

M. L. Dreyer

SERIE A,
N° 1861
N° D'ORDRE
2728

THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES
DE L'UNIVERSITÉ DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS-SCIENCES MATHÉMATIQUES

PAR

GUSTAVE MALÉCOT

*Ancien élève de l'École Normale Supérieure
Agrégé de l'Université*

1^{re} THÈSE. — THÉORIE MATHÉMATIQUE DE L'HÉRÉDITÉ MENDÉLIENNE
GÉNÉRALISÉE.

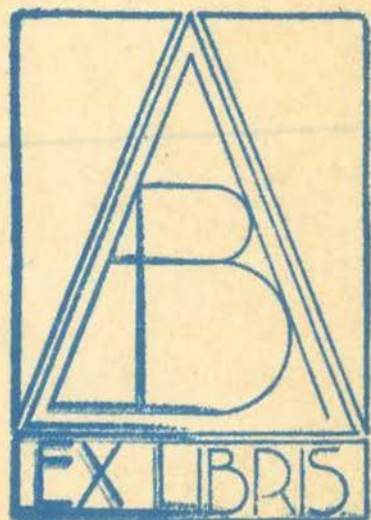
2^e THÈSE. — PROPOSITIONS DONNÉES PAR LA FACULTÉ.

Soutenues le

1939 devant la Commission d'examen.

MM. E. BOREL *Président.*
G. DARMOIS }
G. VALIRON } *Examineurs.*

PARIS
IMPRIMERIE J. GUILHOT
52, Boul. Malesherbes



SÉRIE A,
N° 1861
N° D'ORDRE
2728

THÈSES

PRÉSENTÉES

A LA FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

POUR OBTENIR

LE GRADE DE DOCTEUR ÈS-SCIENCES MATHÉMATIQUES

PAR

GUSTAVE MALÉCOT

*Ancien élève de l'École Normale Supérieure
Agrégé de l'Université*

1^{re} THÈSE. — THÉORIE MATHÉMATIQUE DE L'HÉRÉDITÉ MENDÉLIENNE
GÉNÉRALISÉE.

2^e THÈSE. — PROPOSITIONS DONNÉES PAR LA FACULTÉ.

Soutenues le

1939 devant la Commission d'examen.

MM. E. BOREL *Président.*
G. DARMOIS }
G. VALIRON } *Examineurs.*

PARIS

IMPRIMERIE J. GUILHOT
52, Boul. Malesherbes

—
1939

MM.

Doyen honoraire. . . . M. MOLLIARD.
 Doyen C. MAURAIN, Professeur, Physique du Globe.

Professeurs honoraires	}	H. LEBESGUE.	BLAISE.	G. BERTRAND.
		Émile PICARD.	DANGEARD.	ABRAHAM.
		Léon BRILLOUIN.	LESPIEAU.	Ch. FABRY.
		GUILLET.	MARCHIS.	LÉON BERTRAND.
		PÉCHARD.	VESSIOT.	WINTREBERT.
		FREUNDLER.	PORTIER.	DUBOSQ.
AUGER.	MOLLARD.	BOHN.		
		LAPICQUE.	RABAUD.	

PROFESSEURS

M. CAULLERY T	Zoologie (Évolution des êtres organisés).	H. BÉGHIN T	Mécanique physique et expérimentale.
Émile BOREL T	Calcul des probabilités et Physique mathématique.	FOCH	Mécanique expérimentale des fluides.
Jean PERRIN T	Chimie physique.	PAUTHENIER	Physique (P. C. B.).
E. CARTAN T	Géométrie supérieure.	De BROGLIE T	Théories physiques.
A. COTTON T	Recherches physiques.	CHRÉTIEN	Optique appliquée.
J. DRACH T	Analyse supérieure et Algèbre supérieure.	P. JOB	Chimie générale.
Charles PÉREZ T	Zoologie.	LABROUSTE	Physique du Globe.
M. GUICHARD T	Analyse et mesures chimiques.	PRENANT T	Anatomie et Histologie comparées.
Pau MONTEL T	Théorie des fonctions et Théorie des transformations.	VILLEY	Mécanique physique et expérimentale.
L. BLARINGHEM . . . T	Botanique.	COMBES T	Physiologie végétale.
G. JULIA T	Mécanique analytique et Mécanique céleste.	GARNIER T	Mathématiques générales.
C. MAUGUIN T	Minéralogie.	PÈRES	Mécanique théorique des fluides.
A. MICHEL-LÉVY . . . T	Pétrographie.	HACKSPILL	Chimie (P. C. B.).
H. BÉNARD T	Mécanique expérimentale des fluides.	LAUGIER T	Physiologie générale.
A. DENJOY T	Application de l'analyse à la Géométrie.	TOUSSAINT	Technique Aéronautique.
L. LUTAUD T	Géographie physique et géologie dynamique.	M. CURIE	Physique (P. C. B.).
Eugène BLOCH T	Physique.	G. RIBAUD T	Hautes températures.
G. BRUHAT T	Physique théorique et physique céleste.	CHAZY T	Mécanique rationnelle.
E. DARMOIS T	Enseignement de Physique.	GAULT	Chimie (P. C. B.).
A. DEBIERNE T	Physique Générale et Radioactivité.	CROZE	Recherches physiques.
A. DUFOUR T	Physique (P. C. B.).	DUPONT T	Théories chimiques.
L. DUNOYER	Optique appliquée.	LANQUINE T	Géologie structurale et Géologie appliquée.
A. GUILLIERMOND . . T	Botanique.	VALIRON	Mathématiques générales.
M. JAVILLIER T	Chimie biologique.	BARRABÉ	Géologie structurale et Géologie appliquée.
ROBERT-LÉVY T	Physiologie comparée.	MILLOT	Biologie animale (P. C. B.).
F. PICARD	Zoologie (Évolution des êtres organisés).	F. PERRIN	Théories physiques.
Henri VILLAT T	Mécanique des fluides et applications.	VAVON	Chimie organique.
Ch. JACOB T	Géologie.	G. DARMOIS	Calcul des probabilités et Physique-Mathématique.
P. PASCAL T	Chimie générale.	CHATTON T	Biologie maritime.
M. FRÉCHET T	Calcul différentiel et Calcul intégral.	AUBEL	Chimie biologique.
E. ESCLANGON T	Astronomie.	Jacques BOURCART	Géographie physique et Géologie dynamique.
M ^{me} RAMART-LUCAS . . T	Chimie organique.	M ^{me} JOLIOT-CURIE	Physique générale et Radioactivité.
		PLANTEFOL	Biologie végétale (P. C. B.).
		CABANNES	Enseignement de Physique.
		GRASSÉ	Biologie animale (P. C. B.).
		PRÉVOST	Chimie (P. C. B.).
		BOULIGAND	Mathématiques.

Secrétaire A. PACAUD.
 Secrétaire honoraire D. TOMBECK

A MA MÈRE
A MES BEAUX-PARENTS
A MA FEMME

A MONSIEUR GEORGES DARMOIS

PROFESSEUR A LA SORBONNE

PREMIÈRE THÈSE

THÉORIE MATHÉMATIQUE DE L'HÉRÉDITÉ MENDÉLIENNE GÉNÉRALISÉE

AVANT-PROPOS

C'est à la fin du XIX^e siècle qu'a été entreprise, par le savant anglais F. Galton et ses disciples, en particulier Karl Pearson, l'étude numérique de l'hérédité. C'est cette étude qui les a conduits à préciser les notions fondamentales de ce qui est aujourd'hui « la Statistique Mathématique », en particulier la notion de « corrélation » entre des caractères mesurables (taille par exemple) des 2 individus apparentés.

Ces corrélations ont été l'objet de nombreuses déterminations empiriques, publiés dans la revue *Biométrie*. J'ai indiqué (1) les résultats essentiels obtenus par Galton et Pearson, et rappelé comment, lors de la redécouverte, vers 1900, des lois de Mendel, la question s'est posée de savoir s'il n'y a pas incompatibilité entre ces lois, qui supposent une hérédité « particulière » et font dépendre les enfants des parents seulement, et les résultats de l'école biométrique, qui traduisent une hérédité mélangée et continue, une dépendance en moyenne des enfants à l'égard des parents et aussi des divers ancêtres (2). Cette divergence apparente a été

(1) Annales de l'Université de Lyon, sciences, 1938.

(2) Voir G. Darrois, annales de l'Institut H. Poincaré, 3, 1932-33, p. 213.

levée par Pearson, Snow (voir bibliographie p. 51) et surtout R.-A. Fisher dans son mémoire fondamental des Transactions of the R. S. of Edinburgh (52, 1918, p. 399) (mémoire que nous désignerons toujours, par la suite, par la notation abrégée Ed.). Le but du travail que nous présentons ici est de développer systématiquement les hypothèses et les méthodes qui permettent d'étendre les lois de Mendel à l'explication des modes d'hérédité « mélangée ».

Dans le chapitre I, nous précisons d'abord les notions de base : facteurs mendéliens cumulatifs, aléatoires qui traduisent l'action de ces facteurs, liaison de ces aléatoires dans l'hypothèse usuelle du « random mating », et dans une hypothèse d'« homogamie » pour laquelle nous indiquons une méthode commode de calcul symbolique.

Dans les chapitres II, III et IV, nous parvenons à la détermination générale, dans le cas du « random mating », des corrélations entre individus apparentés (englobant ainsi tous les cas particuliers traités par Pearson, Snow et Fisher) grâce aux propriétés de la « causalité par échelons » (chapitre II) et à la forme symbolique que nous utilisons pour ces corrélations (chapitre III, 2^e cas).

Dans le chapitre V nous indiquons comment on peut calculer, dans l'hypothèse d'homogamie, les corrélations les plus simples, et sous quelles approximations sont valables les expressions simplifiées admises par R.-A. Fisher.

Je tiens à exprimer ici ma très vive reconnaissance à M. G. DARMOIS, qui m'a si judicieusement guidé dans mon travail et n'a cessé de m'encourager. J'adresse des remerciements très sincères à M. Emile BOREL, qui a bien voulu présenter à l'Académie des Sciences les principaux résultats (1) ; et à M. VALIRON, qui a bien voulu se joindre à MM. Borel et Darmois pour constituer le jury. Je remercie aussi MM. TEISSIER et L'HÉRITIER pour les nombreux entretiens qu'ils m'ont accordés. En même temps, je tiens à témoigner toute ma gratitude à M. H. EYRAUD, dont les conseils m'ont été précieux.

(1) C. R. Ac. Sc. Paris : 204, p. 619 ; 206, p. 153 et 404 ; 208, 6 février et 20 février 1939.

CHAPITRE I

LES VARIABLES ALÉATOIRES QUI TRADUISENT L'ACTION DES FACTEURS MENDÉLIENS

Hypothèse des facteurs mendéliens cumulatifs

Considérons une population dont les individus diffèrent par un certain nombre de facteurs mendéliens. Les caractéristiques de chaque individu dépendent de sa constitution héréditaire et des conditions de son développement. Nous supposons que la constitution héréditaire d'un individu peut être caractérisée uniquement par la donnée de l'ensemble des gènes qu'il porte, par sa « formule héréditaire » : si $Aa, A'a'...$ sont les différents couples de gènes alléomorphes qui interviennent, les formules possibles seront :

$AAA'A'...$, $AaA'A'$, $aaA'A'$, $AAA'a'...$, etc.

Soit alors un caractère mesurable y d'un individu. Nous supposons que c'est la somme de 2 termes : $y=x+e$, le premier x représente l'action sur ce caractère des facteurs héréditaires, c'est-à-dire qu'il est déterminé par la formule héréditaire de l'individu ; le deuxième terme e représente l'effet des conditions du développement, du milieu, nous supposons par la suite qu'il est indépendant de la constitution héréditaire ; nous verrons qu'il est en général négligeable par rapport à x et dans ce chapitre nous confondrons x et y .

Nous supposons en général (nous étudierons au chap. IV le cas de la non additivité ou « epistacy ») que *les différents couples de gènes qui constituent la formule héréditaire de l'individu ajoutent leurs effets, c'est-à-dire que x est la somme des contributions $H, K, etc.$, des différents couples. H prendra les*

valeurs i, j , ou k , suivant que le premier couple de l'individu est dans l'état AA, Aa , ou ; aa de même K prendra les valeurs i', j', k' , suivant que le deuxième couple est $A'A', A'a'$, ou $a'a'$; etc.

La variable aléatoire qui correspond à un facteur

Si nous portons notre attention sur le premier couple indépendamment des autres, les individus se répartiront en 3 catégories ou « phases » suivant qu'ils portent le couple : AA, Aa, aa ; désignons par $P, 2Q, R$ la fréquence dans la population de chacune de ces 3 catégories :

$$P + 2Q + R = 1$$

Ce sont aussi les probabilités pour un individu pris au hasard dans la population considérée, d'avoir la constitution AA, Aa, aa . (Elles sont relatives à une population déterminée et à un instant déterminé, et elles peuvent naturellement changer d'une génération à une autre.) La contribution H du premier couple au caractère, pour un individu pris au hasard, apparaît donc comme une variable aléatoire qui peut prendre les valeurs i, j, k , avec les probabilités $P, 2Q, R$, de même K est une aléatoire égale à : i', j', k' , avec les probabilités $P', 2Q', R'$. Dans les hypothèses faites, la contribution x des facteurs héréditaires est donc pour un individu pris au hasard, une aléatoire qui résulte de l'addition d'un certain nombre d'aléatoires du 3^e ordre H, K , etc., qui traduisent les effets des différents couples de gènes (que nous appellerons également facteurs).

On pourrait aller plus loin et supposer que l'effet d'un couple de gènes est la somme des effets des 2 gènes qui le constituent. Alors, en désignant par t la contribution du gène A et par s celle de a , la contribution H du facteur correspondant prendrait l'une ou l'autre des 3 valeurs :

$$i = 2t, \quad j = s + t, \quad k = 2s,$$

correspondant respectivement aux états AA, Aa, aa . La contribution j de l'hétérozygote est alors la moyenne des

contributions i et k des 2 homozygotes. C'est là la condition pour que l'aléatoire du 3^e ordre H puisse être considérée comme résultant de l'addition de 2 aléatoires du 2^e ordre (que nous désignerons par \mathcal{H} et \mathcal{H}'). Dans ce cas x serait la somme d'aléatoires du 2^e ordre représentant les contributions des différents gènes. Mais nous verrons qu'en général on ne peut admettre qu'il en soit ainsi ; l'hétérozygote est presque toujours plus voisin d'un des homozygotes que de l'autre et même il arrive souvent que la présence de l'un des gènes efface presque complètement l'effet de l'autre, c'est le cas observé par Mendel, de la *dominance* complète : l'hétérozygote apparaît identique à l'un des homozygotes. La dominance complète de A sur a en ce qui concerne le caractère mesuré x se traduit par $j=i$. S'il y a dominance incomplète de A , j est plus voisin de i que de k .

Liaison stochastique des divers facteurs

Pour connaître le comportement de la somme de ces aléatoires, il ne suffit pas de se donner chacune d'elles, il faut encore connaître leur liaison stochastique ; le cas le plus simple, reconnu par Mendel dans ses expériences sur les pois est celui de la « répartition indépendante des facteurs » ; c'est le cas où la répartition d'un des facteurs reste la même quand on fixe l'état dans lequel se trouvent les autres. H , K sont alors des variables aléatoires indépendantes. Les différentes valeurs possibles pour $H+K$ par exemple ont les probabilités :

$$PP', 2PQ', PR', 2QP', \text{ etc.}$$

Mendel a donné comme exemple pour les pois, les 2 couples de gènes lisse-ridé et jaune-vert. Si on croise entre eux les hétérozygotes doubles qui constituent F_1 , chaque couple de gène apparaît dans F_2 sous ses 3 états avec les fréquences $1/4$, $1/2$, $1/4$; et on vérifie que les fréquences des 9 états de l'ensemble des 2 couples sont bien sensiblement les coefficients du développement de :

$$(u/4 + v/2 + w/4)^2$$

Nous verrons plus loin que si les croisements ont lieu au hasard, on tend toujours asymptotiquement vers l'indépendance stochastique des facteurs de quelque manière que soit choisie la population initiale.

Homogamie

Mais il arrive souvent qu'il n'en est pas ainsi, par suite du fait très général de « l'homogamie » observé par les biométriciens chez l'homme et chez les diverses espèces animales. Si on examine un certain nombre de couples et qu'on mesure chaque fois un même caractère chez les 2 conjoints, on observe que les 2 mesures ne sont pas indépendantes, qu'il y a entre elles une corrélation positive dont le coefficient est de l'ordre de grandeur de 0,2. Les conjoints des individus ayant un caractère de valeur déterminée leur ressemblent en moyenne plus que s'ils étaient choisis au hasard parmi la population. Si par exemple on considère les individus ayant une taille déterminée supérieure à la moyenne, la taille moyenne de leurs conjoints est aussi supérieure à la moyenne, et par suite chaque facteur a tendance à se trouver dans une phase qui apporte à la taille une contribution positive, plutôt que dans une phase qui apporte une contribution négative. Il y a donc, chez tout couple de parents et par suite aussi chez les enfants, une tendance à l'association des phases des différents facteurs qui produisent sur la taille des effets de même sens. L'homogamie se traduit par une dépendance mutuelle de l'ensemble des facteurs. Dans une population en équilibre, la probabilité, que nous désignerons par (11), des états simultanés i, i', i'' pour les aléatoires H, K, L , est différente du produit $PP'P''$.

Probabilités d'association des facteurs

Nous aurons surtout à considérer l'association 2 à 2 de ces facteurs, par exemple l'association entre H et K ; pour cela il pourra être commode de désigner avec R.-A. Fisher les probabilités des divers états possibles pour l'ensemble HK par :

$$(11) = PP'(1 + f_{11}), (12) = 2PQ'(1 + f_{12}), (13) = PR'(1 + f_{13}), \text{ etc.}$$

on introduit ainsi des coefficients d'association f_{lm} (> -1), variables naturellement avec le couple de facteurs considérés et qui sont nuls dans le cas de l'indépendance.

Ces coefficients sont liés aux probabilités totales de H ou K par les relations :

$$(11) + (12) + (13) = (10)$$

(probabilité pour que H soit dans l'état AA),
qui s'écrit aussi :

$$PP'(1+f_{11}) + 2PQ'(1+f_{12}) + PR'(1+f_{13}) = P$$

$$\left. \begin{array}{l} \text{c'est-à-dire : } P'f_{11} + 2Q'f_{12} + R'f_{13} = 0 \\ \text{et de même : } P'f_{21} + 2Q'f_{22} + R'f_{23} = 0 \\ P'f_{31} + 2Q'f_{32} + R'f_{33} = 0 \\ Pf_{11} + 2Qf_{21} + Rf_{31} = 0 \\ Pf_{12} + 2Qf_{22} + Rf_{32} = 0 \\ Pf_{13} + 2Qf_{23} + Rf_{33} = 0 \end{array} \right\} (1)$$

Ces relations permettent de calculer les 9 coefficients à partir par exemple de f_{11} , f_{13} , f_{31} , f_{33} , pris arbitrairement et supposés suffisamment petits en valeur absolue pour que tous les f soient > -1 , ce qui est nécessaire pour que toutes les fréquences soient positives. (Il n'y a donc que 5 relations indépendantes ; d'ailleurs la relation qui les unit est :

$$P(P'f_{11} + \dots) + 2Q(P'f_{21} + \dots) + \dots = P'(Pf_{11} + \dots) + 2Q'(Pf_{12} + \dots)$$

On peut de même introduire des coefficients d'association 3 à 3. On désignera par exemple les probabilités des divers états possibles pour l'ensemble HKL par :

$$(111) = PP'P''(1+f_{111}), \quad (112) = 2PP'Q''(1+f_{112}), \quad \text{etc.}$$

Ces coefficients sont liés aux probabilités totales et aux coefficients d'association 2 à 2 de H et K , H et L , K et L , soient f_{ijo} , f_{ioj} , f_{oij} , par les relations :

$$\begin{aligned} PP'P''(1+f_{111}) + 2PP'Q''(1+f_{112}) + PP'R''(1+f_{113}) \\ = PP'(1+f_{110}) \end{aligned} \quad \text{etc.}$$

$$\begin{array}{l}
 \text{c'est-à-dire :} \\
 \left. \begin{array}{l}
 P''f_{111} + 2Q''f_{112} + R''f_{113} = f_{110} \\
 \text{et 8 analogues.} \\
 P'f_{111} + 2Q'f_{121} + R'f_{131} = f_{101} \\
 \text{et 8 analogues.} \\
 Pf_{111} + 2Qf_{211} + Rf_{311} = f_{011} \\
 \text{et 8 analogues.}
 \end{array} \right\} (1')
 \end{array}$$

Probabilités d'association des gènes

A côté de ces quantités qui traduisent l'association entre facteurs, introduisons les fréquences des diverses associations entre gènes, autrement dit, des divers types de gamètes produits par les individus, soit dans le cas des 3 facteurs :

$$\begin{array}{lll}
 (111) & \text{fréquence de l'association } AA'A'' \\
 (113) & \text{» } & \text{» } AA'a'' \\
 (131) & \text{» } & \text{» } Aa'A'', \text{ etc.}
 \end{array}$$

La quantité (111) par exemple représente aussi la probabilité pour que 3 gènes pris au hasard dans les 3 couples d'un individu quelconque soient les gènes A , A' et A'' ; or ces 3 gènes peuvent se rencontrer chez l'homozygote $AAA'A'A''A''$ qui fournit leur association $AA'A''$ avec probabilité 1; ou chez les hétérozygotes simples tels que $AAA'a'A''A''$ [de fréquence (121)] qui les fournit avec probabilité $1/2$; chez les hétérozygotes doubles tels que $AAA'a'A''a''$ de fréquence (122) qui les fournit avec probabilité $1/4$; etc. La probabilité totale de l'association $AA'A''$ est donc :

$$\begin{aligned}
 (2) \quad (111) &= (111) + \frac{(112) + (121) + (211)}{2} \\
 &+ \frac{(122) + (212) + (221)}{4} + \frac{(222)}{8}
 \end{aligned}$$

On voit que la fréquence de l'association $AA'A''$ n'est autre que celle d'un groupe qui comprendrait avec les homozygotes correspondants la moitié des hétérozygotes simples, le quart des hétérozygotes doubles, le huitième des hétérozygotes triples, etc. Cela tient à ce que la moitié des gamètes produites par un hétérozygote simple, le quart de celles d'un hétérozygote double, sont d'un type déterminé.

