

ROGER HURON

**Méthode générale d'estimation de la fréquence des gènes.
Application aux groupes sanguins**

Annales de la faculté des sciences de Toulouse 4^e série, tome 22 (1958), p. 159-173

http://www.numdam.org/item?id=AFST_1958_4_22__159_0

© Université Paul Sabatier, 1958, tous droits réservés.

L'accès aux archives de la revue « Annales de la faculté des sciences de Toulouse » (<http://picard.ups-tlse.fr/~annales/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

Méthode générale d'estimation de la fréquence des gènes Application aux groupes sanguins

par **Roger HURON**

INTRODUCTION

L'estimation à partir des observations de la fréquence des gènes conditionnant un caractère déterminé, particulièrement les groupes sanguins, peut se faire tout au moins sur le plan théorique à l'aide de la méthode du maximum de vraisemblance.

Sur le plan pratique et si l'on envisage un groupe un peu complexe, tel le groupe Rhésus, le système non linéaire auquel la méthode conduit est si compliqué qu'il faut introduire des hypothèses particulières et user d'artifices pour en obtenir des solutions approchées.

La méthode que nous indiquons dans ce travail donne par des calculs simples, systématiques, facilement programmable si l'on veut utiliser un ordinateur électronique, les estimations qui rendent maximum la vraisemblance. Elle s'applique à tous les cas sans l'introduction d'hypothèses arbitraires. Elle permet d'utiliser au mieux des résultats d'enquêtes perturbées pour des raisons techniques (épuisement ou inefficacité d'un anti-sérum par exemple).

Dans une première partie nous exposons son fondement théorique⁽¹⁾. Dans la seconde partie nous montrons son adaptation à des cas particuliers. Les exemples que nous avons choisis proviennent de groupages effectués au Sahara par M. le Professeur agrégé J. RUFFIÉ et le Docteur J. Ducos de la Faculté de Médecine et du Centre de Transfusion Sanguine de Toulouse (Mission de 1959).

(1) Cette partie constitue le développement de deux notes publiées aux Comptes Rendus de l'Académie des Sciences, t. 250, pp. 1603-1604 et t. 250, pp. 1790-1791.

PREMIÈRE PARTIE

ÉTUDE THÉORIQUE

1. — Préliminaires : Étude d'un problème d'estimation de paramètres.

Considérons une population parente P formée de plusieurs strates : $S_1, S_2, \dots, S_n$. Soit p_i la probabilité qu'un individu pris au hasard dans P d'appartenir au strate S_i . Nous supposons que :

$$(1) \quad p_i = f_i(t_1, t_2, \dots, t_l) \quad (i = 1, 2, \dots, n)$$

où les f_i sont des fonctions connues des paramètres inconnus : $t_1, t_2, \dots, t_l$. Nous supposons que l'échantillon E extrait au hasard et non exhaustivement de P comprend d'une part des individus parfaitement identifiés en nombre : $a_1, a_2, \dots, a_n$, d'autre part un groupe de b individus pour lesquels on sait seulement qu'ils appartiennent à l'une ou l'autre des classes : $S_1, S_2, \dots, S_k$ ($k < n$).

Considérons l'échantillon E_1 formé par les individus parfaitement identifiés, la probabilité d'obtenir un tel échantillon est proportionnelle à :

$$(2) \quad \varpi = p_1^{a_1} p_2^{a_2} \dots p_n^{a_n}$$

d'où : $\log \varpi = \sum_i a_i \log f_i(t_1, t_2, \dots, t_l)$
et

$$\frac{1}{\varpi} \frac{\partial \varpi}{\partial t_\lambda} = \sum_i \frac{a_i}{f_i} \frac{\partial f_i}{\partial t_\lambda} \quad (\lambda = 1, 2, \dots, l)$$

Il en résulte que les estimations : $t_{10}, t_{20}, \dots, t_{l0}$ calculées à partir de E_1 et rendant maximum la vraisemblance sont solution du système

$$(3) \quad \sum_i \frac{a_i}{f_i} \frac{\partial f_i}{\partial t_\lambda} = 0 \quad (\lambda = 1, 2, \dots, l)$$

A l'aide de ces estimations des t_λ il est possible de calculer des estimations des p_i : $f_{i0} = f_i(t_{10}, t_{20}, \dots, t_{l0})$ puis d'estimer le nombre des individus du groupe E_2 restant, qui auraient dû être classés en : $S_1, S_2, \dots, S_k$ respectivement.

