

CHAPITRE XII

MIGRATIONS

par

DANIEL COURGEAU

Les populations humaines tout à fait isolées sont très rares. Nous avons cité au cours du chapitre 8 le cas de groupes humains, qui se sont développés sans contact avec l'extérieur (Ilot de Pitcairn, peuplade Alakaluf de la Terre de Feu). Mais, en dehors de ces cas extrêmes, les hommes, comme d'ailleurs la plupart des animaux, communiquent entre eux par des migrations. Celles-ci sont fonction d'un grand nombre de facteurs géographiques, sociologiques, économiques, qui déterminent leur répartition.

Si une population reçoit un certain nombre de migrants provenant de groupes humains étrangers, sa structure génétique peut se trouver modifiée. Il suffit pour cela que les fréquences de divers allèles présents en un locus soient différentes dans l'ensemble des immigrants et dans la population réceptrice. La migration est ainsi une cause d'évolution du patrimoine génétique.

C'est cette évolution que nous allons chercher à déterminer au cours de ce chapitre.

En premier lieu, supposant toutes les autres hypothèses de la panmixie parfaite vérifiées, nous considérerons un certain nombre de groupes, chacun d'effectif infini, échangeant des individus, par migration. Un modèle de ce type est en fait peu réaliste, aussi, dans un deuxième temps, nous considérerons des populations d'effectif limité. Nous établirons alors des « modèles stochastiques » qui permettront d'étudier la variance entre groupes, introduite par l'existence de la migration. Dans une troisième partie nous tenterons de montrer, en fonction d'études sur les migrations matrimoniales, quelles hypothèses des modèles utilisés devraient être modifiées, pour cerner de plus près la réalité.

1. — MODÈLES DÉTERMINISTES FAISANT INTERVENIR LA MIGRATION

Soit une population infiniment nombreuse décomposée en m groupes, chacun d'effectif « infini ». Nous supposons qu'à chaque génération g des migrations ont lieu entre ces groupes : on notera $l_r^k(g)$ la proportion de migrants originaires du groupe r et présents dans le groupe k à la génération g . Ceci implique l'hypothèse que les migrations ont toutes eu lieu entre la naissance et le début de la phase reproductrice des individus : pendant cette phase les divers groupes sont supposés panmictiques. La proportion d'autochtones du groupe r est alors notée $l_r^r(g)$ avec la relation évidente :

$$\sum_r l_r^k(g) = 1$$

Il sera commode d'utiliser la notation matricielle :

$$L(g) = (l_r^k(g))$$

où $L(g)$ est une matrice carrée stochastique d'ordre m , l'indice supérieur k étant l'indice ligne, l'indice inférieur r l'indice colonne.

Ainsi, à chaque génération les échanges de migrants qui interviennent entre les m groupes sont définis par la matrice $L(g)$. Cette matrice peut évidemment être modifiée d'une génération à l'autre, les flux migratoires n'étant pas constants; cependant afin de simplifier les formulations nous admettrons dans tous les calculs qui suivent, que les divers termes l_r^k sont les mêmes durant toute la période étudiée, ce qui permet de caractériser les échanges entre groupes par une matrice unique L .

1.1. — Évolution de la structure génique.

Désignons par $s^k(g)$ la structure génique du groupe k à la génération g , c'est-à-dire :

$$s^k(g) = (\varpi_1^k(g), \dots, \varpi_i^k(g), \dots, \varpi_n^k(g))$$

où les ϖ_i^k sont les fréquences des divers allèles dans ce groupe.

Les structures géniques des m groupes sont donc représentées par m vecteurs à n composantes. Pour simplifier l'écriture, nous représenterons par $\Omega(g)$ la matrice rectangulaire de type (mn) dont chacune des m lignes est constituée par les composantes d'un vecteur $s^k(g)$; soit :

$$\Omega(g) = (\varpi_i^k(g)) \quad \text{avec} \quad \sum_i \varpi_i^k(g) = 1$$

En supposant que les fréquences des allèles sont les mêmes dans chaque lot de migrants que dans le groupe dont il est issu, il vient :

$$\omega_i^k(g) = \sum_r l_r^k \omega_i^r (g - 1)$$

relation qui s'écrit plus simplement sous forme matricielle :

$$\Omega(g) = L \cdot \Omega(g - 1)$$

En remontant à la génération initiale, on obtient :

$$(1) \quad \Omega(g) = L^g \Omega(0)$$

Ainsi l'existence d'une structure génique limite, est liée à celle d'une limite de la matrice L^g lorsque $g \rightarrow \infty$. Pour étudier cette limite il est commode d'exprimer la matrice L^g en fonction des valeurs propres et des vecteurs propres de L . On a vu au chapitre 11 que, la matrice L étant stochastique, toutes ses valeurs propres sont de module inférieur ou égal à 1, l'une d'elles au moins étant égale à 1. De plus si tous les éléments de la diagonale principale sont non nuls, cette valeur propre $\lambda = 1$ est la seule de module 1. Cette condition peut en général, être supposée réalisée pour les populations humaines : elle signifie en effet qu'il n'existe pas de colonie envoyant à chaque génération tous ses membres dans les autres colonies, sans en garder aucun.

La matrice L peut alors s'écrire sous la forme :

$$L = USU^{-1}$$

S étant une matrice partitionnée de la forme :

$$S = \begin{pmatrix} I_{\rho_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & I_{\rho_2}(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & I_{\rho_p}(\lambda_p) \end{pmatrix}$$

où $\lambda_1, \dots, \lambda_p$ sont les diverses valeurs propres de L , d'ordre ρ_1, \dots, ρ_p . Les matrices d'ordre ρ_i , $I_{\rho_i}(\lambda_i)$ peuvent s'écrire :

$$I_{\rho_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i \end{pmatrix}$$

La matrice U est non singulière, ce qui entraîne l'existence de U^{-1} . On peut alors calculer $L^g = US^gU^{-1}$ où S^g s'écrit :

$$S^g = \begin{pmatrix} I_{\rho_1}^g(\lambda_1) & 0 & \dots & 0 \\ 0 & I_{\rho_2}^g(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & I_{\rho_p}^g(\lambda_p) \end{pmatrix}$$

avec :

$$I_{\rho_i}^g(\lambda_i) = \begin{pmatrix} \lambda_i^g & \binom{g}{1} \lambda_i^{g-1} & \dots & \binom{g}{\rho_i - 1} \lambda_i^{g-\rho_i+1} \\ 0 & \lambda_i^g & \dots & \binom{g}{\rho_i - 2} \lambda_i^{g-\rho_i+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_i^g \end{pmatrix}$$

Lorsque $g \rightarrow \infty$ toutes les matrices $I_{\rho_i}^g(\lambda_i)$ d'ordre ρ_i correspondant aux valeurs propres $\lambda_i \neq 1$ tendent par conséquent vers des matrices dont tous les termes sont nuls.

Nous admettons ici que la valeur propre 1 est simple, le vecteur propre u_1 lui correspondant, a toutes ses composantes égales. A la limite, la matrice S^g a tous ses éléments nuls à l'exception de la valeur propre 1. Le produit $\lim_{g \rightarrow \infty} US^g$ tend donc vers une matrice dont seule la colonne correspondant au vecteur propre u_1 n'est pas nulle. Enfin, le produit $\lim_{g \rightarrow \infty} US^g U^{-1}$ tend vers une matrice dont toutes les lignes sont identiques.

Le produit de cette matrice par la matrice $\Omega(0)$ donnera encore une matrice Ω du même type. Ainsi à la limite les structures géniques de tous les groupes deviennent identiques. Soit s cette structure limite : elle dépend en général à la fois de la structure initiale $\Omega(0)$ et de la matrice de migration L .

La vitesse avec laquelle la différence $\Omega(g) - \Omega$ tend vers 0 ne dépend que de la plus grande des valeurs $|\lambda_k|$ inférieure à 1.

Cas particuliers. — Le cas le plus simple est celui de deux groupes dont l'un seulement reçoit des migrants issus de l'autre; ces deux groupes sont alors liés par la matrice de migration.

$$L = \begin{pmatrix} 1 - m & m \\ 0 & 1 \end{pmatrix}$$

où m est la proportion des individus du groupe 1 qui proviennent du groupe 2. On voit immédiatement que :

$$L^g = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1 - m)^g & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} (1 - m)^g & 1 - (1 - m)^g \\ 0 & 1 \end{pmatrix}$$

Lorsque $g \rightarrow \infty$, $L^g \rightarrow \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ et la structure du premier groupe, seule variable, tend vers celle du second groupe. Elle lui est sensiblement égale dès que g dépasse l'ordre de grandeur de $\frac{1}{m}$.

Le cas un peu plus complexe où les deux groupes ont des échanges réciproques, c'est-à-dire sont liés par la matrice de migration :

$$L = \begin{pmatrix} 1 - m_1 & m_1 \\ m_2 & 1 - m_2 \end{pmatrix}$$

conduit à la solution d'équilibre :

$$(2) \quad \left\| \right. \quad \mathbf{L}^g \underset{g \rightarrow \infty}{=} \begin{pmatrix} \frac{m_2}{m_1 + m_2} & \frac{m_1}{m_1 + m_2} \\ \frac{m_2}{m_1 + m_2} & \frac{m_1}{m_1 + m_2} \end{pmatrix}$$

Si la structure génique des populations initiales était : $\Omega(0) = \begin{pmatrix} p_1 & 1 - p_1 \\ p_2 & 1 - p_2 \end{pmatrix}$
 Alors la structure génique de la population limite sera :

$$\Omega = \begin{pmatrix} \frac{p_1 m_2 + p_2 m_1}{m_1 + m_2} & 1 - \frac{p_1 m_2 + p_2 m_1}{m_1 + m_2} \\ \frac{p_1 m_2 + p_2 m_1}{m_1 + m_2} & 1 - \frac{p_1 m_2 + p_2 m_1}{m_1 + m_2} \end{pmatrix}$$

Cette structure dépend bien de la structure initiale et de la matrice de migration.

1.2. — Évolution de la structure génotypique.

La structure génotypique des divers groupes tend, comme la structure génique, vers une limite commune à tous les groupes lorsque $g \rightarrow \infty$.

Cependant, avant d'atteindre cette limite, une hétérogénéité, liée au brassage incomplet de la population, se maintient pendant un nombre de générations d'autant plus élevé que les taux de migration sont plus faibles. Cette hétérogénéité peut être mesurée par l'écart entre la structure génotypique réelle \bar{S} de la population considérée dans son ensemble et la structure théorique d'une population panmictique, dont la structure génique serait la même que celle de la population étudiée, c'est-à-dire la structure panmictique équivalente.

Supposons que l'effectif du $k^{\text{ième}}$ groupe, toujours très nombreux, soit N_k et posons $N = \sum_k N_k$. Alors la structure génotypique de l'ensemble des groupes peut s'écrire :

$$\bar{S} = \frac{\sum_k N_k \bar{S}^k}{N} = \frac{\sum_k N_k \bar{S}^{k^2}}{N}$$

La structure génique est :

$$s = \frac{\sum_k N_k s^k}{N}$$

à laquelle correspond une structure panmictique équivalente :

$$\bar{S}_P = \bar{s}^2$$

L'écart entre la structure génotypique réelle de la population et la structure

panmictique équivalente est donc

$$(3) \quad \left\| \bar{S} - \bar{S}_P = \frac{\sum_k N_k \vec{s}^k{}^2}{N} - \vec{s}^2 = \frac{\sum_k N_k (\vec{s}^k - \vec{s})^2}{N} \right.$$

Il correspond aux variances statistiques des diverses fréquences.

