Journal de la société statistique de Paris

Fernand Chartier

Sur l'influence d'une « stratification aléatoire » lors de l'estimation du total d'une population

Journal de la société statistique de Paris, tome 97 (1956), p. 121-130
<http://www.numdam.org/item?id=JSFS_1956__97__121_0>

VI

VARIÉTÉS

Sur l'influence d'une " stratification aléatoire " lors de l'estimation du total d'une population

But de la note

Il s'agit de comparer la variance des estimateurs x' et x'' du total X de la variable additive (X) dans une population donnée, le mode d'obtention de ces estimateurs étant :

pour x' : tirage d'un échantillon d'effectif m parmi les M unités de la population; d'où l'estimateur x' de X (Cf. formule [8]);

pour x'' : « stratification aléatoire » de la population initiale, c'est-à-dire partage en deux strates d'effectifs M_1 et M_2 connus ($M_1 + M_2 = M$) mais de compositions aléatoires (par exemple : tirage sans remise d'un échantillon d'effectif M_1 donné parmi les M unités initiales pour obtenir la 1re strate, les $M_2 = M - M_1$ unités restantes constituant la 2e strate), puis, dans chaque strate, tirage d'un échantillon d'effectif m_1 et m_2 respectivement, permettant d'estimer X_1, total aléatoire de (X) dans la 1re strate, par x'_1 et X_2, total aléatoire de (X) dans la 2e strate, par x'_2, d'où l'estimateur $x'' = x'_1 + x'_2$ de $X = X_1 + X_2$ (Cf. formule [8]).

On supposera évidemment $m_1 + m_2 = m$, soit le même nombre d'unités-échantillon dans les deux procédés. Un cas particulier sera celui où :

$$\frac{m_1}{M_2} = \frac{m_2}{M_1} = \frac{m}{M}.$$

On distinguera deux cas :

Tirage des échantillons *avec* remise,

Tirage des échantillons *sans* remise.

Il s'agit, bien entendu, des échantillons d'effectif m, m_1 et m_2, la stratification aléatoire de la population ne pouvant correspondre qu'à un tirage sans remise de M_1 unités parmi les M, ou de M_2 sur les M. On suppose tous les tirages faits avec d'égales probabilités.

Avant d'étudier les variances de x' et x'', précisons les notations pour la population initiale et les deux strates et étudions les aléatoires que sont les caractéristiques de chaque strate.

* * *

La population initiale et les deux strates aléatoires

Les caractéristiques de la population initiale et des deux strates sont

(1)

		Population	1re strate	2e strate
Effectif.		M	M_1	M_2
N° d'identité des unités constitutives.		$a = 1, 2, \dots M$	$a_1 = 1, 2, \dots M_1$	$a_2 = 1, 2, \dots M_2$
Valeur de (X) sur l'unité.		n° a : X_a	n° a_1 : X_{a_1}	n° a_2 : X_{a_2}
Total de (X) . .		$X = \sum_{a}^{M} X_a$	$X_1 = \sum_{a_1}^{M_1} X_{a_1}$	$X_2 = \sum_{a_2}^{M_2} X_{a_2}$
Moyenne de (X).		$\overline{X} = X/M$	$\overline{X}_1 = X_1/M_1$	$\overline{X}_2 = X_2/M_2$
Dispersion de (X)		$S^2 = Q/(M-1)$	$S_1^2 = Q_1/(M_1-1)$	$S_2^2 = Q_2/(M_2-1)$
		$\sigma^2 = Q/M$	$\sigma_1^2 = Q_1/M_1$	$\sigma_2^2 = Q_2/M_2$
		$Q = \sum_{a}^{M} (X_a - \overline{X})^2$	$Q_1 = \sum_{a_1}^{M_1} (X_{a1} - \overline{X}_1)^2$	$Q_2 = \sum_{a_2}^{M_2} (X_{a2} - \overline{X}_2)^2$

Il est utile d'introduire les deux caractéristiques de dispersion S et σ, l'expression des variances étant plus simple en utilisant S pour un tirage sans remise et σ pour un tirage avec remise.

