

Estimation de la proportion d'erreurs de filiation dans un troupeau ovin à l'aide des marqueurs biochimiques sanguins (Étude théorique)

B. BONAÏTI*

** Station de génétique quantitative et appliquée,
Centre national de Recherches zootechniques, I.N.R.A.,
78350 Jouy en Josas.*

Résumé

Les marqueurs biochimiques sanguins permettent d'estimer sur une population donnée la proportion d'erreurs de filiation : l'étude statistique de cette estimation permet de choisir le marqueur le plus efficace (variance minimum) pour estimer cette proportion puis de calculer les erreurs inhérentes au test d'hypothèse. Ceci donne un moyen pour distinguer, avec une marge d'erreur connue, les troupeaux présentant des filiations suffisamment fiables pour permettre une sélection sur descendance efficace.

Les marqueurs biochimiques sanguins présentent un grand intérêt pour le diagnostic de paternité. Ils permettent de refuser ou non le père d'un individu ou de choisir le père le plus probable parmi un ensemble de pères possibles. L'étude probabiliste de ce problème a été faite par RENDEL (1958). Elle peut aussi être empruntée aux études de génétique humaine (JACQUARD et SALMON, 1971). L'utilisation des marqueurs biochimiques sanguins permet aussi d'évaluer la proportion moyenne d'erreurs de filiation dans un troupeau. Ceci permettrait de connaître a priori l'efficacité de la sélection sur descendance dans les troupeaux de moutons (BONAÏTI, 1974). L'évaluation de cette proportion d'erreurs se fait sur un échantillon de filiations déclarées (*t* ensembles constitués chacun par un père, une mère et leur produit déclaré). L'analyse des marqueurs biochimiques sanguins sur chacun des trois individus permet d'exclure une filiation déclarée, si elle est fausse, avec une probabilité donnée (*p*). Cette note étudie la liaison entre le nombre (*X*) de filiations erronées dans un troupeau d'effectif (*n*) et le nombre (*Y*) de filiations reconnues fausses dans l'échantillon étudié.

I. — Exposé théorique du problème

La liaison entre les variables X et Y est compliquée par le fait que le nombre réel de filiations (Z) erronées dans l'échantillon étudié est variable même si X est constant. La loi de distribution de la variable aléatoire Z , connaissant X , est une loi hypergéométrique de paramètres n , t et X :

$$P(Z = z) = \frac{C_X^z C_{n-X}^{t-z}}{C_n^t} \quad \text{si } 0 \leq z \leq X$$

$$= 0 \quad \text{si } z > X$$

La loi de distribution de la variable aléatoire Y , connaissant Z , est une loi binomiale de paramètres p et Z .

$$P(Y = y | Z = z) = C_z^y p^y (1-p)^{z-y} \quad \text{si } 0 \leq y \leq z$$

$$= 0 \quad \text{si } y < z$$

La loi de distribution de Y se déduit de la loi conditionnée de Y et de la loi de Z .

$$P(Y = y) = \sum_{z=0}^{z=t} P(Y = y | Z = z) \cdot P(Z = z)$$

$$P(Y = y) = \sum_{z=0}^{z=s} C_z^y p^y (1-p)^{z-y} \frac{C_X^z C_{n-X}^{t-z}}{C_n^t}$$

(s est la valeur la plus petite parmi t et X).

L'espérance de Y est égale à l'espérance de l'espérance conditionnée par Z .

$$E(Y | Z) = pZ$$

$$E(Y) = \frac{ptX}{n}$$

La variance de Y est obtenue en utilisant l'identité :

$$\text{var}(Y) = \text{var}^Z \{E(Y | Z)\} + E^Z \{\text{var}(Y | Z)\}$$

$$\text{var}(Y) = \text{var}(pZ) + E\{p(1-p)Z\} = p^2 \text{var}(Z) + p(1-p) E(Z)$$

Soit
$$\text{var}(Y) = \frac{ptX}{n} \left[1 - p + p \left(\frac{n-X}{n} \right) \left(1 - \frac{t-1}{n-1} \right) \right]$$

La loi de distribution de Y tend vers une loi binomiale, de paramètres $\left(\frac{px}{n}\right)$ et t , lorsque n tend vers l'infini puisque le nombre Y tend alors vers le nombre de résultats positifs (mise en évidence d'une erreur de filiation avec une probabilité égale à $\frac{px}{n}$) parmi t épreuves indépendantes. La fonction de répartition exacte de Y peut être facilement calculée grâce à deux relations de récurrence.

