

ANNALES DE L'I. H. P.

DAVID KENDALL

Les processus stochastiques de croissance en biologie

Annales de l'I. H. P., tome 13, n° 1 (1952), p. 43-108

http://www.numdam.org/item?id=AIHP_1952__13_1_43_0

© Gauthier-Villars, 1952, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P. » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

Les processus stochastiques de croissance en biologie

par

David G. KENDALL,
Fellow, Magdalen College, Oxford.

CHAPITRE I.

LA LOI STOCHASTIQUE DE CROISSANCE DES ORGANISMES BACTÉRIENS

1.1. Lorsque l'on veut construire un modèle mathématique pour représenter la croissance réelle des organismes bactériens, il est extrêmement important d'attribuer une forme mathématique précise à la

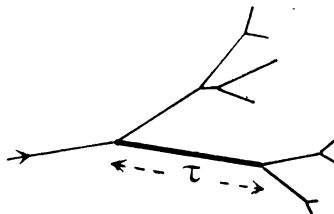


Fig. 1. — Définition du temps de division τ . •

distribution ⁽¹⁾ de ce que nous pourrions appeler *le temps de division* τ . Si une bactérie est née à l'époque t et si sa propre division se place à l'époque $t + \tau$, le temps τ sera son temps de division.

D'après les travaux de Kelly et de Rahn (1932) nous savons quelle

⁽¹⁾ Au lieu de « distribution » d'une variable aléatoire, certains auteurs français préférèrent l'expression « loi de répartition ».

est la distribution réelle du temps de division de certains organismes, en particulier celui du *Bacterium ærogenes* (à une température de 30° C). On trouve une distribution avec un mode unique et positif (égal à peu près, à une demi-heure), et dispersée autour de ce mode avec un coefficient de variation de 15 à 30 %, dont voici la forme approximative :

Dans le modèle mathématique, il faut définir une fonction de répartition $G(\tau)$ pour le temps de division; $dG(t)$ nous donnera alors la valeur de la probabilité de trouver τ dans l'intervalle élémentaire

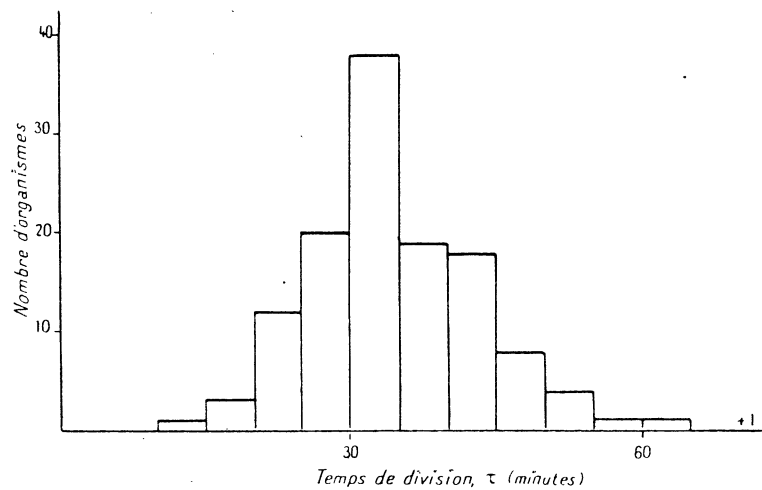


Fig. 2. — Temps de division pour le *B. ærogenes* (Kelly et Rahn : 1932).

$(t, t + dt)$. La fonction G est tout à fait générale, bien qu'elle doive être monotone croissante, continue à droite, et croissante de zéro à l'unité. Étant donné que le temps de division τ n'est pas négatif, par sa définition même, $G(\tau)$ doit être égale à zéro pour toutes les valeurs négatives de τ .

1. 2. Voici maintenant quelques hypothèses particulières concernant la fonction $G(\tau)$:

a. Loi exponentielle. — Pour obtenir un processus markovien de croissance, il faut prendre

$$(1) \quad dG(\tau) = e^{-k\tau} \cdot d\tau \quad (0 < \tau < \infty).$$

Cette distribution est beaucoup trop dispersée pour être une repré-

sentation fidèle des faits empiriques; le coefficient de variation est de 100 % et le mode est égal à zéro. Malgré tout, cette loi présente un grand intérêt mathématique et le processus de croissance qui s'y associe correspond à des situations réelles en Physique. D'ailleurs, lorsque l'on veut faire des calculs plus complexes, on peut arriver, grâce à cette hypothèse, à des simplifications souvent utiles, surtout dans un essai préliminaire.

b. Loi déterministe. — Si l'on pousse le contraste au maximum, on peut affirmer que

$$(2) \quad G(\tau) \equiv 0 \quad \left(\tau < \frac{1}{\lambda} \right) \quad \text{et} \quad G(\tau) \equiv 1 \quad \left(\tau \geq \frac{1}{\lambda} \right),$$

c'est-à-dire que le temps de division est exactement égal à la valeur déterminée $\frac{1}{\lambda}$. (Notons que cette valeur $\frac{1}{\lambda}$ était la valeur moyenne du temps de division dans l'exemple précédent). Le processus de croissance qui s'y associe est un processus purement déterministe, dans lequel le calcul des probabilités n'entre pas en jeu, ce qui est d'ailleurs l'hypothèse classique.

c. Loi à phases multiples. — Entre ces deux hypothèses, on peut en intercaler une autre, plus réaliste (Kendall, 1948c), que voici :

$$(3) \quad dG(\tau) \equiv \frac{k^k \lambda^k}{\Gamma(k)} e^{-k\lambda\tau} \tau^{k-1} d\tau \quad (0 < \tau < \infty).$$

Dans ce cas, on a un mode unique égal à $\frac{1 - \frac{1}{k}}{\lambda}$ et pour la valeur moyenne de τ , on a encore $\frac{1}{\lambda}$. Le coefficient de variation du temps de division est égal à $\frac{1}{\sqrt{k}}$; on obtient ainsi un résultat que nous rencontrerons très souvent. On peut obtenir une représentation de la distribution du temps de division pour le *B. xerogenes* si l'on pose

$$\frac{1}{\lambda} = 30 \text{ m} \quad \text{et} \quad k = 20.$$

Les deux lois précédentes sont contenues dans celle-ci comme cas particuliers. Le processus markovien qui correspond à la loi exponentielle se produit quand k est égal à l'unité. Pour trouver la loi de

croissance déterministe, il suffit de faire tendre la constante k vers l'infini.

On peut en donner une interprétation « microscopique ». Si nous affirmons que le processus de division se compose d'un nombre k de phases *successives*, que chaque phase a besoin d'un temps soumis à la loi de distribution

$$e^{-k\lambda t} k\lambda dt \quad (0 < t < \infty)$$

et que les temps successifs associés aux phases successives sont des variables aléatoires indépendantes, le temps de division (qui sera la somme des temps pour les k phases) aura la distribution (3) ci-dessus. Cette interprétation est grosse de conséquences, car elle permet, par un raffinement artificiel du processus, de le réduire à un processus vectoriel de Markoff. Nous disons « un processus *vectoriel* », parce que dans la description de l'état actuel d'un tel processus, il faut rendre compte du nombre des bactéries dans chacune des phases. Nous n'aurons cependant pas besoin de considérer un tel processus vectoriel pour le moment.

Suivant M. L. Martin de Gembloux, on peut, si l'on veut, donner une signification biologique à ces k phases. Considérons un seul chromosome et supposons qu'il y ait k gènes (ou groupes de gènes) rangés linéairement dessus. Considérons alors une onde de duplication qui partage le chromosome suivant sa longueur, de telle façon que chaque dédoublement d'un gène nécessite un temps aléatoire dont la distribution est celle donnée par la formule ci-dessus. On peut aussi considérer une onde qui s'avance simultanément du centre vers les deux extrémités du chromosome, mais cette hypothèse (comme aussi celle qui tient compte de plusieurs chromosomes) conduit à un modèle mathématique beaucoup plus compliqué. Ce point de vue nous suggère une généralisation importante. Supposons, en effet, que le temps du séjour dans la $s^{\text{ième}}$ phase ait la distribution

$$e^{-\lambda_s t} \lambda_s dt \quad (0 < t < \infty; 1 \leq s \leq k),$$

les constantes λ_s n'étant plus nécessairement égales. En ce cas, le temps de division a la même distribution que la variable aléatoire

$$\frac{u_1}{\lambda_1} + \frac{u_2}{\lambda_2} + \dots + \frac{u_k}{\lambda_k},$$

où les variables u sont toutes indépendantes les unes des autres et chacune a la distribution $e^{-u} du$ ($0 < u < \infty$).

d. Loi de Rahn (1932). — Elle s'énonce de la manière suivante :

$$(4) \quad dG(\tau) \equiv k(1 - e^{-\alpha\tau})^{k-1} e^{-\alpha\tau} \alpha d\tau \quad (0 < \tau < \infty).$$

Notons qu'il s'agit là encore d'une distribution avec un mode unique et positif. Rahn y a été conduit par le raisonnement suivant : Supposons qu'il y ait un nombre k d'entités dans la cellule qui doivent être doublées avant que la cellule ne se divise, et que ces k processus de duplication se déroulent indépendamment et *simultanément*, le temps exigé par le $s^{\text{ième}}$ processus ayant la distribution

$$e^{-\alpha_s t} \alpha_s dt \quad (0 < t < \infty), \quad (1 \leq s \leq k).$$

Dans ce cas la probabilité pour que le temps de division de la cellule elle-même ne dépasse pas la valeur t , sera

$$\prod_{s=1}^k (1 - e^{-\alpha_s t}).$$

Rahn supposa que les constantes α_s étaient toutes égales à une valeur commune, α , ce qui entraîne pour $dG(\tau)$ la formule écrite plus haut. Il identifia le nombre entier k (qu'il appela g) avec le nombre des gènes. Évidemment, si l'on accepte l'hypothèse *c*, ou bien l'hypothèse *d*, on peut associer le nombre k à une caractéristique énumérative du processus de duplication. On ne sait d'ailleurs pas si elle représente le nombre des gènes ou bien le nombre de certains groupes de gènes.

Il est curieux de noter que nous sommes conduits à la formule de Rahn par deux autres raisonnements. J'ai indiqué le premier ailleurs (1952 *a*); quant au deuxième, il est nouveau. Au moyen d'un calcul qui n'est pas très difficile on peut s'assurer que *les deux variables aléatoires*

$$U \equiv \frac{u_1}{1} + \frac{u_2}{2} + \dots + \frac{u_k}{k}$$

et

$$(5) \quad V \equiv \max(u_1, u_2, \dots, u_k),$$

ont la même distribution lorsque les variables u constituent un échantillon pris au hasard de la distribution $e^{-u} du$ ($0 < u < \infty$).

Par conséquent, dans le processus aux phases multiples, si nous posons

$$\lambda_s \equiv s\alpha \quad (1 \leq s \leq k),$$

le temps de division aura la distribution de Rahn.

Pour développer le deuxième raisonnement, il faut considérer le processus de croissance markovien qui découle de la loi exponentielle et que l'on écrira ici sous la forme $e^{-\alpha t} dt$ ($0 < t < \infty$). Supposons que les temps de division pour deux bactéries quelconques de la colonie soient entièrement indépendants et appelons n_t le nombre des bactéries à l'époque t . Si, à l'époque zéro il y a une seule bactérie mère, nous aurons $n_0 = 1$ et il est bien connu, d'après les travaux de Yule (1925) et de Furry (1937), que

$$(6) \quad \text{prob}(n_t = n) = e^{-\alpha t} (1 - e^{-\alpha t})^{n-1} \quad (n \geq 1).$$

En effet, nous avons là la distribution géométrique du processus de Yule et de Furry. Considérons le moment T pendant lequel n_t croît de la valeur k à la valeur $k + 1$. Nous aurons

$$\text{prob}(T \leq t) = \text{prob}(n_t \geq k + 1) = (1 - e^{-\alpha t})^k,$$

qui est l'intégrale indéfinie de la distribution de Rahn; quant à la distribution de T , nous aurons, en effet,

$$k(1 - e^{-\alpha T})^{k-1} e^{-\alpha T} \alpha dT \quad (0 < T < \infty).$$

La distribution de Rahn peut être ainsi expliquée par l'hypothèse d'un tel mécanisme secondaire de croissance multiplicative dans l'intérieur d'une cellule.

Notons pour terminer quelques formules concernant le processus à phases multiples ⁽²⁾

$$\begin{aligned} \text{moyenne } (\tau) &= \sum_{s=1}^k \frac{1}{\lambda_s}; \\ \text{variance } (\tau) &= \sum_{s=1}^k \frac{1}{\lambda_s^2}. \end{aligned}$$

⁽²⁾ Au lieu de « la variance » d'une variable aléatoire, certains auteurs français préfèrent le terme « fluctuation ».

Pour le coefficient de variation de τ , nous aurons

$$(7) \quad \gamma_{\tau} = \frac{\left(\sum \frac{1}{\lambda_s^2} \right)^{\frac{1}{2}}}{\sum \frac{1}{\lambda_s}}.$$

De l'inégalité de Schwartz, on déduit que

$$(8) \quad \gamma_{\tau} \geq \frac{1}{\sqrt{k}},$$

le signe d'égalité ayant lieu seulement lorsque les constantes λ_s sont toutes égales (cas d'un processus à phases multiples *uniformes*).

Pour la distribution de Rahn, on déduit les formules suivantes (en écrivant $s\alpha$ pour λ_s) :

$$\text{moyenne } (\tau) = \frac{1}{\alpha} \sum_{s=1}^k \frac{1}{s} \simeq \frac{(\gamma + \log k)}{\alpha}$$

et

$$(9) \quad \gamma_{\tau} \simeq \frac{\pi}{(\gamma + \log k) \sqrt{6}} \quad (k \rightarrow \infty),$$

où la constante γ est celle d'Euler-Mascheroni (elle a la valeur 0,5772...). Ces formules pour la moyenne et la variance de la distribution de Rahn se trouvent aussi dans un article de Finney et de Martin (1951), qui les ont attribuées à M. R. M. Jones.

1.3. Rahn lui-même avait déjà fait (1932) un premier essai d'examen d'un tel modèle en liaison avec des expériences réelles; ses calculs ont été repris récemment avec plus de précision par Finney et par Martin (1951). De plusieurs séries d'expériences faites par Kelly et par Rahn avec le *B. xerogenes*, Finney et Martin ont tiré pour le nombre k de la distribution de Rahn des valeurs qui varient de 22 à 1010. L'hypothèse un peu douteuse, d'une valeur commune à chaque série pour le deuxième paramètre, α , a donné une valeur de 25 à 35 pour k . Ces auteurs ont insisté sur la difficulté essentielle de la détermination précise de k .

D'un autre côté, j'ai fait (1948c) un calcul fondé à peu près sur les mêmes faits expérimentaux, pour déterminer une valeur de k d'après mon hypothèse c d'un processus de division à phases multiples *uni-*

formes (les constantes λ_s étant toutes égales). J'ai trouvé des valeurs de k qui varient de 12 à 47, avec une valeur moyenne de 22.

Gardons-nous de dire qu'il y a 22 gènes. Peut-être pouvons-nous dire que dans le mécanisme de division d'une telle bactérie il y a 20 ou 30 étapes d'une nature complètement inconnue. Cependant, à mon avis, cette conclusion même n'est pas inévitable. Essentiellement, quelle que soit la formule qui donne $G(\tau)$, on estimera la constante k à l'aide du coefficient de variation de τ , γ_τ . Ainsi, les estimations de Finney et de Martin et les miennes équivalent à un seul fait incontestable :

$$\text{coefficient de variation } (\tau) \simeq 20 \%$$

C'est le plus que l'on puisse en dire à moins que l'on n'accepte comme sûre une hypothèse quelconque sur la nature du mécanisme de duplication.

J'ai raconté ailleurs (1952 a) une histoire qui éclaire la situation. La *British Iron and Steel Research Association* a fait une enquête sur les retards que souffrent les navires chargés de minerais avant qu'ils ne soient déchargés. J'ai le privilège de pouvoir citer un Mémoire encore inédit de M. D. G. Owen sur ce sujet qui a trouvé, pour le temps de déchargement des navires, une distribution avec un seul mode positif et avec un coefficient de variation d'environ 50 %. On peut alors expliquer ce fait, si l'on veut, par un processus à quatre phases uniformes, car les faits expérimentaux sont rares, et il est remarquable qu'ils suffisent à réfuter une telle hypothèse. Mais il ne s'ensuit pas qu'il y ait quatre étapes distinctes dans le processus de déchargement d'un tel bateau. Ce cas est à peu près analogue à celui du nombre de gènes.

1.4. Considérons maintenant le processus de croissance d'une colonie bactérienne, lorsque la fonction $G(\tau)$ a une forme tout à fait générale. Bellman et Harris l'ont discuté en 1948 et en 1951.

\mathcal{E} étant le signe de formation d'une valeur moyenne, nous aurons

$$\begin{aligned} (10) \quad \varphi(z, t) &\equiv \mathcal{E} z^{n_t} \equiv \sum_{n=1}^{\infty} z^n \text{prob}(n_t = n) \\ &= \int_0^t \{ \varphi(z, t-u) \}^2 dG(u) + z \{ 1 - G(t) \}. \end{aligned}$$

On peut arriver à l'équation intégrale (écrite et étudiée de très près

par Bellman et par Harris) en tenant compte de ce qui peut se trouver dans l'intervalle de temps $(0, t)$. L'on suppose ici :

- 1° qu'il n'y a eu qu'une seule bactérie mère à l'instant $t = 0$;
- 2° que cette bactérie mère était, à ce moment-là, nouvellement née;
- 3° que les organismes dans la colonie à chaque époque sont tous indépendants les uns des autres;
- 4° qu'il n'y a aucun lien stochastique entre le temps de division d'un organisme et celui de son organisme générateur.

Pour obtenir l'équation de Bellman et Harris, il suffit de considérer que deux cas seulement peuvent se présenter :

a. La bactérie mère peut se diviser à une époque u antérieure à t . La probabilité d'une telle division dans l'intervalle élémentaire $(u, u+du)$ est égale à $dG(u)$ et, après cette division, nous obtiendrons deux organismes semblables à la bactérie mère;

b. La bactérie mère peut rester entière jusqu'à l'époque t . La probabilité de ce cas est égale à $1 - G(t)$.

