

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

JEAN DUFRÉNOY

Conception statistique de la liberté en biologie

Journal de la société statistique de Paris, tome 78 (1937), p. 260-288

http://www.numdam.org/item?id=JSFS_1937__78__260_0

© Société de statistique de Paris, 1937, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

II'

CONCEPTION STATISTIQUE DE LA LIBERTÉ EN BIOLOGIE

PROBABILITÉS INDÉPENDANTES

Nous considérons, comme l'expression de la plus grande liberté, les séries d'événements concomitants ou successifs tels que la manifestation d'un événement n'influe en rien sur la manifestation ou la non manifestation des événements concurrents.

La probabilité propre de manifestation de chaque événement étant p_1, p_2, p_3, \dots ; la probabilité de manifestation de tous (simultanément ou successivement) est $p_1 \times p_2 \times p_3 \dots$ ou p^n dans le cas où $p_1 = p_2 = p_3 \dots$

Exemples : 1° D'après les statistiques des naissances en Allemagne de 1901 à 1925, les probabilités de naissance d'un enfant, d'un couple de jumeaux, d'un groupe de 3 ou d'un groupe de 4... jumeaux multiples sont respectivement de l'ordre de $p, p/80, p/(80)^2, p/(80)^3$.

Selon la notation des statisticiens allemands, la fréquence relative des naissances de $x = 1$ (enfant), $x = 2$ (jumeaux), $x = 3$ (triplets), $x = 4$ (quadreplets) est donnée par la formule $\frac{1}{(n)^{x-1}}$ où n est le rapport du nombre de naissances d'enfant unique (42.107.657) au nombre (531.541) de naissance de jumeaux.

2° La fréquence des Biologistes présentant une Note à la Société de Biologie pendant un trimestre étant N (N est voisin de 350), la fréquence des Biologistes présentant x notes est respectivement $N \times (0.22)^x$.

La distribution de ces fréquences peut s'ajuster à la distribution $y = a^{(n-x)}$ où n représente le plus grand nombre de notes présentées par un même auteur.

Cette distribution se représente graphiquement par une droite sur un papier semi logarithmique.

Cet exemple illustre la loi des probabilités composées; voir par exemple Darmois (Stat. Math., p. 19) :

Quand un événement résulte du concours de 2 événements A et B de probabilités respectives $a|h, b|h$ indépendantes, sa probabilité est égale aux produits des probabilités de A et de B, soit $ab|h = a|h b|h$.

« Cette indépendance, au sens de la théorie des probabilités, est d'un caractère très spécial. Le fait de savoir que a est vrai ne laisse pas intact le champ des possibilités pour b . Il restreint seulement dans la même mesure le champ total et le champ favorable » (Darmois).

DÉPENDANCE DES PROBABILITÉS

Si les probabilités $a|h$, $b|h$ ne sont plus indépendantes, le quotient

$$\frac{ab|h}{a|h \cdot b|h}$$

n'est plus égal à 1, mais aux quantités :

$$\frac{a|bh}{a|h} = \frac{b|ah}{b|h}.$$

Dans l'exemple précédent, les ajustements des fréquences observées aux fréquences calculées dans l'hypothèse des probabilités indépendantes correspondent à des valeurs de χ^2 très probables pour les fréquences des auteurs présentant 1, 3, 4 notes, mais les fréquences observées sont systématiquement trop élevées pour les auteurs présentant 2 notes ou les auteurs présentant plus de 4 notes : la présentation d'une première note (a étant vrai) paraît augmenter la probabilité b de présentation d'une seconde note ; les auteurs qui présentent plus de 4 notes forment une population statistiquement différente de la population des auteurs présentant 1 à 4 notes.

D'ailleurs, le calcul des moments permet de caractériser l'une de ces distributions (C. R. Soc. Biol. CXV) par les constantes $\bar{x} = 1.354$; $\sigma^2 = 0,678$; $\beta_1 = 14,7$; $\beta_2 = 32,3$

Cette distribution peut se représenter par une courbe du type III.

Causalité et Corrélation.

A la notion de « causalité » qui lie indissolublement un résultat à une cause unique, se substitue la notion de « corrélation » qui exprime numériquement la force de la relation qui lie la « séquence » probable de deux événements.

Cette notion tient compte de ce qu'un événement n'est jamais « déterminé » uniquement par un certain événement préalable ou « facteur » mais par un ensemble d'événements ou de « facteurs » caractérisant l'histoire du système où se situent les événements.

Pour la commodité de l'étude expérimentale de la séquence des phénomènes on fait varier un des « facteurs » en supposant les autres constants, et dans le système ainsi défini, on considère le facteur variable comme la cause des phénomènes observés par suite de sa variation.

Ce fut le propre du génie de Mendel d'utiliser cette abstraction à l'étude de la transmission héréditaire de l'un des caractères d'une paire de caractères opposables, exclusifs l'un de l'autre (allélomorphiques); de substituer, à l'idée métaphysique d'Hérédité, la notion expérimentale d' « unités héréditaires », transmises selon les lois du Hasard; de permettre, par conséquent, de prévoir statistiquement les résultats d'un croisement.

Dans le cas simple des monohybrides tel que celui, étudié par Mendel, du

croissement entre pois ridé et pois rond, la détermination des hybrides est la suivante :

Quant aux caractères : les pois seront ou ronds ou ridés (aucun caractère nouveau n'apparaît quant aux proportions).

La première loi de Mendel est l'application à la génétique de la loi des probabilités indépendantes :

Considérons comme parents deux variétés de pois : l'une produit des grains ronds (nous l'appellerons RR), l'autre produit des pois ridés (nous l'appellerons rr).

fécondons une cellule femelle R du pois rond par un grain de pollen *r* du pois ridé.

Le résultat sera un œuf « hétérozygote » (Rr) qui se développera en un embryon contenu dans un grain de pois rond.

Ce pois rond germera en produisant un hybride de première génération (F₁). Cet hybride F₁, fleurissant à l'abri de toute contamination de pollen étranger, produira des gousses dans chacune desquelles 75 % des grains seront ronds et 25 % ridés « aux erreurs d'échantillonnage près ».

Mendel ayant fait cette observation fondamentale, l'interpréta ainsi : lorsque l'hybride F₁ forme ses cellules sexuelles, chaque cellule sexuelle reçoit le « facteur » R ou le facteur *r*. La fécondation recombine au hasard ces facteurs dissociés, suivant les termes du développement de :

$$(R + r) (R + r) = RR + 2 Rr + rr.$$

Faute de pouvoir distinguer l'un de l'autre (RR) et (Rr), nous ne distinguons que 2 classes « phénotypiques » : ronds (RR et 2 Rr) et ridés (rr) représentées approximativement avec les fréquences respectives 3 : 1.

Caractère héréditaire dominant. — Exemple de Mendel : le pois à grain rond.

1^o Un grain est rond si plus de 33 % des sucres qui s'y accumulent au cours de son développement se transforment en amidon. Il est ridé si les sucres demeurent en grande partie à l'état de « stachyose » soluble dans les solutions vacuolaires.

La cellule sexuelle (mâle ou femelle) d'un pois à grain rond contient et transmet cette aptitude à condenser les sucres en amidon, aptitude qui manque à la cellule sexuelle des pois ridés.

Appelons R la possibilité de condenser les sucres, *r* l'absence.

La fécondation d'une cellule femelle (R) par une cellule mâle (R) reproduira avec certitude une plante (RR); un « génotype » RR est « homozygote » pour le « caractère héréditaire » : *aptitude à condenser ces sucres*, condensation qui se manifeste à nos sens par le « caractère phénotypique » *grain rond*.

La fécondation d'une cellule sexuelle femelle R par une cellule sexuelle mâle *r* produira un génotype hétérozygote Rr, mais la présence de R suffit ici pour permettre une condensation d'amidon telle que le grain de génotype Rr soit aussi *rond* que le grain de génotype RR.

Chaque cellule sexuelle produite par le génotype Rr contiendra soit R, soit *r*;

il se formera autant de cellules sexuelles R (mâles ou femelles) transmettant l'aptitude à condenser l'amidon qu'il se formera de cellules sexuelles r ne la transmettant pas.

Chaque cellule sexuelle R (ou r) ayant autant de chances de rencontrer un partenaire de l'autre sexe, R ou r , les combinaisons à prévoir sont :

$$RR + 2 Rr + r = (R + r)^2.$$

Dans ce cas de « dominance », l'impossibilité de distinguer, à l'œil, le génotype RR des génotypes Rr, les fait réunir dans le même groupe phénotypique « ronds » : la proportion phénotypique devient $\frac{3}{4} + \frac{1}{4}$.

Les proportions mendéliennes de disjonction d'hybrides entre parents différant par 2, 3, n paires de caractères représentent des résultats numériques de la loi des probabilités composées, comme les proportions de la disjonction monohybride représentent ceux de la loi des probabilités simples.

Dans le cas des monohybrides, la proportion des phénotypes dominants et récessifs en F_2 peut s'exprimer sous la forme de la fraction 3 : 1 ou sous la forme de pourcentage $p = 0,75$; $q = 0,25$.

Dans le cas des hybrides entre parents différant par 2 ou n paires de caractères, cette proportion peut se calculer en comparant, au nombre des phénotypes récessifs, celui des phénotypes de tous les autres groupes.

Soit pour les dihybrides :

$$\begin{array}{lll} 15 : 1 & & \\ 9 : 7 & p = 0,5631 & q = 0,4375 \\ 13 : 3 & p = 0,8131 & q = 0,1091 \end{array}$$

Soit pour les trihybrides :

$$27 : 37 \qquad p = 0,4219 \qquad q = 0,5781.$$

Considérer un hybride comme un monohybride, c'est généralement négliger toutes les autres paires de caractères pour lesquelles cet hybride peut être hétérozygote, et ne considérer que la seule paire de caractères qu'il nous convient d'étudier.

Proportions mendéliennes des dihybrides. — Une application directe de la loi des probabilités composées indique la relation typique 9 : 3 : 3 : 1 des dihybrides.

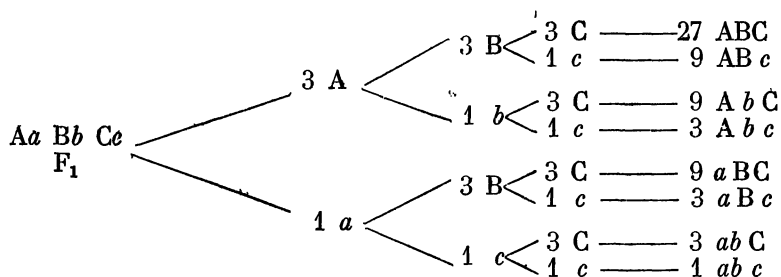
$$\begin{array}{l} 3 A \begin{array}{l} \left\langle \begin{array}{l} 3 B = 9 AB \\ 1 b = 3 A b \end{array} \right. \\ \\ 1 \begin{array}{l} \left\langle \begin{array}{l} 3 B = 3 a B \\ 1 b = 1 a b \end{array} \right. \end{array} \end{array}$$

déterminées mathématiquement par les lois de combinaison et de permutation.