On a évidemment :

pour les quantités connues M_1 et M_2, la relation :

$$M_1 + M_2 = M,$$

pour les quantités aléatoires X_1 et X_2, la relation :

$$X_1 + X_2 = X,$$

puis les espérances mathématiques :

$$(2) \quad E\,X_1 = \frac{M_1}{M} X, \qquad E\,X_2 = \frac{M_2}{M} X, \qquad E\,(X_1 + X_2) = X,$$

et les variances :

$$V\,X_1 = M_1 \frac{M - M_1}{M} S^2, \qquad V\,X_2 = M_2 \frac{M - M_2}{M} S^2$$

soit :

$$\text{(3)} \qquad \mathrm{V\,X_1 = V\,X_2 = \frac{M_1\,M_2}{M}\,S^2},$$

deux variances égales comme il se doit puisque les deux aléatoires X_1 et X_2 sont de somme certaine, X.

pour les quantités aléatoires S_1^2 et S_2^2, la relation :

$$\text{(4)} \qquad (M_1 - 1)\,S_1^2 + (M_2 - 1)\,S_2^2 = (M - 1)\,S^2 - M_1\,(\overline{X}_1 - \overline{X})^2 - M_2\,(\overline{X}_2 - \overline{X})^2$$

puis les espérances mathématiques :

$$\text{(5)} \qquad \mathrm{E}\,S_1^2 = S^2, \qquad \mathrm{E}\,S_2^2 = S^2,$$

pour les quantités aléatoires σ_1^2 et σ_2^2, la relation :

$$\text{(6)} \qquad M_1\,\sigma_1^2 + M_2\,\sigma_2^2 = M\,\sigma^2 - M_1\,(\overline{X}_1 - \overline{X})^2 - M_2\,(\overline{X}_2 - \overline{X})^2$$

puis les espérances mathématiques :

$$\text{(7)} \qquad \mathrm{E}\,\sigma_1^2 = \frac{M_1 - 1}{M_1}\,\mathrm{E}\,S_1^2 = \frac{M_1 - 1}{M_1}\,S^2, \quad \mathrm{E}\,\sigma_2^2 = \frac{M_2 - 1}{M_2}\,S^2$$

* * *

Les échantillons aléatoires, leurs caractéristiques :

Ce sont :

	Échantillons tirés de : Population	1re strate	2e strate
Effectif	m	m_1	m_2
Rang des tirages	$i = 1, 2, \ldots, m$	$i_1 = 1, 2, \ldots, m_1$	$i_2 = 1, 2, \ldots m_2$
Valeur de (X) sur l'unité obtenue au tirage	n° i : X_i	n° i_1 : X_{i_1}	n° i_2 : X_{i_2}
Total de (X)	$x = \overset{m}{\underset{i}{S}}\,X_i$	$x_1 = \overset{m_1}{\underset{i_1}{S}}\,X_{i_1}$	$x_2 = \overset{m_2}{\underset{i_2}{S}}\,X_{i_2}$
Moyenne de (X)	$\bar{x} = x/m$	$\bar{x}_1 = x_1/m_1$	$\bar{x}_2 = x_2/m_2$
(8) Estimateur	de X : $x' = \frac{M}{m}x$	de X_1 : $x_1' = \frac{M_1}{m_1}x_1$	de $X_2 = \frac{M_2}{m_2}x_2$
		de $X = X_1 + X_2$: $x'' = x_1' + x_2'$	

Que les échantillons aléatoires soient tirés avec remise ou sans remise on, a *en espérance mathématique :*

(9) $\mathrm{E}\,x' = X$ pour le 1er procédé.

$\mathrm{E}\,x'' = \mathrm{E}\,x_1' + \mathrm{E}\,x_2'$

(10) $= \mathrm{E}\,X_1 + \mathrm{E}\,X_2$ en prenant l'espérance mathématique liée par la stratification aléatoire, x_1' et x_2' étant considérées comme aléatoires dépendant du tirage des échantillons de m_1 et m_2 dans les strates données,

$= \mathrm{E}\,(X_1 + X_2) = \mathrm{E}\,X = X.$

Ainsi, pour le second procédé, $E\,x'' = X$ quelque soit le processus de division de la population initiale en deux strates, processus aléatoire comme il a été exposé plus haut, assurant de surcroît :

$$E\,X_1 + E\,X_2 = \left(\frac{M_1}{M} + \frac{M_2}{M}\right) X = X,$$

ou processus non aléatoire, tel que classement raisonné des M unités en deux strates, ce qui est le cas usuel, mais pour lequel on ne peut plus définir les espérances mathématiques de X_1 et de X_2; on sait seulement que $X_1 + X_2 = X$.

