REVUE DE STATISTIQUE APPLIQUÉE

P. FERIGNAC

Note sur l'estimation de la proportion de pièces défectueuses dans le contrôle de réception basé sur un échantillon double

Revue de statistique appliquée, tome 5, nº 3 (1957), p. 11-15
<http://www.numdam.org/item?id=RSA_1957__5_3_11_0>

NOTE SUR L'ESTIMATION DE LA PROPORTION DE PIÈCES DÉFECTUEUSES DANS LE CONTROLE DE RÉCEPTION BASÉ SUR UN ÉCHANTILLON DOUBLE

par

P. FERIGNAC

Lorsqu'on réceptionne des lots contenant la proportion p de pièces défectueuses, il faut, à intervalles plus ou moins réguliers, estimer p pour juger de la qualité globale des lots soumis au contrôle.

Dans le cas d'un échantillonnage double, l'estimation moyenne, basée suivant le stade où la décision est obtenue, soit sur le premier échantillon, soit sur les deux échantillons, comporte un biais dû au prélèvement aléatoire du second échantillon.

L'objet de la note de M. Férignac est de montrer comment on peut estimer ce biais pour un plan d'échantillonnage donné.

1. PROBLÈME

Soit un plan de contrôle par double échantillonnage basé sur les données ci-dessous :

1er Échantillon : effectif n_1, acceptation si $k_1 \leqslant a_1$, refus si $k_1 \geqslant r_1$, prélèvement d'un 2e échantillon si $a_1 < k_1 < r_1$; a_1, r_1 et k_1, étant respectivement le nombre de pièces défectueuses acceptable , non acceptable et observé par échantillon.

2e Échantillon : effectif n_2, acceptation si $k_2 \leqslant a_2$, refus si $k_2 \geqslant r_2$, a_2, r_2 et k_2 étant respectivement le nombre de pièces défectueuses acceptable, non acceptable et observé dans l'échantillon total d'effectif $n_1 + n_2$ (a_2 et r_2 sont deux nombres entiers consécutifs).

On a une estimation correcte de p en formant la moyenne des $\frac{k_1}{n_1}$ pour les lots soumis au contrôle; cette estimation n'utilise que l'information des premiers échantillons; si l'on estimait p par la moyenne de la proportion de pièces défectueuses observée dans les premiers ou les premiers et deuxièmes échantillons, suivant le stade auquel la décision sur le lot est obtenue, l'estimation comporterait un biais à cause du prélèvement conditionnel du 2e échantillon.

Il est intéressant de calculer ce biais : c'est le but de cette étude dans le cas où aucun des échantillons n'est tronqué, c'est-à-dire où l'on n'arrête pas l'examen de l'échantillon dès qu'une décision est possible.

2. CALCUL DU BIAIS

Représentant par f_i une fréquence du nombre de pièces défectueuses et par P_i sa probabilité, la valeur théorique de la fréquence moyenne f dans les échantillons est :

$$f = \sum_i P_i \cdot f_i \qquad (1)$$

les i étant étendus à toutes les situations possibles. La somme du membre de droite de (1) peut être décomposée en deux parties :

- les i étendus à toutes les éventualités d'acceptation ou de refus au 1er échantillon, soit f_1,

- les i étendus à tous les cas où l'on doit prélever un 2e échantillon, soit f_2,

$$f = f_1 + f_2$$

1er CAS : ON A UNE DÉCISION D'ACCEPTATION OU DE REFUS AU 1er ÉCHANTILLON

f_i est une variable aléatoire discrète qui peut prendre les valeurs : $\frac{0}{n_1}$, $\frac{1}{n_1}, \ldots, \frac{a_1}{n_1}, \frac{r_1}{n_1}, \frac{r_1+1}{n_1}, \ldots, \frac{n_1}{n_1}$, avec les probabilités respectives : P(0), P(1),...., $P(a_1)$, $P(r_1)$, $P(r_1+1) \ldots P(n_1)$, , où P(x) est la probabilité d'obtenir x pièces défectueuses quand on fait n répétitions ; les probabilités P(x) sont données par la loi binomiale $(p+q)^{n_1}$. On a :

$$f_1 = \frac{0}{n_1} \cdot P(0) + \frac{1}{n_1} \cdot P(1) + \ldots + \frac{a_1}{n_1} P(a_1) + \frac{r_1}{n_1} p(r_1) + \frac{r_1+1}{n_1} \cdot p(r_1+1) + \ldots$$

$$\ldots + \frac{n_1}{n_1} p(n_1) \qquad (2)$$

Tenant compte de la relation

$$\frac{0}{n_1} \cdot P(0) + \frac{1}{n_1} \cdot P(1) + \ldots + \frac{a_1}{n_1} \cdot P(a_1) + \frac{a_1+1}{n_1} \cdot P(a_1+1) + \ldots + \frac{n_1}{n_1} \cdot P(n_1) = p \qquad (3)$$

On peut écrire :

$$f_1 = p - \left[\frac{a_1+1}{n_1} \cdot P(a_1+1) + \frac{a_1+2}{n_1} \cdot P(a_1+2) + \ldots + \frac{r_1-1}{n_1} \cdot P(r_1-1) \right]$$

$$f_1 = p - f' \qquad (4)$$

où

$$f' = \frac{a_1+1}{n_1} \cdot P(a_1+1) + \frac{a_1+2}{n_1} \cdot P(a_1+2) + \ldots + \frac{r_1-1}{n_1} \cdot P(r_1-1) \qquad (5)$$

f' est la fréquence moyenne des pièces défectueuses pour l'intervalle de la coupure dans lequel le premier échantillon ne permet pas de conclure par l'acceptation ou le refus du lot.

