Exercices de démonstrations et corrections

June 2021

1 Montrez que les probabilités d'inclusion du premier et du second ordre peuvent s'écrire : $\pi_k = 1 - (1-p_k)^m$ et $\pi_{kl} = 1 - (1-p_k)^m - (1-p_l)^m + (1-p_k-p_l)^m$

Solution: **Montrons que** $\pi_k = 1 - (1-p_k)^m \;\; k \in U$

Par definition, $\pi_k = \mathbb{P}(k \in s)$
Et P(k $\notin s$) = Proba que k ne soit tiré à aucun des tirages
P(k $\notin s) = (1-p_k)\ldots(1-p_k)$
P(k $\notin s) = (1-p_k)^m$
Ainsi $\pi_k = \mathbb{P}(k \in s) = (1-1-p_k)^m$

Montrons que $\pi_{kl} = 1 - (1-p_k)^m - (1-p_l)^m + (1-p_k-p_l)^m \,,\; k,l \in U, k \neq l$

Par definition, $\pi_{kl} = \mathbb{P}(k \in \text{s et l} \in s)$
$\pi_{kl} = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(A \bigcup B)$
$\pi_{kl} = (1-(1-p_k)^m) + (1-(1-pl)^m) - \mathbb{P}(A \bigcup B)$
$\pi_{kl} = \mathbb{P}(B) + \mathbb{P}(A) - (1 - \mathbb{P}(\bar{A} \bigcap \bar{B}))$
$\pi_{kl} = \mathbb{P}(B) + \mathbb{P}(A) - (1-(1-p_k-p_l)^m)$
$\pi_{kl} = (1-(1-p_k)^m) + (1-(1-pl)^m) - (1-(1-p_k-p_l)^m)$
Ainsi, $\pi_{kl} = 1-(1-p_k)^m - (1-p_l)^m + (1-p_k-p_l)^m)$

2 Montrez que l'estimateur $\hat{V}(\hat{t}_{y\mathrm{HT}})$ defini ci-dessous est un estimateur sans biais de la variance de l'estimateur HT d'un total

$$\hat{V}(\hat{t}_{y\mathrm{HT}}) = \sum_{k\in s}\sum_{l\in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l}.$$

Rappellons que :

$$V(\hat{t}_{y\mathrm{HT}}) \;\; = \;\; \sum_{k\in U}\sum_{l\in U} \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} \tag{1}$$

Solution: **Montrons que** $\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HT}})) = \mathbb{V}(t_{y\mathrm{HT}})$

Reécrivons $\hat{V}(\hat{t}_{y\mathrm{HT}})$=$\sum_{k\in U}\sum_{l\in U} \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} I_k I_l$

$$
\begin{aligned}
\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HT}})) &= \sum_{k\in U}\sum_{l\in U}\frac{\Delta_{kl}}{\pi_{kl}}\frac{y_k}{\pi_k}\frac{y_l}{\pi_l}\mathbb{E}(I_k I_l) \\
&= \mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HT}})) = \sum_{k\in U}\sum_{l\in U}\frac{\Delta_{kl}}{\pi_{kl}}\frac{y_k}{\pi_k}\frac{y_l}{\pi_l}\mathbb{E}(\nVdash k, ls) \\
&= \mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HT}})) = \sum_{k\in U}\sum_{l\in U}\frac{\Delta_{kl}}{\pi_{kl}}\frac{y_k}{\pi_k}\frac{y_l}{\pi_l}\pi_k l \\
&= \mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HT}})) = \sum_{k\in U}\sum_{l\in U}\Delta_{kl}\frac{y_k}{\pi_k}\frac{y_l}{\pi_l} = V(\hat{t}_{y\mathrm{HT}}
\end{aligned}
$$

3 Calculez la variance de l'estimateur de Hansen-Hurwitz (HH)

L'estimateur HH peut s'écrire

$$
\hat{t}_{y\mathrm{HH}} = \sum_{k\in U}\frac{y_k}{m\,p_k}R_k,
$$

où R_k et la variable aléatoire égale au nombre de fois l'unité k a été sélectionnée. Le vecteur $(R_1,\ldots,R_N)'$ suit une loi multinomiale de paramètres m et $(p_1,\ldots,p_N)'$.