Pour former l'expression des *variances* de x' et x'' on doit distinguer les deux cas d'échantillons tirés avec ou sans remise.

1er *cas : Échantillons tirés avec remise.*

Par le 1er procédé (sans-stratification), on a :

$$\text{(11)} \qquad Vx' = M^2\frac{\sigma^2}{m} = \frac{M}{f}\sigma^2 = \frac{M-1}{f}S^2 \text{ en posant } f = \frac{m}{M}.$$

Par le 2e procédé (avec stratification) on doit, pour prendre la variance de x'', considérer d'abord $x'' = x'_1 + x'_2$ comme aléatoire dépendant des tirages de m_1 et m_2 unités dans les strates 1 et 2 supposées données, d'où la variance conditionnelle

$$\text{(12)} \qquad V_c\,x'' = V_c\,x'_1 + V_c\,x'_2$$
$$= M_1^2\frac{\sigma_1^2}{m_1} + M_2^2\frac{\sigma_2^2}{m_2},$$

puis chercher l'espérance mathématique de cette variance conditionnelle qui s'exprime en fonction des aléatoires σ_1^2 et σ_2^2 dépendant de la stratification aléatoire de la population :

$$V\,x'' = E\,V_c\,x'' = \frac{M_1^2}{m_1}E\,\sigma_1^2 + \frac{M_2^2}{m_2}E\,\sigma_2^2$$

$$\text{(13)} \qquad V\,x'' = \frac{M_1}{m_1}(M_1 - 1)\,S^2 + \frac{M_2}{m_2}(M_2 - 1)\,S^2$$

Dans le cas particulier où $\frac{m_1}{M_1} = \frac{m_2}{M_2} = \frac{m}{M} = f$ (même fraction de sondage dans la population initiale et dans les strates aléatoires, ce qui assure bien $m_1 + m_2 = m$), la variance de x'' s'écrit

$$\text{(14)} \qquad Vx'' = \frac{1}{f}(M_1 + M_2 - 2)\,S^2 = \frac{M-2}{f}S^2,$$

alors que

$$Vx' = \frac{f}{M-1}S^2$$

Il y a un très léger gain de précision, les variances de x' et x'' étant proportionnelles à $M - 1$ et $M - 2$ respectivement.

Remarque. — On peut encore écrire la variance conditionnelle de x'' (formule [12]) sous la forme :

$$V_c x'' = \frac{1}{f}(M_1 \sigma_1^2 + M_2 \sigma_2^2) \quad \text{où} \quad \frac{1}{f} = \frac{M_1}{m_1} = \frac{M_2}{m_2}$$

soit, en tenant compte de la relation (6) puis de l'expression (11) de Vx' :

$$V_c x'' = \frac{M}{f}\sigma^2 - \frac{1}{f}\left[M_1(\overline{X}_1 - \overline{X})^2 + M_2(\overline{X}_2 - \overline{X})^2\right]$$

(15) $$Vx' - V_c x'' = \frac{1}{f}\left[M_1(\overline{X}_1 - \overline{X})^2 + M_2(\overline{X}_2 - \overline{X})^2\right],$$

formule classique qui exprime le gain de précision due à la stratification quand les deux strates sont sondées avec la même fraction de sondage (échantillon représentatif). Ici, ce gain est aléatoire comme les différences $\overline{X}_1 - \overline{X}$ et $\overline{X}_2 - \overline{X}$, dépendant du partage aléatoire de la population initiale. Il est positif, nul seulement si $\overline{X}_1 = \overline{X}_2 = \overline{X}$. Son espérance mathématique est donc forcément positive. Elle est exactement :

$$E(Vx' - V_c x'') = \frac{1}{f}(M_1 V\overline{X}_1 + M_2 V\overline{X}_2)$$

$$= \frac{1}{f}(M_2 + M_1)\frac{S^2}{M} \qquad \text{(d'après (3)}$$

(16) $$Vx' - Vx'' = \frac{1}{f}S^2 = \frac{Vx'}{M-1}$$

soit précisément la différence entre Vx'' et Vx' des formules (14) et (11).