Les classes de gamètes formées par un polyhybride représentent toutes les combinaisons qu'un membre d'une paire d'alléomorphes (hétérozygote) peut contracter avec chacun des autres alléomorphes représentés à l'état hétérozygote. Chaque classe est représentée par un nombre égal de gamètes.

La série zygotique F_2 s'obtient en réalisant entre deux séries gamétiques toutes les combinaisons possibles.

Les quatre classes phénotypiques F_2 correspondent, quant à leur nature et à la proportion de leurs représentants, aux classes gamétiques formées par le F_1 , ce qu'exprime la dichotomie suivante :



Les proportions phénotypiques que des études ultérieures ont montré correspondre à des groupements arbitraires trouvent rarement une justification par l'emploi d'un critérium statistique. Au contraire, les groupements phénotypiques que confirment les criteria statistiques trouvent une interprétation logique à la lumière de recherches de Biochimie ou de Cytologie. Ces proportions phénotypiques sont données par les développements des binômes :

$$\begin{array}{l}
 (3 + 1)^1 = 3 + 1 \\
 (3 + 1)^2 = 3^2 + 3 + 3 + 1 \\
 (3 + 1)^3 = 3^3 + 3(3^2) + 3(3) + 1 \\
 \dots \dots \dots \\
 (3 + 1)^n
 \end{array}$$

où n représente le nombre des paires de caractères tels que Rr , où R « domine » r .

Les proportions observées reproduisent des proportions théoriques aux erreurs d'échantillonnage près.

Caractères héréditaires additifs. — Dans l'exemple du pois rond opposé au pois ridé, le caractère R « rond » était opposable au caractère r « ridé ».

Dans d'autres cas chaque caractère d'une paire peut ajouter son effet à celui de l'autre : ce qui distingue alors les 2 caractères d'une paire, des 2 n caractères de n paires différentes, c'est que la présence, dans une cellule sexuelle, d'un caractère d'une paire, exclut nécessairement la présence de l'autre (chez une plante diploïde).

On appelle « allélomorphes » dans le sens le plus général, 2 caractères R et r , tels que la présence de R dans une cellule sexuelle exclut la présence de r (et vice versa).

Pour la commodité de l'écriture on désigne encore, par la majuscule et la minuscule de la même lettre, les 2 « allélomorphes »; les proportions génotypiques du produit de croisement de parents différant par n paires de caractères seront alors donnés par le développement du binôme $\left(\frac{1}{2} + \frac{1}{2}\right)^n$ puisque la présence hétérozygote Rr d'une paire produit la moitié de l'effet que produirait sa présence homozygote RR .

Quand n devient très grand, les « fréquences » dans les diverses classes génotypiques se distribuent selon la courbe binomiale normale.

De même qu'un « événement » n'est jamais déterminé par une « cause unique », un caractère héréditaire n'est pas, en dernière analyse, déterminé par un facteur unique, mais par l'interaction d'une série de facteurs. Dans les cas idéalement simples des monohybrides et des dihybrides, on suppose réalisés et maintenus constamment optima tous les autres facteurs, et on étudie les effets de la substitution, à l'un des facteurs d'une paire, de son alléomorphe, ou, plus exactement, de l'un de ses alléomorphes.

Un caractère, non plus apprécié qualitativement mais quantitativement, c'est-à-dire mesuré, pesé, dosé, se définit : chez un individu hybride, par un chiffre; chez chacun des autres hybrides descendant du même croisement, par un chiffre qui diffère plus ou moins du premier. Ordonnés statistiquement, tous ces chiffres se distribuent selon la « loi des écarts » autour d'une valeur moyenne. Cette valeur moyenne est la « déviation standard » qui mesure le degré de dispersion de ces chiffres autour de la moyenne définissant les « constantes statistiques » de la population que forment les hybrides étudiés.

Ces paramètres ne valent que pour la population étudiée; mais les lois statistiques permettent de transformer ces « constantes statistiques » en paramètres de valeur universelle, c'est-à-dire définissant la population infinie dont la population étudiée représente un échantillon.

Cette notion de population infinie, comprenant idéalement tous les représentants possibles des hybrides possédant les diverses combinaisons théoriques des caractères étudiés, c'est-à-dire comprenant 100 % des cas que notre imagination peut concevoir comme réalisables, correspond à un cas particulier de la certitude philosophique; elle définit, dans le cas particulier, la liberté.

Pratiquement cependant, les manifestations extrêmes de cette liberté, qui se situent le plus loin de la moyenne, sont si rares que nous pouvons les négliger. L'étude de la surface définie par la courbe normale des erreurs montre que les manifestations qui se situent à une distance de la moyenne qui est supérieure à trois fois l'erreur probable représentent moins de 1 % de toutes les manifestations possibles, et généralement on les considère comme négligeables.

La liberté en Biologie, telle qu'elle se manifeste par les formes d'êtres vivants, est définie, pour chaque lignée d'êtres, par le stock des caractères héréditaires des ancêtres et par les possibilités de recombinaison de ces caractères, dans la mesure où ces recombinaisons réalisent des possibilités de survie dans le milieu.

Les produits du croisement entre deux parents sont généralement considérés comme ayant des caractères intermédiaires entre ceux des parents. Chez les hybrides, des manifestations « transgressives » de caractères, prévues théoriquement en 1910 par East à la suite d'une généralisation rationnelle des lois de Mendel, ont été observées en 1912 par Hayes chez les Tabacs; depuis, leur importance est bien reconnue. Ces transgressions, dues à des recombinaisons entre « facteurs multiples », permettent d'étudier expérimentalement les faits groupés sous le vocable métaphysique d'Évolution.

« Limitée à l'étude de l'hérédité proprement dite, c'est-à-dire à l'analyse expérimentale de la constitution de l'organisme individuel, dans ce qu'elle a de stable, ayant permis de discerner des lois précises et générales, là où il n'y

avait que confusion et où il ne semblait régner que le caprice, ayant permis aussi de mieux nous rendre compte des rapports entre les diverses collections d'individus qui constituent l'espèce et par suite d'avoir pris de celle-ci une compréhension plus claire, la génétique — et les lois de Mendel qui en sont le centre — apparaît comme ayant réalisé dans la biologie et dans la science en général, un des progrès les plus substantiels de l'époque contemporaine ». (Caullery).

« Après avoir eu en France certains de ses plus importants précurseurs... la Génétique s'est développée avec une extrême rapidité en Angleterre, aux États-Unis, en Hollande, en Allemagne et dans divers autres pays. Elle n'a malheureusement trouvé chez nous que très peu d'adeptes... »

On pourrait généraliser aux autres applications agronomiques de la Statistique ce jugement formulé par le professeur Caullery; certains préfèrent conserver une représentation mythologique d'un Monde où ils seraient les jouets des caprices de la Nature, plutôt que d'en acquérir une représentation statistique capable de définir la Liberté humaine en fonction du déterminisme. « Un monde conçu comme la manifestation d'un déterminisme scientifique est le monde où le fatalisme a le moins de place » : (Thorndike).

Les Génétistes mettent du matériel vivant homogène, à la disposition des Physiologistes, pour leurs expériences, des Agronomes, pour leurs essais.

Les progrès de la technique permettent d'élever, dans des conditions expérimentalement déterminées, les individus d'une population composée de génotypes identiques ou de génotypes différant par 1, 2... n paires de caractères.

Les progrès de la statistique permettent de déterminer à l'avance les conditions auxquelles doivent satisfaire telle expérience pour fournir telle conclusion valable.

Croisement entre géniteurs différant par 5 paires de caractères (Hayes, 1912).

Hayes a croisé le tabac « Sumatra » formant en moyenne 28.2 feuilles (avec une déviation standard $\sigma = 1.5$) et le tabac « Broadleaf » formant en moyenne 19 feuilles.

Il obtient, à la première génération, des tabacs (F_1) intermédiaires entre les 2 parents et formant en moyenne 23.6 feuilles.

L'autofécondation d'un de ces hybrides F_1 (par exemple d'un hybride formant 24 feuilles) lui a fourni une population de 1.362 hybrides F_2 , formant en moyenne 22.7 feuilles ($\sigma = 2$).

Hayes émet l'hypothèse que les 2 parents diffèrent par un nombre de 10 feuilles; cette différence est imputable à 5 paires de facteurs désignés par les lettres majuscules AA, BB..... EE tels que la présence « homozygote » des 2 facteurs (AA) d'une paire augmente de 2 le nombre des feuilles, la présence « hétérozygote » d'un facteur d'une paire (par exemple Aa) augmentant de 1 le nombre des feuilles.

Selon cette hypothèse, la classe de plus grande fréquence doit être représentée par le nombre moyen des feuilles $\frac{19 + 29}{2} = \frac{48}{2} = 24$; de part et d'autre de cette classe ou « mode », les fréquences doivent se répartir selon les

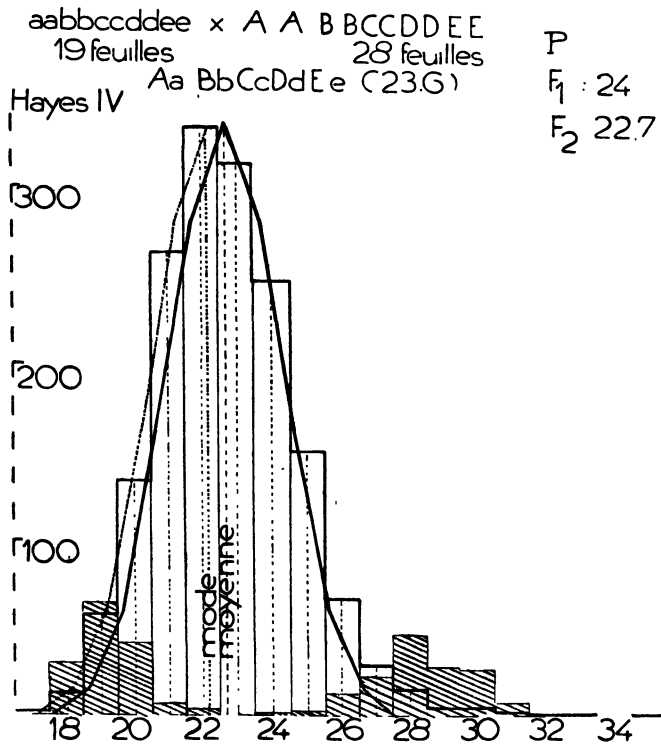
termes du développement du binôme $(0.5 + 0.5)^5$, c'est à dire selon les fréquences respectives (c) obtenues en multipliant par $\frac{1362}{1024} = 1,33$ les nombres de la ligne 5 du triangle de Pascal (tableau 1, p. 23). La fréquence calculée du génotype $aa\ bb\ cc\ dd\ ee$ du parent à 19 feuilles est $c = 1,33$; la fréquence observée est 2.

