1 < V_k$.

Etape i :

- ▶ On génère $v_i \sim \mathcal{U}[0, 1]$, et on sélectionne dans $S^{(st)}$ l'unité k telle que $V_{k-1} \leq v_i + (i - 1) < V_k$.
- ▶ On génère

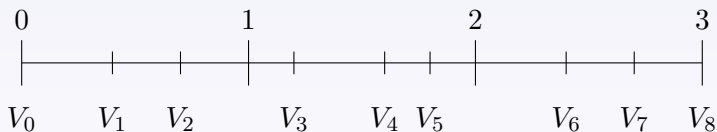
$$u_i \sim \begin{cases} \mathcal{U}[b_{i-1}, 1] & \text{si } k_{i-1} \text{ a été tirée à l'étape } i - 1, \\ \mathcal{U}[0, b_{i-1}] \wedge v_i & \text{sinon .} \end{cases}$$

- ▶ On sélectionne dans $S^{(sd)}$ l'unité k telle que $V_{k-1} \leq u_i + (i - 1) < V_k$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$

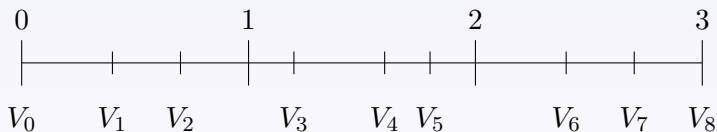


Etape 1:

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



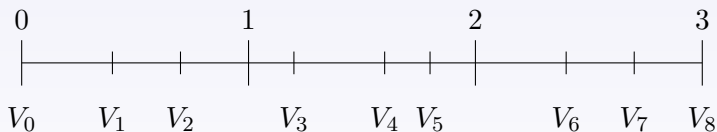
Etape 1:

- ▶ Tirage dans la 1ère microstrate,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



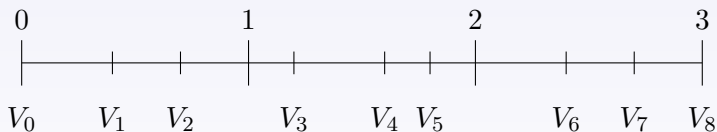
Etape 1:

- ▶ Tirage dans la 1ère microstrate,
- ▶ On génère $u_1 = v_1 \sim \mathcal{U}[0, 1] \Rightarrow u_1 = 0.65$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



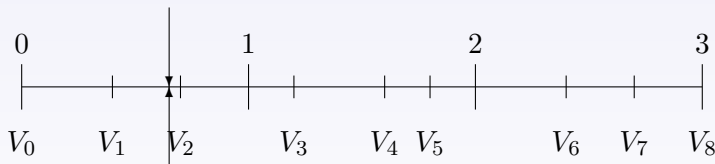
Etape 1:

- ▶ Tirage dans la 1ère microstrate,
- ▶ On génère $u_1 = v_1 \sim \mathcal{U}[0, 1] \Rightarrow u_1 = 0.65$,
- ▶ On sélectionne l'unité 2 dans les échantillons $S^{(st)}$ et $S^{(sd)}$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



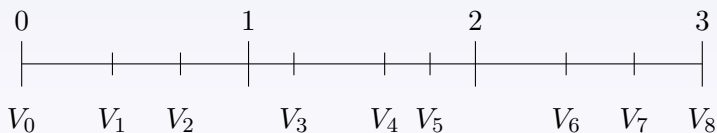
Etape 1:

- ▶ Tirage dans la 1ère microstrate,
- ▶ On génère $u_1 = v_1 \sim \mathcal{U}[0, 1] \Rightarrow u_1 = 0.65$,
- ▶ On sélectionne l'unité 2 dans les échantillons $S^{(st)}$ et $S^{(sd)}$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$

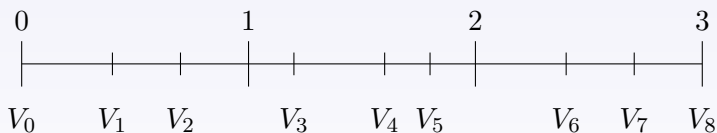


Etape 2:

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



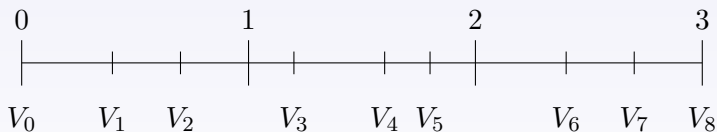
Etape 2:

- ▶ On génère $v_2 \sim \mathcal{U}[0, 1] \Rightarrow v_2 = 0.92$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



