



HAL
open science

Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques

G. Malécot

► **To cite this version:**

G. Malécot. Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques. Annales de l'ISUP, 1952, 1 (3), pp.1-16. hal-04094483

HAL Id: hal-04094483

<https://hal.science/hal-04094483>

Submitted on 11 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES PROCESSUS STOCHASTIQUES ET

LA METHODE DES FONCTIONS GENERATRICES OU CARACTERISTIQUES

par G. MALECOT

Professeur à la Faculté des Sciences

de LYON

On trouve en Physique et en Biologie bien des exemples de systèmes formés de corpuscules élémentaires et qui évoluent de façon non déterministe mais aléatoire, soit par suite d'un indéterminisme essentiel, soit peut-être seulement par suite d'une connaissance insuffisante des conditions internes ou des conditions ambiantes. Si l'on étudie l'évolution au cours du temps de la loi de probabilité des différents états d'un tel système, on se trouve en présence d'un problème de processus aléatoire ou stochastique.

I) - Les plus simples des processus stochastiques que l'on ait étudiés rentrent dans la catégorie des processus conservatifs. Ce sont les processus dans lesquels un nombre constant N de corpuscules, chacun de nature invariable, se déplacent dans un certain espace (qui sera par exemple, dans le cas de la physique, l'espace des phases). Bornons nous au cas où il s'agit d'un espace discret (comme l'espace des états de la mécanique quantique divisé en cellules dont la dimension est réglée par la constante de Planck h ; ou encore l'aire d'extension d'une espèce biologique, divisée en domaines panmictiques, à l'intérieur desquels tous les croisements sont également probables). Nous pouvons alors repérer par un ou plusieurs indices entiers la "position" ou "l'état" d'un corpuscule, d'un individu, d'une unité, dans cet espace; et introduire la probabilité, supposée infiniment petite, qu'un corpuscule occupant la position (ou l'état) A_i à l'époque t passe dans la position A_j à l'époque infiniment voisine $t + dt$; soit $l_{ij}(t) dt$ cette probabilité; elle ne dépendra pas des positions qu'a occupées le corpuscule aux époques antérieures à t si la définition de la "position A_i à l'époque t " est suffisamment précise pour englober tous les éléments de prévision que fournit le passé (par exemple, en mécanique corpusculaire, il faudra considérer la position dans l'espace des phases qui sera l'ensemble de la position au sens ordinaire et de la vitesse) ; on dira alors qu'on a affaire à un Processus de Markoff ; on s'efforce le plus possible, dans les diverses applications, de se ramener à de tels processus. Un cas plus particulier sera celui des "processus additifs", ou à "accroissements indépendants" dans lesquels les accroissements instantanés des coordonnées qui caractérisent la "position", sont indépendants (en probabilité) de la position déjà atteinte (comme dans la "promenade au hasard" suivant le schéma de Polya, ou encore dans la variation du gain total dans un jeu de hasard formé de parties successives indépendantes). Ces processus ont été

très étudiés depuis Laplace (voir en particulier les études de P. Levy sur le mouvement Brownien).

1°) Nous ne traiterons dans ce qui suit que de processus de Markoff, en supposant d'abord que le comportement de chaque corpuscule est indépendant des positions des autres corpuscules. (corpuscules indépendants). Si on multiplie la probabilité $P(n_1, \dots, n_i, \dots, t)$ pour que, à l'époque t , il y ait n_1 corpuscules dans l'état A_1 , etc, n_i corpuscules dans l'état A_i , etc, par $Z_1^{n_1} \dots Z_i^{n_i} \dots$ et qu'on fasse la somme pour tous les systèmes de valeurs des n_i , on obtient par définition la fonction génératrice des n_i à l'époque t soit $\Pi(Z_1, \dots, Z_i, \dots, t)$, qui est forcément pour un processus conservatif, puisque $\sum_i n_i = N$, une fonction homogène de degré N des Z_i .

Chacun des n_i corpuscules occupant l'état A_i aura, dans l'intervalle de temps Δt , la probabilité $l_{ij} \Delta t + o(\Delta t)$ de passer dans l'état A_j ; la fonction génératrice à l'époque $t + \Delta t$ s'obtiendra donc en remplaçant dans Π la variable Z_i par $Z_i + \sum_j (Z_j - Z_i) l_{ij} \Delta t + o(\Delta t)$, ce qui revient à augmenter Π de $\frac{\delta \Pi}{\delta Z_i} \sum_j (Z_j - Z_i) l_{ij} \Delta t + \dots + \frac{\delta \Pi}{\delta Z_i} \sum_j (Z_j - Z_i) l_{ij} \Delta t + \dots + o(\Delta t)$

Il en résulte immédiatement, en divisant par Δt et faisant tendre Δt vers zéro, que la dérivée partielle $\frac{\delta \Pi}{\delta Z_i}$ au temps de la fonction génératrice Π est :

$$(1) \quad \frac{\delta \Pi}{\delta t} = \sum_i \sum_j l_{ij} (Z_j - Z_i) \frac{\delta \Pi}{\delta Z_i} \quad (\text{Bartlett})^1$$

Le système différentiel associé à cette équation aux dérivées partielles peut se mettre sous la forme $-dt = \frac{dy_i}{\mu_i y_i} = \dots = \frac{d y_i}{\mu_i y_i} = \dots$ en désignant par μ_i et y_i les valeurs propres et les solutions propres de la substitution li-

néaire définie par les formules :

$$(2) \quad \begin{cases} Z_1 = \sum_j l_{1j} (Z_j - Z_1) \\ Z_i = \sum_j l_{ij} (Z_j - Z_i) \end{cases}$$

Ce qui donne les intégrales premières $y_i e^{\mu_i t} = \text{constante}$.

L'intégrale générale Π de (1) est donc une fonction forcément homogène et du degré N de $y_1 e^{\mu_1 t}, \dots, y_i e^{\mu_i t}, \dots$ dont la forme est déterminée par les conditions initiales. Si les corpuscules sont par exemple tous initialement dans l'état A_k , on a pour $t = 0$ $\Pi \equiv Z_k^N$, donc, puisque Z_k est une forme linéaire des Y_i , soit $Z_k = \sum_i a_{ki} y_i$, on a pour t quelconque

$$\Pi \equiv \left(\sum_i a_{ki} y_i e^{\mu_i t} \right)^N. \text{ Mais la matrice de la substitution (2) a ses}$$

éléments tous $\gg 0$, sauf ceux de la diagonale principale qui sont tels que la somme des éléments de chaque ligne soit nulle : il en résulte qu'il y a au moins

¹ J. Roy. Stat. Soc. B, 11, 1949, p.222

une valeur propre nulle et que, d'après le théorème de Tambs-Lyche, complété par Fréchet, les valeurs propres non nulles ont des parties réelles négatives.