Si le processus est du type simple de Yule et de Furry, on peut transformer l'équation et la résoudre comme suit. Nous avons (après une transformation initiale de la variable indépendante)

$$\varphi(z, t) = \int_0^t \{\varphi(z, u)\}^2 e^{-\lambda t} \lambda e^{\lambda u} du + z e^{-\lambda t}.$$

Introduisons les nouvelles variables

$$T \equiv e^{-\lambda t}, \quad \Phi(z, T) \equiv e^{\lambda t} \varphi(z, t), \quad U \equiv e^{-\lambda u};$$

nous trouverons l'équation intégrale transformée

$$\Phi(z, T) = \int_T^1 \{\Phi(z, U)\}^2 dU + z.$$

L'équation différentielle suivante lui est équivalente :

$$\frac{\partial}{\partial T} \left(\frac{1}{\Phi} \right) = 1, \\ \Phi(z, 1) = z.$$

Elle a la solution unique

$$\frac{1}{\Phi} = T + \frac{1-z}{z},$$

que l'on peut écrire

$$(11) \quad \varphi(z, t) = \frac{z e^{-\lambda t}}{1 - z(1 - e^{-\lambda t})},$$

qui est la fonction génératrice de la distribution géométrique de Yule-Furry dont j'ai déjà parlé.

Si le processus est du type déterministe, l'équation de Bellman et de Harris conduira à l'équation fonctionnelle

$$(12) \quad \varphi\left(z, \frac{r+1}{\lambda}\right) = \left\{ \varphi\left(z, \frac{r}{\lambda}\right) \right\}^2 \quad [\varphi(z, 0) = z],$$

ce qui veut dire que la population se double rigoureusement après chaque intervalle de temps égal à $\frac{1}{\lambda}$.

En général, on ne peut résoudre l'équation intégrale sans se servir des méthodes plus compliquées de Bellman et de Harris.

Il y a cependant un cas assez général dans lequel une technique spéciale et tout à fait élémentaire nous permet de donner des résultats très utiles. Ce cas se présente lorsque la duplication d'une cellule est produite par l'opération d'un processus à phases multiples (uniformes ou non) (D'après ce que j'ai déjà dit, les calculs qui suivent s'appliquent également lorsque le temps de division a la distribution de Rahn). La technique exposée ci-après est le résultat d'un travail fait en collaboration avec mon ami M. W. A. O'N. Waugh, de Trinity College, Oxford, qui est actuellement en train d'en étudier les applications et de la compléter.

Soit un temps de division composé de la manière suivante :

$$(13) \quad \tau = \frac{u_1}{\lambda_1} + \frac{u_2}{\lambda_2} + \dots + \frac{u_k}{\lambda_k},$$

où les variables u sont toutes indépendantes les unes des autres, chacune ayant la distribution $e^{-u} du$ ($0 < u < \infty$). La fonction $G(\tau)$ étant dérivable, il nous est permis (et il sera utile) d'écrire

$$(14) \quad dG(\tau) \equiv g(\tau) d\tau.$$

Alors, si $g_s(t) \equiv \lambda_s e^{-\lambda_s t}$, nous aurons

$$(15) \quad g(t) = g_1(t) * g_2(t) * \dots * g_k(t),$$

où l'astérisque signifie comme d'habitude l'opération

$$a(t) \star b(t) \equiv \int_0^t a(x) b(t-x) dx.$$

L'équation de Bellman et de Harris deviendra

$$(16) \quad \varphi(z, t) = g_1(t) \star g_2(t) \star \dots \star g_k(t) \star \{[\varphi(z, t)]^2 - z\} + z,$$

et sous cette forme il est facile de la résoudre, à cause de l'identité opérationnelle (1)

$$(17) \quad \left(1 + \frac{1}{\lambda_s} \frac{d}{dt}\right) \{g_s(t) \star \psi(t)\} = \psi(t).$$

Par conséquent, nous avons l'équation différentielle fondamentale

$$(18) \quad \prod_{s=1}^k \left(1 + \frac{1}{\lambda_s} \frac{d}{dt}\right) \varphi(z, t) = \{\varphi(z, t)\}^2,$$

les deux termes z ayant disparu. Il faut ajouter des conditions aux limites; elles sont, en effet,

$$(19) \quad \varphi(z, 0) = z, \quad \left(\frac{d}{dt}\right)^s \varphi(z, 0) = 0 \quad (1 \leq s \leq k-1).$$

Pour s'en assurer, il suffit de considérer les valeurs des expressions

$$\prod_{s=1}^j \left(1 + \frac{1}{\lambda_s} \frac{d}{dt}\right) \varphi(z, t) \quad (0 \leq j \leq k-1)$$

lorsque que t est égal à zéro.

Dans la pratique, c'est surtout la moyenne, $\mathcal{E}n_t$, et le coefficient de variation

$$(20) \quad \gamma_n \equiv \frac{\sqrt{\text{var } n_t}}{\mathcal{E}n_t},$$

qui sont de première importance. Nous écrivons

$$(21) \quad q_1(t) \equiv \mathcal{E}n_t \quad \text{et} \quad q_2(t) \equiv \mathcal{E}n_t(n_t - 1).$$

(1) Pour une étude détaillée de ces « convolution transforms », voir I. I. Hirschman Jr. et D. V. Widder, *Trans. American Math. Soc.*, 66 (1949), 135-201 et 67 (1949), 69-97.

Nous aurons alors

$$q_1(t) = \frac{d}{dz} \varphi(1, t) \quad \text{et} \quad q_2(t) = \left(\frac{d}{dz} \right)^2 \varphi(1, t);$$

$$\mathcal{E} n_t = q_1 \quad \text{et} \quad \text{var}(n_t) = q_2 + q_1 - q_1^2$$

et

$$(22) \quad \gamma_n = \sqrt{\frac{q_2}{q_1^2} + \frac{1}{q_1} - 1}.$$

A partir de l'équation différentielle fondamentale, nous obtenons, pour la moyenne, une équation différentielle linéaire à coefficients constants

$$(23) \quad f(D) q_1(t) = 0,$$

où le polynôme caractéristique $f(x)$ a la valeur

$$(24) \quad f(x) = \prod_{s=1}^k \left(1 + \frac{x}{\lambda_s} \right) - 2$$

et où le symbole D signifie l'opérateur $\frac{d}{dt}$. Les conditions aux limites sont

$$(25) \quad q_1(0) = 1, \quad \left(\frac{d}{dt} \right)^s q_1(0) = 0 \quad (1 \leq s \leq k-1)$$

et la solution unique est donnée par la série finie

$$(26) \quad q_1(t) = \sum_{l=1}^k \frac{e^{\rho_l t}}{\rho_l f'(\rho_l)},$$

où les constantes ρ_1, \dots, ρ_k sont les racines (supposées simples, ce qui est vrai hormis dans des cas exceptionnels) de l'équation algébrique $f(x) = 0$. Pour s'assurer que cette somme est vraiment la solution, il suffit de vérifier que les conditions aux limites sont remplies. Or, la condition $q_1(0)$ égal à l'unité est remplie en vertu de l'identité

$$(27) \quad \frac{1}{f(x)} = \sum_{l=1}^k \frac{1}{f'(\rho_l)(x - \rho_l)}$$

(notons que $f(0)$ est égale à -1).

D'ailleurs, pour les valeurs de x suffisamment grandes, nous aurons

$$\sum_{l=1}^k \frac{1}{f''(\rho^l)} \left(\frac{1}{x} + \frac{\rho^l}{x^2} + \frac{\rho^{2l}}{x^3} + \dots \right) = O\left(\frac{1}{x^k}\right),$$

ce qui rend évident le fait que les autres conditions sont remplies.

Il est facile de voir que l'équation algébrique, $f(x) = 0$, admet une racine positive qui est simple et unique; appelons-la ρ , sans aucun indice. Quant aux autres racines, chacune d'elles a une partie réelle qui est inférieure à ρ . Donc, pour les valeurs assez grandes du temps t , nous aurons la formule asymptotique

$$(28) \quad \mathcal{E} n_t \simeq \frac{e^{\rho t}}{2f'(\rho)}.$$

Le calcul du coefficient de variation commence avec l'équation

$$(29) \quad f(D) q_2(t) = 2[q_1(t)]^2,$$

que l'on déduit aisément de l'équation différentielle fondamentale en différenciant deux fois par rapport à z . Les conditions aux limites sont

$$(30) \quad \left(\frac{d}{dt}\right)^s q_2(0) = 0 \quad (0 \leq s \leq k-1).$$

Une intégrale particulière peut être obtenue par la méthode habituelle; écrivons cette intégrale sous la forme

$$\frac{2}{f(D)} [q_1(t)]^2.$$

Le terme principal sera celui qui est associé au carré du terme principal dans la formule de la moyenne; asymptotiquement, on peut écrire

$$(31) \quad q_2(t) \simeq \frac{2}{f(2\rho)} [q_1(t)]^2,$$

tous les termes de la fonction complémentaire étant d'un ordre de grandeur relativement négligeable. Ainsi, pour les valeurs du temps t assez grandes, nous aurons

$$(32) \quad \gamma_n \simeq \sqrt{\frac{2}{f(2\rho)} - 1}.$$

C'est là notre résultat principal; on peut le déduire aussi des calculs

antérieurs et beaucoup plus généraux de Bellman et de Harris, mais je crois pourtant que notre déduction a un intérêt à cause de sa simplicité et qu'elle sera peut-être utile dans des calculs plus complexes de ce genre.

Plaçons-nous maintenant dans le cas le plus simple, dans lequel il s'agit d'un processus à k phases *uniformes*. L'on aura alors $\lambda_s = k\lambda$ pour chaque valeur de s ; le polynome caractéristique deviendra

$$f(x) \equiv \left(1 + \frac{x}{k\lambda}\right)^k - 2$$

et, par conséquent,

$$(33) \quad \rho = \lambda \alpha_k$$

si l'on définit α_k par

$$\alpha_k \equiv k \left(2^{\frac{1}{k}} - 1\right).$$

Nous aurons, de plus,

$$f'(\rho) = \frac{2^{1-k}}{\lambda}$$

et aussi

$$f(2\rho) = \left(1 + \frac{2\alpha_k}{k}\right)^k - 2.$$

Pour les valeurs assez grandes du temps t , nous aurons

$$(34) \quad \mathcal{E}n_t \simeq \frac{2^{\frac{1}{k}} e^{\alpha_k \lambda t}}{2 \alpha_k},$$

et

$$(35) \quad \frac{\gamma_n}{\gamma_\tau} \equiv \gamma_n \sqrt{k} \rightarrow A_k \quad (t \rightarrow \infty).$$

Ici, la constante A_k est donnée par la formule

$$(36) \quad A_k^2 \equiv k \left[\frac{4 - \left(1 + \frac{2\alpha_k}{k}\right)^k}{\left(1 + \frac{2\alpha_k}{k}\right)^k - 2} \right].$$

J'ai donné ces formules-ci pour le cas uniforme en (1948c), les ayant déduites par une méthode d'une complexité rebutante.

Nous aurons besoin des valeurs numériques des coefficients α_k et A_k .
Or, nous avons

$$(37) \quad \lim_{k \rightarrow \infty} \alpha_k = \log_e 2 \quad \text{et} \quad \lim_{k \rightarrow \infty} A_k = 2^{\frac{1}{2}} \log_e 2$$

et, de plus,

$$(38) \quad \frac{A_k}{A_\infty} = 1 + \frac{0,01377}{k} + \frac{0,01069}{k^2} + \dots$$

A l'aide de ces formules et des calculs directs, j'ai établi le tableau suivant :

Nombre des phases		
k .	α_k .	A_k .
1.....	1	1
2.....	0,8284	0,9890
3.....	7798	9858
4.....	7568	984
5.....	7435	983
6.....	0,7348	0,983
7.....	7286	982
8.....	7241	982
9.....	7205	982
10.....	7177	982
15.....	0,7094	0,981
20.....	7053	981
25.....	7029	981
30.....	7012	981
35.....	7000	981
∞	0,69315	0,980258

On voit que la valeur du facteur A_k est proche de l'unité pour chaque valeur de k et que pour les valeurs de k que l'on rencontre dans la pratique, A_k est constant et α_k ne varie que très peu.

Ces valeurs numériques du facteur A_k m'ont amené à supposer que peut-être :

S'il n'y a qu'une seule bactérie mère à l'instant initial (qui doit être l'instant précis de sa propre naissance), si les temps de division pour tous les organismes sont indépendants et si, de plus, la distribution $dG(\tau)$ est « suffisamment concentrée » autour de sa moyenne, on aura

$$(39) \quad \lim_{t \rightarrow \infty} \gamma_n \simeq A_\infty \gamma_\tau \quad (A_\infty = 0,980\dots)$$

J'ai démontré l'exactitude de cette conjecture pour le processus à k phases uniformes, quel que soit le nombre des phases, et sans aucune condition de concentration. Actuellement, M. Waugh poursuit l'examen détaillé des cas plus généraux à l'aide des résultats que nous avons obtenus ensemble et dont le plus important est constitué par la formule (32) donnant la valeur asymptotique du coefficient de variation.

Considérons les conditions qui interviennent dans la conjecture ci-dessus. La première condition, à savoir qu'il n'y a qu'une seule bactérie mère, fraîchement née, est de la plus haute importance. Évidemment, s'il y a N bactéries mères toutes nées à l'instant initial, il suffit de remplacer la formule asymptotique par la suivante :

$$(49) \quad \lim_{t \rightarrow \infty} \gamma_n \simeq A_\infty \frac{\gamma_\tau}{\sqrt{N}}.$$

Mais, en pratique, il faudra aussi tenir compte des « âges » initiaux des bactéries mères. J'entends ici par l'âge d'une bactérie *la partie épuisée de son temps de division*. [Ce qui constitue également la définition utilisée par Harris (1951). Notons qu'elle est *tout à fait différente* de la définition utilisée par Bartlett et par moi-même et dont il sera question dans la quatrième partie de cet article]. Comme hypothèse naturelle, on peut supposer que les âges initiaux sont répartis selon la distribution *stable* des âges, dont on trouvera une discussion dans un article récent de Harris (1951) [pour la distribution stable des âges dans un processus à k phases uniformes, voir aussi mon article de 1948 (c)]. M. Waugh considère également cette question; ses résultats paraîtront ailleurs.

La deuxième condition (indépendance des temps de division) exige qu'il n'y ait aucune liaison stochastique entre le temps de division d'une bactérie et celui de son organisme générateur. En effet, on trouve quelques indications d'une telle liaison (Kelly et Rahn, 1932; Armitage, 1951; Kendall, 1952 a), mais on ne peut pas la représenter dans la théorie mathématique sans une connaissance plus détaillée de sa nature et de sa grandeur.

C'est M. Martin qui m'a fait apprécier l'importance biologique de ces idées. Nous avons vu que γ_τ est étroitement lié au nombre entier k , nombre assez intéressant, que nous l'identifions au « nombre de gènes » ou non. De plus, on peut par la méthode précédente mesurer γ_τ au

moyen d'une détermination de γ_n , c'est-à-dire, en considérant les fluctuations dans la variable aléatoire qui est la grandeur n d'une colonie que l'on a laissée croître pendant un intervalle déterminé (et assez grand) à partir d'une seule bactérie mère fraîchement née. La réalisation de ce programme présentera sans doute des difficultés de toutes sortes, mais peut-être en vaut-il quand même la peine.

CHAPITRE II.

LES PROCESSUS STOCHASTIQUES ASSOCIÉS AUX MUTATIONS DES BACTÉRIES.

2.1. Dans le premier chapitre j'ai étudié la croissance bactérienne normale du point de vue des processus stochastiques. De la même manière, je vais considérer maintenant les problèmes qui se posent lorsque l'on tient compte d'un autre phénomène bactérien, celui de la mutation.

Qu'elle est l'origine de l'immunité des bactéries lorsqu'elles sont attaquées par des bactériophages, des drogues telle que la péniciline et par d'autres agents antibactériens ? Voilà une question qui a provoqué de chaudes discussions. Deux hypothèses principales tâchent séparément d'expliquer les faits expérimentaux, mais certains auteurs préfèrent une combinaison des deux hypothèses.

D'une part, la première hypothèse suppose que *tous* les organismes ont une même possibilité d'adaptation à un milieu nouveau. Seulement, un nombre aléatoire d'entre eux cependant (variable aléatoire du type de Poisson) réussiront à s'adapter et transmettront l'immunité acquise à leur postérité. Le reste des organismes sera éliminé.

D'autre part, la seconde hypothèse considère que dans n'importe quelle colonie bactérienne il existe des bactéries résistantes qui sont des produits de mutation. Les agents antibactériens (détruisant des bactéries non résistantes) les mettent donc en évidence. Il est extrêmement important de noter que le nombre des bactéries résistantes aura une distribution bien différente de celle de Poisson.

Luria et Delbrück (1943) furent les premiers à noter que si plusieurs

colonies bactériennes sont cultivées pendant un temps déterminé et enfin placées dans un milieu nouveau, l'étude statistique des nombres des bactéries résistantes rendra possible la comparaison des deux hypothèses. Il est beaucoup plus simple d'énoncer les principes d'une telle comparaison que de la mettre à exécution. Des difficultés mathématiques et pratiques se posent en même temps. Même actuellement, il n'existe pas d'accord général sur les faits expérimentaux ⁽³⁾. Par conséquent, un mathématicien doit avancer avec grande circonspection sur un terrain aussi perfide; mais il est tout de même évident que ni l'une ni l'autre de ces hypothèses ne peuvent être jugées correctement avant que les solutions des problèmes associés ne soient connues. Au cours de ce qui suit nous nous occuperons *seulement* de quelques problèmes associés à la deuxième hypothèse : celle de la mutation. Cette restriction n'entraîne nullement un jugement prématuré sur une question qui n'est pas encore tranchée.

Les expériences de Luria et de Delbrück ont été continuées par plusieurs chercheurs, notamment Newcombe (1948). [*Voir*, pour les contributions principales, la compilation récemment faite par Lederberg (1951).] D'autre part, Lea et Coulson (1949), Armitage (1951), Bartlett (1951 *b*) et Harris (1951) avaient fait de grands progrès dans la théorie mathématique. Il y a, en effet, deux problèmes interprétatifs : le premier consiste à faire une comparaison critique entre la distribution réelle des nombres des bactéries résistantes et la distribution donnée par la théorie. Le second problème interprétatif consiste à déduire d'une telle comparaison *un taux de mutation*. Dès l'abord, on vit que ces problèmes étaient intimement liés entre eux. Luria et Delbrück firent deux estimations du taux de mutation; ils utilisèrent :

1° Dans leur première méthode, le nombre des cultures qui ne présentaient pas de mutations;

2° Dans leur deuxième méthode, le nombre moyen (par culture) des bactéries résistantes.