Si nous appelons $A(z, y)$ la quantité :

$$C_z^y p^y (1-p)^{z-y} \frac{C_X^z C_{n-X}^{t-z}}{C_n^t}$$

$$A(z, y) = A(z-1, y) \cdot \frac{(1-p)(X-z+1)(t-z+1)}{(z-y)(n-X-t+z)}$$

$$A(y, y) = A(y-1, y-1) \cdot \frac{p(X-y+1)(t-y+1)}{y(n-X-t+y)}$$

L'estimateur (H) du nombre (X) d'erreurs de filiation dans un troupeau, qui doit être tel que son espérance soit égale à X, sera :

$$H = \frac{n}{pt} Y$$

La variance de cet estimateur (n, x étant fixé) se déduit de celle de Y :

$$\text{var}(H) = \frac{n}{pt} X \left[1 - p + p \left(\frac{n-X}{n} \right) \left(1 - \left(\frac{t-1}{n-1} \right) \right) \right]$$

II. — Conséquences pratiques

I. — Choix des marqueurs biochimiques sanguins

Le coût de l'estimation de la proportion d'erreurs de filiations d'un troupeau dépend des marqueurs biochimiques sanguins choisis auxquels est liée une probabilité d'exclusion (p) et de l'effectif de l'échantillon (t). Plusieurs marqueurs sont actuellement utilisables par le laboratoire de génétique biochimique de l'INRA. Les probabilités d'exclusion en race Ile de France et les coûts d'analyse (c) associés à chacun d'eux sont présentés dans le tableau 1. Ces valeurs sont liées aux protocoles d'analyse utilisés et aux contraintes propres du laboratoire où seront réalisées les analyses. Pour un coût total d'analyse donné (C), le choix des marqueurs biochimiques sanguins devra minimiser la variance de l'estimateur H.

Une solution générale peut être donnée à ce problème. La variance de l'estimateur (H) peut s'écrire sous la forme :

$$\text{Var}(H) = \frac{nX}{pt} \left[1 - p \left(\frac{X-1}{n-1} \right) \right] - \frac{X(n-X)}{n-1}$$

En remplaçant la taille (t) de l'échantillon par $\frac{C}{kc}$, k étant le nombre moyen d'analyses à effectuer pour le contrôle d'une filiation (ce nombre sera compris entre 2 et 3) cette expression devient :

$$\text{var}(H) = \frac{nXk}{C} \frac{c}{p} \left[1 - p \left(\frac{X-1}{n-1} \right) \right] - X \frac{n-X}{n-1}$$

Parmi un ensemble de couples de valeurs (c, p) possibles, Var H sera minimum lorsque l'expression $f = \frac{c}{p} \left[1 - p \left(\frac{X-1}{n-1} \right) \right]$ sera minimum. Le couple de valeurs

TABLEAU I
Marqueurs biochimiques sanguins
(cas de la race Ile de France)
Biochemical blood markers

Marqueurs	Probabilité d'exclusion (p)	Coût d'analyse par individu (c)
1 — Groupes sanguins, Transferrines et Hé- moglobines	0,8	28 F
2 — Groupes sanguins	0,6	20 F
3 — Transferrines et Hé- moglobines	0,5	8 F
4 — Tranferrines	0,44	6 F
5 — Hémoglobine	0,13	2 F
6 — Anhydrase carboni- que et protéine X	0,08	6 F

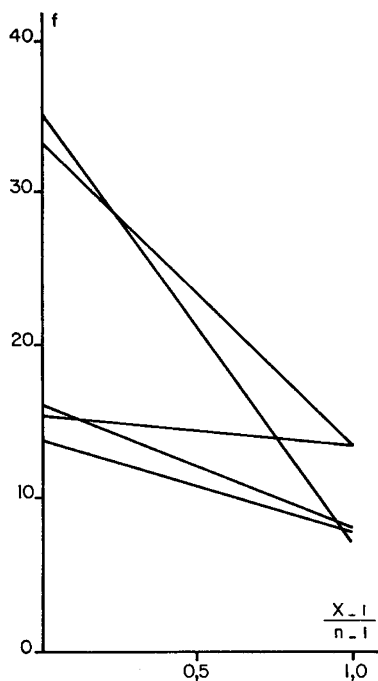


FIG. 1. — Variation of the expression according to the value of parameters c and p .
 Variation de l'expression selon les valeurs des paramètres c et p