Π tend donc alors, quand $t \rightarrow +\infty$, vers la puissance N^e d'une fonction linéaire des y_i , donc des Z_j , soit $\sum p_j Z_j$; la probabilité pour qu'il y ait des nombres donnés n_i des corpuscules dans les états A_i tend donc asymptotiquement vers

$$\frac{N!}{n_1! n_2!} p_1^{n_1} p_2^{n_2} \quad (\text{loi multinominale})$$

Remarque : Dans le cas plus général où les corpuscules seraient initialement dans des état divers, Π serait un produit de puissances de sommes $\sum_i a_{k_i} y_i e^{\mu_i t}$ correspondant à différentes valeurs de k . Mais, s'il y a une seule valeur propre nulle, soit μ_1 , Π tend toujours quand $t \rightarrow +\infty$ vers la puissance N^e de y_1 (à un facteur constant près) et la loi de la répartition des corpuscules entre les différents états A_i tend alors vers une Loi multinomiale indépendante de la répartition initiale (processus de Markoff réguliers au sens de Fréchet)

2°) Un schéma de Markoff un peu moins simple correspondra au cas où les mouvements des différents corpuscules ne sont plus stochastiquement indépendants, et où des interactions sont possibles entre eux (chocs de la mécanique statistique classique, interactions de la mécanique quantique). Si l'on ne tient compte que des interactions des corpuscules 2 à 2 (les autres étant supposées négligeables comme trop improbables), on désignera par l_{rs} , uv dt la probabilité pour que, pendant l'intervalle $t \dots t + dt$, 2 corpuscules d'abord dans les états respectifs A_r et A_s agissent l'un sur l'autre et passent respectivement dans les états A_u et A_v (on peut d'ailleurs englober si l'on veut dans ce cas le cas précédemment traité d'évolution indépendante en remarquant que ce qui était précédemment appelé l_{ij} fournira maintenant la probabilité pour que, le 2e corpuscule restant par exemple dans l'état A_s , le 1^{er} évolue indépendamment de l'état A_i à l'état A_j , par suite l_{ij} devra maintenant être noté l_{is} , js , avec deux indices égaux).

La considération du changement de la fonction génératrice relatif à une durée Δt que l'on fait tendre vers zéro conduit maintenant, au lieu de l'équation (1), à l'équation :

$$(3) \quad \frac{\delta \pi}{\delta t} = \sum_{r s u v} l_{r s u v} (Z_u Z_v - Z_r Z_s) \frac{\delta^2 \pi}{\delta Z_r \delta Z_s}$$

que Moyal étudie dans le cas où s'applique le "principe de microreversibilité" d'après lequel $l_{r s, u v} = l_{u v, r s}$ (c'est la transposition en mécanique quantique de l'hypothèse des chocs parfaitement élastiques) : il remarque qu'alors l'équation (3) est vérifiée par toutes les solutions de l'équation

$$\frac{\delta^2 \pi}{\delta Z_r \delta Z_s} = \frac{\delta^2 \pi}{\delta Z_r \delta Z_v}$$

à partir de quoi Moyal retrouve la loi de distribution des molécules de Maxwell Boltzman, classique en théorie cinétique des gaz. Et moyennant des modifications de l'équation aux dérivées partielles pour l'adapter au cas de corpuscules indiscernables à fonction d'onde symétrique, où au cas de corpuscules à fonction d'onde antisymétrique obéissant au principe d'exclusion de Pauli, il retrouve les statistiques de Bose-Einstein et de Fermi-Dirac. =

II) Passons maintenant aux plus simples des processus multiplicatifs

1°) Imaginons d'abord que chaque corpuscule ait, pendant l'intervalle $t \dots t + \Delta t$ et indépendamment de sa position ainsi que du nombre d'autres corpuscules présents une probabilité $l \Delta t + o(\Delta t)$, de se fragmenter au deux corpuscules identiques; si les corpuscules initiaux sont tous identiques entre eux, le processus se traduira simplement par une multiplication aléatoire de leur nombre au cours du temps, multiplication due aux "naissances" par bipartition. C'est le processus dit de Furry. Si on suppose que pour tous les corpuscules existant à chaque instant la probabilité de bipartition $l \Delta t$ est une constante indépendante de l'ancienneté de la "naissance" de chacun, il suffit de considérer la fonction génératrice du nombre n de corpuscules à l'époque t , soit :

$$\pi(Z, t) = \sum_n p_n(t) Z^n$$

pour remarquer que la fonction $\pi(Z, t + \Delta t)$ s'en déduit en y remplaçant Z par $Z + l(Z^2 - Z)\Delta t + o(\Delta t)$, on a donc, en divisant par Δt et en faisant tendre Δt vers zéro :

$$(4) \quad \frac{\delta \pi}{\delta t} = l(Z^2 - Z) \frac{\delta \pi}{\delta Z}$$

qui donne, en supposant qu'initialement $\pi(Z, 0) \equiv Z$ (un seul corpuscule) :

$$\pi(Z, t) \equiv \frac{Z e^{-lt}}{1 - Z(1 - e^{-lt})}$$

Les $P_n(t)$ sont donc, pour t fixé, les termes d'une série géométrique.

Ce processus admet une extension simple due à Feller (processus de "naissance et décès") en supposant qu'indépendamment de la probabilité de bipartition $l \Delta t$, chaque corpuscule admet, dans l'intervalle $t \dots t + \Delta t$, une probabilité de disparition, de "décès" égale à $\mu \Delta t + o(\Delta t)$

On a alors :

$$(4') \quad \frac{\delta \pi}{\delta t} = [l(Z^2 - Z) + \mu(1 - Z)] \frac{\delta \pi}{\delta Z}$$

qui donne encore, si l'on part d'un seul corpuscule initiale, pour les $p_n(t)$ (pour t fixé et quand $l \neq \mu$) une série géométrique dont la raison qui est

$$\frac{1 - e^{-(l-\mu)t}}{1 - e^{-(l+\mu)t}}$$

croît ou décroît quand t croît suivant que $l > \mu$ où $l < \mu$, d'où l'on peut déduire que, quand $l > \mu$ il y a une probabilité supérieure à zéro pour que le nombre de corpuscules tende vers $+\infty$ avec t , alors que, quand $l < \mu$ (et aussi quand $l = \mu$), il y a presque certitude d'extinction totale.

Si on part initialement de n_0 corpuscules, il faut élever la fonction π précédente à la puissance n_0 et on obtient (Kendall 2) pour les $p_n(t)$ la série de Pascal ou "négative binomial" (qui représente en général très bien la distribution

1 ibid., p. 207

2 ibid., p. 237

des nombres d'individus de différentes espèces animales capturés en un temps déterminé sur un espace donné).

2°) Le processus introduit par Arley comme modèle mathématique des gerbes de rayons cosmiques secondaires suppose deux catégories différentes de corpuscules, photons ou électrons, chaque corpuscule d'une catégorie étant susceptible, dans l'intervalle $t \dots t + \Delta t$, de se transformer, par choc sur la matière, avec probabilité $1 \Delta t + o(\Delta t)$, en deux corpuscules (paire d'électrons ou électron + photon); et aussi de "décéder", c'est-à-dire de perdre son énergie, avec la probabilité $\mu \Delta t$; ce qui donne pour la fonction génératrice $\pi(Z, w, t) = \sum_{m,n} p_{m,n}(t) Z^m w^n$ des nombres m et n de corpuscules des deux catégories à la date t l'équation :

$$(4'') \quad \frac{\delta \pi}{\delta t} = \left[1(Z^2 - w) + \mu(1 - w) \right] \frac{\delta \pi}{\delta w} + \left[1(wZ - Z) + \mu(1 - Z) \right] \frac{\delta \pi}{\delta Z}$$

équation qui n'a été résolue qu'approximativement par Arley et Bartlett. ¹

3°) On obtient un autre type de processus multiplicatif, important en démographie et en biologie, en considérant une population où sont séparées les générations successives F_0, F_1, \dots les unités constituant chaque génération étant, suivant le cas, soit des individus, soit des gamètes, soit des gènes mendéliens.