Si on représente par (110), (130), etc., les probabilités d'association des gènes des 2 premiers couples, on a :

$$(3) \quad \begin{array}{ll} (110) = (111) + (113) & (130) = (131) + (133) \\ (310) = (311) + (313) & (330) = (331) + (333) \end{array}$$

En ajoutant à l'expression (2) de (111) l'expression analogue de (113), on trouve que :

$$(110) = (110) + (120)/2 + (210)/2 + (220)/4$$

c'est bien l'équivalent de la formule (2) dans le cas de 2 gènes seulement.

De même les probabilités des états A et a pour un gène du premier couple sont :

$$(4) \quad (100) = (110) + (130) \quad (300) = (310) + (330)$$

Pour simplifier nous poserons comme Fisher (100) = p et (300) = q et de même (010) = p' , (030) = q' , etc. Ces quantités sont évidemment liées par la relation :

$$(5) \quad (100) + (300) = 1$$

(cette relation, qui peut s'écrire aussi d'après (3) et (4) :

$$(111) + (113) + (131) + (133) + (311) + (313) + (331) + (333) = 1$$

exprime que la probabilité totale de l'ensemble des combinaisons de gènes est 1 ; en ajoutant les relations telles (2), on voit que cela revient à dire que la probabilité totale de toutes les combinaisons possibles des deux facteurs est 1).

Les deux cas qui seront étudiés

L'introduction de ces différentes probabilités ou fréquences, nous permet de spécifier la constitution génétique d'une population déterminée ; nous allons maintenant étudier leur variation au cours du temps de génération en génération, et voir si elles tendent vers une répartition d'équilibre, et quelles sont alors leurs valeurs. Nous désignerons par F_0 une population d'adultes initiale, dont la constitution génétique est supposée donnée ; par F_1 l'ensemble de leurs enfants, F_2 l'ensemble de leurs petits-enfants, etc. La répartition des gènes dans F_1 se déduira de la répartition connue dans F_0 grâce aux lois de Mendel, mais à condition de faire des hypothèses sur la manière dont se forment les couples de parents dans F_0 . Nous examinerons d'abord l'hypothèse de conjoints s'associant au hasard : *random mating* ; puis nous étudierons l'hypothèse proposée par Fisher pour tenir compte du phénomène de l'*homogamie*.

1^{er} CAS :

RANDOM MATING

Stabilité de la proportion des gènes

L'hypothèse « d'accouplement au hasard », ou « random mating », fréquemment faite par les biologistes, consiste à admettre que les individus prennent leurs conjoints au hasard dans la population, sans manifester de préférence, sans rechercher ceux qui leur ressemblent. Alors, la connaissance de l'un des conjoints ne nous apprend rien sur l'autre, il y a indépendance stochastique entre leurs caractères. Les probabilités, pour que les gamètes des deux conjoints qui s'unissent pour former un descendant soient de types déterminés sont donc aussi indépendantes l'une de l'autre, la probabilité d'un descendant déterminé est donc le produit des probabilités des deux gamètes qui le constituent, ou la somme de tels produits s'il peut être constitué de plusieurs façons différentes. Si on considère, par exemple, la répartition des individus de F_0 entre les trois génotypes AA , Aa , aa , dans les proportions P , $2Q$, R , la probabilité pour que les deux conjoints d'un couple aient des génotypes déterminés sera le produit des probabilités de ces génotypes.

Supposons la population nombreuse, les différents types de couples se formeront avec des fréquences sensiblement égales à leurs probabilités (aussi emploierons-nous indistinctement dans ce qui suit les termes « probabilité » et « fréquence »), probabilités qui sont les différents coefficients du développement de $(Pu + 2Qv + Rw)^2$. Si tous les types de couples ont la même fécondité, le même nombre moyen d'enfants, ces coefficients sont également les fréquences de leurs enfants dans F_1 . Il y a donc dans F_1 , d'après les lois de Mendel, des

proportions P^2 d'enfants AA , provenant de parents AA et AA , des proportions PQ de AA et PQ de Aa provenant de parents AA et Aa , ainsi que :

Q^2 de AA , $2Q^2$ de Aa , Q^2 de aa provenant de parents Aa et Aa			
PR de Aa	»	»	AA et aa
QR de Aa et QR de aa	»	»	Aa et aa
R^2 de aa	»	»	aa et aa

Les génotypes AA , Aa , aa , sont donc répartis dans F_1 dans les proportions :

$$P_1 = P^2 + 2PQ + Q^2 = (P+Q)^2 = p^2$$

$$2Q_1 = 2PR + 2Q^2 + 2PR + 2QR = 2(P+Q)(Q+R) = 2pq$$

$$R_1 = Q^2 + 2QR + R^2 = (Q+R)^2 = q^2$$

En appelant comme d'usage $p = P+Q$ et $q = Q+R$ ($p+q=1$) les fréquences des gènes A et a dans F_0 .

On voit ainsi que tout se passe comme si les génotypes de F_1 étaient formés par association au hasard de 2 gènes puisés dans un ensemble qui contiendrait une proportion p de gènes A et une proportion q de gènes a . Il devait bien en être ainsi puisque nous avons supposé que les individus forment chacun la même quantité de gamètes fertiles et que celles-ci s'associent au hasard, les gamètes A et a , issues de la population F_0 , sont alors dans les proportions $p = P+Q$, et $q = Q+R$ et leurs 3 types de combinaisons AA , Aa , aa , dans la population F_1 ont bien les fréquences : p^2 , $2pq$, q^2 . Les proportions de gènes dans F_1 sont donc $p_1 = P_1 + Q_1 = p$ et $q_1 = Q_1 + R_1 = q$; les proportions des génotypes dans la génération suivante F_2 seront donc encore p^2 , $2pq$, q^2 ; F_2 a donc la même composition que F_1 ; à partir de la première génération obtenue par « random mating » la composition de la population relativement à un facteur déterminé reste invariable, les proportions des 2 gènes étant p et q et celles des 3 génotypes p^2 , $2pq$, q^2 .

Passage d'une génération à l'autre, pour les probabilités d'association

Examinons maintenant quelle variation provoque au cours des générations le « random mating » sur l'association entre les différents gènes.

Partons d'une population F_0 , classée pour fixer les idées suivant l'état des 3 facteurs. La fréquence de leurs associations dans F_0 est parfaitement caractérisée par les quantités telles que (111), (112), etc. Nous représentons toujours par (111), (113), etc., les associations entre les gènes dans F_0 . Les lois de Mendel nous permettent alors de calculer les fréquences dans la génération suivante F_1 , fréquences que nous désignerons par (111)', (112)', etc. La fréquence (111)' est celle des homozygotes $AAA'A'A''A''$, qui proviennent de l'union de 2 gamètes $AA'A''$ issues de F_0 , gamètes dont la probabilité est (111). De même les hétérozygotes simples tels que $AAA'A'A''a''$, de fréquence (112)', proviennent des deux combinaisons $AA'A'' \times AA'a''$ et $AA'a'' \times AA'A''$ qui ont chacune pour probabilité (111) (113). Les hétérozygotes doubles tels que $AAA'a'A''a''$, des 4 combinaisons $AA'A'' \times Aa'a''$, $Aa'a'' \times AA'A''$, $AA'a'' \times Aa'A''$, $Aa'A'' \times AA'a''$, etc.

On a donc :

$$\begin{aligned}
 (111)' &= (111)^2 \\
 (112)' &= 2(111)(113) \\
 (122)' &= 2(111)(133) + 2(131)(113) \\
 (222)' &= 2(111)(333) + 2(113)(331) + 2(131)(313) + 2(133)(311)
 \end{aligned}
 \tag{6}$$

Les probabilités des associations entre gènes dans F_1 sont fournies par les équations telles que (2) qui donne :

$$\begin{aligned}
 (111)' &= (111)(111) + (111)(113) + (111)(131) + (111)(311) \\
 &+ (1/2)(111)(133) + (1/2)(131)(113) + (1/2)(111)(313) \\
 &+ (1/2)(311)(113) + (1/2)(111)(331) + (1/2)(131)(311) \\
 &+ (1/4)(111)(333) + (1/4)(113)(331) + (1/4)(131)(313) \\
 &+ (1/4)(133)(311)
 \end{aligned}$$

en tenant compte des relations (3) et (4) cela s'écrit :

$$(7) \quad (\mathbf{111})' = (1/4)(\mathbf{110})(\mathbf{001}) + (1/4)(\mathbf{101})(\mathbf{010}) \\ + (1/4)(\mathbf{011})(\mathbf{100}) + (1/4)(\mathbf{111})$$

Cette dernière équation peut d'ailleurs s'établir directement en remarquant que, quand les 3 gènes $AA'A''$ se trouvent associés chez un individu de F_1 (ce qui se présente avec la fréquence $(\mathbf{111})'$), il y a égale probabilité $2/8$ pour qu'ils soient fournis tous 3 par l'un de ses parents (chez lequel ils ont alors la probabilité $(\mathbf{111})$ d'être associés) ou pour que A et A' soient fournis par le même parent et A'' par l'autre [probabilité de cette réunion $(\mathbf{110}) \times (\mathbf{001})$] ou pour que A et A'' soient fournis par le même parent et A' par l'autre [probabilité $(\mathbf{101}) \times (\mathbf{010})$] ou pour que A' et A'' soient fournis par le même parent et A par l'autre [probabilité $(\mathbf{011}) \times (\mathbf{100})$].

Les équations du type (7) nous fournissent les probabilités d'association triple, à partir des probabilités d'association triple et double dans la génération précédente.

Ajoutant à (7) l'équation analogue :

$$(\mathbf{113})' = (1/4)(\mathbf{110}) \times (\mathbf{003}) + (1/4)(\mathbf{103})(\mathbf{010}) \\ + (1/4)(\mathbf{013}) \times (\mathbf{100}) + (1/4)(\mathbf{113})$$

on a :

$$(\mathbf{110})' = (1/4)(\mathbf{110}) + (1/4)(\mathbf{100}) \times (\mathbf{010}) + (1/4)(\mathbf{010})(\mathbf{100}) \\ + (1/4)(\mathbf{113}) + (1/4)(\mathbf{111})$$

c'est-à-dire :

$$(8) \quad (\mathbf{110})' = (1/2)(\mathbf{110}) + (1/2)(\mathbf{010}) \times (\mathbf{100})$$

Cette équation fournit la probabilité d'association entre A et A' qui ont la même probabilité $2/4$ d'être fournis tous 2 par le même parent, ou chacun par un des parents, ce qui permet de la retrouver directement.

En ajoutant l'équation (8) et l'équation analogue :

$$(\mathbf{130})' = (1/2)(\mathbf{130}) + (1/2)(\mathbf{030})(\mathbf{100})$$

On retrouve l'équation $(\mathbf{100})' = (1/2)(\mathbf{100}) + (1/2)(\mathbf{100}) = (\mathbf{100})$

qui exprime la constance d'une génération à l'autre de la proportion $p = (100)$ du gène A .

Indépendance asymptotique des divers facteurs

La variation au cours du temps de l'association entre 2 gènes nous est fournie par l'itération de l'équation (8) :

$$(110)' = (1/2)(110) + (1/2)pp'$$

$$(110)'' = (1/2)(110)' + (1/2)pp' = (1/4)(110) + (3/4)pp'$$

$$(110)^{(n)} = (1/2^n)(110) + (1 - 1/2^n)pp'$$

On voit que quel que soit le degré d'association initial, la fréquence de l'association AA' tend, au cours du temps, vers la limite pp' , produit des fréquences des 2 gènes.

Il en résulte, d'après l'équation (7), que $(111)^{(n)} = (1/4)(111)^{(n-1)}$ tend vers la limite $(3/4)pp'p''$ par suite la probabilité $(111)^{(n)}$ de l'association $AA'A''$ tend asymptotiquement vers le produit $pp'p''$.

On voit aisément par récurrence qu'il en serait encore de même si on considérait un nombre quelconque de gènes.

Les équations telles que (6) montrent que les fréquences des différents génotypes tendent aussi vers des limites, et par exemple les valeurs asymptotiques des probabilités des génotypes $AAA'A'A''A''$, $AAA'A'A''a''$, $AAA'A'a''a''$ sont respectivement $p^2p'^2p''^2 = PP'P''$, $2p^2p'^2p''q'' = 2PP'Q''$ et $p^2p'^2q''^2 = PP'R''$. La répartition des individus entre les 3 états $A''A''$, $A''a''$, $a''a''$, est donc indépendante de l'état des autres couples de gènes.

En conclusion, si une population nombreuse se reproduit dans les conditions du « random mating », la répartition des gènes tend vers un état d'équilibre dans lequel *la fréquence d'une association est le produit des fréquences de ses constituants, et la connaissance de l'état de l'un d'eux ne modifie pas la fréquence relative des autres. Cela revient à dire que, quand l'état d'équilibre est atteint, les aléatoires H , K , L , etc., sont stochastiquement indépendantes.* La répartition du caractère x

dans la population finale est celle d'une somme d'aléatoires indépendantes du 3^e ordre.

Dans les expériences de Mendel (Blaringhem, « L'hérédité mendélienne », p. 32) cette répartition d'équilibre était atteinte dès la 2^e génération F_2 du fait de la composition particulière de F_1 , qui était formée uniquement d'individus hétérozygotes pour tous les facteurs.

Il y a lieu de remarquer en outre que, quand il n'y a pas dominance c'est-à-dire quand $j=i/2+k/2$, l'aléatoire H se décompose en une somme de 2 aléatoires indépendantes du 2^e ordre \mathcal{H} et \mathcal{H}' , qui prennent chacune les valeurs $i/2$ et $k/2$ avec les probabilités p et q , et qui représentent l'action de chacun des 2 gènes. L'aléatoire x est alors une somme d'aléatoires indépendantes du 2^e ordre. Tout se passe comme si les 2 gènes de chaque couple agissaient indépendamment.

Compatibilité de « l'hérédité mêlée » avec la lois de Mendel

Nous pouvons maintenant voir à quoi correspondent les 2 catégories de caractères étudiées par les généticiens et pour lesquelles on a cru parfois devoir distinguer deux modes d'hérédité différents. D'une part certains caractères ne dépendant que d'un petit nombre de facteurs mendéliens présentent à l'observation une échelle nettement discontinue de valeurs, et leur analyse génétique permet alors dans beaucoup de cas de préciser le rôle des différents facteurs.

Au contraire, d'autres caractères apparaissent susceptibles de prendre une échelle continue de valeurs, le caractère des enfants étant alors en général intermédiaire entre ceux des parents. Ce sont les cas « d'hérédité mêlée » que Pearson opposait aux cas d'hérédité « exclusive » ou « alternative » (Ph. Transact. 1901, A, 195, p. 80).

C'est sur les caractères à hérédité mêlée, dont la taille est le type, que Galton avait établi ses lois. Et il avait trouvé que la répartition de ces caractères dans la population obéit sensiblement à une loi de Gauss (Natural Inheritance, p. 90).

Or ce résultat se justifie parfaitement en admettant que ces caractères résultent de l'addition des effets d'un grand nombre n de facteurs mendéliens indépendants. Cette somme x admet une échelle pratiquement continue de valeurs et si on suppose que les contributions des différents facteurs sont petites par rapport à l'écart-type σ_x de la somme ($\sigma_x^2 = \sigma_H^2 + \sigma_K^2 + \dots$ puisqu'il y a indépendance), hypothèse qui se traduit par :

$$\frac{|i|}{\sigma_x} < \varepsilon, \quad \frac{|j|}{\sigma_x} < \varepsilon, \quad \frac{|k|}{\sigma_x} < \varepsilon, \quad \frac{|i'|}{\sigma_x} < \varepsilon, \quad \text{etc.},$$

le théorème de Liapounoff montre que la loi de répartition de la somme différera peu de la loi de Gauss (Paul Lévy, « Théorie de l'addition des variables aléatoires », p. 241) :

$$\left| P\{x/\sigma_x < v\} - (1/\sqrt{2\pi}) \int_{-x}^v \exp(-v^2/2) dv \right| < 6 \varepsilon^{1/4}$$

On peut montrer de même que la loi de corrélation entre les caractères x_1 et x_2 de 2 individus apparentés diffère peu, sous ces hypothèses, d'une loi de Gauss à 2 variables, ce qui permet de la caractériser par le coefficient de corrélation $M(x_1 x_2) / M(x^2)$.

Ainsi, contrairement à ce qu'ont cru certains biologistes, il n'y a pas incompatibilité entre l'hérédité « mêlée » et l'hérédité mendélienne. Une hérédité mêlée en apparence se ramène à l'action d'un grand nombre de facteurs mendéliens. Cette explication avait déjà été fournie par Nilsson-Ehle, dans ses travaux sur l'hérédité des couleurs des grains d'avoine ; puis par East (maïs et tabac) et par Castle et Lang (lapins) (cf. Blaringham, op. cit.).

Il est curieux de noter que cette idée se trouve en germe chez Galton lui-même (Natural Inheritance, p. 139) : « the blending in stature is due to its being the aggregate of the quasi-independent inheritances of many separate parts, while eye-colour appears to be much less various in its origin... we may with some confidence expect that the law

by which those hereditary contributions are found to be governed, may be widely, and perhaps universally applicable ».

Conservation de la variance

R.-A. Fisher, dans le premier chapitre de son ouvrage « The Genetical Theory of Natural Selection », a mis en évidence une conséquence importante des propriétés démontrées plus haut, conséquence qui fournit une intéressante vérification de la théorie mendélienne de l'hérédité. Nous avons vu qu'après quelques générations de brassage la composition génétique de la population reste sensiblement constante : les différents gènes conservent les mêmes proportions et tendent à se répartir indépendamment. La variabilité de la population (mesurée par exemple par la variance $\sigma_x^2 = \sigma_H^2 + \sigma_K^2 + \dots$) reste constante. C'est là une preuve importante à l'appui de l'hypothèse mendélienne de la ségrégation des gènes, car les autres théories de l'hérédité proposées entraîneraient en général une diminution rapide de la variabilité. Par exemple, dans la théorie de l'hérédité mêlée admise implicitement par Darwin, si les écarts de la partie héréditaire de la taille des parents avec la moyenne étaient x_1 et x_2 , l'écart de la taille de leurs enfants serait à peu près $\frac{x_1 + x_2}{2}$ et on aurait :

$$M\left(\frac{x_1 + x_2}{2}\right)^2 = \frac{M(x_1^2) + M(x_2^2)}{4} = \frac{M(x^2)}{2}$$

et la variance héréditaire de la population diminuerait de moitié à chaque génération. Or cette diminution ne peut être compensée, ni par la variabilité produite par l'environnement, car les expériences de Galton sur les jumeaux ont montré quelle était très faible (et dès lors, comment expliquer, en admettant l'hérédité mêlée, la variabilité observée parmi les enfants d'une même famille?) ni par l'apparition de mutations nouvelles qui devraient alors être très fréquentes, alors que les expériences de Johannsen sur les lignées pures

de haricots ont montré qu'un type héréditaire, une fois isolé, se conserve longtemps sans altération de ses caractères, c'est-à-dire sans apparition de mutations.

Le fait de la conservation de la variance ne peut donc s'expliquer que par une théorie de l'hérédité dans laquelle les dissemblances des parents, au lieu de se fondre chez les enfants, subsistent chez eux à l'état latent et peuvent réapparaître dans la descendance. Or le schéma mendélien est précisément, comme le souligne Fisher (*The Genetical Theory*, pages 8 et 11) un des plus simples parmi de tels schémas.

2^e CAS :

HOMOGAMIE

Hypothèse fondamentale

L'hypothèse de « random mating » que nous venons d'étudier n'est admissible qu'en première approximation ; dans certains cas on ne peut pousser l'analyse qu'en tenant compte du phénomène de l'homogamie. Fisher traduit ce phénomène de la façon suivante : supposons que la ressemblance entre deux individus soit caractérisée par la proximité des deux valeurs chez chacun d'eux d'un même caractère mesurable x , soient x_1 et x_2 ces deux valeurs. La fréquence des couples formés d'un individu de valeur x_1 et d'un individu de valeur x_2 , qui serait dans le cas de random mating $\Pi_1 \Pi_2$ (Π_i étant la fréquence des individus de valeur x_i) sera :

$$\Pi_1 \Pi_2 [1 + f(x_1, x_2)]$$

La fonction $f(x_1, x_2)$ que nous pouvons supposer assez petite, car l'homogamie n'introduit en général que des termes correctifs, doit satisfaire aux conditions suivantes :

1) la somme des fréquences de tous les couples doit être 1, ce qui se traduit par :

$$\sum_1 \sum_2 \Pi_1 \Pi_2 [1 + f(x_1, x_2)] = 1$$

c'est-à-dire :

$$\sum_1 \sum_2 \Pi_1 \Pi_2 f(x_1, x_2) = 0 \quad (\text{puisque } \sum_1 \Pi_1 = \sum_2 \Pi_2 = 1)$$

2) $f(x_1, x_2)$ doit être positive si x_1 et x_2 sont tous deux supérieurs ou tous deux inférieurs à la moyenne, et négative si x_1 et x_2 sont de part et d'autre de la moyenne ; il est donc naturel par continuité de lui imposer d'être nulle quand l'une des quantités x_1 ou x_2 est égale à la moyenne.

