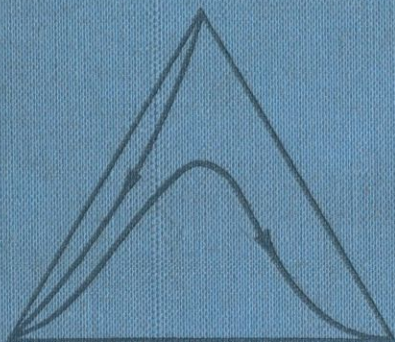


OUVRAGE PUBLIE SOUS LES AUSPICES
DE L'INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

STRUCTURES GÉNÉTIQUES DES POPULATIONS

PAR

Albert JACQUARD



MASSON & CIE

STRUCTURES GÉNÉTIQUES
DES
POPULATIONS

8° S
28214

A LA MÊME LIBRAIRIE

RECUEIL D'EXERCICES DE GÉNÉTIQUE DES POPULATIONS, *avec rappels des définitions et notions fondamentales*, par Jean GENERMONT, 1970, un volume de 150 pages.

ÉLÉMENTS DE GÉNÉTIQUE GÉNÉRALE ET HUMAINE, par J. RUFFIÉ, 1969 un volume de 86 pages avec 27 figures.

HÉMATOLOGIE GÉOGRAPHIQUE. *Écologie humaine. Caractères héréditaires du sang*, par Jean BERNARD et J. RUFFIÉ, 1966, un volume de 436 pages avec 56 figures.

LES MÉTHODES EN GÉNÉTIQUE GÉNÉRALE ET EN GÉNÉTIQUE HUMAINE, par R. HURON et J. RUFFIÉ, 1959, un volume de 556 pages avec 79 figures.

OUVRAGE PUBLIÉ SOUS LES AUSPICES
DE L'INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

PRÉFACE

STRUCTURES GÉNÉTIQUES DES POPULATIONS

par

Albert JACQUARD

Directeur de Recherches
à l'Institut National d'Études Démographiques.

PRÉFACE DE J. SUTTER

MASSON ET C^{ie}, ÉDITEURS
120, BOULEVARD SAINT-GERMAIN, PARIS, VI^e

1970

*Tous droits d'adaptation, de traduction
et de reproduction, par tous procédés,
y compris la photographie et le microfilm
réservés pour tous pays.*

© 1970 Masson et C^{ie}, Paris.

Imprimé en France.



PRÉFACE

Les succès spectaculaires obtenus à tous les niveaux par la biologie cellulaire et la génétique moléculaire, ont tendance, à l'heure actuelle, à laisser dans l'ombre le développement de la génétique physiologique et celui de la génétique de population. Il est bien évident, cependant, que le mécanisme de fonctionnement d'un gène ou d'un enzyme n'a de sens pour l'individu que si l'on arrive à en évaluer le retentissement sur les divers cycles biochimiques ou physiologiques intéressés. Plus avant, la variété et la variabilité des structures individuelles permettent d'estimer les structures génétiques des populations et leur évolution.

Le livre de M. Jacquard intéresse tous les aspects de la génétique des populations envisagée sous l'angle de la théorie probabiliste de la parenté et de la covariance génétique, développée depuis trente ans par M. Gustave Malécot.

L'intérêt capital des conceptions de M. Malécot n'est plus à démontrer. Elles n'ont pu faire réaliser de si grands progrès à la génétique des populations qu'en s'opposant en partie à celles des savants de la génération précédente, Fisher, Haldane et Wright. Comme l'a remarqué Newton E. Morton, ceux-ci pratiquèrent surtout l'analyse de la corrélation et ne pensèrent pas que « le passage des probabilités aux corrélations est beaucoup plus facile que le chemin inverse ». L'approche probabiliste de M. Malécot devait relancer tous les problèmes inhérents et ouvrir de nouveaux horizons à la génétique des populations.

On peut estimer que M. Jacquard a créé un langage mathématique qui permet de simplifier un grand nombre d'opérations et qui doit permettre, à l'avenir, un nouveau développement de cette branche de la génétique. Le succès de son langage parmi la jeune génération de chercheurs et parmi les étudiants, permet tous les espoirs. Ce langage paraît en effet tout à fait bien adapté d'une part aux opérations électroniques : analyse, simulations, etc...; d'autre part, à l'intégration des faits génétiques humains dans les structures démographiques. Ce dernier point doit avoir bientôt une large incidence sur les problèmes de Santé Publique.

D^r Jean SUTTER.

AVANT-PROPOS

SCIENCE relativement récente, — son objet même n'a été défini qu'au début de ce siècle — la génétique des populations a connu un développement si rapide, orienté dans des directions si variées, qu'il n'est possible d'en donner, dans un cadre limité, qu'une vue superficielle. Ayant avant tout pour objectif l'information des étudiants, nous n'avons nullement cherché à évoquer ici l'ensemble des travaux consacrés à cette discipline; nous avons simplement tenté de faciliter l'accès aux ouvrages et articles spécialisés, en aidant le débutant dans la phase la plus pénible pour qui aborde une science nouvelle, la compréhension des concepts de base.

Nous limitant aux populations humaines, nous avons articulé le plan des divers chapitres, qui sont autant de « leçons » d'importance d'ailleurs inégale, autour de la notion de « structure génétique », définie comme l'ensemble des fréquences des gènes ou des génotypes dans une génération.

Quel que soit le facteur étudié, mutation, sélection, migration, le but de la recherche est toujours de définir le processus de passage d'une génération à la suivante, processus qui n'est généralement connu qu'au moyen de lois de probabilités. La génétique est sans doute le domaine où le recours au raisonnement probabiliste est le mieux adapté à la réalité: la structure génétique d'une population, objet du discours, constitue une réalité en fait inaccessible, il ne peut être question d'affirmer qu'elle obéit à telle loi déterministe, il s'agit d'utiliser au mieux, et en toute rigueur, l'information que l'on possède à son sujet en affectant des probabilités aux divers cas possibles; aussi avons-nous fait appel systématiquement à cette forme de raisonnement. Nous espérons que ce parti ne constituera pas un obstacle trop difficile à franchir pour les étudiants peu habitués à cette démarche.

ARTICLE

The first of these is the fact that the law is not a static body of rules, but a living organism which grows and changes with the needs of the community. The second is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts. The third is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts.

The fourth is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts. The fifth is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts.

The sixth is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts. The seventh is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts. The eighth is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts. The ninth is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts. The tenth is the fact that the law is not a mere collection of rules, but a system of principles which guide the action of the courts.

INTRODUCTION

L'Individu. — Deux cellules reproductrices, un ovule, un spermatozoïde, se rencontrent, fusionnent pour ne plus constituer qu'une seule cellule : un être humain nouveau vient d'être créé, doté, dès cet instant, de toutes ses caractéristiques fondamentales.

A partir de cette cellule initiale des milliards de cellules nouvelles, adaptées chacune à une fonction bien déterminée, vont peu à peu se constituer; des milliers de substances chimiques, servant de matériaux pour l'édification de ces cellules ou assurant les liaisons entre elles, vont être synthétisées; des mécanismes délicats de régulation et d'adaptation vont être instaurés; tous ces processus seront gouvernés par un programme précis se déroulant inexorablement : développement de l'embryon, croissance, sénescence, dislocation finale de l'organisme.

Les informations nécessaires pour assurer toutes ces fonctions sont inscrites, dès l'origine, dans la cellule initiale : provenant pour moitié de l'ovule, pour moitié du spermatozoïde, le « code génétique », porté par les chromosomes de cette cellule, contient toutes les instructions permettant de construire peu à peu un être humain, ou plus exactement *tel* être humain.

Car l'information génétique dont il est doté différencie cet être nouveau de tout autre, fonde son individualité : aucun homme vivant ou mort ne possède ou n'a possédé exactement le même héritage. Certes le développement de l'organisme qui peu à peu se constitue, son ampleur, sa vigueur, sont dans une large mesure influencés par l'environnement extérieur : apport de la mère durant la vie utérine, qualité et quantité de nourriture disponible, température et rayonnement subis, ... Mais l'effet de ces facteurs externes est étroitement réglementé par les processus internes définis par l'information génétique. L'aboutissement du développement de l'organisme est fonction des hasards de son *existence*, mais ces hasards agissent selon des lois de probabilités définies par l'héritage génétique que cet organisme a reçu lors de sa conception, et qui constitue son *essence*.

Lors de chaque différenciation cellulaire, la structure de la nouvelle cellule est fonction du rôle qu'elle aura à jouer dans l'organisme (cellule nerveuse, cellule osseuse, cellule sanguine, ...) mais, dans tous les cas, elle reçoit, quel que soit ce rôle, une copie intégrale de l'héritage génétique. Chaque cellule de

l'individu I est dotée ainsi de la totalité des informations nécessaires pour construire l'ensemble de l'organisme, chaque cellule « connaît » le prodigieux secret qui permet de réaliser I, mais aussi chaque cellule « sait » qu'elle appartient à I; elle n'est pas seulement une cellule nerveuse ou une cellule osseuse, ayant telle ou telle fonction, elle est aussi, et avant tout, une cellule de I.

Une exception cependant : les cellules chargées de la reproduction, fabriquées par les organes sexuels, ovules et spermatozoïdes, ne reçoivent que la moitié de l'héritage génétique. Expulsés de l'organisme qui les a créés, dotés d'une information insuffisante, incapables de se différencier, ils meurent presque aussitôt. Dans une proportion infime, certains d'entre eux toutefois ont la chance de rencontrer une cellule reproductrice de l'autre sexe et de fusionner avec elle. Dans la nouvelle cellule ces deux moitiés d'héritages génétiques se complètent et reconstituent un ensemble d'informations comportant tous les éléments nécessaires à la réalisation d'un être humain.

Le développement d'un nouvel individu peut commencer.

La Population. — Chaque être humain ne peut ainsi transmettre à un descendant que la moitié de son héritage génétique, le partage en deux étant réalisé au hasard sans qu'une intervention de l'intéressé soit possible.

Mais, dans une population suffisamment nombreuse, ces partitions et ces fusions successives permettent à l'ensemble de l'héritage génétique collectif de se transmettre de génération en génération : derrière la diversité des individus qui se succèdent, subsiste la constance du patrimoine commun.

Si toutes les informations concernant la synthèse de telle protéine sont portées par les chromosomes d'un même individu, cette protéine est effectivement réalisée et il en porte la marque, favorable ou défavorable : son « génotype », disent les biologistes, peut s'exprimer en un « phénotype ». Si, à la génération suivante, ces informations sont réparties entre plusieurs individus, aucun d'entre eux ne les possédant en totalité, aucun ne sait synthétiser cette protéine, aucun n'en porte la marque, le phénotype correspondant a disparu. Mais la recette de cette synthèse n'est pas pour autant perdue : dans une génération ultérieure, peut être, les informations nécessaires pourront se trouver à nouveau réunies dans les chromosomes d'un même individu, qui manifestera alors le phénotype de son lointain ancêtre.