Dans le cas plus général où l'on a seulement $m_1 + m_2 = m$, les rapports m_1/M_1 et m_2/M_2 n'étant pas égaux (même nombre total d'unités-échantillon, mais réparties non proportionnellement entre les deux strates), on posera :

$$m_1 = f M_1 + e$$

(17) $$m_2 = f M_2 - e$$

au total : $$m_1 + m_2 = f(M_1 + M_2) = m$$

et l'on étudiera la variation de Vx'', formule (13), en fonction de e. Pour des valeurs assez petites de e ;

$$\frac{1}{m_1} = \frac{1}{fM_1 + e} = \frac{1}{fM_1}\left(1 + \frac{e}{fM_1}\right)^{-1} = \frac{1}{fM_1}\left(1 - \frac{e}{fM_1} + \frac{e^2}{f^2M_1^2} - \ldots\right)$$

(18) $$\frac{1}{m_2} = \frac{1}{fM_2}\left(1 + \frac{e}{fM_2} + \frac{e^2}{f^2M_2^2} + \ldots\right)$$

D'où le développement de Vx''

$$Vx'' = \frac{S^2}{f}\left[(M_1 - 1)\left(1 - \frac{e}{fM_1} + \frac{e^2}{f^2M_1^2} - \ldots\right)\right.$$
$$\left. + (M_2 - 1)\left(1 + \frac{e}{fM_2} + \frac{e^2}{f^2M_2^2} + \ldots\right)\right]$$

(19) $$= \frac{M-2}{f}S^2 + e\frac{S^2}{f^2}\left(\frac{1}{M_1} - \frac{1}{M_2}\right) + e^2\frac{S^2}{f^3}\left(\frac{M_1 - 1}{M_1^2} + \frac{M_2 - 1}{M_2^2}\right) + \ldots$$

Ainsi Vx'' ne présente pas un minimum pour $\frac{m_1}{M_1} = \frac{m_2}{M_2}$, c'est-à-dire pour $e = 0$, le coefficient de e n'étant pas nul (à moins que $M_1 = M_2$).

En réalité, le minimum de Vx'' est obtenu pour la valeur de m_1 et m_2 satisfaisant au système

$$\begin{cases} \dfrac{M_1(M_1-1)}{m_1} + \dfrac{M_2(M_2-1)}{m_2} \text{ minimum} \\ m_1 + m_2 = m, \end{cases}$$

soit :

$$\frac{m_1}{\sqrt{M_1(M_1-1)}} = \frac{m_2}{\sqrt{M_2(M_2-1)}} = \frac{m}{\sqrt{M_1(M_1-1)} + \sqrt{M_2(M_2-1)}}, \tag{20}$$

au moins dans la mesure où l'on considère m_1 et m_2 comme des paramètres continus. Or, ce sont évidemment des nombres entiers (effectifs d'échantillons). Dans ces conditions la recherche de la valeur du minimum de Vx'' quand m_1 et m_2 satisfont à (20) n'a pas grand intérêt pratique, non plus que la comparaison de ce minimum à la valeur de Vx'' pour m_1 et m_2 proportionnels à M_1 et M_2. Notons cependant que pour M_1 et M_2 quelque peu élevés, la condition (20) est pratiquement équivalente à

$$\frac{m_1}{M_1} = \frac{m_2}{M_2} = \frac{m}{M}.$$

2e cas : Échantillons tirés sans remise

Pour l'estimateur sans stratification, x', la variance est :

$$Vx' = M^2 \frac{M-m}{M} \frac{S^2}{m} = M \frac{1-f}{f} S^2, \quad \frac{M}{m} = \frac{1}{f}. \tag{21}$$