Leurs données étaient déduites d'expériences sur la bactérie *Escherichia coli B.* et sa résistance à un virus donné. Leurs deux estimations du taux de mutation étaient bien différentes, la deuxième étant la plus

⁽³⁾ *Voir*, par exemple, Dean et Hinshelwood (1952).

grande dans un rapport de 5 environ. Par une analyse nouvelle de leurs données et de quelques autres données de Newcombe, Armitage réduisit ce rapport à approximativement 4. La deuxième méthode se servait de la moyenne d'une distribution d'une forme exceptionnelle et il fallait s'attendre à une erreur statistique considérable. Luria et Delbrück proposèrent une correction qui me semble très douteuse, et à laquelle nous ne nous intéresserons pas dans la suite.

Newcombe (1948) a suggéré une explication de la différence entre les deux estimations : celle-ci pouvait provenir d'un *délai phénotypique* qui varie entre deux et six générations. Si nous supposons qu'un tel délai existe, cela signifie que le caractère résistant acquis par mutation par une bactérie ne se manifeste dans sa postérité que quelques générations *après*.

Dans son précieux Mémoire, Armitage a examiné ces questions entre beaucoup d'autres en supposant la plupart du temps un mode de croissance déterministe. Lea et Coulson (1949) ont donné une solution d'un type mixte, en supposant pour les bactéries normales un mode de croissance déterministe et pour les bactéries mutantes un mode de croissance markovien dans lequel

$$e^{-\lambda\tau} \lambda d\tau \quad (0 < \tau < \infty)$$

est la distribution du temps de division. Une solution entièrement markovienne a été donnée par Bartlett (1951 *b*). Harris (1951) a donné une solution mixte dans laquelle le mode de croissance des bactéries normales est déterministe et celui des bactéries mutantes est d'un type tout à fait général. Parmi ces auteurs, Armitage seul a considéré l'effet d'un délai phénotypique. L'idéal serait une solution qui tiendrait compte du délai phénotypique et dans laquelle les processus de croissance pour les bactéries des deux types seraient des processus à phases multiples. Je me propose uniquement de sonder les possibilités d'une théorie à ce point réaliste.

Il sera utile de classifier les diverses solutions au moyen de symboles composés tels que D/M/0 ou G/M/+, dont la signification est la suivante :

1° La première lettre (qui sera toujours D, ou M, ou G) indiquera que l'on suppose pour les bactéries *normales* un mode de croissance déterministe, markovien ou plus général (par exemple, celui associé au processus à phases multiples);

2° La deuxième lettre indiquera de la même façon le type de croissance que l'on attribue aux bactéries *mutantes*;

3° Quant au troisième symbole, il sera + (si l'on tient compte du délai phénotypique) ou zéro (si l'on n'en tient pas compte).

Les solutions antérieures et celles de cet article peuvent alors être rangées dans le tableau suivant :

Luria et Delbrück (1943).....	D/D/o
Lea et Coulson (1949).....	D/D/o, D/M/o
Armitage (1951).....	D/D/o, D/D/+
Harris (1951).....	D/G/o
Bartlett (1951 <i>b</i>).....	M/M/o
Kendall (§ 2.4).....	M/G/o, M/G/+

Armitage a discuté également le processus D/M/o de Lea et Coulson. Des formules de Harris on pourrait déduire une solution du type D/G/+, mais cet auteur ne l'a pas fait.

2.2. Hypothèses générales. — 1° Nous supposons d'abord qu'à l'instant initial (t égal à zéro) il y ait une seule bactérie mère, née à cet instant-là; elle doit être normale. C'est uniquement quand la croissance des bactéries normales est du type markovien que l'on peut ignorer l'hypothèse que la bactérie mère était fraîchement née à l'époque zéro. (S'il y a N bactéries mères dont chacune se conforme aux hypothèses que nous venons de faire, il suffira d'élever à la puissance $N^{\text{ième}}$ la fonction génératrice de la distribution du nombre des bactéries résistantes.)

2° Nous supposons qu'après l'époque de division d'une bactérie normale, il y ait deux cas possibles et deux seulement :

a. Premier cas, dont la probabilité est égale à p : il y a deux bactéries, l'une étant normale et l'autre mutante;

b. Deuxième cas, dont la probabilité est égale à q : il y a deux bactéries, chacune étant normale.

La somme $p + q$ doit être égale à l'unité. Il existe beaucoup d'autres hypothèses possibles sur l'incidence de la mutation, mais celle que nous

avons choisie est pour nous la plus commode. Sa grande utilité mathématique a été notée par Bartlett (1951 *b*).

3° Nous supposons qu'il n'y ait aucune liaison stochastique entre les temps de division de deux bactéries quelconques.

4° Nous supposons qu'il n'existe pas de mutations dans la direction inverse.

La probabilité p sera pour nous la mesure la plus naturelle du taux de mutation; pratiquement, elle ne sera pas supérieure à une chance sur dix millions.

Idéalement, on voudrait savoir la forme analytique précise de la distribution du nombre des bactéries résistantes. Cependant, il sera utile de connaître, pour une seule colonie, le nombre moyen des bactéries résistantes et la probabilité de n'en trouver aucune.

2.3. Les équations fondamentales. — A l'instant t , soit m_t le nombre des bactéries *résistantes* (qui est en général différent du nombre des bactéries *mutantes*, à cause du délai phénotypique) et soit $\psi(z, t)$ la fonction génératrice

$$(41) \quad \psi(z, t) \equiv \mathcal{E} z^{m_t}.$$

Nous aurons alors pour ψ l'équation intégrale fondamentale

$$(42) \quad \psi(z, t) = \int_0^t [p \varphi(z, t-u) + q \psi(z, t-u)] \psi(z, t-u) dG(u) + 1 - G(t).$$

On obtient cette équation au moyen d'une méthode utilisée dans le premier chapitre, en considérant simplement ce qui peut arriver dans l'intervalle $(0, t)$. Il n'arrivera rien, peut-être, et alors la probabilité sera égale à $1 - G(t)$. D'autre part, si la division de la bactérie mère a lieu à une époque v antérieure à t et est contenue dans un intervalle infinitésimal $(u, u + du)$ [probabilité, $dG(u)$], deux cas seront possibles, suivant qu'une des deux bactéries nouvelles est mutante (probabilité conditionnelle égale à p) ou non (probabilité conditionnelle égale à q).

Lorsqu'on ne tient pas compte de l'existence d'un délai phénotypique, la fonction $\varphi(z, t)$, qui se présente dans la fonction à intégrer, est identique à celle introduite dans le premier chapitre. En général, $\varphi(z, t)$ est définie comme il suit. Soit n_t le nombre (à l'instant t) des

bactéries *résistantes* parmi la postérité d'une bactérie mère *mutante*, fraîchement née d'une parente *normale* à l'époque zéro. $\varphi(z, t)$ sera dans ce cas la fonction génératrice

$$(43) \quad \varphi(z, t) \equiv \mathcal{E} z^{nt}.$$

Elle est la solution d'une équation intégrale dont la forme dépend de la nature précise du délai phénotypique.

Si nous connaissions la fonction ψ , il suffirait de déterminer le coefficient de la $m^{\text{ième}}$ puissance de z pour trouver la probabilité qu'à l'époque t il y ait m bactéries résistantes. La probabilité pour qu'il n'y ait pas de bactéries résistantes est égale à $\psi(0, t)$. Le nombre moyen de telles bactéries est égal à $\frac{\partial}{\partial z} \psi(1, t)$.

Si la croissance des bactéries *normales* a lieu suivant un processus à phases multiples, on peut encore se servir de l'opérateur différentiel qui a été utilisé dans le problème de la croissance normale par M. Waugh et par moi-même, et dont il a été question dans le premier chapitre. On obtient ainsi l'équation différentielle fondamentale

$$(44) \quad \prod_{s=1}^k \left(1 + \frac{1}{\lambda_s} \frac{\partial}{\partial t} \right) \psi = \psi^2 + p\psi(\varphi - \psi),$$

avec les conditions aux limites

$$(45) \quad \psi(z, 0) = 1, \quad \left(\frac{\partial}{\partial t} \right)^s \psi(z, 0) = 0 \quad (1 \leq s \leq k-1).$$

2.4. Les processus des types M/G/o, +. — La connaissance du rôle joué par le mode de croissance est de la plus haute importance dans ces problèmes et il nous faut faire une distinction entre la croissance des bactéries du type normal et celle des bactéries du type mutant. J'ai déjà fait référence à la solution du type D/G/o de Harris; je vais donner maintenant une solution du type M/G/o, de sorte qu'il sera possible de déterminer l'effet sur la solution du mode de croissance des bactéries du type *normal*. Pour les calculs suivants, il ne sera pas nécessaire d'établir à l'avance si l'on tient compte du délai phénotypique ou non. Tout ceci est contenu dans la forme de la fonction φ et nous n'aurons pas besoin de la spécifier. Mais, s'il n'y a pas de délai, la fonction φ sera identique

à la fonction φ du premier chapitre et elle peut être celle qui est associée à un processus de croissance à phases multiples, par exemple.

Comme par le passé, il sera commode d'écrire

$$(46) \quad \Phi T \equiv e^{-\lambda t}, \quad (z, T) \equiv e^{\lambda t} \varphi(z, t), \quad \Psi(z, T) \equiv e^{\lambda t} \psi(z, t),$$

et alors l'équation différentielle fondamentale deviendra

$$(47) \quad \frac{\partial}{\partial T} \left(\frac{1}{\Psi} \right) - p \frac{\Phi}{\Psi} = q,$$

avec la condition $\Psi(z, 1) = 1$. La solution est immédiate, car si nous écrivons

$$(48) \quad \Omega(z, U) \equiv \exp \left[p \int_U^1 \Phi(z, V) dV \right],$$

nous aurons

$$(49) \quad \Psi(z, T) = \frac{\Omega(z, T)}{1 - q \int_T^1 \Omega(z, U) dU}$$

et par conséquent la fonction qui engendre la distribution du nombre des bactéries résistantes sera

$$(50) \quad \psi(z, t) \equiv \frac{e^{-q\lambda t} \chi(z, t)}{1 - \int_0^t e^{-q\lambda u} \chi(z, u) q \lambda du},$$

où χ est la fonction génératrice auxiliaire

$$(51) \quad \chi(z, t) \equiv \exp \left\{ -p \int_0^t [1 - \varphi(z, v)] \lambda dv \right\}.$$

En particulier, si les bactéries des deux types ont la même distribution du temps de division, avec la même valeur de λ , et s'il n'y a pas de délai, nous obtiendrons la solution du type M/M/0 de Bartlett (1951 b)

$$(52) \quad \Phi(z, T) = \frac{z}{1 - z(1 - T)};$$

$$\frac{1}{\psi(z, t)} = 1 + e^{\lambda t} \left(\frac{1 - z}{z} \right) \{ 1 - (1 - z + z e^{-\lambda t})^p \}.$$

En général, nous savons que φ est une fonction génératrice et rien de

plus. Disons que $\{\varphi(z, t)\}$ est une famille mesurable de fonctions génératrices lorsque

$$\varphi(z, t) = \sum_{n=0}^{\infty} p_n(t) z^n \quad (|z| \leq 1; \quad t \geq 0),$$

où

$$p_n(t) \geq 0 \quad (t \geq 0; \quad n = 0, 1, 2, \dots),$$

$$\sum_{n=0}^{\infty} p_n(t) = 1 \quad (t \geq 0)$$

et $p_n(t)$ est une fonction mesurable de t pour chaque nombre entier n . Soit $[\Pi(z, t)]$ une telle famille; on voit aisément que

$$\exp \left\{ -\alpha \int_0^t [1 - \Pi(z, u)] du \right\} \quad (\alpha > 0; t \geq 0)$$

et

$$\frac{e^{-\alpha t} \Pi(z, t)}{1 - \int_0^t e^{-\alpha v} \Pi(z, v) \alpha dv} \quad (\alpha > 0; t \geq 0)$$

sont des familles de même nature. Ceci nous démontre que la fonction donnée par la formule (50) détermine véritablement une distribution de probabilité.

Voyons maintenant ce que l'on peut déduire sans particulariser de la solution générale du type M/G/o, +. Nous savons que la probabilité p est excessivement petite, ce qui nous suggère l'emploi d'approximations. Très souvent cependant la probabilité p est multipliée par le facteur $\frac{1}{T}$ ($= e^{\lambda t}$) qui est grand. Voici les valeurs numériques données par Armitage (1951) pour les expériences de Newcombe :

Expérience.	Nombre initial des bactéries.	λt .	$e^{\lambda t}$.	T.
A à D.....	10	17	3.10^7	3.10^{-8}
E à H.....	10^4	11	4.10^4	3.10^{-5}

Comme je l'ai déjà dit, la probabilité p n'est pas supérieure à 10^{-7} . Considérons d'abord ce qui arrive lorsque p et T tendent vers zéro, de telle sorte que leur rapport, $\frac{p}{T}$, reste constant. J'appellerai ces résultats les *formules approximatives de première espèce*.

Supposons que la valeur absolue de z soit inférieure ou égale à l'unité; nous aurons

$$|U\Phi(z, U)| = |\varphi(z, u)| \leq 1$$

et, par conséquent,

$$|\Phi(z, U)| \leq \frac{1}{U}$$

et

$$\left| p \int_U^1 \Phi(z, V) dV \right| \leq -\frac{p}{U} U \log U.$$

Donc $\Omega(z, T)$ a pour limite l'unité et

$$|\Omega(z, U)| \leq \frac{1}{U^p}.$$

Nous aurons de plus :

$$\left| T \int_T^1 \Phi(z, V) \Omega(z, V) dV \right| \leq T \int_T^1 \frac{dV}{V^{1+p}} = \frac{(T^{-p}-1)T}{p} \rightarrow 0$$

et

$$\begin{aligned} & \left| \int_T^1 V \Phi(z, V) dV \int_V^1 p \Phi(z, W) \Omega(z, W) dW \right| \\ & \leq \int_T^1 dV \int_V^1 p \frac{dW}{W^{1+p}} = \frac{1-T^q}{q} - (1-T) \rightarrow 0 \end{aligned}$$

et, par conséquent,

$$\begin{aligned} \frac{1}{p} \int_T^1 \{ \Omega(z, U) - 1 \} dU &= \int_T^1 V \Phi(z, V) dV - T \int_T^1 \Phi(z, V) \Omega(z, V) dV \\ &+ \int_T^1 V \Phi(z, V) dV \int_V^1 p \Phi(z, W) \Omega(z, W) dW \\ &\rightarrow \int_0^1 V \Phi(z, V) dV = \int_0^\infty \varphi(z, v) e^{-\lambda v} \lambda dv. \end{aligned}$$

Or, nous avons, d'après les formules (46) et (49),

$$\frac{1}{\psi(z, t)} = \frac{1}{\Omega(z, T)} \left\{ 1 - \frac{p}{T} \frac{q}{p} \int_T^1 [\Omega(z, U) - 1] dU + \frac{p}{T} (1-T) \right\}$$

et, par conséquent,

$$(53) \quad \lim_1 \psi(z, t) = \left\{ 1 + p e^{\lambda t} \int_0^\infty [1 - \varphi(z, v)] e^{-\lambda v} \lambda dv \right\}^{-1}.$$

Telle est la fonction génératrice de la distribution limite, parce que ⁽⁴⁾

$$\lim_{x \rightarrow 1-0} \lim_t \psi(x, t) = 1.$$

Lorsqu'il y a N bactéries mères, il suffit d'élever la fonction génératrice limite à sa $N^{\text{ième}}$ puissance

Considérons ensuite *les formules approximatives de deuxième espèce*. Ici, la probabilité p tend vers zéro et le nombre N des bactéries mères augmente indéfiniment, de telle sorte que leur produit Np reste constant, le temps t restant également le même. Nous nous servirons du résultat suivant ⁽⁵⁾ : soit $|\zeta_N| \leq \frac{1}{2} N$; nous aurons alors

$$N \log \left(1 + \frac{\zeta_N}{N} \right) = \zeta_N + o \frac{\zeta_N^2}{N},$$

où le facteur complexe o aura une valeur absolue qui ne dépassera pas l'unité. Par conséquent,

$$(54) \quad \lim_{N \rightarrow \infty} \left(1 + \frac{\zeta_N}{N} \right)^N = \exp \left(\lim_{N \rightarrow \infty} \zeta_N \right)$$

lorsque la limite existe à droite.

Or, lorsqu'il y a N bactéries mères, la fonction génératrice sera

$$\exp \left[Np \int_T^1 \Phi(z, U) dU \right] \left\{ 1 + \frac{1}{N} Np \left[e^{\lambda t} - 1 - q e^{\lambda t} \int_T^1 \frac{\Omega(z, U) - 1}{p} dU \right] \right\}^{-N}.$$

La valeur absolue de z étant inférieure ou égale à l'unité, nous aurons

$$\begin{aligned} & \left| \int_T^1 dU \int_U^1 \Phi(z, V) dV \int_V^1 p \Phi(z, W) \Omega(z, W) dW \right| \\ & \leq \int_T^1 dU \int_U^1 \frac{dV}{V} \int_V^1 \frac{p dW}{W^{1+p}} \\ & \leq \left(\frac{T}{q} \frac{1 - T^{-p}}{p} - T \log T \right) + \frac{p(1 - T)}{q}. \end{aligned}$$

⁽⁴⁾ Voir FELLER (1950), p. 232.

⁽⁵⁾ Cf. CRAMÉR (1946), p. 217.

qui tend vers zéro et, par conséquent,

$$\begin{aligned} \frac{1}{p} \int_T^1 [\Omega(z, U) - 1] dU &= \int_T^1 dU \int_U^1 \Phi(z, V) \Omega(z, V) dV \\ &= \int_T^1 (V - T) \Phi(z, V) dV \\ &\quad + \int_T^1 dU \int_U^1 \Phi(z, V) dV \int_V^1 p \Phi(z, W) \Omega(z, W) dW \\ &\rightarrow \int_T^1 (V - T) \Phi(z, V) dV. \end{aligned}$$

Donc, nous obtenons la formule limite de deuxième espèce

$$(55) \quad \underline{\lim_2 [\psi(z, t)]^N = \exp \{ -Np\lambda [1 - \varphi(z, t)] \star e^{\lambda t} \}}.$$

Lorsqu'il s'agit d'un processus totalement markovien (c'est-à-dire, un processus du type M/M/o), ces deux formules limites [(53) et (55)] se relient aux formules données antérieurement par Bartlett (1951 b).