Lorsque l'équilibre est atteint, ce trimat a tous ses éléments nuls.

Cas de deux allèles.

Dans le cas particulier, où il n'y a que deux allèles, cet écart peut être caractérisé par un unique coefficient. Écrivant en effet

$$s = (p, q) \quad \text{avec} \quad p + q = 1$$

il vient

$$\begin{aligned} \frac{\sum_k N_k (p^k - p)^2}{N} &= \frac{\sum_k N_k (q^k - q)^2}{N} = \sigma^2 \\ \frac{2 \sum_k N_k p^k q^k}{N} - 2pq &= 2\sigma^2 \end{aligned}$$

et l'on peut écrire

$$\bar{S} - \bar{S}_P = \begin{vmatrix} \sigma^2 & & \\ -2\sigma^2 & \sigma^2 & \\ & & \end{vmatrix} = \sigma^2 \begin{vmatrix} 1 & & \\ -2 & 1 & \\ & & \end{vmatrix}$$

Définissant f par la relation :

$$\sigma^2 = fpq$$

on a alors

$$\bar{S} = \bar{S}_P + f \begin{vmatrix} pq & & \\ -2pq & pq & \end{vmatrix} = \bar{S}_P + f \begin{vmatrix} p & & \\ 0 & q & \end{vmatrix} - f \begin{vmatrix} p^2 & & \\ 2pq & q^2 & \end{vmatrix}$$

soit

$$(4) \quad \left\| \bar{S} = (1 - f)\bar{S}_P + f\bar{S}_H \right.$$

\bar{S}_H étant la structure homozygote équivalente de la population.

Cette relation correspond à la classique formule de Wahlund. M. Lamotte a donc pu dire, dans ce cas particulier, que la population présentait une « consanguinité de position »; si l'on rapproche la relation (4) ci-dessus de la relation (18 bis) du chapitre 8 (paragraphe 2.4) on constate en effet qu'elles sont formellement identiques et que le coefficient f joue le même rôle que le coefficient de consanguinité α . Il convient toutefois de remarquer que dans le cas général, où il y a plus de deux allèles, on ne peut plus parler d'un coefficient mesurant la consanguinité de position. Le trimat $\bar{S} - \bar{S}_P$ dépend alors de plus d'un coefficient. De plus il est nécessaire de bien marquer la différence entre les

significations des paramètres α et f : l'un représente une probabilité, l'autre mesure le rapport de la variance des fréquences entre les divers groupes au produit pq des fréquences moyennes dans l'ensemble des groupes.

1.3. — Applications aux populations réelles.

Un tel modèle peut s'appliquer de plusieurs façons, selon les données dont on dispose. En général les structures géniques des diverses populations ne sont connues que sur une courte période. Cependant dans certains cas, diverses sous-populations ont été extraites de populations plus importantes, et placées dans un nouveau milieu (cas de la colonisation). On peut connaître la structure initiale de ces sous-populations, sous l'hypothèse que la structure génique des populations d'où elles ont été extraites, n'a pas varié au cours du temps. Connaissant le nombre de générations écoulées depuis la colonisation, g , les structures géniques des colonisateurs $\Omega(0)$, supposées identiques aux structures actuelles des populations d'origine, et les structures actuelles $\Omega(g)$ de la colonie, on peut donc déduire la matrice de migration L , grâce à la relation

$$\Omega(g) = L^g \Omega(0)$$

Populations noire et blanche des États-Unis. — C'est la méthode qu'ont utilisée B. Glass, C. C. Li, D. F. Roberts, pour déterminer les taux de migration raciale entre noirs et blancs aux États-Unis.

Les calculs de B. Glass et de C. C. Li, ont été effectués pour 7 allèles différents, en supposant que la fréquence d'un allèle parmi les blancs est restée constante au cours des trois derniers siècles, et que le nombre de générations écoulé depuis la mise en contact des colonies est de 10. Ce cas est celui que nous avons traité plus haut où

$$L = \begin{pmatrix} 1 - m & m \\ 0 & 1 \end{pmatrix}$$

Il conduit à des estimations de m assez voisines, comprises entre :

$$0,028 \leq m \leq 0,056.$$

Ces conditions de migrations restant constantes ($m = 0,0358$), la fréquence d'équilibre de l'allèle R^0 , par exemple, ne sera atteinte que dans 60,7 générations, soit à peu près dans 1 670 ans, la durée d'une génération étant prise égale à 27,5 ans. On peut admettre qu'actuellement la proportion des gènes de la population « noire » qui sont originaires de la population « blanche » est de l'ordre de 30 %, ce qui correspond à $m \simeq 0,03$.

D. F. Roberts, partant de données plus précises sur l'origine des noirs, arrive à des valeurs de m comprises entre $0,02 \leq m \leq 0,025$. Dans un article postérieur il fait également intervenir la population indienne.

Populations Nilotes du Soudan. — Un autre cas est celui où, au contraire, on possède des données sur les migrations matrimoniales. Connaissant la

structure de la population à un instant donné, on peut calculer cette structure pour une génération antérieure ou postérieure, sous l'hypothèse que la migration est indépendante du temps.

Ces données montrent en général la faible valeur des migrations matrimoniales entre groupes ⁽¹⁾. Ainsi D. F. Roberts et R. W. Hiorns donnent la matrice des migrations matrimoniales entre trois populations du Soudan, les Nuer, les Dinka et les Shilluk :

$$L = \begin{pmatrix} 0,9850 & 0,0125 & 0,0025 \\ 0,0138 & 0,9775 & 0,0087 \\ 0 & 0,0098 & 0,9902 \end{pmatrix}$$

Il en résulte que, si les structures géniques de ces populations sont différentes, un nombre très important de générations est nécessaire pour atteindre l'équilibre. C'est cette faiblesse des taux de migrations entre groupes qui permet de garder un certain sens à la notion de race humaine, notion qui peut se ramener à celle de la consanguinité de position. Mais il faut bien voir que cette distinction entre races s'efface peu à peu au cours du temps, pour finalement disparaître, lorsque les effectifs sont infinis. Comme nous le verrons dans la seconde partie, ce n'est que dans le cas où les groupes sont d'effectif limité que peut se maintenir une hétérogénéité entre groupes, par le seul jeu du hasard.

1.4. — Modèles déterministes dans lesquels, en plus de la migration, d'autres conditions de la panmixie parfaite ne sont plus vérifiées.

Dans les modèles suivants on admet, comme précédemment, que les effectifs des divers sous-groupes sont très grands et que la matrice de migration est indépendante du temps.

1.4.1. Cas de la migration continue dans le temps. — La migration continue dans le temps peut être définie par une matrice M , telle que, en multipliant tous ses termes par l'intervalle de temps infiniment petit dt , on obtienne les variations de composition de la population considérée pour chacun des allèles.

$$M = (m_j)$$

Il convient de noter que cette matrice n'est plus stochastique comme la matrice L considérée plus haut. En effet ses termes ne sont pas tous positifs, ils vérifient la condition :

$$\sum_j m_j = 0$$

⁽¹⁾ Ces valeurs sont cependant très supérieures aux taux de mutations, de l'ordre de 10^{-5} , comme on l'a vu au chapitre 11.

Supposant que l'intervalle de temps séparant 2 générations soit égal à 1, et que durant cet intervalle la structure génique de la population reste identique pour passer de façon discontinue de la valeur $\Omega(g - 1)$ à $\Omega(g)$ à l'instant g , on peut écrire :

$$\Omega(g) - \Omega(g - 1) = M\Omega(g - 1)$$

Ayant d'autre part la relation : $\Omega(g) = L\Omega(g - 1)$ il vient : $L = M + I$ relation reliant les deux matrices de migration, L dans le cas discontinu, M dans le cas continu.

Revenant à la migration continue dans le temps, on peut écrire

$$\Omega(t + dt) = \Omega(t) + M\Omega(t) dt$$

soit

$$d[\Omega(t)] = M\Omega(t) dt$$

On peut montrer que cette équation a, lorsque toutes les valeurs propres ρ_i de M sont simples, ce que nous supposons par la suite, une solution qui s'écrit

$$\Omega(t) = US^tU^{-1}\Omega(0)$$

S étant une matrice diagonale dont le $i^{\text{ème}}$ terme est égal à $e^{\rho_i t}$.

Comme précédemment l'existence d'une structure limite est liée à celle d'une limite de la matrice S^t lorsque $t \rightarrow \infty$. On a vu plus haut que la matrice $M + I$ était stochastique : il s'ensuit donc que ses valeurs propres sont de module inférieur ou égal à 1. Or, les valeurs propres de cette matrice sont égales à $1 + \rho_i$, ρ_i étant les valeurs propres de M . Il s'ensuit donc que :

$$|1 + \rho_i| \leq 1$$

Si ρ_i est réel, il vient : $\rho_i \leq 0$, donc $e^{\rho_i t}$ tendra, quand $t \rightarrow \infty$ vers 0 si $\rho_i < 0$, et vers 1 si $\rho_i = 0$.

Si ρ_i est imaginaire, écrivons $\rho_i = r_i(\cos \theta + i \sin \theta)$; la condition précédente donne $r_i \cos \theta \leq 0$. Il s'ensuivra que $e^{\rho_i t} = e^{r_i t \cos \theta} e^{i r_i t \sin \theta}$.

Quand $t \rightarrow \infty$ le premier terme, si $\cos \theta < 0$, tend vers 0 tandis que le second peut s'écrire $\cos(r_i t \sin \theta) + i \sin(r_i t \sin \theta)$ et reste toujours fini. Dans ce cas $e^{\rho_i t} \rightarrow 0$. Par contre si $\cos \theta = 0$ il vient $r_i = 0$ d'où $e^{\rho_i t} = 1$.

Donc, S^t tend alors, quand $t \rightarrow \infty$, vers une matrice déterminée et $\Omega(t)$ a une limite.

Comme précédemment les structures géniques de tous les groupes deviennent donc identiques lorsque $t \rightarrow \infty$.

Dans le cas de deux groupes liés par la matrice de migration :

$$M = \begin{pmatrix} -m_1 & m_1 \\ m_2 & -m_2 \end{pmatrix}$$

il vient :

$$S = \begin{pmatrix} 1 & 0 \\ 0 & e^{-(m_1+m_2)} \end{pmatrix}$$

soit

$$\begin{aligned}\Omega(t) &= \begin{pmatrix} 1 & -m_1 \\ 1 & m_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-t(m_1+m_2)} \end{pmatrix} \begin{pmatrix} \frac{m_2}{m_1+m_2} & \frac{m_1}{m_1+m_2} \\ -\frac{1}{m_1+m_2} & \frac{1}{m_1+m_2} \end{pmatrix} \Omega(0) \\ &= \begin{pmatrix} \frac{m_2 + m_1 e^{-t(m_1+m_2)}}{m_1+m_2} & \frac{m_1(1 - e^{-t(m_1+m_2)})}{m_1+m_2} \\ \frac{m_2(1 - e^{-t(m_1+m_2)})}{m_1+m_2} & \frac{m_1 + m_2 e^{-t(m_1+m_2)}}{m_1+m_2} \end{pmatrix} \Omega(0)\end{aligned}$$

quand $t \rightarrow \infty$

$$\Omega(t) \rightarrow \begin{pmatrix} \frac{m_2}{m_1+m_2} & \frac{m_1}{m_1+m_2} \\ \frac{m_2}{m_1+m_2} & \frac{m_1}{m_1+m_2} \end{pmatrix} \Omega(0)$$

Cette théorie basée sur l'hypothèse de continuité a de l'intérêt dans la mesure où le nombre d'années dont on dispose pour l'étude de la population est assez restreint. Par contre, l'étude des limites montre que celles-ci sont identiques dans les modèles continus et discontinus, seul change le rythme avec lequel la population tend vers sa limite. On pourra vérifier l'identité des limites dans la formule 2.

