Généralisation de la méthode du partage des poids à une population échantillonnée continue
Application à l'Inventaire National Forestier

Guillaume Chauvet[1,2] Olivier Bouriaud[3] Philippe Brion[2]

[1]École Nationale de la Statistique et de l'Analyse de l'Information

[2]Irmar

[3]Laboratoire de l'Inventaire Forestier

26/02/2021

Contexte

Dans une enquête, on peut distinguer plusieurs types d'unités statistiques, et notamment les unités d'échantillonnage et les unités d'observation. Dans les cas les plus complexes, ces deux types d'unités sont distincts.

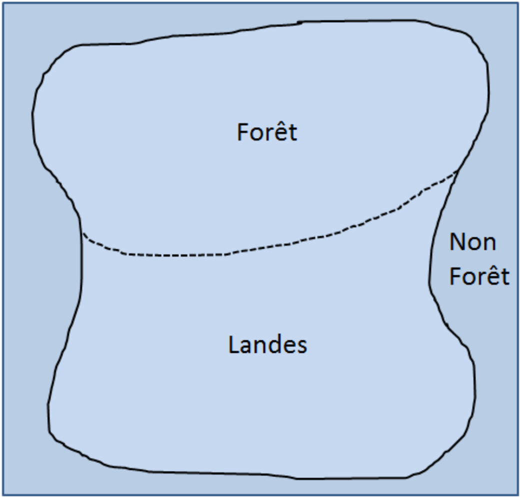
Pour résoudre ce problème, Deville et Lavallée (2006) ont proposé la méthode de partage des poids, basée sur un principe de dualité: une variable d'intérêt est ainsi exprimée sous la forme d'une variable synthétique dans la population échantillonnée.

Nous étendons la méthode de partage des poids au cas d'une population échantillonnée continue et d'une population observée discrète (cas continu-discret). Cela permet en particulier de l'appliquer au cas de l'Inventaire National Forestier français.

Principe simplifié de l'Inventaire National Forestier

Principe de l'Inventaire National Forestier

Territoire d'étude



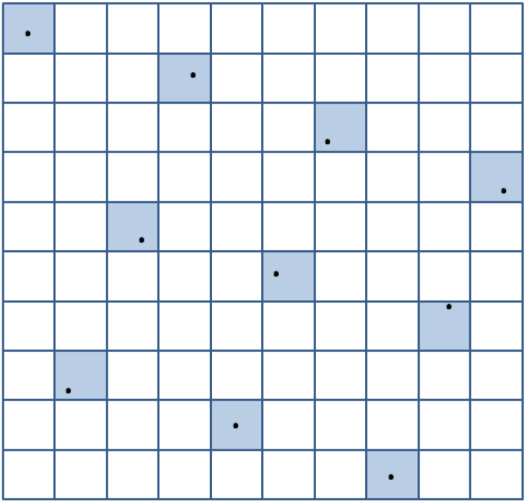
Principe de l'Inventaire National Forestier

Grille à maille carrée + tirage au 1/10

1									
			1						
						1			
									1
		1							
					1				
								1	
	1								
				1					
							1		

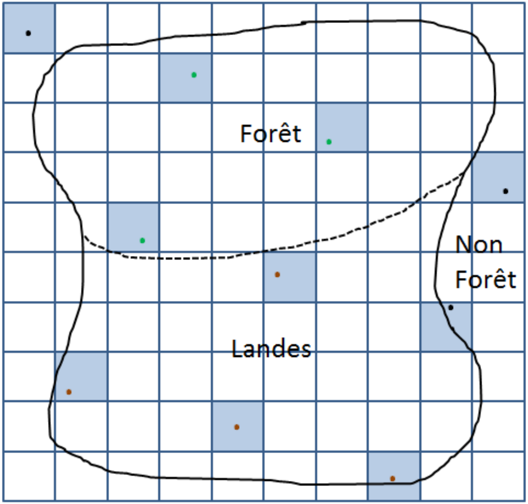
Principe de l'Inventaire National Forestier