2.5. Comparaison avec les solutions des types D/G/o, + de Harris.

— Si les bactéries du type normal croissent selon un processus à phases multiples uniformes, s'il y a une seule bactérie mère fraîchement née à l'époque zéro et si le nombre k des phases a une valeur très grande, nous aurons un mode de croissance presque complètement déterministe : chaque division aura lieu à une époque qui devra être à peu près un multiple entier de $\frac{1}{\lambda}$ et la solution sera approximativement d'un « temps » discret. Du point de vue mathématique, il sera plus commode de supposer un mode de croissance des bactéries normales qui soit à la fois déterministe et *continu*, en écrivant pour leur nombre, à un instant quelconque,

$$(56) \quad l_t \equiv A e^{\rho t},$$

où ρ est le taux relatif de croissance pour les bactéries normales et A est leur nombre initial.

Soit $\varepsilon A e^{\rho u} \rho du$ la probabilité pour que dans l'intervalle infinitésimal $(u, u + du)$ il y ait une mutation, sans aucune diminution du nombre

des bactéries normales. Peut-être n'y aura-t-il pas de mutations dans l'intervalle $(0, u)$; voici la probabilité d'un tel événement :

$$(57) \quad P_u \equiv \exp[-\varepsilon \Lambda (e^{\rho u} - 1)].$$

Soit $\psi(z, t, \varepsilon, \Lambda)$ la fonction génératrice de la distribution du nombre des bactéries résistantes à une époque t ; nous aurons pour cette fonction l'équation intégrale de Harris

$$(58) \quad \psi(z, t, \varepsilon, \Lambda) = \int_0^t \psi(z, t-u, \varepsilon, \Lambda e^{\rho u}) \varphi(z, t-u) P_u \varepsilon \Lambda e^{\rho u} \rho \, du + P_t.$$

La méthode de solution de Harris utilise l'idée ingénieuse d'une « temps », (t, Λ) , à deux dimensions. On peut autrement résoudre l'équation comme il suit.

Il est évident que la fonction ψ ne dépend des quantités ε et Λ que comme fonction de leur produit. Le taux de croissance, ρ , étant fixé, écrivons

$$\begin{aligned} \psi(z, t, \varepsilon, \Lambda) &\equiv e^{\varepsilon \Lambda} f(z, \varepsilon \Lambda e^{\rho t}, \varepsilon \Lambda), \\ \varphi(z, t) &\equiv h(z, e^{\rho t}) \end{aligned}$$

et

$$\alpha \equiv \varepsilon \Lambda, \quad x \equiv \varepsilon \Lambda e^{\rho t}, \quad \gamma \equiv \varepsilon \Lambda e^{\rho u}.$$

L'équation de Harris deviendra

$$f(z, x, \alpha) = \int_{\alpha}^x f(z, x, \gamma) h\left(z, \frac{x}{\gamma}\right) d\gamma + e^{-x},$$

ce qui équivaut à l'équation différentielle

$$\begin{aligned} \frac{df}{d\alpha} &= -h\left(z, \frac{x}{\alpha}\right) f(z, x, \alpha), \\ f(z, x, x) &= e^{-x}. \end{aligned}$$

La solution, unique, est

$$f(z, x, \alpha) = \exp\left[\int_{\alpha}^x h\left(z, \frac{x}{\gamma}\right) d\gamma - x\right]$$

ou

$$(59) \quad \psi(z, t, \varepsilon, \Lambda) = \exp\{-\Lambda \varepsilon \rho [1 - \varphi(z, t)] \star e^{\rho t}\}$$

et c'est la solution du type D/G/0, + de Harris (1951).

La similitude avec notre solution du type M/G/0, + dans les condi-

tions de limite de *deuxième* espèce [éq. (55)] est frappante, mais non inattendue. En effet, il est tout à fait naturel que le mode de croissance des bactéries du type normal n'ait que peu d'effet sur la distribution du nombre des bactéries résistantes, lorsque le nombre des bactéries mères est très grand et le taux de mutation très petit. Notons que pour faire une comparaison précise avec la solution du type M/G/o, +, il faut poser

$$\rho = q\lambda \quad \text{et} \quad \varepsilon = \frac{p}{q},$$

de sorte que $A\varepsilon\rho = Np\lambda$. Dans les conditions de limite de deuxième espèce, il nous est permis d'écrire $\rho = \lambda$, et alors les formules (55) et (59) deviennent identiques.

Par contre, les solutions des types D/G/o, + et M/G/o, + sont tout à fait différentes lorsqu'on se place dans les conditions de limite de *première* espèce. Il sera intéressant de considérer quels sont les cas expérimentaux qui permettent l'application des formules limites de l'une ou de l'autre espèce.

2.6. Étude des solutions asymptotiques des deux espèces. —

1° *Limite de première espèce : solutions des types M/G/o, +.* — S'il y a une seule bactérie mère, la fonction génératrice de la distribution du nombre de bactéries résistantes sera celle donnée par la formule (53). S'il y en a N, il faudra considérer la même fonction élevée à sa N^{ième} puissance. Si l'on veut obtenir la probabilité P₀ pour qu'il n'existe pas de bactéries résistantes, il suffira de donner la valeur zéro à z; en particulier, s'il n'y a pas de délai phénotypique, nous aurons $\varphi(0, t) = 0$ et alors

$$(60) \quad \lim_1 P_0 = (1 + p e^{\lambda t})^{-N}.$$

Soit $x_1(t)$ et $x_2(t)$ la moyenne et la variance de la distribution associée à la fonction génératrice $\varphi(z, t)$ (qu'il y ait un délai ou non). La différentiation nous donne

$$(61) \quad Np e^{\lambda t} \int_0^\infty x_1(v) e^{-\lambda v} \lambda dv$$

pour le nombre moyen des bactéries résistantes et

$$(62) \quad Np e^{\lambda t} \int_0^\infty \{x_2(v) + [x_1(v)]^2\} e^{-\lambda v} \lambda dv + N \left[p e^{\lambda t} \int_0^\infty x_1(v) e^{-\lambda v} \lambda dv \right]^2$$

pour la variance du nombre de telles bactéries. En général, $x_1(t)$ sera de la forme $c e^{\rho t}$, avec une valeur de ρ à peu près égale à λ . Donc la distribution limite de première espèce aura une variance infinie, et peut-être sa moyenne sera-t-elle également infinie.

2° *Limite de deuxième espèce : solutions des types G/G/o, +.* — A la limite de deuxième espèce, nous venons de voir qu'il n'y a aucune différence entre les solutions des types M/G/o, + et D/G/o, +, ce qui nous suggère (comme Harris l'a noté, lui aussi) que dans ce cas limite nous avons trouvé *la solution générale* G/G/o, +.

Comme dans le cas de la limite de première espèce, la probabilité de ne pas trouver de bactéries résistantes a une forme très simple lorsqu'il n'y a pas de délai.

$$(63) \quad \lim_2 P_0 = \exp[-Np(e^{\lambda t} - 1)].$$

Introduisons les moments factoriels,

$$\mu_{[s]}(t) \equiv \left(\frac{\partial}{\partial z}\right)^s \varphi_{\lambda}^*(1 - o, t),$$

pour la distribution dont φ est la fonction génératrice, et les « demi-invariants factoriels »

$$z_{[s]}^*(t) \equiv \left(\frac{\partial}{\partial z}\right)^s \log[\psi(1 - o, t)]^N,$$

pour la distribution du nombre des bactéries résistantes. Nous aurons

$$(64) \quad \lim_2 z_{[s]}^*(t) = Np \int_0^t \mu_{[s]}(t - u) e^{\lambda u} \lambda \, du \quad (s \geq 1).$$

C'est la forme la plus commode d'une relation dont Harris a noté l'existence pour sa solution du type D/G/o.

Il est à noter qu'à la limite de deuxième espèce nous avons

$$\lim_2 [\psi(z, o)]^N = 1,$$

de sorte que la distribution limite se conforme à la condition initiale d'absence de bactéries résistantes à l'époque zéro.

3° *Solutions limites des deux espèces dans le cas M/M/o.* — Dans le cas M/M/o de Bartlett, nous connaissons la solution exacte; elle est

donnée par sa formule (52). Voici les solutions approximatives des deux types :

Première espèce (une bactérie mère) :

$$(65) \quad \lim_1 \psi(z, t) = \left[1 - p e^{\lambda t} \left(\frac{1-z}{z} \right) \log(1-z) \right]^{-1};$$

Deuxième espèce :

$$(66) \quad \lim_2 [\psi(z, t)]^N = [1 - z(1 - e^{-\lambda t})]^{N p e^{\lambda t} \frac{1-z}{z}}.$$

Plusieurs formules de ce genre ont été données par Bartlett (1951 *b*), dans une communication grâce à laquelle j'ai entrevu la possibilité d'arriver aux solutions plus générales que je viens de donner. Les solutions asymptotiques dans les cas D/D/o et D/M/o avaient été déjà données par Lea et par Coulson en 1949 ⁽⁶⁾.

2.7. L'effet d'un délai phénotypique : solution du type M/M +. — Armitage (1951) a considéré plusieurs mécanismes de délai phénotypique pour le processus de mutation du type D/D. Les formules de ce chapitre permettront d'ajouter quelques autres résultats à ceux d'Armitage : en voici un exemple :

Soit π la probabilité pour qu'une mutation produise immédiatement une bactérie résistante (c'est-à-dire sans aucun délai). Supposons qu'après la division d'une bactérie mutante *non* résistante il y ait deux cas possibles et deux seulement :

- a.* Premier cas, probabilité égale à π : une seule des deux nouvelles bactéries est résistante;
- b.* Deuxième cas, probabilité égale à $1 - \pi$: ni l'une ni l'autre des deux nouvelles bactéries n'est résistante.

Enfin, supposons que le mode de croissance pour les bactéries des deux types soit markovien et que la valeur de λ soit la même pour chacun des deux types. Nous aurons alors, pour la fonction $\varphi(z, t)$ de la formule (43), l'expression

$$(67) \quad \varphi(z, t) = \pi \varphi_0(z, t) + (1 - \pi) \varphi_1(z, t),$$

⁽⁶⁾ Il faut avertir le lecteur de la présence de quelques erreurs dans ce Mémoire; elles ont été corrigées par Armitage (1951) et Bartlett (1951 *b*).

où φ_0 est la fonction

$$\frac{z e^{-\lambda t}}{1 - z + z e^{-\lambda t}}$$

de Yule et de Furry et où φ_1 est la fonction

$$\frac{1}{1 + e^{\lambda t} \left(\frac{1-z}{z} \right) \left[1 - (1-z + z e^{-\lambda t}) \pi \right]}$$

de Bartlett [éq. (52)].

La fonction φ étant connue, la distribution du nombre de bactéries résistantes à l'époque t est déterminée par les formules (50) et (51).

Par exemple, trouvons la probabilité $P_0(\pi)$ pour que la colonie n'ait aucune bactérie résistante à cette époque. Nous avons

$$\varphi(0, t) = \frac{(1-\pi) e^{-\lambda t}}{\pi + (1-\pi) e^{-\lambda t}}$$

et

$$\chi(0, t) = \frac{e^{-\rho \lambda t}}{[\pi + (1-\pi) e^{-\lambda t}]^{\rho}},$$

et par conséquent

$$(68) \quad P_0(\pi) = \frac{1}{\left\{ 1 + \frac{\pi}{1-\pi} e^{\lambda t} [1 - (\pi + (1-\pi) e^{-\lambda t})^{\rho}] \right\}^N}.$$

Dans le cas étudié par Bartlett (pas de délai), la probabilité π est égale à l'unité, et alors

$$(69) \quad P_0(1) = \frac{1}{[1 + p(e^{\lambda t} - 1)]^N}.$$

En général, le temps t étant suffisamment grand, nous aurons

$$(70) \quad P_0(\pi) \simeq \frac{1}{\left[1 + p e^{\lambda t} \frac{\pi}{1-\pi} \log \frac{1}{\pi} \right]^N}.$$

Donc, si p est le véritable taux de mutation, et si \hat{p}_1 est la valeur donnée par une méthode pareille à la *première* méthode d'estimation de Luria et de Delbrück, nous aurons

$$(71) \quad \frac{p}{\hat{p}_1} \simeq \frac{1-\pi}{\pi \log \frac{1}{\pi}}$$

lorsque le nombre des colonies étudiées est assez grand pour éliminer les fluctuations statistiques.

Soit maintenant v_1 , v_2 et v_3 les nombres moyens des bactéries qui sont respectivement : 1° normales, 2° mutantes mais non résistantes, et 3° résistantes. Alors, nous aurons

$$(72) \quad \begin{cases} \frac{1}{\lambda} \frac{dv_1}{dt} = q v_1, \\ \frac{1}{\lambda} \frac{dv_2}{dt} = p(1-\pi)v_1 + (1-\pi)v_2, \\ \frac{1}{\lambda} \frac{dv_3}{dt} = p\pi v_1 + \pi v_2 + v_3, \end{cases}$$

de sorte que

$$v_1(t, \pi) = N e^{(1-p)\lambda t}$$

et

$$v_2(t, \pi) = N e^{\lambda t} \frac{p(1-\pi)}{\pi-p} (e^{-p\lambda t} - e^{-\pi\lambda t})$$

et, par conséquent, le nombre moyen des bactéries résistantes sera

$$(73) \quad v_3(t, \pi) = N e^{\lambda t} - N e^{\lambda t} \left[\frac{\pi(1-p)e^{-p\lambda t} - p(1-\pi)e^{-\pi\lambda t}}{\pi-p} \right].$$

Dans le cas étudié par Bartlett ($\pi = 1$), nous aurons

$$(74) \quad v_3(t, 1) = N e^{\lambda t} (1 - e^{-p\lambda t}) \simeq N e^{\lambda t} p \lambda t.$$

En général, nous aurons

$$(75) \quad v_3(t, \pi) \simeq N e^{\lambda t} p \lambda t \left[1 - (1-\pi) \frac{1 - e^{-\pi\lambda t}}{\pi\lambda t} \right].$$

Donc, si \hat{p}_2 est la valeur du taux de mutation donnée par une méthode semblable à la *deuxième* méthode d'estimation de Luria et de Delbrück, nous aurons

$$(76) \quad \frac{\hat{p}_2}{p} = 1 - (1-\pi) \frac{1 - e^{-\pi\lambda t}}{\pi\lambda t}$$

et, d'après (71),

$$(77) \quad \frac{\hat{p}_2}{\hat{p}_1} = \frac{1-\pi}{\pi \log \frac{1}{\pi}} \left[1 - (1-\pi) \frac{1 - e^{-\pi\lambda t}}{\pi\lambda t} \right].$$

Dans ces formules [(71), (76) et (77)] nous négligeons toutes les erreurs statistiques (qui seront très importantes, à moins que le nombre des colonies étudiées ne soit assez grand) et nous n'avons trouvé que les erreurs inévitables qui sont associées aux méthodes de Luria et de

Delbrück quand le délai phénotypique n'est pas égal à zéro. En pratique, le nombre des colonies étudiées peut être petit et il faudra tenir compte des erreurs statistiques et aussi de la possibilité que quelques-unes des bactéries mères peuvent être mutantes non résistantes. Il faut surtout tenir compte du fait que le mécanisme de délai étudié ici n'est qu'un mécanisme parmi plusieurs autres possibles ⁽¹⁾.

CHAPITRE III.

LE PROCESSUS DE NAISSANCE ET DE MORT ET LE PROBLÈME D'ESTIMATION DE SES PARAMÈTRES.

3.1. Nous ne nous sommes occupés jusqu'ici que des processus de naissance; considérons maintenant les processus de naissance *et de mort*. Les problèmes qui y sont associés ont été discutés par plusieurs auteurs ces cinq dernières années. Harris a donné récemment (1951) un excellent résumé des travaux américains et russes; je ne citerai donc que les résultats obtenus en Angleterre par Bartlett, Moran, Ramakrishnan et par moi-même, et en Belgique par Consael. Nous avons étudié pour la plupart quelques exemples d'un seul type, au lieu de chercher des résultats plus généraux mais moins détaillés. L'exemple dont je viens de parler fut décrit avec précision pour la première fois par Feller (1939) dans un article très important quoique peu connu; un problème de même nature était discuté il y a 40 ans par Mc Kendrick (1914). L'équation différentielle qui régit le développement du processus de Feller [discuté plus tard par Arley (1943) et par Arley et Borghsenius (1945)] a été résolue pour la première fois par Palm (1945). Depuis cette époque, ce processus a beaucoup attiré l'attention.

La théorie de la croissance stochastique d'une population a presque 80 ans d'existence; elle est née lors d'une discussion qui eut lieu en 1873 sur la question de l'extinction inévitable des noms de famille. C'est en cette année qu'Alphonse de Candolle (botaniste suisse né à Paris en 1806) a dit: « Évidemment, tous les noms doivent s'éteindre »; et de cette remarque ⁽²⁾, et des travaux de Galton et de Watson (1874)

⁽¹⁾ A ce sujet, voir Armitage (1951) et Kendall (1952 b).

⁽²⁾ Elle n'est pas exacte, à moins que l'on n'y ajoute la supposition que le nombre moyen des fils d'un seul homme soit inférieur ou égal à l'unité.

qui s'y étaient associés, est sortie la théorie moderne des processus d'embranchement. J'ai donné ailleurs (1949) l'histoire de cette théorie et je ne la répéterai pas ici; je me contenterai seulement de citer les conférences faites par M. J. F. Steffensen à l'Institut Henri Poincaré en décembre 1931, dans lesquelles il a donné la première solution exacte du problème central. Les solutions de Watson et de Steffensen et les développements donnés plus récemment par Harris (1948, 1951) et par Otter (1949) utilisent pour la plupart « un temps discret » qui est, en effet, le nombre ordinal de la génération dont il est question. Le processus de naissance et de mort dont je veux parler est tout à fait semblable à ces processus d'embranchement à temps discret, mais il s'exprime au moyen d'un temps continu.

3.2. Considérons d'abord une population homogène d'organismes qui sont entièrement indépendants les uns des autres. Pour chacun d'entre eux, soit

$$(78) \quad e^{-(\lambda+\mu)\tau}(\lambda + \mu) d\tau \quad (0 < \tau < \infty),$$

la distribution du « temps de repos » qui dure du présent jusqu'à la division ou à la mort de l'organisme (qu'elle arrive plus tôt ou plus tard, il n'importe). Enfin, soit $\frac{\lambda}{\lambda + \mu}$ la probabilité pour que la division se produise en premier; par conséquent $\frac{\mu}{\lambda + \mu}$ est la probabilité de mort sans qu'il y ait eu reproduction de l'organisme. Le processus ainsi défini est celui de Feller; c'est évidemment un processus markovien.