Si on suppose que chaque unité a des probabilités déterminées q_s d'engendrer s unités de la génération suivant, et cela indépendamment de ce qui se passe pour les autres unités, on voit aisément que la fonction génératrice $\pi_r(Z) =$

$\sum_n p_n(r) Z^n$ du nombre n d'unités de la r^e génération F_r qui sont issues d'une ou plusieurs unités déterminées de la génération initiale F_0 satisfait à l'équation de récurrence :

$$(5) \quad \pi_{r+1}(Z) = \pi_r \left(\sum_s q_s Z^s \right)$$

[puisque $p_m(r+1)$ est le coefficient de Z^m dans $\sum_n p_n(r) \left(\sum_s q_s Z^s \right)^n$]

Cette équation avait été donnée par Watson, Erlang, et Steffensen pour le problème de l'extinction du hasard des noms de famille; R.A. Fisher et Haldane l'ont utilisée dans le problème de l'extinction de mutations.

De cette équation on peut par passage à la limite déduire une équation "à temps continu" comme celles qui ont été indiquées précédemment. Il suffit de considérer les individus engendrés non plus pendant une durée finie, prise pour unité de temps, mais pendant un temps Δt , et de supposer que la fonction génératrice du nombre total d'unités issues pendant le temps $t \dots t + \Delta t$ de chaque unité existante à la date t est de la forme :

$$\sum_s q_s Z^s \equiv Z + h(Z, t) \Delta t + o(\Delta t)$$

(le 1er terme correspondant à la permanence très probable de l'unité engendrante, et le 2ème terme pouvant comprendre comme dans le cas particulier du 1°), un terme de natalité et un terme de mortalité).

Le 2ème membre de (5) devient alors, après substitution de cette expression :

$$\pi_r(Z) + \frac{\delta \pi_r}{\delta Z} h(Z, t) \Delta t + o(\Delta t)$$

¹ ibid., p. 221

ce qui montre que la fonction génératrice $\pi(Z, t)$ du processus "à temps continu" vérifie l'équation aux dérivées partielles :

$$(6) \quad \frac{\delta \pi}{\delta t} = h(Z, t) \frac{\delta \pi}{\delta Z} \quad (\text{Barlett}) \underline{1}$$

équation qui comprend comme cas particulier l'équation de Feller (4') et qui se résout très aisément dans le cas particulier d'un processus "homogène dans le temps" où $h(Z, t)$ est indépendant de t . On trouve alors, si n_0 est le nombre d'individus de la génération initiale dont on étudie les descendants ($\pi(Z, 0) \equiv Z^{n_0}$) :

$$\pi(Z, t) \equiv Z_1^{n_0} \quad \text{avec } t \int_Z Z_1 \frac{dU}{n(u)}$$

ce qui redonne en particulier tous les résultats du 1°) 1

III) Franchissons maintenant un pas de plus vers l'adaptation des modèles stochastiques au réel en introduisant des processus avec multiplication et limitation.

1°) On sait que le nombre de descendants laissés par un individu d'une population n'est en général (contrairement à ce que nous avons supposé précédemment au (II)) indépendant de l'effectif total : la fécondité en descendants adultes d'un animal ou d'un végétal est beaucoup plus grande quand il s'agit du "peuplement d'une place vide" que quand la population est devenue nombreuse. Après un phase initiale de "multiplication pure" où l'effectif de la population pourrait être regardé comme obéissant à un processus du type (II), intervient une phase de "limitation" où le nombre d'individus déjà existants restreint à chaque instant la multiplication des autres. On peut certes en tenir compte, dans l'équation (4') par exemple, en supposant que la natalité représentée par l diminue ou que la mortalité représentée par μ augmente avec l'effectif N de la population. C'est ce qu'a fait Feller (acta biotheorica, 5, 1939) en prenant pour l une fonction linéaire (décroissante) de N et pour μ une constante, c'est-à-dire en faisant l'hypothèse qui, dans un schéma déterministe et non plus stochastique de l'évolution de l'effectif, conduit à la loi dite "logistique" de Verhulst-Pearl. Comme dans le schéma de Feller N est devenu aléatoire, l'équation (4') doit être elle-même modifiée pour en tenir compte, mais Feller a pu résoudre l'équation ainsi modifiée. (L'étude détaillée reste par contre à faire en suivant quelques aperçus de Kendall 2 dans le cas où μ , au lieu d'être une constante, serait lui aussi une fonction linéaire-croissante de N).

2°) Mais le problème ainsi posé est celui de la fluctuation du nombre total d'individus entre deux bornes correspondant respectivement à un taux de mortalité nul et à un taux de natalité nul. Un problème différent se posera si l'on distingue les individus de la population suivant leur nature (héréditaire), par exemple suivant les gènes mandéliens qu'ils portent. Il y aura alors, outre les fluctuations de l'effectif total, des fluctuations dans les proportions des différents gènes représentés, et ces fluctuations pourront être regardées comme stochastiquement indépendantes de celles de l'effectif dans le cas où la reproduction de chaque individu n'est pas influencée par les gènes dont il est porteur, c'est-à-dire dans le cas où il n'y a pas de sélection des gènes. Je me suis autorisé de cette remarque pour séparer le problème de la proportion des gènes du problème de l'effectif total. Si l'on désigne par $N(t)$ l'effectif total (constant, fonctionnellement variable, ou aléatoire) à la date t , le nombre total de "loci" c'est-à-dire d'emplacements occupables par un gène déterminé est (dans le cas usuel où les individus sont "diploïdes" c'est-à-dire porteurs de deux chromosomes de chaque paire) $2 N(t)$. Le nombre de ceux qui sont occupés effectivement par un gène déterminé que nous désignerons par E est l'entier aléatoire $r(t)$ (les autres étant occupés par les gènes "allèles" de E que nous grouperons sous la désignation e).

(1) ibid., p. 220 - (2) ibid., p. 245.

Il est commode d'introduire le quotient $q(t) = \frac{r(t)}{r \cdot N(t)}$ qui sera appelé la fréquence (aléatoire) du gène E à la date t. Cette fréquence vérifie, dans certains cas, une équation différentielle stochastique qui rentre dans la catégorie de celles qui ont été étudiées par S. Bernstein, mais qu'il est préférable, (pour ne pas avoir à distinguer entre la distribution discontinue de la fréquence et une approximation continue de celle-ci) de remplacer par l'équation aux dérivées partielles que vérifie la fonction caractéristique de q(t), soit $F(s,t) = \mathbb{m}(e^{sq(t)}) \stackrel{1}{=}$

Il est à remarquer que cette fonction caractéristique se ramène immédiatement, quand N(t) n'est pas aléatoire, à la fonction génératrice G(Z,t) de r(t), en posant

$$\frac{s}{c \cdot P N} = Z \quad \text{ce qui donne} \quad G(Z, t) \equiv F(2 N \log Z, t)$$

Cette équation aux dérivées partielles peut se déterminer directement (Bartlett, ibid. p. 220; G. Malecot "Les Mathématiques de l'hérédité", p.45) en partant de la loi de probabilité conditionnée de q(t + Δt) quand on connaît la valeur de q(t) ainsi bien entendu que celle de N(t + t) regardée d'ailleurs comme égale, avec une erreur relative négligeable, à N(t). Il est commode de définir cette loi par les moments conditionnés de q(t + Δt) ou, ce qui revient au même, de Δq = q(t + Δt) - q(t), soit :

$$\mathbb{m}_c(\Delta q) = A(q, N, t) \Delta t + o(\Delta t)$$

$$\mathbb{m}_c(\Delta^2 q) = B^2(q, N, t) \Delta t + o(\Delta t) \quad \mathbb{m}_c(\Delta^3 q) = D^3 \Delta t + o(\Delta t)$$