Pour connaître la rapidité de la convergence vers l'état d'équilibre, on a dans le cas discontinu :

$$\begin{aligned}\Omega(g) &= \begin{pmatrix} 1 & -m_1 \\ 1 & m_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & [1 - (m_1 + m_2)]^g \end{pmatrix} \begin{pmatrix} \frac{m_2}{m_1+m_2} & \frac{m_1}{m_1+m_2} \\ -\frac{1}{m_1+m_2} & \frac{1}{m_1+m_2} \end{pmatrix} \Omega(0) \\ &= \begin{pmatrix} \frac{m_2 + m_1[1 - (m_1 + m_2)]^g}{m_1+m_2} & \frac{m_2(1 - [1 - (m_1 + m_2)]^g)}{m_1+m_2} \\ \frac{m_2(1 - [1 - (m_1 + m_2)]^g)}{m_1+m_2} & \frac{m_1 + m_2[1 - (m_1 + m_2)]^g}{m_1+m_2} \end{pmatrix} \Omega(0)\end{aligned}$$

On constate donc que dans le cas où la migration est supposée discontinue $[\Omega(g) - \Omega(\infty)]$ tend vers 0 comme $(1 - a)^g$, où $a = m_1 + m_2$; dans le cas où, au contraire, elle est supposée continue, cette différence tend vers 0 comme e^{-ag} , en supposant l'unité de temps égale à une génération. Il s'ensuit que le second modèle tend plus rapidement que le premier vers sa limite. Cette différence est d'autant plus grande que a est élevé, c'est-à-dire que le nombre de migrants est important.

1.4.2. Migration et mutation. — Nous supposons maintenant que les gènes subissent également des mutations. La probabilité (ou taux) de mutation d'un allèle en un autre allèle est supposée indépendante du groupe dans lequel cet allèle se trouve : soit v^x la probabilité de mutation de l'allèle x en l'allèle y . (on a $v_x^x = 1 - \sum_{y \neq x} v_y^x$).

La fréquence $\omega_i^k(g)$ de l'allèle i dans le $k^{\text{ième}}$ groupe avant mutation, mais après migration est donnée par

$$\omega_i^k(g) = \sum_r l_r^k \omega_i^r(g-1)$$

Après mutation elle devient égale à :

$$\begin{aligned} \omega_i^k(g) &= \omega_i^k(g) - \sum_{y \neq i} v_y^i \omega_i^k(g) + \sum_{x \neq i} v_i^x \omega_x^k(g) \\ &= \omega_i^k(g) \left[1 - \sum_{y \neq i} v_y^i \right] + \sum_{x \neq i} v_i^x \omega_x^k(g) \\ &= \sum_x v_i^x \omega_x^k(g) \end{aligned}$$

Appelant V la matrice stochastique de terme général v_y^x , d'ordre n , cette relation peut s'écrire :

$$\Omega(g) = L\Omega(g-1)V = L^g\Omega(0)V^g$$

Nous avons vu précédemment que, quand $g \rightarrow \infty$, la matrice $L^g\Omega(0)$ tend, dans le cas où la seule valeur propre de module 1 est 1 et est racine simple de l'équation aux valeurs propres, vers une matrice Ω dont toutes les lignes sont identiques. Sous les mêmes conditions, V^g tend vers une matrice Y dont toutes les lignes sont identiques. Donc, dans ce cas, le produit $L^g\Omega(0)V^g$ tend également vers une matrice dont toutes les lignes sont identiques, et cette matrice n'est autre que Y , comme on peut facilement le voir. Ainsi, dans ce cas, la structure limite est indépendante de la structure initiale de la population et de la matrice de migration L . Elle ne dépend que de la matrice de mutation.

G. Malécot a traité le cas particulier où il n'y a que 2 allèles, en notant u la quantité v_2^1 et v la quantité v_1^2 . La matrice Y est alors égale à :

$$Y = \begin{pmatrix} \frac{v}{u+v} & \frac{u}{u+v} \\ \frac{v}{u+v} & \frac{u}{u+v} \end{pmatrix}$$

Par conséquent, par exemple la fréquence $\omega_1^k(g) \xrightarrow{g \rightarrow \infty} \frac{v}{u+v}$.

La vitesse avec laquelle les $\omega_1^k(g)$ tendent vers leur limite ne dépend que de $(1-u-v)^g$. Ainsi cette limite est sensiblement atteinte lorsque g dépasse l'ordre de grandeur de $\frac{1}{u+v}$. Or, on sait que les taux de mutation sont généralement de l'ordre de 10^{-5} . Le temps nécessaire à l'égalisation des fréquences est donc très élevé, de l'ordre de 10^5 générations.

Par ailleurs, nous avons vu que les taux de migration sont en général beaucoup plus importants. Ainsi les populations se comportent dans un premier temps comme si la mutation ne jouait pas. Elles tendent alors vers un état d'équilibre Ω , fonction de la structure initiale et de la matrice de migration L

(nous avons étudié plus haut cet état). Puis dans un deuxième temps cette structure tend lentement vers la structure d'équilibre Y définie seulement par les taux de mutation. Au cours de cette deuxième période la population dans son ensemble se conforme à la loi de Hardy, alors que pendant la première se manifeste une « consanguinité de position ». Il convient de noter cependant ce qu'un tel schéma a de théorique : il suppose que durant tous les phases de cette évolution les caractéristiques de migration et de mutation restent constantes, ce qui est bien peu réaliste.

2. — MODÈLES STOCHASTIQUES FAISANT INTERVENIR LA MIGRATION

On considèrera maintenant une population finie répartie en m groupes, d'effectif N_i , entre lesquels des échanges ont lieu par migration. Dans un premier temps, les autres conditions de la panmixie étant supposées réalisées, on étudiera la limite vers laquelle tendent les fréquences géniques des divers groupes, ainsi que les moments d'ordre deux qui donnent une mesure des fluctuations existant entre eux, une fois atteinte la position d'équilibre. Dans un second temps on considèrera divers modèles où d'autres conditions de la panmixie ne seront plus supposées réalisées.

Pour définir le tirage des $2N$ gamètes utiles de la génération g il est nécessaire de faire un certain nombre d'hypothèses. En premier lieu on suppose que le nombre des gamètes, émis par la génération $g - 1$ est très grand. Il s'ensuit que le tirage se fait parmi un nombre supposé infini de gamètes présents dans les proportions $\varpi_i(g - 1)$. Il est également nécessaire de faire l'hypothèse que la contenance des deux lots de gamètes mâles et femelles est identique. Enfin le tirage des gamètes utiles dans chaque lot est supposé effectué dans une urne à composition constante : le fait qu'un individu ait déjà subi le tirage d'un certain nombre de ses gamètes, au i ème tirage, n'intervient pas pour le $(i + 1)$ ième tirage.

Ces conditions étant réalisées, le nombre de gamètes utiles mâles (ou femelles) $N_i \varpi_i(g)$ porteurs du i ème allèle obéit à une loi binômiale prenant la valeur entière n_i (avec $0 \leq n_i \leq N_i$) avec la probabilité :

$$P_{n_i} = \binom{N_i}{n_i} \varpi_i^{n_i}(g - 1) [1 - \varpi_i(g - 1)]^{N_i - n_i}$$

Comme nous l'avons vu à propos des populations d'effectif limité, les espérances et variances de $\varpi_i(g)$ conditionnées par la connaissance de $\varpi_i(g - 1)$ sont

$$E_{g-1} [\varpi_i(g)] = \varpi_i(g - 1), \quad \sigma_{g-1}^2 [\varpi_i(g)] = \frac{\varpi_i(g - 1)[1 - \varpi_i(g - 1)]}{2N_i}$$

Voyons maintenant comment l'introduction de la migration modifie ces relations.

2.1. — Migration.

Les notations seront les mêmes que dans la première partie pour la matrice de migration L , indépendante de la génération. D'autre part, la relation (1) du paragraphe 1-1 montre qu'il est possible de se limiter à l'étude de l'évolution d'un seul allèle. Nous ne considérons donc ici que le $i^{\text{ème}}$ allèle dont les fréquences dans les m groupes sont données sous forme d'un vecteur colonne.

$$\varpi_i(g) = \begin{pmatrix} \varpi_i^1(g) \\ \varpi_i^2(g) \\ \vdots \\ \varpi_i^m(g) \end{pmatrix}$$

On supprimera par la suite l'indice i devenu superflu.

Il sera enfin nécessaire de supposer que la migration a lieu avant la période reproductrice des individus, ceux-ci restant par la suite dans le même groupe pendant toute cette période. Ce modèle s'applique en particulier dans le cas où les seules migrations effectuées par les individus sont des migrations matrimoniales.

2.1.1. Étude des espérances des fréquences géniques. — La migration intervenant avant la phase reproductrice, la structure génique, qui est soumise au tirage au sort des gamètes utiles de la génération g , est :

$$\varpi'(g) = L\varpi(g-1)$$

Après ce tirage au sort, la structure génique réellement obtenue peut s'écrire

$$\varpi(g) = L\varpi(g-1) + e(g)$$

où $e(g)$ est un vecteur colonne dont le $i^{\text{ème}}$ terme $\varepsilon^i(g)$ représente la variation fortuite de la fréquence de l'allèle considéré dans le $i^{\text{ème}}$ groupe. Son espérance et sa variance conditionnées par la connaissance de la génération $g-1$, vérifient les relations :

$$E_{g-1}[\varepsilon^i(g)] = 0, \quad \sigma_{g-1}^2[\varepsilon^i(g)] = \frac{\varpi'^i(g)[1 - \varpi'^i(g)]}{2N_i} = \sigma_{g-1}^2[\varpi^i(g)]$$

Il s'ensuit pour l'étude des espérances :

$$E[\varpi(g)] = E\{E_{g-1}[\varpi(g)]\} = LE[\varpi(g-1)]$$

Si l'on suppose la génération initiale connue, il vient en résolvant la récurrence précédente

$$E[\varpi(g)] = L^g\varpi(0)$$

Ainsi l'étude de l'espérance se ramène à celle du modèle déterministe étudié dans la première partie. Lorsque $g \rightarrow \infty$, $E[\varpi(g)] \rightarrow \varpi$ vecteur dont tous les éléments sont identiques. Ce vecteur dépend en général de la matrice de migration et de la structure génique initiale des diverses populations.

Dans le cas particulier, où il existe un groupe envoyant des migrants dans tous les autres sans lui même en recevoir aucun, on voit facilement que les fréquences de l'allèle considéré dans les autres groupes tendent toutes vers celle du groupe « émetteur ».

2.1.2. Étude des moments d'ordre 2. — Écrivons les variances et covariances entre les divers groupes sous la forme :

$$\begin{aligned} u^{ij}(g) &= \text{cov} [\varpi^i(g), \varpi^j(g)] = u^{ji}(g) \\ u^{ii}(g) &= \sigma^2[\varpi^i(g)] \end{aligned}$$

Introduisons également les variances et covariances conditionnées par la connaissance de la génération $(g - 1)$ que l'on notera $u'^{ij}(g - 1)$.