Si nous supposons pour simplifier que x_1 et x_2 sont évalués en prenant pour origine leur valeur moyenne (c'est-à-dire que $\sum_1 \Pi_1 x_1 = \sum_2 \Pi_2 x_2 = 0$), on peut satisfaire simplement cette dernière condition en prenant avec Fisher :

$$f(x_1, x_2) = \lambda x_1 x_2$$

et la première condition est alors satisfaite car :

$$\sum_1 \sum_2 \Pi_1 \Pi_2 f(x_1, x_2) = \lambda \sum_1 \Pi_1 x_1 \times \sum_2 \Pi_2 x_2 = 0$$

Nous prendrons dorénavant pour la fonction f cette forme-là ; nous ferons donc l'hypothèse suivante :

Hypothèse fondamentale :

Si les fréquences des groupes d'individus de valeurs x_1 et x_2 supposées rapportées à la moyenne comme origine sont respectivement Π_1 et Π_2 , la fréquence des accouplements entre deux individus provenant chacun d'un de ces groupes sera :

$$\Pi_1 \Pi_2 (1 + \lambda x_1 x_2)$$

Si nous caractérisons l'écart au « random mating » par « le coefficient de choix » f obtenu en mettant la fréquence d'accouplement sous la forme $\Pi_1 \Pi_2 (1 + f)$, alors que $\Pi_1 \Pi_2$ est la fréquence dans le cas de « random mating », nous pouvons aussi énoncer ainsi l'hypothèse :

Le coefficient de choix f est proportionnel au produit des écarts x à la moyenne des deux individus (le coefficient de choix f joue un rôle analogue aux coefficients d'association entre gènes f_{lmn} définis plus haut).

La formule du choix ; sa représentation symbolique

Considérons alors les accouplements entre deux catégories quelconques d'individus pris dans la population ; la première catégorie ayant une fréquence totale $\Pi' = \sum \Pi_1$ et une

moyenne $x' = \frac{\Sigma' \Pi_1 x_1}{\Sigma' \Pi_1}$ (le symbole Σ' représentant la somme étendue à la catégorie considérée); la deuxième catégorie ayant de même la fréquence $\Pi'' = \Sigma'' \Pi_2$ et la moyenne $x'' = \frac{\Sigma'' \Pi_2 x_2}{\Sigma'' \Pi_2}$; la fréquence totale des accouplements entre ces deux catégories sera, en décomposant chaque catégorie en groupes d'individus de même valeur :

$$\begin{aligned} \Sigma' \Sigma'' \Pi_1 \Pi_2 (1 + \lambda x_1 x_2) &= \Sigma' \Pi_1 \times \Sigma'' \Pi_2 + \lambda \Sigma' \Pi_1 x_1 \times \Sigma'' \Pi_2 x_2 \\ &= \Pi' \Pi'' (1 + \lambda x' x'') \end{aligned}$$

Donc, dans l'hypothèse fondamentale, la fréquence totale des accouplements entre deux catégories quelconques est : $\Pi' \Pi'' (1 + \lambda x' x'')$, alors que s'il y avait random mating elle serait $\Pi' \Pi''$.

On peut dire aussi que le coefficient de choix entre deux catégories est proportionnel au produit des valeurs moyennes de ces catégories (la moyenne générale étant toujours supposée prise pour origine).

Remarquons qu'il résulte en particulier de cette propriété que la fréquence d'accouplement entre un individu de valeur déterminée x' et un individu quelconque de la population ($\Pi'' = 1, x'' = 0$) est égale à la fréquence du premier individu. Donc, dans la loi d'homogamie adoptée aucun individu n'est handicapé, tous ont la même probabilité de trouver des conjoints mais ces conjoints sont choisis de façon différente suivant la valeur du caractère x chez l'individu.

Il nous sera commode de représenter la fréquence d'accouplement $\Pi' \Pi'' (1 + \lambda x' x'')$ par le produit symbolique $\Pi' \star \Pi''$. L'opération représentée par le symbole \star traduit l'accouplement des deux groupes facteurs Π' et Π'' , elle se réduit au produit ordinaire dans le cas du random mating (c'est-à-dire quand $\lambda = 0$).

Ce produit symbolique est évidemment commutatif comme les produits ordinaires qui le composent. Il est aussi distributif car, si la catégorie Σ'' est la somme de 2 catégories

Σ'''' et Σ'''''' ($\Pi'' = \Pi'''' + \Pi''''''$, $\Pi''x'' = \Sigma''\Pi_2x_2 = \Sigma''''\Pi_2x_2 + \Sigma''''''\Pi_2x_2 = \Pi''''x'' + \Pi''''''x''''$), on a :

$$\begin{aligned} \Pi''\star(\Pi'''' + \Pi''''') &= \Pi''\Pi'''' + \lambda\Pi''x''\Pi''''x'' = \Pi''\Pi'''' + \lambda\Pi''x''\Pi''''x'' \\ &+ \Pi''\Pi'''''' + \lambda\Pi''x''\Pi''''''x'''' = \Pi''\star\Pi'''' + \Pi''\star\Pi'''''' \end{aligned}$$

C'est en somme cette propriété de distributivité que nous avons utilisée pour déduire la formule du choix de l'hypothèse fondamentale.

Rappel de la méthode de Fisher

Nous devons signaler que Fisher a obtenu cette formule d'une autre façon (Ed. p. 410) en partant de l'hypothèse que la corrélation entre les caractères x des 2 conjoints est gaussienne et de coefficient μ , la variance de chacun d'eux étant V . La fréquence d'accouplement entre 2 groupes compris respectivement entre x_1 et $x_1 + dx_1$ et entre x_2 et $x_1 + dx_2$ (groupes de fréquences respectives

$$\Pi_1 = \frac{1}{\sqrt{2\Pi V}} \exp[-x_1^2/2V] dx_1 \text{ et } \Pi_2 = \frac{1}{\sqrt{2\Pi V}} \exp[-x_2^2/2V] dx_2$$

est alors :

$$F = \frac{\Pi_1\Pi_2}{\sqrt{1-\mu^2}} \exp\left[-\frac{\mu^2 x_1^2 - 2\mu x_1 x_2 + \mu^2 x_2^2}{2V(1-\mu^2)}\right]$$

(Si μ est assez petit pour que μ^2 soit négligeable, la partie principale de cette expression est : $\Pi_1\Pi_2 \exp. [\mu x_1 x_2 / V]$ ~~≠~~ $\Pi_1\Pi_2(1 + \mu x_1 x_2 / V)$, et on retrouve alors l'hypothèse fondamentale, mais c'est une hypothèse différente si μ n'est pas très petit). Si on considère 2 catégories de fréquence totale Π' et Π'' , réparties suivant la loi de Gauss avec une variance presque égale à V autour de moyennes respectives x' et x'' , les fréquences élémentaires dans chaque catégories seront

$$\Pi_1 = \frac{\Pi'}{\sqrt{2\Pi V}} \exp. [-(x_1 - x')^2 / 2V] dx_1$$

$$\text{et } \Pi_2 = \frac{\Pi''}{\sqrt{2\Pi V}} \exp. [-(x_2 - x'')^2 / 2V] dx_2$$

et la fréquence totale des accouplements entre les deux catégories sera donnée par l'intégrale double de F étendue aux deux catégories, c'est-à-dire :

$$\int_{-x}^{+x} \int_{-x}^{+x} \frac{\Pi_1 \Pi_2}{\sqrt{1-\rho^2}} \exp. \left[-\frac{\rho^2 x_1^2 - 2\rho x_1 x_2 + \rho^2 x_2^2}{2V(1-\rho^2)} \right] =$$

$$\Pi' \Pi'' \int_{-x}^{+\infty} \int_{-x}^{+\infty}$$

$$\exp. \left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2 - 2(1-\rho^2)(x_1 x_1' + x_2 x_2') + (1-\rho^2)(x_1'^2 + x_2'^2)}{2V(1-\rho^2)} \right]$$

$$\left[\frac{dx_1 dx_2}{2\Pi \sqrt{1-\rho^2}} \right]$$

en posant $x_1 = y_1 + x' + \rho x''$ et $x_2 = y_2 + x'' + \rho x'$, les termes du premier degré en y_1 et y_2 dans le numérateur de l'exposant disparaissent ; les termes indépendants de y_1 et y_2 se réduisent à $-2x'x''\rho(1-\rho^2)$ qui fournit dans l'intégrale le facteur constant $\exp. [\rho x'x''/V]$ et les termes du deuxième degré fournissent une intégrale égale à 1 (intégrale sur tout le plan de la loi de Gauss à 2 variables) ; la fréquence d'accouplement entre les 2 catégories est donc $\Pi' \Pi'' \exp. [\rho x'x''/V]$ qui est sensiblement égale à $\Pi' \Pi'' (1 + \rho x'x''/V)$ si x' et x'' sont petits par rapport à $\sqrt{V} = \sigma_x$.

Il en est ainsi si x résulte, comme page 17, de l'addition d'effets de même ordre de grandeur d'un grand nombre de couples de gènes, et si les 2 catégories s'obtiennent en choisissant les individus pour lesquels quelques couples de gènes seulement sont dans un état déterminé. Alors la variance V n'est sensiblement pas modifiée, et les nouvelles moyennes sont faibles par rapport à l'écart-type σ_x . La formule du choix est donc valable pour de telles catégories, même si ρ n'est pas très petit, et la propriété de distributivité montre que cette formule se conserve quand on groupe entre elles certaines de ces catégories. Cette remarque éclaire les résultats déduits par Fisher de son hypothèse.

Passage d'une génération à l'autre

Pour fixer les idées considérons 3 couples de gènes et supposons la population classée en 27 catégories, suivant l'état de ces 3 couples.

Soient toujours (111), (112), etc., les fréquences de ces catégories, et soient x_{111} , x_{112} , etc., les valeurs moyennes du caractère dans chaque catégorie. La fréquence ou probabilité (111)' de l'homozygote AAA'A'A''A''' dans la génération suivante F_1 s'obtiendra en multipliant la probabilité (fournie par la formule du choix) de chaque accouplement possible dans F_0 par la probabilité (fournie par la loi de Mendel) pour qu'il donne cet homozygote, et en ajoutant tous ces termes. On obtient ainsi :

$$\begin{aligned}
 & (111) \star (111) \times 1 \\
 3 \text{ termes tels que } & 2(111) \star (112) \times 1/2 \\
 3 \text{ termes tels que } & (112) \star (112) \times 1/4 \\
 3 \text{ termes tels que } & 2(112) \star (121) \times 1/4 \\
 3 \text{ termes tels que } & 2(111) \star (122) \times 1/4 \\
 3 \text{ termes tels que } & 2(112) \star (221) \times 1/8 \\
 6 \text{ termes tels que } & 2(112) \star (122) \times 1/8 \\
 & \quad - 2(111) \star (222) \times 1/8 \\
 3 \text{ termes tels que } & 2(112) \star (222) \times 1/16 \\
 3 \text{ termes tels que } & (122) \star (122) \times 1/16 \\
 3 \text{ termes tels que } & 2(122) \star (212) \times 1/16 \\
 3 \text{ termes tels que } & 2(122) \star (222) \times 1/32 \\
 & (222) \star (222) \times 1/64
 \end{aligned}$$

(les autres accouplements ne peuvent pas fournir l'homozygote considéré). Or les propriétés de distributivité et de commutativité du produit symbolique montrent que la somme de ces termes n'est autre que le développement du produit symbolique par elle-même de la quantité :

$$(111) + \frac{(112)}{2} + \frac{(121)}{2} + \frac{(211)}{2} + \frac{(122)}{4} + \frac{(212)}{4} + \frac{(221)}{4} + \frac{(222)}{8}$$

quantité qui n'est autre que la fréquence (111) de la formule (2) (p. 8), fréquence d'association des gènes AA'A'', ou

obtenues en assignant un état donné à 2 ou un seul gène satisfait à :

$$(11) \quad \left\{ \begin{array}{l} (110)\mathbf{x}_{110} = (111)\mathbf{x}_{111} + (113)\mathbf{x}_{113} \\ (100)\mathbf{x}_{100} = (110)\mathbf{x}_{110} + (130)\mathbf{x}_{130} \end{array} \right.$$

(et, pour que la moyenne générale soit nulle, on doit avoir :

$$\mathbf{x}_{000} = (100)\mathbf{x}_{100} = (300)\mathbf{x}_{300} = 0)$$

Nous avons alors l'équation, généralisation de (7) :

$$(12) \quad (111)' = (1/4)(110)\star(001) + (1/4)(101)\star(010) \\ + (1/4)(011)\star(100) + (1/4)(111)$$

en ajoutant à cette équation l'équation analogue qui donne $(113)'$ et grâce encore à la distributivité, nous avons l'équation pour 2 gènes fixés seulement (cf. équation (8)).

$$(13) \quad (110)' = (1/2)(110) + (1/2)(010)\star(100)$$

et l'équation pour un seul gène s'obtient en ajoutant encore l'expression de $(130)'$: $(100)' = (100)$.

On voit donc qu'avec la loi d'homogamie adoptée, *la proportion de chaque gène reste constante au cours des générations* comme dans le cas du random mating. Nous désignerons toujours ces proportions par les notations :

$$p = (100), \quad q = (300), \quad p' = (010), \text{ etc.}$$

Valeurs asymptotiques des probabilités d'association

Les équations telles que (13), (12), nous fournissent la variation d'une génération à l'autre des fréquences d'association entre gènes, à condition de connaître chaque fois les quantités telles que \mathbf{x}_{100} , \mathbf{x}_{110} , etc., qui représentent la valeur moyenne du caractère quand 1 gène, 2 gènes, etc., sont fixés.

Les hypothèses que l'on pourrait faire sur la variation de ces quantités étant arbitraires, nous supposons simplement qu'elles tendent vers une valeur stable.

Alors l'équation (13), qui s'écrit :

$$(110)' - (010)\star(100) = (1/2)[(110) - (010)\star(100)]$$

montre que (110) tend vers la limite :

$$(14) \quad (110) = (010)\star(100) = pp'(1 + \lambda\mathbf{x}_{100}\mathbf{x}_{010})$$

De même l'équation (12) montre que (111) tend vers la limite :

$$(15) \quad (111) = (1/3)(110) \star (001) + (1/3)(101) \star (010) \\ + (1/3)(011) \star (100)$$

Nous pourrions de même obtenir les fréquences d'association entre plus de 3 gènes, mais cela ne nous sera pas nécessaire pour la suite.

Les formules (10) nous montrent alors que les fréquences d'association entre les différents facteurs tendent aussi vers des limites qui sont :

$$(16) \quad (111) = (111)^2 (1 + \lambda \mathbf{x}_{111}^2) \\ (112) = 2(111)(113) (1 + \lambda \mathbf{x}_{111} \mathbf{x}_{113})$$

Les formules (14), (15), (16) expriment ainsi la composition génétique de la population lors de l'équilibre.

Notations et résultats de Fisher

Pour montrer l'analogie avec les résultats de Fisher, reprenons ses notations indiquées au début de ce chapitre ; représentons par $i, j, k ; i', j', k', \text{etc.}$, les contributions possibles des différents facteurs $H, K, \text{etc.}$ au caractère x quand la population a atteint son état d'équilibre génétique. On peut évidemment compter les contributions de chaque facteur à partir d'une origine arbitraire, sous réserve que la moyenne générale $M(x)$ de x dans toute la population soit nulle (on a vu p. 21 que cette condition est nécessaire pour que la formule de choix adoptée soit admissible). Or :

$$M(x) = M(H) + M(K) + \dots \text{etc.}$$

On peut donc supposer que les origines pour $H, K, \text{etc.}$, sont choisies de telle façon que :

$$(17) \quad \left\{ \begin{array}{l} M(H) = Pi + 2Qj + Rk = 0 \\ M(K) = P'i + 2Q'j' + R'k' = 0, \text{ etc.} \end{array} \right.$$

Nous supposerons toujours par la suite qu'il en est ainsi.

(Si la population n'est pas en état d'équilibre, pour que la propriété $M(x) = 0$, ou encore les formules (17), se conservent d'une génération à l'autre, il serait nécessaire de supposer

L'écart des proportions des 3 phases avec les proportions p^2 , $2pq$, q^2 , qu'elles auraient s'il y avait random mating, est mesuré par le terme correctif $(\lambda/4)(PI-RK)^2$.

Ces relations ont été fournies approximativement par Fisher (Ed., p. 411). De même, les fréquences d'association entre 2 gènes non homologues, par exemple A et A' , sont données par l'équation (14) :

$$(21) \quad pp'(1+f_{11}) = (\mathbf{110} \dots) = pp'(1 + \lambda \mathbf{x}_{100} \dots \mathbf{x}_{010} \dots) \\ = pp'(1 + \lambda TT')$$

f_{13} , f_{31} , f_{33} , sont donnés par des équations analogues qui se résument en :

$$pp'f_{11} = -pq'f_{13} = -qp'f_{31} = qq'f_{33} = (\lambda/4)(PI-RK)(P'I - R'K')$$

mais les coefficients d'association f_{ij} entre les phases de 2 facteurs, coefficients qui s'obtiennent à partir d'équations du type (16), n'ont pas d'expression rigoureuse simple, et les expressions données par Fisher (équation XXI) n'en sont que des approximations, obtenues en écrivant dans l'équation (XIX) (p. 412) $I+I'$ au lieu de x_{11} , etc.

Conclusion

Ce chapitre a eu pour but d'étudier, lorsque l'équilibre est atteint, la liaison entre les aléatoires H , K , etc., dont la somme est le caractère x considéré. L'aléatoire x est dès lors connue et nous pourrons alors dans les chapitres suivants déterminer la corrélation entre les caractères x chez 2 individus apparentés, pris dans une population stable, c'est-à-dire résoudre dans la théorie mendélienne le problème étudié empiriquement par Karl Pearson et son école. Nous examinerons d'abord le cas où il y a « random mating », puis nous étudierons quelles modifications y introduit l'homogamie exprimée par l'hypothèse faite plus haut.

CHAPITRE II

LES CORRÉLATIONS ENTRE APPARENTÉS DANS LE CAS DU "RANDOM MATING" EN L'ABSENCE DE DOMINANCE

Nous avons vu p. 16 que, quand il n'y a pas de dominance, le caractère considéré, que nous désignerons dans ce chapitre par z , peut être envisagé comme une somme d'aléatoires indépendantes du 2^e ordre :

$$z = \mathcal{H} + \mathcal{H}' + \mathcal{K} + \mathcal{K}' + \dots$$

\mathcal{H} et \mathcal{H}' pouvant prendre les valeurs $t=i/2$ et $s=k/2$ avec les probabilités p et q ;

\mathcal{K} et \mathcal{K}' pouvant prendre les valeurs $t'=i'/2$ et $s'=k'/2$ avec les probabilités p' et q' , etc. (on suppose toujours que les valeurs moyennes sont nulles : $ps+qt=0$, $p's'+q't'=0\dots$).

Nous désignerons toujours le « couple de gènes » $\mathcal{H} + \mathcal{H}'$ par H , etc.

Corrélation entre parent (1) et enfant ; principe fondamental

Désignons la valeur du caractère chez un des parents, le père pour fixer les idées, par :

$$z = \mathcal{H} + \mathcal{H}' + \mathcal{K} + \mathcal{K}' \dots$$

chaque terme représentant la valeur de la contribution d'un des gènes.

Désignons de même la valeur chez la mère par :

$$\bar{z} = \bar{\mathcal{H}} + \bar{\mathcal{H}}' + \bar{\mathcal{K}} + \bar{\mathcal{K}}' \dots$$

(1) Nous emploierons toujours dans ce qui suit le mot "parent" pour désigner "père ou mère", sinon nous dirons "apparentés".

(Il résulte de l'« association au hasard des conjoints » que les aléatoires $\overline{\mathcal{H}}$, $\overline{\mathcal{H}'}$, etc., sont indépendantes des aléatoires \mathcal{H} , \mathcal{H}' , \mathcal{K} , etc.)

Désignons le caractère chez un des enfants de ce couple par :

$$z_1 = \mathcal{H}_1 + \mathcal{H}'_1 + \mathcal{K}_1 + \mathcal{K}'_1 + \dots$$

Les lois de Mendel reviennent à dire que les 2 gènes homologues de l'enfant sont tirés au hasard, l'un parmi les 2 gènes correspondants du père, l'autre parmi les 2 gènes de la mère.