Pour une population, la « réalité » n'est donc pas ce qui apparaît dans les individus, mais ce qui est inscrit dans les chromosomes lovés dans les noyaux de leurs cellules sexuelles; alternativement réunies ou éparpillées, selon les hasards des rencontres de ces cellules, les informations portées par ces chromosomes se manifestent par des caractères apparents ou au contraire restent cachées, inexprimées, mais toujours réelles.

Au delà des individus qui *existent*, l'*essence* d'une population est constituée par l'héritage génétique collectif.

Par son comportement cependant chacun peut agir, dans une faible mesure, sur ce patrimoine collectif : selon qu'il choisit, ou supporte, d'avoir peu ou beaucoup d'enfants, les informations génétiques dont il est porteur seront transmises en plus ou moins grande proportion à la génération suivante. La

fréquence moyenne de chaque type d'information dans la population peut ainsi évoluer d'une génération à l'autre.

L'objet de la Génétique de Population est avant tout l'étude de cette évolution; son but est de répondre à la question : « quels sont les effets des divers comportements et des diverses conditions d'existence sur le patrimoine génétique collectif ? »

BIBLIOGRAPHIE

A la suite de chaque chapitre nous donnerons une bibliographie sommaire des ouvrages et articles consacrés à l'objet du chapitre. Le lecteur trouvera ci-après une liste de quelques livres plus généraux qui peuvent constituer une introduction à la génétique dans son ensemble, ou, plus spécialement, à la Génétique des populations humaines.

- BURDETTE W., 1962, *Methodology in Human Genetics*, 436 pages, Holden Day, San Francisco.
- FALCONER D. S., 1960, *Introduction to quantitative Genetics*, 366 pages, Oliver and Boyd, London.
- HURON R. et RUFFIÉ J., 1959, *Les méthodes en génétique animale et en génétique humaine*, 556 pages, Masson, Paris.
- L'HÉRITIER Ph., 1959, *Traité de génétique*, 518 pages, Presses Universitaires de France, Paris.
- KEMP THORNE O., 1966, *An Introduction to genetic statistics*, 545 pages, John Wiley, New York.
- LAMY M. et DE GROUCHY J., 1967, *L'homme et l'hérédité*, 281 pages, Hachette, Paris.
- LI C. C., 1955, *Population Genetics*, 366 pages, The University of Chicago Press, Chicago.
- Mc KUSICK V. A., 1964, *Human Genetics*, 148 pages, Prentice Hall, Englewood Cliff, N. J.
- MALECOT G., 1966, *Probabilités et Hérité*, 356 pages, Presses Universitaires de France, Paris.
- PETIT C. et PRÉVOST G., 1967, *Génétique et Evolution*, 392 pages, Hermann, Paris.
- SINNOTT E. W., DUNN L. C. et DOBZHANSKY T., 1958, *Principles of Genetics*, 459 pages, Mc Graw Hill, New-York.
- SRB A. M., OWEN R. D. et EDGAR R. S., 1965, *General Genetics*, 557 pages, W. H. Freeman, San Francisco.
- STERN C., 1960, *Principles of Human Genetics*, 753 pages, W. H. Freeman, San Francisco.
-

PREMIÈRE PARTIE

LES FAITS ET LES CONCEPTS

L'objectif de la génétique des populations est d'étudier les transformations du patrimoine héréditaire au niveau d'une collectivité; il paraît toutefois utile de rappeler l'essentiel des phénomènes biologiques qui, au niveau de l'individu, ou même au niveau de la cellule, réalisent la transmission de ce patrimoine d'une génération à l'autre.

Cette description permettra de définir de façon précise la terminologie ainsi que les divers concepts et les notations qui seront systématiquement utilisés.

CHAPITRE PREMIER

LES FONDEMENTS LOGIQUE ET BIOLOGIQUE DE LA GÉNÉTIQUE

1. — FONDEMENT LOGIQUE : LES LOIS DE MENDEL

Tous les raisonnements effectués au sujet de la transmission des caractères individuels et de l'évolution des caractéristiques globales d'une population utilisent des concepts définis il y a exactement un siècle par Johann Mendel, en religion Père Grégor (1822-1884).

Disposant comme outil de travail du petit ⁽¹⁾ jardin de son monastère de Brno, en Moravie, et de graines de pois, il sut imaginer des principes expliquant de façon simple les résultats de ses expériences.

En opposition avec toutes les idées admises, ces principes n'eurent naturellement aucune audience, et le mémoire de 1865 dans lequel il les exposait resta totalement ignoré ⁽²⁾. En 1900, seize ans après la mort de G. Mendel, son œuvre fut enfin redécouverte, d'abord par le hollandais Hugo de Vries, puis par l'allemand Correns et l'autrichien Von Tschermak; la Génétique put alors, notamment sous l'impulsion de Morgan aux États-Unis, se constituer progressivement en science; mais trente-cinq précieuses années avaient été perdues.

1.1. — *Première loi de Mendel. Loi de ségrégation.*

Croisant des pois à cotylédons jaunes que nous désignerons par le symbole [J] et des pois à cotylédons verts [V], Mendel obtenait uniquement des hybrides à cotylédones jaunes [J]. Croisant alors ces hybrides entre eux (ou plutôt les laissant s'auto-féconder) il obtenait à nouveau les deux sortes de pois; le caractère [V] caché par le caractère [J] dans la première génération, réapparaissait dans la seconde, où se produisait une « ségrégation » de ces deux

⁽¹⁾ En tout 250 m².

⁽²⁾ « Versuche über Pflanzen hybriden », paru dans le *Bulletin d'Histoire naturelle de Brno*, dont la diffusion était évidemment assez restreinte.

caractères. De plus dans cette seconde génération les proportions obtenues étaient dans toutes les expériences très proches de : 3 [J] pour 1 [V]. Des expériences analogues basées sur six autres caractères (graines lisses ou ridées, inflorescences axiales ou terminales,...) donnaient lieu aux mêmes constatations.

L'hypothèse posée par Mendel pour expliquer ces résultats peut être formulée ainsi (avec une terminologie évidemment différente de la sienne) :

— chaque caractère d'un individu est gouverné par deux « facteurs », les « gènes », qui lui ont été transmis l'un par l'individu mâle, l'autre par l'individu femelle dont il est issu;

— lorsque, chez un même individu, deux gènes gouvernant un même caractère sont différents, l'un est « dominant » et son action l'emporte sur celle de l'autre, qui est « récessif »;

— les cellules reproductrices émises par un individu contiennent, pour chaque caractère, un et un seul des deux gènes dont il est doté.

Dans l'exemple des pois, les plantes de la lignée jaune [J] sont dotées de deux gènes « jaunes »; leur constitution génétique peut donc être représentée par le symbole (JJ).

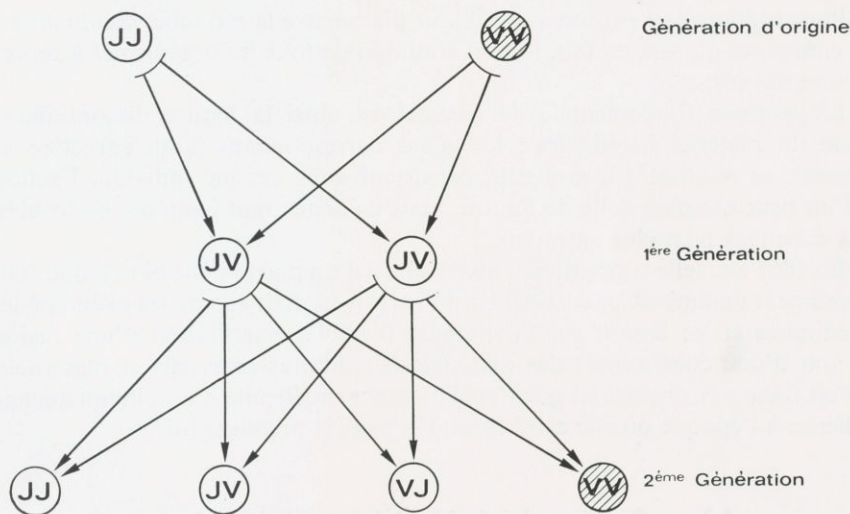


FIG. 1. — « L'expérience fondamentale de G. Mendel. »

Les pois de la lignée verte [V] ont la constitution génétique (VV). (Nous emploierons systématiquement la notation consistant à définir :

— une composition génétique par deux lettres comprises entre deux parenthèses, chaque lettre désignant l'un des gènes,

— le caractère manifesté par l'individu par une lettre comprise entre deux crochets.)

Chez les pois de la première génération hybride, un gène provient du parent de constitution génétique (JJ) et l'autre du parent (VV); la formule génétique

de ces pois hybrides est donc (JV); le gène J étant dominant sur le gène V tous sont jaunes [J].

La deuxième génération est issue de croisements entre hybrides, l'on peut représenter ces croisements par la formule :

$$(JV) \times (JV).$$

Chacun des parents fournit des cellules reproductrices dont la moitié contient le gène J et l'autre moitié le gène V. Les pois issus de ce croisement ont donc trois formules génétiques possibles :

(JJ) pour un quart d'entre eux.

(JV) ou (VJ) pour deux quarts d'entre eux.

(VV) pour un quart d'entre eux.

En raison de la dominance de J sur V, les pois (VJ) ou (JV) sont jaunes, ce qui explique la proportion trouvée : 1 pois vert pour 3 pois jaunes.

L'intuition de G. Mendel était basée, essentiellement, sur la constatation que le caractère récessif inapparent chez les pois de la première génération hybride, ici le caractère « cotylédons verts », doit y être toujours représenté de quelque façon puisqu'il réapparaît à la génération suivante.

Pour expliquer ce paradoxe il suffisait d'admettre la « double commande » des caractères qui est, en fait, le trait commun de tous les organismes à reproduction sexuée.

L'hypothèse fondamentale de Mendel est ainsi la nature discontinue et stable du matériel héréditaire : les gènes correspondant à un caractère se séparent, se réunissent à nouveau, coexistent dans chaque individu, l'action de l'un peut masquer celle de l'autre, mais ils demeurent inaltérés, insécables, sans échanges possibles entre eux.

En 1865 une telle conception « quantique » d'un phénomène biologique était proprement inadmissible; en 1900 Karl Pearson étudiait encore les phénomènes héréditaires en se basant sur l'hypothèse formulée par Galton d'une fusion (et non d'une coexistence) des matériels héréditaires paternels et maternels. Il n'est donc pas surprenant que l'explication mendélienne n'ait obtenu aucune audience à l'époque où elle a été formulée pour la première fois.

1.2. — Deuxième loi de Mendel. Loi d'indépendance.