Phase 1 : 1 point par maille



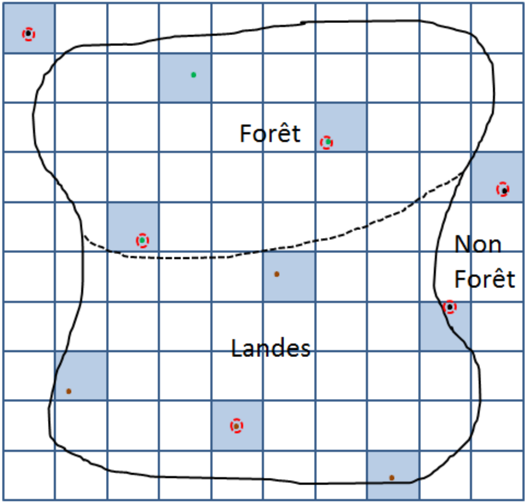
Principe de l'Inventaire National Forestier

Phase 1 : 1 point par maille



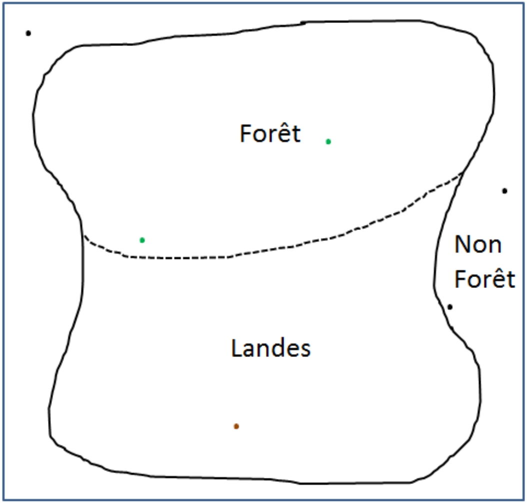
Principe de l'Inventaire National Forestier

Echantillon de phase 2



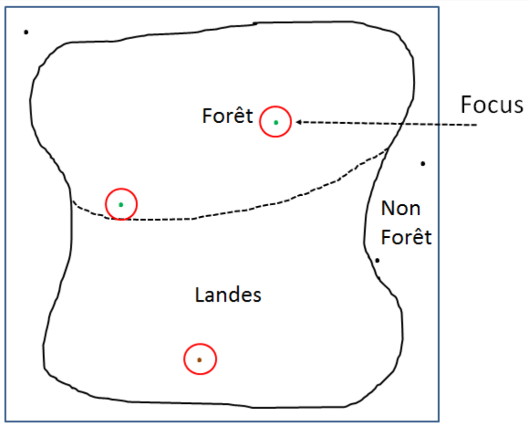
Principe de l'Inventaire National Forestier

Echantillon de phase 2



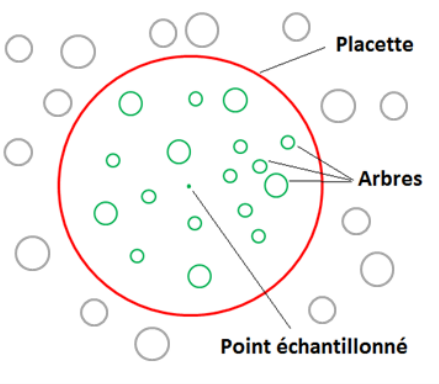
Principe de l'Inventaire National Forestier

Les placettes



Principe de l'Inventaire National Forestier

Une placette

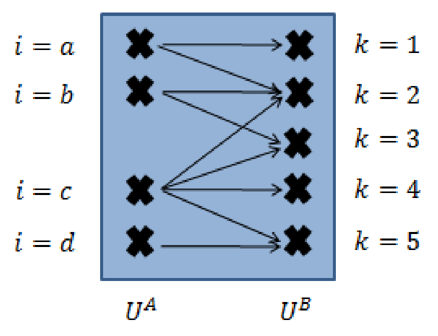


On trace une placette de rayon 25 m autour du point échantillonné. On enquête (taille, volume, essence, ...) les arbres situés à l'intérieur de la placette.

Comment passer d'un poids de tirage $d^A(x)$ pour le point x échantillonné à un poids de tirage pour chaque arbre enquêté?

Partage des poids
cas discret-discret

Partage des poids : cas discret-discret



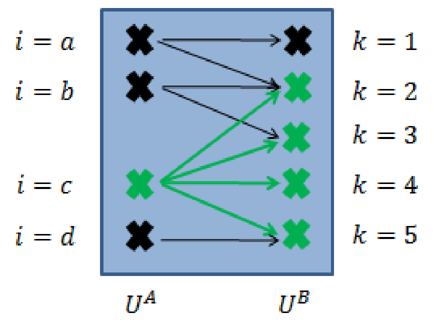
On s'intéresse à une pop. discrète U^B, avec une var. d'intérêt y_k^B dont on veut estimer le total

$$\tau_y^B = \sum_{k \in U^B} y_k^B.$$

Pas de base de sondage pour U^B, mais elle est liée à une pop. U^A avec une base de sondage. On note

$$L_{ik} = \begin{cases} 1 & \text{si } i \in U^A \text{ et } k \in U^B \\ & \text{sont liées,} \\ 0 & \text{sinon.} \end{cases}$$

