Plans à probabilités inégales

Exercice 1: Une population est composée de 6 ménages de tailles respectives 2 ; 4 ; 3 ; 9 ; 1 et 2 (la taille x_k d'un ménage k est le nombre de personnes physiques qu'il comprend). On tire 3 ménages sans remise, avec une probabilité proportionnelle à leur taille.

1. On a pour tout k : $\pi_k = n\frac{x_k}{t_x}$ avec $n = 3$ et $t_x = 2+4+3+9+1+2 = 21$ donc

$$\pi_k = \frac{x_k}{7}, \quad k \in U$$

pour $k = 4$ on voit bien que $\pi_4 = \frac{9}{7} > 1$ alors on assigne la valeur 1 à π_4 ($\pi_4 = 1$)

et

$$\pi_k = (n-1)\frac{x_k}{t_x - 9} \quad \text{pour } k = 1;\ 2;\ 3;\ 5;\ 6$$

donc on recalcule les π_k pour les autres unités, selon :

$$\pi_k = 2\,\frac{x_k}{12} = \frac{x_k}{6}$$

Finalement, les probabilités d'inclusion sont présentées dans le tableau suivant

k	1	2	3	4	5	6
π_k	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{2}$	1	$\frac{1}{6}$	$\frac{1}{3}$

On peut vérifier que $\sum_k \pi_k = 3$.

2. On tire un réel aléatoire u entre 0 et 1, et on s'intéresse aux probabilités cumulées présentées dans le tableaux suivant

k	1	2	3	4	5	6
$c_k = \sum_{j=1}^{k} \pi_j$	$\frac{1}{3}$	1	$\frac{3}{2}$	$\frac{5}{2}$	$\frac{16}{6}$	$\frac{18}{6}$

Pour $u = 0,4$ on a $c_1 = \frac{1}{3} \leq 0,4 < 1 = c_2$ donc on sélectionne le 2.

Pour $u + 1 = 1,4$ on a $c_2 = 1 \leq 1,4 < \frac{3}{2} = c_3$ donc on sélectionne le 3.

On s'arrête puisqu'on doit choisir 3 éléments et puisque le 4 est choisi d'office l'échantillon comporte le 2, 3 et 4.

3. On a :

$$\begin{aligned}
\overline{x} &= \frac{1}{6}\sum_{k\in s}\frac{x_k}{\pi_k} \\
&= \frac{1}{6}\left(\frac{x_2}{\pi_2}+\frac{x_3}{\pi_3}+\frac{x_4}{\pi_4}\right) \\
&= \frac{1}{6}\left(\frac{4}{\frac{2}{3}}+\frac{3}{\frac{1}{2}}+9\right) \\
&= \frac{1}{6}(6+6+9)=\frac{21}{6}=3,5
\end{aligned}$$

Comme $\pi_k = \frac{x_k}{6}$ $\forall\, k$ alors les π_k et les x_k sont parfaitement proportionnels, par construction on a une variance nulle, donc un estimateur est parfait pour la moyenne.

Exercice 2: Puisqu'il y 'a une relation à peu près proportionnelle entre le chiffre d'affaire et l'effectif salarié, le tirage à probabilité inégales proportionnellement au chiffres d'affaires est justifié.

On a pour tout k : $\pi_k = n\frac{x_k}{t_x}$ avec $n=3$ et $t_x = 60$

Puisque $\pi_1 = 3\times\frac{40}{60} = 2 > 1$ alors l'unité 1 est sélectionnée d'office et éliminée de la poplation, ce qui entraine

$$t_x = \sum_{k\in U\setminus\{1\}} x_k = 20$$

et

$$\pi_2 = \frac{(n-1)x_2}{t_x} = 2\times\frac{10}{20} = 1$$

l'unité 2 est aussi sélectionnée d'office et éliminée de la population. Il reste à sélectionner une unité parmi les identifiants 3, 4, 5 et 6.

Comme

$$t_x = \sum_{k\in U\setminus\{1,2\}} x_k = 10$$

$$\pi_3 = \frac{8}{10} = 0,8 \quad \pi_4 = 0,1 \quad \pi_5 = 0,05 \quad \text{et } \pi_6 = 0,05$$

Les probabilités d'inclusion cumulées valent (en notant $c_k = \sum_{j=1}^{k} \pi_j$)

$$c_3 = 0,8 \quad c_4 = 0,9 \quad c_5 = 0,95 \quad \text{et } c_6 = 1$$

Comme $c_3 = 0,8 \leq 0,83 < c_4 = 0,9 \Rightarrow$ l'unité 4 est sélectionnée. L'échantillon finalement sélectionné est $s = \{1, 2, 4\}$

Si on modifie l'ordre du fichier les 2 plus grosses unités ($x = 40$) et ($x = 10$) sont toujours retenues d'office, quel que soit l'ordre initial. Avec le nombre tiré au hasard entre 0 et 1, tout dépend de la position de l'unité ($x = 8$) lorsqu'on considère les 4 unités restantes ($x = 0,5;\ 0,5;\ 1;\ 8$). Si l'unité ($x = 8$) est dans la position 2, 3 ou 4 alors elle est toujours retenue (facile à vérifier). Si elle est en position 1, alors on peut retenir n'importe quelle valeur des 3 restantes.