On a ainsi :

$$(\Delta a_j)_0 = b \frac{f_{j0}}{\sum_j f_{j0}} \quad (j = 1, 2, \dots, k)$$

En remplaçant les a_i par leurs valeurs corrigées $a_i + (\Delta a_i)_0$ s'il y a lieu, on a pour calculer une nouvelle estimation des t_i un système analogue à (3) :

$$(4) \quad \sum_i \frac{a_i + (\Delta a_i)_0}{f_i} \frac{\partial f_i}{\partial t_\lambda} = 0$$

où certains $(\Delta a_i)_0$ sont nuls.

Le procédé d'itération se poursuit à partir des nouvelles valeurs : $t_{11}, t_{21}; \dots t_{l1}$ fournies par le système (4).

2. — Propriété des estimations.

Si le procédé converge et si θ_λ est la limite t_λ on aura pour valeur limite des Δa_j :

$$(1) \quad \Delta a_j = b \frac{f_j (\theta_1 \theta_2 \dots \theta_l)}{\sum_j f_j (\theta_1 \theta_2 \dots \theta_l)}$$

et les θ_λ seront solutions du système suivant déduit de la relation (4) du numéro précédent :

$$(2) \quad \sum_i \frac{a_i}{f_i} \frac{\partial f_i}{\partial t_\lambda} + \frac{b}{\sum_j f_j} \sum_j \frac{\partial f_j}{\partial t_\lambda} = 0 \quad (\lambda = 1, 2 \dots l)$$

Ceci posé nous pouvons considérer l'échantillon E comme formé de deux échantillons indépendants, l'un E'_1 dont les valeurs observées seraient :

$$a_1 a_2 \dots a_k, a'_{k+1}, a'_{k+2}, \dots a'_n$$

et dont l'autre E'_2 comprendrait les b individus appartenant à l'union des classes S_j auxquels seraient adjoints des individus en nombre :

$$a''_{k+1}, a''_{k+2}, \dots a''_n$$

des classes : $S_{k+1}, S_{k+2}, \dots S_n$ tels que

$$a'_{k+1} + a''_{k+2} = a_{k+1}; \dots; a'_n + a''_n = a_n$$

La probabilité du couple $E'_1 E'_2$ est alors proportionnelle à :

$$\begin{aligned} \varpi' &= f_1^{a_1} f_2^{a_2} \dots f_k^{a_k} f_{k+1}^{a'_{k+1}} \dots f_n^{a'_n} (f_1 + f_2 + \dots + f_k)^b f_{k+1}^{a''_{k+1}} \dots f_n^{a''_n} \\ &= f_1^{a_1} f_2^{a_2} \dots f_n^{a_n} (f_1 + f_2 + \dots + f_k)^b \end{aligned}$$

On a :

$$\log \varpi' = \sum_i a_i \log f_i + b \log (f_1 + f_2 + \dots + f_k)$$

d'où en dérivant par rapport à t_λ :

$$\frac{1}{\varpi'} \frac{\partial \varpi'}{\partial t_\lambda} = \sum_i \frac{a_i}{f_i} \frac{\partial f_i}{\partial t_\lambda} + \frac{b}{\sum_j f_j} \sum_j \frac{\partial f_j}{\partial t_\lambda} \quad (\lambda = 1, 2 \dots l)$$

On voit que les valeurs des t_λ qui rendent cette probabilité maximum sont solution du système (2).

En définitive, nous pouvons dire que les estimations des t_i obtenues par le procédé d'itération décrit ci-dessus rendent extremum la quantité :

$$(3) \quad f_1^{a_1} f_2^{a_2} \dots f_n^{a_n} (f_1 + f_2 + \dots + f_k)^b$$

3. — Généralisation.

Le groupe E_2 peut se fragmenter en :

un groupe E_{21} d'individus appartenant à l'une ou l'autre des classes :

$S_1, S_2, \dots, S_{k_1}$ et en nombre b_1 ;

un groupe E_{22} d'individus appartenant à l'une ou l'autre des classes :

$S_{k_1+1}, \dots, S_k$ et en nombre b_2 .