La fréquence du « phénotype » à 20 feuilles est 13,3; 2,66 individus appartenant à chacun des 5 génotypes $Aa\ bb\ cc\ dd\ ee$; $aa\ bb\ cc\ dd\ ee... aa\ bb\ cc\ dd\ eE$; la fréquence observée est 13...

Comparons les fréquences calculées (c) aux fréquences observées (f) : la branche ascendante de la courbe des fréquences observées peut s'ajuster à la branche ascendante de la courbe binomiale pour le même nombre d'individus pour la même moyenne et la même déviation-standard (graphique 1).

Par contre la branche descendante de la distribution des fréquences observées s'étale sur un nombre de classes plus grand que celle des fréquences calculées.

La distribution des fréquences observées est dissymétrique; dans son en-



Graphique 1 : Histogrammes hachurés : distribution des fréquences de nombre de feuilles chez les Tabacs Sumatra et Broadleaf. (HAYES, 1912.)

Au centre, histogramme des distributions de fréquences de nombre de feuilles chez 1.362 hybrides F_2 provenant de l'autofécondation d'un hybride F_1 Sumatra \times Broadleaf possédant 24 feuilles.

Ajustement, à la distribution observée, du polygone de fréquences correspondant au développement du binôme $(0,5 + 0,5)^5$.

En ordonnées, fréquences; en abscisses, nombre de feuilles.

semble elle peut s'ajuster à une courbe du type I, de part et d'autre du mode situé à 22.15.

La comparaison de l'histogramme de la distribution observée et du polygone des fréquences calculées par le développement binomial nous apprend :

1° La variance (4.16) de la distribution observée représente la somme de la variance de la distribution binomiale qui représente la « variance interclasse » pour 5 classes ($\sigma_s^2 = 2.5$) et de la variance « intraclasse ».

Nous pouvons adopter comme valeur moyenne de la variance intraclasse celle qui s'observe dans la classe des hybrides $F_1 Aa Bb Cc Dd Ee$ et qui est 1.7.

La somme $1.7 + 2.5 = 4.2$ est très voisine de la variance totale 4.16.

2° La distribution observée est dissymétrique, du fait que, dans chaque classe, la variance intraclasse est d'autant plus grande que le nombre des facteurs mis en jeu est plus grand, ce que permet de comprendre la théorie de Goldschmidt ou théorie des « seuils de réaction » .

En d'autres termes la dissymétrie de la distribution provient de ce que la variance d'un caractère mesurable est d'autant plus grande que son intensité moyenne est plus grande.

Dans l'hypothèse où la production de 2 feuilles dépend de la *présence* des 2 facteurs d'une paire, les tabacs de la classe « 18 feuilles » se caractérisent par le présence de 9 paires telles que :

ZZ; YY; XX RR

les tabacs de la classe 28 feuilles possèdent, en outre de ces 9 paires, 77.....RR les 5 paires AA..... EE.

Or, la présence d'un facteur tel que E constitue seulement une possibilité de manifestation du caractère observable ou phénotypique : « formation de la 28^e feuille par exemple »,

Signification statistique de la sélection.

On sait que la courbe en cloche, dite courbe normale des erreurs (ou la courbe en ogive, qui représente la courbe d'intégration de la précédente) représente graphiquement les résultats de tirage des boules d'une urne, lorsque les boules tirées sont remises dans l'urne, ou lorsque le nombre des boules tirées et non remises est petit par rapport au nombre des boules à tirer.

Si, au contraire, le nombre des boules « sorties » du jeu par tirage est relativement grand, la distribution des résultats du tirage se représente, non plus par une courbe normale, mais par la courbe IV de Pearson.

La biologie en général, la biochimie en particulier, nous offrent de nombreux exemples de modification irréversible. Une telle modification, affectant une proportion importante des individus d'une population finie, correspond statistiquement à un tirage des boules d'une urne sans remise des boules tirées. Conformément à ce que la théorie permettait de prévoir, la fréquence des manifestations de ces phénomènes en fonction du temps se représente par une courbe IV de Pearson.

La courbe normale des erreurs représente théoriquement la distribution des individus prélevés *au hasard* dans une population infinie; le lot des individus dont on étudie la distribution représente un « échantillon » de cette population

infinie, étant implicitement entendu que le hasard a joué librement dans le groupement des individus composant cet échantillon.

Il est une Science, la Génétique, qui a précisément pour objet de restreindre cette liberté du hasard dans la composition des populations : la technique qu'elle met en œuvre à cet effet est la sélection.

La sélection doit permettre d'isoler d'une population infinie, non plus un échantillon au hasard, mais un échantillon groupant des individus situés vers une certaine zone de la distribution de la population infinie.

En fait, la sélection n'est efficace que si elle porte sur une population hétérogène, sur un mélange de populations distribuées chacune normalement.

Dès lors, théoriquement, la sélection ne peut que permettre d'identifier telle ou telle distribution normale qui était intégrée dans la distribution générale de la population infinie.

Une distribution normale étant définie par les moments de 1^{er} ordre (moyenne) et de second ordre (σ), la sélection doit permettre d'isoler, de la distribution générale définie par sa moyenne générale $M_{\bar{x}}$ et sa déviation standard σ , les distributions élémentaires caractérisées chacune par une moyenne \bar{x} et une déviation standard σ_x : sélectionnant, à partir d'une même population originale, des maïs riches en protéines et des maïs pauvres. Winter obtient, au cours d'années successives, des populations distribuées autour de valeurs moyennes de plus en plus élevées (ou de plus en plus basses) en protéines avec une déviation standard de plus en plus réduite; ces distributions ne diffèrent jamais significativement de distributions normales.

Sélection du Tabac.

1^o Recherches de East.

EAST a croisé 2 variétés de *Nicotiana longiflora* différant par la longueur x de la corolle des fleurs.

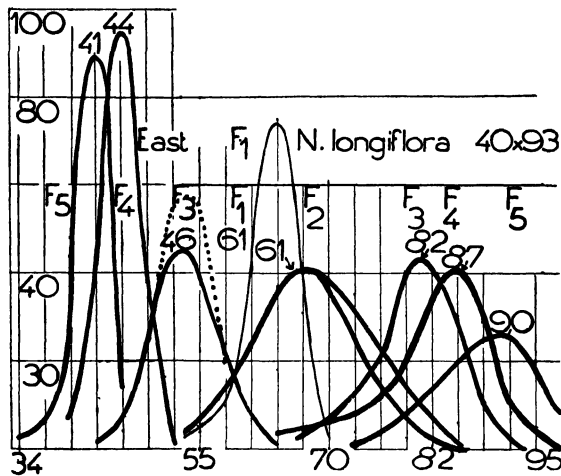
$$\begin{array}{l} \text{Variété 383, } \bar{x}_1 = 40,46, \sigma_{x_1} = 1,75 \\ \text{330, } \bar{x}_2 = 93,22, \sigma_{x_2} = 2,3 \end{array}$$

Les hybrides F_1 forment une population distribuée autour d'une moyenne intermédiaire $\bar{x} = 63,8$ avec une déviation standard de l'ordre de celle des parents $\sigma = 2,9$

Les populations F_2 se distribuent à peu près normalement ($\beta_1 = 0,042$, $\beta_2 = 3,06$ pour F_2 383 \times 330; 1) autour de moyennes voisines de 67, et avec des déviations standards de l'ordre de $\sigma = 6$; c'est-à-dire que 98% des variants environ sont compris entre la moyenne et $\pm 2 \sigma$.

Un parent F_4 ayant des fleurs de 90 mm donne une population F_5 distribuée symétriquement autour d'une moyenne de 87,8 avec une déviation standard $\sigma = 5,5$; ($\beta_1 = 0$, $\beta_2 = 2,5$, $K_2 = 0$; courbe type II) (graphique 2).

Un parent F_4 ayant des fleurs de 40 mm donne une population F_5 distribuée autour d'une moyenne 42 avec $\sigma = 5,5$ suivant une courbe du type IV.



Graphique 2 : 1° Vers la droite, effets de la sélection de parents à fleurs longues (F₁ = 61 mm ; F₂ = 82 mm. ; F₃ = 87 mm. ; F₄ = 90 mm.) sur la distribution des fréquences par classes de longueurs de fleurs chez les descendants.

La population des F₁ provenant du croisement entre *Nicotiana longiflora* à fleur de 40,5 (σ = 2) et *N. longiflora* à fleur de 93,2 (σ = 2,3) se distribue autour de 63,6 ; (σ = 3).

L'autofécondation d'un hybride F₁ à fleur de 61 donne une population de F₂ normalement distribuée autour de 67,5 ; (σ = 6).

L'autofécondation d'un hybride F₂ à fleur de 82 donne une population de F₃ distribuée autour de 80,2 ; (σ = 4,8).

L'autofécondation d'un hybride F₃ à fleur de 87 donne une population distribuée autour de 83 ; (σ = 5,8).

L'autofécondation d'un hybride F₄ à fleur de 90 donne une population F₅ distribuée autour de 88 ; selon une courbe du type II (β₁ = 0,354 ; β₂ = 2,7 ; σ = 5,5).

En ordonnées, fréquences ; en abscisses, classes de longueur des fleurs en millimètres. Les chiffres 61, 82, 87, 90 indiquent la longueur des fleurs des parents.

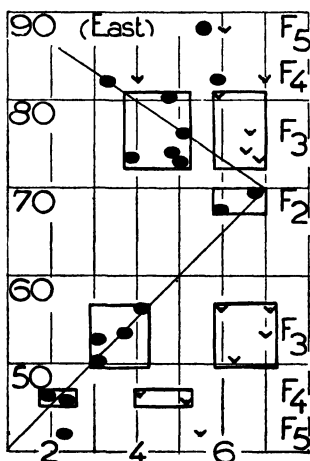
2° Vers la gauche, effets de la sélection des parents à petites fleurs : F₂ = 48 mm. ; F₃ = 44 ; F₄ = 41.

L'autofécondation d'un F₂ à fleur de 46 donne une population de F₃ distribuée autour d'une moyenne de 53 ; (σ = 3).

L'autofécondation d'un F₃ à fleur de 44 donne une population de F₄ distribuée autour d'une moyenne de 45,7 ; (σ = 2,37).