2^e CAS : LA DÉCISION D'ACCEPTATION OU DE REFUS EST PRISE AU 2° ÉCHANTILLON

La probabilité de prélèvement du 2° échantillon est :

$$P_2 = P(a_1 + 1) + P(a_1 + 2) + \ldots + P(r_1 - 1) \qquad (6)$$

Le 2° échantillon n'étant pas tronqué il comporte, en moyenne, $n_2 p$ pièces défectueuses et la fréquence moyenne des pièces défectueuses après le 2° échantillon est :

$$f_2 = \frac{a_1 + 1 + n_2 p}{n_1 + n_2} \cdot P(a_1 + 1) + \frac{a_1 + 2 + n_2 p}{n_1 + n_2} \cdot P(a_1 + 2) + \ldots + \frac{r_1 - 1 + n_2 p}{n_1 + n_2} \cdot P(r_1 - 1)$$

ou d'après (6) :

$$f_2 = \frac{n_2 p}{n_1 + n_2} \cdot P_2 + \left[\frac{a_1 + 1}{n_1} \cdot P(a_1 + 1) + \frac{a_1 + 2}{n_1} \cdot P(a_1 + 2) + \ldots + \frac{r_1 - 1}{n_1} \cdot P(r_1 - 1)\right] \frac{n_1}{n_1 + n_2}$$

soit, en tenant compte de (5) :

$$f_2 = \frac{n_2 p}{n_1 + n_2} \cdot P_2 + \frac{n_1}{n_1 + n_2} \cdot f' \qquad (7)$$

Finalement, des égalités (4) et (7) on déduit :

$$f = p - f' + \frac{n_2 p}{n_1 + n_2} \cdot P_2 + \frac{n_1 f'}{n_1 + n_2}$$

$$f = p + \frac{n_2}{n_1 + n_2}(p P_2 - f') \qquad (8)$$

La relation (8) prouve que l'espérance mathématique de la fréquence théorique des pièces défectueuses, calculée dans l'ensemble des 2 échantillons possibles, est différente de sa vraie valeur, p, dans le lot.

L'estimation de p qui tient compte de toutes les pièces prélevées au lieu des seules pièces des premiers échantillons, comporte un biais :

$$\varepsilon = f - p = \frac{n_2}{n_1 + n_2}(P_2 \cdot p - f')$$

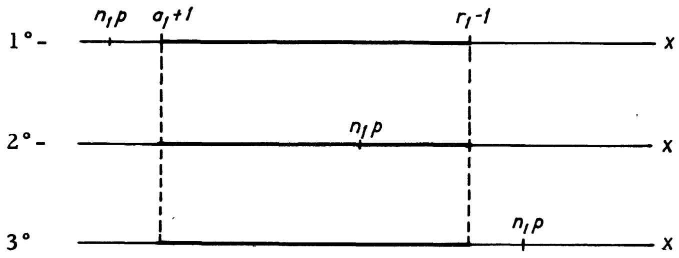
Dans l'équation (8), P_2 et f' dépendent de p et il n'est pas facile de tirer p en fonction de f. Dans l'exemple numérique que nous développons plus loin, on voit que la solution graphique de cette question est facile.

Le signe du biais varie. Il peut être prévu dans un certain domaine facile à déterminer. Pour cela, portons notre attention sur le signe de $(P_2 . p - f')$, qui est du même signe que le biais, ou $E = (P_2 . n_1 p - n_1 f')$. Cette dernière expression, tenant compte des relations (5) et (6) peut s'écrire :

$$E = \left[n_1 p - (a_1 + 1)\right] . P(a_1 + 1) + \left[n_1 p - (a_1 + 2)\right] P(a_1 + 2) + \ldots + \left[n_1 p - (r_1 - 1)\right] P(r_1 - 1)$$

qui représente la valeur moyenne théorique des écarts du nombre des pièces défectueuses x par rapport à sa valeur probable, $n_1 p$, changée de signe, soit $E = -$ Moyenne $(x_1 - n_1 p)$, lorsque $a_1 + 1 \leq x \leq r_1 - 1$, c'est-à-dire lorsque x est situé dans l'intervalle d'indécision pour le lot après le 1er échantillon.