(c, p) rendant minimum $\text{Var}(H)$ ne sera donc dépendant que du rapport $(X - 1) / (n - 1)$. Mais la figure 1, présentant, pour chacun des couples de valeurs (c, p) présentés au tableau 1, les variations de f en fonction du rapport $(X - 1) / (n - 1)$, montre que si ce rapport est inférieur à 0,97, l'expression f sera minimum pour le couple ($p = 0,44, c = 6$ francs). Cette condition sera généralement réalisée. L'utilisation des transferrines rendra donc minimum la variance de l'estimateur du nombre d'erreurs de filiations. Ce choix est seulement valable pour la race Ile de France d'où proviennent les valeurs du paramètre p utilisées dans cette étude. Mais il est indépendant de la taille du troupeau et de la proportion d'erreur pré-supposée.

2. — *Possibilité de détection des troupeaux présentant une proportion d'erreurs supérieure à un niveau donné*

L'utilisation des marqueurs biochimiques sanguins peut permettre aux organismes chargés de la sélection dans l'espèce ovine de reconnaître les troupeaux dont les filiations ne sont pas suffisamment fiables pour permettre une sélection sur descendance efficace. Pour chaque troupeau, il peut s'agir de choisir entre deux hypothèses :

- la proportion d'erreur $\left(\frac{X}{n}\right)$ est inférieure à q_0 (Hypothèse H_0)
- la proportion d'erreur $\left(\frac{X}{n}\right)$ est supérieure à q_0 (Hypothèse H_1).

Dans ce but une vérification des liens de parenté est effectuée sur un échantillon de t descendants. Si le nombre d'erreurs relevé (Y) est inférieur ou égal à une limite L , l'hypothèse H_0 sera retenue. Dans le cas contraire, l'hypothèse H_1 le sera.

Les deux risques d'erreurs inhérents à ce test d'hypothèse peuvent être calculés. Ils dépendent tous deux de la valeur supposée du rapport $\frac{X}{n}$. Le risque de première espèce (probabilité d'exclure l'hypothèse H_0 alors qu'elle est vraie) sera majoré en prenant la valeur q_0 pour le rapport $\frac{X}{n}$ (c'est le niveau du test). Le risque de seconde espèce (probabilité de retenir l'hypothèse H_0 alors qu'elle est fautive) peut être donné pour une valeur q_1 du rapport $\frac{X}{n}$. Ce risque de seconde espèce particulier est un majorant du risque de seconde espèce pour l'ensemble des troupeaux dont la proportion d'erreur $\frac{X}{n}$ est supérieure à q_1 . Le risque de seconde espèce est, par contre, plus élevé pour les troupeaux dont la proportion est comprise entre q_0 et q_1 . La valeur q_1 doit donc être choisie de telle sorte qu'il soit peu important de situer dans l'une ou l'autre catégorie les troupeaux dont la proportion d'erreurs est comprise entre q_0 et q_1 .

Nous avons calculé ces risques d'erreur pour quelques valeurs particulières des paramètres $\left(n, \frac{X}{n}, p\right)$: tableau 2. La limite L a été choisie pour que le risque de première espèce (α_1) soit le plus proche de 5 p. 100, tout en restant inférieur à 5 p. 100. Les risques α_2 et α_3 sont des risques de seconde espèce calculés respecti-

TABLEAU 2

Risques de première et de seconde espèce
Probability of first and second type errors

$1 / p = 0,44$ (utilisation des transferrines).

t	L	$N \rightarrow \infty$			$N = 300$			$N = 100$		
		α_1	α_2	α_3	α_1	α_2	α_3	α_1	α_2	α_3
1	0	4,4	78,0	99,0	4					
8	1	4,5	44,6	78,3	4,5	44,5	78,3	4,3	44,4	78,4
9	2	4,9	17,8	65,1	4,6	17,5	65,1	4,1	17,0	65,1
1	3	4,6	6,7	55,1	4,2	6,4	55,0	3,4	5,8	54,7
5	4	4,7	1,9	43,9	4,1	1,7	43,5	2,9	1,4	42,4
0	5	4,8	0,4	34,1	4,1	0,4	33,3	2,4	0,2	31,3
6	6	5,0	0,1	25,6	4,0	0,1	24,3	1,9	0,0	21,3
2	7	5,0	0,0	19,4	3,8	0,0	17,8	1,3	0,0	14,0