L'autofécondation d'un F₄ à fleur de 41 donne une population F₅ distribuée autour d'une moyenne de 42 ; (σ = 2,30).

(D'après EAST.)



Graphique 3 : Diminution de la grandeur du coefficient de variation (figurée par le signe ρ) et de la déviation standard (figurée par le signe .), (échelle des valeurs numériques portée en abscisses), en fonction du nombre des autofécondations successives fournissant les générations F₃, F₄, F₅ à partir de la F₂.

La diminution est beaucoup plus rapide dans les populations sélectionnées pour les petites fleurs, que dans les populations sélectionnées pour les grandes fleurs. (En ordonnées, à gauche, échelle des longueurs de fleurs de *N. longiflora*, d'après EAST).

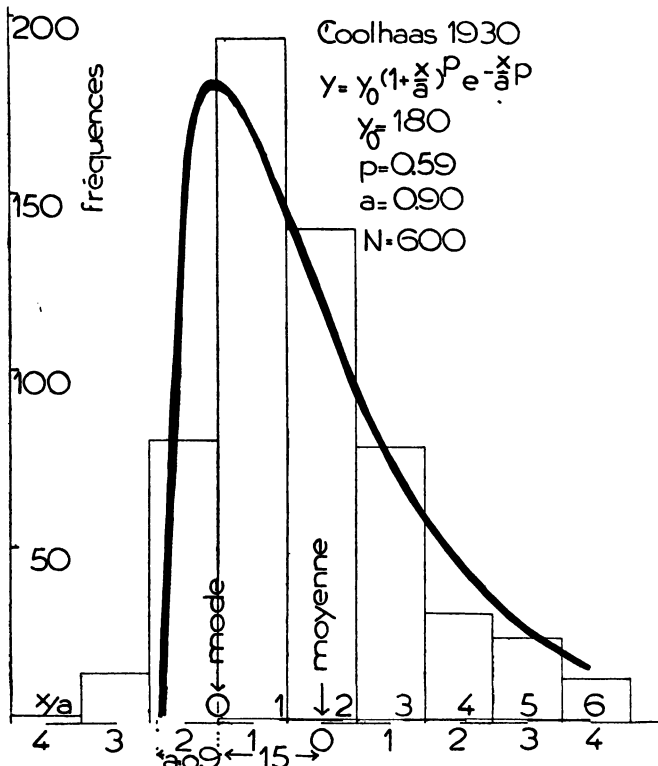
2° Recherches de Coolhaas.

La sélection, pour être efficace, doit porter sur les caractères mêmes dont dépend la qualité de la plante pour les usages alimentaires ou industriels.

Aux Indes Néerlandaises, pendant le siècle dernier, les porte graines de tabac formaient leurs fleurs sans protection; les croisements naturels entre les divers types cultivés avaient donné origine à un certain nombre de races locales dont l'ensemble formait une « population impure » à grande variance.

Au début de ce siècle, JENSEN et LODEWIJKS commencèrent de protéger les inflorescences des porte graines contre le pollen étranger, et dès 1925, dans une population de la variété Kanari qui paraissait homogène à l'œil, COOLHAAS pouvait distinguer une série de génotypes différant quant à la « vitesse de combustion » et à la « composition chimique » de leurs feuilles.

Ainsi apparaissait le danger de choisir, dans le champ, sans épreuve technique de la qualité et sur sa seule belle apparence extérieure, chaque plante à réserver comme porte graines, et de la soumettre à l'auto fécondation : une telle sélection tend à diminuer, dans l'ensemble de la progéniture, les chances de posséder les qualités techniques désirables, et, à mesure que les lignées deviennent de plus en plus homozygotes, ce sont les défauts qui tendent à être fixés.



Le polygone de fréquence des durées de combustion en secondes pour 600 échantillons de tabac d'après COOLHAAS (1930) : les classes de plus grandes fréquences correspondent aux courtes durées de combustion. Au contraire, le caractère désirable, longue durée de combustion, n'est possédé que par quelques individus.

La distribution peut s'ajuster à une courbe du type III de Pearson puisque $\beta_1 = 2.513$ et $\beta_2 = 6.82$ ce qui satisfait approximativement l'égalité :

$$2 \beta_2 = 6 + 3 \beta_1$$

soit :

$$2 \times 6,82 = 6 + 3 \times 2,513,$$

où :

$$13,64 = 6 + 7,54.$$

$$\gamma = \frac{2 \mu_2}{\mu_3} = \frac{7,24}{10,92} = 0,66$$

$$p = \left(\frac{4}{\beta_1} \right) - 1 = \frac{4}{2,513} - 1 = 0,59.$$

$$a = \frac{p}{\gamma} = 0,89. \text{ définit la distance du mode à la}$$

limite de la distribution.

Prenons pour origine le mode qui est à une distance $\frac{1}{\gamma} = 1.51$ de la moyenne.

L'équation de la courbe III peut s'écrire :

$$y = y_0 \left(1 + \frac{x}{a} \right)^p e^{-\frac{x}{a} \cdot p}$$

$$y_0 = p^p e^{-p} \frac{\gamma}{\Gamma(p+1)} \times 597 \text{ (Cf. Darmois p. 117).}$$

Utilisant la table de $\Gamma(p+1)$ (Darmois) notre regretté collègue M. Veizian a calculé :

log 597	=	2.775 9743
$p \log p$	=	1. 864 8027
$- p \log e$	=	1. 743 7662
log γ	=	1. 819 5439
$-\log \Gamma(p+1)$	=	0.049 4267
log y_0	=	2.253 5138
y_0	=	179.27

L'équation de la courbe III peut maintenant s'écrire :

$$y = y_0 e^{-p \left(\frac{x}{a} \right)} \left(1 + \frac{x}{a} \right)^p$$

où :

$$y_0 = 179.27 \quad p = 0.59 \quad a = 0,90$$

$$\log y = \log y_0 + p \log A$$

$$\log A = \log \left(1 + \frac{x}{a} \right) - \frac{x}{a} \log e$$

Sélection du coton :

Une « sélection pedigree » commencée au Congo belge en 1928, sans « isolement » des lignées, avait produit en 1931-1932, par suite d'hybridations naturelles, une population hétérogène.

A partir de 1931-1932, la sélection est orientée vers l'allongement et la régularité de la fibre; la dispersion des moyennes diminue progressivement avec l'allongement de la fibre. Les diagrammes publiés par WÆLKENS font nettement ressortir la marche de ce progrès.

D'année en année la moyenne des distributions se déplace vers les classes de grande longueur de fibres, sa variance diminue, et la forme même de la distribution semble se modifier.

*La variance comme criterium de la pureté d'une population :
application à la sélection de blés résistants aux « Rouilles ».*

Jusque dans ces récentes années, l'idéal pour les résultats expérimentaux paraissait être d'obtenir un ajustement aussi parfait que possible entre distributions théorique et observée, un tassement aussi complet que possible, autour de la moyenne, de tous les résultats numériques entrant dans le calcul de cette moyenne; les perfectionnements expérimentaux tendaient à réduire la dispersion des résultats numériques autour de cette moyenne, à réduire la variance, considérée comme une déféctuosité.

Il a fallu les travaux de Pearson, puis les recherches de Fisher sur l'*Analyse de la Variance*, pour faire comprendre aux agronomes l'intérêt qui s'attache à l'étude de la variance en elle-même et à la discrimination des différentes causes qui contribuent à donner à la variance sa valeur totale observée.

La variance affectant la moyenne de l'expression d'un caractère agronomique désirable mesure : 1^o le degré de pureté génétique de la population chez laquelle on étudie ce caractère, 2^o le degré de susceptibilité des individus composant cette population vis-à-vis des facteurs aduerses.

Un exemple nous en est fourni par les travaux classiques de U. S. Department of Agriculture : des croisements tels que Hope-Cerès, Hope-Marquis, Pentad-Marquis, et la sélection, parmi les produits de ces croisements, a permis d'obtenir des hybrides résistants aux « rouilles » causées par les divers *Puccinia*, possédant une haute qualité agronomique, et produisant du blé excellent pour la meunerie et par suite, de la farine excellente pour la boulangerie.

Le comportement d'un blé vis-à-vis des races de *Puccinia graminis tritici*, ou de *P. glumarum*, peut être apprécié d'après le nombre des lésions et l'aspect des lésions de rouille observées à la suite d'infections expérimentales; numériquement, il est commode de considérer les classes 0 %, 1 %, 5 %, 10 %, 15 %... de rouille.

Parmi les blés les plus cultivés aux États-Unis, la variété de *T. vulgare* appelée « Marquis » se caractérise, sur l'échelle conventionnelle d'appréciation de la susceptibilité aux attaques de *Puccinia graminis*, par une moyenne de 50 % de rouille. Dans le cours de l'hiver 1932-1933, trois croisements ont été effectués dans les serres du U. S. Department of Agriculture entre ce blé Marquis et trois génotypes issus du croisement « H₄₄-Cerès » caractérisés respecti-

vement par leur moyenne de rouille : 0,01 ; 3.66 ; 10.99 %. Les hybrides de première génération (F_1) furent cultivés en plein champ au cours de l'été 1933, ceux de deuxième génération (F_2) en serre au cours de l'hiver 1933-1934 et ceux de troisième génération (F_3) en champ en 1934. Ces hybrides de troisième génération se caractérisaient quant à la rouille par la moyenne (M) et l'erreur standard (E).

Croisements	M	E	Croisements	M	E
(0,1 %) × (50 %) :	10,22	± 1,10	(0,1 %) × (50 %) :	10,22	± 1,10
(3,66) × (50 %) :	21,16	± 1,45	(10,99 %) × (50 %) :	39,94	± 1,69
Différences :	10,94 ± 1,82		Différences :	29,72 ± 2,01	

Croisements	M	E
(3,66 %) × (50 %) :	21,16	± 1,45
(10,99 %) × (50 %) :	39,94	± 1,69
Différences :	18,78 ± 2,22	

L'importance des différences entre Moyennes (M) statistiquement appréciées en fonction des erreurs standards (E) a permis à Clark et Smith de conclure que les 3 génotypes utilisés pour les 3 croisements avec Marquis différaient quant à leur constitution génotypique. En appréciant les résultats des croisements antérieurs, soit quant aux distributions des hybrides par classes de fréquence de pourcentage de rouille, soit quant aux valeurs moyennes et aux déviations standards représentant ces distributions, les auteurs confirment l'importance des méthodes statistiques (Year book Agriculture, 240-260, Washington, 1936).