Pour obtenir la variance de l'estimateur avec stratification on doit considérer d'abord $x'' = x'_1 + x'_2$ comme aléatoire dépendant des tirages de m_1 et m_2 unités dans les strates 1 et 2 supposées données, d'où la variance conditionnelle :

$$\begin{aligned} V_c x'' &= V_c x'_1 + V_c x'_2 \\ &= \frac{M_1^2}{m_1} \frac{M_1 - m_1}{M_1} S_1^2 + \frac{M_2^2}{m_2} \frac{M_2 - m_2}{M_2} S_2^2, \end{aligned} \tag{22}$$

puis chercher l'espérance mathématique de cette variance conditionnelle, aléatoire dépendant de la stratification aléatoire de la population par S_1^2 et S_2^2 :

$$Vx'' = E\, V_c x'' = \left(\frac{M_1^2}{m_1} \frac{M_1 - m_1}{M_1} + \frac{M_2^2}{m_2} \frac{M_2 - m_2}{M_2} \right) S^2$$

puisque $E\, S_1^2 = E\, S_2^2 = S^2$ (formule [5]), soit encore en regroupant les termes entre parenthèses :

$$Vx'' = \left(\frac{M_1^2}{m_1} + \frac{M_2^2}{m_2} - M \right) S^2 \tag{23}$$

Les deux paramètres m_1 et m_2 étant astreints à la condition :

$$m_1 + m_2 = m \text{ donnée,}$$

le minimum de Vx'' a lieu quand :

$$\frac{m_1}{M_1} = \frac{m_2}{M_2} = \frac{m}{M} = f,$$

c'est-à-dire *pour une même fraction de sondage* f *dans les deux strates.*

Ce minimum est alors :

$$Vx'' = \left[\frac{1}{f}(M_1 + M_2) - M\right] S^2 = M\,\frac{1-f}{f}\,S^2 \tag{24}$$

soit exactement la valeur de Vx', variance de l'estimateur sans stratification aléatoire (formule [21]).

Ainsi, quand les échantillons sont tirés sans remise, la stratification aléatoire de la population initiale n'apporte aucun gain en précision.

On peut étudier la variation de Vx'' autour de son minimum en posant encore :

$$m_1 = f\,M_1 + e$$
$$m_2 = f\,M_2 - e,$$

ce qui donne les développements (18) pour $1/m_1$ et $1/m_2$ quand e est petit. Le développement de Vx'' est alors :

$$Vx'' = M\,\frac{1-f}{f}\,S^2 + e^2\,\frac{S^2}{f^3}\left(\frac{1}{M_1} + \frac{1}{M_2}\right) + \ldots\ldots \tag{25}$$

Cette fois il n'y a pas de terme en e ; le minimum a lieu pour m_1 et m_2 proportionnels à M_1 et M_2 et c'est $M\,\dfrac{1-f}{f}\,S^2$, soit Vx'.

Remarque. — Dans le cas d'un tirage sans remise on ne peut exprimer simplement la différence entre Vx' et V_cx'' en fonction des différences $(\overline{X}_1 - \overline{X})^2$ et $(\overline{X}_2 - \overline{X})^2$ comme dans le cas d'un tirage avec remise (formule [15]). En effet, si l'on suppose $m_1 = f\,M_1$ et $m_2 = f\,M_2$:

$$Vx' - V_cx'' = \frac{1-f}{f}\,(M\,S^2 - M_1\,S_1^2 - M_2\,S_2^2), \tag{26}$$

alors que la relation entre S_1^2, S_2^2 et S^2 (formule [4]) fait intervenir les coefficients $M_1 - 1$, $M_2 - 1$ et $M - 1$. On voit, en revanche, sur (26) que l'espérance mathématique de $Vx' - V_cx''$ est nulle exactement puisque $E\,S_1^2 = E\,S_2^2 = S^2$.