$2 / p = 0,77$ (utilisation des transferrines et des groupes sanguins)

t	L	$N \rightarrow \infty$			$N = 300$			$N = 100$		
		α_1	α_2	α_3	α_1	α_2	α_3	α_1	α_2	α_3
1	0									
4	1	3,2	50,1	83,1	3,1	50,1	83,1	3,0	50,0	83,3
11	2	4,7	14,0	64,2	4,5	13,7	64,3	4,0	13,1	64,3
18	3	4,5	4,3	53,4	4,1	4,0	53,2	3,3	3,5	52,9
26	4	4,6	1,0	42,1	4,0	0,8	41,5	2,8	0,6	40,4
35	5	4,9	0,2	31,0	4,1	0,1	30,0	2,4	0,1	27,6
44	6	5,0	0,0	23,1	4,0	0,0	21,7	1,8	0,0	18,2
53	6			9,3				5,5		4,8
	7	4,9		17,4	3,7	0,0	15,7	1,2	0,0	11,5
62	7			7,0				3,6		2,7
	8	4,7		13,2	3,3	0,0	11,3	0,6	0,0	6,8
72	8			4,8				2,2		1,1
	9	4,9		9,2	3,2	0,0	7,2	0,2	0,0	3,1
82	8			1,5				6,5		0,1
	10	5,0		6,4	3,1	0,0	4,5	0	0,0	1,2
92	9			1,1				3,1		0,0
	11	5,0		4,5	2,9	0,0	< 4,5	0	0,0	0,4

vement pour $\frac{X}{n} = 0,5$ et $\frac{X}{n} = 0,25$. Ces risques d'erreur augmentent avec la taille (n) du troupeau. Leurs limites supérieures, lorsque n tend vers l'infini, sont déduites de la loi binominale. Ces résultats montrent l'influence relativement importante de la taille du troupeau. Il semble donc difficile de se contenter de

l'approximation obtenue avec la loi binomiale pour des troupeaux de faibles effectifs.

Avec un échantillon de 45 descendants, il est donc possible de distinguer les troupeaux dont la proportion d'erreurs est supérieure ou inférieure à 10 p. 100 avec des risques d'erreur inférieurs à 5 p. 100 (méthode n° 1 du tableau 3). Mais la valeur du risque d'erreur de seconde espèce étant calculée pour $\frac{X}{n} = 0,5$, on ne distingue vraiment par cette méthode que les troupeaux dont la proportion d'erreurs est inférieure à 10 p. 100 ou supérieur à 50 p. 100. Les autres troupeaux se classent indifféremment dans l'une ou l'autre catégorie. Pour un troupeau de 100 brebis, il suffirait de 32 descendants dans l'échantillon pour obtenir le même résultat.

Une autre méthode d'analyse (méthode n° 2 du tableau 3) peut permettre les mêmes résultats sur un échantillon de 18 filiations seulement. Selon que le nombre d'erreurs sera inférieur ou supérieur à 2, la proportion d'erreurs du troupeau sera considérée ou non comme satisfaisante. Si il est égal à 3, une analyse complémentaire devrait être faite sur le même échantillon à l'aide des groupes sanguins. Dans ce cas les deux risques d'erreur de première et de seconde espèce sont inférieurs à

TABLEAU 3

Méthode de vérification de la proportion d'erreurs d'un troupeau
Method of verification of proportion of errors in a flock

MÉTHODE N° 1 : Échantillon de 45 filiations étudiées avec les transferrines.

Si $y < 4$ l'hypothèse H_0 est retenue avec un risque d'erreur inférieur à 1,9 p. 100.
 Si $y > 4$ l'hypothèse H_1 est retenue avec un risque d'erreur inférieur à 4,7 p. 100.