Des blés des classes 0, 1, 5 % de rouille possèdent un « caractère dominant » d'immunité et, génotypiquement, se caractérisent par la présence d'une paire de facteurs d'immunité II (plus exactement, d'une paire de facteurs d'inhibition de la susceptibilité) qu'ils ont hérité de leur ancêtre *Triticum dicoccum*.

C'est en effet d'un croisement entre *T. dicoccum* et un *T. vulgare* (le blé dur de printemps Marquis) que dérivent les hybrides Hope et H_{44} .

Le croisement de Hope (blé quasi immune parce que possédant II) avec Marquis (blé susceptible parce que possédant une paire de facteurs dominants de susceptibilité SS) donne une population d'hybrides dont les produits de ségrégation se répartissent, à la troisième génération (F_3), conformément à ce que permettait de prévoir la disjonction mendélienne des 2 paires de facteurs invoqués :

II d'immunité, c'est-à-dire d'inhibition de la susceptibilité;
SS de susceptibilité.

Par exemple, les hybrides fixés du type Hope ou H_{44} issus du croisement des génotypes II SS × *ii ss* (*Triticum dicoccum* × *T. vulgare* var. *Marquis*) appartiennent à l'un des génotypes II *ss*, II SS ou *ii ss*.

1° Parmi les descendants H_{44} un premier génotype II *ss*, possédant le facteur dominant II d'immunité et le facteur récessif *ss* de résistance (absence de caractère SS de susceptibilité) appartient aux classes 0, ou 1 % de rouille; plus exactement, il sera caractérisé par une moyenne 0,1 % de rouille avec une variance très faible. Le croisement avec Marquis, manifestant en moyenne

49,4 % de rouille, permettra d'obtenir les produits de la ségrégation des deux génotypes correspondants : $II\ ss + ii\ ss$, c'est-à-dire qu'en F_3 les hybrides se distribueront autour d'une moyenne 10,22 % de rouille (avec une erreur standard 1,10) : sur 96 lignées étudiées, 27 appartiennent aux classes quasi immunes (0,1 % de rouille) et 4 seulement aux classes aussi susceptibles que le parent Marquis (40 à 55 % de rouille).

2° Un second génotype $II\ SS$, possédant à la fois le facteur de susceptibilité SS et facteur II d'inhibition de la susceptibilité appartiendra aux classes 1-10 % de rouille avec une moyenne $M = 3,66$.

Le croisement $II\ SS \times ii\ SS$ met en jeu une seule différence monofactorielle; en F_3 , les hybrides se distribuent en 2 groupes : le premier, surtout représenté par la classe 5 % de rouille; le second, par la classe 20 %; la moyenne générale étant 21,16 % avec une erreur standard de 1.45.

3° Enfin, le troisième génotype $ii\ ss$ ne possède pas la paire de facteurs dominants SS de susceptibilité; en d'autres termes, il possède la paire de facteurs récessifs alléломorphes ss de résistance. Le croisement $ii\ ss \times ii\ SS$ réalise une recombinaison monofactorielle; cependant, l'erreur standard (1,69) est plus élevée que dans le cas précédent, ce qui est corrélatif d'une susceptibilité moyenne plus élevée de la population.

En général, plus une population comprend d'individus susceptibles à un agent défavorable, plus grande est la variance que peut introduire l'effet de cet agent.

La caractéristique d'un fait pathologique, c'est d'exagérer la variance.

La « variance potentielle » d'une population représentée par un lot de plantes est d'autant plus grande que ce lot de plantes est plus hétérogène, plus « hétérozygote ».

Le « variance observée » pour une même variance potentielle sera d'autant plus grande que cette population hétérogène sera exposée à l'action de parasites, capables d'affecter les unes plus que les autres.

Les récents travaux effectués coopérativement par les « Céréalistes » du U. S. Department of Agriculture et du Dominion of Canada, en même temps qu'ils offrent une démonstration éloquentes des services que peut rendre l'analyse judicieuse de la variance dans la sélection des blés résistants aux rouilles et de qualité boulangère désirable, font comprendre la signification générale de la variance.

Modification des distributions, par la sélection.

Darwin, dans son *Origin of Species* et Galton dans ses *Laws of Inheritance*, pensant que la sélection des meilleurs individus dans une espèce permet d'améliorer les variétés de cette espèce, avaient conduit à croire que la sélection pouvait progressivement modifier une race en l'enrichissant successivement des petites mutations provoquées directement par la sélection elle-même, de par sa vertu propre de sélection.

Johannssen, cependant, a précisé que la sélection peut modifier seulement les races qui, génétiquement, sont hétérozygotes, qu'elle est sans effet sur les lignées homozygotes. La sélection ne peut pas modifier le génotype d'un indi-

vidu, mais peut seulement éliminer, de la descendance d'un hétérozygote, certains produits de la ségrégation des caractères parentaux; elle peut, par purifications successives, faire d'une race, représentée originairement par une population hétérogène, une race formée d'individus tous identiques et homozygotes pour les caractères étudiés.

Confirmant les résultats de Vilmorin, Johannssen a précisé qu'on peut purifier une race, en faire une lignée pure, par quelques autofécondations successives suivies de la sélection, parmi les phénotypes, de celui qui est reconnu correspondre à la présence du facteur recherché à l'état homozygote.

Dès 1916-1917, Jennings montrait qu'à partir d'une plante hétérozygote Aa , la proportion de zygotes des différents génotypes serait, après n autofécondations successives :

$$AA; \frac{2n-1}{2n+1} \quad Aa; \frac{1}{2n} \quad aa; \frac{2n-1}{2n+1}$$

c'est-à-dire qu'après 6 autofécondations, la proportion des homozygotes de constitution AA ou aa serait 98,4 % alors qu'elle n'était que 50 % à la première génération (F_2). Si la sélection avait éliminé les récessifs aa , la race aurait donc une pureté de 98,4 %, c'est-à-dire serait représentée par 98,4 % d'homozygotes AA (v. graphique 6, p. 23).

Robbins (1918), Wright (1921-1930)... Haldane et Waddington (1931) ont étudié statistiquement les conséquences des divers types de croisement et des divers taux de sélection.

Croisons un parent homozygote avec un autre parent homozygote, différent du 1^{er} par les caractères de n paires :

1° Les hybrides de première génération ou F_1 se distribuent autour d'une moyenne avec une variance du même ordre que la variance des parents.

(Cette variance est la somme de la variance intraclasse et de la variance due aux erreurs d'échantillonnage).

2° Un hybride F_1 pris vers une extrémité de la distribution des F_1 doit donner des hybrides F_2 dont la distribution se superpose (aux erreurs d'échantillonnage près) à la distribution des F_2 issus de n'importe quel autre F_1 (pris sur la moyenne ou à l'autre extrémité de la distribution des F_1).

3° La variance de la population F_2 est beaucoup plus grande que la variance des populations des parents ou des F_1 par suite de la manifestation de la variance interclasse. Généralement et surtout si les caractères pour lesquels les parents différaient sont cumulatifs, la distribution des F_2 tend à être dissymétrique, et à prendre la forme de courbe I par suite de la variance intraclasse plus grande dans les classes correspondant aux plus grandes intensités de manifestation des caractères (v. graphique 1).

La distribution des F_2 peut transgresser les distributions de l'un et l'autre parent.

4° Un individu situé à une extrémité de la distribution des F_2 donne, en F_3 , une population dont le mode et la moyenne diffèrent significativement du mode et de la moyenne d'une population provenant d'un individu situé à l'autre extrémité de la distribution des F_2 .

5° Dans les générations successivement obtenues par autofécondation, la variance devient plus petite que la variance en F_2 par suite de la diminution progressive des degrés de liberté entre classes. Théoriquement on peut imaginer que la distribution de chaque population F_n issue de l'autofécondation d'une hybride F_{n-1} tend vers le type IV;

Enfin lorsque les autofécondations successives ont rendu une population pure (ou homozygote), sa distribution doit devenir symétrique avec une variance (inférieure à npq) qui tend vers un minimum, correspondant à la somme des variances dues aux fluctuations et aux erreurs d'échantillonnage.

Diminution de la variance par sélection (courbe IV).

D'une urne contenant à l'origine A boules (dont A_p blanches A_q noires) on fait N tirages successifs en ne remettant pas les boules : la loi de répartition est très voisine d'une loi normale dont la variance $N \frac{A-N}{A} pq$ serait plus petite que la variance $N pq$ de la distribution normale (DARMOIS, *Stat. Math.*, p. 82).

Soit une courbe symétrique ($\beta_1 = 0$); si $\beta_2 = 3$ la courbe est normale et sa variance est npq . Si $\beta_2 > 3$ la courbe se rétrécit en élevant son sommet, et la variance devient inférieure à npq (courbe type VII).

Polya a donné une interprétation de certaines courbes de fréquence représentant les résultats de n tirages d'une urne contenant originalement $N\varrho$ boules rouges et $N\sigma$ boules noires ($\varrho + \sigma = 1$); après chaque tirage, on remplace la boule tirée par $1 + N\delta$ boules de la même couleur; $\delta = 0$ (boule remise); événements usuels indépendants (courbe normale).

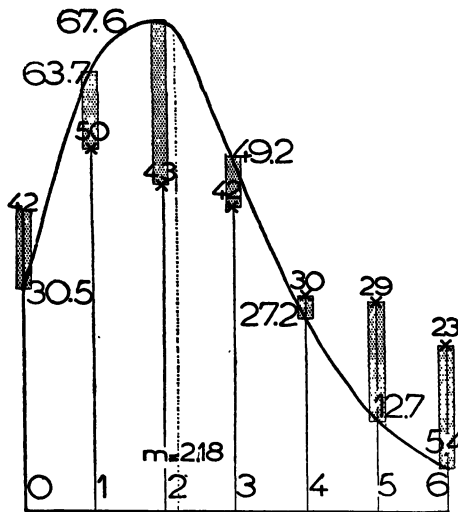
$N\delta = -1$ (boule non remise), la distribution tend vers le type IV de Pearson lorsque $\frac{n}{N}$ devient suffisamment grand.

$n\varrho = h$ constant; $\delta = 0$; événements rares indépendants, séries de Poisson (voir DARMOIS, *loc. cit.*).

ϱ constant; $N\delta = d$ constant; événements usuels avec faible contagion : la dispersion diffère de la normale, le sens de la différence étant celui de d .

Applications à la biologie et à l'épidémiologie : la sélection ou l'isolement tendent à empêcher les événements de devenir contagieux; des mesures prophylactiques prises au début d'une épidémie, dans une population finie, empêchent la distribution des fréquences de cas de prendre l'allure d'une courbe IV et lui imposent l'allure d'une distribution des séries de Poisson ou d'une distribution de probabilités composées.