Résumé et conclusion

On peut schématiser comme suit la comparaison de Vx' (sans stratification) et Vx'' (avec stratification) :

m_1 et m_2	Échantillons tirés avec remise	Échantillons tirés sans remise
proportionnels à M_1 et M_2.	$Vx' > Vx''$	$Vx' = Vx''$
non proportionnels à M_1 et M_2	$Vx' \gtrless Vx''$ (en fait $<$)	$Vx' < Vx''$

Précisons l'importance des écarts entre Vx' et Vx'' par quelques exemples numériques pour lesquels on a calculé les rapports :

$$\frac{Vx''}{Vx'} = \frac{\dfrac{M_1(M_1-1)}{m_1} + \dfrac{M_2(M_2-1)}{m_2}}{\dfrac{M(M-1)}{m}} \qquad \textbf{échantillons avec remise}$$

$$\frac{Vx''}{Vx'} = \frac{\dfrac{M_1(M_1-m_1)}{m_1} + \dfrac{M_2(M_2-m_2)}{m_2}}{\dfrac{M(M-m)}{m}} \qquad \textbf{échantillons sans remise.}$$

Ces rapports sont donnés au tableau ci-après, ainsi que la part de Vx'' due au sondage dans chacune des deux strates.

On constate que l'avantage de la stratification est très faible ou nul dans le cas optimum où l'on sonde les deux strates avec une même fraction.

Dans le cas de fractions inégales dans les deux strates, l'avantage de la stratification disparaît complètement devant l'augmentation considérable de Vx'' due précisément à cette représentation non proportionnelle des deux strates, ainsi que le veut la théorie de l'estimation.

COMPARAISON de Vx' et de Vx''.

M effectif total , m effectif échantillon dans population initiale.
M_1 — , m_1 — dans 1^re^ strate
M_2 — , m_2 — dans 2^e^ strate
$(M_1 + M_2 = M)$ $(m_1 + m_2 = m)$

		ÉCHANTILLONS AVEC REMISE			ÉCHANTILLONS SANS REMISE		
m_1	m_2	I	II	Vx''/Vx'	I	II	Vx''/Vx'
1^er^ exemple M = 12, M_1 = 6, M_2 = 6; m = 6 (soit $f = m/M = 1/2$).							
1	5	1,364	0,273	1,636	2,500	0,100	2,600
2	4	0,682	0,341	1,023	1,000	0,250	1,250
3	*3*	*0,455*	*0,455*	*0,909*	*0,500*	*0,500*	*1,000*
2^e^ exemple : M = 12, M_1 = 8, M_2 = 4; m = 6 (soit $f = m/M = 1/2$).							
1	5	2,545	0,109	2,655	/////	/////	/////
2	4	1,273	0,136	1,409	2.000	—	2,000
3	3	0,849	0,182	1,030	1,111	0,111	1,223
4	*2*	*0,636*	*0,273*	*0,909*	*0,667*	*0,333*	*1,000*
a) 4, 10	1,90	(0,621)	(0,287)	(0,908)	(.....)	(.....)	(.....)
5	1	0,509	0.545	1,055	0,400	1,000	1,400
3^e^ exemple : M = 100, M_1 = 50, M_2 = 50; m = 10 (soit $f = m/M = 1/10$).							
1	9	2,475	0,275	2,749	2,722	2,533	2,976
2	8	1,237	0,309	1 546	1,333	0,292	1,626
3	7	0,825	0,354	1,179	0,870	0,341	1,211
4	6	0,619	0,412	1,031	0,639	0,408	1,047
5	*5*	*0,495*	*0,495*	*0,990*	0,500	0,500	1,000
4^e^ exemple : M = 100, M_1 = 70, M_2 = 30 ; m = 10 (soit f = m/M = 1/10)							
1	9	4,879	0,098	4,977	5,367	0,078	5,444
2	8	2,439	0,110	2,549	2,644	0,092	2,737
3	7	1,626	0,125	1,752	1,737	0,110	1,847
4	6	1,220	0,146	1,367	1,283	0,133	1,417
5	5	0,976	0,176	1,152	1,011	0,167	1,178
6	4	0,813	0,220	1,033	0,830	0,217	1,047
7	*3*	*0,697*	*0,293*	*0,990*	*0,700*	*0,300*	*1,000*
8	2	0,610	0,439	1,049	0.603	0,467	1,070
9	1	0,542	0,879	1,421	0,527	0,967	1,493

I et II, parts de Vx'' (en fraction de Vx') dues au sondage dans les 1^re^ et 2^e^ strates respectivement.
Lignes en italique : correspondent à une même fraction de sondage dans les deux strates.
a) valeurs optima de m_1 et m_2 dans le cas d'échantillons avec remise. (Pour M = 100, M_1 = 70, M_2 = 30, ce sont $m_1 = 7{,}02$ et $m_2 = 2{,}98$, valeurs très voisines de la répartition proportionnelles, $m_1 = 7$, $m_2 = 3$).