Partage des poids : cas discret-discret



Pour $i \in U^A$, ses *descendants* sont les individus de U^B liés à i:

$$Des(i) \quad = \quad \{k \in U^B; \ L_{ik} = 1\}.$$

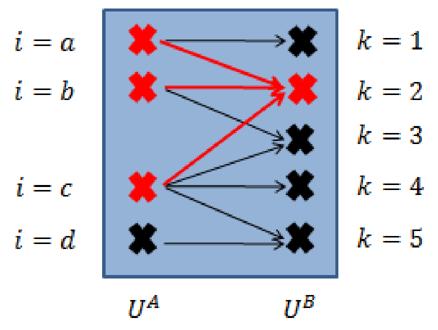
Pour un indiv. $k \in U^B$, ses *ancêtres* sont les indiv. de U^A liés à k:

$$Anc(k) \quad = \quad \{i \in U^A; \ L_{ik} = 1\}.$$

On suppose que tout $k \in U^B$ a au moins un ancêtre :

$$N_k^+ \quad = \quad \sum_{i \in U^A} L_{ik} > 0.$$

Partage des poids : cas discret-discret



Pour $i \in U^A$, ses *descendants* sont les individus de U^B liés à i:

$$Des(i) \quad = \quad \{k \in U^B;\ L_{ik} = 1\}.$$

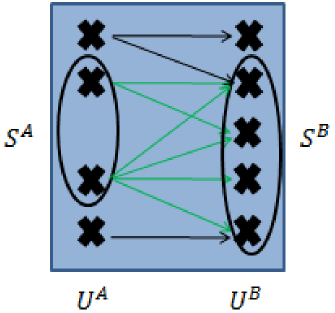
Pour un indiv. $k \in U^B$, ses *ancêtres* sont les indiv. de U^A liés à k:

$$Anc(k) \quad = \quad \{i \in U^A;\ L_{ik} = 1\}.$$

On suppose que tout $k \in U^B$ a au moins un ancêtre :

$$N_k^+ \quad = \quad \sum_{i \in U^A} L_{ik} > 0.$$

Partage des poids : cas discret-discret

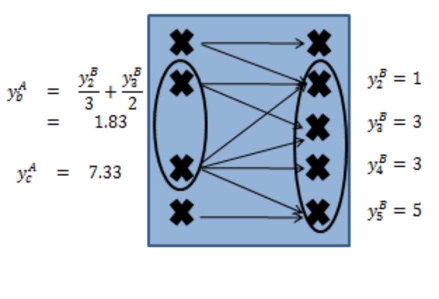


On tire un échantillon $S^A \subset U^A$.

On enquête dans U^B l'ensemble des descendants des indiv. de S^A:

$$S^B = \left\{ k \in U^B;\ L_{ik} = 1 \text{ pour au moins un } i \in S^A \right\}.$$

Partage des poids : cas discret-discret



Selon le principe de dualité, la var. y_k^B peut être distribuée sur les unités de U^A :

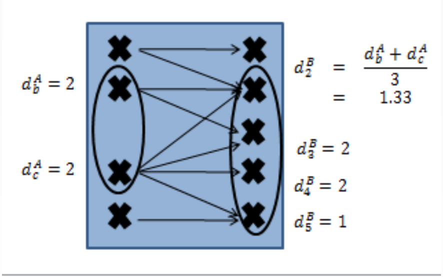
$$\tau_y^B = \sum_{k \in U^B} y_k^B = \sum_{i \in U^A} y_i^A$$

avec

$$y_i^A = \sum_{k \in U^B} \frac{L_{ik} y_k^B}{N_k^+}.$$

Chaque y_k^B est réparti à parts égales sur les ancêtres de k.