Examinant ensuite le résultat du croisement de lignées de pois différant par deux caractères : les uns à cotylédons jaunes et à graines lisses [J — L], les autres à cotylédons verts et à graines ridées [V — R], Mendel constata que la première génération hybride ne comportait que des pois du type [J — L], mais que la seconde génération comprenait les quatre types de pois résultant de la combinaison de ces deux caractères; leurs proportions étaient : 9/16 de [J — L], 1/16 de [V — R], 3/16 de [J — R], 3/16 de [V — L]. En renouvelant l'expérience pour sept caractères différents, il fit dans chaque cas la même constatation.

Compte tenu de la première loi, il suffit pour expliquer ce résultat d'admettre l'hypothèse suivante qui constitue la deuxième « loi » de Mendel :

— les gènes gouvernant des caractères différents ont des ségrégations indépendantes.

En effet les formules génétiques des lignées initiales peuvent s'écrire (JJ—LL) et (VV — RR), et celle de la première génération hybride (JV — LR). Le gène J étant dominant sur le gène V, et le gène L sur le gène R, ces pois sont jaunes et lisses. Les cellules reproductrices émises par ces hybrides de première génération sont de quatre types : (J — L), (V — L), (J — R) et (V — R). Les gènes gouvernant la couleur des cotylédons étant indépendants de ceux gouvernant l'aspect de la graine, ces quatre types de cellules ont la même fréquence; chacune représente 1/4 du total. Le résultat de leur croisement peut être représenté par le tableau I dans lequel les cases correspondent aux

Tableau 1. — SÉGRÉGATIONS INDÉPENDANTES DE DEUX CARACTÈRES.

Cellules reproductrices	♂	(J — L)	(J — R)	(V — L)	(V — R)
♀					
(J — L)		(JJ — LL) [JL]	(JJ — LR) [JL]	(JV — LL) [JL]	(JV — LR) [JL]
(J — R)		(JJ — LR) [JL]	(JJ — RR) [JR]	(JV — LR) [JL]	(JV — RR) [JR]
(V — L)		(JV — LL) [JL]	(JV — LR) [JL]	(VV — LL) [VL]	(VV — LR) [VL]
(V — R)		(JV — LR) [JL]	(JV — RR) [JR]	(VV — LR) [VL]	(VV — RR) [VR]

rencontres des divers types de gamètes mâles ♂ avec les divers types de gamètes femelles ♀. Dans chaque case on a inscrit entre parenthèses le génotype correspondant à cette rencontre et entre crochets le phénotype résultant des relations de dominance-récessivité entre les gènes.

Les pois d'apparence jaunes et lisses [JL] correspondent, en raison de la dominance du gène de J sur le gène V et du gène L sur le gène R, à quatre formules génotypiques, (JJ — LL), (JJ — LR), (JV — LL) et (JV — LR), que l'on trouve dans 9 des 16 cases du tableau.

Les pois jaunes et ridés [JR] correspondent aux deux formules (JJ — RR) et (JV — RR) que l'on trouve dans 3 cases.

Les pois verts et lisses [VL] correspondent aux deux formules (VV — LL) et (VV — LR) que l'on trouve dans 3 cases.

Enfin les pois verts et ridés [VR] correspondent à l'unique formule (VV — RR) que l'on trouve dans 1 case.

Les croisements ayant lieu au hasard, les rencontres de cellules repro-

ductrices représentées par chacune des cases du tableau sont également probables : les fréquences des quatre types possibles de pois sont donc bien : $9/16$, $3/16$, $3/16$ et $1/16$.

1.3. — *Limites de la deuxième loi. Le Linkage.*

Loin d'avoir la portée universelle de la première, la deuxième loi de Mendel est en fait le plus souvent démentie par l'expérience. Utilisant comme matériel non plus le pois mais la mouche *Drosophile*, Morgan constata que la ségrégation indépendante des caractères est loin d'être le cas général. Au contraire il est possible de classer les divers caractères d'un organisme en un certain nombre de groupes de liaisons, de « linkage » : deux caractères appartenant à deux groupes distincts ont des ségrégations indépendantes (comme ceux analysés par Mendel); deux caractères appartenant à un même groupe sont liés, c'est-à-dire qu'ils ont « tendance » à rester associés chez les enfants comme ils l'étaient chez les parents. Cette « tendance » est plus ou moins marquée selon le degré de liaison; elle peut être mesurée par la proportion des nouvelles combinaisons de caractères apparaissant dans la descendance.

Reprenons l'exemple des pois : si l'on croise des pois à fleurs pourpres et à pollen long [P — L] avec des pois à fleurs rouges et à pollen court [R — C], on obtient une première génération hybride à fleurs pourpres et à pollen long : les formules génétiques des deux lignées initiales étaient (PP — LL) et (RR — CC), celle de la première génération hybride est (PR — LC), P est dominant sur R, L sur C, le phénotype est donc [P — L]. Dans la deuxième génération hybride on trouve bien, comme on s'y attend, quatre types de pois [P — L], [R — C], [P — C] et [R — L], mais les deux premières catégories, qui correspondent aux associations de caractères présentes dans les lignées initiales, sont beaucoup plus fréquentes que ne l'indique la deuxième loi de Mendel. Au lieu de trouver les proportions $\frac{9}{16} = 55\%$, $\frac{1}{16} = 7\%$, $\frac{3}{16} = 19\%$, $\frac{3}{16} = 19\%$, on obtient respectivement 70 %, 20 %, 5 %, 5 %.

Comme le montre le tableau 2, ces proportions peuvent s'expliquer en admettant que, lors de la création des cellules reproductrices des pois de la première génération hybride, dont la formule génétique est (PR — LC), les cellules comportant les mêmes associations (P — L) et (R — C) que les lignées initiales sont dix fois plus fréquentes que les cellules comportant les associations (P — C) et (R — L) correspondant à des recombinaisons; autrement dit parmi 11 cellules reproductrices contenant le gène P, 10 portent le gène L et 1 seulement le gène C, et parmi 11 cellules portant le gène R, 1 seulement contient le gène L et 10 le gène C.

Les caractères : couleur pourpre ou rouge de la fleur, forme longue ou ronde du pollen, appartiennent donc à un même groupe de liaison. La proportion 10/1 trouvée est très éloignée de la proportion 1/1 caractéristique de l'indépendance, l'intensité de leur « linkage » est donc très forte.

Tableau 2. — SÉGRÉGATIONS DE DEUX CARACTÈRES LIÉS.

Gamètes → ♂ ↓ ♀	Fréquence → ↓	(P — L)	(P — C)	(R — L)	(R — C)
		10/22	1/22	1/22	10/22
(P — L)	10/22	0,207 [PL]	0,021 [PL]	0,021 [PL]	0,207 [PL]
(P — C)	1/22	0,021 [PL]	0,002 [PC]	0,002 [PL]	0,021 [PC]
(R — L)	1/22	0,021 [PL]	0,002 [PL]	0,002 [RL]	0,021 [RL]
(R — C)	10/22	0,207 [PL]	0,021 [PC]	0,021 [RL]	0,207 [RC]
[PL] : 0,704, [RC] : 0,206, [PC] : 0,045, [RL] : 0,045					

En étudiant ainsi la ségrégation simultanée d'un grand nombre de caractères, on peut mettre en évidence l'existence chez la *Drosophile* de quatre groupes de liaisons, de sept chez les pois, de dix chez le maïs.

On a vu que Mendel avait étudié sept caractères du pois. Par un hasard assez remarquable, ces sept caractères appartenaient chacun à l'un des sept groupes de liaisons présents chez le pois (ce qui n'avait *a priori* qu'environ une chance sur 163 de se produire); ils avaient par conséquent des ségrégations indépendantes; cette circonstance malheureuse (qui, toutefois, facilitait l'interprétation des résultats) a empêché Mendel de découvrir le « linkage ».

1.4. — Terminologie.

Comme toute science, la génétique, en devenant autonome, s'est donné un langage parfois inutilement spécifique, basé sur des définitions souvent imprécises, ou même variables d'un auteur à l'autre; le lecteur non initié rencontre de ce fait dans la lecture des ouvrages de génétique des difficultés parfaitement artificielles tenant non pas, comme il peut le croire, à la complexité des raisonnements tenus, mais à l'ambiguïté des termes employés. Il est donc important de préciser dès à présent les principaux « mots-clés » :

Gène. — Un gène est la quantité d'information concernant un caractère élémentaire, transmise par un parent à son descendant.

Dans le cas des organismes à reproduction sexuée, tout individu possède pour chaque caractère élémentaire, deux gènes provenant l'un de son père, l'autre de sa mère.

L'ensemble de tous les gènes possédés par les individus constituant une population peut être classé selon divers critères : individu auquel ils appartiennent, caractère sur lequel ils agissent, nature de l'action qu'ils ont sur ce caractère. A chacun de ces critères correspond la définition d'une relation d'équivalence, donc d'une partition de cet ensemble.

Génôme. — L'ensemble des gènes possédés par un individu constituent son génôme.

Pour moitié ce génôme lui a été transmis par son père et pour moitié par sa mère; lorsqu'il procrée, il transmet, à son tour, la copie d'une moitié de ce génôme à chacun de ses descendants.

Locus. — Les gènes agissant sur un même caractère élémentaire sont dits « appartenant à un même locus » ou « situés en un même locus » ⁽¹⁾.

Chaque individu possède pour chaque locus deux gènes transmis l'un par son père, l'autre par sa mère; il transmet à son tour, à chacun de ses descendants, la copie d'un des deux gènes qu'il possède en chaque locus.

Le locus est ainsi l'unité insécable de transmission de l'information héréditaire.

On peut dire également que les gènes situés en un même locus sont « homologues » d'où « appartiennent à la même classe d'homologie » (M. Gillois).

Allèles. — L'ensemble des gènes situés en un même locus, c'est-à-dire agissant sur un même caractère élémentaire, et ayant la même action sur ce caractère constitue une « classe d'allélisme ».

En un même locus il y a donc autant de « classes d'allélisme », ou plus simplement autant d'« allèles ». qu'il y a de modalités d'action sur le caractère concerné. Jusqu'ici nous n'avons cité que des locus comportant deux allèles (jaune-vert, lisse-ridé, ...) mais le nombre des allèles peut être quelconque; il n'est pas rare de trouver des locus occupés par plus d'une dizaine d'allèles. Par la suite nous désignerons généralement par les symboles $(a_1, \dots, a_i, \dots, a_n)$ les différents allèles, ou, lorsqu'il n'y en a que deux, par a et a' .

On peut dire également que les gènes représentant le même allèle sont « isoactifs » ou « appartiennent à la même classe d'isoaction » (M. Gillois).

Les deux gènes d'un même individu peuvent naturellement appartenir ou ne pas appartenir à la même classe d'allélisme; le génôme de cet individu est dit, selon le cas, « homozygote » ou « hétérozygote » pour ce locus.