Les corrections à apporter aux a_i seront alors et à l'itération d'ordre μ du type :

$$(\Delta a_{j_1})_\mu = b_1 \frac{f_{j_1^\mu}}{\sum_{j_1} f_{j_1^\mu}} (j_1 = 1, 2, \dots, k_1)$$

$$(\Delta a_{j_2})_\mu = b_2 \frac{f_{j_2^\mu}}{\sum_{j_2} f_{j_2^\mu}} (j_2 = k_1 + 1, \dots, k)$$

L'équation 2 du n° 2 s'écrira maintenant :

$$(1) \quad \sum_i \frac{a_i}{f_i} \frac{\partial f_i}{\partial t_k} + \frac{b_1}{\sum_{j_1} f_{j_1}} \sum_{j_1} \frac{\partial f_{j_1}}{\partial t_k} + \frac{b_2}{\sum_{j_2} f_{j_2}} \sum_{j_2} \frac{\partial f_{j_2}}{\partial t_k} = 0$$

et ses solutions rendront extrême la quantité :

$$(2) \quad f_1^{a_1} f_2^{a_2} \dots f_n^{a_n} (f_1 + f_2 + \dots + f_{k_1})^{b_1} (f_{k_1+1} + \dots + f_k)^{b_2}$$

La généralisation se poursuit facilement

4. — Méthode générale d'estimation des fréquences géniques.

Nous exposons la méthode dans le cas assez simple du groupe sanguin A B O, c'est-à-dire du cas où il y a trois gènes allèles : A, B, R de probabilités respectives p_1, p_2, p_3 et où les phénotypes observés sont :

(A) = A A ou A R; (B) = B B ou B R; (AB) = A B; (O) = R R

Si nous étions capables d'identifier tous les génotypes, nous aurions pour les fréquences observées un tableau tel que le suivant :

Échantillon E_1						
Fréquences :	A A	A R	B B	B R	A B	R R
Génotypes :	a_{11}	a_{13}	a_{22}	a_{23}	a_{12}	a_{33}
						n_1

La probabilité de E_1 est proportionnelle à :

$$(1) \quad \alpha = p_1^{a_{11}} p_1^{a_{22}} p_3^{a_{33}} (2p_1 p_2)^{a_{12}} (2p_1 p_2)^{a_{13}} (2p_2 p_3)^{a_{23}}$$

Pour obtenir les estimations des p_i par la méthode du maximum de vraisemblance, il faut chercher les extremum de

$$\left\{ \begin{array}{l} \Phi = 2a_{11} \log p_1 + 2a_{22} \log p_2 + 2a_{33} \log p_3 \\ \quad + a_{12} (\log p_1 + \log p_2) + a_{13} (\log p_1 + \log p_3) \\ \quad + a_{23} (\log p_2 + \log p_3) \end{array} \right.$$

sachant que l'on a

$$p_1 + p_2 + p_3 = 1$$

La méthode des multiplicateurs de Lagrange conduit à annuler les dérivées

partielles par rapport aux p_i de

$$\Phi - \nu (p_1 + p_2 + p_3 - 1)$$

où ν est un paramètre. Désignons par x_1, x_2, x_3 les estimations respectives de : p_1, p_2, p_3 on a :

$$x_1 = \frac{1}{\nu} (2 a_{11} + a_{12} + a_{13})$$

et des formules analogues pour x_2 et x_3 . La condition : $x_1 + x_2 + x_3 = 1$ donne alors $\nu = 2n_1$ d'où :

$$(2) \quad \left\{ \begin{array}{l} x_1 = \frac{1}{2 n_1} (2 a_{11} + a_{12} + a_{13}) \\ x_2 = \frac{1}{2 n_1} (2 a_{22} + a_{12} + a_{23}) \\ x_3 = \frac{1}{2 n_1} (2 a_{33} + a_{13} + a_{23}) \end{array} \right.$$

Soit maintenant un échantillon réel où les individus ne sont connus que par leurs phénotypes; le tableau des fréquences observées est :

Échantillon E_2				
Phénotypes :	(A)	(B)	(AB)	(O)
Fréquences :	a	b	c	d
				Total
				n_2

La probabilité d'un tel échantillon est proportionnelle à :

$$(3) \quad \beta = (p_1^2 + 2p_1 p_3)^a (p_2^2 + 2p_2 p_3)^b (2p_1 p_2)^c p_3^{2d}$$