Si \mathcal{H}_1 est par exemple la contribution d'un gène fourni par le père, \mathcal{H}_1 a donc la probabilité 1/2 d'être identique à \mathcal{H} , et la probabilité 1/2 d'être identique à \mathcal{H}' . Si le père est fixé, on a donc pour les enfants l'espérance mathématique :

$$M(\mathcal{H}_1) = (\mathcal{H} + \mathcal{H}') / 2$$

De même, si la mère est fixée, on a pour la contribution qu'elle fournit aux enfants :

$$M(\mathcal{H}'_1) = (\overline{\mathcal{H}} + \overline{\mathcal{H}'}) / 2$$

d'où :
$$M(H_1) = M(\mathcal{H}_1) + M(\mathcal{H}'_1) = (H + \overline{H}) / 2$$

Donc, si on suppose fixées les contributions du père et de la mère, l'espérance mathématique de la contribution des enfants est la moyenne des contributions des deux parents.

C'est ce principe fondamental, traduction mathématique des lois de Mendel, qui va nous servir de base pour le calcul des corrélations entre apparentés.

On a, en supposant toujours fixés le père et la mère :

$$\begin{aligned} M(z_1) &= M(H_1) + M(K_1) + \dots \\ &= (H + K + \dots) / 2 + (\overline{H} + \overline{K} + \dots) / 2 \end{aligned}$$

$$M(z_1) = (z + \overline{z}) / 2$$

Donc quand on fixe la composition génétique du père et de la mère, ou évidemment, plus généralement, quand on fixe les valeurs z et \overline{z} du caractère chez le père et chez la mère, l'espérance mathématique du caractère chez les enfants est la moyenne des valeurs chez les 2 parents.

(Remarquons que c'est cette quantité, $z_m = (z + \overline{z}) / 2$, que Galton a appelée la valeur du « midparent ».)

Si on fixe seulement la composition génétique du père, celle de la mère en est indépendante du fait du « random mating », donc alors $M(\bar{H})=0$ et

$$M(H_1)=H/2 \quad M(z_1)=M(H_1)+M(K_1)+\dots=z/2$$

formule encore valable si on fixe chez le père la valeur z seulement.

La régression des enfants par rapport à un des parents est donc linéaire et de coefficient 1/2.

Inversement d'ailleurs, si on se donne chez un enfant $H_1=\mathcal{H}_1+\mathcal{H}'_1$ et si on désigne par \mathcal{H} le gène du père qui a été transmis à cet enfant, \mathcal{H} est égal avec la probabilité 1/2, à \mathcal{H}_1 ou à \mathcal{H}'_1 , et \mathcal{H}' en est indépendant, on a donc alors :

$$M(\mathcal{H})=(\mathcal{H}_1+\mathcal{H}'_1)/2 \text{ et } M(\mathcal{H}')=0$$

donc : $M(H)=M(\mathcal{H})=H_1/2$

et par suite : $M(z) = z_1/2$

La régression inverse est donc, elle aussi, linéaire et de coefficient 1/2. *Le coefficient de corrélation parent-enfant est donc égal à la valeur commune des coefficients de régression, c'est-à-dire 1/2.*

Pour la corrélation entre un enfant et le *midparent* de Galton ($z_m=(z+\bar{z})/2$), nous avons vu que $M(z_1)=z_m$. Inversement, si on se donne z_1 , on a :

$$M(z)=z_1/2=M(\bar{z})=M(z_m)$$

et le coefficient de corrélation est la moyenne géométrique des coefficients de régression, c'est-à-dire $1/\sqrt{2}$.

Corrélations ancestrales; théorèmes sur l'enchaînement des corrélations

Passons à la corrélation des petits-enfants. D'après les lois de Mendel, la contribution H_2 d'un couple de gènes chez les petits-enfants ne dépend de la contribution H chez le père que par l'intermédiaire de la contribution H_1 chez le fils ; en d'autres termes, *les aléatoires H et H_2 deviennent indépen-*

dantes quand la valeur de l'aléatoire H_1 est fixée ; quand on fixe la constitution génétique du fils, il n'y a plus de liaison stochastique entre le grand-père et le petit-fils. Il est facile dès lors de calculer le coefficient de corrélation r_{02} entre H et H_2 connaissant les corrélations $H-H_1$ et H_1-H_2 (r_{01} et r_{12}).

En effet, désignons par $M'(H)$ et $M'(H_2)$ les espérances mathématiques de H et H_2 quand H_1 est fixée. On a :

$$r_{02} = \frac{M(HH_2) / M(H^2)^{1/2} M(H_2^2)^{1/2}}{M[M'(H) \times M'(H_2)] / M(H^2)^{1/2} M(H_2^2)^{1/2}}$$

ce qui traduit simplement le fait qu'on peut calculer la valeur moyenne du produit en fixant d'abord H_1 (on a alors le produit des valeurs moyennes puisqu'il y a indépendance) puis en donnant à H_1 toutes les valeurs possibles.

Si maintenant nous supposons de plus qu'une des régressions par rapport à la variable intermédiaire H_1 est linéaire, c'est-à-dire que par exemple (en supposant les aléatoires rapportées à leur moyenne)

$$M'(H_2) = KH_1$$

(ce qui est le cas dans l'application considérée : $M'(H_2) = H_1/2$) il vient alors :

$$r_{02} = \frac{KM[H_1 M'(H)] / M(H^2)^{1/2} M(H_2^2)^{1/2}}{KM(HH_1) / M(H^2)^{1/2} M(H_2^2)^{1/2}}$$

or on a : $K = r_{12} \frac{M(H_2^2)^{1/2}}{M(H_1^2)^{1/2}}$

et par suite :

$$r_{02} = r_{01} \times r_{12}$$

Nous avons donc démontré le théorème suivant, qui nous servira à plusieurs reprises :

Etant données 3 aléatoires H , H_1 , H_2 rapportées à leurs moyennes, et telles que 2 d'entre elles (H et H_2 par exemple) deviennent indépendantes quand la 3^e (H_1) est fixée, si une des régressions de H ou H_2 par rapport à la variable intermédiaire H_1 est linéaire, le coefficient de corrélation r_{02} est le produit des coefficients intermédiaires r_{01} et r_{12} .

En appliquant ce résultat au problème que nous nous posons, on voit immédiatement que la corrélation entre H et H_2 est $(1/2)^2$.

Donc la corrélation entre les caractères z et z_2 du grand-père et du petit-fils est (en désignant par σ l'écart-type de l'aléatoire H , et en tenant compte de ce que les différents couples de gènes sont indépendants) :

$$M(zz_2)/\Sigma\sigma^2 = \Sigma M(HH_2)/\Sigma\sigma^2 = 1/4$$

On peut remarquer en outre que la régression des petits-enfants par rapport aux grands-parents est encore linéaire.

Cela résulte d'un théorème un peu différent du précédent.

Si les régressions de H_2 en H_1 et de H_1 en H sont linéaires (toutes ces aléatoires étant supposées rapportées à leur moyenne) on a $H_1 = \lambda H + h_1$, l'aléatoire h_1 ayant une moyenne nulle quand H est fixé ; et $H_2 = \mu H_1 + h_2$, h_2 ayant une moyenne nulle quand H_1 est fixé ; si de plus H_2 ne dépend de H que par l'intermédiaire de H_1 , h_2 a aussi une moyenne nulle quand H est fixé (la probabilité h_2/H étant alors :

$$\frac{\Sigma h_2/H_1 \times H_1/H}{H_1}$$

et on a alors :

$$H_2 = \mu\lambda H + \mu h_1 + h_2$$

l'aléatoire $\mu h_1 + h_2$ ayant une moyenne nulle quand H est fixé.

On voit donc que la régression de H_2 en H est encore linéaire, et que les coefficients de régression se multiplient.

En particulier, dans l'application que nous considérons, on a, quand H est fixé : $M(H_2) = (1/2^2)H$ et par suite, quand la constitution génétique du père est fixée :

$$M(z_2) = \Sigma M(H_2) = (1/2)^2 \Sigma H = (1/2)^2 z$$

formule qui subsiste évidemment quand on ne fixe chez le père que la valeur z du caractère ; il en est de même pour la régression de z en z_2 ; nous voyons donc que les régressions grand-parentales sont linéaires, et nous retrouvons le coefficient de corrélation $(1/2)^2$.

Tous ces résultats se généralisent immédiatement pour la

corrélation d'un individu avec un ancêtre de rang n quelconque. Nos théorèmes sont applicables du fait que les contributions, chez l'un et chez l'autre, d'un couple de gènes, deviennent stochastiquement indépendantes quand on fixe un ancêtre intermédiaire. Soient H' , H'' , H''' ces contributions chez 3 individus appartenant à une ligne de parenté directe, et rangés par ordre d'âge. Alors si les régressions $H'-H''$ et $H''-H'''$ sont linéaires, la régression $H'-H'''$ est linéaire et son coefficient est le produit des 2 autres. Par récurrence, on en déduit immédiatement que *la corrélation entre un individu et un ancêtre de rang n est $(1/2)^n$, et que les régressions correspondantes sont linéaires.*

Cas des frères; théorème plus général

La relation entre 2 frères est l'exemple le plus simple d'un cas de parenté où les 2 individus sont reliés par l'intermédiaire de leurs 2 parents, et non plus par un seul comme dans le cas de la parenté en ligne directe. Il résulte du mécanisme mendélien que, quand on connaît la constitution génétique des 2 parents, la constitution (aléatoire) de leurs descendants est indépendante de la connaissance génétique d'autres individus apparentés. Soient toujours H et \bar{H} les contributions d'un couple de gènes chez 2 conjoints; quand elles sont fixées, les contributions H_1 et H_2 chez 2 de leurs enfants sont donc indépendantes; on a vu qu'elles ont pour valeur moyenne $(H + \bar{H})/2$; la régression de H_2 par rapport à l'ensemble $H\bar{H}$ est donc linéaire; nous sommes donc amenés à généraliser le théorème du paragraphe précédent au cas de régressions multiples linéaires:

Soit une aléatoire H_2 en régression linéaire par rapport à l'ensemble H, \bar{H}, \dots ; supposons aussi que H, \bar{H}, \dots soient en régression linéaire par rapport à un ensemble H_1, H'_1, \dots ; cela se traduit par:

$$(1) \quad M_0(H_2) = aH + \bar{a}\bar{H} + \dots \quad \text{quand } H, \bar{H}, \dots \text{ sont fixés.}$$

$$(2) \quad \left\{ \begin{array}{l} M_1(H) = \bar{b}H_1 + \bar{b}'H'_1 + \dots \\ M_1(\bar{H}) = \bar{b}H_1 + \bar{b}'H'_1 + \dots \end{array} \right\} \quad \text{quand } H_1, H'_1, \dots \text{ sont fixés.}$$

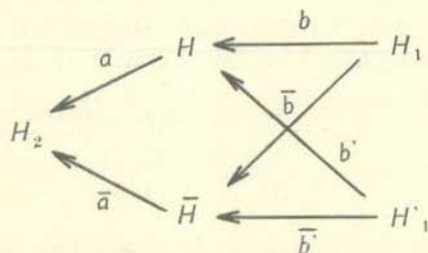
si de plus H_2 ne dépend de H_1, H'_1, \dots , que par l'intermédiaire de H, \bar{H}, \dots , ⁽¹⁾ on a (les quantités entre crochets étant des probabilités suivant la notation de Keynes, et en supposant, pour simplifier les notations, qu'il n'y ait que des couples de variables) :

$$[H_2/H_1, H'_1] = \sum_u \sum_{\bar{H}} [H_2/H, \bar{H}] \times [H, \bar{H}/H_1, H'_1]$$

d'où la moyenne de H_2 quand on fixe H_1, H'_1, \dots :

$$\begin{aligned} M_1(H_2) &= \sum H_2 \times [H_2/H_1, H'_1] \\ &= \sum_u \sum_{\bar{H}} M_0(H_2) \times [H, \bar{H}/H_1, H'_1] \\ &= \sum_u \sum_{\bar{H}} (aH + \bar{a}\bar{H}) [H, \bar{H}/H_1, H'_1] \\ &= aM_1(H) + \bar{a}M_1(\bar{H}) \\ &= (ab + \bar{a}\bar{b})H_1 + (ab' + \bar{a}\bar{b}')H'_1 \end{aligned}$$

La régression résultante est donc aussi linéaire, et chacun des coefficients de régression correspondants est la somme des produits des coefficients composants effectués le long de tous les trajets de liaison correspondants, trajets représentés sur le schéma ci-dessous :



(cette représentation schématique de l'influence stochastique est inspirée de la théorie des path-coefficients de S. Wright).

Nous avons supposé pour alléger les formules qu'il n'y avait que des couples de variables, mais cette propriété serait évidemment encore exacte pour des régressions multiples à un nombre quelconque de variables.

(1) Expression qu'il nous sera commode d'employer avec le sens précis suivant : les aléatoires extrêmes deviennent indépendantes quand on fixe la valeur des aléatoires intermédiaires. Nous appellerons cette hypothèse : hypothèse de "causalité par échelons".

Cela revient à dire, en vertu de la formule

$$M_1(H_2) = aM_1(H) + \bar{a}M_1(\bar{H})$$

que, sous l'hypothèse de « causalité par échelons », on obtient la régression résultante, à partir des régressions linéaires composantes définies par les formules (1) et (2), simplement en remplaçant dans (1) les quantités H , \bar{H} ..., par leurs moyennes fournies par (2).

Il en serait encore évidemment de même dans tout schéma de « causalité par échelons » comportant un nombre quelconque d'échelons intermédiaires au lieu d'un seul. Il suffirait de procéder par substitutions successives dans les formules représentant les régressions intermédiaires (à condition toujours de supposer ces régressions linéaires).

Ce théorème nous permettra d'obtenir la corrélation pour n'importe quel degré de parenté.

Corrélation entre frères

Revenons d'abord au cas des frères qui, ne dépendent l'un de l'autre que par l'intermédiaire de leurs 2 parents. Avec les notations adoptées plus haut pour ce cas, on a :

$$(1) \quad M_0(H_2) = (H + \bar{H}) / 2$$

$$(2) \quad \begin{cases} M_1(H) = H_1 / 2 \\ M_1(\bar{H}) = \bar{H}_1 / 2 \end{cases}$$

d'où, quand on fixe la valeur, H_1 , pour l'un des frères :

$$M_1(H_2) = (1/2)M_1(H) + (1/2)M_1(\bar{H}) = H_1 / 2$$

par suite $M_1(z_2) = \Sigma H_1 / 2 = z_1$, quand on fixe la constitution génétique, ou seulement la valeur z_1 du caractère, chez un des frères.

La régression frère-frère est donc linéaire, et le coefficient de corrélation est 1/2.

Remarquons ici que, pour des *demi-frères*, c'est-à-dire des individus qui ne sont apparentés que par l'intermédiaire de

leur père par exemple, les résultats du 1^{er} paragraphe montrent qu'ils sont liés à lui par des régressions linéaires de coefficient $1/2$, et par suite qu'ils sont liés entre eux par une *régression linéaire de coefficient $1/4$* .

Corrélations collatérales

Elles se déduisent immédiatement des précédentes.

L'oncle et le neveu ne sont liés que par l'intermédiaire d'un des parents de ce dernier, la corrélation oncle-neveu résulte donc de la composition, suivant le théorème de p. 35, d'une corrélation frère-frère avec une corrélation père-fils. La régression résultante est *linéaire*, comme les régressions composantes. Son coefficient est le même, que ce soit l'oncle ou le neveu qu'on fixe ; sa valeur $(1/2)^2$ représente donc *le coefficient de corrélation oncle-neveu*.

On aurait de même, pour un oncle et un petit neveu de rang n : $(1/2)^{n+1}$.

La liaison entre cousins germains résulte de la composition d'une liaison fils-père, une liaison frère-frère, une liaison père-fils. Il y a donc encore *des régressions linéaires, avec un coefficient $(1/2)^3$* .

Des cousins quelconques sont des individus tels que le n^e ancêtre de l'un et le p^e ancêtre de l'autre sont frères. En supposant qu'il n'y a pas d'autre trajet de liaison que celui qui passe par ces 2 ancêtres, on trouve encore *des régressions linéaires avec un coefficient $(1/2)^{n+p+1}$* .

Ce calcul s'étend aussi à tout couple d'individus qui sont apparentés par l'intermédiaire d'un nombre quelconque de liaisons père-fils ou frère-frère, mais ceci à *condition que le trajet de liaison soit unique*. Le coefficient de corrélation est alors $(1/2)^{n+1}$, n désignant le nombre d'échelons intermédiaires ; (étant entendu que dans le trajet de liaison 2 frères représentent 2 échelons consécutifs, sans que l'on fasse intervenir leurs parents comme intermédiaires).

Ce cas d'un trajet de liaison unique représente, dans une

population nombreuse se reproduisant par « random mating », le type de parenté général.

C'est en effet le cas où 2 individus quelconques pris sur le trajet, ou bien sont frères, ou bien sont apparentés seulement par l'intermédiaire d'un des parents de l'un. Or le cas où, au contraire, ils seraient apparentés par l'intermédiaire de l'un et l'autre de leurs parents est un cas exceptionnel, puisqu'il suppose que 2 individus choisis en raison de la parenté d'un de leurs parents présentent en même temps une parenté par l'intermédiaire de l'autre, parenté qui, en raison du « random mating » n'a qu'une probabilité faible.

Nous allons néanmoins donner quelques indications sur ce cas.

Cas de parentés plus complexes

Examinons d'abord le cas des « doubles cousins germains », c'est-à-dire des individus dont, par exemple, les pères sont frères et les mères sont sœurs. Cette fois-ci le trajet de liaison, défini comme ci-dessus, n'est plus unique ; il y en a 2, un qui passe par les pères et un qui passe par les mères. Désignons par H_1 et H'_1 les contributions chez les 2 individus ; par H et \bar{H} , H' et \bar{H}' , les contributions chez leurs parents respectifs. H'_1 est lié à H_1 par l'intermédiaire des régressions successives suivantes :

$$\begin{array}{l} M(H) = H_1/2 \\ M(\bar{H}) = H_1/2 \end{array} \left. \vphantom{\begin{array}{l} M(H) = H_1/2 \\ M(\bar{H}) = H_1/2 \end{array}} \right\} \text{ quand } H_1 \text{ est fixé.}$$

$$\begin{array}{l} M(H') = H/2 \\ M(\bar{H}') = \bar{H}/2 \end{array} \left. \vphantom{\begin{array}{l} M(H') = H/2 \\ M(\bar{H}') = \bar{H}/2 \end{array}} \right\} \text{ quand } H \text{ et } \bar{H} \text{ sont fixés.}$$

Il en résulte, d'après le théorème de p. 38 que, quand H_1 est fixé, on a :

$$M(H'_1) = H_1/4$$

La régression entre double cousins est donc linéaire et de coefficient 1/4. Ce calcul se généralise et donne un procédé

de récurrence pour calculer la corrélation dans le cas où les 2 parents de chaque individu sont apparentés respectivement aux 2 parents de l'autre, car il permet de déduire la corrélation entre les 2 individus considérés des corrélations de leurs parents ; en effet, si nous conservons les mêmes notations, et si nous désignons par :

$$\begin{aligned} M(H') &= rH \\ M(\bar{H}') &= \bar{r}\bar{H} \end{aligned}$$

les régressions entre chaque parent de H'_1 et les parents correspondants de H_1 , le calcul précédent donne immédiatement :

$$M(H'_1) = (H' + \bar{H}') / 2 = (rH + \bar{r}\bar{H}) / 2 = (r + \bar{r})H_1 / 4$$

On déduit ainsi la corrélation des enfants de celle des parents, et en remontant ainsi de proche en proche on finira toujours par arriver à un type de parenté pour lequel on connaît la corrélation.

Nous avons supposé dans tout ce qui précède qu'il n'y avait *jamais de parenté entre les conjoints* (sauf naturellement celle qui peut résulter exceptionnellement de l'association au hasard). Mais les résultats sont tout différents si les conjoints se choisissent précisément en raison de leur parenté (on est alors dans le cas du « consanguine mating » : choix de conjoints en parenté étroite, pratiqué par certains botanistes, et étudié en détail par Sewall Wright) ou en raison de leur ressemblance (c'est l'homogamie, que nous étudierons dans le dernier chapitre).

CHAPITRE III

LES CORRÉLATIONS DANS LE CAS DU "RANDOM MATING" AVEC DOMINANCE

Contributions additives des gènes, et résidu de dominance

Pour traiter le cas plus général où il y a une certaine dominance, il faut, comme nous avons vu (p. 5), supposer que la contribution H d'un couple de gènes n'est plus la somme des contributions \mathcal{H} et \mathcal{H}' fournies par chacun des 2 gènes ; l'aléatoire H est alors une aléatoire du 3^e ordre quelconque, prenant des valeurs i, j, k (alors que, s'il n'y a pas dominance, H est la somme de 2 aléatoires du 2^e ordre \mathcal{H} ou \mathcal{H}' prenant chacune les 2 valeurs l ou s ; les 3 valeurs $i=2l, j=s+l, k=2s$, que peut prendre H , sont alors en progression arithmétique).