On en déduit que, d'après la règle des moyennes conditionnées et pour s fini quelconque :

$$\begin{aligned} F(s, t + \Delta t) &= \mathbb{m}[e^{sq(t + \Delta t)}] = \mathbb{m}[e^{sq(t)} \mathbb{m}_c(e^{s \Delta q})] = \\ &= \mathbb{m}[e^{sq(t)} \mathbb{m}_c(1 + s \Delta q + \frac{s^2 \Delta q^2}{2!} + \frac{s^3 \Delta q^3}{3!} + \dots)] \\ &= \mathbb{m}[e^{sq(t)} (1 + s A \Delta t + \frac{s^2}{2!} B^2 \Delta t + \frac{s^3}{3!} D^3 \Delta t + \text{etc} + o(\Delta t))] \\ &= F(s, t) + [s \mathbb{m}(A e^{sq(t)}) + \frac{s^2}{2} \mathbb{m}(B^2 e^{sq(t)}) + \\ &\quad \frac{s^3}{3!} \mathbb{m}(D^3 e^{sq(t)}) + \dots] \Delta t + o(\Delta t) \end{aligned}$$

La fonction caractéristique F(s,t) vérifie donc une équation aux dérivées partielles qui peut s'écrire symboliquement, en posant

$$\mathbb{m}[\varphi(q) e^{sq}] = \mathbb{m}[\varphi(\frac{\delta}{\delta s}) e^{sq}] = \varphi(\frac{\delta}{\delta s}) \mathbb{m}(e^{sq}) = \varphi(\frac{\delta}{\delta s}) F \stackrel{2}{=} :$$

$$(7) \quad \frac{\delta F}{\delta t} = s A(\frac{\delta}{\delta s}, N, t) F + \frac{s^2}{2} B^2(\frac{\delta}{\delta s}, N, t) F + \frac{s^3}{3!} D^3(\frac{\delta}{\delta s}, N, t) + \dots \stackrel{3}{=}$$

(1) Il est inutile ici de prendre comme d'habitude l'exposant isq, car aucune question de convergence ne se pose, q variant entre 0 et 1.

(2) F est indéfiniment dérivable % à s, puisque q, reste dans l'intervalle 0...1. Et on a évidemment

$$(\frac{\delta}{\delta s}) F = \mathbb{m}(q e^{sq}), \quad (\frac{\delta}{\delta s})^m F = \mathbb{m}(q^m e^{sq})$$

(3) Cette équation a été donnée par Bartlett ibid. p. 220.

ou, si l'on suppose $N(t)$ non aléatoire et si l'on introduit la fonction génératrice de l'entier r , soit $G(Z,t) \equiv F(2N \log Z, t)$, et en tenant compte de ce que

$$\frac{\delta F}{\delta s} \equiv \frac{Z}{2N} \frac{\delta G}{\delta Z}$$

$$(7') \quad \frac{\delta G}{\delta t} = 2N \log Z A \left(\frac{Z}{2N} \frac{\delta}{\delta Z}, N, t \right) G + \frac{(2N \log Z)^2}{2} B^2 \left(\frac{Z}{2N} \frac{\delta}{\delta Z}, N, t \right) G + \dots$$

Les équations (7) et (7') ne seraient du 1er ordre (comme l'étaient les équations (1) et (6) des processus conservatifs et multiplicatifs à corpuscules indépendants) que

si A, B^2, D^3 , etc, étaient des fonctions linéaires de q (on ne peut d'ailleurs faire rentrer dans (7') les équations des processus précédents, puisque le passage de (7) à (7') suppose que $N(t)$ n'est pas aléatoire).

Mais, dans la plupart des applications A et B^2 sont % à q des polynomes de degré supérieur à 1, (7) et (7') sont alors même quand D^3 est négligeable comme souvent le cas des équations d'ordre supérieur à 1, $\frac{1}{2}$. Nous le constatons sur quelques cas particuliers :

a) si tous les gènes ont égale probabilité de laisser des descendants et si ces descendants sont de même nature qu'eux (pas de "mutations"), on a

$$m_c [q(t+\Delta t)] = q(t) \text{ donc } A = 0, \text{ mais } B^2 \Delta t, \text{ qui est alors la variance}$$

conditionnée de $q(t+\Delta t)$, ne peut être nul, car $q(t+\Delta t)$ fluctue autour de sa moyenne $q(t)$ du fait que les individus qui décèdent et ceux qui se reproduisent pendant l'intervalle $t - t + \Delta t$ sont regardés comme choisis au hasard, $\frac{2}{2}$ ainsi que les gènes qu'ils portent. Si le nombre moyen de ces individus est supposé proportionnel à l'effectif $N(t)$ et à Δt , donc de la forme $KN \Delta t$, le nombre total de gènes qu'ils portent sera le double, $2KN \Delta t$, et la variante du nombre de gènes E qu'ils portent sera, d'après la formule de Bernoulli,

$q(t) [1-q(t)] 2KN \Delta t$; par suite la variance de la proportion de ces gènes dans la population totale, c'est-à-dire aussi la variance conditionnée $B^2 \Delta t$ de $q(t+\Delta t)$ que nous cherchons, sera le quotient par $(2N)^2$ du nombre précédent d'où, (

$$(8) \quad B^2(q, N, t) \equiv K \frac{(1-q)}{2N}$$

(il appartient à la zoologie ou à la démographie de fournir, dans chaque cas, une évaluation numérique raisonnable de la valeur de la constante K).

L'équation (7) s'écrit alors :

$$(9) \quad \frac{\delta F}{\delta t} = \frac{K s^2}{4N} \left(\frac{\delta F}{\delta s} - \frac{\delta^2 F}{\delta s^2} \right) + \frac{K s^3}{3!} \quad 0 \left(\frac{1}{N^2} \right)$$

$$\text{car } D^3 \text{ est de la forme } \frac{K}{(2N)^2} q(1-q)(1-2q)$$

b) on étend aisément la formule précédente à un schéma de sélection du gène E (toujours sans mutation), à condition de supposer que la sélection ne dépend que des proportions relatives q et $1-q$ des gènes E et e , et non de leur nombre total $2N$. Suivant la schéma adopté usuellement par les généticiens, on admettra que ,

¹ G. Malecot, "les Mathématiques de l'hérédité", p.47, équation (3)

² Ce choix étant souvent plus concentré sur certains âges, une étude plus poussée devrait prendre en considération les fluctuations des proportions de gènes entre les différents "groupes d'âge", pour voir dans quels cas ces fluctuations peuvent avoir un effet important.

par l'effet de la sélection, le rapport $\frac{q}{1-q}$ de ces proportions augmente en moyenne, pendant le temps Δt , de $\sigma \Delta t$, σ étant une constante (ou plus généralement une fonction donnée du temps)¹. Cette hypothèse fournit alors :

$$m_c [q(t + \Delta t)] = q(t) + \sigma q(t) [1 - q(t)] \Delta t + o(\Delta t)$$

d'où (10) $A(q, N, t) \equiv \sigma q(1 - q)$

et, si l'on conserve pour B^2 la formule (8) (le terme correctif qu'y introduirait la sélection serait en général négligeable, parce que de l'ordre de Δt), on aura :

$$(11) \quad \frac{\partial F}{\partial t} \equiv \left(\sigma s + \frac{K s^2}{4 N} \right) \left(\frac{\partial F}{\partial s} - \frac{\partial^2 F}{\partial s^2} \right) + \frac{K s^3}{3!} \quad 0 \left(\frac{1}{N^2} \right)$$