D'après ce qui a été dit aux numéros précédents, l'estimation des p_i en utilisant E_1 et E_2 , pourrait se faire par un procédé d'approximations successives où les valeurs initiales seraient :

$$(4) \quad \left\{ \begin{array}{l} x_{10} = \frac{1}{2 n_1} (2 a_{11} + a_{12} + a_{13}) ; \\ x_{20} = \frac{1}{2 n_1} (2 a_{22} + a_{12} + a_{23}) ; \\ x_{30} = \frac{1}{2 n_1} (2 a_{33} + a_{13} + a_{23}) ; \end{array} \right.$$

A partir de ces valeurs on calculerait

$$(5) \quad \left\{ \begin{array}{l} (\Delta a_{11})_0 = \frac{a x_{10}^2}{x_{10}^2 + 2 x_{10} x_{30}} \\ (\Delta a_{13})_0 = \frac{2 a x_{10} x_{30}}{x_{10}^2 + 2 x_{10} x_{30}} \end{array} \right.$$

et de la même façon : $(\Delta a_{22})_0; (\Delta a_{23})_0$. Il faut ensuite remplacer dans (1) :

$$\begin{array}{l} a_{11} \text{ par } a_{11} + (\Delta a_{11})_0 \\ a_{22} \text{ par } a_{22} + (\Delta a_{22})_0 \text{ etc...} \end{array}$$

La somme totale des nouvelles fréquences est $n_1 + n_2$ et les formules (4) donnent :

$$(6) \quad x_{11} = \frac{1}{2(n_1 + n_2)} \left\{ 2[a_{11} + (\Delta a_{11})_0] + [a_{12} + (\Delta a_{12})_0] + [a_{13} + (\Delta a_{13})_0] \right\},$$

et des formules analogues pour x_{21} et x_{31} .

Ces nouvelles valeurs permettent à leur tour de calculer :

$$(7) \quad (\Delta a_{11})_1 = -\frac{a x_{11}^2}{x_{11}^3 + 2x_{11} x_{31}}; \text{ etc.}$$

Le procédé d'itération se poursuit toujours de la même façon, les valeurs limites obtenues pour : x_1, x_2, x_3 rendant extremum la quantité $\alpha\beta$.

5. — Cas particulier fondamental.

Si nous prenons les a_{ij} nuls, alors $n_1 = 0$. Nous voyons, d'après (4) que les x_{i0} sont indéterminés et que $\alpha\beta$ se réduit à β , il en résulte que les valeurs limites des x_i obtenues par le procédé d'itération sont celles qui rendent extremum β , c'est-à-dire que ce sont celles que donnerait la méthode du maximum de vraisemblance.

Dans ce cas particulier on peut simplifier les écritures en supprimant les symboles Δ et en posant $n_2 = n$.

DEUXIÈME PARTIE

EXEMPLES D'APPLICATION

A. — Groupe M N

1. — Étude d'un cas particulier.

Le cas usuel où les génotypes sont tous connus n'a pas d'intérêt ici puisque l'équation obtenue par la méthode du maximum de vraisemblance donne des estimations d'un type déjà obtenu au n° 4 de la première partie (formules 2). Nous étudierons le cas plus particulier où pour une raison ou une autre certains individus ne sont connus que par leurs phénotypes. C'est le cas par exemple où le sérum anti-N est épuisé. Les données sont alors :

Phénotypes :	M	MN	N	M+	Total
Fréquences					
observées :	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	T

Il suffit d'appliquer la méthode du n° 1 de la première partie. Soient p et $q = 1 - p$ les probabilités des gènes M et N respectivement, x et y leurs estimations. Faisant abstraction du lot M+ les valeurs qui rendent maximum

$$p^{2a} (2pq)^b q^{2c}$$

sont :

$$x_0 = \frac{2a + b}{2(a + b + c)}; \quad y_0 = \frac{2c + b}{2(a + b + c)}; \quad x_0 + y_0 = 1$$

A l'aide de ces estimations de p et q nous calculons les corrections à apporter à a et b

$$(1) \quad \begin{cases} (\Delta a)_0 = \frac{dp_0^2}{p_0^2 + 2p_0 q_0} = \frac{dp_0}{p_0 + 2q_0} = \frac{d(1 - q_0)}{1 + q_0} \\ (\Delta b)_0 = \frac{2 dp_0 q_0}{p_0^2 + 2p_0 q_0} = \frac{2 dq_0}{p_0 + dq_0} = \frac{2 dq_0}{1 + q_0} \end{cases}$$

Nous en déduisons le nouveau tableau :

Phénotypes :	M	MN	N	Total
Fréquences :	$a + (\Delta a)_0$	$b + (\Delta b)_0$	c	T

et les nouvelles estimations de p et q

$$(2) \quad \begin{cases} x_1 = \frac{2[a + (\Delta a)_0] + b}{2T} \\ y_1 = \frac{2c + b + [\Delta b]_0}{2T} \end{cases}$$

Le processus d'itération se poursuit identiquement jusqu'à la stabilisation des valeurs de x et de y .