La covariance entre R_k et R_l est égale à $(R_k,R_l) = -mp_kp_l$ pour $k\neq l$. La variance de l'estimateur HH est donc égale à

$$
\begin{aligned}
V(\hat{t}_{y\mathrm{HH}}) &= \sum_{k\in U}\left(\frac{y_k}{m\,p_k}\right)^2 V(R_k) + \sum_{k\in U}\sum_{l\neq k\in U}\frac{y_k}{m\,p_k}\frac{y_l}{m\,p_l}(R_k,R_l) \\
&= \sum_{k\in U}\left(\frac{y_k}{m\,p_k}\right)^2 mp_k(1-p_k) - \sum_{k\in U}\sum_{l\neq k\in U}\frac{y_k}{m\,p_k}\frac{y_l}{m\,p_l}mp_kp_l \\
&= \sum_{k\in U}\left(\frac{y_k}{m\,p_k}\right)^2 mp_k - \sum_{k\in U}\left(\frac{y_k}{m\,p_k}\right)^2 mp_k^2 - \sum_{k\in U}\sum_{l\neq k\in U}\frac{1}{m}y_ky_l \\
&= \frac{1}{m}\sum_{k\in U}\left(\frac{y_k}{p_k}\right)^2 p_k - \frac{1}{m}(\sum_{k\in U}y_k - \sum_{k\in U}\sum_{l\neq k\in U}y_ky_l) \\
&= \frac{1}{m}(\sum_{k\in U}\left(\frac{y_k}{p_k}\right)^2 p_k - t_{yU}^2)
\end{aligned}
$$

4 Montrez que l'estimateur $\hat{V}(\hat{t}_{y\mathrm{HH}})$ defini ci-dessous est un estimateur sans biais de la variance de l'estimateur HH d'un total

$\hat{V}(\hat{t}_{y\mathrm{HH}}) = \frac{1}{m(m-1)}\left(\sum_{k\in U}\left(\frac{y_k}{p_k}\right)^2 R_k - m\hat{t}_{y\mathrm{HH}}^2\right)$

Solution: Montrons que $\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HH}})) = \mathbb{V}(t_{y\mathrm{HH}})$

$\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HH}}))$=$\mathbb{E}[\frac{1}{m(m-1)}\left(\sum_{k\in U}\left(\frac{y_k}{p_k}\right)^2 R_k - m\hat{t}_{y\mathrm{HH}}^2\right)]$

$\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HH}})) = \frac{1}{m(m-1)}\left(\sum_{k\in U}\left(\frac{y_k}{p_k}\right)^2\mathbb{E}(R_k) - m(\mathbb{E}(\hat{t}_{y\mathrm{HH}})^2 + \mathbb{V}(\hat{t}_{y\mathrm{HH}})\right)$
$\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HH}})) = \frac{1}{m(m-1)}\left(\sum_{k\in U}\left(\frac{y_k}{p_k}\right)^2 mp_k - mt_y^2 - \mathbb{V}(t_{y\mathrm{HH}})\right)$
$\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HH}})) = \frac{1}{m(m-1)}\left(m\mathbb{V}(t_{y\mathrm{HH}}) - \mathbb{V}(t_{y\mathrm{HH}})\right)$

Ainsi nous avons $\mathbb{E}(\hat{V}(\hat{t}_{y\mathrm{HH}})) = \frac{1}{m(m-1)}\left(m-1\right)\mathbb{V}(t_{y\mathrm{HH}}) = \mathbb{V}(t_{y\mathrm{HH}})$

5 Dans le cadre d'un plan SY on utilise $\hat{V}(\hat{t}_{y\mathrm{HT}}) = N^2\left(1-\frac{n}{N}\right)\frac{S_y^2}{n}$

Rappellons que dans le cadre d'un plan simple (SI) on a ,

$$V(\hat{t}_{y\mathrm{HT}}) = \sum_{k\in U}\sum_{l\in U}\Delta_{kl}\frac{y_k}{\pi_k}\frac{y_l}{\pi_l}.$$

Aussi dans un plan SI de taille n avec une population de taille N, les probalités de premier ordre d'inclusion sont $\pi_k = n/N$ et celles de second ordre sont $k \neq l$,

$$\pi_{kl} = \frac{n(n-1)}{N(N-1)}$$

on obtient de meme les covariances $\Delta_{kl} = \pi_{kl} - \pi_k\pi_l$ pour $k \neq l$:

$$\Delta_{kl} = -\frac{n\,(1-n/N)}{N(N-1)}.$$

Le plan est de taille fixe, donc nous pouvons utiliser la formuler de variance suivante :

$$V(\hat{t}_{y\mathrm{HT}}) = -\frac{1}{2}\sum_{k\in U}\sum_{l\in U}\Delta_{kl}\left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l}\right)^2. \tag{2}$$

De facon eq́uivalente,

$$V(\hat{t}_{y\mathrm{HT}}) = -\frac{1}{2}\sum_{k\in U}\sum_{l\in U, l\neq k}\Delta_{kl}\left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l}\right)^2. \tag{3}$$