L'ordre de fichier influe donc sur l'échantillon sélectionné.

<u>Exercice 3:</u> Soit une population de 5 unités. On veut sélectionner par un tirage systématique à probabilités inégales un échantillon de deux unités avec des probabilités d'inclusion proportionnelles aux valeurs x_i suivantes 1 ; 1 ; 6 ; 6 ; 6.

1. Avec $n = 2$, on a

$$t_x = \sum_{k=1}^{5} x_k = 20, \quad \pi_k = n\frac{x_k}{t_x} \quad k = 1; \ldots ; 5$$

2. Le nombre de permutation est $\mathcal{C}_5^2 = \frac{5!}{2!3!} = \frac{5\times4}{2} = 10$

1	1	1	6	6	6
2	1	6	1	6	6
3	1	6	6	1	6
4	1	6	6	6	1
5	6	1	1	6	6
6	6	1	6	1	6
7	6	1	6	6	1
8	6	6	1	1	6
9	6	6	1	6	1
10	6	6	6	1	1

On peut remarquer que les permutations se décomposent en deux catégories

— les permutations où les 2 petites valeurs sont consécutives.

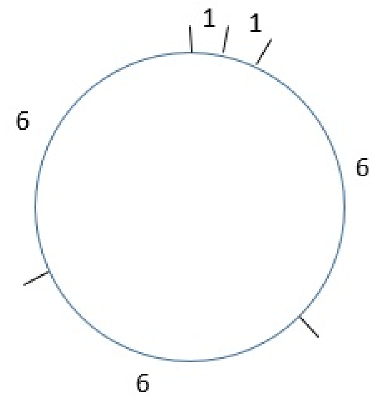

Figure 1 – Cas 1 : les deux petites valeurs sont adjacentes

— les permutations où les 2 petites valeurs sont séparées par une grande valeur.

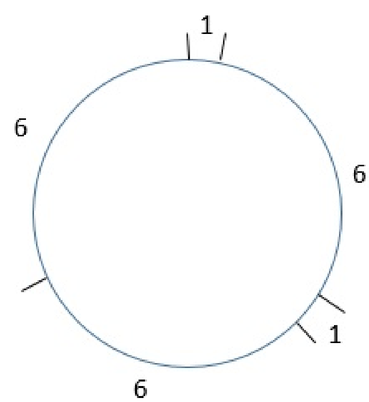

Figure 2 – Cas 2 : les deux petites valeurs ne sont pas adjacentes

Donc pour examiner toutes les permuations, il suffit de se borner à examiner les 2 cas suivant

1	1	6	6	6
1	6	1	6	6

Les probabilités d'inclusion à l'ordre un valent respectivement pour les deux ordres

0,1	0,1	0,6	0,6	0,6
0,1	0,6	0,1	0,6	0,6

Dans les deux cas, il est impossible de choisir conjointement les 2 petites valeurs parce que le pas de tirage est 1, donc ceci est valide pour toutes les permutations possibles et donc leur probabilité d'inclusion d'ordre 2 est nulle. La conséquence principale est qu'il n'y a pas d'estimateur sans biais de la variance.

Exercice 4: Soit une population U composée de 6 unités. On connaît les valeurs prises par un caractère auxiliaire x sur toutes les unités de la population :

$$x_1 = 200,\ x_2 = 80,\ x_3 = 50,\ x_4 = 50,\ x_5 = 10,\ x_6 = 10.$$

1. On a pour tout k : $\pi_k = n\frac{x_k}{t_x}$ avec $n = 4$ et $t_x = 400$

   Puisque $\pi_1 = 4 \times \frac{200}{400} = 2 > 1$ alors l'unité 1 est sélectionnée d'office et éliminée de la poplation, ce qui entraine

   $$t_x = \sum_{k \in U \setminus \{1\}} x_k = 200$$

   et

   $$\pi_2 = (n-1) \times \frac{x_2}{t_x} = 3 \times \frac{80}{200} = 1,2 > 1$$

   l'unité 2 est aussi sélectionnée d'office et éliminée de la population.