On peut avoir une des dispositions relatives ci-dessous :



$n_1 p \leqslant a_1 + 1$

$a_1 + 1 < n_1 p < r_1 - 1$

$n_1 p \geqslant r_1 - 1$

au 1°, tous les écarts étant positifs, le biais est négatif,

au 2°, le sens des écarts est variable, le signe du biais dépend de la valeur des écarts et de $P(x)$,

au 3°, tous les écarts $(x - n_1 p)$ étant négatifs, le biais est positif.

En résumé, lorsque p croît, le biais passe d'une valeur négative à une valeur positive en passant par la valeur zéro pour une valeur de p comprise entre $\frac{a_1 + 1}{n_1}$ et $\frac{r_1 - 1}{n_1}$

3. EXEMPLE NUMÉRIQUE

Soit le plan de double échantillonnage défini par :

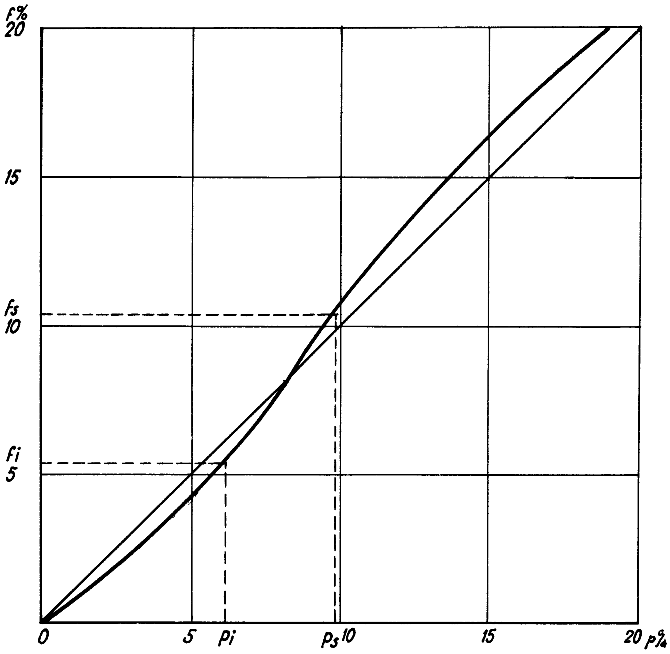
$n_1 = 35$ $\qquad a_1 = 1$ $\qquad r_1 = 5$

$n_2 = 70$ $\qquad a_2 = 4$ $\qquad r_2 = 5$

On calcule le biais et la fréquence moyenne f pour diverses valeurs de p. Les résultats sont consignés dans le tableau ci-dessous où P_2 et f' sont tirés des tables de la distribution binomiale.

P	P_2	f'	$\frac{n_2}{n_1 + n_2}$	Biais	Fréquence moyenne	Correction relative
0,02	0,1521	0,0095	$\frac{2}{3}$	- 0,0043	0,0157	- 21,5%
0,04	0,3987	0,0282	"	- 0,0082	0,0318	- 20,5%
0,07	0,6185	0,0492	"	- 0,0039	0,0661	- 5,6%
0,10	0,6084	0,0526	"	+ 0,0055	0,1055	+ 5,5%
0,15	0,3564	0,0336	"	+ 0,0132	0,1632	+ 8,8%
0,20	0,1395	0,0139	"	+ 0,0090	0,2090	+ 4,5%

Les résultats contenus dans le tableau sont traduits dans le graphique ci-dessous.



Valeur moyenne de la fréquence f
en fonction de la vraie valeur p

Si l'on applique souvent le même plan d'échantillonnage double, on peut construire un graphique analogue au précédent sur lequel on lira la vraie valeur de p correspondant à la valeur moyenne de la fréquence des pièces défectueuses estimée d'après le total des pièces prélevées.

On a représenté sur le graphique la droite de biais nul. D'après ce que nous avons vu sur le signe du biais, selon la situation de $n_1 p$ par rapport aux nombres $a_1 + 1$ et $r_1 - 1$, il s'annule pour une certaine valeur de p. Dans le cas envisagé, on voit sur le graphique que le biais est nul lorsque p = 0,08, l'estimation de p est correcte lorsque le lot contient 8 % de pièces défectueuses, dans tous les autres cas, elle est entachée d'un biais (Le cas p = 0 est omis car son estimation est toujours correcte).

4. CONCLUSION

L'estimation de la proportion de pièces défectueuses, p, d'après toutes les pièces prélevées est plus précise que celle qui porte seulement sur les premiers échantillons, car le nombre des observations est plus grand. Les limites de confiance de l'estimation de f sont plus resserrées ; soient f_s et f_i les limites de confiance supérieure et inférieure, pour obtenir les limites de confiance de p il suffira de lire sur le graphique les abscisses p_s et p_i qui correspondent à f_s et f_i. Les valeurs p_s et p_i sont débarrassées du biais qui entachait f_s et f_i. Cette manière de procéder permet d'utiliser la totalité de l'information recueillie au cours de l'échantillonnage.

Le schéma de calcul appliqué dans le cas d'un échantillonnage double se généralise au cas d'un échantillonnage multiple non tronqué.