On remple les expressiosn de π_k et Δ_{kl} pour le plan SI dans l' equation (3) and obtain:

$$\begin{aligned} V(\hat{t}_{y\mathrm{HT}}) &= \frac{1}{2}\frac{N^2}{n^2}\frac{n\,(1-n/N)}{N(N-1)}\sum_{k\in U}\sum_{l\in U, l\neq k}(y_k - y_l)^2 \\ &= \frac{N^2}{n}\left(1-\frac{n}{N}\right)\frac{1}{2N(N-1)}\sum_{k\in U}\sum_{l\in U, l\neq k}(y_k - y_l)^2. \end{aligned} \tag{4}$$

On a :

$$\begin{aligned} \frac{1}{2N(N-1)}\sum_{k\in U}\sum_{l\in U, l\neq k}(y_k - y_l)^2 &= \frac{1}{2N(N-1)}\sum_{k\in U}\sum_{l\in U}(y_k - y_l)^2 \\ &= \frac{1}{2N(N-1)}\sum_{k\in U}\sum_{l\in U}(y_k - \overline{y_U} + \overline{y_U} - y_l)^2 \\ &= \frac{2}{2N(N-1)}\sum_{k\in U}N(y_k - \overline{y_U})^2 + \frac{2}{2N(N-1)}\sum_{k\in U}(y_k - \overline{y_U})\sum_{l\in U}(\overline{y_U} - y_l) \\ &= \frac{1}{N-1}\sum_{k\in U}(y_k - \overline{y_U})^2 \\ &= S_y^2 \end{aligned}$$

Puisque $\sum_{k\in U}(y_k - \overline{y_U}) = \sum_{k\in U} y_k - N\overline{y_U} = \sum_{k\in U} y_k - \sum_{k\in U} y_k = 0$.

On obtient bien l'equation (4)

$$\hat{V}(\hat{t}_{y\text{HT}}) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}$$

6 Montrez que l'estimateur $\hat{V}(\hat{t}_{y\text{HT}})_{Poisson} = \sum_{l\in s} \frac{1-\pi_k}{\pi_k^2} y_k^2$ defini ci-dessous est un estimateur sans biais de la variance de l'estimateur HT d'un total dans un plan de poisson

Rappellons que :

$$V(\hat{t}_{y\text{HT}})_{Poisson} = \sum_{k\in s} \frac{1-\pi_k}{\pi_k} y_k^2$$

Solution : Montrons que $\mathbb{E}(\hat{V}(\hat{t}_{y\text{HT}}))_{Poisson} = \mathbb{V}(\hat{t}_{y\text{HT}})_{Poisson}$

Reécrivons $\hat{V}(\hat{t}_{y\text{HT}})_{Poisson} = \sum_{k\in U} \frac{1-\pi_k}{\pi_k^2} y_k^2 I_k$

$$\begin{aligned}
\mathbb{E}(\hat{V}(\hat{t}_{y\text{HT}}))_{Poisson} &= \sum_{k\in U} \frac{1-\pi_k}{\pi_k^2} y_k^2 \mathbb{E}(I_k) \\
&= \sum_{k\in U} \frac{1-\pi_k}{\pi_k^2} y_k^2 \pi_k \\
\mathbb{E}(\hat{V}(\hat{t}_{y\text{HT}}))_{Poisson} &= \sum_{k\in U} \frac{1-\pi_k}{\pi_k} y_k^2
\end{aligned}$$

7 Montrez que la variance de l'estimateur de Horvitz-Thompson dans un plan de poisson s'écrit : $V(\hat{t}_{y\text{HT}})_{poisson} = \sum_{k\in U} \left(\frac{1-\pi_k}{\pi_k}\right) y_k^2$.

Rappellons que :

$$\hat{t}_{y\text{HT}} = \sum_{k\in s} \frac{y_k}{\pi_k} \tag{5}$$

On peut écrire la variance de l'estimateur de Horvitz-Thompson sous la forme :

$V(\hat{t}_{y\text{HT}}) = \sum_{k\in U} \left(\frac{y_k}{\pi_k}\right)^2 V(I_k) + \sum_{k\in U} \sum_{l\in U} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} Cov(I_k, I_l)$

Dans le plan de poisson, les unités de la population sont sélectionnées dans l'échantillon de manière indépendante et donc $\pi_{kl} = \pi_k \pi_l, \forall k, l \in U, k \neq l$. Donc $\Delta_{kl} = \pi_k l - \pi_k \pi_l = 0 = Cov(I_k, I_l)$

On obtient donc :

$$\begin{aligned}
V(\hat{t}_{y\text{HT}})_{poisson} &= \sum_{k\in U} \left(\frac{y_k}{\pi_k}\right)^2 V(I_k) \\
&= \sum_{k\in U} \left(\frac{y_k}{\pi_k}\right)^2 \pi_k(1-\pi_k) \\
V(\hat{t}_{y\text{HT}})_{poisson} &= \sum_{k\in U} \left(\frac{1-\pi_k}{\pi_k}\right) y_k^2
\end{aligned}$$