Cette équation parabolique peut s'étudier facilement si σ , N et K sont constants, et si on néglige les terme en $1/N^2$. En effet elle montre que les valeurs $F(0, t)$ et $F(-\frac{4 N \sigma}{K}, t)$ sont indépendantes de t donc égales à leurs valeurs F_0 et F_1 à l'instant initial $t=0$, valeurs qui se calculent à partir de la fréquence initiale q_0 du gène E ; puisque $F(s, 0) = M(e^{sq_0}) = e^{sq_0}$, on a :

$$F_0 = 1 \quad F_1 = e^{-\frac{4 N \sigma q_0}{K}}$$

En outre l'équation (11) admet des solutions stationnaires (c'est-à-dire indépendantes du temps, donc conduisant à un processus stationnaire) données par :

$$\frac{\partial F}{\partial s} - \frac{\partial^2 F}{\partial s^2} = 0 \text{ c'est-à-dire } F = \alpha e^s + \beta ;$$

il y en a une, \bar{F} , qui prend pour $s = 0$ et $s = s_1 = -4 N \sigma / K$ les mêmes valeurs F_0 et F_1 que la solution $F(s, t)$ qui correspond à la donnée initiale q_0 : c'est celle qui correspond à α et β donnés par

$$\alpha + \beta = F_0 = 1$$

$$\alpha e^{-\frac{4 N \sigma}{K}} + \beta = F_1 = e^{-\frac{4 N \sigma q_0}{K}}$$

$$\text{on a donc } \bar{F} \equiv \frac{1 - \exp(-\frac{4 N \sigma q_0}{K})}{1 - \exp(-\frac{4 N \sigma}{K})} (e^s - 1) + 1$$

Si l'on désigne par F_r l'excès sur \bar{F} de la solution cherchée $F(s, t)$, F_r sera une solution de (11) qui s'annule à la fois pour $s = 0$ et pour $s = s_1$; or toutes les solutions de (11) qui jouissent de cette propriété sont des séries de Sturm-Liouville de la forme

$$S = \sum_{i=1}^{i=\infty} a_i f_i(s) g_i(t)$$

chaque produit $f_i g_i$ vérifiant l'équation (11) avec :

$$\frac{g_i'}{g_i} \equiv -l_i \equiv \left(\sigma s + \frac{K s^2}{4 N} \right) \frac{f_i' - f_i''}{f_i}$$

et f_i étant nul pour $s = 0$ et $s = s_1$, ce qui ne peut être réalisé que pour une in-

¹ Cela revient à admettre que les taux de multiplication de 2 gènes sont dans un rapport constant voisin de 1.

finité dénombrable (au plus) de constante l_i qui sont les valeurs propres relatives à l'intervalle $0 \dots s_1$ de l'équation différentielle :

$$(K/4N) s (s - s_1) (f''_i - f'_i) - l_i f_i = 0.$$

Or cette dernière équation peut se mettre de la façon habituelle sous la forme de Sturm :

$$\frac{d}{ds} (e^{-s} f'_i) + \frac{e^{-s} l_i f_i}{(K/4N) s (s_1 - s_n)} = 0$$

et les théorèmes de Sturm montrent que, les coefficients de f_i et f'_i étant positifs, les valeurs propres l_i sont nécessairement toutes positives; donc toutes les séries

$$S = \sum_1^{\infty} a_i e^{-l_i t} f_i(s)$$

tendent vers 0 quand $t \rightarrow +\infty$ (avec une vitesse définie par l'inverse de la plus petite valeur propre). F_r tend donc certainement vers 0, par suite

$$F(s, t) \rightarrow \bar{F} \text{ quand } t \rightarrow +\infty.$$

Or \bar{F} est la fonction caractéristique d'une distribution de probabilité dans laquelle q est égal à 1 avec la probabilité

$$\omega = \frac{1 - \exp. \left(\frac{-4 N \sigma q_0 / K}{-4 N \sigma / K} \right)}{1 - \exp. \left(\frac{-4 N \sigma / K}{-4 N \sigma / K} \right)}$$

et à 0 avec la probabilité complémentaire $1 - \omega$, ce qui signifie que asymptotiquement, quand $t \rightarrow +\infty$, la population devient génétiquement homogène, avec extension à tous les individus du gène E (avec probabilité ω) ou d'un allèle e (avec probabilité $1 - \omega$). C'est l'effet Hagedoorn d'extinction des gènes au hasard 1 (on peut développer ω sous la forme

$$\omega = q_0 + 2 N \sigma q_0 (1 - q_0) K + \dots ;$$

ω est égal à q_0 quand il n'y a pas de sélection et supérieur à q_0 quand σ positif.

c) Il en va tout autrement s'il y a des mutations ayant pour effet de transformer, pendant chaque intervalle Δt , une proportion $u \Delta t$ des gènes E en allèles e, et de faire apparaître le gène E par mutation d'une proportion $v \Delta t$ des gènes e. On aura alors, s'il n'y a pas de sélection 2

$$m_c [q(t + \Delta t)] = q(t) + u q(t) \Delta t + v [1 - q(t)] \Delta t$$

$$\text{d'où (10')} \quad A(q, N, t) = -u q + v (1 - q) \quad \underline{\underline{3}}$$

que nous écrirons pour simplifier

$$(10'') \quad A = -k(q - c) \text{ avec } k = u + v \text{ et } c = v/u + v.$$

Si l'on conserve pour B^2 la même expression (8) qu'avec a) et b), on aura:

$$(12) \quad \frac{\delta F}{\delta t} = -s k \left(\frac{\delta F}{\delta s} - \bar{q} F \right) + \frac{K s^2}{4 N} \left(\frac{\delta F}{\delta s} - \frac{\delta^2 F}{\delta s^2} \right) + \frac{K s^3}{3!} \frac{1}{N^2}$$

équation qui n'est pas commode à résoudre correctement, mais qui donne aisément les coefficients succesifs du développement de $F(s, t)$ en série entière de puissances de s , c'est-à-dire les moments $M_p(t)$ de la loi de probabilité de q .

$$\text{Il suffit de poser } F(s, t) = \sum_{p=0}^{\infty} \frac{M_p(t) s^p}{p!}$$

1 effet déjà signalé par de Candolle pour l'extinction des noms de famille, qui s'héritent comme des gènes portés par le chromosome mâle Y.

2 s'il y avait en outre sélection, il faudrait prendre pour A la somme des expressions (10) et (12), à l'ordre de Δt près.

la convergence uniforme et celle des séries dérivées étant assurée du fait que, q restant dans l'intervalle $0 \dots 1$, tous les $M_p(t)$ sont aussi dans cet intervalle. (12) donne alors :

$$(13) \quad \frac{d M_p}{dt} = -k p (M_p - c M_{p-1}) + \frac{p(p-1)}{4N} K (M_{p-1} - M_p) + O\left(\frac{1}{N^2}\right) \stackrel{1}{\approx}$$

et en particulier, puisque $M_0 = 1$:

$$(13_1) \quad \frac{d M_1}{dt} = -k (M_1 - c) \stackrel{2}{\approx}$$

La valeur moyenne M_1 de q tend donc, quand $t \rightarrow +\infty$, vers la limite c ; sa différence avec sa limite est de l'ordre de e^{-kt} , elle ne tend donc que très lentement vers zéro si $k = u+v$, somme des "taux de mutation", est faible; la valeur moyenne varie peu dans tout intervalle de temps qui est faible par rapport à $1/k$.