Le problème à traiter consiste à voir s'il y a une relation de récurrence entre $u^{ij}(g)$ et $u'^{ij}(g - 1)$. Ce point établi, il faudra ensuite voir si cette relation a une limite lorsque $g \rightarrow \infty$.

Examinons tout d'abord le cas des covariances conditionnées. Les fréquences de la génération $(g - 1)$ étant supposées connues, il vient :

$$\begin{aligned} u'^{ij}(g - 1) &= \sum_z l_z^i [\varpi^z(g - 1) - E\{\varpi^z(g - 1)\}] \\ &\quad \sum_w l_w^j [\varpi^w(g - 1) - E\{\varpi^w(g - 1)\}] + E_{g-1}(\varepsilon^i \varepsilon^j) \end{aligned}$$

Supposant l'indépendance en probabilité entre les tirages dans les « réservoirs » i et j , alors :

$$\begin{aligned} \text{si } i \neq j, \quad E_{g-1}(\varepsilon^i \varepsilon^j) &= 0 \\ \text{si } i = j, \quad E_{g-1}(\varepsilon^i \varepsilon^j) &= \frac{\varpi^i(g)[1 - \varpi^i(g)]}{2N_i} \end{aligned}$$

L'espérance de la covariance conditionnée nous donnera la covariance à priori :

$$u^{ij}(g) = \sum_z \sum_w l_z^i l_w^j u^{zw}(g - 1) + \delta_j^i \frac{E\{\varpi^i(g)\} - E[\{\varpi^i(g)\}^2]}{2N_i}$$

où :

$$\delta_j^i = 0 \quad \text{si } i \neq j \quad \text{et} \quad \delta_i^i = 1$$

Or, on sait que

$$\begin{aligned} E\{\varpi^i(g)\} &= \sum_j l_j^i E\{\varpi^j(g - 1)\} \\ E[\{\varpi^i(g)\}^2] &= [E\{\varpi^i(g)\}]^2 + E\{(\varpi^i(g) - E\{\varpi^i(g)\})^2\} \\ &= \left[\sum_j l_j^i E\{\varpi^j(g - 1)\} \right]^2 + \sum_z \sum_w l_z^i l_w^i u^{zw}(g - 1) \end{aligned}$$

Ainsi la covariance à priori de la génération g s'exprime en fonction des covariances et des moyennes de la génération $(g - 1)$.

Nous avons vu au paragraphe précédent que $E\{\varpi^i(g)\}$ tendait vers une limite ϖ : lorsque g sera supérieur à un certain nombre, fonction uniquement

de la valeur propre de plus grand module, autre que 1, on pourra supposer cette limite atteinte. Nous nous placerons désormais dans ce cas. Alors la relation donnée plus haut, s'écrit plus simplement sous la forme :

$$u^{ij}(g) = \sum_z \sum_w l_z^i l_w^j \left(1 - \frac{\delta_j^i}{2N_i} \right) u^{zw}(g-1) + \delta_j^i \frac{\bar{w} - \bar{w}^2}{2N_i}$$

Introduisons les covariances et les espérances sous forme vectorielle (vecteurs colonnes de dimension m^2) (1).

$$v(g) = \begin{pmatrix} u^{11}(g) \\ u^{12}(g) \\ \vdots \\ u^{1m}(g) \\ u^{21}(g) \\ \vdots \\ u^{2m}(g) \\ u^{31}(g) \\ \vdots \\ u^{mm}(g) \end{pmatrix} \quad \mathcal{R} = \begin{pmatrix} \frac{\bar{w} - \bar{w}^2}{2N_1} \\ 0 \\ \vdots \\ 0 \\ \frac{\bar{w} - \bar{w}^2}{2N_2} \\ \vdots \\ 0 \\ \frac{\bar{w} - \bar{w}^2}{2N_3} \\ \vdots \\ 0 \end{pmatrix}$$

et la matrice Λ , d'ordre m^2 dont les éléments sont (en numérotant par le couple $(i, j) = r$ les lignes, et par le couple $(z, w) = s$ les colonnes)

$$\lambda_s^r = l_z^i l_w^j \left(1 - \frac{\delta_j^i}{2N_i} \right)$$

On voit alors que la relation entre les variances à la génération g et $g-1$ peut s'écrire

$$(5) \quad v(g) = \Lambda v(g-1) + \mathcal{R}$$

Soit encore en supposant que $E[\bar{w}^i(g)]$ est égal à sa limite à partir de la génération g :

$$(6) \quad v(g+n) = \Lambda^n v(g-1) + \sum_{j=0}^{j=n} \Lambda^j \mathcal{R}$$

La somme des éléments de chaque ligne de Λ est égale à 1, si $i \neq j$, à $1 - \frac{1}{2N_i}$ si $i = j$. Il s'ensuit donc que les valeurs propres de cette matrice sont de module inférieur ou égal à 1. Dans le cas général où toutes les colonies échangent des migrants entre elles, aucune valeur propre ne pourra être égale

(1) Il serait possible de ne donner que des vecteurs de dimension $\frac{m(m-1)}{2}$, du fait que $u^{ij}(g) = u^{ji}(g)$. La forme utilisée facilite la construction de la matrice Λ suivante, mais pour les calculs, il serait préférable de prendre le vecteur de dimension minimum.

à 1. En effet, supposons qu'il existe une valeur propre égale à 1. Soit $v^r = ce^{i\theta}$ la composante du vecteur propre correspondant de plus grand module. Ce module vérifie la relation

$$C \leq \sum_k \lambda_k^j |u^k| \leq c \sum_k \lambda_k^j \leq c$$

La première égalité n'est vérifiée que si toutes les composantes du vecteur propre ont même argument θ , la seconde si tous les modules des composantes du vecteur propre sont égaux à C , la dernière si la r ième ligne de la matrice ne correspond pas à $r = (i, i)$. Or, toutes les composantes du vecteur propre vérifiant cette relation, la dernière condition ne sera pas vérifiée. Il s'ensuit donc qu'aucune des valeurs propres ne sera égale à 1. Nous savions déjà qu'aucune des valeurs propres ne pouvait être de module 1 et d'argument différent de $2k\pi$, le terme λ_i^j étant différent de zéro. Dans ce cas-là on peut affirmer, de façon plus restrictive, que toutes les valeurs propres de Λ sont de module inférieur à 1.

On a vu précédemment que Λ pouvait être écrite

$$\Lambda = USU^{-1}$$

où S est une matrice partitionnée de la forme :

$$S = \begin{pmatrix} I_{\rho_1}(\lambda_1) & 0 & \dots & 0 \\ 0 & I_{\rho_2}(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & I_{\rho_p}(\lambda_p) \end{pmatrix}$$

— et $\lambda_1, \dots, \lambda_p$ sont les valeurs propres différentes de Λ d'ordre ρ_1, \dots, ρ_p . Les matrices d'ordre ρ_i , $I_{\rho_i}(\lambda_i)$ s'écrivent

$$I_{\rho_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & & \lambda_i \end{pmatrix}$$

De plus, la matrice U est non singulière, ce qui entraîne l'existence de U^{-1} . On peut alors calculer :

$$\Lambda^n = US^nU^{-1}$$

où S^n s'écrit

$$S^n = \begin{pmatrix} I_{\rho_1}^n(\lambda_1) & 0 & \dots & 0 \\ 0 & I_{\rho_2}^n(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & I_{\rho_p}^n(\lambda_p) \end{pmatrix}$$

avec

$$I_{\rho_i}^n(\lambda_i) = \begin{pmatrix} \lambda_i^n & \binom{n}{1} \lambda_i^{n-1} & \dots & \binom{n}{\rho_i-1} \lambda_i^{n-\rho_i+1} \\ 0 & \lambda_i^n & & \binom{n}{\rho_i-2} \lambda_i^{n-\rho_i+2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \lambda_i^n \end{pmatrix}$$

On sait déjà que lorsque $n \rightarrow \infty$, Λ^n tend vers une matrice dont tous les termes sont nuls. Il nous reste à voir vers quelle limite tend la série

$$\sum_{j=0}^n \Lambda^j = U \left(\sum_{j=0}^n S^j \right) U^{-1} \quad \text{lorsque } n \rightarrow \infty$$

La matrice $\sum_{j=0}^n S^j$ aura ses termes de la forme :

$$1 + r\lambda + \frac{r(r+1)}{2} \lambda^2 + \dots + \binom{n}{r-1} \lambda^{n-r+1}$$

Lorsque $r = 1$, cette série se résoud immédiatement

$$1 + \lambda + \lambda^2 + \dots + \lambda^n + \dots = \frac{1}{1-\lambda} = \varphi_1(\lambda)$$

Lorsque $r = 2$, on a la série

$$1 + 2\lambda + 3\lambda^2 + \dots + n\lambda^{n-1} + \dots$$

On voit facilement qu'elle se déduit de la précédente par dérivation par rapport à λ . Donc sa somme est égale à

$$\varphi_2(\lambda) = \frac{d[\varphi_1(\lambda)]}{d\lambda} = \frac{1}{(1-\lambda)^2}$$

De proche en proche on voit que toutes ces séries seront convergentes :

$$\varphi_r(\lambda) = \frac{d^{r-1}[\varphi_1(\lambda)]}{d\lambda^{r-1}} = \frac{(r-1)!}{(1-\lambda)^r}$$

Donc :

$$\lim_{n \rightarrow \infty} \sum_{j=0}^n I_{\rho_i}^j(\lambda_i) = \begin{pmatrix} \frac{1}{1-\lambda_i} & \frac{1}{(1-\lambda_i)^2} & \dots & \frac{(\rho_i-1)!}{(1-\lambda_i)^{\rho_i}} \\ 0 & \frac{1}{1-\lambda_i} & \dots & \frac{(\rho_i-2)!}{(1-\lambda_i)^{\rho_i-1}} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \frac{1}{1-\lambda_i} \end{pmatrix}$$

De même la matrice $\sum_{j=0}^n \Lambda^j$ tend vers une limite finie Λ' .

Ainsi dans la mesure où toutes les valeurs propres de la matrice Λ sont de module inférieur à 1, le vecteur $\mathcal{V}(g+n)$ tend vers un vecteur déterminé lorsque $n \rightarrow \infty$. Ce vecteur s'écrit

$$\mathcal{V} = \Lambda' \mathcal{R}$$

Il dépend donc de toutes les valeurs propres de la matrice Λ , et du vecteur \mathcal{R} . Une autre façon de le calculer consiste à écrire la relation (5), en faisant $\mathcal{V}(g) = \mathcal{V}(g-1) = \mathcal{V}$, $\mathcal{V} = \Lambda \mathcal{V} + \mathcal{R}$. La résolution du système de Cramer correspondant à cette équation matricielle donne \mathcal{V} .

L'étude de la vitesse avec laquelle les variances tendent vers leurs valeurs asymptotiques montre que cette vitesse est fonction d'une part de la vitesse avec laquelle la moyenne tend vers sa limite (fonction de la plus grande des valeurs propres inférieures à 1 de L) et de la vitesse avec laquelle $\sum_{j=0}^n \Lambda^j \mathcal{R}$ tend vers sa limite. Cette vitesse est en particulier fonction de la plus grande des valeurs propres de Λ .