Homozygotes. — Un individu est dit homozygote pour un certain locus, si les deux gènes qu'il possède en ce locus appartiennent à la même classe d'allélisme, c'est-à-dire ont la même action sur le caractère concerné. Sa formule génétique est de la forme $(a_i a_i)$.

S'il existe n classes d'allélisme, le nombre d'homozygotes distincts est $n : (a_1 a_1) \dots (a_i a_i) \dots (a_n a_n)$.

Hétérozygotes. — Un individu est dit hétérozygote pour un certain locus, si les deux gènes qu'il possède en ce locus n'appartiennent pas à la même classe d'allélisme, c'est-à-dire n'ont pas la même action sur le caractère concerné. Sa formule génétique est de la forme $(a_i a_j)$.

⁽¹⁾ La plupart des auteurs écrivent, au pluriel, des « loci » — cependant, une fois adopté par la langue française, un mot suit les règles de cette langue; il semble donc inutile de le décliner comme en latin. Nous dirons des « locus », comme l'on dit des « cumulus » ou des « sinus » —

S'il existe n classes d'allélisme, le nombre d'hétérozygotes distincts est $\frac{n(n-1)}{2} : (a_1a_2) \dots (a_1a_n) (a_2a_3) \dots (a_{n-1}a_n)$.

Génotypes. — Pour un locus occupé par n allèles, il peut y avoir n homozygotes et $\frac{n(n-1)}{2}$ hétérozygotes distincts. L'ensemble des formules génétiques possibles pour ce locus est dit l'ensemble des « génotypes » : il y a donc $\frac{n(n+1)}{2}$ génotypes.

Si les associations deux à deux des allèles dans les divers génotypes sont ainsi distinguées en autant de génotypes, leurs effets sur le caractère concerné ne sont pas obligatoirement différents. Les divers allèles peuvent en effet avoir entre eux des relations de dominance et de récessivité.

Dominance-Récessivité. — On dit que l'allèle a_i est dominant sur l'allèle a_j , ou, ce qui est équivalent, que l'allèle a_j est récessif devant l'allèle a_i , lorsque l'action d'un gène appartenant à la classe d'allélisme a_i rend inapparente celle d'un gène appartenant à la classe a_j , quand ils sont associés dans un génotype hétérozygote (a_ia_j) .

Lorsque cette dominance est totale, le génotype hétérozygote (a_ia_j) a le même effet sur le caractère concerné que le génotype homozygote (a_ia_i) .

Mais cette dominance peut être partielle, le génotype (a_ia_j) a alors un effet sur le caractère concerné plus proche de l'effet du génotype (a_ia_i) que de l'effet du génotype (a_ja_j) .

Il est bien évident que le plus souvent la présence d'un gène a_j , même s'il appartient à une classe d'allèle récessive, n'est pas sans conséquence sur l'organisme; la notion de dominance est donc fonction de l'acuité de l'observation ou de la précision avec laquelle est défini le « phénotype ».

Phénotype. — Le caractère effectivement manifesté par un individu correspond à son phénotype : alors que le génotype représente une réalité inaccessible, le phénotype est observable, et, éventuellement, mesurable.

Pour un caractère élémentaire gouverné par un locus comportant n allèles, le nombre de phénotypes est variable selon les relations de dominance existant entre ces allèles; il est compris entre n (si la relation de dominance-récessivité constitue une relation d'ordre entre les allèles et si cette dominance est totale) et $\frac{n(n+1)}{2}$ (autant que de génotypes).

Caractère élémentaire. — Un caractère élémentaire est un caractère dont la transmission se fait conformément à la Première Loi de Mendel : c'est-à-dire qui est gouverné par des gènes situés en un seul locus.

Cette dernière définition met en évidence la tautologie sous-jacente à la théorie exposée jusqu'ici. Cette théorie ne serait qu'un « modèle » logique purement abstrait si les découvertes biologiques, postérieures aux travaux de

Mendel, n'avaient montré que les mots « gènes » ou « locus » (et aussi « groupes de liaison ») peuvent désigner non seulement, comme nous l'avons vu jusqu'ici, des concepts permettant d'expliquer les observations concernant la transmission des caractères, mais aussi des objets matériels parfaitement concrets.

Comme en bien d'autres domaines, le raisonnement logique a fourni un schéma explicatif des phénomènes de l'hérédité avant même que soient connus les mécanismes qui les réalisent; les découvertes cytologiques ultérieures ont permis d'affirmer que les principes mendéliens constituaient non seulement un modèle décrivant de façon cohérente les déterminismes impliqués dans la transmission des caractères, mais correspondaient à une réalité biologique.

2. — FONDEMENT BIOLOGIQUE : LES CHROMOSOMES

Une dizaine d'années après les travaux de Mendel, les cytologistes constatèrent la présence, dans les noyaux des cellules, de structures filamenteuses décelables par teinture lors des divisions cellulaires : les chromosomes. Toute cellule vivante possède un ensemble de n paires de chromosomes, le nombre n étant caractéristique de chaque espèce : quatre chez la drosophile, sept chez le pois, dix chez le maïs.

Constatant la rigueur du comportement de ces chromosomes, les cytologistes pensèrent qu'ils pouvaient jouer un rôle important dans les phénomènes héréditaires; dans les dernières années du XIX^e siècle, ils commencèrent à étudier les règles qui devraient gouverner la transmission des caractères si effectivement les chromosomes en étaient le support. Lorsque fut retrouvé le mémoire de Mendel, on put constater que ces règles étaient justement celles qu'il avait indiquées.

Sans entrer dans les détails d'un domaine où de nouvelles connaissances s'accumulent constamment, il paraît nécessaire de décrire ici très sommairement la réalité physiologique dont la « génétique de population » est le reflet.

2.1. — *Comportement des chromosomes. Mitose et Méiose.*

Chaque fois qu'une cellule s'apprête à se dédoubler, ce qui se produit des milliards de fois au cours de la croissance d'un organisme, elle est le siège de phénomènes complexes dont le principal concerne les chromosomes : invisibles au sein du noyau durant le repos de la cellule, ceux-ci apparaissent et prennent progressivement l'aspect de bâtonnets, de formes diverses, classables par paires (phase 1 de la figure 2).

Puis chaque bâtonnet se dédouble en deux « chromatides » identiques qui restent attachés, en un point, le « centromère », cependant que la membrane qui entoure le noyau disparaît (phase 2).

Entraînés par leurs centromères les chromosomes ainsi dédoublés viennent

se placer à l'équateur de la cellule (phase 3); puis le centromère se dédouble, les chromatides « sœurs » se séparent; comme attirées par les deux extrémités de la cellule, elles se regroupent alors en deux ensembles nouveaux comportant l'un et l'autre un exemplaire de chacun des chromosomes (phases 4 et 5).

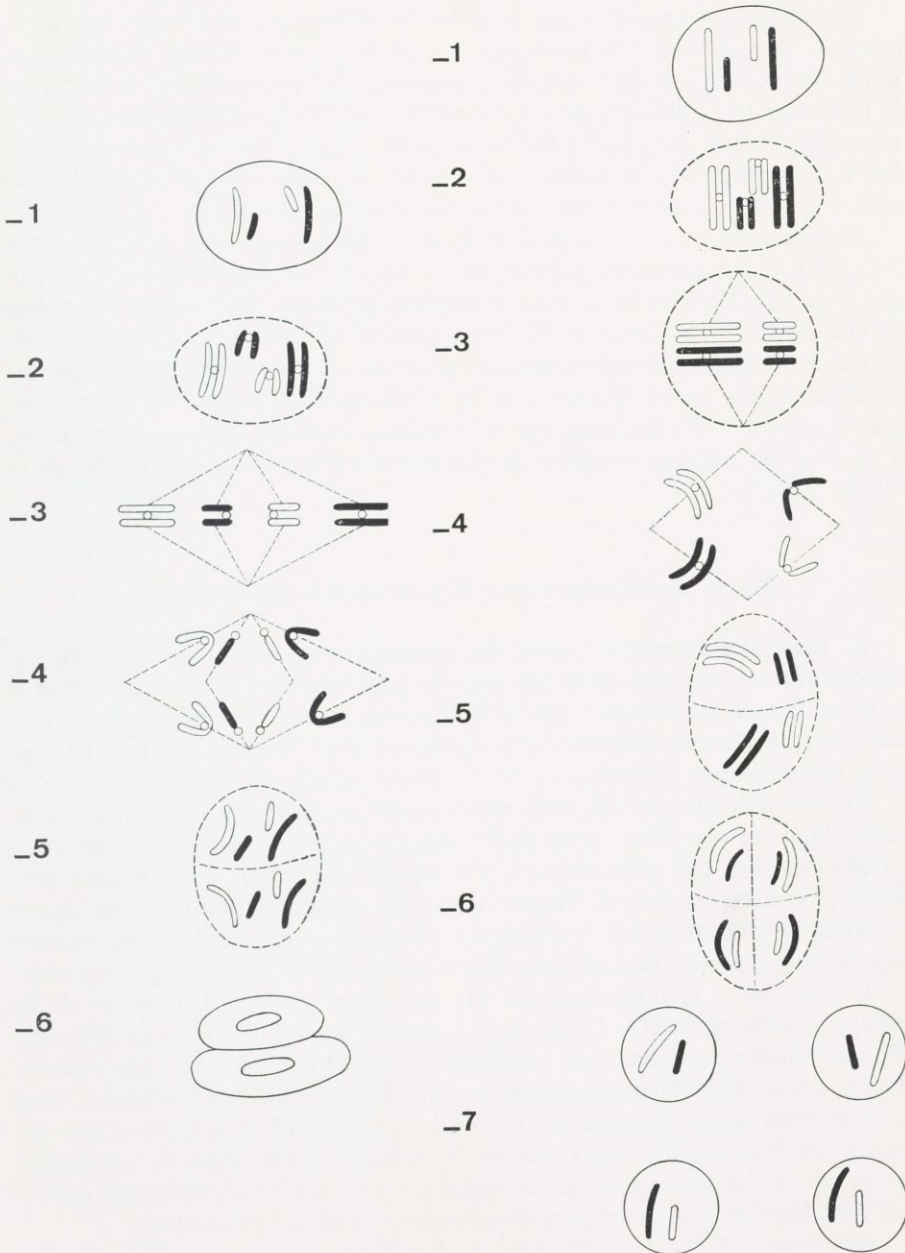


FIG. 2. — « Schéma de la Mitose. »

FIG. 3. — « Schéma de la Méiose. »

Une membrane se reconstitue autour de chacun de ces ensembles pour constituer deux nouveaux noyaux dans lesquels peu à peu les chromosomes redeviennent invisibles. Chacune des nouvelles cellules possède ainsi, grâce à ce ballet bien réglé, que les cytologistes nomment la *Mitose*, une copie intégrale et parfaite de tous les chromosomes de la cellule initiale.