Comment expliquer le léger avantage de la stratification dans le cas d'échantillons avec remise et d'effectifs proportionnels aux effectifs des strates?

La stratification étant aléatoire, chaque strate est en moyenne aussi hétérogène que la population totale. Ce n'est donc pas dans ce sens que l'on doit chercher la raison de la réduction de variance de l'estimation de X.

Le seul effet de la stratification aléatoire semble être de réduire, dans une mesure dépendant du nombre de strates, la possibilité d'observer plusieurs fois la même unité de la population. S'il n'y a pas de strates, la même unité peut sortir jusqu'à m fois dans l'échantillon; s'il y a deux strates, elle peut sortir un nombre de fois au plus égal au plus grand des nombres m_1 et m_2.

En poussant la stratification aléatoire à l'extrême, c'est-à-dire en faisant m strates, de même effectif $M_h = M/m$ et dans chacune desquelles on prélèverait $m_h = 1$ unité seulement ($h = 1, 2, \ldots, m$), on aurait l'estimation $x^{(m)}$ de X:

$$x^{(m)} = \sum_h^m \frac{M_h}{m_h} X_h \qquad \text{où } X_h \text{ est l'observation unique faite dans la strate } h,$$

$$= \frac{M}{m} \sum_h^m X_h,$$

dont la variance serait :

$$Vx^{(m)} = E\, V_e\, x^{(m)} = \frac{M^2}{m^2} \sum_h^m E\,\sigma_h^2$$

où σ_h^2 est la variance aléatoire dans la strate h, d'espérance mathématique (cf. formule [7]) :

$$E\,\sigma_h^2 = \frac{M_h - 1}{M_h} S^2 \text{ avec } M_h = M/m,$$

d'où

$$Vx^{(m)} = \frac{M^2}{m} \frac{M - m}{M} S^2,$$

soit Vx' où x' est l'estimation de X dans le cas d'un échantillon sans remise d'effectif m tiré dans la population de M (formule [21]). Dans ce cas extrême, l'effet de la stratification est de ramener les tirages avec remise à des tirages sans remise (1).

---

(1) Remarquer la loi de formation de la variance :

Nombre de strates aléatoires	Estimation de X	Variance de l'estimateur
1 (population initiale)	x'	$Vx' = \frac{M^2}{m} \frac{M-1}{M} S^2$
2	x''	$Vx'' = \frac{M^2}{m} \frac{M-2}{M} S^2$
.....		
h	$x^{(h)}$	$Vx^{(h)} = \frac{M^2}{m} \cdot \frac{M-h}{M} S^2$
.....		
m	$x^{(m)}$	$Vx^{(m)} = \frac{M^2}{m} \frac{M-m}{M} S^2$

Il est évident que si les tirages sont déjà effectués sans remise, l'effet de la stratification est nul : $Vx'' = Vx'$.

Donnons encore un exemple concret. Supposons que la population soit constituée de M boules dont certaines sont rouges et les autres blanches. Si l'on tire m boules avec remise et que l'on observe 2 rouges, on n'est pas certain qu'il y ait 2 rouges dans la population : ce peut être la même boule unique qui est sortie deux fois. Si maintenant on répartit ces M boules en deux strates en comprenant M_1 et M_2 $(M_1 + M_2 = M)$, le partage étant fait en ignorant les couleurs, puis que l'on tire m_1 boules avec remise parmi les M_1 et que l'on obtienne 1 rouge au moins une fois, tandis qu'un tirage avec remise de m_2 boules parmi les M_2 donne aussi 1 rouge au moins une fois, on est assuré qu'il y avait dans la population initiale de M au moins 2 boules rouges. On a évidemment la même certitude si l'on a tiré sans remise m boules parmi les M et observé 2 rouges.

Paris, le 1er décembre 1955.

F. Chartier.

***