Supposons qu'il y ait un seul organisme à l'instant zéro, et soit n_t le nombre des organismes à un temps t quelconque. Si nous écrivons

$$(79) \quad \varphi(z, t) \equiv \mathcal{E} z^{n_t}$$

pour la fonction génératrice de la distribution du nombre aléatoire n_t , nous aurons

$$\varphi(z, t) = \int_0^t \{ \lambda [\varphi(z, t-u)]^2 + \mu \} e^{-(\lambda+\mu)u} du + z e^{-(\lambda+\mu)t},$$

équation intégrale que l'on peut transformer en équation différentielle,

$$(80) \quad \left. \begin{aligned} \frac{d\varphi}{dt} &= (\varphi - 1)(\lambda\varphi - \mu), \\ \varphi(z, 0) &= z. \end{aligned} \right\}$$

Nous obtenons donc la fonction génératrice

$$(81) \quad \varphi(z, t) = \frac{\mu(E-1) - (\mu E - \lambda)z}{(\lambda E - \mu) - \lambda(E-1)z},$$

de la distribution de Palm,

$$(82) \quad p_0(t) = \xi_t, \quad p_n(t) = [1 - p_0(t)](1 - \eta_t)\eta_t^{n-1} \quad (n \geq 1),$$

où

$$E \equiv e^{(\lambda - \mu)t}, \quad \frac{\xi_t}{\mu} \equiv \frac{\eta_t}{\lambda} \equiv \frac{E-1}{\lambda E - \mu}.$$

Lorsqu'il y a n organismes à l'instant zéro, il faut élever la fonction génératrice à sa $N^{\text{ième}}$ puissance. Ainsi, la probabilité d'extinction avant l'époque t sera

$$(83) \quad p_0(t) = \left[\frac{\mu(E-1)}{\lambda E - \mu} \right]^N$$

et la distribution de la durée du processus, T , sera

$$(84) \quad \frac{d}{dT} p_0(T) dT \quad (0 < T < \infty)$$

dans le cas d'une extinction inévitable. On voit, en effet, que l'extinction sera « presque certaine » quand λ est inférieur ou égal à μ , et que dans le cas contraire il y aura toujours une probabilité positive

$$(85) \quad \left(\frac{\mu}{\lambda} \right)^N$$

pour que la population s'éteigne.

Quand λ est égal à μ , il faut modifier les calculs ci-dessus, mais cela ne présente aucune difficulté; je ne citerai que la formule donnant $p_0(t)$

$$(86) \quad p_0(t) = \left(\frac{\lambda t}{1 + \lambda t} \right)^N \quad (\lambda = \mu).$$

La différentiation nous donne aisément la valeur moyenne et la variance du nombre aléatoire n_t

$$(87) \quad \mathcal{E} n_t = NE \quad \text{et} \quad \text{var}(n_t) = N \left(\frac{\lambda + \mu}{\lambda - \mu} \right) E(E-1).$$

Notons que le processus de Feller est *toujours* instable; même lorsque les paramètres λ et μ sont égaux, la population est condamnée à l'extinction inévitable, malgré la constance du nombre moyen des organismes.

C'est un résultat exactement analogue à celui de l'extinction des noms de famille, décrit par de Candolle et auquel Watson et Steffensen ont donné une forme quantitative.

Pour trouver un processus stable, il faut ajouter [comme je l'ai fait en 1948 (b)] un mécanisme de récupération que l'on peut nommer « l'immigration ». Notons que les paramètres λ et μ sont, en effet, les taux respectifs de natalité et de mortalité. Soit, de plus, κ le taux d'une immigration qui arrive au hasard; alors nous aurons l'équation intégrale

$$(88) \quad \varphi(z, t; \kappa) = \int_0^t \varphi(z, t-u) \varphi(z, t-u; \kappa) \kappa e^{-\kappa u} du + e^{-\kappa t}$$

pour la fonction génératrice, $\varphi(z, t; \kappa)$, de la distribution du nombre des organismes à un moment t quelconque. Ici l'on suppose qu'il n'y a pas d'organismes à l'instant zéro; la fonction $\varphi(z, t)$ qui apparaît dans la formule (88) est la fonction (81) ci-dessus. La différentiation nous donne

$$(89) \quad \begin{cases} \frac{\partial}{\partial t} \varphi(z, t; \kappa) = \kappa \varphi(z, t; \kappa) [\varphi(z, t) - 1], \\ \varphi(z, 0; \kappa) = 1; \end{cases}$$

donc nous aurons

$$(90) \quad \lim_{t \rightarrow \infty} \varphi(z, t; \kappa) = \left(\frac{\mu - \lambda z}{\mu - \lambda} \right)^{-\frac{\kappa}{\lambda}}$$

lorsque λ est inférieur à μ . Voilà donc un processus stable; la distribution du nombre des organismes est celle d'un binome négatif. Notons, de plus, le résultat

$$(91) \quad \varphi(z, t; \kappa) = (1 + \lambda t - \lambda t z)^{-\frac{\kappa}{\lambda}} \quad (\lambda = \mu).$$

Ces calculs équivalent à peu près à ceux que McKendrick a fait en 1914.

Il y a, en général, deux méthodes pour l'analyse d'un tel processus. Celle que nous avons utilisée jusqu'ici s'occupe des conditions initiales, et l'autre des conditions finales. Elles sont reliées [comme Bartlett (1951 a) l'a noté] aux équations différentielles « en avant » et « en arrière » de Kolmogoroff. Pour le processus simple de naissance et de mort, par exemple, la deuxième méthode nous donne pour la fonction génératrice l'équation différentielle

$$(92) \quad \frac{\partial \varphi}{\partial t} = (z - 1)(\lambda z - \mu) \frac{\partial \varphi}{\partial z}.$$

On peut aisément déduire cette équation en considérant ce qui peut arriver dans un intervalle infinitésimal de temps $(t, t + dt)$. Supposons qu'il y ait, à l'époque t , n organismes dans la colonie, et considérons les probabilités suivantes d'événements pouvant se trouver dans cet intervalle :

1° Une naissance : $n_{t+dt} = n + 1$; probabilité, $n\lambda dt$;

2° Une mort : $n_{t+dt} = n - 1$; probabilité, $n\mu dt$;

3° Un événement multiple; probabilité, $o(dt)$;

de sorte que la probabilité pour qu'un tel intervalle soit sans incidents sera $1 - n(\lambda + \mu) dt + o(dt)$. (Dans les cas plus généraux, il faut définir le processus de cette façon, chacun des taux λ et μ étant une fonction du nombre n des organismes et du temps t .) Si le nombre aléatoire n_t a une valeur n fixée à l'avance, nous aurons, comme moyenne conditionnelle,

$$\begin{aligned} \mathcal{E} z^{n_t+dt} &= z(n\lambda dt)z^n + \frac{1}{z}(n\mu dt)z^n + [1 - n(\lambda + \mu) dt]z^n + o(dt), \\ &= \left[\lambda dt z^2 \frac{\partial}{\partial z} + \mu dt \frac{\partial}{\partial z} + 1 - (\lambda + \mu) dt z \frac{\partial}{\partial z} \right] z^n + o(dt), \end{aligned}$$

et, par conséquent, en prenant les valeurs moyennes par rapport à n_t , nous aurons

$$\varphi(z, t + dt) = \varphi(z, t) + dt \left[\lambda z^2 \frac{\partial}{\partial z} + \mu \frac{\partial}{\partial z} - (\lambda + \mu) z \frac{\partial}{\partial z} \right] \varphi(z, t) + o(dt),$$

d'où découle sans difficulté aucune l'équation différentielle (92). Cette méthode de raisonnement est due à Bartlett (1946).

Chacune de ces méthodes a ses avantages. Nous n'aurons que la première méthode à notre disposition, à moins que le processus ne soit du type markovien. La deuxième méthode est obligatoire lorsque l'on veut discuter certains processus markoviens plus compliqués. Je l'ai utilisée en 1948 (a) pour déterminer la fonction génératrice lorsque les taux λ et μ sont donnés par n'importe quelle fonction positive du temps. Un problème encore plus général, celui qui consiste à déterminer la fonction génératrice lorsque l'on a

$$\lambda \equiv \lambda(t), \quad \mu \equiv \mu(t) \quad \text{et} \quad z \equiv z(t)$$

n'est pas encore résolu, quoique Consael (1948-1950) ait déjà discuté

plusieurs cas particuliers. Si les paramètres λ et μ sont constants, nous avons la solution très élégante de Bartlett (1949) :

$$(93) \quad \varphi(z, t; z) = \exp \left\{ - \int_0^t [1 - \varphi(z, u)] z(t-u) du \right\},$$

où $\varphi(z, t)$ est la fonction de la formule (81) et

$$\int_{t_1}^{t_2} z(t) dt$$

est le nombre moyen des organismes immigrés dans l'intervalle (t_1, t_2) ; le nombre réel des immigrés étant une variable aléatoire du type de Poisson.

3.3. Lorsque les organismes dans une population homogène dépendent les uns des autres, les problèmes de la croissance stochastique sont bien plus difficiles. Soient $\lambda(n)$ et $\mu(n)$ les taux instantanés de natalité et de mortalité lorsque le nombre total des organismes est égal à n ; appliquons la deuxième méthode. Nous trouverons pour la fonction génératrice l'équation différentielle

$$(94) \quad \frac{d\varphi}{dt} = \left(1 - \frac{1}{z}\right) \left[z\lambda\left(z \frac{\partial}{\partial z}\right) - \mu\left(z \frac{\partial}{\partial z}\right) \right] z \frac{d\varphi}{dz},$$

avec la condition initiale

$$\varphi(z, 0) = z^N.$$

Deux cas spéciaux ont été discutés. J'ai donné (1949) la solution limite (lorsque le temps t tend vers l'infini) pour le processus « logistique » dans lequel les taux sont

$$(95) \quad \lambda(n) \equiv \alpha(N_2 - n) \quad \text{et} \quad \mu(n) \equiv \beta(n - N_1) \quad (0 < N_1 \leq N \leq N_2)$$

(α et β étant constants). [C'est un processus que Feller (1939) a introduit pour généraliser un type bien connu de croissance déterministe.] Prendiville (1949) a donné la solution complète pour le processus dans lequel

$$(96) \quad \lambda(n) \equiv \alpha \left(\frac{N_2}{n} - 1 \right) \quad \text{et} \quad \mu(n) \equiv \beta \left(1 - \frac{N_1}{n} \right) \quad (0 < N_1 \leq N \leq N_2).$$

Voilà donc deux modèles possibles pour chacun des phénomènes biologiques suivants : le surpeuplement et le cannibalisme.

Notons que, chaque fois que le nombre des états possibles est fini (ce qui est vrai pour les deux processus ci-dessus), on peut appliquer un théorème qui se trouve dans le livre célèbre de M. Fréchet (1938) pour s'assurer qu'il y a vraiment une solution limite, $\varphi(z, \infty)$.

Supposons que, comme dans les processus étudiés par Prendiville et par moi-même, les fonctions $\lambda(n)$ et $\mu(n)$ soient égales à zéro en dehors de l'intervalle (N_1, N_2) et que dans cet intervalle elles soient des polynômes ayant les propriétés suivantes :

$$\lambda(n) > 0 \quad (0 < N_1 \leq n < N_2), \quad \lambda(N_2) = 0$$

et

$$\mu(n) > 0 \quad (0 < N_1 < n \leq N_2), \quad \mu(N_1) = 0.$$

La solution limite sera alors unique, et elle ne dépendra pas du nombre initial des organismes. Elle sera, de plus, une solution de l'équation différentielle

$$(97) \quad z \lambda(\delta) \delta \varphi = \mu(\delta) \delta \varphi$$

(où δ est l'opérateur $\frac{zd}{dz}$) pour laquelle il n'y aura qu'une seule solution admissible. Ces conclusions sont tirées de la théorie générale des chaînes finies de Markoff, exposée par M. Fréchet dans son traité déjà cité (2^e livre, p. 248-252).

3.4. Tournons-nous maintenant vers le problème qui est peut-être le plus difficile et certainement le plus intéressant de notre sujet : celui qui se pose lorsqu'il y a deux espèces d'organismes ayant une liaison biologique entre elles, et lorsque la loi de croissance du nombre des organismes de première espèce dépend stochastiquement du nombre de la deuxième espèce et réciproquement. Voici quelques exemples :

1^o Park (1948) (*) a fait une étude détaillée de la lutte entre les scarabées *Tribolium confusum* (Duval) et *Tribolium castaneum* (Herbst) dans ses populations expérimentales. La théorie mathématique de la lutte pour la vie du point de vue déterministe a été développée avec beaucoup d'élégance par Volterra, par Lotka et par bien d'autres, mais l'analyse des expériences de Park demande une théorie stochastique.

(*) Voir aussi T. PARK et MARIAN B. FRANK, *The population history of Tribolium free of sporozoan infection* (*J. Animal Ecology*, t. 19, 1950, p. 95-105).

2° Jusqu'à ces derniers temps, on a eu l'habitude, dans l'analyse démographique, de traiter l'espèce humaine comme si elle était composée d'éléments du sexe féminin seulement. Récemment, Karmel (1947-1948) a discuté « *le problème des deux sexes* » du point de vue déterministe, et je l'ai également discuté ailleurs (1949) du point de vue stochastique en signalant qu'il s'agit d'un processus du type décrit ci-dessus.

3° Bartlett (1949) et Bailey (1950) ont discuté le difficile problème d'une épidémie stochastique. Les difficultés qui y sont associées sont encore du même type.

Pour le problème d'une épidémie stochastique, Bartlett a donné l'équation différentielle résultant de la « deuxième méthode » [la même qui fournit l'équation (92) pour le processus de naissance et de mort]. Je veux indiquer maintenant pour « *le problème des deux espèces* » l'équation différentielle résultant de la « première méthode » [la même qui a donné l'équation (80) pour le processus de naissance et de mort]. Ni l'une ni l'autre de ces deux équations n'ont été jusqu'ici résolues. Il faut noter que Feller (1951) a discuté plusieurs problèmes d'une difficulté analogue, qui se posent en génétique [voir aussi Malécot (1949-1950)].

L'idée du processus que nous allons élaborer est contenue implicitement dans le travail de pionnier de Feller (1939). Bien que nous puissions construire les équations d'un processus très général de ce type, il suffira d'en donner l'exemple le plus simple. Soit pour les deux espèces la notation suivante :

Espèce.	Nombre initial.	Nombre au temps t .	Taux de naissance.	Taux de mortalité.
Rouge.....	M	m_t	λc	λn_t
Bleue.....	N	n_t	λc	λm_t

On voit aisément quelles sont les propriétés caractéristiques de notre processus :

- a. Il y a une symétrie parfaite entre les deux espèces ;
- b. Si les organismes rouges dépassent le nombre critique c , les organismes bleus auront un taux *net* de croissance qui sera négatif, et réciproquement.

c. Si les organismes rouges disparaissent, la colonie bleue croîtra indéfiniment, et réciproquement.

Soit

$$(98) \quad \varphi(z, w; t; M, N) \equiv \mathcal{E}(z^{m_t} w^{n_t} \mid m_0 = M, n_0 = N),$$

et soit, de plus,

$$(99) \quad \psi(z, w; t; Z, W) \equiv \sum_0^{\infty} \sum_0^{\infty} Z^M W^N \varphi(z, w; t; M, N),$$

de sorte que la condition initiale sera

$$(100) \quad \psi(z, w; 0; Z, W) \equiv \frac{1}{(1 - Zz)(1 - Ww)}.$$

Je vais maintenant donner une équation différentielle pour la fonction ψ , dans laquelle z et w ne seront que des paramètres.

Notons d'abord deux cas spéciaux importants :

1° $z = 0, w = 1$; la condition initiale devient

$$\psi(0, 1; 0; Z, W) = \frac{1}{1 - W},$$

et lorsque t vers l'infini, nous aurons

$$\psi(0, 1; \infty; Z, W) = \sum_0^{\infty} \sum_0^{\infty} Z^M W^N \text{prob}[m_{\infty} = 0 \mid M, N].$$

Donc, s'il y a au début M organismes rouges et N organismes bleus, la probabilité pour que l'espèce rouge s'éteigne sera le coefficient de $Z^M W^N$ dans la fonction limite, $\psi(0, 1; \infty; Z, W)$.

2° $z = 1, w = 0$, nous obtiendrons de la même façon la probabilité pour que l'espèce bleue disparaisse.

Notons, de plus, le cas (3°) : $z = 0, w = 0$. La condition initiale devient

$$\psi(0, 0; 0; Z, W) = 1$$

et nous aurons

$$\psi(0, 0; \infty; Z, W) = \sum_0^{\infty} \sum_0^{\infty} Z^M W^N \text{prob}[m_{\infty} = n_{\infty} = 0 \mid M, N].$$

L'intuition nous suggère que dans ce cas

$$(101) \quad \psi(0, 0; \infty; Z, W) = 1$$

(à cause de l'impossibilité d'une extinction simultanée, exception faite

du cas où $M = N = 0$). L'intuition nous suggère, de plus, que nous devons avoir

$$(102) \quad \psi(0, 1; \infty; Z, W) + \psi(1, 0; \infty; Z, W) = 1 + \frac{1}{(1-Z)(1-W)}$$

(puisque l'une seule des deux espèces doit s'éteindre, exception faite du cas où $M = N = 0$).

Le problème se réduit donc essentiellement à celui de déterminer la seule fonction

$$\psi(0, 1; \infty; Z, W),$$

mais on ne sait pas encore comment le faire.

Les arguments usuels nous donnent l'équation intégrale ⁽⁸⁾

$$(103) \quad \varphi(M, N) = \int_0^t [Mc\varphi(M+1, N) + MN\varphi(M-1, N) \\ + Nc\varphi(M, N+1) + MN\varphi(M, N-1)]_{t-u} \\ \times e^{-(Mc+Nc+2MN)\alpha u} du + z^M w^N e^{-(Mc+Nc+2MN)\alpha t},$$

de laquelle, par différentiation, nous obtenons l'équation différentielle

$$(104) \quad \frac{1}{z} \frac{d\varphi(M, N)}{dt} = Mc[\varphi(M+1, N) - \varphi(M, N)] - MN[\varphi(M, N) - \varphi(M-1, N)] \\ + Nc[\varphi(M, N+1) - \varphi(M, N)] - MN[\varphi(M, N) - \varphi(M, N-1)].$$

Si nous multiplions chaque équation par $Z^M W^N$ (avec les valeurs appropriées de M et de N), nous obtiendrons, en les ajoutant,

$$(105) \quad \frac{1}{z} \frac{\partial \psi}{\partial t} = c \left[(1-Z) \frac{\partial \psi}{\partial Z} - \frac{\psi(Z, W) - \psi(0, W)}{Z} \right] \\ + c \left[(1-W) \frac{\partial \psi}{\partial W} - \frac{\psi(Z, W) - \psi(Z, 0)}{W} \right] \\ + ZW \left(\frac{\partial \psi}{\partial Z} + \frac{\partial \psi}{\partial W} \right) - ZW(2-Z-W) \frac{\partial^2 \psi}{\partial Z \partial W}.$$

Les fonctions $\psi(0, W)$ et $\psi(Z, 0)$ qui se présentent dans l'équation différentielle sont connues. D'après la théorie du processus de Yule et de Furry (donnée dans le premier chapitre), nous avons

$$(106) \quad \psi(0, W) = \sum_{N=0}^{\infty} W^N \left[\frac{w e^{-\alpha ct}}{1 - w(1 - e^{-\alpha ct})} \right]^N \\ = \frac{1 - w(1 - e^{-\alpha ct})}{1 - w(1 - e^{-\alpha ct}) - wW e^{-\alpha ct}};$$

pour la fonction $\psi(Z, 0)$, nous avons une formule analogue.