Mais ce ballet comporte une variante, la *Méiose*, qui se produit lors de la création des cellules reproductrices, ou gamètes. Les choses commencent, semble-t-il, comme pour la mitose : apparition des chromosomes (phase 1 de la figure 3), dédoublement en 2 chromatides liées par un centromère (phase 2), migration vers le plan équatorial de la cellule; mais, lors de cette migration, les deux chromosomes homologues de chaque paire viennent l'un contre l'autre réalisant ainsi des groupes de quatre chromatides (phase 3).

Une première division se produit ensuite : les chromatides, toujours attachées par leur centromère, migrent aux deux pôles de la cellule de telle façon que chaque pôle reçoit un élément (dédouble) et un seul de chacune des paires de chromosomes (phases 4 et 5). Deux cellules se constituent provisoirement autour de ces pôles; ces cellules se divisent ensuite chacune en deux, en séparant les chromatides sœurs (phases 6 et 7). Finalement, au terme de la méiose, quatre cellules reproductrices ont été réalisées, chacune possédant une copie d'un, et d'un seul, des membres de chacune des paires de chromosomes de la cellule initiale.

2.2. — Conséquences pour la transmission des caractères.

La régularité du comportement des chromosomes permet de poser l'hypothèse suivante : les caractères biologiques transmissibles d'une génération à la suivante sont sous la dépendance d'éléments matériels portés par les chromosomes; à chaque caractère correspondent deux éléments portés par les deux chromosomes homologues d'une même paire.

Chaque individu possède deux séries parallèles de n chromosomes, transmises l'une par son père, l'autre par sa mère; si les éléments reçus de ses deux parents ne sont pas semblables, il y a conflit et l'action de l'un peut prédominer et masquer celle de l'autre; mais, lors de la création de la génération suivante, les chromosomes homologues, paternels et maternels, se retrouvent séparés et le caractère masqué peut à nouveau se manifester : il y a *ségrégation*.

D'autre part, la migration des chromosomes vers les deux pôles de la cellule, lors des phases (3, 4 et 5) de la « méiose » se réalise indépendamment dès leur origine maternelle ou paternelle. Chacune des deux cellules intermédiaires qui se dédoublent pour donner les cellules reproductrices finales, peut donc recevoir, dans l'exemple très simple de la figure 3 où $n = 2$, aussi bien des copies des deux chromosomes paternels que des copies des deux chromosomes maternels ou des copies d'un chromosome paternel et d'un chromosome maternel : les ségrégations des divers chromosomes sont *indépendantes*.

Ces deux notions de ségrégation et d'indépendance sont justement celles des « lois » de Mendel : ainsi, ce qui n'était qu'une explication formelle se

trouve être en totale conformité avec une réalité physiologique; il est équivalent, dans la description de l'hérédité, d'admettre les principes de Mendel ou de poser l'hypothèse que les chromosomes sont les supports des caractères héréditaires.

Les « gènes » qui étaient chez Mendel des « facteurs » explicatifs de l'hérédité peuvent alors être vus comme des « éléments » physiques portés par les chromosomes.

Reste cependant à expliquer le phénomène du « linkage ».

2.3. — Linkage et Crossing-over ⁽¹⁾.

Selon le schéma simplifié de la méiose qui vient d'être décrit, deux caractères subiraient des ségrégations soit indépendantes, s'ils sont sous la dépendance de gènes situés sur deux paires de chromosomes distinctes, soit totalement liées, s'ils sont sous la dépendance de gènes situés sur une même paire de chromosomes.

En réalité, lorsque les chromosomes d'une même paire, après s'être dédoublés, sont accolés l'un à l'autre (phase 3 de la méiose), des échanges peuvent se produire entre les chromatides.

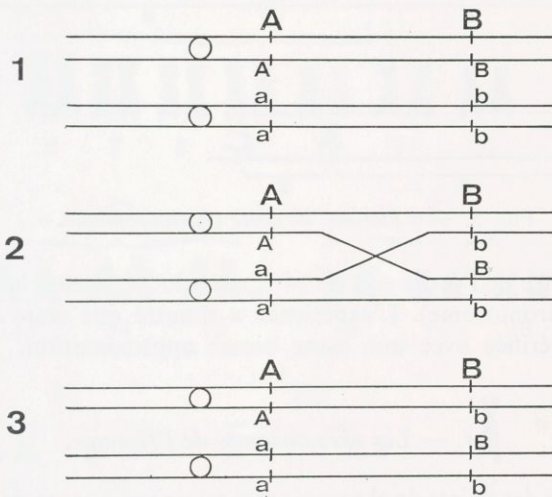


FIG. 4. — « Schéma du Crossing over. »

La figure 4 illustre le cas où les chromosomes d'origine paternelle portent, pour deux caractères, des gènes A et B, et les chromosomes d'origine maternelle des gènes a et b.

Après l'échange de deux segments entre deux chromatides (échange que

⁽¹⁾ L'usage a consacré l'emploi de ces deux vocables d'origine anglaise. Nous admettons donc que ce sont des mots aussi français que poker ou rugby et nous les utilisons sans chercher à les franciser.

les cytologistes ont dénommé *crossing-over*), les chromosomes qui migrent vers les deux extrémités de la cellule comportent des chromatides où les caractères sont restés associés comme chez les parents (AB et ab), d'autres où ils se trouvent « recombinés » (Ab et aB).

Plus grande est la distance entre les points du chromosome qui portent les gènes gouvernant deux caractères, plus grande est la chance qu'un *crossing-over* se produise entre eux au cours de la méiose ⁽¹⁾, plus grande est donc la proportion de cellules reproductrices dans lesquelles ces caractères se trouvent recombinés; le *linkage* constaté lors de la ségrégation des caractères peut donc servir à mesurer la distance entre les points du chromosome. De ce fait des « cartes chromosomiques » ont été établies en recherchant le pourcentage de recombinaison des divers caractères pris deux à deux; la distance entre les points correspondant est alors exprimée, non au moyen d'une classique unité de mesure, mais au moyen d'un pourcentage.

Pour que cette mesure ait un sens, il faut évidemment que les pourcentages de recombinaisons soient, approximativement, additifs : si, par exemple, la distance entre A et B est de 6 %, entre B et C de 8 %, la distance entre

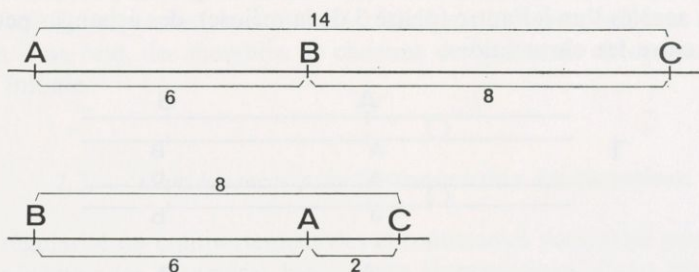


FIG. 5. — « Élément de carte chromosomique. »

A et C doit être de l'ordre soit de 14 %, soit de 2 % (selon la disposition des gènes sur le chromosome). L'expérience a montré que cette condition était effectivement vérifiée avec une assez bonne approximation.

2.4. — Les chromosomes de l'Homme.

Le nombre n des paires de chromosomes est une constante de chaque espèce. Sa connaissance est donc fondamentale. Pourtant ce n'est qu'en 1956 que le véritable nombre chromosomique de l'espèce humaine a été connu.

Depuis les travaux de Winiwarter en 1912, il était admis que chez l'homme $n = 24$. Le manque de précision des techniques employées, et sans doute aussi le respect excessif des affirmations antérieures, ont empêché la remise en cause de ce résultat jusqu'à ce que J. H. Tjio et Levan montrent, grâce à de nouveaux

⁽¹⁾ Nous ne nous préoccupons pas ici de la complication introduite par la réalisation de multiples *crossing-over*.

procédés de mise en évidence des chromosomes, qu'en réalité $n = 23$. L'espèce humaine a ainsi 46 chromosomes.

Ces 23 paires ont des formes maintenant bien connues permettant de les classer de façon rigoureuse selon une classification qui a fait l'objet d'une convention internationale. Cette convention distingue la paire XX ou XY de chromosomes « sexuels », ou « gonosomes », et les 22 autres paires de chromosomes dits « autosomes ». Ces derniers sont numérotés en fonction de leur taille et de la position de leur centromère, de 1 pour les chromosomes les plus grands, à centromère médian, jusqu'à 22 pour les chromosomes les plus petits à centromère extrême. La série complète des chromosomes d'un individu constitue son « Caryotype ».

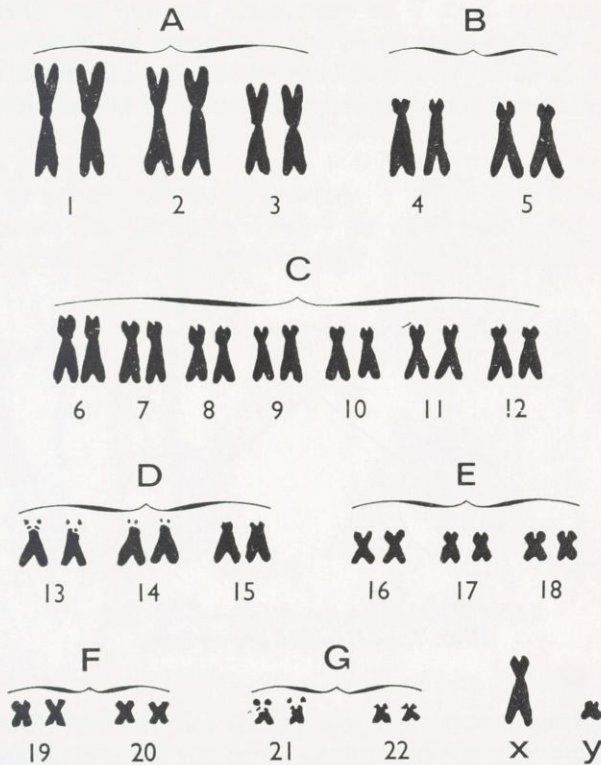


FIG. 6. — Caryotype d'un homme. »

On sait que le caryotype humain n'est pas une constante absolue, et que ses variations expliquent certaines maladies : ainsi en 1959, Lejeune et Turpin ont montré que le « Mongolisme » était dû à la présence d'un chromosome 21 en excédent, cette maladie est donc une « Trisomie 21 » (3 exemplaires du n° 21 au lieu des 2 normaux). De multiples exemples d'aberrations chromosomiques sont maintenant connus; de nombreuses observations ont permis de lier notamment certains types de cancer ou de leucémie à de telles aberrations.

2.5. — Les chromosomes sexuels.

Les chromosomes sexuels, ou gonosomes, différencient fondamentalement les caryotypes des femmes de ceux des hommes : chez les femmes la 23^e paire de chromosomes se présente sous la forme normale de deux chromosomes semblables, dits XX (que leur longueur apparente aux chromosomes 6), chez les hommes, au contraire cette 23^e paire comporte deux chromosomes de tailles très différentes, l'un de même apparence que les X des femmes et l'autre Y, très petit (semblable à un chromosome 21 ou 22).