A l'itération d'ordre μ et d'après (1) et (2) on aura :

$$(3) \quad y_{\mu} = \frac{2c + b}{2T} + \frac{d}{T} \cdot \frac{y_{\mu-1}}{1 + y_{\mu-1}}$$

A la limite :

$$y_{\mu} = y_{\mu-1} = y \quad \text{d'où :}$$

$$y = \frac{2c + b}{2T} + \frac{d}{T} \cdot \frac{y}{1 + y}$$

soit en réduisant :

$$(4) \quad 2Ty^2 + (2a + b)y - (2c + b) = 0$$

Une seule racine est acceptable, celle comprise entre 0 et 1

$$(5) \quad y = \frac{1}{4T} \left[-(2a + b) + \sqrt{(2a + b)^2 + 8T(2c + b)} \right]$$

On pourra remarquer que conformément à la théorie générale la recherche du maximum de

$$p^{2a} (2pq)^b q^{2c} (p^2 + 2pq)^d$$

avec $p + q = 1$, conduit justement à (4).

2. — Exemple :

Données : Tribu saharienne Tédas

Phénotypes :	M	MN	N	M+	Total
Fréquences :	42	51	54	68	215

L'équation (4) s'écrit ici :

$$430y^2 + 135y - 159 = 0$$

D'où par (5) :

$$y = 0,4711 \quad x = 1 - y = 0,5219$$

B. — Groupe A B O

I. — CAS DE TROIS GENES ALLELES : A, B, R

1. — Données.

Phénotypes :	(A)	(B)	(AB)	(O)	Total
Fréquences :	b_1	b_2	b_3	b_4	n

2. — Formules générales.

Soient p , q et r les probabilités des gènes A, B, et R respectivement. On sait que :

$$Pr(A) = p^2 + 2pr; \quad Pr(B) = q^2 + 2qr; \quad Pr(AB) = 2pq; \quad Pr(O) = r^2$$

Désignons par : $x_1 \ x_2 \ x_3$ les estimations de $p \ q \ r$.

Partant de valeurs arbitraires dont la somme est l'unité posons :

$$\begin{aligned} B_1 &= x_1^2 + 2x_1 x_3 & ; & & c_1 &= b_1/B_1 \\ B_2 &= x_2^2 + 2x_2 x_3 & ; & & c_2 &= b_2/B_2 \end{aligned}$$

L'application des formules générales de la première partie donne alors :

$$\begin{aligned} \text{Nouvelle valeur de } x_1 &= \frac{b_1 + b_3}{2n} + \frac{c_1 x_1}{2n} \\ \text{Nouvelle valeur de } x_2 &= \frac{b_2 + b_3}{2n} + \frac{c_2 x_2}{2n} \\ \text{Nouvelle valeur de } x_3 &= \frac{2b_3}{2n} + \frac{c_1 + (2x_1 x_3) + c_2 (2x_2 x_3)}{2n} \end{aligned}$$

Après avoir vérifié que la somme des nouvelles valeurs des x_i est bien égale à l'unité, on recommence l'ensemble du calcul avec ces nouvelles valeurs et ceci jusqu'à ce que l'on obtienne la stabilisation des résultats.

3. — Exemple.

Données : Tribu saharienne Tédas

(A)	(B)	(AB)	(O)	Total
47	46	4	125	222

Valeurs arbitraires de départ :

$$x_1 = 0,1 \quad ; \quad x_2 = 0,1 \quad ; \quad x_3 = 0,8$$

Valeurs à la fin des itérations successives :

Ordre des itérations	x_1	x_2	x_3
1	0,12109	0,11870	0,76020
2	0,12267	0,12011	0,75721
3	0,12279	0,12022	0,75697
4	0,12280	0,12023	0,75696

On pourra donc admettre comme estimations de p , q et r les valeurs des x_i obtenues à l'itération d'ordre 4.

II. — CAS DE CINQ ALLELES : A_i , A_1 , A_2 , B, R

1. — Données.