   Il reste de sélectionner 2 unités parmi 4 :

   on a

   $$t_x = \sum_{k \in U \setminus \{1,\, 2\}} x_k = 120$$

   ce qui entraine

   $$\pi_3 = \pi_4 = 2 \times \frac{50}{120} = \frac{5}{6} \quad \text{et} \quad \pi_5 = \pi_6 = 2 \times \frac{10}{120} = \frac{1}{6}$$

Les probabilités d'inclusion cumulées valent (en notant $c_k = \sum_{j=3}^{k} \pi_j$)

$$c_3 = \frac{5}{6} \quad c_4 = \frac{10}{6} \quad c_5 = \frac{11}{6} \quad \text{et } c_6 = 2$$

d'où on sélectionne l'échantillon $\{1;\ 2;\ 3;\ 4\}$

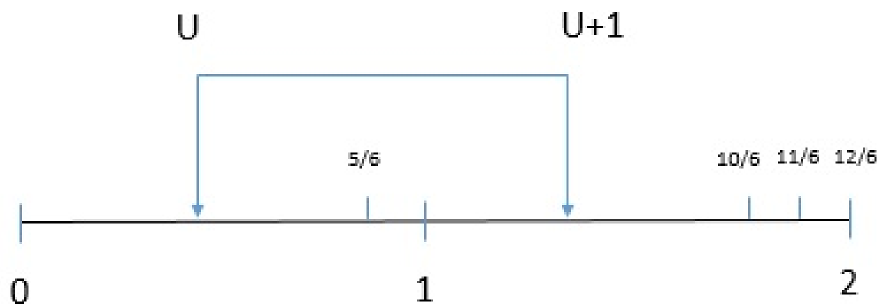

FIGURE 3 – Tirage systématique de 2 unités

2. Remarquons que pour tous k et l, si $\pi_k = 1$ alors $\pi_{kl} = \pi_l$.

   Soit u la valeur donnée par $\mathcal{U}(0;1)$

   — si $0 \leq u \leq \frac{4}{6}$ alors on tombe dans les intervalles n° 1 et 2, les deux unités sélectionnées sont $\{3;\ 4\}$

   — si $\frac{4}{6} < u \leq \frac{5}{6}$ alors on tombe dans les intervalles n° 1 et 3, les deux unités sélectionnées sont $\{3;\ 5\}$

   — si $\frac{5}{6} < u \leq 1$ alors on tombe dans les intervalles n° 2 et 4, les deux unités sélectionnées sont $\{4;\ 6\}$

   donc

$$\pi_{34} = \frac{4}{6} \quad \text{et} \quad \pi_{35} = \pi_{46} = \frac{1}{6}$$

Les autres combinaisons donnent bien $\pi_{kl} = 0$. En fin la matrice d'inclusion d'ordre deux est donnée par.

$$\begin{pmatrix} - & 1 & \frac{5}{6} & \frac{5}{6} & \frac{1}{6} & \frac{1}{6} \\ 1 & - & \frac{5}{6} & \frac{5}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{5}{6} & \frac{5}{6} & - & \frac{4}{6} & \frac{1}{6} & 0 \\ \frac{5}{6} & \frac{5}{6} & \frac{4}{6} & - & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & - & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & 0 & - \end{pmatrix}$$

3. d'après ce qui précède il y'a que 3 échantillons possible de taille 4 à savoir $\{1;\ 2;\ 3;\ 4\}$, $\{1;\ 2;\ 3;\ 5\}$ et $\{1;\ 2;\ 4;\ 6\}$. Le vrai total $t_y = 200$

   On rappelle que

$$\widehat{t}_{y,\pi} = \sum_{k \in S} \frac{y_k}{\pi_k}.$$

$$\widehat{\mathbf{Var}}\left(\widehat{t}_{y,\pi}\right) = -\frac{1}{2} \sum_{k \in s} \sum_{\substack{l \in s \\ l \neq k}} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l}\right)^2 \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}}$$

   On sait bien que $\widehat{t}_{y,\pi}$ est sans biais de t_y.

s	$\widehat{t}_{y,\pi}$	$\widehat{\mathbf{Var}}\left(\widehat{t}_{y,\pi}\right)$	$p(s)$
$\{1;\ 2;\ 3;\ 4\}$	196	0,75	$\frac{4}{6}$
$\{1;\ 2;\ 3;\ 5\}$	226	-48	$\frac{1}{6}$
$\{1;\ 2;\ 4;\ 5\}$	190	0	$\frac{1}{6}$

   conformément à la théorie $\mathbb{E}[\widehat{t}_{y,\pi}] = \sum p(s)\widehat{t}_{y,\pi}(s) = 200$ et $\mathbb{E}[\widehat{\mathbf{Var}}[\widehat{t}_{y,\pi}]] < 0$ ce qui implique $\mathbb{E}[\widehat{\mathbf{Var}}[\widehat{t}_{y,\pi}]] \neq \mathbf{Var}[t_{y,\pi}] \geq 0$ car $\pi_{kl} = 0$.

M. BADAOUI