Les chromosomes X et Y ne comportent pas que des gènes chargés de définir le sexe, d'autres caractères y sont également localisés. En raison de la dissymétrie du X et du Y chez l'homme, le mode d'hérédité de ces caractères est évidemment différent de celui des caractères liés aux « autosomes ».

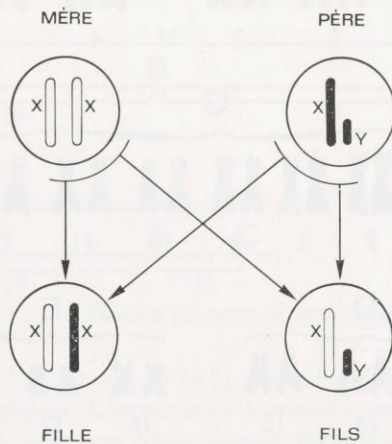


FIG. 7. — Hérédité liée au sexe.

En effet un garçon ne peut avoir reçu que de son père son chromosome Y, son X vient donc obligatoirement de sa mère; aucun des caractères gouvernés par un gène porté par l'X du père ne peut donc lui être transmis. Pour les filles, au contraire, les deux X viennent l'un du père, l'autre de la mère et les seuls caractères dont elles ne peuvent hériter sont ceux gouvernés par un gène porté par le Y de leur père ⁽¹⁾.

La dissymétrie entraînée par l'hétérogénéité de la paire XY chez l'homme

⁽¹⁾ Alors que de nombreux gènes sont situés sur le chromosome X (hémophilie, daltonisme), très peu de gènes portés par le chromosome Y ont pu jusqu'à présent être décelés.

rend ainsi nécessaire une étude spéciale de la transmission des caractères gouvernés par un gène porté par les chromosomes X ou Y, ou caractères « liés au sexe ».

2.6. — Structure des chromosomes. L'ADN.

Depuis une quinzaine d'années des progrès remarquables ont été accomplis dans la connaissance de la structure des chromosomes, structure qui doit permettre d'expliquer leurs deux propriétés fondamentales :

- auto-reproduction lors de chaque division cellulaire;
- commande des processus biologiques.

Grâce notamment aux travaux de Watson et Crick, l'on sait qu'un chromosome est constitué de deux longues molécules, enroulées en hélice l'une autour de l'autre, d'acide desoxyribonucléique, ADN.

Chaque molécule est constituée d'une suite d'éléments, les nucléotides, qui peuvent être de quatre types : A, si la base qui s'y trouve incorporée est l'Adénine, G si cette base est la Guanine, T si elle est la Thymine, C si elle est la Cytosine. Une molécule d'ADN est ainsi définie par la séquence des nucléotides qu'elle comporte, par exemple TCGAGCAAGCC...

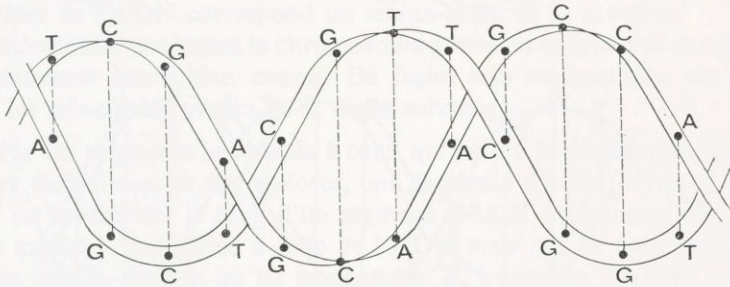


FIG. 8. — « La double molécule d'ADN. »

Les nucléotides sont ainsi les 4 « lettres » d'un alphabet avec lequel on « écrit » les molécules d'ADN.

De plus les deux molécules d'ADN qui se font face dans cette hélice sont complémentaires l'une de l'autre : à un nucléotide A de l'une correspond un nucléotide T de l'autre, à un G correspond un C.

En face du segment de molécule d'ADN indiqué ci-dessus se trouvera donc un segment AGCTCGTTCGG...

Auto-reproduction. — Lors de la première phase de la mitose ou de la méiose les deux molécules d'ADN se séparent. Mais la cellule contient à ce stade de nombreux nucléotides de type A, G, C ou T qui ont été préalablement synthétisés et sont restés à l'état libre; lorsque ces nucléotides se trouvent près de l'une des deux molécules disjointes d'ADN, ils sont « accrochés » par les

nucléotides de celle-ci qui leur sont complémentaires : T est accroché par A, C par G, etc...

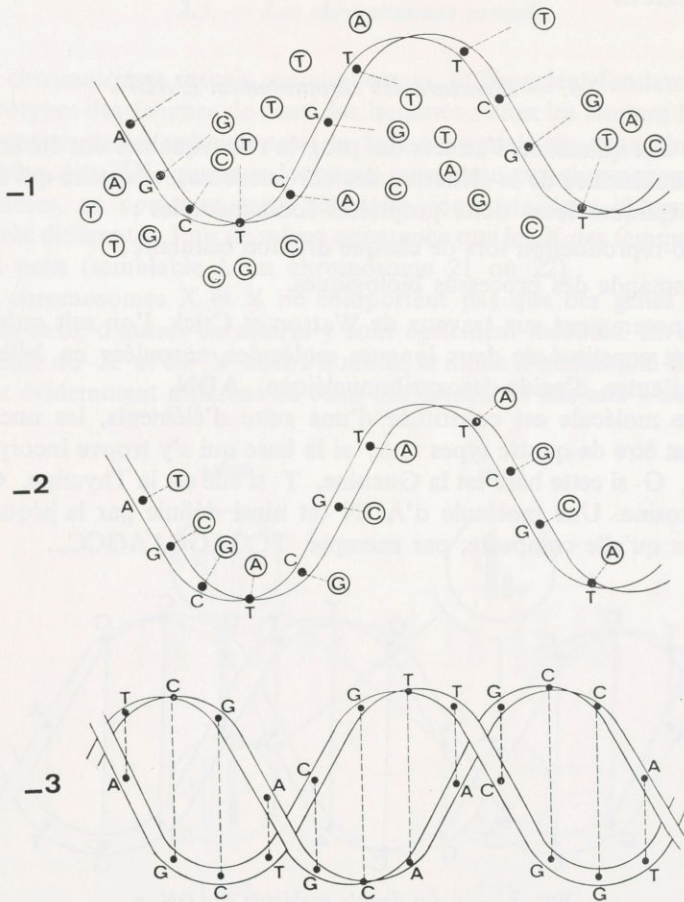


FIG. 9. — « Auto-reproduction de l'ADN. »

Puis, sous l'action probablement d'une enzyme, tous les nucléotides ainsi « piégés » se réunissent pour former à leur tour une molécule d'ADN, dont la séquence reproduit identiquement celle de la molécule initiale : chaque molécule d'ADN sert ainsi de moule lors de la synthèse de la molécule complémentaire.

Commande des processus bio-chimiques. — Les protéines, qui constituent l'essentiel de la matière vivante, sont des molécules énormes dont le poids moléculaire peut dépasser plusieurs millions d'unités. Rigides et compactes ces molécules sont formées de longues « chaînes polypeptidiques » repliées sur elles-mêmes en une géométrie rigoureuse, imposée par la structure même de ces chaînes. Les éléments qui forment ces « chaînes » sont des composés

organiques bien définis, les amino-acides, dont il existe 20 types auxquels ont été affectés des symboles alphabétiques :

Alanine	L	Leucine	U
Arginine	R	Lysine	I
Asparagine	N	Méthionine	M
Acide Aspartique	A	Phénylalanine	F
Cystéine.....	C	Proline	P
Acide glutamique	G	Sérine	S
Glutamine	Q	Thréonine	E
Glycine	Y	Tryptophane	T
Histidine	H	Tyrosine.....	O
Isoleucine	W	Valine.....	V

De même qu'une molécule d'ADN est définie par la séquence des nucléotides qui la constituent, une chaîne polypeptidique de protéine est définie par la séquence des amino-acides qu'elle comporte.

Les amino-acides sont ainsi les 20 « lettres » d'un alphabet avec lequel on « écrit » les protéines.

Les deux séquences « ADN » et « Protéine » sont, en fait, reliées l'une à l'autre par l'intermédiaire du « code génétique ». A chaque groupe de trois nucléotides de l'ADN correspond un amino-acide de la protéine.

Le mécanisme par lequel le chromosome gouverne la synthèse des protéines est maintenant assez bien connu. De façon très schématique, on peut en décrire les principales phases de la façon suivante :

1^o Par un processus semblable à celui qui assure la duplication de l'ADN au cours des mitoses et des méioses, une molécule d'acide oxyribonucléique, ARN, est synthétisée le long d'un segment d'ADN chromosomique. L'ARN est une molécule semblable à celle de l'ADN mais qui ne comprend qu'une hélice au lieu de deux et où les nucléotides T à base de Thymine sont remplacés par des nucléotides U à base d'Uracyle. Une séquence — AGCTCGAA... — d'ADN provoque donc la synthèse d'une séquence complémentaire — UCGAGCUU... — d'ARN.

2^o Cette molécule d'ARN, complémentaire de celle d'ADN, dite ARN messenger, sort du noyau, et pénètre dans le cytoplasme de la cellule. L'un des ribosomes présents dans ce cytoplasme vient alors se placer à l'une des extrémités de l'ARN.

Les divers types d'acides-amino sont présents dans la cellule soit à l'état libre, soit accolés à des molécules d'« ARN de transfert » ou ARN_t. Ces ARN de transfert sont de courtes molécules capables, d'une part de reconnaître et d'« accrocher » l'un des amino-acides, d'autre part de reconnaître par complémentarité un groupe de 3 nucléotides; tel ARN_t est capable de fixer une molécule de phénylalanine et de reconnaître la séquence UUU sur l'ARN messenger, tel autre de fixer une molécule de Sérine et de reconnaître la séquence UCA, ...