Partage des poids : cas discret-discret



Soit d_i^A le poids de sondage de $i \in S^A$. En termes d'estimation, le principe de dualité conduit à

$$\hat{\tau}_y^B = \sum_{i \in S^A} d_i^A y_i^A = \sum_{k \in S^B} d_k^B y_k^B$$

avec

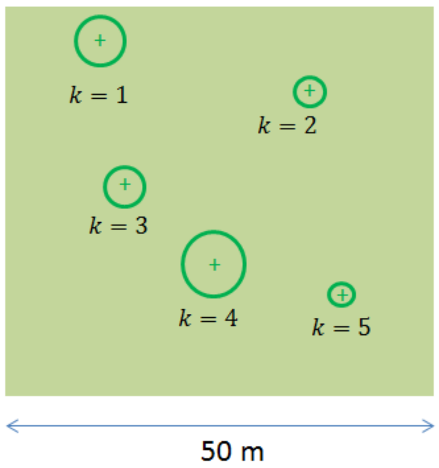
$$d_k^B = \frac{1}{N_k^+} \sum_{i \in S^A} L_{ik} d_i^A.$$

Chaque $k \in S^B$ reçoit la somme des poids de ses ancêtres de S^A, en divisant par le nb d'ancêtres N_k^+.

On doit avoir N_k^+ pr tt $k \in S^B$.

Partage des poids
cas continu-discret

Partage des poids : cas continu-discret



On s'intéresse à une pop. discrète U^B (arbres), avec une var. d'intérêt y_k^B dont on veut estimer le total

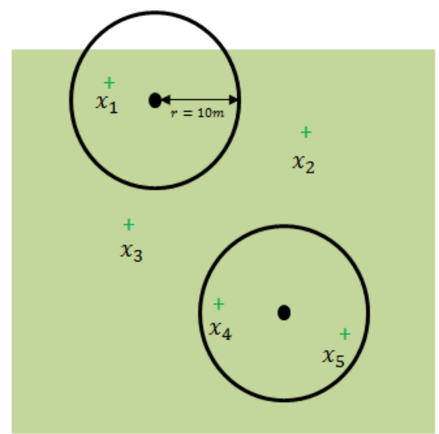
$$\tau_y^B = \sum_{k \in U^B} y_k^B.$$

Pas de base de sondage pour U^B, mais elle est liée à une pop. continue $\mathcal{U}^A$ (territoire). On note

$$L_k(x) = \begin{cases} 1 & \text{si } x \in \mathcal{U}^A \text{ et} \\ & k \in U^B \text{ sont liés,} \\ 0 & \text{sinon.} \end{cases}$$

$$= \begin{cases} 1 & \text{si } x_k \in C(x, r) \\ 0 & \text{sinon.} \end{cases}$$

Partage des poids : cas continu-discret



Pour $x \in \mathcal{U}^A$, ses *descendants* sont les indiv. de U^B liés à x:

$$Des(x) = \{k \in U^B;\ x_k \in C(x, r)\}.$$

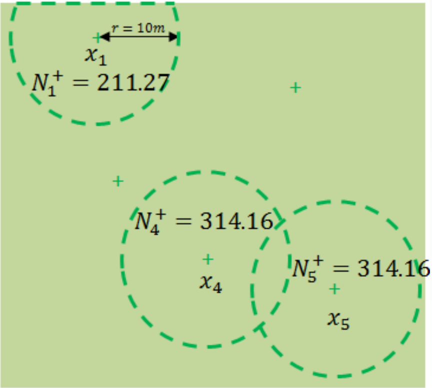
Pour un indiv. $k \in U^B$, ses *ancêtres* sont les points de $\mathcal{U}^A$ liés à k:

$$\begin{aligned} Anc(k) &= \{x \in \mathcal{U}^A;\ x \in C(x_k, r)\} \\ &= C(x_k, r) \cap \mathcal{U}^A. \end{aligned}$$

On suppose que tout $k \in U^B$, la surface de $Anc(k)$ est non nulle :

$$N_k^+ = \int_{x \in \mathcal{U}^A} L_k(x)dx > 0.$$

Partage des poids : cas continu-discret



Pour $x \in \mathcal{U}^A$, ses *descendants* sont les indiv. de U^B liés à x:

$$Des(x) \;=\; \{k \in U^B;\; x_k \in C(x,r)\}.$$

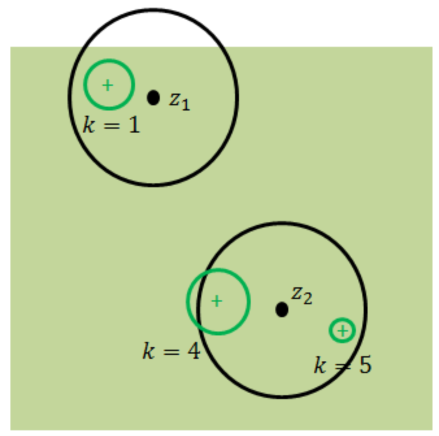
Pour un indiv. $k \in U^B$, ses *ancêtres* sont les points de $\mathcal{U}^A$ liés à k:

$$\begin{aligned} Anc(k) &= \{x \in \mathcal{U}^A;\; x \in C(x_k, r)\} \\ &= C(x_k, r) \cap \mathcal{U}^A. \end{aligned}$$