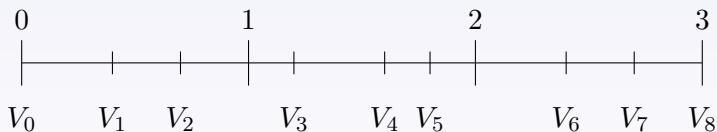
Etape 2:

- ▶ On génère $v_2 \sim \mathcal{U}[0, 1] \Rightarrow v_2 = 0.92$,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge v_2 \Rightarrow u_2 = v_2 = 0.92$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



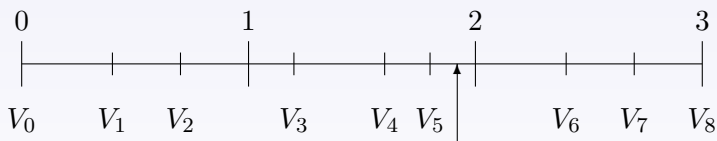
Etape 2:

- ▶ On génère $v_2 \sim \mathcal{U}[0, 1] \Rightarrow v_2 = 0.92$,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge v_2 \Rightarrow u_2 = v_2 = 0.92$,
- ▶ On sélectionne l'unité 6 dans les échantillons $S^{(st)}$ et $S^{(sd)}$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



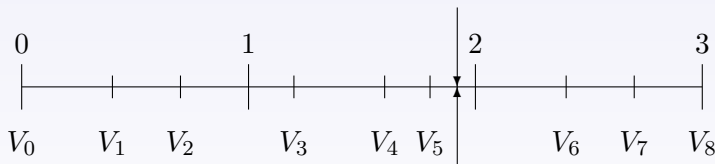
Etape 2:

- ▶ On génère $v_2 \sim \mathcal{U}[0, 1] \Rightarrow v_2 = 0.92$,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge v_2 \Rightarrow u_2 = v_2 = 0.92$,
- ▶ On sélectionne l'unité 6 dans les échantillons $S^{(st)}$ et $S^{(sd)}$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



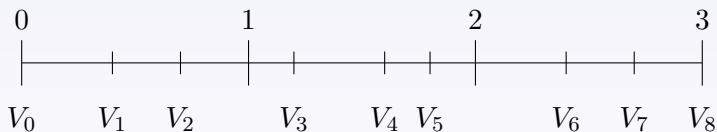
Etape 2:

- ▶ On génère $v_2 \sim \mathcal{U}[0, 1] \Rightarrow v_2 = 0.92$,
- ▶ On génère $u_2 \sim \mathcal{U}[0, 0.2] \wedge v_2 \Rightarrow u_2 = v_2 = 0.92$,
- ▶ On sélectionne l'unité 6 dans les échantillons $S^{(st)}$ et $S^{(sd)}$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$

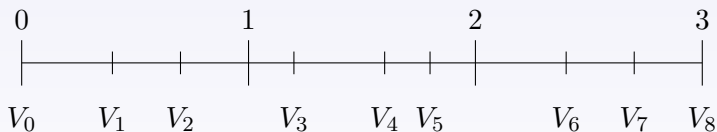


Etape 3:

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



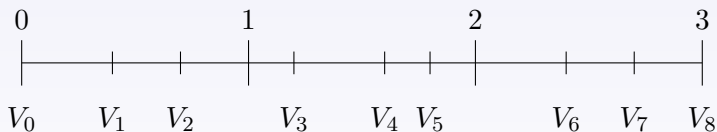
Etape 3:

- ▶ On génère $v_3 \sim \mathcal{U}[0, 1] \Rightarrow v_3 = 0.25$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



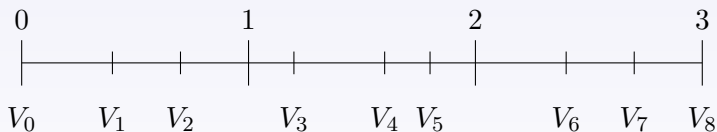
Etape 3:

- ▶ On génère $v_3 \sim \mathcal{U}[0, 1] \Rightarrow v_3 = 0.25$,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



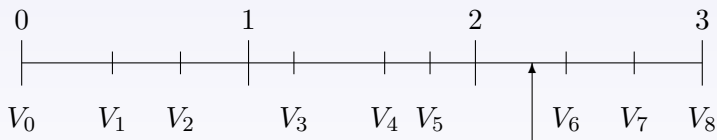
Etape 3:

- ▶ On génère $v_3 \sim \mathcal{U}[0, 1] \Rightarrow v_3 = 0.25$,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,
- ▶ On sélectionne l'unité 6 dans $S^{(st)}$