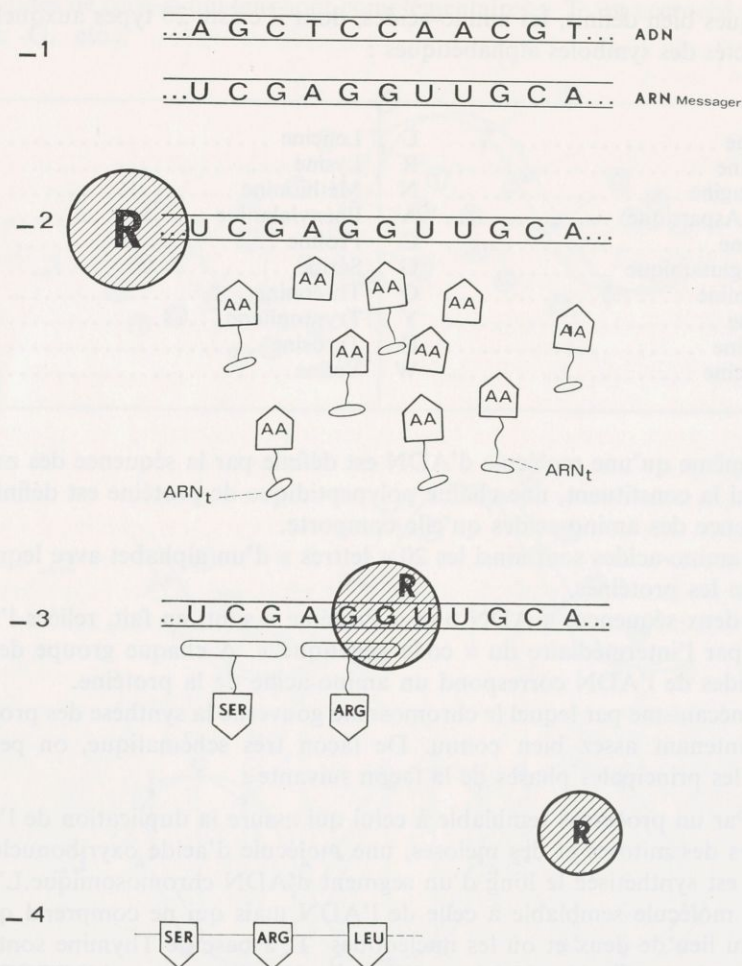


FIG. 10. — « Synthèse des chaînes polypeptidiques. »

La correspondance ainsi réalisée par les ARN_t entre les 20 amino-acides et les $4^3 = 64$ séquences de 3 nucléotides constitue le *code génétique*.

3° Le ribosome qui s'était fixé à une extrémité de la molécule d'ARN messenger progresse le long de cette molécule; dès qu'il a parcouru 3 nucléotides un ARN_t correspondant à cette séquence vient s'y fixer en entraînant l'amino-acide auquel il est accroché.

4° Les divers amino-acides, qui ont été ainsi disposés côte à côte dans un ordre imposé par celui des nucléotides de l'ARN messenger, s'unissent pour former une chaîne polypeptidique qui se libère des ARN_t.

Code génétique. — Une des particularités du code génétique maintenant connu dans son ensemble est d'être « dégénéré »; il comporte 64 séquences

distinctes, or il n'y a que 20 amino-acides : en fait certaines séquences ne correspondent à aucun amino-acide, ce sont des séquences « non sens » qui servent de « ponctuation » délimitant le début et la fin des chaînes polypeptidiques; d'autre part la plupart des amino-acides sont définis par plusieurs codes : UUG et CUC correspondent ainsi tous deux à la Leucine. Il semble d'ailleurs que ces codes multiples soient agencés de façon à protéger, autant que possible, les amino-acides les plus importants pour la cellule des erreurs de code provenant des mutations.

2.7. — Mutations.

La stabilité des chromosomes n'est pas absolue : avec une fréquence extrêmement faible mais non négligeable des modifications se produisent. Sous l'action de rayonnements, de substances chimiques ou de virus, ou par suite d'une simple erreur lors de la duplication de l'ADN, la séquence des nucléotides peut être changée : remplacement d'un nucléotide par un autre, élimination d'un ou plusieurs nucléotides, insertion d'un ou plusieurs nucléotides nouveaux.

Toutes les cellules ultérieures descendant de cette cellule « mutée » (le *clône* disent les biologistes) comporteront évidemment la même modification de leur ADN: s'il s'agit d'une cellule somatique l'individu concerné sera une « mosaïque », l'héritage génétique d'une certaine partie de son organisme sera différent de l'héritage génétique initial (tel semble bien être le cas de certaines cellules cancéreuses); s'il s'agit au contraire d'une cellule reproductrice, l'individu auquel elle donnera naissance aura un héritage génétique comportant un élément nouveau qui n'existait chez aucun de ses deux parents : une mutation apparaît, qui sera ensuite transmise aux descendants.

Ainsi le simple changement d'un nucléotide de l'ADN peut avoir des conséquences immenses pour l'ensemble de l'organisme; ce changement peut enlever tout sens à la partie correspondante de l'héritage génétique, rendant impossible la création de la protéine dont ce segment d'ADN gouvernait la synthèse; s'il s'agit, par exemple, d'une enzyme, toutes les activités biochimiques qu'elle conditionne ne pourront plus avoir lieu.

Un facteur important de stabilité vient corriger cette sensibilité de l'organisme aux modifications de l'ADN : pour chaque fonction la cellule dispose de deux gènes situés l'un sur le chromosome transmis par le père, l'autre sur celui transmis par la mère. Si l'un de ces gènes a subi une mutation qui annihile son action, cette circonstance reste le plus souvent cachée, car l'autre gène, intact, contient toutes les informations permettant d'assurer les fonctions normales de la cellule : les conséquences de cette dissymétrie, de cette « hétérozygotie », sont alors faibles ou nulles : ainsi s'explique le phénomène de récessivité et de dominance constaté par Mendel; un gène muté incapable de diriger la synthèse d'une protéine est récessif devant le gène qui a gardé cette capacité : si la quantité de protéine résultant de l'action de ce seul gène suffit au fonctionnement normal de l'organisme, la récessivité est totale;

si au contraire elle est insuffisante la récessivité est partielle, et le phénotype de l'hétérozygote diffère de celui de l'homozygote dominant.

Cependant lors d'une génération ultérieure, une cellule reproductrice portant le gène muté peut s'unir à une cellule reproductrice provenant d'un organisme ayant subi une mutation identique, l'individu résultant de cette union sera « homozygote » pour ce gène et l'effet de la mutation apparaîtra.

2.8. — *Diversité des individus.*

Les mutations ne sont pas nécessairement néfastes pour l'organisme. Elles constituent des tentatives de changement, réalisées sans doute au hasard, mais qui peuvent se révéler bénéfiques. Elles sont la source de la diversité des individus.

Pour donner une mesure de cette diversité, il suffit de rappeler que si les membres d'une population peuvent différer par N caractères ayant chacun deux aspects, il y aura au total 2^N types possibles d'individus.

La population totale de la terre est actuellement de $3 \cdot 10^9$ hommes, le nombre total des hommes ayant vécu depuis le début de l'humanité est sans doute inférieur à 10^{11} : il suffirait donc qu'ils aient différé par 40 caractères seulement pour qu'ils puissent être tous distincts.

Le nombre total de spermatozoïdes émis par tous les hommes ayant vécu est de l'ordre de 10^{22} : or les spermatozoïdes d'un homme hétérozygote pour 75 caractères, ce qui n'est pas considérable, peuvent être de $2^{75} = 10^{23}$ types distincts.

Comment cette diversité se maintient-elle? Comment, en fonction des conditions de vie et du comportement des hommes, les proportions des divers types d'hommes évoluent-elle? A ces questions la génétique de population s'efforce de répondre.

BIBLIOGRAPHIE.

- GEDDA L., 1955, *Novant anni delle Leggi Mendeliane*, 494 pages, Istituto Gregorio Mendel, Roma.
- RIEGER R., MICHAELIS A. et GREEN M. M., 1968, *A glossary of genetics and cytogenetics*, 507 pages, Springer-Verlag, Berlin.
- ROBERT J. M., 1966, La génétique de la vie, *Cahiers Médicaux Lyonnais*, Lyon, **42**, 123-224.
-

CHAPITRE II

CONCEPTS ET NOTATIONS STRUCTURES GÉNÉTIQUES DES POPULATIONS ET DES INDIVIDUS

1. — PROBABILITÉS

Lorsque les chromosomes d'une même paire se séparent au cours de la méiose pour migrer vers des cellules reproductrices distinctes, la séparation des chromosomes d'origine paternelle et des chromosomes d'origine maternelle s'effectue « au hasard ». Lorsque les dizaines de millions de spermatozoïdes émis simultanément cherchent à fusionner avec l'ovule, un seul y parvient et le choix semble résulter du « hasard ».

L'étude des faits concernant l'hérédité implique ainsi constamment le recours aux concepts forgés pour raisonner au sujet des événements soumis « au hasard », c'est-à-dire aux probabilités.

Certes, des déterminismes rigoureux interviennent, en fait, pour commander la réalisation de tel ou tel événement, mais ces déterminismes sont inaccessibles à l'observation; seul est observable leur effet global, qui peut revêtir des aspects divers auxquels sont attachées des probabilités : il est impossible en examinant un spermatozoïde provenant d'un homme hétérozygote à un locus, de savoir si ce spermatozoïde porte le gène a , mais l'on peut affirmer, par exemple, connaissant son ascendance, que la probabilité pour qu'il porte a est $1/2$, pour qu'il porte a' $1/2$. Cette affirmation correspond évidemment à un certain état de l'information; si, par la suite, l'enfant produit par ce spermatozoïde a le génotype (aa), on saura que le spermatozoïde portait le gène a , à la probabilité aura succédé une certitude.

Les raisonnements probabilistes permettent de tirer les conséquences d'une information incomplète. Ils n'ont de sens que si cette information est définie, ils deviennent caducs si elle est modifiée.

La génétique représente sans doute le domaine où la méthode probabiliste correspond le mieux à la réalité des phénomènes étudiés. Paradoxalement, les premiers théoriciens de génétique de population n'ont guère eu recours à cette forme de raisonnement et se sont surtout préoccupés de modèles statistiques; depuis 1948, G. Malécot a repris et approfondi l'ensemble de ces modèles en utilisant systématiquement les concepts probabilistes.

Nous croyons utile de rappeler tout d'abord ici les éléments essentiels du calcul des probabilités.

1.1. — Définition des probabilités.

Considérons un événement E dont la réalisation dépend du résultat d'une « épreuve » qui n'a pas encore eu lieu, ou dont l'issue n'est pas encore connue; cette épreuve peut, par exemple, être une loterie.

Compte tenu des opinions, ou des informations, que nous avons sur la nature de cette épreuve, nous pouvons attacher, à chacun de ses divers résultats possibles e_i , un nombre p_i qui traduit le niveau de notre confiance dans sa réalisation. Par convention, nous choisissons des nombres p_i positifs et tels que

$$(1) \quad \sum_i p_i = 1$$

La probabilité d'un événement qui est réalisé si le résultat de l'épreuve est, par exemple, e_k , e_l ou e_m sera, par définition :

$$P(E) = p_k + p_l + p_m,$$

ou, de façon générale :

$$(2) \quad P(E) = \sum_{i \in R} p_i,$$

où R est l'ensemble des résultats de l'épreuve qui entraînent la réalisation de l'événement E .