On suppose que tout $k \in U^B$, la surface de $Anc(k)$ est non nulle :

$$N_k^+ \;=\; \int_{x \in \mathcal{U}^A} L_k(x)dx > 0.$$

Partage des poids : cas continu-discret

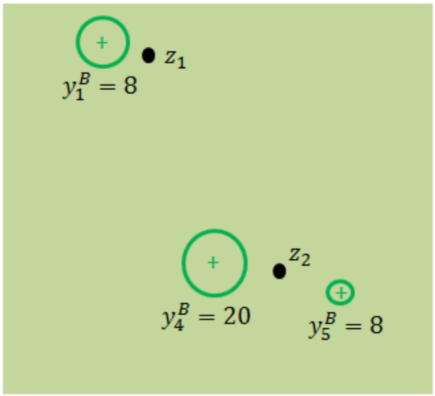


On tire un échantillon $\mathcal{S}^A = \{z_1, \ldots, z_n\}$ dans $\mathcal{U}^A$.

On enquête dans U^B l'ensemble des descendants des points de $\mathcal{S}^A$:

$$\begin{aligned} S^B &= \left\{k \in U^B;\ L_{ik} = 1 \text{ pour au moins un } x \in \mathcal{S}^A\right\}. \\ &= \bigcup_{z \in \mathcal{S}^A} Des(z). \end{aligned}$$

Partage des poids : cas continu-discret



Selon le principe de dualité, la var. y_k^B peut être distribuée sur les points de $\mathcal{U}^A$:

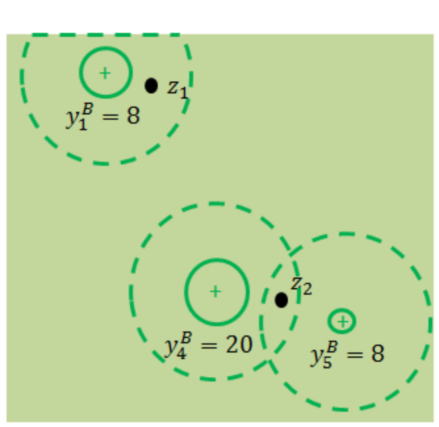
$$\tau_y^B = \sum_{k \in U^B} y_k^B = \int_{x \in \mathcal{U}^A} y^A(x) dx$$

avec

$$y^A(x) = \sum_{k \in U^B} \frac{L_k(x) y_k^B}{N_k^+}.$$

Chaque y_k^B est réparti à parts égales sur les ancêtres de k.

Partage des poids : cas continu-discret



Selon le principe de dualité, la var. y_k^B peut être distribuée sur les points de $\mathcal{U}^A$:

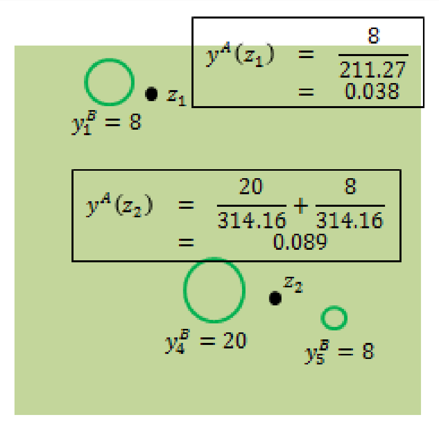
$$\tau_y^B = \sum_{k \in U^B} y_k^B = \int_{x \in \mathcal{U}^A} y^A(x)dx$$

avec

$$y^A(x) = \sum_{k \in U^B} \frac{L_k(x) y_k^B}{N_k^+}.$$

Chaque y_k^B est réparti à parts égales sur les ancêtres de k.