Un cas particulier est intéressant à étudier : c'est celui où il existe un groupe i d'effectif infini envoyant des migrants dans tous les autres groupes sans lui-même en recevoir aucun. On a vu que tous les groupes tendent vers une structure génique identique à celle du groupe i . Voyons si les variances tendent également vers une limite. Le groupe « émetteur » étant d'effectif infini $\frac{\sigma - \sigma^2}{2N_i} = 0$. De même, ne recevant aucun élément extérieur, on a pour sa variance : $v^{ii}(g) = 0$ et pour les covariances $v^{ij}(g) = 0$. La relation entre les covariances, aux générations g et $(g - 1)$, peut s'écrire, en négligeant ce groupe émetteur

$$v'(g) = \Lambda' v'(g - 1) + \mathcal{R}'$$

les vecteurs considérés étant de dimension $m^2 - m = m(m - 1)$, et la matrice Λ' d'ordre $m(m - 1)$. De plus le groupe i envoyant des migrants dans tous les autres il s'ensuit que : $\sum_{j=m+1}^{m^2} \lambda_j < 1$. Il en résulte que le vecteur $v'(g + n)$ tend vers un vecteur déterminé lorsque $n \rightarrow \infty$.

Traisons le cas correspondant à 2 groupes seulement dont l'un est de population infinie et de composition constante. L'autre d'effectif N , reçoit dans la proportion m des migrants provenant du groupe précédent. Ce cas correspond au « modèle de l'île » de Sewal Wright. On voit facilement que seule la variance $u(g)$ du groupe d'effectif N est non nulle. Elle vérifie la relation,

$$u(g) = (1 - m)^2 \left(1 - \frac{1}{2N} \right) u(g - 1) + \frac{p(1 - p)}{2N}$$

où p est la fréquence du gène considéré dans la population isolée. Sa limite u lorsque $g \rightarrow \infty$ vérifie l'équation :

$$u \left[1 - (1 - m)^2 \left(1 - \frac{1}{2N} \right) \right] = \frac{p(1 - p)}{2N}$$

soit

$$u = \frac{p(1 - p)}{1 + 4Nm - 2m - 2Nm^2 + m^2}$$

Lorsque m est faible on peut négliger les termes en m , Nm^2 et m^2 pour obtenir la valeur

$$(7) \quad \parallel u = \frac{p(1 - p)}{1 + 4Nm}$$

Étudions la vitesse avec laquelle $u(g) - u$ tend vers 0. Elle dépend nous l'avons vu, du plus grand module des valeurs propres de L inférieures à 1, qui est ici égal à m . La quantité $(E\{\varpi(g)\} - p)$ peut être considérée comme pratiquement nulle lorsque g dépasse l'ordre de grandeur de $\frac{1}{m}$. D'autre part, supposant cette quantité nulle on peut écrire

$$\begin{aligned} u(g+n) - u &= (1-m)^2 \left(1 - \frac{1}{2N}\right) [u(g+n-1) - u] \\ &= (1-m)^{2n} \left(1 - \frac{1}{2N}\right)^n [u(g) - u] \simeq \left(1 - 2m - \frac{1}{2N}\right)^n [u(g) - u] \end{aligned}$$

On voit alors que, lorsque n dépasse l'ordre de grandeur de $\frac{1}{2m + \frac{1}{2N}}$,

on peut être assuré que cette différence est nulle. Ainsi après un nombre de génération de l'ordre de $\frac{1}{m} + \frac{2N}{1 + 4Nm}$ l'équilibre est sûrement atteint.

Il convient de noter qu'il s'agit là d'une estimation maximum du nombre de générations devant s'écouler avant d'atteindre l'équilibre.

Signalons l'étude faite par Bodmer et Cavalli Sforza, qui suppose que les fréquences géniques restent proches de la valeur $\frac{1}{2}$. Utilisant la transformation suivante : $\varpi^k(g) = \sin^2\theta^k(g)$ où θ est exprimé en radians, on trouve pour la valeur de la variance conditionnée :

$$\sigma_{g-1}^2[\varepsilon^k(g)] = \frac{1}{8N_k} + 0 \left(\frac{1}{N_k}\right)^2$$

valeur indépendante des fréquences géniques à la génération g .

Il devient alors inutile de supposer que la limite de $E\{\varpi^k(g)\}$ est atteinte et l'on peut étudier l'évolution des variances de la population, de façon simple dès la première génération. Il vient, en posant

$$\mathcal{R}' = \begin{bmatrix} \frac{1}{8N_1} \\ 0 \\ \vdots \\ 0 \\ \frac{1}{8N_2} \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathcal{V}(g) = \wedge^g \mathcal{V}(0) + \sum_{j=0}^g \wedge^j \mathcal{R}' = \sum_{j=0}^g \wedge^j \mathcal{R}'$$

Les calculs effectués sur diverses matrices de migration, comportant des groupes qui envoient des migrants dans les autres groupes, mais n'en reçoivent pas eux-mêmes, montrent que les résultats restent satisfaisants tant que les fréquences géniques ne sont pas proches de 0 ou de 1. La variance croît rapi-

dement en fonction du temps et atteint sa valeur d'équilibre à partir d'un nombre de générations de l'ordre de $\frac{1}{\alpha}$, α étant le plus faible des pourcentages de migrants reçus des groupes extérieurs.

2.1.3. Autres possibilités d'étude de l'effet des migrations. — Dans les paragraphes précédents, nous avons introduit la migration entre groupes d'effectifs donnés. A l'intérieur de chaque groupe les conditions de la panmixie parfaite, exception faite de leur effectif limité, sont supposées réalisées. Cette condition peut sembler peu réaliste et l'on conçoit l'avantage d'introduire la migration sous forme continue dans l'espace. G. Malécot a développé divers types de modèles faisant intervenir la probabilité pour qu'un parent d'un individu né en un lieu Q , soit lui-même né dans une aire dS_P centrée au point P . Cette probabilité est supposée être une fonction continue des coordonnées des deux points P et Q . L'introduction de cette loi de migration permet d'écrire la moyenne et la variance des fréquences d'un gène à la génération g en fonction de la génération $g - 1$. Les limites quand $g \rightarrow \infty$ de ces deux quantités vérifient des équations intégrales, dont la résolution n'est en général possible que par approximations successives.

Une autre méthode pour l'étude des moments d'ordre 2 consiste à étudier l'évolution du coefficient de parenté entre les emplacements i et j . Ce coefficient φ_{ij} est défini comme la probabilité pour qu'un gène pris au hasard dans le réservoir gamétique de l'emplacement i soit identique à un gène pris au hasard dans le réservoir gamétique de l'emplacement j . Dans la mesure où la population aura atteint son état stationnaire, on montre que les variances et covariances a priori entre groupes peuvent s'exprimer en fonction du coefficient de parenté.

En effet, introduisons les variables aléatoires indicatrices X_y^i , qui sont égales à 1 ou à 0 selon que $y^{\text{ème}}$ gamète de l'emplacement i (il y en a $2N_i$ au total) est l'allèle considéré ou un autre allèle. On écrit alors

$$\varpi^i(g) = \frac{\sum_y X_y^i(g)}{2N_i}, \quad \varpi^j(g) = \frac{\sum_z X_z^j(g)}{2N_j}$$

On sait que, si l'état d'équilibre est atteint, $E[\varpi^i(g)] = E[\varpi^j(g)] = \varpi$. D'autre part :

$$\text{cov} [\varpi^i(g), \varpi^j(g)] = \frac{1}{4N_i N_j} \sum_y \sum_z E[(X_y^i(g) - \varpi)(X_z^j(g) - \varpi)]$$

Deux cas se présentent alors. Ou bien les gamètes y et z tirés sont identiques, ce qui peut se produire avec la probabilité $\varphi_{ij}(g)$, alors l'espérance mathématique correspondante s'écrit :

$$E[\{X_y(g) - \varpi\}^2] = E[\{X_y(g)\}^2] - \varpi^2 = \varpi - \varpi^2$$

car

$$X_y^2 = X_y$$

ou bien les gamètes y et z tirés sont non identiques, ce qui peut se produire

avec la probabilité $1 - \varphi_{ij}(g)$; mais alors la relation

$$E\{(X_i^i(g) - \varpi)(X_j^j(g) - \varpi)\} = 0$$

est vérifiée en raison de l'indépendance stochastique des tirages. Il vient donc :

$$\text{cov} [\varpi^i(g), \varpi^j(g)] = \frac{1}{4N_i N_j} \times 4N_i N_j (\varpi - \varpi^2) \varphi_{ij}(g) = (\varpi - \varpi^2) \varphi_{ij}(g)$$

On trouve de même que :

$$\sigma^2[\varpi^i(g)] = (\varpi - \varpi^2)\varphi_{ii}(g) + \frac{\varpi - \varpi^2}{2N_i} [1 - \varphi_{ii}(g)]$$

Ces relations montrent la possibilité d'étudier les moments d'ordre 2 par l'intermédiaire des coefficients de parenté.

2.2. — Modèles stochastiques dans lesquels, en plus de la migration, d'autres conditions de la panmixie parfaite ne sont plus vérifiées.

Nous envisageons plus particulièrement dans ce paragraphe les effets de la mutation et de la sélection linéarisée venant se superposer à ceux de la migration. Cette simplification est possible car on ne considère que l'état stationnaire supposé atteint par la population.

2.2.1. Migration et mutation. — Introduisons maintenant la possibilité de mutation avec les taux introduits précédemment (paragraphe 1-4-2). On fait l'hypothèse qu'il y a en premier lieu migration d'individus, puis mutation sur les gamètes émis, en nombre supposé infini, par ces individus, enfin tirage au sort des gamètes utiles. Il sera nécessaire dans ce paragraphe de réintroduire tous les allèles, sous forme matricielle.

Nous avons déjà vu que, avant le tirage au sort des gamètes utiles :

$$\varpi_i^k(g) = \sum_x \sum_r v_x^r l_r^k \varpi_x^r(g-1)$$

Il s'ensuit que

$$E\{\Omega(g)\} = LE\{\Omega(g-1)\}V = L^g \Omega(0)V^g$$

L'évolution de la moyenne est identique à celle étudiée pour le modèle déterministe. Lorsque $g \rightarrow \infty$, $E\{\Omega(g)\}$ tend vers une matrice Y dont toutes les lignes sont identiques.

Pour étudier les covariances, on écrit :

$$\text{cov}(\varpi_i^k, \varpi_j^l) = u_{ij}^{kl} \quad \text{et} \quad \sigma^2(\varpi_i^k) = u_{ii}^{kk}$$

Les covariances conditionnées correspondantes sont u'_{ij}^{kl} et u'_{ii}^{kk} .