Phénotypes :	(A _i)	(A ₁)	(A ₂)	(B)	(A _i B)	(A ₁ B)	(A ₂ B)	(O)	Total
Fréquences									
observées :	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	n

2. — Formules générales.

Soient : p_1, p_2, p_3, q et r les probabilités respectives des gènes A_i, A_1, A_2, B et R et x_1, x_2, x_3, x_4, x_5 leurs estimations. On a maintenant :

<i>Phénotypes</i>	<i>Probabilités</i>
(A_i)	$p_1^2 + 2 p_1 p_2 + 2 p_1 p_3 + 2 p_1 r$
(A_1)	$p_2^2 + 2 p_2 p_3 + 2 p_2 r$
(A_2)	$p_3^2 + 2 p_3 r$
(B)	$q^2 + 2 qr$
($A_i B$)	$2 p_1 q$
($A_1 B$)	$2 p_2 q$
($A_2 B$)	$2 p_3 q$
(0)	r^2

Soient x_1, x_2, x_3, x_4, x_5 des valeurs de départ arbitraires telles que leur somme soit égale à l'unité, posons :

$$\begin{aligned} B_1 &= x_1^2 + 2 x_1 x_2 + 2 x_1 x_3 + 2 x_1 x_5 ; & c_1 &= b_1 / B_1 \\ B_2 &= x_2^2 + 2 x_2 x_3 + 2 x_2 x_5 ; & c_2 &= b_2 / B_2 \\ B_3 &= x_3^2 + 2 x_3 x_5 ; & c_3 &= b_3 / B_3 \\ B_4 &= x_4^2 + 2 x_4 x_5 ; & c_4 &= b_4 / B_4 \end{aligned}$$

On a maintenant :
Nouvelle valeur de x_1 :

$$\frac{b_1 + b_5}{2 n} + \frac{c_1 x_1^2}{2 n}$$

Nouvelle valeur de x_2 :

$$\frac{b_1 + b_6}{2 n} + \frac{c_1 (2 x_1 x_2) + c_2 x_2^2}{2 n}$$

Nouvelle valeur de x_3 :

$$\frac{b_3 + b_7}{2 n} + \frac{c_1 (2 x_1 x_3) + c_2 (2 x_2 x_3) + c_3 x_3^2}{2 n}$$

Nouvelles valeurs de x_4 :

$$\frac{b_4 + b_8 + b_6 + b_7}{2 n} + \frac{c_1 x_1^2}{2 n}$$

Nouvelle valeur de x_5 :

$$\frac{2 b_8}{2 n} + \frac{c_1 (2 x_1 x_5) + c_2 (2 x_2 x_5) + c_3 (2 x_3 x_5) + c_4 (2 x_4 x_5)}{2 n}$$

3. — Exemple :

<i>Données :</i>	Tribu des Dazas	(A_i)	(A_1)	(A_2)	(B)	($A_i B$)	($A_1 B$)	($A_2 B$)	(0)	Total
		89	60	19	160	0	9	2	261	600

Ordre des itérations	x_1	x_2	x_3	x_4	x_5
Départ	0,20	0,20	0,20	0,20	0,20
1 ^{re}	0,0848	0,0887	0,0640	0,1869	0,5756
6 ^e	0,07775	0,06581	0,02111	0,15625	0,15625
7 ^e	0,07775	0,06581	0,02113	0,67908	0,67906

C. — SYSTÈME RHESUS

1. — Notations.

Nous nous plaçons dans le cas où l'on a utilisé les quatre sérums :
anti C; anti c; anti D; anti E

Le tableau suivant rappelle les diverses notations selon la théorie génétique adoptée :

	Fisher	Wiener	Probabilités
	CDE	R_z	p_1
	CDe	R_1	p_2
	CdE	r''	p_3
(1)	Cde	r'	p_4
	cDE	R_2	p_5
	cDe	R_0	p_6
	cdE	r''	p_7
	cde	r	p_8

Les douze phénotypes sérologiques et leurs probabilités respectives sont alors :