La distribution des Séries de Poisson peut être déformée par l'effet de la sélection (graphique 5, p. 19) :



Sur 259 espèces bactériennes récoltées sur filtre d'égout d'eaux résiduaires de sucrerie, 42 ne peuvent utiliser aucun sucre; 50, 43, 42, 30, 29, 23 utilisent respectivement 1; 2; 3; 4; 5; 6; parmi les 6 sucres : sucrose, maltose, lactose, dextrose, galactose, lévulose.

Supposons que le pouvoir d'utiliser chacun de ces sucres (conféré par la présence de la diastase correspondante) soit distribué au hasard entre les espèces : les fréquences d'espèces possédant 0, 1, 6 diastases doivent se distribuer selon la courbe de série de Poisson calculée pour le nombre moyen de diastase par espèce, ici $m = 2.18$.

Les différences entre fréquences calculées et observées (figurées sur le graphique par la hauteur des rectangles grisés) sont faibles pour 0; 1; 2; 3; 4 sucres et correspondent à des valeurs de χ^2 dont la probabilité est respectivement 0.04; 0.10; 0.005; 0.20; 0.60; la déficience des espèces n'utilisant qu'un, deux ou trois sucres est corrélative d'un excès significatif des espèces capables d'utiliser cinq ou six sucres : ce qui était à prévoir, puisqu'un microorganisme capable d'utiliser l'un quelconque de nombreux aliments carbonés a plus de chances de survivre et de pulluler dans les sols (d'où proviennent les microorganismes des betteraves de sucrerie) qu'un microorganisme qui ne peut vivre qu'avec telle forme d'aliment carboné.

(Beaucoup des 42 espèces n'utilisant aucun sucre pullulent en utilisant les sels d'acides organiques libérés par les autres microorganismes utilisant 1, 2, 6 sucres.) (D'après CUTLER et CRUMP, *Problems in Soil Microbiology*, Rothamsted Monog. Agr. Sc., 1935, p. 46-47.)

De façon générale, si $\delta > 0$ chaque succès obtenu favorise les chances de succès à venir, le succès est contagieux.

Si $\delta < 0$ chaque succès obtenu diminue les chances de succès à venir.

Nous pouvons exprimer les mêmes notions ainsi :

1° L'introduction de nouveaux degrés de liberté dans une population augmente la variance en modifiant la distribution vers le type I de Pearson.

2° La destruction des degrés de liberté diminue la variance en modifiant la distribution vers le type IV ou le type VII de Pearson (type III de Polya).

La distribution de χ^2 qui est du type I lorsque le nombre des degrés de liberté est petit (inférieur à 10) tend à devenir normale lorsque le nombre des degrés de liberté augmente.

CORRÉLATION ENTRE VARIANCE ET NOMBRE DE DEGRÉS DE LIBERTÉ

Le jeu à pile ou face avec $n = 1, 2, \dots, n$ paires de pièces permet de réaliser, aux erreurs d'échantillonnages près, les fréquences respectives de pile ou face qui sont données par les chiffres du triangle de Pascal (tableau I, p, 23).

Par rapport à la fréquence de la classe du mode, les fréquences des classes inférieures ou supérieures se distribuent symétriquement, et cette distribution se rapproche de la normale à mesure que n augmente.

Si nous considérons n comme le nombre de degrés de liberté, nous constatons que le quotient, par le nombre total des variants de la population, des déviations par rapport à la moyenne, c'est-à-dire la variance σ^2 , augmente en même temps que n (graphique 6).

En génétique, East et Hayes ont montré que n désignant le nombre des caractères héréditaires par lesquels diffèrent deux parents quant à l'intensité de manifestation d'un caractère mesurable, les chiffres du triangle de Pascal indiquent les fréquences respectives avec lesquelles, à la deuxième génération d'hybrides (F_2), se manifesteront l'intensité moyenne (la plus probable) et les intensités supérieures ou inférieures à la moyenne (et qui sont d'autant moins probables qu'elles dévient davantage de la moyenne) (Tableau I).

1° La variance imputable à la différence de constitution génétique des individus d'une population est donc fonction du nombre des degrés de liberté représentés chacun par une paire de facteurs héréditaires transmis selon les lois de Mendel.

La constitution génotypique constitue une somme de possibilités de manifestations de caractères mesurables, mais, dans une population constituée d'individus ayant le même génotype, c'est-à-dire la même somme de possibilités, ces possibilités s'extériorisent, par les caractères mesurables, avec des intensités qui se distribuent normalement autour de la valeur moyenne la plus probable; les génétistes appellent « fluctuations » cette variation due aux erreurs « d'échantillonnage ».

2° La variance imputable aux fluctuations représente la « variance résiduelle » qui est le minimum vers lequel tend la variance d'une population originellement hétérogène progressivement purifiée par autofécondation et sélection.

Si les individus d'un même génotype, qui constituent une population pure (« homozygote »), constituent une « classe » quant au caractère étudié, la variance résiduelle correspond à la corrélation intraclasse par opposition à la variance entre génotypes différents, qui constitue la variance interclasse.

L'isolement des « populations pures » ou « lignées pures » d'un mélange de

populations correspond donc à une « analyse de la variance » puisqu'elle permet de discriminer entre la variance intraclasse et la variance interclasse.

Une population de plantes semblables en apparence (comme celles d'un champ de maïs où même d'un champ de tabacs) paraît se distribuer à peu près normalement autour d'une moyenne M avec une variance σ^2 .

Une sélection permettra de discriminer dans cette population une collection de n populations pures, caractérisées chacune par une moyenne $\bar{x}_1, \dots, \bar{x}_n$, ces diverses moyennes se distribuant à peu près normalement autour de la moyenne générale M ; mais quelle sera la distribution des individus de chacune des n populations pures autour de sa moyenne respective $\bar{x}_1 \dots \bar{x}_n$.

La sélection porte nécessairement sur un caractère ou un ensemble de caractères accessibles aux moyens d'observation, de mesures ou d'analyses dont dispose le sélectionneur.

Trop souvent, la sélection porte sur un caractère, ou un ensemble de caractères évidents; mais s'il existe une corrélation plus ou moins étroite entre certains caractères (par exemple chez le tabac, nombre de feuilles et hauteur de la plante) par contre, beaucoup de caractères sont indépendants.

La sélection ayant en vue d'isoler d'une population impure (distribuée autour d'une valeur moyenne M d'un certain caractère) une population pure (distribuée autour d'une certaine valeur moyenne \bar{x}_R de ce même caractère), les questions à résoudre sont donc :

1° quelle sera la distribution de la population pure autour de cette moyenne \bar{x}_R quant au caractère considéré.

2° quelle sera la distribution de la population quant à la manifestation des autres caractères liés au premier par une certaine corrélation, et des caractères indépendants du premier.

Plus strictement, pour l'une ou l'autre de ces 2 questions, la sélection a-t-elle pour effet simplement de déplacer la moyenne et de réduire la variance, la distribution restant symétrique, ou la sélection peut-elle écarter la classe du mode de celle de la moyenne?

La sélection portant sur un certain caractère peut produire une population pure pour ce caractère, mais pour ce caractère seulement : d'après les théories de East et de Hayes, une population pure doit être distribuée normalement pour le caractère pour lequel elle est pure.

Il ne s'ensuit nullement que cette population soit pure pour les autres caractères, et l'étude de ces autres caractères peut conduire et conduit généralement à trouver une distribution non normale. Il est facile d'en donner statistiquement la raison.

Sélectionner en vue d'un certain caractère, c'est supprimer tous les degrés de liberté qui existaient primitivement entre les diverses classes (génotypes différents) qui constituaient la population initiale; c'est supprimer la variance interclasse pour ne laisser subsister que la variance résiduelle correspondant à des fluctuations normalement distribuées.

Mais par suite des corrélations qui existent entre le caractère pour lequel on a sélectionné et les autres caractères, la sélection a joué pour ces autres caractères comme des tirages au sort sans remise des boules dans l'urne; elle doit donc aboutir à une distribution non normale pour ces caractères.

Soient n tentatives, (« épreuves » ou « coups ») réalisant n « possibilités » d'un événement dont la probabilité de manifestation (ou chance de succès est p à chaque coup.

(La probabilité de non manifestation, ou chance d'échec étant $q = 1 - p$.)

La probabilité que l'évènement se manifestera à chaque coup est $p \times p \times \dots$
 $p = p^n$.

La probabilité que l'évènement se manifestera $(n-1)$ fois et manquera une fois de se manifester est $n(p \times p \dots \times p \times q) = np^{n-1}q \dots$

La probabilité que l'évènement se manifestera $(n - k)$ fois et échouera k fois sera $\frac{n(n-1)(n-2) \dots [n-(k-1)]}{1, 2, \dots k} p^{n-k} q^k$.

De façon générale, sur n possibilités d'un événement dont la probabilité de manifestation pour chaque possibilité est p , la probabilité respective de n , $(n-1) \dots (n - k)$ manifestations est donnée par le terme correspondant du développement du binôme $(p + q)^n$.

Appelons « fréquence » chaque valeur correspondant à chacune de ces probabilités. « Distribuons » ces fréquences comme autant d'ordonnées pour représenter graphiquement le nombre de fois qu'un événement se manifeste.

La progression géométrique qui décrit le nombre des « suites » dans une expérience directe, réalisée par le jeu de pile ou face, par le jeu de dés ou par tirage au sort des boules d'une urne, réalise une courbe de fréquence dont l'équation est de la forme $y = N a^x$.

La variable x peut être, comme la variable y , mesurée en fonction de n , de p et de q , et le plus aisément en fonction de $n \cdot p \cdot q$.

C'est ce produit que Fisher désigne sous le nom de « variance », que nous représenterons par le symbole σ^2 . Sa racine carrée $\sigma = \sqrt{n \cdot p \cdot q}$ représente la déviation standard, c'est-à-dire la valeur minima que peut prendre la somme des carrés des déviations de chaque valeur de x par rapport à la valeur moyenne \bar{x} . Pour une même valeur de n , la variance $\sigma^2 = n \cdot p \cdot q$ prend sa plus grande valeur pour $p = q = \frac{1}{2}$.

Si nous définissons la « liberté » par la faculté pour les variants x de s'écarter de la valeur moyenne \bar{x} , la liberté se mesure le plus logiquement par la variance.

Pour une même valeur de p , la variance est une fonction linéaire de n , et $(n-1)$ mesure le maximum de degrés de liberté qui peuvent exister entre n variants (graphique 6).

Les colonnes 1 à 6 du tableau I indiquent d'après East :

- 1° le nombre (n) de paires de caractères par lesquelles diffèrent deux parents;
- 2° les proportions numériques des individus dans les diverses classes (ou phénotypes) qui peuvent être distinguées parmi les hybrides de deuxième génération (F_2).