(On suppose toujours pour simplifier les calculs que les valeurs de chaque aléatoire sont comptées à partir de la valeur moyenne comme origine, c'est-à-dire que :

$$(1) \quad p^2i + 2pqj + q^2k = 0)$$

Ainsi dans le cas général, si l'aléatoire H résulte bien toujours de l'action combinée des 2 gènes du couple, les effets de chacun de ces gènes ne sont plus additifs. (D'ailleurs, s'ils sont additifs quand on les mesure d'une certaine manière, il suffit en général de les mesurer sur une autre échelle pour que cela cesse.) Suivant l'idée de R.-A. Fisher, nous continuerons à représenter les contributions des 2 gènes par 2 aléatoires additives \mathcal{H} et \mathcal{H}' , prenant de valeurs fictives s ou l , que nous déterminerons au mieux. $\mathcal{H} + \mathcal{H}'$ prend alors les 3 valeurs $2l, s+l, 2s$. Pour représenter l'aléatoire H , il faut alors corriger la somme $\mathcal{H} + \mathcal{H}'$ en lui ajoutant un « résidu » d qui traduit l'effet de non additivité, de la domi-

nance. L'aléatoire d prendra alors avec les probabilités :

$$p^2, 2pq, q^2, \text{ les valeurs } i-2l, j-s-l, k-2s.$$

On voit qu'il n'est possible de la rendre certainement nulle que si $j=(i+k)/2$, c'est-à-dire s'il n'y a pas dominance (il suffit alors de prendre $t=i/2$, $s=k/2$). Quand il y a dominance, on ne peut l'annuler complètement, on ne peut que se proposer de réduire au minimum son importance, par un choix convenable de s et l . R.-A. Fisher a proposé de choisir pour s et l des valeurs telles que :

$$M(d^2) = p^2(i-2l)^2 + 2pq(j-s-l) + q^2(k-2s)^2$$

soit minimum.

Nous allons voir que cette condition de moindres carrés (que nous désignerons par (C)) introduit dans les calculs des simplifications remarquables. Mais faisons auparavant quelques remarques sur la décomposition obtenue pour H .

On a :

$$H = \mathcal{H} + \mathcal{H}' + d$$

L'aléatoire \mathcal{H} représente la contribution d'un des gènes du couple ; elle est égale à t quand il est dans l'état A , à s quand il est dans l'état a , ces 2 cas ayant des probabilités respectives p et q ; l'aléatoire \mathcal{H}' contribution de l'autre gène obéit à la même loi ; nous avons vu que, du fait du « random mating », ces 2 gènes sont stochastiquement indépendants et, par suite, aussi les aléatoires \mathcal{H} et \mathcal{H}' . Quant à l'aléatoire d , elle ne dépend que des aléatoires \mathcal{H} et \mathcal{H}' (elle est connue quand \mathcal{H} et \mathcal{H}' sont connus).

Soit un individu apparenté à celui que nous considérons, et désignons la contribution qui lui correspond par :

$$H_1 = \mathcal{H}_1 + \mathcal{H}'_1 + d_1$$

La liaison entre $\mathcal{H}_1 + \mathcal{H}'_1$ et $\mathcal{H} + \mathcal{H}'$ (aléatoires que nous désignerons respectivement par \mathbf{H}_1 et \mathbf{H} et que nous appellerons « partie génétique » de la contribution) est précisément celle qui a été étudiée au chapitre précédent, mais pour en déduire la corrélation entre H et H_1 , il faut examiner le rôle du « résidu de dominance » d , ce qui résultera de l'étude que nous allons faire maintenant.

Détermination de l'ajustement par moindres carrés (C); propriété caractéristique

Pour exprimer que s et t sont choisis de manière que $M(d^2)$ soit minimum (condition (C)), nous écrivons que les dérivés partielles de cette expression par rapport à s et t sont nulles, ce qui donne :

$$(2) \quad \begin{cases} p^2(i-2t) + pq(j-s-t) = 0 \\ pq(j-s-t) + q^2(k-2s) = 0 \end{cases}$$

ou en représentant par $i=2t$, $j=s+t$, $k=2s$, les valeurs de la « partie génétique » \mathbf{H} :

$$(2') \quad \begin{cases} p^2(i-i) + pq(j-j) = 0 \\ pq(j-j) + q^2(k-k) = 0 \end{cases}$$

en ajoutant ces 2 équations, on voit que la moyenne de \mathbf{H} est égale à celle de H , c'est-à-dire à 0. On a donc :

$$M(\mathbf{H}) = M(\partial\mathcal{C}) + M(\partial\mathcal{C}') = 2(pt + qs) = 0$$

donc
$$M(\partial\mathcal{C}) = pt + qs = 0$$

En retranchant maintenant les 2 équations, on a :

$$p^2i - q^2k = p^2i - q^2k = 2(p^2i + pqj) = -2(pqj + q^2k)$$

[en tenant compte de la relation (1)].

s et t sont donc fournis par :

$$\begin{cases} pt + qs = 0 \\ p^2t - q^2s = p^2i + pqj = -pqj - q^2k \end{cases}$$

ce qui donne :

$$(2'') \quad \begin{cases} t = pi + qj \\ s = pj + qk \end{cases}$$

Nous voyons donc que les équations (2) fournissent pour s et t une solution unique.

Nous supposons dorénavant que \mathbf{H} et d sont ainsi choisis. On peut alors interpréter les conditions (2) de la façon suivante :

Le résidu d a une valeur moyenne nulle chaque fois qu'on fixe la valeur d'une des aléatoires $\mathcal{J}\mathcal{C}$ ou $\mathcal{J}\mathcal{C}'$: $M_t(d) = M_s(d) = 0$

Il en résulte immédiatement que :

$$M(\mathcal{J}\mathcal{C}d) = p t M_t(d) + q s M_s(d) = 0$$

$$M(\mathbf{H}d) = M(\mathcal{J}\mathcal{C}d) + M(\mathcal{J}\mathcal{C}'d) = 0$$

ce qui revient à dire que les coefficients de corrélation entre $\mathcal{J}\mathcal{C}$ et d , entre $\mathcal{J}\mathcal{C}'$ et d , entre \mathbf{H} et d , sont nuls.

Conséquences

1° Pour la variance

La variance de \mathbf{H} est la somme de la variance de la partie génétique \mathbf{H} et de la variance du résidu d . En effet :

$$M(H^2) = M[(\mathbf{H} + d)^2] = M(\mathbf{H}^2) + M(d^2) \quad \text{puisque} \quad M(\mathbf{H}d) = 0$$

Il nous sera utile de calculer ces 2 composantes de la variance. On a :

$$M(\mathbf{H}^2) = M(\mathcal{J}\mathcal{C}^2) + M(\mathcal{J}\mathcal{C}'^2) = 2(pt^2 + qs^2)$$

$$= 2(p/q)t^2 = (p^2i - q^2k)^2 / 2pq$$

$$(\text{puisque } -qs = pt = p^2i + pqj = -pqj - q^2k = (p^2i - q^2k) / 2).$$

Quant à d , ses valeurs $i - 2t$, $j - s - t$, $k - 2s$, sont, en remplaçant s et t par leurs valeurs (2''), et éliminant j à l'aide de (1) :

$$i - (p^2i - q^2k) / p \quad -pi - qk \quad k + (p^2i - q^2k) / q$$

c'est-à-dire :

$$(q/p)(pi + qk) \quad -(pi + qk) \quad (p/q)(pi + qk)$$

ce qui donne :

$$M(d^2) = (pi + qk)^2 (p^2q^2/p^2 + 2pq + q^2p^2/q^2) = (pi + qk)^2$$

Une autre forme de cette quantité est, comme on vérifie facilement : $M(d^2) = p^2q^2(i - 2j + k)^2$ (il suffit de remplacer $-2j$ par $(p^2i + q^2k) / pq$).

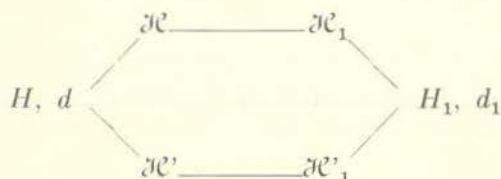
2° Pour les corrélations

La corrélation entre les contributions $H = \mathcal{H} + \mathcal{H}' + d$ et $H_1 = \mathcal{H}_1 + \mathcal{H}'_1 + d_1$ chez 2 individus apparentés est fournie par :

$$M(HH_1) = M(\mathbf{H}\mathbf{H}_1) + M(\mathbf{H}d_1) + M(d\mathbf{H}_1) + M(dd_1)$$

Le terme $M(\mathbf{H}\mathbf{H}_1)$ représente la corrélation « génétique » étudiée au chapitre précédent.

Nous allons voir que les termes $M(\mathbf{H}d_1)$ et $M(d\mathbf{H}_1)$ sont nuls si on considère toujours des couples où les conjoints ne sont pas apparentés ; les 2 parents de tout individu sont alors au plus apparentés séparément avec les 2 parents d'un autre (comme dans le cas des frères par exemple), mais sans l'être entre eux ; les 2 gènes de chaque individu n'ont respectivement de liaison stochastique qu'avec chacun des 2 gènes de l'autre. Par exemple il n'y a de liaison qu'entre \mathcal{H} et \mathcal{H}'_1 et entre \mathcal{H}' et \mathcal{H}_1 (\mathcal{H}_1 étant indépendant de \mathcal{H}' et \mathcal{H}'_1 , \mathcal{H}'_1 étant indépendant de \mathcal{H} et \mathcal{H}_1). Comme H (et d) ne dépendent que de \mathcal{H} et \mathcal{H}' , le schéma de liaison est le suivant :



d ne dépend de \mathcal{H}'_1 que par l'intermédiaire de \mathcal{H} .

Quand \mathcal{H} est fixé, d et \mathcal{H}'_1 deviennent des aléatoires indépendantes, donc la valeur moyenne de $d\mathcal{H}'_1$ est le produit des valeurs moyennes de \mathcal{H}'_1 et de d , cette dernière étant alors nulle. Donc $M(d\mathcal{H}'_1) = 0$ quand \mathcal{H} est fixé, et par suite aussi quel que soit H . De même $M(d\mathcal{H}'_1) = 0$, donc $M(d\mathbf{H}_1) = 0$; par suite aussi $M(d_1\mathbf{H}) = 0$.

On voit donc alors que :

$$M(HH_1) = M(\mathbf{H}\mathbf{H}_1) + M(dd_1)$$

relation parallèle à celle du 1° :

$$M(H^2) = M(\mathbf{H}^2) + M(d^2)$$

Ces équations relatives à la contribution d'un couple de gènes s'étendent à la somme de toutes les contributions, $x=H+K+\dots$, qui représente le caractère mesuré. Décomposons en effet ce caractère en une « partie génétique » $z=\Sigma\mathbf{H}$, et un « résidu de dominance » $D=\Sigma d$. On a alors :

$$x=\Sigma(\mathbf{H}+d)=z+D$$

Pour un individu apparenté : $x_1=\Sigma(\mathbf{H}_1+d_1)=z_1+D_1$.

Et on a, en vertu de l'indépendance stochastique des différents couples de gènes (qui entraîne $M(HK)=M(H)\times M(K)=0$ et $M(HK_1)=0$) :

$$M(x^2)=M(H^2)+M(K^2)+\dots=\Sigma M(\mathbf{H}^2)+\Sigma M(d^2)$$

$$\text{c'est-à-dire} \quad M(x^2)=M(z^2)+M(D^2) \quad (3)$$

$$\text{et de même} \quad M(xx_1)=M(zz_1)+M(DD_1)$$

ce qui s'écrit, en introduisant les coefficients de corrélation :

$$(4) \quad r(xx_1)M(x^2)=r(zz_1)M(z^2)+r(DD_1)M(D^2)$$

Ces 2 relations nous fournissent $r(xx_1)$, coefficient de corrélation entre les valeurs x et x_1 du caractère chez 2 individus quand on tient compte de la dominance, à condition de connaître :

1° le rapport $M(z^2)/M(x^2)=1-M(D^2)/M(x^2)$, quantité comprise entre 0 et 1, que nous désignerons par δ^2 , et qui caractérise le degré de dominance (elle est d'autant plus voisine de 1 que $M(D^2)$ est plus petit par rapport à $M(x^2)$, c'est-à-dire que la dominance est plus faible) ;

2° la corrélation $r(zz_1)$ entre les parties génétiques du caractère, qui a été étudiée au chapitre précédent pour les types usuels de parenté ;

3° la corrélation $r(DD_1)$ entre les résidus de dominance des 2 individus, que nous allons calculer maintenant, en distinguant 2 cas, suivant que les individus ne sont apparentés que par l'intermédiaire d'un de leurs parents, ou de leurs 2 parents.

1^{er} CAS :

**CORRÉLATIONS ENTRE DES INDIVIDUS
QUI NE SONT APPARENTÉS
QUE PAR L'INTERMÉDIAIRE
D'UN DES PARENTS DE L'UN**

C'est le cas, par exemple, de la corrélation avec un ancêtre, avec un oncle, avec un cousin ordinaire. La liaison se fait uniquement par un des parents de l'individu, ce qui revient à dire qu'un seul des gènes de cet individu est en liaison stochastique avec l'autre individu.

Valeur de $r(DD_1)$

Si on pose : $x = \Sigma(\mathcal{H} + \mathcal{H}' + d)$ pour le premier et $x_1 = \Sigma H_1 = \Sigma(\mathbf{H}_1 + d_1)$ pour le second, on peut supposer par exemple que \mathcal{H} est en corrélation avec H_1 et que \mathcal{H}' en est indépendant, donc d n'est lié à H_1 (et par suite à d_1) que par l'intermédiaire de \mathcal{H} . Or \mathcal{H} est une aléatoire du 2^e ordre, et par suite les régressions de d et d_1 par rapport à cette variable intermédiaire \mathcal{H} sont certainement linéaires, on peut donc appliquer le théorème de p. 34 et on a :

$$\begin{aligned} r(dd_1) &= r(d\mathcal{H}) \times r(\mathcal{H}d_1) = 0 \text{ puisque } r(d\mathcal{H}) = 0 \\ \text{donc} \quad M(dd_1) &= 0 \quad M(DD_1) = \Sigma M(dd_1) = 0 \\ & \quad r(DD_1) = 0 \end{aligned}$$

Nous obtenons ainsi un résultat important : *dans le cas où la liaison se fait par un seul parent, le coefficient de corrélation entre les résidus de dominance D et D₁ des 2 individus est nul.*

On en déduit alors le coefficient de corrélation $r(xx_1)$ entre les caractères.

$$(5) \quad r(xx_1) = r(zz_1)M(z^2) / M(x^2) = r(zz_1)\hat{z}^2$$

Une autre méthode

Signalons une autre manière de démontrer cette formule. Le schéma de liaison entre H et H_1 est le suivant :

$$H - \mathbf{H} - \mathcal{C} - \mathbf{H}_1 - H_1$$

(2 aléatoires de ce schéma ne sont liées stochastiquement que par l'intermédiaire de celles qui se trouvent en elles.)

Or il résulte du chapitre précédent que les régressions de H par rapport à \mathbf{H} et à \mathbf{H}_1 , qui correspondent à des corrélations génétiques, sont linéaires. On a donc, toujours en vertu du théorème de p. 34 :

$$r(H\mathcal{C}) = r(\mathbf{H}\mathbf{H}) \times r(\mathbf{H}\mathcal{C}) \quad r(\mathcal{C}H_1) = r(\mathcal{C}\mathbf{H}_1) \times r(\mathbf{H}_1H_1)$$

et comme les régressions de H et \mathbf{H} en \mathcal{C} sont linéaires (puisque \mathcal{C} est du 2^e ordre) :

$$\begin{aligned} r(HH_1) &= r(H\mathcal{C}) \times r(\mathcal{C}H_1) \\ &= r(\mathbf{H}\mathbf{H}) \times r(\mathbf{H}\mathcal{C}) \times r(\mathcal{C}\mathbf{H}_1) \times r(\mathbf{H}_1H_1) \\ &= r(\mathbf{H}\mathbf{H}) \times r(\mathbf{H}\mathbf{H}_1) \times r(\mathbf{H}_1H_1) \end{aligned}$$

Donc le coefficient de corrélation entre les 2 contributions H et H_1 est le produit du coefficient de corrélation entre les parties génétiques \mathbf{H} et \mathbf{H}_1 , par les coefficients de corrélation entre chaque contribution et sa partie génétique. On peut donc dire que la corrélation se fait par l'intermédiaire des contributions génétiques \mathbf{H} et \mathbf{H}_1 , mais que les contributions mesurées H et H_1 ne fournissent les contributions génétiques qu'avec une certaine erreur, ce qui a pour effet de diminuer la corrélation observée dans le rapport $r^2(H\mathbf{H})$. Ce rapport de réduction a pour expression :

$$r(HH_1) / r(\mathbf{H}\mathbf{H}_1) = [M(H\mathbf{H})]^2 / M(H^2)M(\mathbf{H}^2) = M(\mathbf{H}^2) / M(H^2)$$

On en déduit facilement, en passant à la somme de toutes les contributions, le rapport de réduction global :

$$r(xx_1) / r(zz_1) = \Sigma M(\mathbf{H}^2) / \Sigma M(H^2) = \zeta^2$$

On retrouve bien la formule (5).

Valeurs des corrélations usuelles

Indiquons maintenant les valeurs des corrélations usuelles que cette formule fournit, en tenant compte des résultats du chapitre précédent :

père-fils :	$(1/2)\zeta^2$
grand-père - petit-fils :	$(1/4)\zeta^2$
ancêtre d'ordre n :	$(1/2)^n\zeta^2$
demi-frères :	$(1/4)\zeta^2$
oncle-neveu :	$(1/4)\zeta^2$
cousins germains :	$(1/8)\zeta^2$
cousins quelconques :	$(1/2)^{n+p+1}\zeta^2$

Rappels bibliographiques ; comparaison des méthodes

La valeur de la corrélation ancestrale dans l'hypothèse mendélienne, $(1/2^n)\zeta^2$, a été fournie pour la première fois par K. Pearson en 1903 dans son mémoire des Philosophical Transactions (203 A, p. 53) dans un cas particulier, car il était parti d'hypothèses très restrictives ; il supposait qu'il y avait dominance complète, la contribution j d'un couple de gènes hétérozygotes étant égale à k par exemple. Il supposait aussi que les contributions des n couples de gènes avaient la même importance ($i=i'=\dots$; $k=k'=\dots$) et que les 3 phases de chacun se trouvaient dans les proportions : $p^2=(1/4)$, $2pq=1/2$, $q^2=1/4$. Il trouvait ainsi une corrélation parentale de $1/3$, et une corrélation ancestrale de $(1/3)\times 1/2^n$, car dans ces hypothèses le rapport de réduction ζ^2 devient égal à $2/3$. (1) fournit en effet $i=-3k$, et (2) donne alors $-l=s=k$; on voit alors que $M(\mathbf{H}^2)=2k^2$ et $M(H^2)=3k^2$.

Dans un mémoire ultérieur (R.-S. proceedings ; B 81, 1909, p. 225) traitant le cas de la corrélation ancestrale et complété par Snow (R.-S. proceedings, B 83, 1911, p. 37) pour le cas des frères, oncles, et cousins germains, Pearson a repris le calcul par une autre méthode ; il a remarqué qu'il suffisait de calculer les corrélations dans le cas où la variabilité dépend d'un seul couple de gènes, et il est parti de propor-

tions p^2 , $2pq$, q^2 des 3 phases ; sa méthode consiste à construire pour chaque degré de parenté, la table à 9 cases fournissant les probabilités d'avoir pour l'un et pour l'autre individu, 2 phases déterminées ; par exemple :

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>AA</i>			
<i>Aa</i>			
<i>aa</i>			

De ces tables, Pearson et Snow déduisent de la manière habituelle les coefficients de corrélation dans le cas de la dominance complète (c'est-à-dire quand il n'y a que deux valeurs de la contribution, l'une correspondant à *AA* et à *Aa* et l'autre à *aa*) et aussi dans le cas où il n'y a pas de dominance (c'est-à-dire quand la contribution de *Aa* est moyenne de celles de *AA* et *aa*).

Pour la corrélation ancestrale, Pearson trouve dans le 1^{er} cas $2q/(p+2q)2^n$ (nous pouvons facilement vérifier que $\xi^2=2q/(p+2q)$, en faisant $j=i$ dans (1) et (2)) et dans le 2^e cas $1/2^n$. Mais ces valeurs paraissent l'une trop faible et l'autre trop forte pour rendre compte des corrélations mesurées, aussi Yule a-t-il suggéré qu'on pourrait obtenir une valeur plus correcte en supposant la dominance incomplète. Ce calcul plus général a été fait par R.-A. Fisher dans l'ouvrage déjà mentionné (Ed., p. 399), toujours suivant la même méthode, mais en assignant aux contributions des 3 phases des valeurs quelconques i , j , k (de valeur moyenne nulle). Il a ainsi imaginé l'ajustement par moindres carrés (C), et a mis en évidence le rôle du rapport $\Sigma M(\mathbf{H}^2)/\Sigma M(H^2)$ qu'il appelle le

« taux de dominance » τ^2/σ^2 (« dominance ratio ») ; il a ainsi obtenu pour la corrélation parent-enfant la valeur $(1/2)\tau^2/\sigma^2$, et pour les corrélations entre oncle et neveu et entre cousins germains les valeurs $(1/4)\tau^2/\sigma^2$ et $(1/8)\tau^2/\sigma^2$, ce qui faisait pressentir la diminution de moitié de la corrélation pour chaque degré d'éloignement, et la similitude du rôle de la dominance avec celui d'une erreur accidentelle dans la mesure du caractère. Mais la méthode de Fisher, reposant toujours sur l'établissement de la table à 9 cases qui traduit l'association des 3 phases chez les 2 individus, n'est pas susceptible de généralisation. Les calculs sont laborieux et deviennent impraticables pour un degré de parenté un peu éloigné. Tandis que les méthodes que nous avons indiquées permettent de calculer par échelons les corrélations génétiques, et d'en déduire les corrélations réelles par la formule générale : $r(xx_1) = r(zz_1) \times \frac{1}{2}$ valable pour n'importe quelle parenté ayant lieu par l'intermédiaire d'un seul des parents d'un des individus ; et nous avons vu la raison profonde de cette formule : elle repose sur le fait qu'alors les résidus de dominance D et D_1 des 2 individus ont un coefficient de corrélation nul.