On peut alors écrire la relation suivante, donnant la covariance conditionnée :

$$u'_{ij}^{kl}(g-1) = \sum_x \sum_r v_x^r l_r^k (\varpi_x^r(g-1) - \varpi_x) \sum_y \sum_s v_y^s l_s^l [\varpi_y^s(g-1) - \varpi_y] + E_{g-1}(\epsilon_i^k \epsilon_j^l)$$

La covariance à priori est alors égale à

$$\begin{aligned}
 u_{ij}^{kl}(g) &= \sum_x \sum_r \sum_y \sum_s v_i^x v_j^y l_r^k l_s^l u_{xy}^{rs}(g-1) - \frac{\delta_{ij}^{kl} E(\varpi_i^{k'} \varpi_j^{l'})}{2N_k} + \delta_{ij}^{kl} \frac{E[\varpi_i^{k'}(1 - \varpi_i^{l'})]}{2N_k} \\
 &= \sum_x \sum_r \sum_y \sum_s v_i^x v_j^y l_r^k l_s^l \left(1 - \frac{\delta_l^k}{2N_k}\right) u_{xy}^{rs}(g-1) - \delta_{ij}^{kl} \frac{\varpi_i \varpi_j}{2N_k} \\
 &\quad + \delta_{ij}^{kl} \frac{(\varpi_i - \varpi_i^2)}{2N_k}
 \end{aligned}$$

où

$$\begin{aligned}
 \delta_{ij}^{kl} &= 0 \text{ si } k \neq l \text{ ou si } k = l \text{ et } i = j; & \delta_{ij}^{kk} &= 1 \text{ si } i \neq j \\
 \delta_{ij}^{kl} &= 0 \text{ si } k \neq l \text{ ou si } i \neq j; & \delta_{ii}^{kk} &= 1 \\
 \delta_i^k &= 0 \text{ si } k \neq l; & \delta_k^k &= 1
 \end{aligned}$$

Cette relation peut s'écrire plus simplement sous forme matricielle en supposant que les moyennes ont déjà atteint leur état stationnaire à la génération g . Les matrices suivantes sont définies en numérotant par le couple supérieur les lignes et le couple inférieur des colonnes.

Λ matrice carrée d'ordre m^2 , dont le terme général est $l_r^k l_s^l \left(1 - \frac{\delta_l^k}{2N_k}\right)$

W matrice carrée d'ordre n^2 , dont le terme général est $v_i^x v_j^y$

$U(g)$ matrice rectangulaire du type (m^2, n^2) de terme général u_{xy}^{rs}

R matrice de type (m^2, n^2) dont seuls les termes r_{ij}^{kl} sont non nuls et égaux à :

$$-\frac{\varpi_i \varpi_j}{2N_k} \quad \text{si } i \neq j, \quad \frac{\varpi_i - \varpi_i^2}{2N_k} \quad \text{si } i = j$$

On peut alors écrire

$$U(g) = \Lambda U(g-1)W + R$$

Relation qui donne

$$U(g+n) = \Lambda^n U(g-1)W^n + \sum_{j=0}^n \Lambda^j R W^j$$

Dans la mesure où Λ n'a pas de valeurs propres égales à 1, $\Lambda^n \rightarrow 0$ et la limite lorsque $n \rightarrow \infty$ de $U(g+n)$, est celle de la somme :

$$U = \sum_{j=0}^{\infty} \Lambda^j R W^j$$

Cette limite peut encore se calculer comme solution U de l'équation matricielle :

$$U = \Lambda U W + R$$

2.2.2. Migration, mutation (cas de 2 allèles seulement), sélection linéarisée.

— Les hypothèses de ce paragraphe sont voisines de celles du précédent. On suppose que la migration a d'abord lieu. Elle est suivie d'une étape correspondant à la mutation et à la sélection zygotique et gamétique linéarisée. Enfin intervient le tirage au sort des gamètes. La formalisation d'un tel modèle

est simple du fait que l'on peut s'en tenir à la fréquence d'un seul allèle, n'ayant considéré que des locus où seulement deux allèles sont présents. L'indice correspondant à cet allèle sera supprimé par la suite.

Partant de la valeur $\varpi^i(g)$ dans le $i^{\text{ème}}$ groupe il vient après migration

$$\varpi_1^i(g) = \sum_r l_r^i \varpi^r(g)$$

La mutation et la sélection linéarisée introduisent le coefficient de rappel vers la position d'équilibre k , et cette position d'équilibre ϖ_0 . Leur application conduit à la nouvelle valeur

$$\varpi_2^i(g) = \varpi_1^i(g) - k[\varpi_1^i(g) - \varpi_0]$$

Enfin le tirage au sort des $2N_i$ gamètes donne finalement, introduisant la dérive fortuite

$$\varpi^i(g+1) = \varpi_2^i(g) + \varepsilon^i(g)$$

Comme précédemment on a :

$$E_g\{\varepsilon^i(g)\} = 0, \quad \sigma_g^2\{\varepsilon^i(g)\} = \frac{\varpi_2^i(g) [1 - \varpi_2^i(g)]}{2N_i}$$

Sous forme matricielle le passage de la génération g à la génération $(g+1)$ peut s'écrire :

$$\varpi(g+1) - \varpi = (1-k)L[\varpi(g) - \varpi] + \varepsilon(g)$$

en désignant par ϖ le vecteur colonne dont tous les éléments sont égaux à ϖ_0 .

L'étude de la moyenne se fait simplement et conduit à la relation :

$$E[\varpi(g+1) - \varpi] = (1-k)LE[\varpi(g) - \varpi]$$

Cette relation montre que quand $g \rightarrow \infty$, $E[\varpi(g) - \varpi] \rightarrow 0$ donc à la limite $E[\varpi(g)]$ va tendre vers ϖ . L'espérance de la fréquence d'une gène dans un groupe tend ainsi vers une limite indépendante du groupe considéré.

Pour pousser plus avant les calculs, nous allons traiter un exemple concernant une matrice de migration simple. On suppose les groupes répartis linéairement, chaque groupe d'effectif constant N n'échangeant une proportion fixe de migrants qu'avec les deux groupes voisins, comme cela est indiqué par la figure 61 :

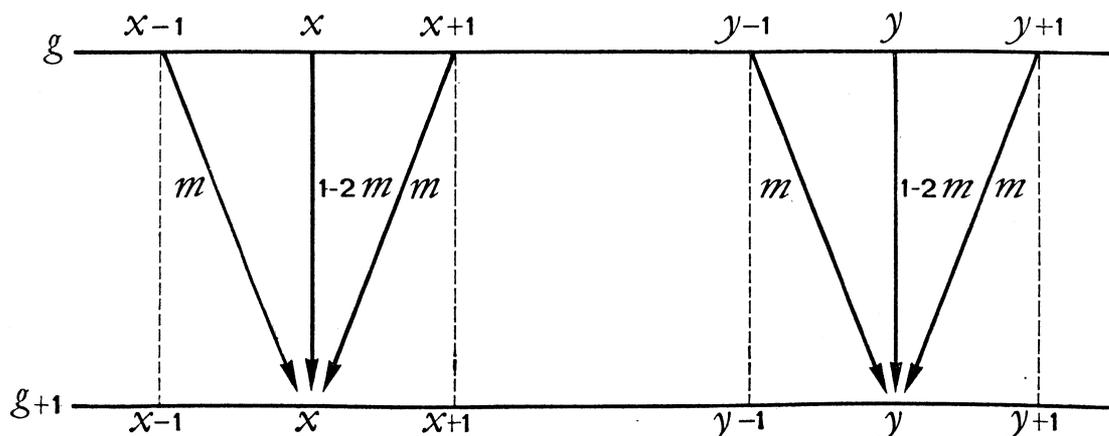


FIG. 61. — Migrations contiguës à coefficient constant.

Les effectifs de tous ces groupes sont supposés égaux. L'état asymptotiquement stationnaire sera défini par les relations du type suivant :

$$\begin{aligned} v_y^x &= (1 - k)^2 \{ m^2 [v_{y-1}^{x-1} + v_{y+1}^{x-1} + v_{y-1}^{x+1} + v_{y+1}^{x+1}] \\ &\quad + m(1 - 2m) [v_y^{x-1} + v_y^{x+1} + v_{y-1}^x + v_{y+1}^x] \\ &\quad + (1 - 2m)^2 v_y^x \} \left(1 - \frac{\delta_y^x}{2N} \right) + \frac{\delta_y^x (\varpi - \varpi^2)}{2N} \end{aligned}$$

En fait les covariances entre colonies ne sont fonction que de la distance les séparant. On a donc en posant la distance xy égale à d , et négligeant les termes en m^2

$$\begin{aligned} v(d) &= (1 - k)^2 \{ 2m[v(d+1) + v(d-1)] + (1 - 4m)v(d) \} + \frac{\delta_y^x (\varpi - \varpi^2)}{2N} \\ &\quad - \frac{\delta_y^x (1 - k)^2}{2N} [(1 - 4m)v(0) + 4mv(1)] \end{aligned}$$

Dans le cas où $d \neq 0$, $v(d)$ est défini par une récurrence du deuxième ordre dont on sait que les solutions sont de la forme h^d , h vérifiant l'équation

$$h^d = (1 - k)^2 \{ 2mh^{d+1} + 2mh^{d-1} + (1 - 4m)h^d \}$$

Négligeant les termes en k^2 et en km , cette équation s'écrit plus simplement :

$$h = (1 - 4m - 2k)h + 2m + 2mh^2$$

Soit :

$$h^2 - \left(2 + \frac{k}{m} \right) h + 1 = 0$$

dont les racines sont :

$$h = 1 + \frac{k}{2m} \pm \sqrt{\frac{k^2}{4m^2} + \frac{k}{m}}$$

La solution $v(d)$ lorsque $d \rightarrow \infty$ devant rester bornée, seule la racine inférieure à 1 devra être considérée et l'on a :

$$v(d) = \lambda \left(1 + \frac{k}{2m} - \sqrt{\frac{k^2}{4m^2} + \frac{k}{m}} \right)^d$$

λ étant une constante à déterminer en fonction de $v(0)$.

Or, pour $d = 1$, on a la relation :

$$v(1) = (1 - 4m + 2k)v(1) + 2mv(0) + 2mv(2)$$

remplaçant $v(1)$ et $v(2)$ par leurs valeurs, il vient :

$$2m\lambda h^2 - (4m + 2k)\lambda h + 2mv(0) = 0 \quad \text{soit} \quad \lambda = v(0)$$

Calculons maintenant la variance $v(0)$, qui est donnée par l'équation :

$$v(0) = (1 - 4m - 2k)v(0) + 4mv(1) + \frac{\varpi - \varpi^2 - (1 - 4m - 2k)v(0) - 4mv(1)}{2N}$$

ou encore

$$(4m + 2k)v(0) - 4mv(1) = \frac{\varpi - \varpi^2 - v(0)}{2N - 1}$$

$$2\sqrt{4mk + k^2}v(0) = \frac{\varpi - \varpi^2 - v(0)}{2N - 1}$$

Supposant 1 négligeable devant N, il vient finalement

$$v(0) = \frac{\varpi - \varpi^2}{1 + 4N\sqrt{4mk + k^2}}$$

et la solution générale s'écrit :

$$(8) \quad \left\| v(d) = \frac{\varpi - \varpi^2}{1 + 4N\sqrt{4mk + k^2}} \left(1 + \frac{k}{2m} - \sqrt{\frac{k^2}{4m^2} + \frac{k}{m}} \right)^d \right.$$

Dans le cas où k est très petit devant m , c'est-à-dire où la pression de migration est très supérieure aux pressions de mutation et de sélection, cette solution se simplifie :

$$v(d) = \frac{\varpi - \varpi^2}{1 + 8N\sqrt{mk}} \left(1 - \sqrt{\frac{k}{m}} \right)^d$$

On peut approcher cette formule par une exponentielle :

$$(9) \quad \left\| v(d) = \frac{\varpi - \varpi^2}{1 + 8N\sqrt{mk}} e^{-d\sqrt{\frac{k}{m}}} \right.$$

mettant en évidence la décroissance exponentielle du coefficient de corrélation $\frac{v(d)}{v(0)}$ en fonction de la distance. On constate en particulier qu'il est indépendant de l'effectif N dans divers groupes.