On a aussi :

$$(13_2) \quad \frac{d M_2}{dt} = -2k (M_2 - c M_1) + \frac{K}{2N} (M_1 - M_2) \stackrel{2}{\approx}$$

ce qui donne pour la variance $V = M_2 - M_1^2$ de q :

$$\frac{d V}{dt} = - (2k + K/2N) V + (K/2N) (M_1 - M_1^2)$$

et cette formule montre que, dans un groupe peu nombreux où $K/2N$ supposé constant est très grand % à k , la variance V se stabilise beaucoup plus rapidement que la

moyenne M_1 : l'écart entre V et la valeur $V_S = \frac{M_1 - M_1^2}{1+4Nk/K}$ est en effet de l'ordre

de $\exp. [-(2k + K/2N)t]$; la valeur V_S subit elle-même, aussi longtemps que M_1 n'est pas stabilisé, une variation, mais cette variation est beaucoup plus lente, puisqu'elle est réglée par $\exp. [-Kt]$

Les autres moments pourraient faire l'objet de calculs analogues et on démontrerait que, quand $t \rightarrow +\infty$ et si N et K sont constants, ils tendent bien, avec des vitesses réglées au moins par $\exp. [-kt] \stackrel{3}{\approx}$, vers des valeurs limites qui sont solutions des équations (13) avec leurs membres nuls, et qui sont donc liées par la formule de récurrence suivante (dans laquelle nous remplaçons k et c en fonction de U et v) :

$$M_p = M_{p-1} \frac{p-1 + 4Nv/K + O(1/N)}{p-1 + 4N(u+v)/K + O(1/N)}$$

Les quantités ainsi définies ne sont autres, si l'on néglige les termes $O(1/N)$, que les moments de la "distribution β de Pearson", distribution absolument continue dont la probabilité élémentaire est :

$$(14) \quad \varphi(q) dq \approx cte \times q^{4Nv/K-1} (1-q)^{4Nu/K-1} dq \quad (\text{S. Wright } \frac{4}{5}).$$

J'ai montré $\stackrel{5}{\approx}$ qu'on peut suivre toujours avec la même approximation l'évolution au cours du temps de la distribution de probabilité de q , supposée absolument continue,

¹ s'il y avait une sélection, cette équation ferait en outre intervenir M_{p+1} , et la résolution par récurrence serait impossible.

² on peut montrer que les termes de l'ordre de $1/N^2$, dans les équations (13₁) et (13₂) sont toujours nuls.

³ Cf "les mathématiques de l'hérédité", p.54. $\stackrel{5}{\approx}$ "Les mathématiques de l'hérédité" p.48-51

⁴ Genetics 16, 1931; Statistical Genetics, actualités scient. et ind., 802, 1939; la formule (14) y est déterminée seulement par une condition de stationarité du 2^e ordre.

et obtenir ainsi, par une autre voie, la distribution limite (14), en déduisant de l'équation (12) [en remplaçant $\frac{\delta}{\delta s}$ par q puis s par $-\frac{\delta}{\delta q}$, en vertu des propriétés de la transformation de Laplace] l'équation aux dérivées partielles de Fokker-Planck-Kolmogoroff que vérifie la densité de probabilité quand elle existe et a des dérivées nulles aux bornes. Mais les Méthodes basées sur la fonction caractéristique ne supposent pas l'existence d'une densité de probabilité, et elles s'appliquent, comme nous l'avons vu au b), à des cas qui échappent à la méthode de Kolmogoroff. Elles permettraient d'ailleurs dans le cas présent, de corriger la formule (14) en tenant compte des termes négligés en $1/N, 1/N^2$, etc.

L'équation (14) montre qu'asymptotiquement la densité de probabilité de la fréquence q du gène E est répartie, dans le cas où u et v sont $< K / 4N$, suivant une "courbe en U" qui fournira une probabilité appréciable pour que q soit voisin de 0 ou 1, c'est-à-dire qu'un des gènes soit près de s'éteindre (extinction qui n'est incomplète qu'en raison de la réintroduction par mutation). Dans le cas contraire où u et v sont $> K / 4N$, la densité est répartie suivant une "courbe en cloche" d'autant plus serrée autour de sa moyenne que la variance V_s est petite c'est-à-dire que $4N(u+v) / K$ est grand.

IV) Une extension notable des processus à limitation et mutation étudiés au paragraphe III) sera obtenue en les combinant avec un processus de déplacement au hasard analogue à ceux étudiés au paragraphe I).

Supposons que les unités considérées (individus, ou gènes) soient réparties à la date t entre p groupes distincts A_1, \dots, A_p , et que quand t varie certaines unités se déplacent d'un groupe à un autre. Admettons que, pendant l'intervalle de temps $t \dots t + \Delta t$, chaque individu du groupe A_i ait, indépendamment des autres et indépendamment de sa constitution génétique, λ_{ij} la probabilité $\lambda_{ij} \Delta t + o(\Delta t)$ de passer dans le groupe A_j ; si l'on désigne par N_j le nombre d'individus dans le groupe A_j à la date t , la proportion moyenne "d'immigrants" arrivés dans ce groupe, pendant le temps Δt , depuis le groupe A_i , sera

$$\frac{N_i \lambda_{ij} \Delta t}{N_j} + o(\Delta t),$$

que nous représenterons dorénavant par $k_{ij} \Delta t + o(\Delta t)$, les k_{ij} étant appelés les "taux d'immigration".

Si l'on désigne par $q_j(t)$ la proportion du gène E dans le groupe A_j à la date t , l'immigration précédente aura pour effet de changer cette proportion en une autre dont la valeur moyenne est : $q_j(t) + \sum_i k_{ij} [q_i(t) - q_j(t)] \Delta t + o(\Delta t)$. A ce déplacement de la moyenne pourra s'en ajouter un autre dû à la mutation ou à la sélection (et éventuellement aux deux). Si nous désignons ce dernier [comme précédemment au paragraphe III, 2°), b) ou c), équation (10) au (10'')] par $A_j(q_j, N_j, t) \Delta t + o(\Delta t)$, ou plus brièvement par $A_j(q_j) \Delta t + o(\Delta t)$, on aura en désignant par C les conditions qui consistent en la donnée de

$$q_1(t), \dots, q_p(t) : m_c(\Delta q_j) = m_c [q_j(t + \Delta t)] - q_j(t) = A_j(q_j) \Delta t + \sum_i k_{ij}(q_i - q_j) \Delta t + o(\Delta t).$$

Si l'on désigne les 2^e et 3^e moments conditionnés par

¹ la discussion est la même dans le cas plus général où il y a sélection (ibid. p. 49) car la sélection complète seulement la formule (14) par un facteur exponentiel sans aucune discontinuité.

$$m_c (\Delta^2 q_j) = B_j^2 (q_j) \Delta t + o(\Delta t)$$

$$m_c (\Delta^3 q_j) = D_j^3 \Delta t + o(\Delta t)$$

les hypothèses faites au paragraphe III, 2°), a), montrent encore que $B_j^2 (q_j)$ est donné, à l'ordre de Δt près, par la formule :

$$(8) \quad B_j^2 (q_j) = \frac{k_j q_j (1-q_j)}{2 N_j}$$

et que D_j^3 n'est que de l'ordre de $1/N_j^2$.