Partage des poids : cas continu-discret



Selon le principe de dualité, la var. y_k^B peut être distribuée sur les points de $\mathcal{U}^A$:

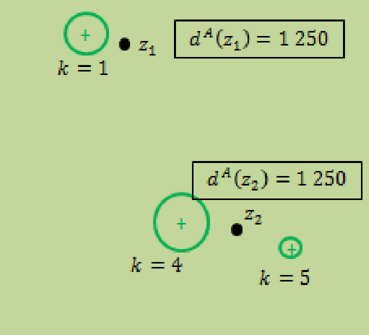
$$\tau_y^B = \sum_{k \in U^B} y_k^B = \int_{x \in \mathcal{U}^A} y^A(x)dx$$

avec

$$y^A(x) = \sum_{k \in U^B} \frac{L_k(x) y_k^B}{N_k^+}.$$

Chaque y_k^B est réparti à parts égales sur les ancêtres de k.

Partage des poids : cas continu-discret



Soit $d^A(x)$ le poids de sondage de $x \in \mathcal{S}^A$. En termes d'estimation, le principe de dualité conduit à

$$\hat{\tau}_y^B = \sum_{x \in \mathcal{S}^A} d^A(x) y^A(x) = \sum_{k \in S^B} d_k^B y_k^B$$

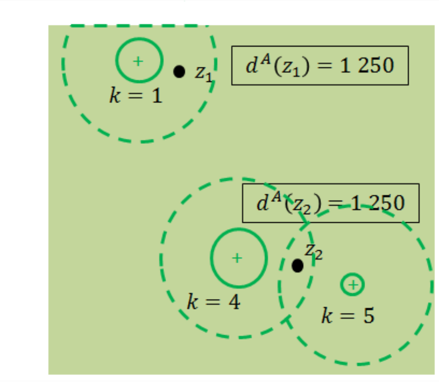
avec

$$d_k^B = \frac{1}{N_k^+} \sum_{x \in S^A} L_k(x) d^A(x).$$

Chaque $k \in S^B$ reçoit la somme des poids de ses ancêtres de $\mathcal{S}^A$, en divisant par le nb d'ancêtres N_k^+.

On doit avoir N_k^+ pr tt $k \in S^B$.

Partage des poids : cas continu-discret



Soit $d^A(x)$ le poids de sondage de $x \in \mathcal{S}^A$. En termes d'estimation, le principe de dualité conduit à

$$\hat{\tau}_y^B = \sum_{x \in \mathcal{S}^A} d^A(x) y^A(x) = \sum_{k \in S^B} d_k^B y_k^B$$

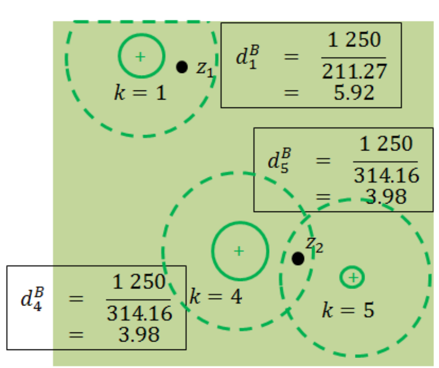
avec

$$d_k^B = \frac{1}{N_k^+} \sum_{x \in S^A} L_k(x) d^A(x).$$

Chaque $k \in S^B$ reçoit la somme des poids de ses ancêtres de $\mathcal{S}^A$, en divisant par le nb d'ancêtres N_k^+.

On doit avoir N_k^+ pr tt $k \in S^B$.

Partage des poids : cas continu-discret



Soit $d^A(x)$ le poids de sondage de $x \in \mathcal{S}^A$. En termes d'estimation, le principe de dualité conduit à

$$\hat{\tau}_y^B = \sum_{x \in \mathcal{S}^A} d^A(x) y^A(x) = \sum_{k \in S^B} d_k^B y_k^B$$

avec

$$d_k^B = \frac{1}{N_k^+} \sum_{x \in S^A} L_k(x) d^A(x).$$

Chaque $k \in S^B$ reçoit la somme des poids de ses ancêtres de $\mathcal{S}^A$, en divisant par le nb d'ancêtres N_k^+.

On doit avoir N_k^+ pr tt $k \in S^B$.

Conclusion et travaux en cours

La généralisation du partage des poids offre un cadre théorique pour l'étude d'une population sans base de sondage, mais située dans un territoire délimité. C'est une méthode particulièrement intéressante pour le sondage des ressources naturelles.

Travaux en cours :

- Formalisation dans un contexte général (population continue quelconque).
- Développement d'estimateurs de totaux + estimateurs de variance pour l'inventaire national forestier, en tenant compte du redressement des estimateurs (post-stratification en seconde phase).
- Pas de difficulté particulière apparente pour le cas continu-continu : application?

Merci pour votre attention
Des questions?