Phénotypes	CcDE	Fréquences observées	Probabilités
P_1	—+—	b_1	p_8^2
P_2	—++	b_2	$p_6^2 + 2p_6p_8$
P_3	—++	b_3	$p_7^2 + 2p_7p_8$
P_4	+++	b_4	$p_2^2 + 2(p_5p_6 + p_5p_7 + p_5p_8 + p_6p_7)$
P_5	++—	b_5	$2p_4p_8$
P_6	++—	b_6	$2(p_2p_6 + p_2p_8 + p_4p_6)$
P_7	++—	b_7	$2(p_3p_7 + p_3p_8 + p_4p_7)$
P_8	+++	b_8	$2(p_1p_5 + p_1p_6 + p_1p_7 + p_1p_8 + p_2p_5 + p_2p_7 + p_3p_5 + p_3p_6 + p_4p_5)$
P_9	+—	b_9	p_4^2
P_{10}	+—	b_{10}	$p_2^2 + 2p_2p_4$
P_{11}	+—	b_{11}	$p_2^2 + 2(p_1p_2 + p_1p_3 + p_1p_4 + p_2p_3)$
P_{12}	+—	b_{12}	$p_3^2 + 2p_3p_4$

2. — Formules générales.

Partant de *valeurs arbitraires*

$x_1 x_2 x_3 \dots x_8$ avec $\sum x_i = 1$, on pose maintenant ⁽¹⁾ :

$$B_2 = x_6^2 + 2 x_6 x_8 \quad ; \quad B_3 = x_7^2 + 2 x_7 x_8$$

$$B_4 = x_5^2 + 2 x_5 x_6 + 2 x_5 x_7 + 2 x_5 x_8 + 2 x_6 x_7$$

$$B_6 = 2 x_2 x_6 + 2 x_2 x_8 + 2 x_4 x_6$$

$$B_7 = 2 x_3 x_7 + 2 x_3 x_8 + 2 x_4 x_7$$

$$B_8 = 2 x_1 x_5 + 2 x_1 x_6 + 2 x_1 x_7$$

$$+ 2 x_1 x_8 + 2 x_2 x_5 + 2 x_2 x_7 + 2 x_3 x_5 + 2 x_3 x_6 + 2 x_4 x_5$$

$$B_{10} = x_2^2 + 2 x_2 x_4$$

$$B_{11} = x_1^2 + 2 x_1 x_2 + 2 x_1 x_3 + 2 x_1 x_4 + 2 x_2 x_3$$

$$B_{12} = x_3^2 + 2 x_3 x_4$$

et

$$c_2 = \frac{b_2}{B_2} \quad ; \quad c_3 = \frac{b_3}{B_3} \quad ; \quad c_4 = \frac{b_4}{B_4}$$

$$c_6 = \frac{b_6}{B_6} \quad ; \quad c_7 = \frac{b_7}{B_7} \quad ; \quad c_8 = \frac{b_8}{B_8}$$

$$c_{10} = \frac{b_{10}}{B_{10}} \quad ; \quad c_{11} = \frac{b_{11}}{B_{11}} \quad ; \quad c_{12} = \frac{b_{12}}{B_{12}}$$

On en déduit les nouvelles valeurs des x_i :

Nouvelles valeurs de x_1

$$\frac{1}{2n} \left[c_{11} (2 x_1^2 + 2 x_1 x_2 + 2 x_1 x_3 + 2 x_1 x_4) \right. \\ \left. + c_8 (2 x_1 x_5 + 2 x_1 x_6 + 2 x_1 x_7 + 2 x_1 x_8) \right]$$

Nouvelles valeurs de x_2

$$\frac{1}{2n} \left[c_6 (2 x_2 x_6 + 2 x_2 x_8) + c_8 (2 x_2 x_5 + 2 x_2 x_7) \right. \\ \left. + c_{10} (2 x_2^2 + 2 x_2 x_4) + c_{11} (2 x_1 x_2 + 2 x_2 x_3) \right]$$

Nouvelles valeurs de x_3

$$\frac{1}{2n} \left[c_7 (2 x_3 x_7 + 2 x_3 x_8) + c_8 (2 x_3 x_5 + 2 x_3 x_6) \right. \\ \left. + c_{11} (2 x_1 x_3 + 2 x_2 x_3) + c_{12} (2 x_3^2 + 2 x_3 x_4) \right]$$

Nouvelles valeurs de x_4

$$\frac{1}{2n} \left[b_5 + b_9 + 2 x_4 x_6 c_6 + 2 x_4 x_7 c_7 + 2 x_4 x_5 c_8 \right. \\ \left. + 2 x_2 x_4 c_{10} + 2 x_1 x_4 c_{11} + 2 x_3 x_4 c_{12} \right]$$

(1) Il est inutile de calculer : B_1 , B_5 et B_9 ainsi que : c_1 , c_5 et c_9 .