Ces proportions sont données par les chiffres du triangle de Pascal, et correspondent au développement du binôme $(p + q)^n$.

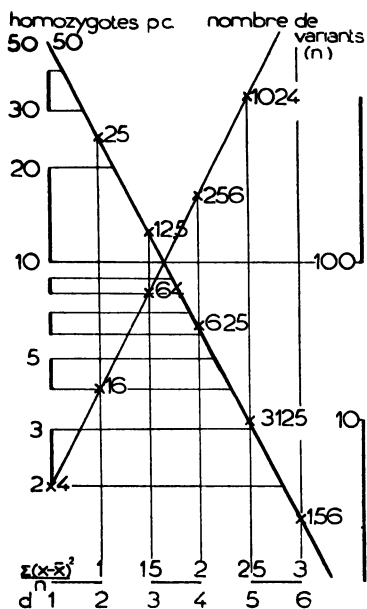
Les colonnes 3, 4, 5 du tableau I indiquent : (3) les nombres (N) des individus (c'est-à-dire la somme des nombres de la ligne correspondante du triangle de Pascal); (4) les nombres d'individus « homozygotes » c'est-à-dire de ceux de

l'autofécondation donne une « population pure » et (5) le pourcentage des homozygotes.

Enfin, la colonne (6) indique les valeurs de la variance $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$ qui croissent en même temps que n .

TABLEAU I

(1)	(2)	(3)	(4)	(5)	(6)
1	1 2 1 P P P	4	2	50	0,5
2	1 4 6 4 1 P P P P P	16	4	25	1
3	1 6 15 20 15 6 1 P P P P P P P	64	8	12,5	1,5
4	1 8 28 56 70 56 28 8 1 P P P P P P P P P	256	16	6,25	2
5	1 10 45 120 210 252 210 120 45 10 1 P P P P P P P P P P P	1.024	32	3,125	2,5
6	1 12 66 220 495 792 924 792 495 220 66 12 1 P P P P P P P P P P P P P	4.096	64	1.562	3



Graphique 6 : En abscisses, sur échelles arithmétiques :

ligne supérieure, valeurs de $\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$;

ligne inférieure, nombre (d) de degrés de liberté ;
En ordonnées, sur échelles log. : à droite, nombres (n) de variants; à gauche, pour cent d'homozygotes.

CONCLUSION

Selon la physique classique, certaines opérations sont nécessaires pour prendre une particule occupant une certaine place à l'instant initial et pour lui communiquer une certaine vitesse; nous pourrions la saisir entre des pinces, la conduire à la place en question puis la pousser dans la direction voulue.

La mécanique quantique nous apprend que si une telle opération se fait, avec une particule, un grand nombre de fois, le même résultat n'est pas invariablement atteint même si l'opération se répète identiquement... mais les résultats ne sont pas quelconques. Répétons la même expérience un premier

million de fois, et enregistrons la fréquence avec laquelle se manifeste chacun des divers résultats possibles; au cours du deuxième million, les expériences reproduiront chacun de ces divers résultats avec sensiblement la même fréquence (Schrodinger 1935).

La variance correspondant à la distribution de ces fréquences, (ou variance intraclasse), est la limite vers laquelle tend la variance dans une population de n variants lorsque n augmente indéfiniment, chacune des n possibilités étant aussi identiques que le permettent les conditions de l'expérience.

C'est une variance résiduelle ou « variance due aux erreurs »; elle n'est imputable à aucune différence, mesurable ou accessible à l'observation, entre les causes présumées des résultats observés.

La variance reconnaît deux origines :

1^o la variance « interclasse » tirant origine de différences que nos sens permettent de reconnaître, et qui nous permettent de répartir les variants par classes, que nous puissions caractériser, chacune, par une moyenne significativement différente de la moyenne qui caractérise chacune des autres classes.

2^o La variance « intraclasse » définie plus haut.

La variance est essentiellement une valeur d'intégration dont K. Pearson et Fisher nous ont appris à faire l'analyse.

L'analyse de la variance permet de discriminer ce qui, dans la variance des résultats expérimentaux autour d'une moyenne présumée, reconnaît, comme origine : des différences qui dépendent de l'expérimentateur, ou des différences qui échappent à son contrôle; l'importance de ces dernières diminue avec les progrès de la technique, ce qui permet à Compton d'écrire :

« Avant l'aube de la conscience sociale, Dieu tenait dans ses mains l'entière responsabilité de l'évolution de la vie sur cette planète. Graduellement cette responsabilité est transportée sur nos épaules. A mesure que la science progresse, il semble inévitable que ce transfert d'autorité ne s'accélère... » (Compton, *The freedom of man*, p. 115).

Les progrès de la science ont donc pour but et pour résultat de diminuer le nombre des degrés de liberté qui restent indépendants de l'expérimentateur.

J. DUFRÉNOY.

DISCUSSION

M. le Président félicite M. DUFRÉNOY pour sa très intéressante communication dont la matière n'avait encore fait l'objet d'aucune étude à la Société et est heureux de constater que la statistique intervient de plus en plus dans tous les domaines.

M. LHOSTE voudrait tenter une explication du phénomène de dispersion asymétrique autour d'une moyenne des produits obtenus par le croisement de tabacs à 19 et à 24 feuilles : si l'on admet que l'examen porte sur un nombre de cas suffisamment grand, on peut penser que la loi qui apparaît devrait être une simple variante de la loi de Gauss, et il semble y avoir contradiction avec

la réalité. Toutefois, cette contradiction n'est qu'apparente, car la loi de Gauss vaut pour l'intervalle $-\infty$ à $+\infty$, alors que dans les conditions de l'expérience l'intervalle est seulement 0 à $+\infty$. Pour ramener la courbe expérimentale à la courbe de Gauss, il suffira vraisemblablement de prendre comme variable, non pas x , mais $\log. x$ qui varie de $-\infty$ à $+\infty$ quand x varie de 0 à $+\infty$.

En outre, M. LHOSTE demande si les travaux signalés par M. Dufrénoy ont reçu une application pratique en France.

M. DUFRÉNOY répond à M. Lhoste qu'il y a eu une réalisation : l'Institut expérimental des tabacs de Bergerac dirigé par M. Gisquet. Cependant, il est possible que le décès de M. Vezian, sous les ordres duquel s'est élaboré l'application des méthodes statistiques aux tabacs, en ait modifié l'évolution. Néanmoins grâce à l'existence de cet Institut, qui a officiellement pour but l'étude de l'application des méthodes d'amélioration de la qualité du tabac, il serait certainement possible d'obtenir des chiffres intéressants. A côté de cette création, il n'existe en France que des projets.

Jusqu'à présent, c'est surtout dans les pays anglo-saxons que ces méthodes ont reçu la plus grande expansion. Aux États-Unis, notamment, toute l'économie est édifiée sur une utilisation indéniable de ces méthodes.

M. le D^r ICHOK souligne que parmi les principes d'ordre moral qui découlent de la communication de M. Dufrénoy, il y en a un qui l'a particulièrement frappé : c'est celui de la responsabilité dans la modification d'une distribution naturelle.

Deux grands problèmes se posent à son esprit :

1^o Il mentionne rapidement les expériences qui se font actuellement en Allemagne en vue de la pureté de la race, expériences qui peuvent mener très loin au double point de vue scientifique et politique ;

2^o Mais il est une autre question que M. le D^r ICHOK considère comme plus importante encore : c'est celle de l'assistance aux malades, aux infirmes, à toute une série de personnes qui grèvent les budgets. Ne serait-il pas préférable de prévenir cet état de choses que de chercher à y remédier tardivement et partiellement ? Trop souvent, il y a des familles entières atteintes de maladies, comme la syphilis ou la tuberculose ; n'y aurait-il pas lieu dans ces cas d'essayer de limiter la natalité afin non seulement de diminuer les charges de la société, mais surtout d'éviter des souffrances à ceux qui n'ont pas demandé de venir au monde.

M. DUFRÉNOY avait bien pensé au problème que vient de poser M. le D^r ICHOK. Il estime que les caractéristiques de l'homme dépendent d'un nombre très grand de facteurs et qu'en l'état actuel de nos connaissances il ne paraît pas possible de prévoir ce que donnera la descendance d'un géniteur humain. D'un autre côté, quand on se trouve en face de facteurs pathologiques, de parasites dont la présence connue entraîne des conséquences également connues, et connaissant, en outre, les déterminants de l'avenir, on semble fondé à intervenir.

M. BARRIOL désirerait connaître ce qui se produit dans les croisements des tabacs à 12 et des tabacs à 24 chromosomes.

M. DUFRÉNOY, avant de répondre directement à la question de M. Barriol, examine en détail ce qui se passe lorsque se trouvent en présence des cellules mâles et des cellules femelles à 12 chromosomes.

Un tabac à 12 paires de chromosomes (GG) tel que *Nicotiana glutinosa*, forme des cellules sexuelles mâles contenant 12 chromosomes : $G_1, G_2 \dots G_{12}$, des cellules sexuelles femelles contenant 12 chromosomes $G_1, G_2 \dots G_{12}$, homologues chacun à chacun des 12 chromosomes mâles.

La fécondation de la cellule femelle par la cellule mâle met en présence de chaque chromosome femelle $G_1 \dots G_{12}$ le chromosome mâle homologue G_1 à G_{12} et permet, dans le noyau de l'œuf puis dans le noyau de chaque cellule de l'embryon et de la plante qui en procède, la formation des 12 couples chromosomiques $GG_1, GG_2 \dots GG_{12}$.

Si, au lieu d'être fécondée par une cellule mâle de même espèce, une cellule femelle de *N. glutinosa* (contenant 12 chromosomes G) est fécondée par une cellule mâle d'autre espèce à 12 chromosomes, il se formera ou non des couples, suivant que cette cellule mâle contient ou non des chromosomes homologues chacun à chacun des 12 chromosomes G.

Suivant que chaque accouplement représente un événement probable indépendant ou un événement rare indépendant, les nombres d'appariages observés se distribuent selon une courbe binomiale ou suivant celle des séries de Poisson.

La forme de la distribution et le mode définissent donc le degré d'affinité entre deux espèces de *Nicotiana*.

Par exemple, pour le croisement *Glutinosa* × *tomentosa*, 150 cas observés par Goodspeed se distribuent autour de la moyenne $\bar{x} = 5,17$, avec une erreur standard $\sigma = 1,7$; elle se calcule par le développement du binôme $150(0,43 + 0,57)^{12}$.