⁽⁸⁾ J'ometts z, w et t dans $\varphi(z, w; t; M, N), \dots$, pour rendre l'écriture plus aisée.

Il est évidemment possible d'appliquer les théories de « la promenade au hasard » et des processus de diffusion, mais je ne discuterai pas cette possibilité ici [pour ces théories, voir Lévy (1948) et Feller (1951)].

3.5. Le problème d'estimation de tous les paramètres associés à un tel processus n'a été considéré que pour le processus *simple* de naissance et de mort, c'est-à-dire pour le premier processus que nous venons de discuter. Supposons d'abord que l'on ait observé la croissance réelle d'un grand nombre N de colonies. Les faits expérimentaux constitueront alors un échantillon d'une grandeur N et la méthode de « maximum likelihood » de Fisher sera applicable pourvu que le nombre des colonies soit assez grand. J'ai discuté ailleurs (1949) le processus simple de naissance de ce point de vue, mais je ne citerai pas ici les résultats ainsi obtenus parce qu'en pratique, le problème est tout à fait différent. En pratique, la grandeur N de l'échantillon est égale à l'unité, ce qui est un nombre bien trop petit pour que l'on puisse se servir des méthodes asymptotiques de Fisher. La question de la validité des méthodes fishériennes lorsqu'il s'agit d'une estimation associée à un processus stochastique a été étudiée d'une façon très générale par Wald (1948) et par Grenander (1950). Heureusement, dans le cas spécial du processus de naissance et de mort, il existe une solution exacte, aussi simple que complète et c'est de cette solution que nous allons nous occuper dans ce qui suit.

En premier lieu, il faut que l'on se rende compte du fait que le processus de naissance et de mort équivaut à un procédé d'échantillonnage réitéré, dont voici la forme précise :

1° Soit u une variable aléatoire dont la distribution a la forme suivante :

$$e^{-(\lambda+\mu)u} (\lambda + \mu) du \quad (0 < u < \infty);$$

2° Soit v une variable aléatoire discontinue, à deux valeurs possibles, dont la distribution est

$$p_0 = \frac{\mu}{\lambda + \mu}, \quad p_1 = \frac{\lambda}{\lambda + \mu};$$

3° Soit, de plus, les variables $u_1, v_1, u_2, v_2, \dots$ tout à fait indépendantes les unes des autres;

4° Choisissons les variables $u_1, v_1, u_2, v_2, \dots$ dans l'ordre indiqué, et calculons successivement les quantités $t_0, N_0; t_1, N_1; \dots$, en nous servant des formules

$$t_0 = 0; \quad N_0 = N;$$

$$t_r = t_{r-1} + \frac{u_r}{N_{r-1}} \quad (r \geq 1).$$

$$N_r = N_{r-1} + 2v_{r-1} \quad (r \geq 1);$$

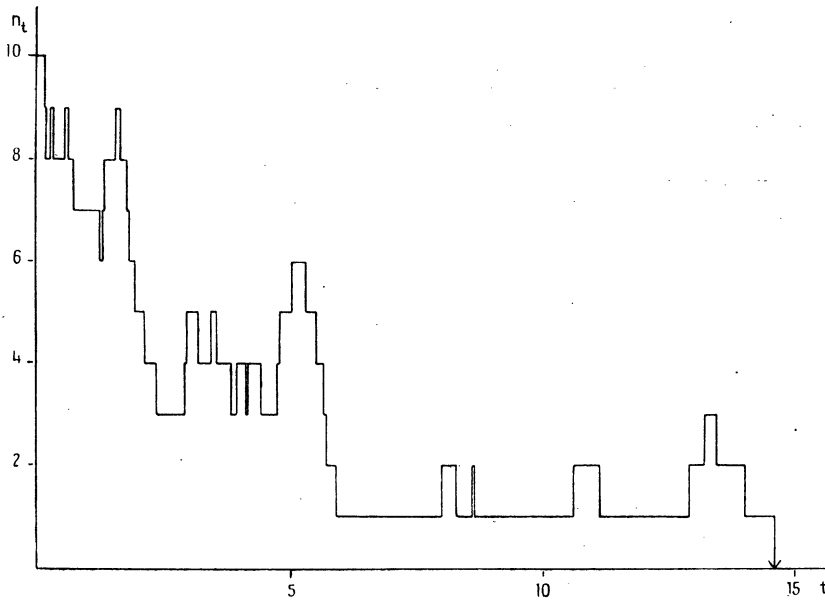


Fig. 3. — La grandeur n_t d'une population artificielle depuis le moment initial ($t = 0$) jusqu'à l'extinction ($\lambda = \frac{1}{3}$, $\mu = \frac{2}{3}$; $N = 10$; $T = 14,6$).

Évidemment, la fonction aléatoire définie par la règle

$$(107) \quad n_t \equiv N_r \quad (t_r \leq t < t_{r+1}; r \geq 0)$$

donnera le nombre des organismes dans une colonie croissante selon le processus de naissance et de mort.

J'ai décrit cette équivalence (1950a) en l'utilisant pour construire un exemple artificiel d'un tel processus. Dans la figure 3 le lecteur trouvera un registre du développement d'une telle population artificielle depuis le moment initial jusqu'à l'extinction. Moran (1951a) a démontré qu'on peut l'utiliser également dans une méthode d'estimation du paramètre $(\lambda + \mu)$.

La méthode de Moran est fondée sur l'idée essentielle qu'il faut accepter, dès le début même de l'expérience, une règle définitive suivant laquelle on se décidera à l'arrêter. Il y a autant de méthodes d'estimation qu'il y a de règles d'arrêt et celles-ci sont des deux types suivants :

a. La décision de continuer l'échantillonnage (ou l'expérience) ne dépend que des nombres des naissances et des morts qui ont déjà eu lieu ;

b. La décision de poursuivre l'expérience dépend aussi du temps qui s'est écoulé depuis le commencement.

Comme Moran, nous ne discuterons que les règles d'arrêt du premier type. Dans ces conditions, Moran a démontré que la quantité

$$\begin{aligned}
 (108) \quad 2(\lambda + \mu) \sum_{r=1}^E u_r &= 2(\lambda + \mu) \sum_{r=1}^E N_{r-1}(t_r - t_{r-1}) \\
 &= 2(\lambda + \mu) \sum_{r=1}^E \int_{t_{r-1}}^{t_r} n_t dt \\
 &= 2(\lambda + \mu) \int_0^T n_t dt
 \end{aligned}$$

aura la distribution conditionnelle de χ_{2E}^2 , le nombre $2E$ des degrés de liberté étant égal à deux fois le nombre total E des naissances et des morts qui ont eu lieu pendant la durée T de l'expérience. Cela nous permet de tirer une conclusion concernant la valeur du paramètre $(\lambda + \mu)$.

L'observation de la croissance « réelle » d'une d'entre mes populations artificielles m'a donné les valeurs « expérimentales » suivantes :

$$E = 80 \quad \text{et} \quad \int_0^T n_t dt = 73,86,$$

desquelles Moran a déduit « l'intervalle de confiance »

$$0,855 \leq \lambda + \mu \leq 1,33$$

pour $(\lambda + \mu)$, le degré de confiance que l'on y associe étant égal à 95 %. La valeur exacte était donnée par $\lambda + \mu = 1$. Voici les autres détails de « l'expérience » :

Nombre initial des organismes : 10 ;

Taux de naissance (λ) : $\frac{1}{2}$;

Taux de mort (μ) : $\frac{1}{2}$;

Règle d'arrêt : arrêter dès que 80 naissances-et-morts auront eu lieu ou que la population sera éteinte (selon que l'un ou l'autre arrive le premier).

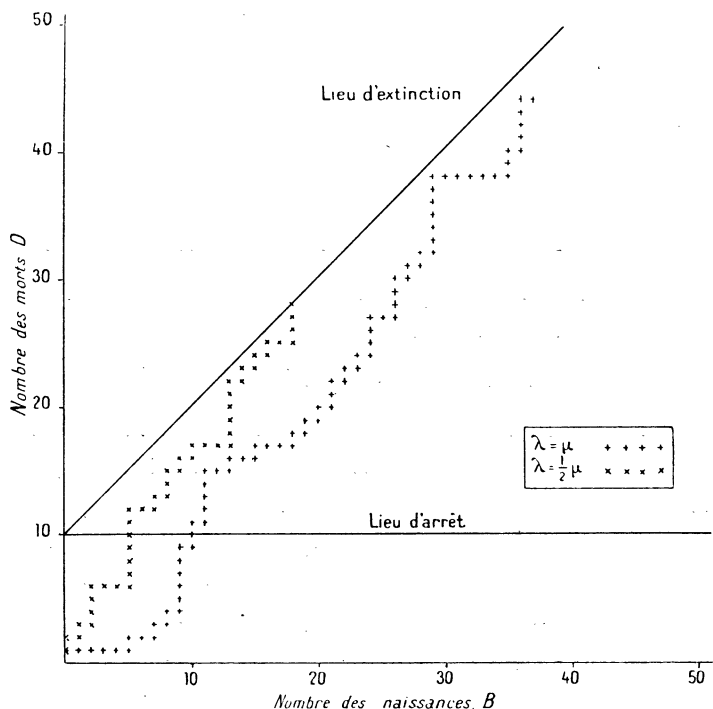


Fig. 4. — Registre des naissances et des morts pour deux populations artificielles.
(Nombre initial d'organismes = 10).

Notons que le résultat de Moran nous donne la solution complète du problème de l'estimation du paramètre unique d'un processus de naissance seulement.

Voyons maintenant si nous pouvons trouver une méthode pour l'estimation de l'autre paramètre $\frac{\lambda}{\mu}$, lorsqu'il s'agit d'un processus de naissance et de mort. Comme on le verra bientôt, une telle estimation est vraiment possible, et elle est tout à fait indépendante de l'estimation de

$(\lambda + \mu)$, mais les détails de la méthode dépendent de la nature précise de la règle d'arrêt.

Il suffit d'enregistrer les naissances et les morts successives sur un diagramme cartésien, en y portant après chaque incident le point dont les coordonnées sont le nombre cumulé des naissances, B, et le nombre cumulé des morts, D. Soit N le nombre initial des organismes. La droite

$$(109) \quad N + B - D = 0$$

sera alors *le lieu d'extinction* et, en général, elle constituera une partie du *lieu d'arrêt*.

Les propriétés statistiques d'un tel diagramme ne dépendent que du seul paramètre $\frac{\lambda}{\mu}$, et la totalité de l'information fournie par notre expérience concernant ce paramètre y sera contenue. En effet, le diagramme équivaut à un registre de l'échantillon $v_1, v_2, v_3, \dots, v_E$. Par conséquent, le problème de l'estimation du paramètre $\frac{\lambda}{\mu}$ se réduit à celui de la détermination de *la distribution des points d'arrêt sur le lieu d'arrêt*. Une solution *exacte* n'est possible que si la règle a une forme assez simple.

Je ne veux discuter ici qu'un seul exemple. [Les solutions plus générales sont l'objet d'une étude que M. F. J. Anscombe (*) et ses collaborateurs font actuellement à Cambridge; plusieurs résultats du même genre nous sont aussi promis par M. E. R. Immel, en Californie.] Voici l'exemple dont je viens de parler :

(110) *Règle d'arrêt*. — L'expérience se termine dès que le nombre cumulé des morts, D, sera égal au nombre initial des organismes, N.

Il s'agit en ce cas d'un échantillonnage binomial inverse (Haldane, 1945 *a, b*; Tweedie, 1945; Finney, 1947, 1949) et, comme l'indique le diagramme, le lieu d'extinction est tout à fait inaccessible (*). Nous pouvons nous servir immédiatement d'un théorème de Finney (1949) pour déterminer un intervalle de confiance pour le rapport des taux, $\frac{\lambda}{\mu}$.

(*) F. J. ANSCOMBE, *Sequential Estimation* [*J. Roy. Statist. Soc. (B)*, sous presse].

(*) Exception faite du point (0, N).

Nous aurons, en effet,

$$\text{prob} \left[B \leq b \mid N, \frac{\lambda}{\mu} \right] = \sum_{d=N}^{N+b} {}^{N+b}C_d \frac{\mu^d \lambda^{N+b-d}}{(\lambda + \mu)^{N+b}}$$

et l'on n'a qu'à se servir d'un tableau (*) de la distribution binomiale si l'on veut construire approximativement un intervalle de confiance

pour $\frac{\mu}{\lambda + \mu}$.

Je donnerai enfin deux exemples numériques fondés sur l'observation de la croissance « réelle » des deux populations artificielles que j'ai construit à cet effet :

	Exemple I.	Exemple II.
Les taux $\lambda (\mu)$	$\frac{1}{2} \left(\frac{1}{2} \right)$	$\frac{1}{3} \left(\frac{2}{3} \right)$
Nombre initial N.....	10	10
Nombre des naissances B.....	10	5
Nombre des morts D.....	10	10
Intervalle de confiance (95 %) pour $\frac{\mu}{\lambda + \mu}$	0,27; 0,71	0,38; 0,87
Valeur exacte de $\frac{\mu}{\lambda + \mu}$	0,50	0,67

CHAPITRE IV.

LA MÉTHODE DE LA FONCTIONNELLE CARACTÉRISTIQUE DANS L'ANALYSE DES PROCESSUS STOCHASTIQUES MULTIPLICATIFS.

4.1. L'idée de décrire un processus stochastique à l'aide d'une *fonctionnelle caractéristique* se trouve pour la première fois dans les travaux indépendants de Bochner (1947) et de Le Cam (1947). Bochner étudiait les processus stochastiques d'un point de vue très général, mais c'est dans l'étude de Le Cam concernant l'effet des chutes de pluie sur le débit d'un fleuve que l'on trouvera le point de vue le plus proche de celui que nous choisissons ici. Il faut noter aussi le travail de

(*) Celui de Hartley et de Fitch (1951) est très commode à cet effet.

M^{lle} Mourier (1949-1951), car je crois qu'il sera d'une grande importance pour les problèmes que nous allons discuter.

Considérons encore le processus simple de naissance et de mort, et distinguons entre les organismes selon leurs âges respectifs. Ici l'on utilise le terme « âge » dans un sens *tout à fait différent* de celui dont je me suis servi dans le premier chapitre (qui était celui utilisé dans l'étude de Bellman et Harris). Il serait facile de confondre les questions dont je veux parler avec celles discutées dans un article récent de Harris (1951), mais il faut mettre l'accent sur le fait que par « distribution des âges » nous entendons, en général l'un et l'autre, des choses bien différentes. Soit t_1 l'époque de la première division dans un processus markovien de naissance seulement. Du point de vue de Harris, un organisme à l'âge t_1 meurt à cet instant, et il est remplacé par deux organismes nouveaux à l'âge zéro. De notre point de vue, un organisme à l'âge t_1 continue à vivre, et un seul organisme nouveau à l'âge zéro s'y ajoute.

Voici une représentation graphique; elle montre la distribution des âges (selon les deux définitions) avant et après la première et la deuxième division dans un processus de naissance seulement :

Première définition (Bellman et Harris) :

$$0 \rightarrow t_1 \left| \begin{array}{c} 0 \rightarrow t_2 \\ 0 \end{array} \right. \left| \begin{array}{c} t_2 \rightarrow \dots \\ 0 \\ 0 \end{array} \right.$$

Deuxième définition (Bartlett et Kendall) :

$$0 \rightarrow t_1 \left| \begin{array}{c} t_1 \rightarrow t_1 + t_2 \\ 0 \end{array} \right. \left| \begin{array}{c} t_1 + t_2 \rightarrow \dots \\ t_2 \\ 0 \end{array} \right.$$

C'est la deuxième définition dont il est question ici et le processus stochastique sera un processus simple de naissance et de mort à paramètres constants, λ et μ .

Supposons qu'à l'époque initiale il y ait un seul organisme, dont l'âge initial est égal à X . Soit $N(x, t)$ le nombre des organismes à l'époque t dont l'âge est inférieur à la valeur x . Nous aurons alors

$$N(x, 0) \equiv 0 \quad (x \leq X), \quad N(x, 0) \equiv 1 \quad (x > X),$$

et à un instant quelconque le nombre des organismes dont les âges se

trouvent dans l'intervalle $x_1 \leq x < x_2$ sera donné par l'intégrale de Stieltjes

$$\int_{x_1}^{x_2} d_x N(x, t).$$

Discutons maintenant le problème posé par Bartlett en 1946 :

L'instant t étant donné, quelle est la loi de répartition de la fonction aléatoire $N(x, t)$?

La solution qui suit est celle que j'ai donnée en 1950.

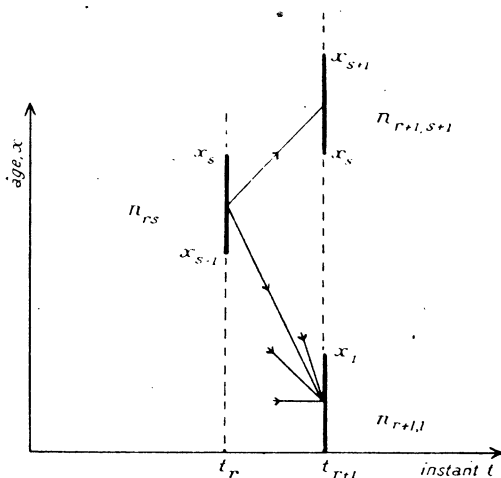


Fig. 5. — Le développement de la distribution des âges.