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



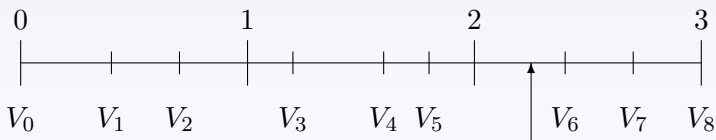
Etape 3:

- ▶ On génère $v_3 \sim \mathcal{U}[0, 1] \Rightarrow v_3 = 0.25$,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,
- ▶ On sélectionne l'unité 6 dans $S^{(st)}$

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T .$$



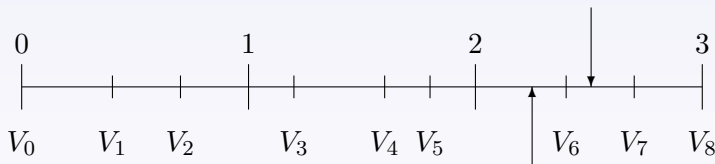
Etape 3:

- ▶ On génère $v_3 \sim \mathcal{U}[0, 1] \Rightarrow v_3 = 0.25$,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,
- ▶ On sélectionne l'unité 6 dans $S^{(st)}$ et la 7 dans $S^{(sd)}$.

Exemple (suite)

Population U de taille $N = 8$, avec $n = 3$ et

$$\boldsymbol{\pi} = [0.4 \ 0.3 \ 0.5 \ 0.4 \ 0.2 \ 0.6 \ 0.3 \ 0.3]^T.$$



Etape 3:

- ▶ On génère $v_3 \sim \mathcal{U}[0, 1] \Rightarrow v_3 = 0.25$,
- ▶ On génère $u_3 \sim \mathcal{U}[0.4, 1] \Rightarrow u_3 = 0.51$,
- ▶ On sélectionne l'unité 6 dans $S^{(st)}$ et la 7 dans $S^{(sd)}$.

Résultats Obtenus

On note $\hat{Y}^{(sd)} = \sum_{i=1}^n \hat{Y}_i^{(sd)}$ et $\hat{Y}^{(st)} = \sum_{i=1}^n \hat{Y}_i^{(st)}$.

Résultats Obtenus

On note $\hat{Y}^{(sd)} = \sum_{i=1}^n \hat{Y}_i^{(sd)}$ et $\hat{Y}^{(st)} = \sum_{i=1}^n \hat{Y}_i^{(st)}$.

On a

$$\begin{aligned} E[\hat{Y}^{(sd)}] &= E[\hat{Y}^{(st)}] \\ &= \sum_{i=1}^n E[\hat{Y}_i^{(st)}] \\ &= \sum_{i=1}^n Y_i = Y. \end{aligned}$$

Résultats Obtenus

On note $\hat{Y}^{(sd)} = \sum_{i=1}^n \hat{Y}_i^{(sd)}$ et $\hat{Y}^{(st)} = \sum_{i=1}^n \hat{Y}_i^{(st)}$.

On a

$$\begin{aligned} E[\hat{Y}^{(sd)}] &= E[\hat{Y}^{(st)}] \\ &= \sum_{i=1}^n E[\hat{Y}_i^{(st)}] \\ &= \sum_{i=1}^n Y_i = Y. \end{aligned}$$

et

$$V[\hat{Y}^{(st)}] = \sum_{i=1}^n V[\hat{Y}_i^{(st)}].$$

Résultats Obtenus

Propriété

On suppose que

Résultats Obtenus

Propriété

On suppose que

$$C1 : \max_i \pi_{k_i} = O(n^{-1/2}),$$

Résultats Obtenus

Propriété

On suppose que

$$C1 : \max_i \pi_{k_i} = O(n^{-1/2}),$$

$$C2 : \max_i \left[\frac{a_i \left(\frac{y_{k_i}}{\pi_{k_i}} - Y_i \right)^2}{V[\hat{Y}_i^{(st)}]} \right] \rightarrow 0 \quad \max_i \left[\frac{b_i \left(\frac{y_{k_i}}{\pi_{k_i}} - Y_{i+1} \right)^2}{V[\hat{Y}_{i+1}^{(st)}]} \right] \rightarrow 0.$$