Pour le croisement *Glutinosa* × *sylvestris*, 150 cas observés par Goodspeed se distribuent autour d'une moyenne $\bar{x} = 1,56$ avec une erreur standard $\sigma = 0,92$, et se calcule par le développement du binôme $150(0,13 + 0,87)^{12}$ ou par l'équation :

$$x = 150 \frac{1}{0,92 \times 2,5} e^{-\frac{(x - 1,56)^2}{1,84}}$$

La fécondation d'une cellule sexuelle à 12 chromosomes $G_1 \dots G_{12}$ par une cellule sexuelle à 24 chromosomes de *Nicotiana tabacum*, $T_1 \dots T_{24}$ ne peut permettre que la formation de 12 couples $G_1 T_1 \dots G_{12} T_{12}$ en admettant que chaque chromosome $T_1 \dots T_{12}$ soit homologue d'un chromosome $G_1 \dots G_{12}$ et puisse s'apparier avec lui.

Dans tous les cas, 12 chromosomes $T_{13} \dots T_{24}$ demeurent sans partenaires, demeurent univalents.

La présence de ces univalents peut introduire des perturbations telles que les chromosomes $G_1 T_1; \dots G_{12} T_{12}$ soient empêchés de s'apparier.

Au lieu de 12 paires, nombre maximum possible, il peut ne s'en former que 12-1, 12-2, 12-12.

En dénombrant, au moment où les hybrides de *N. glutinosa* et de *N. tabacum* forment leurs cellules sexuelles, les fréquences avec lesquelles se forment 12, 11, 10. . . 1 paires de chromosomes GT, Clausen et Goodspeed ont observé, pour 300 cas, une distribution qui est superposable à la distribution calculée pour la somme des distributions des 150 cas observés pour les hybrides de *Glutinosa* × *tomentosa* et des 150 cas observés pour les hybrides *glutinosa* × *sylvestris*.

Si nous considérons les fréquences d'appariage comme les résultats de tirage au sort des boules d'une urne, nous pouvons statistiquement considérer que l'urne « *glutinosa* × *tabacum* » contient le mélange des urnes « *glutinosa* × *tomentosa* » et « *glutinosa* × *sylvestris* ». C'est génétiquement considérer que *N. tabacum* contient la somme des possibilités héréditaires, c'est-à-dire la somme des chromosomes de *N. tomentosa* et de *N. sylvestris*; nous donnons ainsi une démonstration statistique de la théorie de Clausen et Goodspeed : *N. tabacum* représente le produit du doublement du nombre des chromosomes chez un hybride entre *N. tomentosa* et *N. sylvestris*. Cet hybride n'aurait pas été viable sans ce doublement, car il ne contenait pas de chromosomes capables de s'apparier chacun à chacun. Le doublement permet à chaque chromosome d'origine *sylvestris* S₁ à S₁₂ de trouver son homologue S₁ à S₁₂ à chaque chromosome d'origine *tomentosa* T₁ à T₁₂ de trouver son homologue T₁ à T₁₂. *N. tabacum* mérite alors le nom d'hybride « amphidiploïde ».

Un hybride amphidiploïde est un hybride parfaitement équilibré, fertile, représentant l'origine d'une espèce nouvelle, par doublement du nombre des chromosomes d'un hybride déséquilibré, donc stérile.

Nul n'a vu le *N. tabacum* prendre origine par doublement du nombre des chromosomes chez son ancêtre présumé, l'hybride *sylvestris* × *tomentosa*; mais Clausen a obtenu dans ses cultures un hybride amphidiploïde, qu'il a appelé *N. digluta* et qui représente une nouvelle espèce.

Revenant alors à l'accouplement de cellules à 24 chromosomes à des cellules à 12 chromosomes, M. DUFRÉNOY fait connaître qu'il y a, dans ce cas, 12 chromosomes qui trouvent des partenaires, mais qu'il en reste 12 non satisfaits qui introduisent des troubles très graves empêchant l'évolution de la plante. Cependant, dans certains cas très rares on obtient une nouvelle espèce synthétique parfaitement équilibrée et c'est le phénomène auquel s'attache de plus en plus l'activité des génétistes.

M. DUFRÉNOY souligne enfin que c'est grâce aux encouragements de M. Dar-mois qu'il a été amené à faire l'exposé que nous venons d'entendre.

M. le Dr SOREL voudrait mettre en garde ses collègues contre la tendance à généraliser les conclusions données par l'auteur à la fin de sa savante communication, en passant du règne végétal, le tabac, au règne animal dont l'homme fait partie. Pour les fleurs comme pour les chevaux de course, on cherche à conserver ou à développer une qualité. Chez l'homme les aptitudes de l'intelligence sont multiples, on peut voir des hommes fort bien bâtis et bêtes et des hommes fort intelligents au corps mal construit. A titre d'except-

tion on connaît des familles où l'art musical semble héréditaire comme chez les Bach.

Le danger réside dans la facile inclination de certains hommes d'État à transformer en obligations des généralisations hâtives; nous avons des exemples chez certains dictateurs qui prétendent appliquer des lois édictées par eux pour se conformer au résultat de la science (!) comme la stérilisation.

Hélas! nous connaissons des exemples d'hommes de génie comme Van Beethoven issu d'un père ivrogne invétéré. Quel malheur pour l'humanité si alors avait régné M. Hitler qui aurait stérilisé ce père.

Pour la syphilis, laissons en liberté les médecins chercher le remède à la syphilis héréditaire et gardons-nous de demander au Parlement de voter une loi supprimant tous les syphilitiques ou au moins leur descendance. Les médecins ayant une longue pratique connaissent des enfants de syphilitiques qui sont normaux physiquement et mentalement et même quelquefois sont des sujets remarquables. Ne nous hâtons donc pas d'édicter de nouveaux dogmes hors desquels il n'y aurait pas de salut.

Il est fréquent d'entendre parler de dégénérescence. Mais si, depuis l'âge quaternaire, depuis des milliers d'années l'homme dégénérât, aujourd'hui il n'y aurait plus personne pour entendre la communication de M. Dufrénoy, o horreur! qui serait lui-même dégénéré. Quand on voit le magnifique essor de l'esprit humain, les splendides découvertes scientifiques et les merveilleuses productions de l'art et même les belles performances sportives, on est bien obligé d'admettre ou que la dégénérescence a du bon ou que la nature arrange fort bien les choses sans l'intervention des législateurs.

Attendons pour agir le temps nécessaire pour trouver la vérité des conclusions de l'expérience et hâtons-nous lentement.

M. DARMOIS, après avoir vivement félicité le conférencier, signale que l'application de la statistique à la biologie devient de plus en plus fréquente en France et mentionne que récemment il eut à examiner une thèse sur un sujet de cette nature.

BIBLIOGRAPHIE

- BABCOCK (E. B.), Genetic evolution processes. (*Proceed. Nat. Acad. Sciences*, 20 : 510-515, 1934.)
CLARCK (A. J.), The mode of action of drugs on cells. (London; Edward Arnold et C^{ie}, 1933.)
BAUR, Grundlagen der Pflanzenzüchtung; Borntraeger, p. 20:
COMPTON (A. H.), The freedom of man. (New-Haven, Yale University Press, 1935.)
— Statistique et applications. (Doin et C^{ie}, Paris, 1928.)
DARMOIS (G.), Statistique mathématique. (Armand Colin, Paris, 1934.)
DUFRENOY (J.), Méthodes statistiques appliquées à la pathologie végétale. (*Ann. Épi-phyties et Phytogénétique*. I : 147-256, 1934-1935.)
DUFRENOY et VEZIAN, Représentation des phénomènes biochimiques et épidémiologiques par la courbe IV de Pearson (*Rev. Microb. appl* (1937).
EAST (E. M.), The genetics of the genus Nicotiana.
EGGENBERGER (F.) und POLYA (F.), Über die Statistik verketteter Vorgänge. (*Zeist. angew. mathematik und mechanik*, 3 : 279-289, 1923.)
— Die Wahrscheinlichkeitsansteckung. (*Eidgenössischen Technischen Hochschule in Zurich*, 1924.)
— Sur l'interprétation de certaines courbes de fréquence. (*C. R. Acad. Sciences*, 187 : 870, 1928).

- FISHER (R. A.), Has Mendel's work been rediscovered? (*Annals of Science*, 1 : 115-137, avril 1936.)
— The genetical theory of natural selection. (Oxford, Clarendon press, 1930.)
— Uncertain Inference (*Proceed. Amer. Acad. Arts and Science*, 71 : 245-257, 1936).
— The relation between variability and abundance... (*Proceed. Roy. Soc. London*, 122 : 1-26, 1937).
- GREENWOOD (Major), Epidemics and crowd diseases. (London; Williams and Norgate Ltd, 1935.)
- HALDANE (J. B. S.), The amount of heterozygosis to be expected in an approximately pure line. (*Journ. Genetics*, 32 : 375, jul. 1936.)
- HALDANE (J. B. S.), The causes of evolution. (London; Longmans, Green and Co, 1935.)
- HARTLEY (C.) et GRAVATT (A. R.), Some effects of plant diseases on variability of yields (*Phytopathology*, 27 : 159-172, 1937).
- HAYES (H. K.), Correlation and inheritance in *Nicotiana tabacum*. (*Conn. Agric. Expt. stat. bull.* 171, may 1912.)
- JONES, Genetics. (New-York, 1925, p. 213.)
- LOVE (H. H.), Application of statistical methods to agricultural research. (Shanghai, Commercial Press, 1936.)
- MORGAN (Thomas-Hunt), The scientific basis of evolution. (New-York, Norton & Co, 2^e éd., 1935.)
- PEARL (R.), Introduction to medical biometry and statistics. (Philadelphia & London; Saunders Co, 1930.)
- PITTEY (R.), Quelques données sur l'expérimentation cotonnière. (Publications de l'Institut national pour l'étude agronomique du Congo belge, série technique n° 8, 1936.)
- POLYA (G.), Eine Wahrscheinlichkeitsaufgabe in der Pflanzensoziologie. (*Vierteljahrsschrift der naturforsch. Gesellsch. Zurich*, 75 : 211-219, 1930.)
- TORRIE (J. H.), Inheritance studies of several qualitative and quantitative characters in spring wheat crosses between varieties relatively susceptible and resistant to drought. (*Canad. Journ. Res.* : 14 : 368-385, 1936.)
- VEZIAN et DUFRÉNOY, Problèmes statistiques relatifs à la désinfection des semences (*Rev. Microb. appl.*, 2 : 326-333, 1936).
- WÆLKENS (W.), La purification du « Triumph big boll » dans l'Uele. (Publications de l'Institut national pour l'étude agronomique du Congo belge, série technique n° 9, 1936.)
— Travaux de sélection du Coton. (Publications de l'Institut national pour l'étude agronomique du Congo belge, série technique n° 5, 1936.)
- WINTER (F. L.), The mean and variability as affected by continuous selection for composition in corn. (*Jour. Agric. Res.*, 39 : 451-476, 1929).
-