Introduisons la fonctionnelle caractéristique

$$(111) \quad G[\theta(x); t] \equiv \mathcal{E} \exp [i \Sigma_t \theta(x)].$$

Ici

$$\Sigma_t \theta(x) \equiv \theta(x_1) + \theta(x_2) + \theta(x_3) + \dots,$$

où (x_1, x_2, x_3, \dots) est l'ensemble des âges des organismes qui se trouvent dans la population à l'époque t et $\theta(x)$ est une fonction réelle arbitraire (mais elle doit être bornée et intégrable au sens de Riemann dans chaque intervalle fini). Nous trouverons que la fonctionnelle caractéristique est donnée par

$$(112) \quad G[\theta(x); t] = 1 + \left\{ \frac{[e^{i\theta(x+t)} - 1] + [e^{i\theta(t)} - 1] \star \lambda e^{\lambda t}}{e^{\mu t} - [e^{i\theta(t)} - 1] \star \lambda \left(\frac{\lambda e^{\lambda t} - \mu e^{\mu t}}{\lambda - \mu} \right)} \right\}$$

4. 2. Pour arriver au résultat ci-dessus, considérons d'abord un problème analogue formulé par rapport à un « temps » *discret*. A un instant $t_r \equiv r\delta$, l_r sera égal à l'unité si l'organisme générateur est encore en vie, et égal à zéro dans le cas contraire.

Soit $n_{r,s}$ le nombre des organismes qui à l'instant t_r ont un âge situé dans l'intervalle

$$(s-1)\delta \leq x < s\delta \quad (s = 1, 2, \dots, r)$$

et posons enfin

$$N_r \equiv \sum_{s=1}^r n_{r,s} + l_r$$

(qui représente la grandeur de la colonie à l'instant t_r). Évidemment; nous aurons $N_0 = l_0 = 1$. (Je conseille au lecteur de considérer la figure 5 avant de continuer.)

Les conditions à l'instant t_r étant données, les distributions conditionnelles qui se rapportent à l'instant ultérieur t_{r+1} , seront celles que voici :

1° *a.* Si la variable l_r est égale à zéro, on aura $l_{r+1} = 0$ également;

b. Si la variable l_r est égale à l'unité, la distribution de la variable l_{r+1} sera

$$p_0 = 1 - e^{-\mu\delta}, \quad p_1 = e^{-\mu\delta};$$

2° *a.* Si la variable $n_{r,s}$ est égale à zéro, on aura également

$$n_{r+1,s+1} = 0;$$

b. Si la variable $n_{r,s}$ a une valeur positive, la distribution de la variable $n_{r+1,s+1}$ sera celle engendrée par le binôme

$$(1 - e^{-\mu\delta} + e^{-\mu\delta})^{n_{r,s}};$$

3° *a.* Si la variable N_r est égale à zéro, on aura

$$n_{r+1,1} = 0;$$

b. Lorsque la grandeur N_r de la colonie est positive, on peut déterminer comme suit la distribution du nombre d'organismes nouveau-nés.

Classifions les organismes nouveau-nés selon les âges de leurs organismes générateurs : soit $m_{r+1,s}$ le nombre d'organismes nouveau-nés qui seront les produits des $n_{r,s}$ organismes âgés de $(s-1)\delta$ à $s\delta$ à

l'instant t_r , et soit $m_{r+1,0}$ le nombre d'organismes nouveau-nés qui seront les produits de l'organisme générateur primitif dans l'intervalle (t_r, t_{r+1}) . Nous aurons alors

$$(113) \quad n_{r+1,1} = m_{r+1,0} + \sum_{s=1}^r m_{r+1,s}$$

et il y aura une liaison stochastique entre t_{r+1} et $m_{r+1,0}$, entre $n_{r+1,2}$ et $m_{r+1,1}$, entre $n_{r+1,3}$ et $m_{r+1,2}$, et ainsi de suite. A part cela, ces variables seront tout à fait indépendantes. Or, la distribution conditionnelle du couple $(m_{r+1,1}, n_{r+1,2})$, par exemple, sera celle engendrée par la fonction génératrice

$$(114) \quad [\psi(w; z)]^{n_{r+1,2}} \equiv \mathcal{E}(w^{n_{r+1,2}} z^{m_{r+1,1}} | n_{r+1,2}) = \left[1 + \frac{(w-1) + (z-1)(e^{\lambda\delta} - 1)}{e^{\mu\delta} - (z-1)\lambda \left(\frac{e^{\lambda\delta} - e^{\mu\delta}}{\lambda - \mu} \right)} \right]^{n_{r+1,2}}$$

(Pour la démonstration, voir mon article de 1950; les principes dont je me suis servi sont ceux décrits dans le troisième chapitre.)

Soit maintenant

$$\varphi_r(w; z_1, \dots, z_r) \equiv \mathcal{E}(w^{l_r} z_1^{n_{r,1}} \dots z_r^{n_{r,r}})$$

la fonction génératrice *absolue* des variables $l_r, n_{r,1}, \dots, n_{r,r}$. Nous aurons évidemment

$$\begin{aligned} \varphi_{r+1}(w; z_1, \dots, z_{r+1}) &= \mathcal{E}[\psi(w; z_1)]^{l_r} [\psi(z_2; z_1)]^{n_{r,1}} \dots [\psi(z_{r+1}; z_1)]^{n_{r,r}}, \\ &= \varphi_r[\psi(w; z_1); \psi(z_2; z_1), \dots, \psi(z_{r+1}; z_1)] \quad (r \geq 1), \end{aligned}$$

et

$$(115) \quad \varphi_1(w; z_1) = \psi(w; z_1),$$

d'où l'on peut déterminer les fonctions génératrices l'une après l'autre. En effet, une argumentation inductive prouve que nous aurons

$$(116) \quad \varphi_r(w; z_1, \dots, z_r) = 1 + \left[\frac{(w-1) + \sum_{s=1}^r (z_s-1) \alpha_s^{(r)}}{e^{\mu\delta} - \sum_{s=1}^r (z_s-1) \beta_s^{(r)}} \right],$$

où les coefficients α et β restent à déterminer. Ces coefficients sont les solutions uniques de certaines équations à différences finies dont j'ai

donné une discussion complète dans mon article déjà cité. Lorsque λ et μ sont égaux, nous aurons

$$\alpha_s^{(r)} = e^{(r-s)\lambda\delta}(e^{\lambda\delta} - 1)$$

et

$$(117) \quad \beta_s^{(r)} = \lambda\delta [e^{\lambda\delta} + (r-s)(e^{\lambda\delta} - 1)] e^{(r-s)\lambda\delta};$$

dans le cas général, les formules sont un peu plus complexes.

Écrivons maintenant

$$\begin{array}{ll} t \equiv t_r & \text{et} \quad x_s \equiv s\delta \quad (0 \leq s \leq r), \\ w \equiv e^{i\theta(X+t)} & \text{et} \quad z_s \equiv e^{i\theta(x_{s-1})} \quad (1 \leq s \leq r) \end{array}$$

et définissons

$$(118) \quad \left\{ \begin{array}{ll} \theta(x|\delta) \equiv \theta(x_{s-1}) & (x_{s-1} \leq x < x_s; 1 \leq s \leq r), \\ \equiv \theta(x) & (x \geq t). \end{array} \right.$$

Nous aurons alors

$$(119) \quad C[\theta(x|\delta); t] = \varphi_r(w; z_1, \dots, z_r)$$

et, de plus, à cause de l'intégrabilité des fonctions $\cos\{\theta(x)\}$ et $\sin\{\theta(x)\}$, nous trouverons pour

$$\lim_{\delta \rightarrow 0} \varphi_r(w; z_1, \dots, z_r)$$

l'expression donnée au second membre de la formule (112). Donc, pour prouver l'exactitude de cette formule, il suffira de démontrer l'égalité

$$(120) \quad \lim_{\delta \rightarrow 0} C[\theta(x|\delta); t] = C[\theta(x); t].$$

4.3. Soit maintenant

$$U \equiv \Sigma_t \theta(x)$$

et soit, de plus,

$$U_\delta \equiv \Sigma_t \theta(x|\delta).$$

Nous aurons alors

$$|U_\delta - U| \leq \sum_{s=1}^r n_{rs} \omega_s,$$

où la quantité ω_s est l'oscillation de la fonction $\theta(x)$ dans le $s^{\text{ième}}$ intervalle $x_{s-1} \leq x < x_s$.

Or, d'après les calculs ci-dessus, nous aurons

$$\mathcal{E} n_{rs} = \alpha_s^{(r)} e^{-\mu r \delta} < e^{(\lambda - \mu) r \delta} \lambda \delta$$

et, par conséquent,

$$\mathcal{E} (|U_\delta - U|) \leq \lambda e^{(\lambda - \mu)t} \sum_{s=1}^r \omega_s (x_s - x_{s-1}).$$

Mais l'expression de droite tend vers zéro avec δ , à cause de l'intégrabilité de $\theta(x)$; donc

$$\lim_{\delta > 0} \mathcal{E} (|U_\delta - U|) = 0.$$

J'écrirai ρ pour une quantité complexe dont le module ne dépasse pas l'unité; ρ ne représentera pas nécessairement la même quantité chaque fois qu'on le rencontrera. Nous avons, la fonction $\theta(x)$ étant réelle,

$$e^{iU_\delta} = e^{iU} + \rho |U_\delta - U|$$

et, par conséquent,

$$\mathcal{E} e^{iU_\delta} = \mathcal{E} e^{iU} + \rho \mathcal{E} (|U_\delta - U|).$$

Mais

$$C[\theta(x); t] = \mathcal{E} e^{tU} \quad \text{et} \quad C[\theta(x|\delta); t] = \mathcal{E} e^{iU_\delta},$$

d'où l'on déduit le résultat cherché (120).

Un détail dans la démonstration ci-dessus mérite d'être un peu plus développé : c'est le fait que nous nous sommes servis des valeurs moyennes des expressions telles que $|U_\delta - U|$ et e^{iU} , dont la variable U fait partie. Or, toute mention de la variable U doit être précédée de la définition d'une loi de probabilité dans l'espace abstrait dont la fonction $N(x, t)$ est un *point*. On peut se rapporter à un Mémoire récent de Bhabha (1950), qui traite des processus multiplicatifs associés au rayonnement cosmique. Il faut noter, de plus, quelques observations de Moyal (1949) et enfin, les travaux de M^{lle} Mourier (1949-1951) sur les lois de probabilité dans un espace de Banach seront peut-être applicables dans ce cas. D'autre part, on peut considérer la formule (112) comme une *définition même* du processus de naissance et de mort. Il serait intéressant d'entrer plus avant dans ces questions, mais je n'ai pas l'intention de le faire ici.

4.4. Considérons maintenant quelques-unes des conséquences de la formule (112) pour la fonctionnelle caractéristique. Soit ⁽¹⁰⁾

$$0 < x_1 < x_1 + \delta x_1 < x_2 < x_2 + \delta x_2 < t$$

et choisissons la fonction $\theta(x)$ comme suit :

$$(121) \quad \begin{cases} \theta(x) \equiv u_j \equiv -i \log z_j & (x_j \leq x < x_j + \delta x_j) \quad (j = 1, 2); \\ \equiv 0, & \text{ailleurs.} \end{cases}$$

Donc, $C[\theta(x); t]$ sera la *fonction* caractéristique des deux variables aléatoires (n_1, n_2) , où

$n_1 \equiv$ nombre d'organismes dont l'âge se trouve dans l'intervalle $x_1 \leq x < x_1 + \delta x_1$, à l'instant t ;

$n_2 \equiv$ nombre d'organismes dont l'âge se trouve dans l'intervalle $x_2 \leq x < x_2 + \delta x_2$, à l'instant t .

De cette façon, on trouve comme fonction génératrice ⁽¹¹⁾ de la distribution des deux variables (n_1, n_2) la fonction

$$(122) \quad \mathcal{G}(z_1^{n_1} z_2^{n_2}) = 1 + \frac{\sum_{j=1}^2 G_j(z_j - 1)}{1 - \sum_{j=1}^2 H_j(z_j - 1)},$$

où nous avons

$$(123) \quad G_1 \equiv e^{(\lambda - \mu)t} e^{-\lambda x_1} (1 - e^{-\lambda \delta x_1})$$

et

$$(124) \quad H_1 \equiv \frac{\lambda}{\lambda - \mu} [e^{(\lambda - \mu)t} e^{-\lambda x_1} (1 - e^{-\lambda \delta x_1}) - e^{-\mu x_1} (1 - e^{-\mu \delta x_1})],$$

les fonctions G_2 et H_2 étant données par des formules analogues les unes aux autres.

⁽¹⁰⁾ Voici une nouvelle notation sans aucun rapport avec la précédente (δ, x_s, \dots).

⁽¹¹⁾ Nous avons supposé ci-dessus que la fonction $\theta(x)$ était réelle; donc la valeur absolue de chaque nombre z doit être égale à l'unité. Or, une fonction génératrice est déterminée par les valeurs qu'elle prend sur la circonférence du cercle de rayon un.

Les valeurs moyennes, les variances et la covariance des variables n_1 et n_2 seront données par les formules suivantes :

$$(125) \quad \begin{cases} \mathcal{E}n = G, & \text{var}(n) = G + 2GH - G^2, \\ \text{cov}(n_1, n_2) = G_1H_2 + G_2H_1 - G_1G_2. \end{cases}$$

Par conséquent, lorsque les quantités δx_1 et δx_2 sont très petites, nous avons

$$\mathcal{E}n = \alpha(x, t)\delta x + o(\delta x) = \text{var}(n)$$

et

$$(126) \quad \text{cov}(n_1, n_2) = \gamma(x_1, x_2; t)\delta x_1\delta x_2 + o(\delta x_1\delta x_2),$$

où les fonctions α et γ sont faciles à déterminer à partir des fonctions G et H . Dans le cas le plus simple, les taux λ et μ étant égaux, nous aurons

$$\alpha(x, t) = \lambda e^{-\lambda x} \quad (x < t)$$

et

$$(127) \quad \gamma(x_1, x_2; t) = \lambda^2 [1 + \lambda(2t - x_1 - x_2)] e^{-\lambda(x_1 + x_2)} \quad (x_1, x_2 < t).$$

J'ai donné ces formules-ci, et quelques autres plus générales, dans mon article de 1949. Je les avais dérivées d'une série d'équations intégrales pour les fonctions α et γ (équations qui restent vraies lorsque le taux de naissance λ est une fonction positive *quelconque* de l'âge x de l'organisme générateur). L'équation de α est d'un type bien connu, tandis que la solution générale de l'équation donnant γ n'a été trouvée que récemment par M. M. M. Crum du New College, Oxford, dans une étude encore inédite.

Voilà tout ce que je veux dire en ce moment au sujet de la caractéristique fonctionnelle, sans oublier de mentionner l'étude indépendante et très intéressante de Ramakrishnan (1950) sur les fonctions semblables à α et γ introduites en liaison avec les processus d'embranchement qui se présentent dans la théorie du rayonnement cosmique. Il faut noter aussi qu'une équation intégrale a été donnée pour la fonctionnelle caractéristique (Bartlett et Kendall, 1951; Bartlett, 1951a) à l'aide d'une méthode proposée pour la première fois par Jánossy (1950) à propos des mêmes problèmes du rayonnement cosmique. Notre équation intégrale est valable même si les taux λ et μ sont des fonctions quel-

conques de l'âge x de l'organisme, mais la solution dans ce cas général reste encore inconnue ⁽¹²⁾.

Pour conclure ce chapitre revenons au problème qui tire son origine de la remarque de de Candolle : « Évidemment, tous les noms doivent s'éteindre ». Comme je l'ai déjà dit, la théorie qui en découle a été donnée par Steffensen dans les *Annales* de cet Institut, et je vais simplement la rappeler. Soit

$$(128) \quad f(z) \equiv \sum_{r=0}^{\infty} p_r z^r$$

la fonction génératrice de la distribution du nombre des fils dans la postérité d'un seul homme. Alors, la quantité

$$(129) \quad m \equiv \sum_{r=0}^{\infty} r p_r = f'(1-0)$$

sera la valeur moyenne du nombre de ses fils et d'après Steffensen, nous connaissons les faits suivants :

1° Si m est inférieure à l'unité, et si les distributions de probabilité restent les mêmes de génération à génération, sa postérité en ligne masculine directe, s'éteindra « presque certainement » ;

2° L'extinction sera encore inévitable lorsque m a la valeur critique de l'unité ;

3° La probabilité d'extinction lorsque m est supérieure à l'unité sera égale à la racine unique de l'équation

$$(130) \quad f(x) = x,$$

dans l'intervalle, $0 \leq x < 1$. [J'ometts toute mention de quelques cas exceptionnels; ainsi je suppose que $f(x) \neq x$.]

Cela étant, considérons une application du « théorème des noms de famille » de Watson et Steffensen à un problème que l'on trouve à première vue sans aucun rapport avec notre sujet. C'est un problème de M. É. Borel (1942) dans la théorie des queues. Considérons un guichet devant lequel des hommes arrivent par hasard et attendent à la queue leu

(12) Une solution itérative a été donnée récemment par Bartlett (1951 a).

leu. Soit leur arrivée réglée par un processus de Poisson, de sorte que le nombre de clients qui arrivent pendant un temps t sera une variable du type de Poisson, avec une valeur moyenne égale à $\frac{t}{a}$. Soit v le temps de service par client; ce sera en général une variable aléatoire; soit b sa valeur moyenne. Soit $\rho \left(= \frac{b}{a} \right)$ le nombre moyen d'arrivées pendant un temps moyen de service b . Supposons que le premier client arrive à l'époque $t = 0$ et supposons enfin qu'un nombre total M de clients doivent être servis avant que le guichet soit à nouveau libre. Borel a étudié la distribution de la variable aléatoire M lorsque le temps de service v a une valeur fixée.

Soit P la probabilité que la « série » ait une longueur finie, c'est-à-dire soit

$$P = \text{prob}[M < \infty].$$

Soit, de plus,

$$x = \rho \quad \text{et} \quad x = \xi$$

les deux racines positives de l'équation

$$(131) \quad x e^{-x} = \rho e^{-\rho}.$$

Dans le cas spécial d'un temps de service qui est fixé et égal à b , Borel a démontré que

$$(132) \quad P = \min\left(1, \frac{\xi}{\rho}\right).$$

Maintenant, cela étant, je vais donner un résultat plus général en me servant du théorème des noms de famille.