Étant donné cette définition, si l'événement E est réalisé quel que soit e_i , c'est-à-dire si sa réalisation est certaine on a, d'après (1) :

$$P(E \text{ certain}) = 1,$$

si, au contraire, E n'est réalisé pour aucun des e_i , c'est-à-dire si sa réalisation est impossible, on a :

$$P(E \text{ impossible}) = 0.$$

La probabilité d'un événement est ainsi un nombre compris entre 0 et 1.

Exemple. — L'« épreuve » consiste à tirer au hasard une carte dans un jeu de bridge ; il y a 52 résultats possibles e_i ; compte tenu de la façon dont la carte est choisie, nous pouvons admettre que tous ces résultats sont également probables, on a donc :

$$p_1 = p_2 = \dots = p_{52} = \frac{1}{52}.$$

Considérons l'événement E_1 : « la carte tirée est un carreau », cet événement est réalisé pour 13 des 52 résultats de l'épreuve, la probabilité de E est donc :

$$P(E_1) = 13 \times \frac{1}{52} = \frac{1}{4}.$$

Considérons l'événement E_2 : « la carte tirée est un roi », cet événement est réalisé pour 4 des 52 résultats de l'épreuve, la probabilité de E_2 est donc

$$P(E_2) = 4 \times \frac{1}{52} = \frac{1}{13}.$$

1.2. — Principe des probabilités totales.

Deux événements E_1 et E_2 sont dits *incompatibles* lorsque leur réalisation est subordonnée à des résultats non communs de l'épreuve. Par exemple si E_1 est réalisé lorsque ce résultat est soit e_3 soit e_5 , et E_2 lorsqu'il est soit e_4 soit e_8 , E_1 et E_2 ne peuvent être réalisés simultanément : ils sont incompatibles.

Un événement E_R est dit la *réunion* des deux événements E_1 et E_2 , s'il a lieu soit lorsque E_1 est réalisé, soit lorsque E_2 est réalisé, soit lorsque E_1 et E_2 sont simultanément réalisés (dans le cas où E_1 et E_2 sont incompatibles cette troisième éventualité est impossible); on écrit alors :

$$E_R = E_1 \cup E_2.$$

Il est facile de voir, d'après la formule de définition (2) que, si deux événements sont incompatibles, la probabilité de leur réunion est égale à la somme de leurs probabilités. De façon générale :

— si les événements E_i sont incompatibles deux à deux, on a :

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i).$$

Exemple. — Dans l'épreuve consistant à tirer une carte dans un jeu de bridge, les événements E_1 « la carte est un carreau », et E_3 « la carte est un cœur » sont incompatibles. La probabilité de l'événement E_4 « la carte est d'une couleur rouge » est donc

$$P(E_4) = P(E_1) + P(E_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

1.3. — Principe des probabilités composées.

Un événement E_I est dit l'*intersection* des deux événements E_1 et E_2 s'il a lieu lorsque à la fois E_1 et E_2 sont réalisés et l'on écrit :

$$E_I = E_1 \cap E_2,$$

(dans le cas où E_1 et E_2 sont incompatibles, la réalisation de E_I est évidemment impossible).

On appelle *probabilité conditionnelle* de l'événement E_1 , sachant que E_2 est réalisé, le rapport de la probabilité de leur intersection à la probabilité de E_2 ; on désigne cette probabilité conditionnelle par la notation $P(E_1/E_2)$.

Autrement dit :

$$(4) \quad P(E_1/E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)},$$

définition valable à la seule condition que $P(E_2) \neq 0$.

Compte tenu de cette définition, on peut écrire les égalités ci-après qui expriment le « principe des probabilités composées » :

$$(5) \quad \boxed{P(E_1 \cap E_2) = P(E_1) \times P(E_2/E_1) = P(E_2) \times P(E_1/E_2)},$$

relation valable même si $P(E_1)$ ou $P(E_2)$ sont nuls (en effet dans ce cas tous les termes des égalités (5) sont nuls, bien que $P(E_2/E_1)$ ou $P(E_1/E_2)$ ne soient pas définis).

Événements indépendants. — Un événement E_1 est dit indépendant d'un événement E_2 , si la probabilité de E_1 n'est pas modifiée par une information concernant la réalisation de E_2 , autrement dit si :

$$P(E_1/E_2) = P(E_1)$$

Il en résulte, d'après (5), que : $P(E_1 \cap E_2) = P(E_1) \times P(E_2)$:

— si deux événements sont indépendants la probabilité de leur intersection est égale au produit de leurs probabilités.

Exemple. — Dans l'épreuve consistant à tirer une carte dans un jeu de bridge, considérons les événements :

E_1 : la carte est un carreau;

E_2 : la carte est un roi;

E_5 : la carte est un honneur (as, roi, dame, valet) de carreau;

$$\text{on a } P(E_1) = \frac{13}{52} = \frac{1}{4}, \quad P(E_2) = \frac{4}{52} = \frac{1}{13}, \quad P(E_5) = \frac{4}{52} = \frac{1}{13}.$$

On constate facilement en cherchant les résultats de l'épreuve qui correspondent aux événements intersection $E_1 \cap E_2$, $E_2 \cap E_5$ et $E_1 \cap E_5$, que

$$P(E_1 \cap E_2) = \frac{1}{52}, \quad P(E_2 \cap E_5) = \frac{1}{52}, \quad P(E_1 \cap E_5) = \frac{1}{13}$$

En appliquant la formule de définition (4) on obtient les six probabilités conditionnées :

$$\begin{aligned} P(E_1/E_2) &= \frac{1}{4} = P(E_1), & P(E_2/E_1) &= \frac{1}{13} = P(E_2) \\ P(E_2/E_5) &= \frac{1}{4}, & P(E_5/E_2) &= \frac{1}{4} \\ P(E_1/E_5) &= 1, & P(E_5/E_1) &= \frac{4}{13} \end{aligned}$$

Les événements E_1 et E_2 sont indépendants, mais non les événements E_1 et E_5 ou E_2 et E_5 .

1.4. — Théorème de Bayes.

Considérons un événement R qui peut résulter d'un certain nombre d'événements deux à deux incompatibles $C_1, \dots, C_n$, que nous désignons comme les « causes » de R . La connaissance des mécanismes impliqués nous permet d'attribuer, dans l'hypothèse où l'une ou l'autre des causes C_i se produit, une probabilité $P(R/C_i)$ à R ; d'autre part nous supposons que nous sommes en mesure d'attribuer aux divers causes des probabilités *a priori* $P(C_i)$.

Une expérience aboutit à la réalisation de R ; comment cette information nouvelle nous permet-elle de modifier les probabilités *a priori* $P(C_i)$? Autrement dit, quelles sont les probabilités $P(C_i/R)$ de C_i sachant que R est réalisé?

D'après (5) nous pouvons écrire :

$$P(C_1 \cap R) = P(C_1) \times P(R/C_1) = P(R)P(C_1/R).$$

Mais, R n'ayant lieu que si l'un des C_i s'est produit, on a :

$$P(R) = P(R \cap C_1) + P(R \cap C_2) + \dots + P(R \cap C_n) = \sum_i P(C_i)P(R/C_i)$$

d'où :

$$(6) \quad P(C_1/R) = \frac{P(C_1)P(R/C_1)}{\sum_i P(C_i)P(R/C_i)}$$

Cette formule qui exprime le « Théorème de Bayes », est souvent désignée comme « formule de la probabilité des causes », elle permet de tenir compte du résultat d'une expérience pour modifier l'opinion que nous avions initialement sur la probabilité d'un événement.

Exemple. — Soient deux jeux de bridge sur une table. Le premier J_1 est normal, dans le second J_2 deux jokers sont substitués à deux as. On mélange les deux jeux, on tire une carte, c'est un as; avec quelle probabilité appartient-elle au jeu J_1 ?

La probabilité *a priori* de prendre une carte dans un jeu ou dans l'autre est :

$$P(J_1) = P(J_2) = \frac{52}{104} = \frac{1}{2}.$$

Les probabilités de tirer un as dans chacun des jeux sont :

$$P(A/J_1) = \frac{4}{52} = \frac{1}{13}, \quad P(A/J_2) = \frac{2}{52} = \frac{1}{26}.$$

On cherche $P(J_1/A)$; l'application de la relation (6) donne :

$$P(J_1/A) = \frac{P(J_1)P(A/J_1)}{P(J_1)P(A/J_1) + P(J_2)P(A/J_2)} = \frac{1/2 \times 1/13}{1/2 \times 1/13 + 1/2 \times 1/26} = \frac{2}{3}.$$

REMARQUE. — Tous ces raisonnements reposent sur l'hypothèse initiale de l'existence d'une épreuve dont les divers résultats sont probabilisables; on ne peut tenir de raisonnement probabiliste que dans la mesure où le « mécanisme », dont dépend l'épreuve élémentaire, a été défini de façon telle que des probabilités puissent être affectées aux divers résultats.

Dans les exemples donnés, ce « mécanisme » était le choix d'une carte dans un tas, les résultats étaient le tirage des diverses cartes; on admettait que la façon de tirer une carte rendait tous ces résultats équiprobables, mais d'autres hypothèses auraient fort bien pu être adoptées.

1.5. — Variables aléatoires.

A tout résultat e_i de l'épreuve envisagée associons un nombre réel x_i ; nous connaissons les probabilités p_i de tous les résultats e_i , il est donc possible de calculer la probabilité P_k pour que le nombre associé au résultat soit $X = x_k$. Nous définissons ainsi une « variable aléatoire réelle X associée à l'épreuve ».

On désigne par « Loi de la variable aléatoire X » l'ensemble des probabilités P_i des événements $X = x_i$, les x_i étant les diverses valeurs que peut prendre X .

Exemple. — L'épreuve consiste à tirer une carte dans un jeu de bridge; la variable X est la « valeur Albaran » 1 pour le valet, 2 pour la dame, 3 pour le roi, 4 pour l'as, 0 pour les autres cartes. Les valeurs x_i sont : 0, 1, 2, 3, 4, et leurs probabilités sont :

$$P_0 = P(X = 0) = 9/13, \quad P_1 = P_2 = P_3 = P_4 = 1/13.$$

Ces cinq P_i définissent la « loi de X ».

1.6. — Espérance et variance d'une variable aléatoire.

On définit l'espérance d'une variable aléatoire comme la moyenne des diverses valeurs que peut prendre cette variable, pondérées par les probabilités attachées à ces valeurs :

$$(7) \quad E(X) = \sum_{i=1}^k P_i x_i.$$

Ainsi dans l'exemple de la « valeur d'une carte », l'espérance de cette valeur est

$$E(X) = 0 \times \frac{9}{13} + 1 \times \frac{1}{13} + 2 \times \frac{1}{13} + 3 \times \frac{1}{13} + 4 \times \frac{1}{13} = \frac{10}{13} = 0,769.$$

On voit sur cet exemple que l'espérance, qui est par définition un nombre certain, peut fort bien n'être pas l'une des valeurs x_i .