Résultats Obtenus

Propriété

On suppose que

$$C1 : \max_i \pi_{k_i} = O(n^{-1/2}),$$

$$C2 : \max_i \left[\frac{a_i \left(\frac{y_{k_i}}{\pi_{k_i}} - Y_i \right)^2}{V[\hat{Y}_i^{(st)}]} \right] \rightarrow 0 \quad \max_i \left[\frac{b_i \left(\frac{y_{k_i}}{\pi_{k_i}} - Y_{i+1} \right)^2}{V[\hat{Y}_{i+1}^{(st)}]} \right] \rightarrow 0.$$

Alors :

$$\frac{E \left[\hat{Y}^{(sd)} - \hat{Y}^{(st)} \right]^2}{V \left[\hat{Y}^{(st)} \right]} \rightarrow 0.$$

Etude par simulations

Cadre

Deux populations, de taille $N = 23$ (respectivement, $N = 43$). Trois jeux de probabilités d'inclusion sont générés (constantes, loi gamma, loi uniforme). Pour chaque jeu de probabilités, échantillon de taille $n = 3, 5$ ou 7 .

Cadre

Deux populations, de taille $N = 23$ (respectivement, $N = 43$). Trois jeux de probabilités d'inclusion sont générés (constantes, loi gamma, loi uniforme). Pour chaque jeu de probabilités, échantillon de taille $n = 3, 5$ ou 7 .

On compare la vraie matrice de variance-covariance Δ à l'approximation proposée Δ_{app} , correspondant à un tirage "stratifié".

Cadre

Deux populations, de taille $N = 23$ (respectivement, $N = 43$). Trois jeux de probabilités d'inclusion sont générés (constantes, loi gamma, loi uniforme). Pour chaque jeu de probabilités, échantillon de taille $n = 3, 5$ ou 7 .

On compare la vraie matrice de variance-covariance Δ à l'approximation proposée Δ_{app} , correspondant à un tirage "stratifié".

Dans la décomposition $\Delta_{app} \mathbf{z} = \lambda \Delta \mathbf{z}$, les valeurs propres $\lambda_{min}, \lambda_{max} > 0$ représentent la situation où l'approximation de variance est la pire (Deville et Tillé, 2005).

Résultats obtenus

Pop de taille $N = 23$

	EGAL		GAMMA		UNIFORM	
n=3	0,91	1,11	0,91	1,10	0,90	1,11
n=5	0,77	1,32	0,63	1,60	0,71	1,40
n=7	0,64	1,56	0,53	1,88	0,63	1,61

Résultats obtenus

Pop de taille $N = 23$

	EGAL		GAMMA		UNIFORM	
n=3	0,91	1,11	0,91	1,10	0,90	1,11
n=5	0,77	1,32	0,63	1,60	0,71	1,40
n=7	0,64	1,56	0,53	1,88	0,63	1,61

Résultats obtenus

Pop de taille $N = 23$

	EGAL		GAMMA		UNIFORM	
n=3	0,91	1,11	0,91	1,10	0,90	1,11
n=5	0,77	1,32	0,63	1,60	0,71	1,40
n=7	0,64	1,56	0,53	1,88	0,63	1,61

Pop de taille $N = 43$

	EGAL		GAMMA		UNIFORM	
n=3	0,95	1,06	0,90	1,11	0,81	1,24
n=5	0,84	1,19	0,84	1,19	0,74	1,36
n=7	0,82	1,22	0,67	1,47	0,64	1,55

Résultats obtenus

Pop de taille $N = 23$

	EGAL		GAMMA		UNIFORM	
n=3	0,91	1,11	0,91	1,10	0,90	1,11
n=5	0,77	1,32	0,63	1,60	0,71	1,40
n=7	0,64	1,56	0,53	1,88	0,63	1,61

Pop de taille $N = 43$

	EGAL		GAMMA		UNIFORM	
n=3	0,95	1,06	0,90	1,11	0,81	1,24
n=5	0,84	1,19	0,84	1,19	0,74	1,36
n=7	0,82	1,22	0,67	1,47	0,64	1,55

Perspectives

Estimation de variance.

Perspectives

Estimation de variance.

Cas de probabilités d'inclusion quelconques.

Perspectives

Estimation de variance.

Cas de probabilités d'inclusion quelconques.

Cas d'un tirage systématique de Deville sur fichier randomisé?

Bibliographie

Deville, J.-C. (1998). *Une nouvelle (encore une!) méthode de tirage à probabilités inégales*, Rapport Technique, Insee.

Deville, J.-C., and Tillé, Y. (2005). *Variance approximation under balanced sampling*. Journal of Statistical Planning and Inference, 128, pages 569-591.

Hajek, J. (1964). *Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population*. Annals of Mathematical Statistics, 35, 1491-1523.

Tillé, Y. (2006). *Sampling Algorithms*. New-York, Wiley.