Nouvelles valeurs de x_5

$$\frac{1}{2n} \left[c_4 (2x_5^2 + 2x_5x_6 + 2x_5x_7 + 2x_5x_8) \right. \\ \left. + c_8 (2x_1x_5 + 2x_2x_5 + 2x_3x_5 + 2x_4x_5) \right]$$

Nouvelles valeurs de x_6

$$\frac{1}{2n} \left[c_2 (2x_6^2 + 2x_6x_8) + c_4 (2x_5x_6 + 2x_6x_7) \right. \\ \left. + c_6 (2x_2x_6 + 2x_4x_6) + c_8 (2x_1x_6 + 2x_3x_6) \right]$$

Nouvelles valeurs de x_7

$$\frac{1}{2n} \left[c_3 (2x_7^2 + 2x_7x_8) + c_4 (2x_5x_7 + 2x_6x_7) \right. \\ \left. + c_7 (2x_3x_7 + 2x_4x_7) + c_8 (2x_1x_7 + 2x_2x_7) \right]$$

Nouvelles valeurs de x_8

$$\frac{1}{2n} \left[2b_1 + b_5 + 2x_6x_8c_2 + 2x_7x_8c_3 + 2x_5x_8c_4 \right. \\ \left. + 2x_2x_8c_6 + 2x_3x_8c_7 + 2x_1x_8c_8 \right]$$

Vérification et poursuite du calcul

On doit s'assurer que l'on a bien pour les nouvelles valeurs

$$x_1 + x_2 + \dots + x_8 = 1$$

On reprend ensuite le calcul avec ces valeurs des x_i . On fait autant d'itérations qu'il est nécessaire pour obtenir des valeurs des x_i stables.

3. — Premier exemple :

Tirailleurs

Phénotypes	: P ₁	P ₂	P ₄	P ₆	P ₈	Total
Fréquences	: 10	119	29	13	3	174

Ici : $c_3 = c_5 = c_7 = c_9 = c_{10} = c_{11} = c_{12} = 0$; les formules générales sont très simplifiées. Nous avons obtenu :

	Valeurs de départ	1 ^{re} itération	18 ^e itération	19 ^e itération
x_1	0,1	0,0040	0,00462	0,00462
x_2	0,1	0,0227	0,04133	0,04133
x_3	0,1	0,0027	0,00000	0,00000
x_4	0,1	0,0167	0,00003	0,00003
x_5	0,1	0,0615	0,09193	0,09194
x_6	0,3	0,6419	0,62214	0,62212
x_7	0,1	0,0405	0,00007	0,00006
x_8	0,1	0,2100	0,23988	0,23990

4. — Deuxième exemple :

Tribu saharienne des Kamadjas

Ici des accidents de groupage ont conduit aux résultats suivants :

P ₁	P ₂	P ₃	P ₄	P ₆	P ₈	P ₁₀	(P ₆ ou P ₁₀)	Total
20	256	1	42	37	4	5	17	382

Partant de valeurs initiales arbitraires pour les x_i , nous calculons les *corrections* à apporter à b_6 et b_{10}

$$\Delta b_6 = 17 \frac{B_6}{B_6 + B_{10}} \quad \Delta b_{10} = 17 \frac{B_{10}}{B_6 + B_{10}}$$

On fait alors la première itération comme précédemment mais en remplaçant partout :

$$b_6 \text{ par } b'_6 = b_6 + \Delta b_6 \\ b_{10} \text{ par } b'_{10} = b_{10} + \Delta b_{10}$$

et en tenant compte que : $b_5 = b_7 = b_9 = b_{11} = b_{12} = 0$. A chaque itération on recalcule les corrections à apporter aux *valeurs observées* 37 et 5 de b_6 et b_{10} respectivement. Nous avons obtenu :

	Valeurs de départ	1 ^{re} itération	27 ^e itération	28 ^e itération
x_1	0,1	0,0024	0,00000	0,00000
x_2	0,1	0,0529	0,09005	0,09007
x_3	0,1	0,0016	0,00000	0,00000
x_4	0,1	0,0360	0,00007	0,00006
x_5	0,1	0,0404	0,05790	0,05790
x_6	0,3	0,6346	0,61860	0,61816
x_7	0,1	0,0284	0,00566	0,00566
x_8	0,1	0,2037	0,22771	0,22773