Appelons le premier client « l'aïeul ». Ses « fils » seront les clients nouveaux qui arrivent pendant le temps que l'on met à le servir. Ses petits-fils seront les clients qui arrivent pendant le temps que l'on met à servir les fils, et ainsi de suite. Alors le nombre M sera à la fois : 1° la grandeur de la série, au sens de Borel, et 2° le nombre total des personnes masculines dans la famille dont le premier client est l'ancêtre éponyme. Soit $dB(v)$ la distribution du temps de service pour un seul client et soit

$$(133) \quad K(z) \equiv \int_0^{\infty} e^{-\frac{(1-z)v}{a}} dB(v).$$

Cette fonction-ci jouera le rôle de la fonction génératrice $f(z)$ dans le problème transformé. Par conséquent, d'après le théorème de Watson et de Steffensen, nous aurons le résultat suivant : *la probabilité P est égale à la plus petite racine de l'équation $K(x) = x$, dans l'intervalle $0 \leq x \leq 1$.*

Si le temps de service a une valeur fixée, nous aurons

$$(134) \quad K(z) \equiv e^{-\rho(1-z)}$$

et le théorème de Borel suit immédiatement. Notons (comme dans le cas considéré par Borel) qu'en général le nom de famille s'éteindra et la série se terminera seulement si l'on a

$$(135) \quad \rho \equiv \frac{\mathcal{E}v}{a} \leq 1.$$

Quelques autres résultats de ce genre se trouvent dans mon article de 1951 et dans la Communication subséquente de M. I. J. Good.

BIBLIOGRAPHIE.

Cette bibliographie comprend tous les Mémoires cités dans le texte et, de plus, plusieurs Mémoires non cités, mais qui sont en relation avec le sujet de cet article.

ARLEY (N.) :

1943. *On the Theory of Stochastic Processes, and their Application to the Theory of Cosmic Radiation* (Copenhague).

1949. *On the birth and death process* (*Skand. Akt. Tidsskrift*, p. 21-26).

ARLEY (N.) et BORCHSENIUS (V.) :

1945. *On the theory of infinite systems of differential equations, and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics* (*Acta Math.*, t. 76, p. 261-322).

ARMITAGE (P.) :

1951. *The statistical theory of bacterial populations subject to mutation* [*J. Roy. Statist. Soc. (B)*, sous presse, bibliographie].

BAILEY (N. T. J.) :

1950. *A simple stochastic epidemic* (*Biometrika*, t. 37, p. 193-202).
 1951. *On estimating the size of mobile populations from recapture data* (*Biometrika*, t. 38, p. 293-306).

BARTLETT (M. S.) :

1946. *Stochastic Processes* (conférences inédites).
 1949. *Some evolutionary stochastic processes* [*J. Royal Statist. Soc. (B)*, t. 11, p. 211-229, 265-266 et 280-281, bibliographie].
 1951 a. *The dual recurrence relation for multiplicative processes* (*Proc. Cambridge Phil. Soc.*, t. 47, p. 821-825).
 1951 b. *Voir Armitage* (1951).

BARTLETT (M. S.) et KENDALL (D. G.) :

1951. *On the use of the characteristic functional in the analysis of some stochastic processes occurring in physics and biology* (*Proc. Cambridge Phil. Soc.*, t. 47, p. 65-76).

BELLMAN (R.) et HARRIS (T. E.) :

1948. *On the theory of age-dependent stochastic branching processes* (*Proc. Nat. Acad. Sc.*, t. 34, p. 601-604). (Une version plus détaillée paraîtra bientôt).

BERNARDELLI (H.) :

1941. *Population waves* (*J. Burma Research Soc.*, t. 31, p. 1-18).

BHABHA (H. J.) :

1950. *On the stochastic theory of continuous parametric systems and its application to electron cascades* [*Proc. Roy. Soc. London (A)*, t. 202, p. 301-322].

BHABHA (H. J.) et RAMAKRISHNAN (A.) :

1950. *The mean square deviation of the number of electrons and quanta in the cascade theory* (*Proc. Indian Acad. Sc.*, t. 32, p. 141-153).

BOGNER (S.) :

1947. *Stochastic processes* (*Ann. of Math.*, t. 48, p. 1014-1061).

BOREL (É.) :

1942. *Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Applications au problème de l'attente à un guichet* (*C. R. Acad. Sc.*, t. 214, p. 452-456).

CHITTY (D.) et LESLIE (P. H.) :

1951. *The estimation of population parameters from data obtained by means of the capture-recapture method* (*Biometrika*, t. 38, p. 269-292).

COCHRAN (W. G.) :

1949. *The present status of Biometry* (2^e Congrès intern. de Biométrie, Genève, p. 1-19, bibliographie).

CONSABL (R.) :

1948. *Sur une généralisation du processus de Pólya* (*Bull. Classe des Sc., Acad. Roy. de Belgique*, 5^e série, t. 34, p. 863-876).
 1949. *Sur quelques processus stochastiques discontinus à deux variables aléatoires* (*Ibid.*, t. 35, p. 399-416 et 743-755).
 1950. *Sur quelques points de la théorie des processus stochastiques* (*Ibid.*, t. 36, p. 870-879).

CRAMÉR (H.) :

1946. *Mathematical Methods of Statistics* (Princeton).

DÉAN (A. C. R.) et HINSHELWOOD (C. N.) :

1952. *The applicability of the statistical fluctuation test* [*Proc. Roy. Soc. London* (B), t. 139, p. 236-250].

DE CANDOLLE (A.) :

1873. *Histoire des Sciences et des Savants depuis deux siècles* (Genève). Voir p. 388.

ERLANG (A. K.) :

1929. *Matemo Tidsskrift* (B.), p. 36.

FELLER (W.) :

1939. *Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung* (*Acta Biotheoretica*, t. 5, p. 11-40).

1940. *On the integrodifferential equations of purely discontinuous Markoff processes* (*Trans. American Math. Soc.*, t. 48, p. 488-515).

1941. *On the integral equation of renewal theory* (*Ann. Math. Statist.*, t. 12, p. 243-267).

1943. *On a general class of « contagious » distributions* (*Ann. Math. Statist.*, t. 14, p. 389-400).

1949. *On the theory of stochastic processes, with particular reference to applications* (*Proc. 1st Berkeley Symposium on Math. Statist. and Probability*, p. 403-432, bibliographie).

1950. *An Introduction to Probability Theory and its Applications*. t. I, New-York, chap. 11 et 17).

1951. *Diffusion processes in genetics* (*Proc. 2nd Berkeley symposium on Math. Statist. and Probability*, p. 227-246, bibliographie).

FINNEY (D. J.) :

1947. *Errors of estimation in inverse sampling* (*Nature*, London, t. 160, p. 195).

1949. *On a method of estimating frequencies* (*Biometrika*, t. 36, p. 233-234).

FINNEY (D. J.) et MARTIN (L.) :

1951. *A re-examination of Rahn's data on the number of genes in bacteria* (*Biometrics*, t. 7, p. 133-144).

FISHER (R. A.) :

1922. *On the dominance ratio* (*Proc. Roy. Soc. Edin.*, t. 42, p. 321-341).

1930. *The Genetical Theory of Natural Selection* (Oxford).

FRÉCHET (M.) :

1938. *Recherches théoriques modernes sur le Calcul des Probabilités*, t. II, Paris, (bibliographie).

FURRY (W. H.) :

1937. *On fluctuation phenomena in the passage of high energy electrons through lead* (*Phys. Rev.*, t. 52, p. 569-581).

GALTON (F.) et WATSON (H. W.) :

1874. *On the probability of the extinction of families* (*J. Anthropol. Inst.*, t. 4, p. 138-144).

GOOD (I. J.) :

1949 a. *The number of individuals in a cascade process* (*Proc. Cambridge Phil. Soc.*, t. 45, p. 360-363).

- 1949 *b.* *J. Roy. Statist. Soc. (B)*, t. 11, p. 271-273 et t. 12, p. 1.
 1951. *Ibid.*, t. 13, p. 182-183.
- GRAY (J.) et TOCHER (J. F.) :
1902. *The ethnology of Buchan, IV. The frequency and pigmentation value of surnames in East Aberdeenshire* (*Trans. Buchan Field Club*, t. 7).
- GRENANDER (U.) :
1950. *Stochastic processes and statistical inference* (*Ark. Mat.*, t. 1, p. 195-277, bibliographie).
- HALDANE (J. B. S.) :
- 1945 *a.* *A labour-saving method of sampling* (*Nature*, London, t. 155, p. 49-50).
 1945 *b.* *On a method of estimating frequencies* (*Biometrika*, t. 33, p. 222-225).
 1949. *Some statistical problems arising in genetics* [*J. Roy. Statist. Soc. (B)*, t. 11, p. 1-9].
- HAMMERSLEY (J. M.) :
1950. *Electronic computers and the analysis of stochastic processes* (*Math. Tables, and Aids to Computation*, t. 4, p. 56-57).
- HARRIS (T. E.) :
1948. *Branching processes* (*Ann. Math. Statist.*, t. 19, p. 474-494, bibliographie).
 1951. *Some mathematical models for branching processes* (*Proc. 2nd Berkeley Symposium on Math. Statist. and Probability*, p. 305-328, bibliographie).
- HARTLEY (H. O.) et FITCH (E. R.) :
1951. *A chart for the incomplete beta-function and the cumulative binomial distribution* (*Biometrika*, t. 38, p. 423-426).
- IMMEL (E. R.) :
1951. *Problems of estimation and hypothesis-testing in connexion with birth-and-death stochastic processes* (*Ann. Math. Statist.*, t. 22, p. 485).
- JÁNOSSY (L.) :
1950. *Note on the fluctuation problem of cascades* (*Proc. Phys. Soc. London (A)*, t. 63, p. 241-249).
- KARMEL (P. H.) :
1947. *The relations between male and female reproduction rates* (*Population Studies*, t. 1, p. 249-274).
 1948 *a.* *The relations between male and female nuptiality in a stable population* (*Ibid.*, t. 1, p. 353-387).
 1948 *b.* *An analysis of the sources and magnitude of inconsistencies between male and female net reproduction rates in actual populations* (*Ibid.*, t. 2, p. 240-273).
- KELLY (C. D.) et RAHN (O.) :
1932. *The growth rate of individual bacterial cells* (*J. Bacteriol.*, t. 23, p. 147-153).
- KENDALL (D. G.) :
1947. *A review of some recent work on discontinuous Markoff processes with applications to biology, physics and actuarial science* (*J. Roy. Statist. Soc.*, t. 110, p. 130-137).
 1948 *a.* *On the generalised birth-and-death process* (*Ann. Math. Statist.*, t. 19, p. 1-15).
 1948 *b.* *On some modes of population growth leading to R. A. Fisher's logarithmic series distribution* (*Biometrika*, t. 35, p. 6-15).

- 1948 c. *On the role of variable generation time in the development of a stochastic birth process.* (*Ibid.*, t. 35, p. 316-330).
1949. *Stochastic processes and population growth* [*J. Roy. Statist. Soc. (B)*, t. 11, p. 230-264, 266-267 et 281-282, bibliographie].
- 1950 a. *An artificial realisation of a simple birth-and-death process* (*Ibid.*, t. 12, p. 116-119).
- 1950 b. *Random fluctuations in the age-distribution of a population whose development is controlled by the simple birth-and-death process.* (*Ibid.*, t. 12, p. 278-285).
1951. *Some problems in the theory of queues* (*Ibid.*, t. 13, p. 151-173 et 184-185, bibliographie).
- 1952 a. *On the choice of a mathematical model to represent normal bacterial growth* (*Ibid.*, sous presse).
- 1952 b. *Stochastic processes and bacterial growth* (*Soc. for Exp. Biol.*, sous presse.)
- KOLMOGOROFF (A.) :
1936. *Sulla teoria di Volterra della lotta per l'esistenza* (*Giorn. Ist. Ital. Attuari*, t. 7, p. 74-80).
- KOLMOGOROFF (A.), PETROVSKY (I.) et PISCOUNOFF (N.) :
1937. *Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique* (*Bull. Univ. Etat. Moscou* (A), (I), t. 6, p. 1-25).
- KOLMOGOROFF (A.) et DMITRIEV (N. A.) :
1947. *Branching stochastic processes* [*C. R. Acad. Sc. U.R.S.S. (N. S.)*, t. 56, p. 5-8 (anglais)].
- KOLMOGOROFF (A.) et SAVOSTYANOV (B. A.) :
1947. *The calculation of final probabilities for branching random processes* [*Ibid.*, t. 56, p. 783-786 (russe)].
- LEA (D. E.) et COULSON (C. A.) :
1949. *The distribution of the number of mutants in bacterial populations* (*J. Genetics*, t. 49, p. 264-285).
- LE CAM :
1947. *Un instrument d'étude des fonctions aléatoires : la fonctionnelle caractéristique* (*C. R. Acad. Sc.*, t. 224, p. 710-711).
- LEDERBERG (J.) (sous la direction de) :
1951. *Papers in Microbial Genetics* (Madison, É.-U.).
- LESLIE (P. H.) :
1945. *On the use of matrices in certain population mathematics* (*Biometrika*, t. 33, p. 183-212).
1948. *Some further notes on the use of matrices in population mathematics* (*Ibid.*, t. 35, p. 213-245).
- LÉVY (P.) :
1948. *Processus Stochastiques et Mouvement Brownien* (Paris), (bibliographie).
- LEWIS (E. G.) :
1942. *On the generation and growth of a population* (*Sankhya*, t. 6, p. 93-96).
- LOTKA (A. J.) :
1931. *The extinction of families* (*J. Washington Acad. Sc.*, t. 21, p. 377-380 et 453-459).

1939. *Analyse démographique avec application particulière à l'espèce humaine* (*Acta Sc. Ind.*, p. 780, Paris).
1945. *Population analysis as a chapter in the mathematical theory of evolution* (*Essays on Growth and Form presented to d'Arcy Wentworth Thompson* rédigés par W. E. LE GROS CLARK et P. B. MEDAWAR, Oxford) (bibliographie).
1948. *Application of recurrent series in renewal theory* (*Ann. Math. Statist.*, t. 19, p. 190-206).
- LURIA (S. E.) et DELBRÜCK (M.) :
1943. *Mutations of bacteria from virus sensitivity to virus resistance* (*Genetics*, t. 28, p. 491-511).
- MALÉCOT (G.) :
1948. *Les Mathématiques de l'Hérédité* (Paris).
1949. *Les processus stochastiques de la génétique* (*Coll. intern. du C. N. R. S.*, t. 13, p. 121-126).
1950. *Quelques schèmes probabilistes sur la variabilité des populations naturelles* [*Ann. de l'Univ. de Lyon* (A), t. 13, p. 37-60].
- MARTIN (L.) :
- 1948-1949. *Évolution de la Biométrie* (*Bull. de l'Inst. Agronom. de Gembloux*, t. 17, p. 43-66, bibliographie).
- Mc KENDRICK (A. G.) :
1914. *Studies on the theory of continuous probabilities, with special reference to its bearing on natural phenomena of a progressive nature* (*Proc. London Math. Soc.*, 2^e série, t. 13, p. 401-416).
- MORAN (P. A. P.) :
1950. *Some remarks on animal population dynamics* (*Biometrics*, t. 6, p. 250-258).
- 1951 a. *Estimation methods for evolutive processes* (*J. Roy. Statist. Soc. (B)*, t. 13, p. 141-146).
- 1951 b. *A mathematical theory of animal trapping* (*Biometrika*, t. 38, p. 307-311).
1952. *The Statistical analysis of game-bird records* (*J. Animal Ecology*, t. 21, p. 154-518).
- MOURIER (E.) :
1949. *Sur l'espérance mathématique d'un élément aléatoire dans un espace de Banach* (*C. R. Acad. Sc.*, t. 229, p. 1300-1301).
1950. *Propriétés des caractéristiques d'un élément aléatoire dans un espace de Banach* (*Ibid.*, t. 231, p. 28-29).
1951. *Lois des grands nombres et théorie érodique* (*Ibid.*, t. 232, p. 923-925).
- MOYAL (J. E.) :
1949. *J. Roy. Statist. Soc. (B)*, t. 11, p. 265.
- NEWCOMBE (H. E.) :
1948. *Delayed phenotypic expression of spontaneous mutations in Escherichia coli* (*Genetics*, t. 33, p. 447-476, bibliographie).
- OTTER (R.) :
1949. *The multiplicative process* (*Ann. Math. Statist.*, t. 20, p. 206-224).
- PALM (C.) :
1945. Voir Arley et Borchsenius (1945), (p. 298-299).

- PARK (T.) :
1948. *Experimental studies of interspecies competition, I* (*Ecol. Monographs*, t. 18, p. 265-308).
- PÓLYA (G.) :
1930. *Sur quelques points de la théorie des probabilités* (*Ann. Inst. H. Poincaré*, t. 1, p. 117-161).
- PRENDIVILLE (B. J.)
1949. *J. Roy. Statist. Soc. (B)*, t. 11, p. 273.
- RAHN (O.) :
1932. *A chemical explanation of the variability of the growth rate* (*J. Gen. Physiol.*, t. 15, p. 257-277).
1934. *Chemistry of cell growth, II*, (*Cold Spring Harbour Symp. on Quantitative Biol.*, t. 2, p. 63-63).
- RAMAKRISHNAN (A.) :
1950. *Stochastic processes relating to particles distributed in a continuous infinity of states* (*Prob. Cambridge Phil. Soc.*, t. 46, p. 595-602).
1951. *Some simple stochastic processes* [*J. Royal Statist. Soc. (B)*, t. 13, p. 131-140].
- SAVOSTYANOV (B. A.) :
1948. *On the theory of branching random processes* [*C. R. Acad. Sc. U. R. S. S.*, t. 59, p. 1407-1410 (russe)].
- STEFFENSEN (J. F.) :
1930. *Om sandsynligheden for at afkommet uddør* (*Matem. Tidsskrift (B)*, p. 19-23).
1933. *Deux problèmes du calcul des probabilités* (*Ann. Inst. H. Poincaré*, t. 3, p. 319-344).
- THOMPSON (W. R.) et SOPER (H. E.) :
1931. *On the reproduction of organisms with overlapping generations* (*Bull. Entom. Research.*, t. 22, p. 147-172).
- TWEEDIE (M. C. K.) :
1945. *Inverse statistical variates* (*Nature*, London, t. 155, p. 453).
- VOLTERRA (V.) :
1931. *Leçons sur la théorie mathématique de la lutte pour la vie* (Paris).
- WALD (A.) :
1948. *Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process* (*Ann. Math. Statist.*, t. 19, p. 40-46).
- WOODWARD (P. M.) :
1947. *A statistical theory of cascade multiplication* (*Proc. Cambridge Phil. Soc.*, t. 44, p. 404-412).
- YAGLOM (A. M.) :
1947. *Certain limit theorems of the theory of branching random processes* (*C. R. Acad. Sc. U. R. S. S.*, t. 56, p. 795-798 (russe)).
- YULE (G. U.) :
1925. *A mathematical theory of evolution* [*Phil. Trans. Roy. Soc. London (B)*, t. 213, p. 21-87].