2^e CAS :

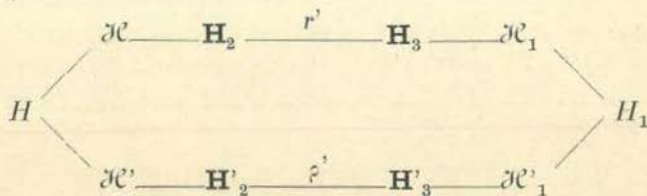
**CORRÉLATION ENTRE DEUX INDIVIDUS
DONT LES PARENTS
SONT RESPECTIVEMENT APPARENTÉS**

Données du problème

Supposons maintenant que chacun des parents d'un des individus soit apparenté à un des parents de l'autre, mais à un seul (ce qui exclut, comme nous l'avons fait remarquer, la parenté des conjoints entre eux, et par suite nous permet d'utiliser les résultats de p. 47).

Alors, si on désigne encore par $\mathcal{H} + \mathcal{H}'$ et $\mathcal{H}_1 + \mathcal{H}'_1$ les contributions « génétiques » d'un couple de gènes chez les 2 individus, il y a des liaisons stochastiques déterminées entre les gènes qui proviennent de parents apparentés, par exemple *une corrélation r entre \mathcal{H} et \mathcal{H}_1 et une corrélation ρ entre \mathcal{H}' et \mathcal{H}'_1 , tandis qu'il y a indépendance stochastique entre \mathcal{H} ou \mathcal{H}_1 , d'une part, et \mathcal{H}' ou \mathcal{H}'_1 , d'autre part.*

Indiquons dès maintenant comment se déterminent r et ρ . Désignons par \mathbf{H}_2 et \mathbf{H}'_2 la contribution génétique du couple de gènes considéré chez les 2 parents du premier individu, et par \mathbf{H}_3 et \mathbf{H}'_3 celle chez les 2 parents du second (\mathbf{H}_2 correspondant au parent qui fournit le gène \mathcal{H} , \mathbf{H}'_2 à celui qui fournit le gène \mathcal{H}' , etc.). La connaissance des degrés de parenté de ces parents nous fournit, d'après le chapitre II, leur corrélation génétique, que nous désignons par r' pour la liaison $\mathbf{H}_2 - \mathbf{H}_3$ et par ρ' pour la liaison $\mathbf{H}'_2 - \mathbf{H}'_3$. Pour plus de clarté, représentons le schéma de liaison entre les 2 individus :



On voit que la liaison entre \mathcal{H} et \mathcal{H}_1 s'obtiendra en composant les régressions linéaires (qui résultent du chapitre II) :

$$M(\mathbf{H}_2) = \mathcal{H} \quad \text{quand } \mathcal{H} \text{ est fixé.}$$

$$M(\mathbf{H}_3) = r' \mathbf{H}_2 \quad \text{quand } \mathbf{H}_2 \text{ est fixé.}$$

$$M(\mathcal{H}_1) = \mathbf{H}_3 / 2 \quad \text{quand } \mathbf{H}_3 \text{ est fixé.}$$

Donc, d'après le théorème de p. 38 : $M(\mathcal{H}_1) = (r' / 2) \mathcal{H}$ quand \mathcal{H} est fixé.

$$\text{Donc} \quad r = r' / 2 \quad \text{et} \quad \rho = \rho' / 2.$$

Exemples :

Comme exemples de ce mode de parenté, nous avons déjà cité le cas des frères (\mathbf{H}_2 coïncide alors avec \mathbf{H}_3 , et \mathbf{H}'_2 avec \mathbf{H}'_3 , donc $r' = \rho' = 1$, $r = \rho = 1 / 2$) et le cas des doubles cousins germains (\mathbf{H}_2 et \mathbf{H}_3 sont frères, ainsi que \mathbf{H}'_2 et \mathbf{H}'_3 , donc $r' = \rho' = 1 / 2$, $r = \rho = 1 / 4$).

Ces 2 cas ont été étudiés par Fisher (Ed.) toujours par la méthode de la table d'association, et il a reconnu que dans ces cas la réduction à faire subir à la corrélation génétique pour obtenir la corrélation réelle est moins élevée que dans ceux du paragraphe précédent. Cela tient, comme nous allons le voir, à ce que la corrélation $r(DD_1)$ des résidus de dominance des 2 individus est maintenant positive, et non plus nulle comme dans le paragraphe précédent ; or la corrélation réelle $r(xx_1)$ est liée à $r(DD_1)$ par la relation (4), qui peut s'écrire :

$$(4') \quad r(xx_1) = \lambda^2 r(zz_1) + (1 - \lambda^2) r(DD_1)$$

Calcul des probabilités d'association des gènes homologues

Nous allons donc calculer, sous les hypothèses précisées ci-dessus (corrélations r et ρ entre les gènes des 2 individus) la corrélation $r(xx_1)$, et donner ainsi l'expression générale des résultats obtenus par Fisher dans le cas particulier des frères et des doubles cousins.

$$\text{On a :} \quad r(xx_1) = M(xx_1) / M(x^2) = \Sigma M(HH_1) / M(x^2)$$

et nous déterminerons $M(HH_1)$ à l'aide des probabilités d'association des gènes \mathcal{H} et \mathcal{H}_1 , \mathcal{H}' et \mathcal{H}'_1 , que nous allons calculer en premier lieu. Désignons-les par :

$$\begin{aligned} p_{tt} &= \text{probabilité pour que } \mathcal{H} = t \text{ et } \mathcal{H}_1 = t \\ p_{ts} &= \text{'' '' } \mathcal{H} = t \text{ et } \mathcal{H}_1 = s \\ p_{st} &= \text{'' '' } \mathcal{H} = s \text{ et } \mathcal{H}_1 = t \\ p_{ss} &= \text{'' '' } \mathcal{H} = s \text{ et } \mathcal{H}_1 = s \end{aligned}$$

et de même :

$$p'_{tt} = \text{'' '' } \mathcal{H}' = t \text{ et } \mathcal{H}'_1 = t, \text{ etc.}$$

Les 4 probabilités d'associations entre \mathcal{H} et \mathcal{H}_1 peuvent se calculer en connaissant les répartitions marginales de \mathcal{H} et \mathcal{H}_1 (qui sont constituées ainsi : valeur t avec la probabilité p , et valeur s avec la probabilité q) et le coefficient de corrélation entre \mathcal{H} et \mathcal{H}_1 , que nous avons désigné par r . On a en effet :

$$\begin{aligned} p_{tt} + p_{ts} &= p \\ p_{st} + p_{ss} &= q \\ p_{tt} + p_{st} &= p \\ p_{ts} + p_{ss} &= q \\ r &= t^2 p_{tt} + st(p_{st} + p_{ts}) + s^2 p_{ss} / (pt^2 + qs^2) \end{aligned}$$

(en tenant compte de ce que $M(\mathcal{H}) = pt + qs = 0$ d'après la p. 45).

Les 4 premières équations donnent :

$$p_{st} = p_{ts} \quad p_{tt} = p - p_{st} \quad p_{ss} = q - p_{st}$$

et la 5^e peut par suite s'écrire :

$$r = pt^2 + qs^2 - p_{st}(t^2 - 2st + s^2) / (pt^2 + qs^2)$$

d'où :

$$p_{ts} = p_{st} = (1-r)(pt^2 + qs^2) / (t-s)^2 = (1-r)pq$$

(en remplaçant s par $-pt/q$),

et par suite :

$$\begin{aligned} p_{tt} &= p - (1-r)pq = p(p + rq) \\ p_{ss} &= q - (1-r)pq = q(q + rp) \end{aligned}$$

Et on aura de même, pour la liaison entre $\mathcal{H}\mathcal{C}'$ et $\mathcal{H}\mathcal{C}'_1$, les probabilités marginales restant toujours p et q :

$$\begin{aligned} p'_{ts} &= p'_{st} = (1-\varphi)pq \\ p'_{tt} &= p(p+\varphi q) \\ p'_{ss} &= q(q+\varphi p) \end{aligned}$$

Forme symbolique de $M(HH_1)$

Et maintenant la connaissance de ces probabilités va nous fournir $M(HH_1)$.

En effet : pour que $H=k$ et $H_1=k_1$ (k_1 étant égal à k , mais il nous sera utile d'affecter de l'indice 1 les contributions qui se rapportent à H_1),

$$\text{il faut que : } \begin{cases} \mathcal{H} = s \\ \mathcal{H}' = s \end{cases} \quad \text{et} \quad \begin{cases} \mathcal{H}\mathcal{C}'_1 = s_1 \\ \mathcal{H}\mathcal{C}' = s_1 \end{cases}$$

la probabilité de ce cas est donc : $p_{ss}p'_{ss}$, ce qui fournit dans $M(HH_1)$ le terme : $p_{ss}p'_{ss}kk_1$.

Pour que $H=k$ et $H_1=j_1$, il faut que :

$$\begin{cases} \mathcal{H} = s \\ \mathcal{H}' = s \end{cases} \quad \text{avec} \quad \begin{cases} \mathcal{H}\mathcal{C}'_1 = s_1 \\ \mathcal{H}\mathcal{C}' = l_1 \end{cases} \quad \text{ou} \quad \begin{cases} \mathcal{H}\mathcal{C}'_1 = l_1 \\ \mathcal{H}\mathcal{C}' = s_1 \end{cases}$$

ce qui fournit dans $M(HH_1)$ les 2 termes $(p_{ss}p'_{st} + p_{st}p'_{ss})kj_1$.

Nous voyons ainsi que tous les termes de $M(HH_1)$ pourront s'obtenir par le calcul symbolique suivant :

Aux différentes associations possibles :

$$\begin{cases} \mathcal{H} = s \\ \mathcal{H}\mathcal{C}'_1 = s_1 \end{cases} \quad \begin{cases} \mathcal{H} = s \\ \mathcal{H}\mathcal{C}'_1 = l_1 \end{cases} \quad \begin{cases} \mathcal{H} = l \\ \mathcal{H}\mathcal{C}'_1 = s_1 \end{cases} \quad \begin{cases} \mathcal{H} = l \\ \mathcal{H}\mathcal{C}'_1 = l_1 \end{cases}$$

nous ferons correspondre respectivement les quantités (I) : $p_{ss}ss_1$; $p_{st}st_1$; $p_{ts}ts_1$; $p_{ll}ll_1$.

De même, aux associations $\begin{cases} \mathcal{H}' = s \\ \mathcal{H}\mathcal{C}'_1 = s_1 \end{cases}$, etc.

nous ferons correspondre les quantités (I') : $p'_{ss}ss_1$, etc.

En multipliant une des quantités (I) par une des quantités (I'), on obtient des termes (II) tels que par exemple : $p_{st}p'_{ss}st_1ss_1$, qui correspond à la probabilité pour que $\mathcal{H} = s$,

$\mathcal{C}_1 = t_1$, $\mathcal{C}' = s$, $\mathcal{C}'_1 = s_1$. Il suffit d'y remplacer s^2 par k et $s_1 t_1$ par j_1 pour avoir le terme de $M(HH_1)$ qui correspond à ces associations.

D'où la méthode suivante pour former $M(HH_1)$: on effectue le produit de la somme des quantités (I) par la somme des quantités (I'), étant entendu que l'on doit remplacer dans le résultat :

$$s^2, st, t^2, s^2_1, s_1 t_1, t^2_1 \text{ par } k, j, i, k_1, j_1, i_1,$$

puis faire $k_1 = k$, $j_1 = j$, $i_1 = i$.

$M(HH_1)$ est donc représentée par le produit symbolique :

$$(6) (ss_1 p_{ss} + st_1 p_{st} + ts_1 p_{ts} + tt_1 p_{tt})(ss_1 p'_{ss} + st_1 p'_{st} + ts_1 p'_{ts} + tt_1 p'_{tt})$$

Calcul de $M(HH_1)$

Nous n'avons plus qu'à remplacer dans cette formule p_{ss} , p_{st} , etc., par leurs expressions obtenues à la page précédente. Le premier facteur de (6) devient alors une fonction linéaire de r :

$$q^2 ss_1 + pqt_1 + pqls_1 + p^2 tt_1 + rpq(ss_1 - st_1 - ts_1 + tt_1)$$

c'est-à-dire :

$$(pt + qs)(pt_1 + qs_1) + rpq(s - t)(s_1 - t_1)$$

Le 2^e facteur peut s'écrire de même :

$$(pt + qs)(pt_1 + qs_1) + r'pq(s - t)(s_1 - t_1)$$

et le produit (6) devient alors :

$$(6a) \frac{(pt + qs)^2 (pt_1 + qs_1)^2 + pq(r + r')(pt + qs)(s - t)(pt_1 + qs_1)(s_1 - t_1) + p^2 q^2 r r' (s - t)^2 (s_1 - t_1)^2}{}$$

Pour en déduire $M(HH_1)$, il faut dans chaque terme effectuer les produits et faire les remplacements indiqués plus haut, ce qui donne, une fois supprimés les indices 1 :

$$(6b) M(HH_1) = (p^2 i + 2pqj + q^2 k)^2 + pq(r + r')[qk + (p - q)j - pi]^2 + p^2 q^2 r r' (k - 2j + i)^2$$

Le 1^{er} terme est nul du fait du choix de la moyenne de H pour origine (condition (1)). $M(HH_1)$ doit bien alors s'annuler quand $r=\varphi=0$, c'est-à-dire quand les 2 individus sont indépendants.

Le coefficient de $pq(r+\varphi)$ s'écrit (toujours d'après (1)) :

$$\begin{aligned} [(q^2k+pqj)/q - (p^2i+pqj)/p]^2 &= [(q^2k-p^2i)(1/q+1/p)/2]^2 \\ &= (q^2k-p^2i)^2/2pq \end{aligned}$$

c'est-à-dire d'après le résultat de p. 46 : $M(\mathbf{H}^2)/2pq$.

Le coefficient de $r\varphi$ n'est autre que $M(d^2)$ (voir p. 46). Cela résulte d'ailleurs immédiatement de ce que, si H_1 coïncide avec H , c'est-à-dire si $r=\varphi=1$, l'équation (6b) fournit pour ce coefficient la valeur $M(H^2)-M(\mathbf{H}^2)=M(d^2)$.

On a donc : $M(HH_1)=M(\mathbf{H}^2)(r+\varphi)/2+r\varphi M(d^2)$.

Formule générale de la corrélation

D'où en faisant la somme pour tous les couples de gènes (il résulte du chapitre II que les valeurs de $r=r'/2$ et $\varphi=\varphi'/2$ sont les mêmes pour tous les couples de gènes, elles sont déterminées dès qu'on connaît le mode de parenté) :

$$M(xx_1)=\Sigma M(HH_1)=M(z^2)(r+\varphi)/2+r\varphi M(D^2)$$

par suite en divisant par $M(x^2)$:

$$(7) \quad r(xx_1)=z^2(r+\varphi)/2+r\varphi(1-z^2)$$

La quantité $(r+\varphi)/2=(r'+\varphi')/4$ n'est autre (p. 41) que la corrélation génétique $r(\mathbf{HH}_1)$ ou $r(zz_1)$. On retrouve bien que $r(xx_1)$ se réduit à $r(zz_1)z^2$ quand l'une des quantités r ou φ est nulle.

En comparant avec (4'), on voit que :

$$r(DD_1)=r\varphi$$

Il y a donc, dans le cas général, une corrélation positive entre les résidus de dominance des 2 individus. Elle n'est nulle que quand la parenté se fait par l'intermédiaire d'un seul parent.

Mais elle est très faible dès qu'une des 2 liaisons entre parents est un peu éloignée.

Exemples :

C'est dans le cas des *frères* qu'elle est la plus importante. On a alors (puisque les parents coïncident) $r' = \rho' = 1$, donc $r = \rho = 1/2$, $r(DD_1) = 1/4$.

$$r(xx_1) = (1/2)\delta^2 + (1/4)(1 - \delta^2) = (1 + \delta^2)/4 = 1/2 - (1/4)(1 - \delta^2)$$

C'est bien le résultat trouvé par Fisher, p. 406 ($\delta^2 = \tau^2/\sigma^2$, $1 - \delta^2 = \varepsilon^2/\sigma^2$). La corrélation entre frères est donc supérieure à la corrélation père-fils.

La valeur génétique est la même, $1/2$, mais la réduction due à la dominance est moitié moindre dans le cas des frères ($(1 - \delta^2)/4$).

Pour des « *doubles cousins* », dont les parents sont frères, on a :

$$r' = \rho' = 1/2, \quad r = \rho = 1/4$$

donc :

$$r(xx_1) = (1/4)\delta^2 + (1/16)(1 - \delta^2) = (1 + 3\delta^2)/16 = 1/4 - (3/16)(1 - \delta^2)$$

La corrélation entre doubles cousins est donc un peu supérieure à la corrélation oncle-neveu $(1/4)\delta^2$, à cause du terme en $r(DD_1)$. Mais ce terme est déjà faible.

Cette formule a été donnée par Fisher (p. 407). Comme Snow, il a fait le calcul en formant la table d'association des phases, ce qui ne permet pas de dégager la loi générale, chaque cas particulier nécessitant un traitement spécial, qui devient d'ailleurs inextricable pour une parenté un peu éloignée. Nous voyons maintenant que le problème général est résolu par la formule (7), pour n'importe quel mode de parenté, tant qu'on ne prélève que des conjoints sans parenté systématique.

Comme autre application de cette formule, considérons des enfants ayant le même père, mais dont les mères sont sœurs.

On a : $r' = 1$ et $\rho' = 1/2$, donc $r(xx_1) = (3/8)\delta^2 + (1/8)(1 - \delta^2)$.

CHAPITRE IV

CONSÉQUENCES DE LA NON-ADDITIVITÉ DES GÈNES DANS LE CAS DU "RANDOM MATING"

Comme nous l'avons indiqué au chapitre précédent, il est nécessaire, pour étudier les corrélations de caractères conditionnés par les lois de Mendel, de tenir compte du phénomène de dominance, c'est-à-dire de non-additivité des effets des 2 gènes d'un couple. Il est naturel de songer à une généralisation de cette hypothèse, et d'étudier le cas où les effets des différents couples de gènes ne sont eux-mêmes pas additifs. C'est ce que Fisher a appelé l'hypothèse « d'épistacy ». Nous allons alors examiner la généralisation de l'ajustement par moindres carrés du chapitre précédent.

Solution de la condition de moindres carrés dans le cas de 2 couples

Considérons par exemple 2 couples de gènes, symbolisés par les aléatoires H et K , H prenant toujours les valeurs désignées par i, j, k , avec les probabilités $p^2, 2pq, q^2$, K les valeurs i', j', k' , avec les probabilités $p'^2, 2p'q', q'^2$. L'effet résultant du concours de ces 2 couples prendra, si les effets partiels ne s'ajoutent pas, au lieu des valeurs :

$$i+i' \quad i+j' \quad i+k' \quad j+i' \quad j+j' \quad j+k' \quad k+i' \quad k+j' \quad k+k'$$

les valeurs quelconques :

$$a_{11} \quad a_{12} \quad a_{13} \quad a_{21} \quad a_{22} \quad a_{23} \quad a_{31} \quad a_{32} \quad a_{33}$$

avec les probabilités :

$$p^2 p'^2 \quad 2p^2 p'q' \quad p^2 q'^2 \quad 2pqp'^2 \quad 4pqp'q' \quad 2pqq'^2 \quad q^2 p'^2 \quad 2q^2 p'q' \quad q^2 q'^2$$

(car la répartition indépendante des divers gènes résulte

seulement du random mating, et n'est pas affectée par la non-additivité.)

Nous procéderons alors comme pour la dominance. Nous représenterons l'effet résultant, désigné par A , par la somme d'effets partiels fictifs H et K , corrigée par un résidu E :

$$A = H + K + E$$

Les valeurs de E seront alors : $a_{11} - i - i'$, $a_{12} - i - j'$, etc. ; si les valeurs a_{ij} de l'effet global sont quelconques, on ne peut pas déterminer i , j , k , i' , j' , k' , de façon que les valeurs de E soient nulles ; mais on peut toujours les déterminer de façon que :

$$\begin{aligned} M(E^2) = & p^2 p'^2 (a_{11} - i - i')^2 + 2p^2 p' q' (a_{12} - i - j')^2 \\ & + p^2 q'^2 (a_{13} - i - k')^2 + 2p q p'^2 (a_{21} - j - i')^2 + \dots \\ & + \dots \end{aligned}$$

soit minimum (condition (C)).