3. — OBSERVATIONS DES MIGRATIONS DANS LES POPULATIONS RÉELLES

Tous les modèles utilisés jusqu'ici supposaient connus les taux de migration entre les divers groupes. De plus dans l'étude de l'état stationnaire vers lequel tendent ces groupes, il a été nécessaire de faire l'hypothèse que ces taux sont constants durant toute la période étudiée. Enfin, au cours du développement des modèles proposés, certaines hypothèses ont dû être faites sur les modalités de la migration (migrations ayant lieu avant la phase reproductrice des individus.)

Il convient maintenant de confronter ces hypothèses avec les résultats des observations effectuées dans divers pays sur la migration des individus.

L'étude des migrations humaines s'est fortement développée au cours de

ces dernières années, et de nombreux modèles ont été proposés en vue de permettre une meilleure description des migrations et une prévision de leur évolution à venir.

3.1. — Modèles migratoires.

Avant d'analyser les divers types de modèles proposés il est nécessaire de définir avec précision la signification du terme « migration ». Considérons l'ensemble des déplacements effectués par un individu, nous admettons que constituent des « migrations », ceux qui entraînent un changement de résidence de cet individu. On néglige donc ainsi tous les déplacements temporaires (navettes, tournées etc...); cette définition n'a de sens que dans la mesure où la résidence de l'individu est connue tout au long de son existence.

Les « migrations matrimoniales » constituent une catégorie particulière de migration; elles peuvent être définies

— soit en comparant le lieu de naissance d'un individu et le lieu de naissance de ses enfants (deux migrations distinctes peuvent alors être envisagées selon que l'on considère la mère ou le père de l'individu).

— soit en comparant les lieux de résidence des époux avant le mariage.

La définition utilisée dépend bien entendu des données statistiques dont on dispose pour mesurer ces « migrations matrimoniales ».

3.1.1. Principaux types de modèles. — L'analyse des données empiriques montre vite que la distance séparant deux zones a et b intervient dans le nombre Y_b^a de migrants échangés, ainsi que les effectifs des populations des zones d'origine et de destination. Le modèle le plus simple et le plus souvent utilisé pour relier Y à ces divers paramètres est le modèle dit « du type Pareto » qui peut s'écrire sous la forme

$$Y_b^a = k \frac{\delta(a)\delta(b)}{r^\alpha} ds(a) ds(b)$$

où $\delta(x)$ est la densité de population au point x , r la distance séparant les deux lieux a et b , k et α des constantes.

S. Stouffer a tenté de substituer au paramètre « distance géographique » un paramètre « distance sociale », caractérisé par le nombre total de migrants présents dans les zones intermédiaires entre a et b . Ces deux variables, distance sociale et distance géographique, sont bien entendu très fortement corrélées et seule l'application des deux modèles à une même population permettra de conclure à la supériorité de l'un ou de l'autre.

Dans une autre voie, T. Hägerstrand, admettant qu'un lien très étroit existe entre les migrations actuelles et les migrations antérieures, a élaboré un modèle faisant intervenir le nombre de migrants originaires de la zone a et installés dans la zone b durant une période antérieure de n années. La distance est introduite indirectement dans ce modèle sous la forme des migrations

de son fait
ds(a) ds(b)

« actives ». Ces migrations sont celles effectuées par les individus qui migrent sans être attirés par des personnes connues d'eux dans le lieu de destination, elles sont supposées être de faible importance.

D'autres auteurs ont élaboré des modèles plus complexes introduisant un grand nombre de variables socio-économiques. En particulier G. Olsson étudiant le lien entre la distance géographique parcourue et 8 caractéristiques définissant le migrant, montre que cette distance géographique est bien déterminée par ces caractéristiques. Il prouve ainsi que la distance constitue une variable permettant de rendre compte correctement de nombreux facteurs intervenant dans la migration.

Les modèles testés sur des migrations matrimoniales sont, le plus souvent, des types précédents; ils montrent qu'il existe un isomorphisme entre ces migrations et les migrations générales.

3.1.2. Application de ces modèles aux populations réelles. — Le modèle le plus souvent testé, est celui du type Pareto. Appliqué à des migrations tant américaines, que françaises, suédoises ou japonaises, il a donné des résultats généralement satisfaisants. L'exposant α , que l'on pensait initialement devoir être une constante, s'est trouvé en fait être variable selon les régions et selon les périodes étudiées.

Ainsi, pour les migrations suédoises, il varie de 0,4 à 3,3, les fortes valeurs correspondant, en général, aux migrations des populations rurales, les faibles valeurs à celles des populations urbaines. Hägerstrand (1957) étudiant l'évolution de l'exposant α sur 100 ans, pour plusieurs paroisses suédoises, a prouvé qu'il est le plus souvent une fonction décroissante du temps.

Notre étude (1969) sur les migrations françaises de 1896 à 1962, montre qu'il est par contre possible de garder α égal à 2 pendant toute cette période mais à condition d'introduire un coefficient correctif l , fonction du recensement considéré; la loi de Y peut alors s'écrire :

$$Y_b^a = \delta(a)\delta(b) \left[\frac{k}{r^2} + l \right] \delta(a)\delta(b)$$

grâce au paramètre supplémentaire l'on peut obtenir un bon ajustement aux données étudiées.

Cette variation des coefficients au cours du temps contredit dans le cas des populations humaines l'hypothèse de base de nombreux modèles de génétique de population.

Un des inconvénients du modèle du type Pareto, est qu'il n'est pas applicable sur tout l'intervalle $0, \infty$: il donne en effet un nombre infini de non migrants ($r = 0$). Il est possible d'éviter cette difficulté, en introduisant une constante β , telle que la loi de Pareto s'écrive :

$$Y_b^a = \frac{k\delta(a)\delta(b)}{(\beta + r)^\alpha} \delta(a)\delta(b)$$

Pour pouvoir écrire la loi de probabilité d'une migration d'origine a et d'amplitude r , il est nécessaire de connaître la forme du territoire où se produit cette

migration, ainsi que sa densité de population en chaque point. Deux cas particuliers donnent des résultats simples. Le territoire peut être considéré comme unidimensionnel (cas d'une vallée par exemple) et la densité constante. La loi de probabilité est alors :

$$f(r) = \frac{(\alpha - 1)\beta^{\alpha-1}}{(\beta + r)^\alpha} \quad \text{pour } r \geq 0$$

Ou bien le territoire peut être considéré comme une surface infinie et de densité δ constante. La loi de probabilité est alors :

$$f(r) = \frac{(\alpha - 1)(\alpha - 2)\beta^{\alpha-2}r}{(\beta + r)^\alpha}$$

Il convient de noter ici le modèle de type voisin utilisé par Cavalli Sforza (1962) pour étudier les migrations matrimoniales dans la vallée de Parme :

$$f(r) = \frac{k^\alpha}{\Gamma(\alpha)} e^{-kr} r^{\alpha-1}$$

où

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$$

Cette fonction du type III de Pearson peut s'ajuster plus facilement à des données empiriques que la loi précédente.

Le modèle de Stouffer a été testé sur des données américaines et suédoises conjointement avec un modèle du type Pareto. De nombreux essais n'ont pas permis de conclure à sa supériorité sur le précédent. Par contre, le modèle de T. Hägerstrand s'est montré capable de prédire avec une précision bien plus grande, que les deux précédents, les migrations suédoises. Nous l'avons également appliqué, avec quelques modifications aux migrations françaises : il nous a permis de décrire l'évolution des migrations de 1896 à 1962. Ce modèle n'a malheureusement pas encore été appliqué à des problèmes de génétique de population.

3.2. — *Compatibilité des modèles génétiques avec les modèles migratoires.*

Les modèles de génétique de population et les modèles migratoires se sont développés, le plus souvent, de façon indépendante. Il importe donc de voir si les hypothèses, qui sont à la base des modèles génétiques, ne sont pas en contradiction avec les résultats obtenus par les géographes ou les démographes. Cette analyse est évidemment nécessaire, avant tout développement des recherches dans cette voie.

3.2.1. Migration indépendante du temps. — Dans les modèles précédemment développés nous avons fait l'hypothèse, que la migration est une caractéristique des populations étudiées, indépendante du temps. Cette hypothèse n'est pas

nécessaire pour l'étude des modèles déterministes. En effet supposant une matrice de migration $L(g)$, variable avec la génération g , on peut écrire la relation :

$$\Omega(g) = L(g)\Omega(g - 1) = L(g)L(g - 1) \dots L(1)\Omega(0)$$

Ce modèle permettrait d'étudier l'évolution d'une population pendant une période où les taux de migration seraient communs, ou même de faire des prévisions à court terme, dans la mesure où les taux de migration sont des fonctions variant lentement avec le temps.

Par contre, nous avons développé les modèles stochastiques en faisant l'hypothèse que les populations avaient atteint leur état stationnaire. Dans la mesure où, à l'époque actuelle, les taux de migration ne sont pas stabilisés, on peut être certain que les populations ne sont pas dans cet état stationnaire.

Or, les recherches sur les migrations suédoises ont montré que si les taux de migration étaient stables durant la période 1785-1870, ceux-ci ont assez fortement évolué de 1870 à nos jours. De même notre étude sur les migrations françaises révèle une évolution de ce taux à partir de 1896.

Il est donc possible qu'un état d'équilibre ait été atteint dans les populations européennes avant 1870 : cet équilibre aurait pu être étudié avec un modèle stochastique. Mais depuis 1870 l'évolution des taux de migration va déplacer cet état d'équilibre, et le nouvel état stable vers lequel on tend, semble loin d'être atteint.

D'autres populations peuvent cependant ne pas avoir connu ce bouleversement et l'application des modèles stochastiques serait possible : mais il est nécessaire de vérifier ce fait sur une longue période avant de pouvoir admettre qu'elles sont suffisamment proches de leur état stationnaire. Par exemple, d'après Cavalli Sforza (1962) la population de la province de Parme semble avoir gardé des taux de migration identiques depuis trois siècles.

On voit ainsi que caractériser l'état stationnaire n'est en général pas suffisant : il paraît nécessaire de pousser plus avant l'étude de l'évolution de la variance, dans la phase qui précède cet état d'équilibre. C'est dans ce sens que sont allés Bodmer et Cavalli Sforza (1968), en cherchant l'évolution génération par génération de la variance, avec diverses matrices de migration. Ils n'ont cependant pas envisagé une modification des taux de migration au cours du temps, ce qui restreint la portée de leur étude.

3.2.2. Lois de migration utilisées. — Nous avons développé surtout des modèles génétiques introduisant la migration sous forme matricielle. Les études sur la migration, par contre, ont essayé, en majorité, de l'introduire sous forme d'une fonction de l'espace parcouru. Nous avons cependant indiqué une possibilité de développer des modèles de génétique, en introduisant la migration sous forme d'une fonction des lieux d'arrivée et de départ.

De nombreux modèles de Malécot vont dans cette voie, mais les lois de migrations utilisées (en général des lois normales) ne correspondent guère à la réalité. Il serait donc nécessaire d'essayer divers types de lois assez voisines des lois trouvées pour la migration, pour voir dans quelles mesures les résultats seraient modifiés par ce changement.

Un second point important est que les densités de population sont dans la plupart des modèles génétiques supposées constantes sur tout le territoire étudié. Cette hypothèse est en fait bien peu vérifiée par les populations réelles où les variations de densité de population sont très importantes, en particulier entre zones urbaines et rurales.

Enfin la loi de migration utilisée est supposée indépendante des lieux de départ et d'arrivée. Or, les études sur les migrations ont montré que sur un territoire donné les lois d'émigration (ou d'immigration) d'un lieu, pouvaient varier assez fortement en fonction de ce lieu.