En outre, si l'on admet l'indépendance stochastique des fluctuations aléatoires relatives à 2 groupes différents, on aura

$$m_c [\Delta q_i \Delta q_s] = m_c (\Delta q_i) \times m_c (\Delta q_j) = o(\Delta t) \text{ si } i \neq j, \text{ etc..}$$

Or il est facile d'exprimer, en fonction des moments conditionnés, l'équation aux dérivées partielles que vérifie la fonction caractéristique conjointe de $q_1(t), \dots, q_p(t)$. Si l'on désigne cette fonction par

$$F(s_1, \dots, s_p, t) = m[e^{s_1 q_1(t) + \dots + s_p q_p(t)}],$$

une extension immédiate des calculs faits dans le cas $p = 1$ donne

$$\begin{aligned} F(s_1, \dots, t + \Delta t) &= m \left[e^{\sum_j s_j q_j(t)} m_c e^{\sum_j s_j \Delta q_j(t)} \right] \\ &= m \left[e^{\sum_j s_j q_j(t)} \left\{ 1 + \sum_{ij} s_j l_{ij} (q_i - q_j) \Delta t + \sum_j s_j A_j (q_j) \Delta t + \sum_j \frac{s_j^2}{2!} B_j^2 (q_j) \Delta t + \right. \right. \\ &\quad \left. \left. + \sum_j \frac{s_j^3}{3!} D_j^3 (q_j) \Delta t + \dots + o(\Delta t) \right\} \right] \\ &= F(s_1 \dots t) + \sum_{ij} s_j l_{ij} \left(\frac{\delta F}{\delta s_i} - \frac{\delta F}{\delta s_j} \right) \Delta t + \sum_j s_j A_j \left(\frac{\delta F}{\delta s_j} \right) \Delta t + \sum_j \frac{s_j^2}{2!} B_j^2 \left(\frac{\delta F}{\delta s_j} \right) \Delta t \\ &\quad + \dots + o(\Delta t) \end{aligned}$$

d'où

$$(15) \quad \frac{\delta F}{\delta t} = \sum_{ij} s_j l_{ij} \left(\frac{\delta F}{\delta s_i} - \frac{\delta F}{\delta s_j} \right) + \sum_j s_j A_j \left(\frac{\delta F}{\delta s_j} \right) + \sum_j \frac{s_j^2}{2!} B_j^2 \left(\frac{\delta F}{\delta s_j} \right) + \dots$$

(dans le cas absolument continu où les q_j possèdent une densité de probabilité conjointe nulle sur la frontière, cette densité, admettant F pour transformée, de Laplace, obéit à l'équation, généralisation de celle de Kolmogoroff, qui se déduit de (15) en y remplaçant $\frac{\delta F}{\delta s_j}$ par q_j puis s_j par $-\frac{\delta}{\delta q_j}$).

Donc, en remplaçant les moments conditionnés par leurs valeurs et supposant D_j^3 , qui est de l'ordre de $1/N_j^2$, négligeable % à $B_j^2 (q_j)$ qui est de l'ordre de

1 (de la page 16) En supposant que les valeurs moyennes \bar{q}_j ; de tous les groupes considérés se sont stabilisées à une valeur commune que nous désignons par \bar{q} (cette valeur pourra n'être pas supposée égale à $\frac{c}{c+u}$ dans le cas où l'acheminement extrêmement lent vers cette dernière valeur sous l'influence des mutations ne serait pas achevé.

$1/N_j$, on aura, dans le cas de mutations à taux constant sans sélection où, suivant (10"), $A_j(q_j) = -k(q_j - c)$:

$$(16) \frac{\delta F}{\delta t} = \sum_{i,j} s_j l_{ij} \left(\frac{\delta F}{\delta s_i} - \frac{\delta F}{\delta s_j} \right) - k \sum_j s_j \left(\frac{\delta F}{\delta s_j} - cF \right) + \sum_j \frac{K_j}{4N_j} s_j^2 \left(\frac{\delta F}{\delta s_j} - \frac{\delta^2 F}{\delta s_j^2} \right) + \sum_j \frac{s_j^3}{3!} O\left(\frac{1}{N_j^2}\right)$$

Le développement de F en série entière toujours uniformément convergente ainsi que les séries dérivées) permet alors d'obtenir des équations différentielles récurrentes pour les moments.

Les moments d'ordre 1, $m(q_j)$, que nous désignerons par \bar{q}_j , sont les coefficients des s_j dans le développement de F , ou les termes constants dans les développements des $\frac{\delta F}{\delta s_j}$; d'où :

$$\frac{d\bar{q}_j}{dt} = \sum_i l_{ij} (\bar{q}_i - \bar{q}_j) - k (\bar{q}_j - c)$$

Dans le cas usuel où les taux d'immigration l_{ij} entre groupes voisins sont beaucoup plus élevés que la somme $k = u + v$ des taux de mutation, les moyennes $\bar{q}_j(t)$ relatives à des groupes voisins tendent d'abord à devenir égales entre elles, leur limite commune évoluant ensuite lentement pour tendre vers $c = v/(u+v)$ avec une vitesse réglée par e^{-kt} .

Plutôt que de faire le calcul des moments d'ordre supérieur, il nous sera plus commode de faire celui des cumulants, c'est-à-dire des coefficients du développement en série entière de $H = \log F$. On sait en effet que les cumulants du 2e ordre ne sont autres que les covariances $m_{ij} = m[(q_i - \bar{q}_i)(q_j - \bar{q}_j)]$

[ou la variances $m[(q_j - \bar{q}_j)^2]$ quand $i = j$]; et ils résulteront du développement de $H = \log F$ par la formule :

$$(17) H = \sum_j s_j \bar{q}_j + \frac{1}{2!} \left[\sum_i s_i^2 m_{jj} + 2 \sum_{(ij)} s_i s_j m_{ij} \right] + \dots$$

qui donne $\frac{\delta H}{\delta s_j} = \bar{q}_j + \sum_r s_r m_{rj}$

Or la transformation $F = e^H$ dans (16) montre que H vérifie l'équation :

$$(18) \frac{\delta H}{\delta t} = \sum_{i,j} s_j l_{ij} \left(\frac{\delta H}{\delta s_i} - \frac{\delta H}{\delta s_j} \right) - k \sum_j s_j \left(\frac{\delta H}{\delta s_j} - c \right) + \sum_j \frac{K_j}{4N_j} s_j^2 \left[\frac{\delta H}{\delta s_j} - \frac{\delta H}{\delta s_j} \right]^2 - \frac{\delta^2 H}{\delta s_j^2} + \sum_j s_j^3 O\left(\frac{1}{N_j^2}\right)$$

D'où, en substituant (17) dans (18) :

$$(19) \frac{d m_{ij}}{dt} = \sum_k (l_{kj} m_{ik} + l_{ki} m_{jk}) - m_{ij} \sum_k (l_{ki} + l_{kj}) - 2(u+v) m_{ij} + s_{ij} \frac{K_j}{2N_j} \left[\bar{q}_j - \bar{q}_j^2 - m_{ij} \right]$$

S_{ij} étant égal à 0 quand $i \neq j$ et à 1 quand $i = j$.