En effet nous traduisons cette condition en exprimant que les dérivées partielles % à i , j , k , i' , j' , k' , sont nulles, ce qui donne le système :

$$(2) \left\{ \begin{array}{l} p^2 [p'^2 (a_{11} - i - i') + 2p q' (a_{12} - i - j') + q'^2 (a_{13} - i - k')] = 0 \\ 2p q [p'^2 (a_{21} - j - i') + 2p' q' (a_{22} - j - j') + \dots] = 0 \\ q^2 [p'^2 (a_{31} - k - i') + \dots] = 0 \\ p'^2 [p^2 (a_{11} - i - i') + 2p q (a_{21} - j - i') + \dots] = 0 \\ 2p' q' [\quad \quad \quad] = 0 \\ q'^2 [\quad \quad \quad] = 0 \end{array} \right.$$

Nous avons là, pour déterminer les 6 inconnues, 6 équations, mais elles ne sont pas indépendantes. En effet, si on fait la somme des 3 premières, ou la somme des 3 dernières, on obtient 2 fois la même équation, qui est :

$$p^2 i + 2p q j + q^2 k + p'^2 i' + 2p' q' j' + q'^2 k' = p^2 p'^2 a_{11} + 2p^2 p' q' a_{12} + \dots$$

et qui exprime que : $M(H) + M(K) = M(A)$, c'est-à-dire que $M(E) = 0$.

Nous pouvons supposer toujours que *l'origine des mesures de l'effet observé A a été choisie de telle manière que sa valeur moyenne soit nulle* : $M(A)=0$. Il en résulte alors que :

$M(H)+M(K)=0$, mais nous ne pouvons en déduire que $M(H)$ et $M(K)$ soient chacune nulle ; seule leur somme doit être nulle. Cela tient à ce que, dans chaque terme de l'expression à rendre minimum figure la somme d'une valeur de H et d'une valeur de K , somme qui ne change pas si on modifie simultanément les origines de repérage de H et de K , de façon à augmenter d'une certaine quantité les valeurs de l'une, et à diminuer de la même quantité les valeurs de l'autre. Nous décelons ainsi la nature de l'indétermination qui intervient dans la solution du problème. Elle réside dans un déplacement arbitraire de l'origine de H , compensé dans la somme par un déplacement opposé de l'origine de K . Nous pouvons alors nous débarrasser de ce paramètre arbitraire sans intérêt, en choisissant les 2 origines de façon que :

$$(1) \quad \begin{cases} M(H)=p^2i+2pqj+q^2k=0 \\ M(K)=p'^2i'+2p'q'j'+q'^2k'=0 \end{cases}$$

Dès lors le système (2) se résout immédiatement. Grâce aux conditions (1), chaque équation fournit une des inconnues :

$$(2'') \quad \begin{cases} i=p'^2a_{11}+2p'q'a_{12}+q'^2a_{13} \\ j=p'^2a_{21}+2p'q'a_{22}+q'^2a_{23} \\ k=p'^2a_{31}+\dots \\ i'=p^2a_{11}+2pqa_{21}+q^2a_{31} \\ j'=p^2a_{12}+\dots \\ k'=p^2a_{13}+\dots \end{cases}$$

Remarquons inversement que, si on connaît les premiers membres de (2'') on peut choisir arbitrairement a_{11} , a_{13} , a_{31} , a_{33} , et les 5 autres valeurs a sont déterminées. L'effet de la non-additivité peut donc être caractérisé, une fois connues les contributions additives, par les résidus relatifs aux 4 individus doublement homozygotes.

Généralisation pour plus de 2 couples

Le calcul précédent s'étend facilement pour l'ajustement de contributions additives H, K, L , etc., à l'effet global d'un nombre quelconque de couples. Plus généralement encore, considérons un effet global G résultant du concours d'un certain nombre d'événements aléatoires, supposés d'ordres finis, et *indépendants*; supposons les au nombre de 3 pour fixer les idées, et symbolisons-les par les aléatoires :

H , prenant les valeurs h_r , avec les probabilités :

$$p_r \quad (r=1,2,\dots, l)(\sum p_r=1)$$

H' , prenant les valeurs h'_s , avec les probabilités :

$$p'_s \quad (s=1,2,\dots, m)(\sum p'_s=1)$$

H'' , prenant les valeurs h''_t , avec les probabilités :

$$p''_t \quad (t=1,2,\dots, n)(\sum p''_t=1)$$

Alors que $H+H'+H''$ prend des valeurs telles que : $h_r+h'_s+h''_t$, nous supposons que l'effet global G prend des valeurs quelconques G_{rst} , mais avec les mêmes probabilités $p_r p'_s p''_t$. Si on pose encore :

$$G=H+H'+H''+E$$

nous nous imposerons toujours la condition :

$$(C) \quad M(E^2) = \sum_{rst} p_r p'_s p''_t (G_{rst} - h_r - h'_s - h''_t)^2 \text{ minimum}$$

condition qui détermine de façon unique (aux origines de repérage près) les h, h', h'' , car elle se traduit par le système obtenu en annulant les $l+m+n$ dérivées partielles :

$$(2b) \quad \left\{ \begin{array}{l} \sum_{st} p_r p'_s p''_t (G_{rst} - h_r - h'_s - h''_t) = 0 \text{ pour } r=1,2,\dots, l \\ \sum_{rt} p_r p'_s p''_t (G_{rst} - h_r - h'_s - h''_t) = 0 \text{ pour } s=1,2,\dots, m \\ \sum_{rs} p_r p'_s p''_t (G_{rst} - h_r - h'_s - h''_t) = 0 \text{ pour } t=1,2,\dots, n \end{array} \right.$$

ces équations ne sont pas toutes indépendantes, car la somme

des l premières, celle des m suivantes, et celle des n dernières, nous donnent toutes 3 le même résultat :

$$\Sigma p_r h_r + \Sigma p'_s h'_s + \Sigma p''_t h''_t = \Sigma_{rst} p_r p'_s p''_t G_{rst}$$

ce qui exprime que la moyenne de l'effet global observé G (que l'on peut supposer nulle, ce que nous ferons dorénavant), est la somme des moyennes des effets fictifs partiels, que nous pouvons supposer dorénavant nulles, en choisissant convenablement le repérage de chacun de ceux-ci :

$$\Sigma p_r h_r = \Sigma p'_s h'_s = \Sigma p''_t h''_t = 0$$

Dès lors le système (2b) se résout immédiatement :

$$(2''b) \quad \left\{ \begin{array}{l} h_r = \Sigma_{st} p'_s p''_t G_{rst} \\ h'_s = \Sigma_{rt} p_r p''_t G_{rst} \\ h''_t = \Sigma_{rs} p_r p'_s G_{rst} \end{array} \right.$$

Propriété caractéristique de l'ajustement (C); analyse de la variance et des corrélations

Les équations (2) ou (2b), ont une interprétation immédiate : elles expriment que *la valeur moyenne du résidu E est nulle chaque fois qu'on fixe la valeur d'une des aléatoires composantes, H, H' ou H''*.

(Il en résulte en particulier que la moyenne générale de E est nulle, ce qui met une fois de plus en évidence la coïncidence entre la moyenne de l'effet observé G et celle de l'effet ajusté $H+H'+H''$).

D'où les conséquences suivantes :

$$1^{\circ} \text{ On a } M(EH) = 0 \quad M(EH') = 0 \quad M(EH'') = 0$$

Donc, en désignant par $\mathbf{G} = H + H' + H''$ l'effet ajusté :

$$G = \mathbf{G} + E \quad \text{avec} \quad M(\mathbf{G}E) = 0$$

D'où :

$$(3) \quad \begin{aligned} M(G^2) &= M(\mathbf{G}^2) + M(E^2) = \\ &M(H^2) + M(H'^2) + M(H''^2) + M(E^2) \end{aligned}$$

Comme dans le cas particulier du chapitre précédent, les parties additives et le résidu ont un coefficient de corrélation nul, et leurs variances s'ajoutent.

2° Si on considère un autre effet global, G_1 , dont les effets composants H_1, H'_1, H''_1 , ne sont en liaison stochastique qu'avec H, H', H'' , respectivement (ce qui est le cas si H et H_1, H' et H'_1 , etc., sont des couples de facteurs homologues chez 2 individus apparentés, dans le cas du « random mating »), on a :

$$(4) \quad \begin{aligned} M(GG_1) &= M[(H+H'+H''+E)(H_1+H'_1+H''_1+E_1)] \\ &= M(HH_1) + M(H'H'_1) + M(H''H''_1) + M(EE_1) \end{aligned}$$

Car par hypothèse, les termes tels que :

$$M(HH'_1) = M(H) \times M(H'_1)$$

sont nuls ; d'autre part on voit, en reprenant le raisonnement de la p. 47, que les termes tels que $M(HE_1)$ sont nuls aussi :

E_1 ne dépend que de H_1, H'_1, H''_1 , et par suite ne dépend de H que par l'intermédiaire de H_1 . Quand H_1 est fixé, H et E_1 deviennent indépendants, donc alors $M(HE_1) = M(H) \times M(E_1)$ et le 2^e facteur est alors nul par suite de l'ajustement (C) ; en étendant la moyenne pour toutes les valeurs de H_1 , on a bien encore $M(HE_1) = 0$.

Application à la non-additivité des effets de 2 couples de gènes

En revenant au cas du 1^{er} paragraphe, et considérant, chez 2 individus apparentés, la contribution de 2 couples de gènes, soit :

$$A = H + K + E \quad \text{et} \quad A_1 = H_1 + K_1 + E_1$$

A et A_1 sont 2 aléatoires obéissant à la même loi ; leur variance se décompose en :

$$(3') \quad M(A^2) = M(H^2) + M(K^2) + M(E^2)$$

et leur liaison se traduit par :

$$(4') \quad M(A^2)r_{(AA_1)} = M(AA_1) = M(H^2)r_{(HH_1)} + M(K^2)r_{(KK_1)} \\ + M(E^2)r_{(EE_1)}$$

Les 2 premiers termes, qui correspondent à la liaison entre couples de gènes homologues, ont été étudiés au chapitre précédent ; si on les décompose en la partie génétique et la partie due à la dominance, on a :

$$(4'') \quad M(A^2)r_{(AA_1)} = \Sigma M(\mathbf{H}^2)r_{(\mathbf{EE}_1)} + \Sigma M(d^2)r_{(dd_1)} + M(E^2)r_{(EE_1)} \\ [r(\mathbf{HH}_1) = (r + \rho)/2, \quad r(dd_1) = r\rho]$$

et parallèlement :

$$(3'') \quad M(A^2) = \Sigma M(\mathbf{H}^2) + \Sigma M(d^2) + M(E^2)$$

les Σ étant étendues aux divers couples de gènes considérés, soit 2 dans notre cas.

Nous mettons ainsi en évidence, dans la variance et la corrélation, les termes qui traduisent les effets partiels indépendants des gènes, ceux qui traduisent l'effet de dominance, et celui qui traduit « l'épistacy ». Nous aurions pu isoler d'emblée le premier groupe de termes, en ajustant aux effets globaux a_{11} , a_{12} , a_{13} , etc., une somme de contributions additives des divers gènes : $2l + 2l'$, $2l + s' + l'$, $2l + 2s'$, etc. ; mais l'avantage qu'il y a à passer par l'intermédiaire de contributions $i + i'$, $i + j'$, $i + k'$, etc., des divers couples de gènes, est que le résidu de non-additivité se trouve ainsi décomposé en 2 parties, celle qui provient de la dominance et est déjà connue, et celle qui traduit « l'épistacy » et que nous allons déterminer maintenant.

Forme symbolique de $M(AA_1)$

Il nous suffira pour cela de poursuivre le calcul symbolique du chapitre précédent. Reportons-nous à l'expression (6a) (p. 58) dans laquelle nous supposons que :

$$s^2, \quad sl, \quad l^2, \quad s^2_1, \quad s_1l_1, \quad l^2_1$$

sont remplacés par :

$$\bar{k}, \bar{j}, \bar{i}, \bar{k}_1, \bar{j}_1, \bar{i}_1$$

(Nous désignons maintenant par des caractères surlignés ces quantités, qui n'ont qu'une signification symbolique, pour les distinguer des effets ajustés k, j, i , etc., déterminés au début du présent chapitre.)

Nous avons vu que dans cette expression le coefficient de $\bar{k}\bar{j}_1$, représente la probabilité pour que $H=k$ et $H_1=j_1$. Si nous considérons l'expression analogue relative au couple de gènes K , expression que nous écrivons immédiatement, en accentuant simplement toutes les lettres de (6a), et que nous désignons par (6'a), le coefficient dans cette expression de $\bar{j}'\bar{i}'_1$ par exemple représente de même la probabilité pour que $K=j'$ et $K_1=i'_1$. Donc, par suite de l'indépendance des divers couples, la probabilité pour qu'on ait à la fois $H=k, K=j', H_1=j_1, K_1=i'_1$ (c'est-à-dire qu'on ait $A=a_{32}$ et $A_1=a_{21}$) est le produit des 2 précédentes, c'est-à-dire le coefficient de $\bar{k}\bar{j}_1\bar{j}'\bar{i}'_1$ dans le développement du produit de (6a) par (6'a) ; si on multiplie ce coefficient par $a_{32}a_{21}$, on obtient un des termes qui constituent $M(AA_1)$.

Nous voyons maintenant comment déterminer $M(AA_1)$, et par suite l'expression (4') (dont d'ailleurs nous est inconnu seulement le dernier terme $M(EA_1)$): nous formerons le produit de (6a) par (6'a), et nous y remplacerons :

$$(5) \quad \left\{ \begin{array}{ll} t^2 t'^2 = \bar{i}\bar{i}' & \text{et } t^2_1 t'^2_1 = \bar{i}_1\bar{i}'_1 \quad \text{par } a_{11} \\ t^2 s' t' = \bar{i}\bar{j}' & \text{et } t^2_1 s'_1 t'_1 = \bar{i}_1\bar{j}'_1 \quad \text{par } a_{12} \\ t^2 s'^2 = \bar{i}\bar{k}' & \text{et } t^2_1 s'^2_1 = \bar{i}_1\bar{k}'_1 \quad \text{par } a_{13} \\ s t t'^2 = \bar{j}\bar{i}' & \text{et } s_1 t_1 t'^2_1 = \bar{j}_1\bar{i}'_1 \quad \text{par } a_{21} \end{array} \right.$$

Calcul de $M(AA_1)$ et $M(EA_1)$

Ecrivons donc ce produit, en y supprimant les indices 1, puisque les remplacements à effectuer sont les mêmes sur les lettres sans indice et sur les lettres avec indice 1 :

$$\begin{aligned}
& [(pt+qs)^2(p't'+q's')^2]^2 + (r+\rho) \{ pq[(pt+qs)(s-t)(p't'+q's')^2]^2 \\
& \quad + p'q'[(p't'+q's')(s'-t')(pt+qs)^2]^2 \} \\
(7) \quad & + r\rho \{ p^2q^2[(s-t)^2(p't'+q's')^2]^2 + p'^2q'^2[(s'-t')^2(pt+qs)^2]^2 \} \\
& + (r+\rho)^2 ppq'q'[(pt+qs)(s-t)(p't'+q's')(s'-t')]^2 \\
& + (r+\rho)r\rho ppq'q' \{ pq[(s-t)^2(s'-t')(p't'+q's')]^2 \\
& \quad + p'q'[(s'-t')^2(s-t)(pt+qs)]^2 \} \\
& + r^2\rho^2 p^2q^2p'^2q'^2[(s-t)^2(s'-t')^2]^2
\end{aligned}$$

étant entendu alors que le carré de chaque crochet ne peut être effectué qu'une fois faits, dans le contenu de ce crochet, les remplacements (5).

En faisant ces remplacements dans le premier crochet, il devient $M(A)$, quantité que nous supposons nulle. Alors l'expression de $M(AA_1)$ s'annule bien quand $r=\rho=0$, c'est-à-dire quand A et A_1 sont indépendants.

Les termes suivants, en $(r+\rho)$ et en $r\rho$, s'expriment uniquement à partir des contributions additives i, j, k, i', j', k' , fournies par (2').

Il résulte en effet des équations (2'') qu'on peut, au lieu de remplacer par les a , remplacer $t^2(p't'+q's')^2$ par i , $st(p't'+q's')^2$ par j , etc., ce qui donne :

$$\begin{aligned}
(r+\rho) \{ & pq[-pi+(p-q)j+qk]^2 + p'q'[-p'i'+\dots]^2 \} \\
& + r\rho \{ p^2q^2(k-2j+i)^2 + p'^2q'^2(k'+\dots)^2 \}
\end{aligned}$$

On reconnaît là l'expression (6b) du chapitre précédent, qui représente $M(HH_1)$, et l'expression analogue pour $M(KK_1)$. Nous avons donc là l'expression des 2 premiers termes de (4'). L'expression du dernier terme de (4'), $M(EE_1)$, qui fournit la corrélation cherchée entre les résidus « d'epistacy », est donnée alors par les 3 derniers termes de (7). Par suite ces termes ne doivent dépendre que des valeurs que peuvent prendre ces résidus, soit $e_{11}=a_{11}-i-i'$, $e_{12}=a_{12}-i-j'$ etc., et non des valeurs i, j, \dots des contributions additives. Cela nous permet de simplifier l'expression de ces termes, en supposant que les a se réduisent aux e , c'est-à-dire que les contributions additives sont nulles : $i=j=k=i'=j'=k'=0$; c'est-à-dire encore, d'après (2''), en supposant nulles les

6 quantités telles que $p'^2 a_{11} + 2p'q'a_{12} + q'^2 a_{13}$; ceci revient évidemment à supposer, dans notre calcul symbolique, que :

$$(8) \quad \begin{cases} p'^2 \bar{i}' + 2p'q' \bar{j}' + q'^2 \bar{k}' = 0 \\ p^2 \bar{i} + 2pq \bar{j} + q^2 \bar{k} = 0 \end{cases}$$

Donc, sous cette nouvelle hypothèse, la formule (7) doit se réduire (H et K étant alors certainement nuls) aux seuls termes traduisent l'epistacy, c'est-à-dire aux trois derniers (cela se vérifie immédiatement). Et la forme de ceux-ci se simplifie en y éliminant \bar{j}' et \bar{j} par les relations (8). Les groupements qui y figurent deviennent alors :

$$\begin{aligned} (pt + qs)(s - t) &= (q^2 \bar{k} - p^2 \bar{i}) / 2pq \\ (s - t)^2 &= (p \bar{i} + q \bar{k}) / pq \end{aligned}$$

et de même pour les quantités accentuées.

La formule (7) se réduit donc alors à :

$$(9) \quad \begin{aligned} M(EE_1) &= ((r + \rho)^2 / 16ppq'q') [(q^2 \bar{k} - p^2 \bar{i})(q'^2 \bar{k}' - p'^2 \bar{i}')]^2 \\ &+ (r + \rho)r\rho \left\{ (1/4pq) [(q^2 \bar{k} - p^2 \bar{i})(p' \bar{i}' + q' \bar{k}')]^2 \right. \\ &\quad \left. + (1/4p'q') [(q'^2 \bar{k}' - p'^2 \bar{i}') (p \bar{i} + q \bar{k})]^2 \right\} \\ &+ (r^2 \rho^2 / pqp'q') [(p \bar{i} + q \bar{k})(p' \bar{i}' + q' \bar{k}')]^2 \end{aligned}$$

étant entendu que les a sont supposés réduits aux e , c'est-à-dire qu'une fois effectués les produits qui figurent dans chaque crochet, on y remplace maintenant $\bar{i} \bar{i}'$ par e_{11} , $\bar{i} \bar{k}'$ par e_{13} , $\bar{k} \bar{i}'$ par e_{31} et $\bar{k} \bar{k}'$ par e_{33} .

L'effet de l'epistacy est alors exprimé en fonction des 4 résidus e_{11} , e_{31} , e_{13} , e_{33} , qui, comme nous l'avons signalé p. 63, sont les 4 paramètres arbitraires qui caractérisent l'epistacy.

Applications :

Nous obtenons alors facilement les corrélations pour les types simples de parenté.

La corrélation père-fils s'obtient en faisant $r=1$ et $\rho=0$

(puisque'un de leurs 2 gènes leur est commun, les autres étant indépendants). (9) se réduit alors à son premier terme, et on obtient le résultat donné par Fisher (Ed. p. 409) :

$$(10) \quad M(EE_1) = (1/16 p q p' q') [p^2 p'^2 e_{11} - p^2 q'^2 e_{13} - q^2 p'^2 e_{31} + q^2 q'^2 e_{33}]^2$$

On voit que, contrairement à ce qui se passe dans ce cas pour le résidu de dominance, la corrélation entre E et E_1 n'est pas nulle.

Pour la *corrélation ancestrale*, on a, $r = (1/2)^{n-1}$, $\rho = 0$, et par suite $M(EE_1)$ est la fraction $(1/2)^{2n-2}$ de la forme (10).

$M(EE_1)$ décroît donc plus rapidement, quand n augmente, que $M(HH_1) + M(KK_1)$ (qui décroît comme $(1/2)^n$, d'après p. 51). La relation entre les corrélations ancestrales successives perd donc sa simplicité.

Pour la corrélation entre *oncle et neveu*, entre *cousins germains*, entre *cousins quelconques*, on trouve encore la fraction $1/4$, ou $1/16$, ou $(1/2)^{2(n+p)}$, de la forme (10). D'une façon générale, en désignant cette forme par F , on a pour la corrélation entre individus apparentés par l'intermédiaire d'un de leurs parents seulement ($\rho = 0$) :

$$M(AA_1) = (r/2) [M(\mathbf{H}^2) + M(\mathbf{K}^2)] + r^2 F^2$$

Nous voyons donc que la corrélation entre oncle et neveu est encore identique à la corrélation grand-parentale ($r = 1/2$) et que la corrélation entre cousins est identique à la corrélation ancestrale comportant le même nombre d'échelons.