Toutes ces conditions rendent donc difficile le traitement d'une population répartie de façon continue sur l'aire de migration.

Il semble dans un premier temps nécessaire de revenir au traitement des données sous forme matricielle. Décomposant la population considérée en sous-populations (plus leur nombre sera grand, plus l'étude sera fine) on peut calculer les taux de migration entre les diverses sous-populations d'effectifs et de fréquences géniques connus. L'évolution de ces groupes pourra alors être prévue sur une période plus ou moins longue, selon la rapidité de variation des taux de migration en fonction du temps.

3.2.3. Stabilité de l'individu après son mariage. — Cette hypothèse, qu'il a été nécessaire de faire pour développer des modèles en générations discontinues, n'est en général pas vérifiée par les populations humaines. Pour introduire plusieurs migrations dans la vie d'un individu, il sera nécessaire de se placer dans un modèle, où les générations ne seraient pas distinctes. Plusieurs études ont été faites, qui permettent de calculer la probabilité qu'un individu effectue une nouvelle migration, n années après la migration initiale. On pourra supposer, pour simplifier, que la répartition dans l'espace de ces nouvelles migrations est identique à celle des migrations initiales, mais cette hypothèse nécessiterait une confirmation pratique.

Il serait ainsi possible de réintroduire ce phénomène dans le modèle génétique, mais on peut penser *a priori* que son effet est minime.

4. — CONCLUSIONS

L'observation des migrations dans les populations réelles enlève beaucoup de portée à certains résultats établis dans les deux premières parties de ce chapitre sur les modèles de génétique faisant intervenir la migration. Les modèles stochastiques, développés par G. Malécot, s'ils sont applicables à certaines populations animales (*Cepaea Nemoralis*, étudiés par M. Lamotte), paraissent peu utiles à l'étude des populations humaines, pour lesquelles les conditions de migration évoluent au cours du temps. La méconnaissance de la limite, si elle existe, vers laquelle tendraient les taux de migration, ne rend possible que des prévisions, à moyen terme, de l'évolution de la structure génique des populations. Pour que cette prévision soit possible, il est d'ailleurs nécessaire de

développer, non pas l'étude de l'état stationnaire faite par G. Malécot, mais celle des étapes intermédiaires. Cette étude nécessite la connaissance précise des taux de migrations matrimoniales entre groupes. La recherche de ces taux, entreprise par le Dr Sutter, nous a permis de constater qu'ils étaient très proches des taux de migrations générales. Cet isomorphisme ouvre la possibilité d'utiliser dans les modèles de génétique des résultats obtenus sur les migrations générales.

Un autre point, que nous avons laissé jusqu'ici de côté, mérite d'être évoqué. Nous avons supposé dans les modèles faisant intervenir, en plus de la migration, la mutation et la sélection linéaire, que les taux de mutation et de sélection ne dépendaient que de l'allèle considéré, mais non du groupe auquel ils s'appliquaient. Cette hypothèse est en fait discutable et G. Malécot (1959) a développé un modèle où ces taux seraient fonction des divers groupes. L'utilité d'un tel modèle réside dans le fait que la connaissance de la variance entre groupes ne permet plus de déterminer les taux de migration, supposés inconnus, sans connaître les taux de mutation et de sélection des divers groupes. Inversement la connaissance des taux de migration permettrait la détermination indirecte des taux de mutation et de sélection chez l'homme. Il convient cependant de noter que ce modèle, comme les précédents, n'a été traité qu'en supposant l'état stationnaire atteint.

Un dernier point enfin doit être signalé. Tous les modèles développés dans ce chapitre considéraient que la migration s'effectuait entre les réservoirs gamétiques de chaque groupe (supposés d'effectif infini) et non au niveau des individus. Cette hypothèse permettait de considérer la fréquence d'un gamète dans le groupe après migration comme déterministe, la seule aléatoire intervenant au niveau du tirage au sort des gamètes utiles.

Or, en fait, la migration s'effectue au niveau des parents, introduisant ainsi une deuxième aléatoire dans le système étudié. En effet la fréquence du gène correspondant, dans le groupe de migrants, ne sera qu'en moyenne égale à sa fréquence dans le groupe initial : il est donc nécessaire d'introduire sa variance.

De plus, si l'on considère, non plus comme déterminé mais comme aléatoire, le nombre de migrants échangés par deux colonies, il est nécessaire d'introduire une troisième aléatoire dans le modèle étudié.

La formulation d'un modèle faisant intervenir un grand nombre d'aléatoires devient vite fort complexe; aussi, dans la voie qui vient d'être indiquée, le recours à des modèles de simulation, tels que les modèles dits « de Monte Carlo », semble s'imposer.

BIBLIOGRAPHIE.

Ouvrages démographiques ou géographiques sur les migrations.

De CANI John S., « On the construction of Stochastic Models of Population Growth and Migration », *Journal of Regional Science*, vol. 3, 1961, 1-13.

CAVALLI SFORZA Luigi, « The Distribution of Migration Distances : Models and Applications to Genetics », *Les Déplacements Humains, Entretiens de Monaco en Sciences Humaines*, 24-29 mai 1962, Hachette.

- COURGEAU Daniel, « Les Champs Migratoires en France », Thèse de doctorat de 3^e cycle, *Faculté des Lettres et Sciences Humaines*, Paris, 1969.
- HÄGERSTRAND Torsten, « Migration and Area », *Lund Studies in Geography*, n° 13, 1957, 27-158.
- KATZ Alvin and HILL Reuben, « Residential Propinquity and Marital Selection : A Review of Theory, Method and Fact », *Les Déplacements Humains, Entretiens de Monaco en Sciences Humaines*, 24-29 mai 1962, Hachette, 41-60.
- OLSSON Gunnar, « Distance and Human Interaction : A Review and Bibliography », *Regional Science Research Institute*, 1965; « Distance and Human Interaction: A Migration Study », *Geografiska Annaler*, vol. 47, 1965, 3-43.
- STOUFFER Samuel, « Intervening Opportunities : A Theory Relating Mobility and Distance », *American Sociological Review*, vol. 5, 1940, 845-867; « Intervening Opportunities and Competing Migrants », *Journal of Regional Science*, vol. 2, 1960, 1-26.
- SUTTER Jean, « Évolution de la distance séparant le domicile des futurs époux », *Population*, 13^e année, 1958, 227-258.
- SUTTER Jean et LUU MAU THANH, « Contribution à l'étude de la répartition des distances séparant les domiciles des époux dans un département français. Influence de la Consanguinité », *Les Déplacements Humains. Entretiens de Monaco en Sciences Humaines*, 24-29 mai 1962, Hachette, 123-137.
- ZIPF George Kinsley, « The $P_1 P_2/D$ Hypothesis : on the Intercity Movement of Persons », *American Sociological Review*, vol 11, 1946, 677-686.

Ouvrages de génétique de population sur le rôle de la migration.

- ALSTRÖM Carl Henry, « First-Cousin Marriages in Sweden 1750-1844 and a Study of the Population Movement in Some Swedish Subpopulations from the Genetic-Statistical Viewpoint », *Acta Genetica et Statistica Medica*, vol. 8, 1958, 295-369.
- BODMER Walter F. and CAVALLI-SFORZA Luigi, « A migration Matrix Model for the Study of Random Genetic Drift », *Genetics*, vol. 59, 1968, 565-592.
- CAVALLI-SFORZA Luigi, KIMURA M. and BARRAI I., « The Probability of Consanguineous Marriages », *Genetics*, vol. 54, 1966, 37-60.
- GLASS Bentley, « On the Unlikelihood of Significant Admixture of Genes from the North American Indians in the Present Composition of the Negroes of the United States », *The American Journal of Human Genetics*, vol. 7, 1955, 368-385.
- GLASS B. and LI CC., « The Dynamics of Racial Intermixture. An Analysis based of the American Negro », *The American Journal of Human Genetics*, vol. 5, 1953, 1-20.
- HALDANE J. B. S., « The Theory of a Cline », *Journal of Genetics*, 1948, 277-284.
- HIORNS R. W. HARRISSON G. A., BOYCE A. J. and KUCHEMANN C. F., « A Mathematical Analysis of the Effects of Movement on the Relatedness between Populations », *Annals of Human Genetics*, vol. 32, 1969, 237-250.
- KIMURA M. and WEISS G. H., « The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance », *Genetics*, vol. 49, 1964, 461-576.
- LAMOTTE Maxime, « Recherches sur la structure génétique des populations naturelles de *Cepaea nemoralis*, » *Supplément au Bulletin Biologique de France et de Belgique*, Supplément 35, 1951.
- MALECOT G. — « Les mathématiques de l'hérédité », Masson et Cie, Paris, 1948, in « *Probabilités et Hérédité* », Cahier n° 47, *Travaux et Documents I.N.E.D.*, Presses Universitaires de France, 1966, 151-160; « Les processus stochastiques de la

- génétique», *Colloque international du C.N.R.S.*, **13**, 1949, 121-126; « Quelques schémas probabilistes sur la variabilité des populations naturelles », *Ann. Univ. Lyon. Sciences*, section A, **13**, 1950, 37-60; « Un traitement stochastique des problèmes linéaires (mutation, linkage, migration) en Génétique de Population », *Ann. Univ. Lyon, Sciences*, Section A, **14**, 1951, 79-117; « Les Processus stochastiques et la méthode des fonctions génératrices ou caractéristiques », *Publ. Inst. Statist. Paris*, 1, F. 3, 1952, 1-16; « Sur les modèles stochastiques linéaires asymptotiquement stationnaires », *Ann. Univ. Lyon. Sciences*, Section A, **17**, 1954, 19-35; « Modèles déterministes de la génétique de Population » in « Probabilités et Hérité », Cahier n° 47, *Travaux et documents I.N.E.D., Presses Universitaires de France*, 1966, 163-189; « The decrease of relationship with distance », *Cold Spring Harbor Symposium on Quantitative Biology*, **20**, 1966, 52-53; « Les Modèles stochastiques en génétique de population », *Publ. Inst. Statist. Paris*, 8, F3, 1959, 173-210, in *Probabilités et Hérité*, Cahier n° 47, *Travaux et Documents I.N.E.D., P.U.F.*, 1966, 151-160; « Sur l'estimation des taux de mutation et de sélection, compte tenu des migrations » *Bulletin de l'Institut International de Statistique*, **37**, 3, 1960, 1-11; « Migration et parenté génétique moyenne », Entretiens de Monaco en Sciences Humaines, 1962, 205-212, in *Probabilités et Hérité*, cahier n° 47. *Travaux et documents I.N.E.D., P.U.F.* 1966, 345-354.
- ROBERTS D. F. and HIORNS R. W., « The Dynamics of Racial Intermixture » *The American Journal of Human Genetics*, vol. **14**, 1962, 261-277.
- SALDANHA P. H., « Gene Flow from White into Negro Populations in Brazil », *The American Journal of Human Genetics*, vol. **9**, 1957, 299-309.
- SMITH Cedric, « Local Fluctuations in Gene Frequencies », *Annals of Human Genetics* vol. **32**, 1969, 251-260.
- WRIGHT Sewal, « Isolation by Distance », *Genetics*, **28**, 1943, 114-138; « Isolation by Distance under Diverse Systems of Mating », *Genetics*, **31**, 1946, 39-59; « The Genetical Structure of Populations », *Annals of Eugenics*, vol. **15**, 1951, 323-354.
-