MÉTHODE N° 2 : Échantillon de 18 filiations étudiées avec les transferrines et si nécessaire avec les groupes sanguins.

a/ Si $y < 2$, l'hypothèse H_0 est retenue avec un risque d'erreur inférieur à 5,6 p. 100.
 b/ Si $y > 2$, l'hypothèse H_1 est retenue avec un risque d'erreur inférieur à 4,2 p. 100.
 c/ Si $y = 2$, un contrôle complémentaire doit être réalisé sur le même échantillon avec les groupes sanguins.

Si $y < 3$, l'hypothèse H_0 est retenue avec un risque d'erreur inférieur à 4,3 p. 100.
 Si $y > 3$, l'hypothèse H_1 est retenue avec un risque d'erreur inférieur à 4,5 p. 100.

MÉTHODE N° 3 : Échantillon de 92 filiations étudiées avec les transferrines et si nécessaire avec les groupes sanguins.

a/ Si $y < 6$, l'hypothèse est H_0 retenue avec un risque d'erreur inférieur à 5,3 p. 100
 b/ Si $y > 7$, l'hypothèse H_2 est retenue avec un risque d'erreur inférieur à 5,0 p. 100
 c/ Si $y = 6$ ou 7, un contrôle complémentaire doit être réalisé sur le même échantillon avec les groupes sanguins.

Si $y < 11$, l'hypothèse H_0 est retenue avec un risque d'erreur inférieur à 4,5 p. 100
 Si $y > 11$, l'hypothèse H_2 est retenue avec un risque d'erreur inférieur à 5,0 p. 100

y = nombre d'erreurs observées dans l'échantillon de filiations étudiées.

H_0 = la proportion d'erreur du troupeau est inférieure à 10 p. 100.

H_1 = la proportion d'erreur du troupeau est supérieure à 50 p. 100.

H_2 = la proportion d'erreur du troupeau est supérieure à 25 p. 100.

5,6 p. 100. Cette seconde méthode sera plus avantageuse que la première si le contrôle complémentaire n'est pas nécessaire dans plus de 45 p. 100 des cas. Pour distinguer les troupeaux présentant une proportion d'erreurs inférieure à 10 p. 100 ou supérieure à 25 p. 100, un échantillon de 92 filiations se révèle nécessaire avec un contrôle complémentaire à l'aide des groupes sanguins si le nombre d'erreurs révélé par les transferrines est égal à 6 ou 7 (méthode n° 3 du tableau 3). Les risques d'erreur donnés dans le tableau 3 ont été calculés pour un effectif de troupeau infini. Ce sont donc des risques d'erreur maximum.

Conclusion

Les marqueurs biochimiques sanguins permettent donc de reconnaître les troupeaux qui ne réalisent pas un contrôle des luttés suffisant pour assurer une bonne efficacité du contrôle de descendance. L'utilisation des transferrines est dans le cas de la race Ile de France la méthode la plus efficace pour évaluer la proportion d'erreurs de filiation dans un troupeau. Des échantillons de 45 ou de 18 descendants selon les marqueurs analysés, permettent de distinguer les troupeaux dont la proportion d'erreur est inférieure à 10 p. 100, de ceux dont la proportion est supérieure à 50 p. 100. Cette méthode de contrôle de la proportion d'erreurs de filiation pourra ultérieurement être améliorée par l'utilisation d'un plan d'échantillonnage progressif.

Reçu pour publication en juillet 1977.

Remerciements

L'auteur remercie MASSON et MILLIER, station de Biométrie I.N.R.A., Nancy, pour leurs critiques et remarques.

Summary

Estimation of the proportion of errors of relationship in a flock based biochemical blood markers. (Theoretical study).

With the biochemical blood markers, the proportion of errors of relationship can be estimated. The statistical study of the estimation allow to choose the most efficient marker (minimum variance) and then to calculate the errors of the test of hypothesis. This give a mean, with a known amount of error, to distinguish the flocks giving relationship enough reliable to make the progeny test efficient.

Références bibliographiques

- JACQUARD A., SALMON O., 1971. Sur le diagnostic de paternité. *Population*, **26**, 677-690.
 RENDEL J., 1958. Studies of cattle Blood Groupes. II. Parentage tests. *Act. Agric. Scand.*, **8**, 131-161.
 BONAÏTI B., 1974. Influence du taux d'erreur dans les filiations paternelles sur la valeur de la sélection sur descendance chez les ovins. *1^{res} Journées de la Recherche ovine et caprine*. II : 166-172. ITOVIC, Paris.