Si l'on suppose que les taux d'immigration sont constants au cours du temps ainsi que les N_j et les K_j , on voit que les m_{ij} sont donnés par un système linéaire à coefficients constants non homogène. La matrice formée par les coefficients des m_{hk} dans les $\frac{dm_{ij}}{dt}$ a les éléments de sa diagonale principale (c'est-à-dire

les coefficients des m_{ij}) égaux à

$$- \sum_{k \neq i} l_{ki} - \sum_{k \neq j} l_{kj} - 2(u+v) - S_{ij} \frac{K_j}{2 N_j},$$

alors que tous les autres éléments de chaque ligne sont positifs ou nuls, la somme de tous les éléments de chaque ligne étant toute fois toujours négative puisqu'égale à $-2(u+v) - \frac{S_{ij} K_j}{2 N_j}$; il résulte alors du théorème de Tambs-Lyche que les

racines de l'équation caractéristique du système rendu homogène ont toutes leurs parties réelles négatives, donc que la solution du système (19) satisfaisant à des conditions initiales quelconques tend quand $t \rightarrow +\infty$ vers une limite indépendante de l'état initial, les valeurs limites des m_{ij} étant définies par le système algébrique linéaire non homogène obtenu en remplaçant par zéro, dans la formule (19), les dérivées qui figurent au 1er membre (ce système a toujours une solution unique, car son déterminant ne peut être nul, sans quoi il existerait une racine caractéristique nulle).

Il résulte de cette étude que, si l'on considère une population fragmentée en "groupe panmictiques" ayant des effectifs constants et des coefficients de migration constants, les fréquences dans ces différents groupes d'une gène mendélien ne donnant pas prise à la sélection seront des fonctions aléatoires du temps dont les variances et covariances tendent asymptotiquement vers des valeurs stationnaires dont le calcul numérique peut se faire facilement par les formules de Cramer. Et on déduit immédiatement les valeurs asymptotiques de coefficients de corrélation entre les fréquences dans les différents groupes, coefficient dont les valeurs expérimentales dans une population donnée peuvent dans certains cas être déterminés par les biologistes ou les démographes, ce qui permettra en particulier, une fois évalués les taux de migration, de faire une estimation des taux de mutation.

Toutefois, dans le cas où le nombre p de groupes panmictiques est très grand, la détermination expérimentale de tous les taux de migration des groupes 2 à 2, quoique théoriquement nécessaire pour la construction numérique du schéma stochastique, deviendra pratiquement inextricable. On recourra alors à l'hypothèse simplificatrice de migration homogène entre groupes égaux. Je me bornerai à la formuler ici dans le cas d'une seule dimension d'espace 1: nous supposons alors que les p groupes sont d'effectifs égaux (soit N), répartis de façon équidistante sur une courbe fermée, et numérotés en parcourant cette courbe dans un sens déterminé, de sorte que chaque groupe sera repéré par un indice entier j défini seulement "modulo p "; le taux d'immigration l_{ij} sera supposé une fonction - périodique et de période p - de $i-j$ seulement; soit $l(i-j)$; la valeur $l(0)$ sera fixée arbitrairement. 2 La matrice des l_{ij} sera alors invariante si l'on effectue sur ses lignes et ses colonnes une même permutation circulaire; il en résulte que les valeurs asymptotiques des m_{ij} , définies comme solutions uniques du système de Cramer indiqué précédemment, seront invariantes dans les mêmes conditions, donc seront fonctions de période p de la différence $i-j$ seulement. En particulier m_{ii} sera indépendant de i et représentera la variance V de q dans un groupe quelconque. Le quotient $\frac{m_{ij}}{V}$

1 pour le schéma à 2 dimensions v. G. Malecot, Annales de l'Univ. de Lyon-Sciences Math. 1950, p. 52.

2 Car on verra aisément qu'elle n'intervient pas effectivement dans les formules; nous ne l'introduisons que pour alléger les notations.

sera le coefficient de corrélation entre les fréquences q dans deux groupes séparés par $(i-j)$ intervalles, nous le désignerons par $f(i-j)$ avec $f(0) = 1$. Le système de Cramer (système (19) à I^{ERS} membres nuls) donnera alors, en divisant tous les membres par V :

$$(20) \sum_k [f(i-k) l(k-j) + f(j-k) l(k-i)] - f(i-j) \sum_k [l(k-i) + l(k-j)] - 2(u+v)f(i-j) = -s_{ij} \frac{K}{2N} \left[\frac{\bar{q} - \bar{q}^2}{V} - 1 \right] \quad \text{= (voir note page 13)}$$

Et cette formule conduit à une relation très simple entre la "fonction génératrice des f " soit :

$$F(\alpha) = \sum_{j=1}^{j+p} \alpha^j f(j)$$

et la "fonction génératrice des l " soit $L(\alpha) = \sum_{j=1}^{j+p} \alpha^j l(j)$.

Il suffit de supposer que α est l'une quelconque des racines $p^{\text{èmes}}$ de l'unité pour constater qu'alors $F(\alpha)$ et $L(\alpha)$ ne dépendent pas du choix de l'entier j , et que par suite la formule (20) donne, après multiplication par α^{i-j} et sommation par rapport à j et k :

$$(21) F(\alpha) [L(\alpha) + L(1/\alpha) - 2L(1) - 2(u+v)] = -\frac{K}{2N} \left[\frac{\bar{q} - \bar{q}^2}{V} - 1 \right].$$

Les p équations obtenues en remplaçant α , dans (21), par chacune des racines $p^{\text{èmes}}$ de l'unité, se résolvent très facilement et fournissent les expressions explicites de la variance V et des $p-1$ coefficients de corrélation $f(i-j)$. Il suffit en effet de poser $w_k = e^{2ik\pi/p}$ et de remarquer que la somme de p puissances successives de w_k est nulle dès que $w_k \neq 1$; en ajoutant membre à membre les équations (21) écrites pour $\alpha = w_0, \dots, \alpha = w_k, \dots, \alpha = w_{p-1}$ après les avoir résolues o/ oà F puis divisées respectivement par $w_0^s, \dots, w_k^s, \dots, w_{p-1}^s$ on a :

$$(22) p f(s) = -\frac{K}{2N} \left(\frac{\bar{q} - \bar{q}^2}{V} - 1 \right) \sum_{k=0}^{k=p-1} \frac{w_k^{-s}}{L(w_k) + L(1/w_k) - 2L(1) - 2(u+v)}$$

et en particulier, pour $s = 0$:

$$p = -\frac{K}{2N} \left(\frac{\bar{q} - \bar{q}^2}{V} - 1 \right) \sum_{k=0}^{k=p-1} \frac{1}{L(w_k) + L(1/w_k) - 2L(1) - 2(u+v)}$$

Si l'on fait augmenter indéfiniment le nombre p de colonies, on obtient le schéma de la "migration homogène illimitée" à une dimension. La formule (21) est alors vérifiée par continuité pour tout $\alpha = e^{i\varphi}$ de module 1. La fonction $L(\alpha)$ étant connue en fonction des "taux d'immigration", la fonction $F(\alpha)$ est définie par la formule (21) sur le cercle $\alpha = e^{i\varphi}$; en effectuant un développement en série de Laurent au voisinage de ce cercle, on obtient les valeurs des $f(i-j)$ comme coefficients de ce développement; on vérifie aisément qu'elles sont données par les formules déduites de (22) en remplaçant les sommes par des intégrales; cela permet en particulier de montrer que le coefficient de corrélation $f(i-j)$ décroît sensiblement en fonction exponentielle de la distance $(i-j)$ ^{1*}.

^{1*} Ce résultat établi pour la 1ère fois par G. Malécot dans le Compte-Rendu du ⁼ international de Probabilités de Lyon (éditions du C.N.R.S., 1949) pour une forme particulière de la loi d'immigration, a été généralisé et ses limites de validité ont été précisées par le même auteur dans les Annales de l'Univ. de Lyon, Sciences Math., XIII (1950) p.50 et XIV (1951) p. 116.