Cas d'individus apparentés par leurs 2 parents

Pour avoir la corrélation $M(EE_1)$ entre des individus apparentés à la fois par leurs 2 parents, il est commode de revenir à la forme (6a), que nous pouvons écrire en laissant de côté son premier terme qui n'intervient pas dans $M(EE_1)$, et en éliminant \bar{j} , \bar{j}' , \bar{j}_1 , \bar{j}'_1 , à l'aide de (8) :

$$(11) \quad (pq/4) \left\{ \bar{k} \bar{k}_1 [(r+\rho)q^2/p^2 + 4r\rho q/p] + \bar{i} \bar{i}_1 [(r+\rho)p^2/q^2 + 4r\rho p/q] \right. \\ \left. + (\bar{i} \bar{k}_1 + \bar{k} \bar{i}_1)(4r\rho - r - \rho) \right\}$$

expression qu'on doit alors multiplier par la forme analogue accentuée, pour faire ensuite les remplacements indiqués plus haut. On obtient ainsi $M(EE_1)$.

Dans le cas de la *corrélation entre frères*, cela nous donne, puisque $r=\rho=1/2$, et en supprimant l'indice 1, inutile ici :

$$\sigma_{EE_1}^2 = (1/16ppq'q')(q^3 \bar{k}^2 + p^3 \bar{i}^2)(q'^3 \bar{k}'^2 + p'^3 \bar{i}'^2)$$

ce qui fournit, en faisant les remplacements, l'expression donnée par Fisher :

$$M(EE_1) = (1/16ppq'q')(p^3 p'^3 e_{11}^2 + p^3 q'^3 e_{13}^2 + q^3 p'^3 e_{31}^2 + q^3 q'^3 e_{33}^2)$$

Cette méthode fournit une expression commode pour la *variance due à l'epistacy*, $M(E^2)$. Il suffit pour cela de supposer que E_1 correspond au même individu que E , c'est-à-dire que $r=\rho=1$.

L'expression (11) devient alors :

$$(pq/4)[2\bar{k}\bar{k}_1(q/p^2)(q+2p) + 2\bar{i}\bar{i}_1(p/q^2)(p+2q) + 2(\bar{i}\bar{k}_1 + \bar{k}\bar{i}_1)]$$

et, en multipliant par l'expression analogue accentuée et faisant les remplacements, on trouve l'expression de la variance en fonction des 4 paramètres arbitraires e_{11} , e_{13} , e_{31} , e_{33} ; c'est l'expression qui a été donnée par Fisher (Ed., p. 409).

Pour les *doubles cousins* ($r=\rho=1/4$), la forme (11) devient :

$$(pq/4)[(\bar{k}\bar{k}_1/4)(2q+p)q/p^2 + (\bar{i}\bar{i}_1/4)(2p+q)p/q^2 - (\bar{i}\bar{k}_1 + \bar{k}\bar{i}_1)/4]$$

ce qui donne pour $M(EE_1)$ une expression très analogue à celle de la variance.

Nous voyons donc que, dans la *corrélation entre apparentés*, le terme correspondant à l'epistacy a des expressions très diverses suivant le type de parenté envisagé. Il a été calculé pour les types les plus simples par Fisher, toujours par la méthode de la table d'association entre phases. Mais la méthode générale que nous avons employée fournit son expression pour n'importe quel type de parenté. Et surtout elle s'étend, sans difficulté autre que la longueur des formules, au cas où l'on considère plus de 2 couples de gènes.

Généralisation : cas de 3 couples

Considérons maintenant 3 couples par exemple, d'effets partiels H, K, L , et d'effet global G . S'il n'y a pas additivité des effets partiels on a $G=H+K+L+E$, le résidu E prenant des valeurs $e_{111}=a_{111}-i-i'-i''$, $e_{112}=a_{112}-i-i'-j''$, etc.

Nous avons vu que l'ajustement de E par (C) entraîne la relation (4) :

$$M(GG_1)=M(HH_1)+M(KK_1)+M(LL_1)+M(EE_1)$$

et nous obtenons l'expression de ces différents termes toujours par le même calcul symbolique. Il suffit de partir de la forme (6a) relative au couple H , que nous écrivons pour simplifier :

$$(6a) \quad x+\beta(r+\rho)+\gamma r\rho$$

et de la multiplier par les formes analogues relatives aux 2 autres couples, que nous obtenons en récrivant la précédente avec des $p, q, \bar{i}, \bar{j}, \bar{k}$, accentués, ce que nous représentons par :

$$\begin{aligned} x' + \beta'(r+\rho) + \gamma' r\rho \\ x'' + \beta''(r+\rho) + \gamma'' r\rho \end{aligned}$$

Le raisonnement fait pour 2 couples (p. 68) se généralise immédiatement : on obtient $M(GG_1)$ en effectuant le produit de ces 3 formes et en y remplaçant $i \bar{i} \bar{i}''$ (et $\bar{i}_1 \bar{i}'_1 \bar{i}''_1$) par a_{111} , $\bar{i} \bar{i}'' \bar{j}''$ par a_{112} , etc. Si nous représentons entre crochets les résultats de ces produits symboliques nous avons :

$$(12) \quad \begin{aligned} M(GG_1) = & [xx'x''] + (r+\rho)\Sigma[ax'\beta''] + r\rho\Sigma[ax'\gamma''] \\ & + (r+\rho)^2\Sigma[\alpha\beta'\beta''] + (r+\rho)r\rho\Sigma[\alpha\beta'\gamma''] + r^2\rho^2\Sigma[\alpha\gamma'\gamma''] \\ & + (r+\rho)^3[\beta\beta'\beta''] + (r+\rho)^2r\rho\Sigma[\gamma\beta'\beta''] \\ & + (r+\rho)r^2\rho^2\Sigma[\gamma\gamma'\beta''] + r^3\rho^3[\gamma\gamma'\gamma''] \end{aligned}$$

les Σ étant étendues à tous les termes qui se déduisent par permutation des accents.

Interprétation des différents termes

Comme page 69, on voit que le premier terme :

$$[xx'\alpha''] = [(pl+qs)^2(p'l'+q's')^2(p''l''+q''s'')^2]^2$$

représente le carré de $M(G)$, quantité que nous supposons toujours nulle.

De même les 2 termes suivants ne font intervenir que les contributions additives des différents couples. En effet ces contributions sont définies par les équations (2''b) qui donnent par exemple :

$$\begin{aligned} i'' &= \text{symboliquement } l''^2(pl+qs)^2(p'l'+q's')^2 \\ j'' &= s''l''(pl+qs)^2(p'l'+q's')^2, \text{ etc.} \end{aligned}$$

d'où les coefficients de $(r+\rho)$, par exemple

$$\begin{aligned} [xx'\beta''] &= p''q''[(pl+qs)^2(p'l'+q's')^2(p''l''+q''s'')(s''-l'')]^2 \\ &= p''q''[-p''i''+(p''-q'')j''+q''k'']^2 = M(\mathbf{L}^2)/2 \end{aligned}$$

On a de même pour les coefficients de $r\rho$, la forme :

$$\begin{aligned} [xx'\gamma''] &= p''^2q''^2[(pl+qs)^2(p'l'+q's')^2(s''-l'')]^2 \\ &= p''^2q''^2(k''-2j''+i'')^2 = M(d''^2) \end{aligned}$$

Les premiers termes de (12) s'écrivent donc :

$$((r+\rho)/2)\Sigma M(\mathbf{H}^2) + r\rho\Sigma M(d^2) \quad \text{c'est-à-dire } \Sigma M(HH_1)$$

Nous voyons donc que les termes restants, ceux du degré ≥ 2 en $(r+\rho)$ et en $r\rho$, représentent $M(EE_1)$.

Indications générales et conclusion

Dans le cas de non-additivité entre un nombre quelconque n de couples de gènes, on voit immédiatement comment se généralise l'équation (12). On fait le produit symbolique de n formes analogues à (6a), et le 2^e membre de (12) est alors un polynôme de degré n en $(r+\rho)$ et $r\rho$.

L'expression des corrélations usuelles se complique donc quand n croît. La corrélation ancestrale par exemple ($\varphi=0$) dépend de toutes les puissances de r jusqu'à la n^{me} . Mais il est intéressant de souligner que la corrélation entre 2 individus ne dépend, quel que soit leur mode de parenté, que de r et de φ .

Il en résulte que les coefficients de corrélation entre oncle et neveu et entre grand-père et petit-fils doivent toujours être les mêmes (dans les 2 cas on a $r=1/2$, $\varphi=0$); plus généralement, la corrélation est la même avec un ancêtre quelconque qu'entre des cousins du même ordre ($r=(1/2)^{n-1}$, $\varphi=0$).

Nous voyons que l'égalité respective des corrélations en ligne directe et en ligne collatérale, que nous avons déjà signalée, subsiste encore dans le cas de l'epistacy la plus générale.

Elle est caractéristique du « random mating » et disparaît dans le cas de l'homogamie.

CHAPITRE V

LES CORRÉLATIONS USUELLES DANS LE CAS DE L'HOMOGAMIE

Examinons maintenant comment l'hypothèse homogamique de Fisher modifie les corrélations obtenues dans le cas du « random mating ». Les méthodes de calcul deviennent alors bien moins simples, du fait de la liaison que l'homogamie introduit entre tous les couples de gènes. Pour obtenir des formules à peu de paramètres, utilisables pratiquement, il nous faudra faire diverses approximations.

Approximations introduites

Nous supposons que l'homogamie est faible, c'est-à-dire que le *coefficient de choix* λ est petit, de sorte que nous puissions négliger son carré. Cela revient à supposer que la fréquence d'accouplement $\Pi_1\Pi_2(1+\lambda x_1x_2)$ de 2 individus de valeurs x_1 et x_2 est voisine du produit $\Pi_1\Pi_2$ des fréquences de chacun même si x_1 et x_2 ne sont pas petits. Nous admettons donc que des écarts notables du caractère chez les 2 conjoints ne modifient que peu la probabilité de leur accouplement ; ainsi est précisée la portée pratique de notre hypothèse.

(Nous utiliserons en outre l'hypothèse déjà faite p. 17, d'après laquelle le nombre n de couples de gènes est très grand, la contribution de chacun d'eux étant petite par rapport à l'ordre de grandeur de leur somme ; avec plus de précision, nous supposons λ/n négligeable.)

Déterminons alors des valeurs approchées des fréquences (14), (15), (16), du chapitre I ; leur partie principale est évidente :

$$(111) = pp'p'' + 0(\lambda), \quad (111) = p^2p'^2p''^2 + 0(\lambda), \text{ etc.}$$

Mais nous voyons que dans le coefficient de λ figurent des quantités telles que \mathbf{x}_{100} , \mathbf{x}_{110} , \mathbf{x}_{111} , etc. Il faut alors déterminer la partie principale de ces dernières. Nous avons :

$$\begin{aligned} (100)\mathbf{x}_{100} &= (100)i + ((200)/2)j + (110)i' + (120)j' + \\ & \quad (130)k' + (101)i'' + (102)j'' + (103)k'' + \dots \\ \text{en posant par exemple : } (110) &= (110) + (210)/2 \end{aligned}$$

La partie principale de $(100)\mathbf{x}_{100}$ est donc fournie par les 2 premiers termes, $Pi + Qj$, car d'après les conditions (17), les autres ne sont que de l'ordre de λ :

$$\begin{aligned} \mathbf{x}_{100} &= (Pi + Qj) / p + 0(\lambda) \\ \text{de même } \mathbf{x}_{010} &= (P'i' + Q'j') / p' + 0(\lambda) \\ \text{et } (110)\mathbf{x}_{110} &= (110)i + ((210)/2)j + (110)i' + ((120)/2)j' + (111)i'' + \\ & \quad (112)j'' + (113)k'' + \dots = p'(Pi + Qj) + p'(P'i' + Q'j') + 0(\lambda) \end{aligned}$$

On voit donc que, en posant comme p. 29 :

$$\mathbf{x}_{100} = T, \quad \mathbf{x}_{010} = T', \quad \text{etc.}$$

(avec les relations $pT + qS = 0$, etc.), on a :

$$\mathbf{x}_{110} = T + T' + 0(\lambda) \quad \text{et aussi } \mathbf{x}_{130} = T + S' + 0(\lambda), \quad \text{etc.}$$

de même on aurait : $\mathbf{x}_{111} = T + T' + T'' + 0(\lambda)$, etc.,

et nous avons alors :

$$(14') \quad (110) = pp'(1 + \lambda TT')$$

$$(15') \quad (111) = pp'p''[1 + \lambda(TT'' + TT'' + T'T'')] + 0(\lambda^2), \quad \text{etc.}$$

et

$$(16') \quad \left\{ \begin{aligned} (11) &= p^2p'^2[1 + \lambda(T + T')^2 + 2\lambda TT'] + 0(\lambda^2) \\ (12) &= 2p^2p'q'[1 + \lambda(T + T')(T + S') + \lambda TT' + \lambda TS'] + 0(\lambda^2) \\ (22) &= 4ppq'q'[1 + \lambda(T + T')(S + S') + \lambda TT' + \lambda SS'] + 0(\lambda^2) \end{aligned} \right.$$

etc.

Remarquons que ces équations nous fournissent les valeurs approchées adoptées par Fisher pour ses coefficients f . En effet :

$$\begin{aligned} (11) &= PP'(1 + f_{11}) = p^2p'^2(1 + f_{11} + \lambda T^2 + \lambda T'^2) + 0(\lambda^2) \\ \text{d'où } f_{11} &= 4\lambda TT' \quad (\text{équation XXI de Fisher}) \end{aligned}$$

de même :

$$(12) = 2PQ'(1 + f_{12}) = 2p^2p'q'(1 + f_{12} + \lambda T^2 + \lambda T'S') + 0(\lambda^2)$$

$$\text{d'où } f_{12} = (T + T')S' + 2TT' + 0(\lambda^2)$$

etc.

Nous avons donc maintenant les expressions, quand λ^2 est négligeable, des probabilités et des coefficients d'association.

Rôle de la dominance; ajustement par moindres carrés

Considérons la contribution H d'un couple de gènes (valeurs i, j, k , avec probabilités $P, 2Q, R$). Comme au chapitre III, nous décomposerons H en 2 parties, l'une \mathbf{H} prenant les valeurs $\mathbf{i}=2t, \mathbf{j}=t+s, \mathbf{k}=2s$ et correspondant aux contributions additives des 2 gènes ($\mathbf{H} = \mathcal{H} + \mathcal{H}'$), l'autre d étant un résidu qui traduit l'effet de la dominance. Nous déterminerons encore ce résidu par une condition de moindres carrés. Mais, cette fois, les différents couples de gènes H, K , etc., n'étant plus indépendants, il est naturel de ne pas ajuster séparément \mathbf{H} à H, \mathbf{K} à K , etc. Nous poserons encore :

$$x = \Sigma H = \Sigma \mathbf{H} + \Sigma d = z + D$$

et nous déterminerons t, s, t', s' , etc., de façon que :

$M(D^2)$ soit minimum (condition Γ)

$$\text{Or } M(D^2) = (111\dots)(i+i'+\dots-2t-2t'\dots)^2 +$$

$$(211\dots)(j+i'+\dots-t-s-2t'-\dots)^2 +$$

$$(311\dots)(k+i'+\dots-2s-2t'\dots)^2 +$$

$$(121\dots)(i+j'+\dots-2t-t'-s'\dots)^2 +$$

$$(221\dots)(j+j'+\dots-t-s-t'-s'\dots)^2 +$$

$$(321\dots)(k+j'+\dots-2s-t'-s'\dots)^2$$

+ etc.

En annulant les dérivées partielles, on obtient, relativement à chaque couple de gènes, 2 équations. On a par exemple, par rapport à s et t :

$$(1) \quad \frac{\partial}{\partial t} [M(D^2)] = 2(111\dots)(x_{111} - z_{111}) + (211)(x_{211} - z_{211}) +$$

$$2(121)(x_{121} - z_{121}) + (221)(x_{221} - z_{221}) + \dots = 0$$

en désignant par x_{lmn} et z_{lmn} , pour abrégé, les valeurs de x et z pour l'association correspondante de couples.

On voit que cela peut encore s'écrire, avec les notations de p. 26 et 27 :

$$2(100)(x_{100}-z_{100})+(200)(x_{200}-z_{200})=2(100)(\mathbf{x}_{100}-\mathbf{z}_{100})=0$$

$$\text{ou encore } p(\mathbf{T}-\mathbf{T}')=(\mathbf{PI}+\mathbf{QJ})-(\mathbf{PI}+\mathbf{QJ}')=0$$

et de même à partir de $\frac{\partial}{\partial s} [M(D^2)] :$

$$q(\mathbf{S}-\mathbf{S}')=(\mathbf{QJ}+\mathbf{RK})-(\mathbf{QJ}+\mathbf{RK}')=0$$

les quantités soulignées étant celles qui correspondent à \mathbf{H} , par exemple :

$$(2) \quad \mathbf{I}=\mathbf{i}+\Sigma'P'(1+f_{11})\mathbf{i}'+2Q'(1+f_{12})\mathbf{j}'+R'(1+f_{13})\mathbf{k}', \text{ etc.}$$

Les équations qui définissent notre ajustement sont donc :

$$(3) \quad \left\{ \begin{array}{l} \mathbf{PI}+\mathbf{QJ}=\mathbf{PI}+\mathbf{QJ}' \\ \mathbf{QJ}+\mathbf{RK}=\mathbf{QJ}+\mathbf{RK}' \\ \mathbf{P}'\mathbf{I}+\mathbf{Q}'\mathbf{J}'=\mathbf{P}'\mathbf{I}+\mathbf{Q}'\mathbf{J}' \\ \mathbf{Q}'\mathbf{J}'+\mathbf{R}'\mathbf{K}'=\mathbf{Q}'\mathbf{J}'+\mathbf{R}'\mathbf{K}' \\ \text{etc.} \end{array} \right.$$

Ces équations ne sont pas indépendantes, car en les ajoutant 2 à 2 on trouve chaque fois la même équation :

$$\mathbf{PI}+2\mathbf{QJ}+\mathbf{RK}=\mathbf{PI}+2\mathbf{QJ}+\mathbf{RK}=0 \text{ (équation (19), chap. I)}$$

(cela peut aussi s'écrire, d'après (2) :

$$(4) \quad \Sigma P\mathbf{i}+2Q\mathbf{j}+R\mathbf{k}=0 \text{ c'est-à-dire } \Sigma p\mathbf{i}+q\mathbf{k}=0$$

Résolution approchée du système (3)

Nous allons résoudre ce système quand λ est petit. (On peut montrer que, si λ n'est pas supposé petit, la solution peut comporter une certaine indétermination.)

Pour cela, au lieu de déterminer directement \mathbf{i} , \mathbf{j} , \mathbf{k} , etc., nous introduirons les résidus de dominance :

$$\mu=i-\mathbf{i} \quad \nu=j-\mathbf{j} \quad \theta=k-\mathbf{k}, \quad \text{etc.}$$

A partir de ces quantités, on a :

$$\begin{aligned} P(I-\mathbf{I}) &= P\mu + \Sigma'(11)\mu' + (12)\nu' + (13)\theta' \\ 2Q(J-\mathbf{J}) &= 2Q\nu + \Sigma'(21)\mu' + (22)\nu' + (23)\theta', \text{ etc.} \end{aligned}$$

La première équation (3) devient alors :

$$P\mu + Q\nu + \Sigma'(11)\mu' + (12)\nu' + (13)\theta' = 0$$

que nous écrivons : (3') $P\mu + Q\nu = -\Delta$

On a : (11) = (11) + (21) / 2 = $pp'^2(1 + 2\lambda TT' + \lambda T'^2) + 0(\lambda^2)$
d'après (16')

$$(12) = (12) + (22) / 2 = 2pp'q'(1 + \lambda S'(T + T') + \lambda TT') + 0(\lambda^2)$$

$$(13) = (13) + (23) / 2 = pq'^2(1 + 2\lambda TS'^2 + \lambda S'^2) + 0(\lambda^2)$$

ce qui nous permet, en tenant compte de ce que, d'après (4), $\Sigma P\mu + 2Q\nu + R\theta = 0$, et de ce que $q'S' = -p'T'$, d'évaluer Δ :

$$\Delta = 2\lambda pT\Sigma'p'T'[(p'\mu' + q'\nu') - (p'\nu' + q'\theta')] + 0(\lambda^2)$$

La 2^e équation (3) s'écrit de même :

$$(3'') \quad Q\nu + R\theta = -\Delta_3 \quad \text{avec} \quad \Delta_3 = -\Delta + 0(\lambda^2)$$

Nous n'avons plus qu'à chercher l'ordre de grandeur de Δ . Pour cela, remarquons que (3') et (3'') s'écrivent, d'après les équations (16') du chapitre I :

$$\begin{cases} p\mu + q\nu + \lambda pT^2(\mu - \nu) = -\Delta / p \\ p\nu + q\theta + \lambda pTS(\nu - \theta) = \Delta / q + 0(\lambda^2) \end{cases}$$

d'où, par soustraction :

$$p(\mu - \nu) + q(\nu - \theta) + \lambda[pT^2(\mu - \nu) + qS^2(\nu - \theta)] + \Delta / pq = 0(\lambda^2)$$

On voit donc que la quantité $p(\mu - \nu) + q(\nu - \theta)$ que nous désignerons dorénavant par ρ est de l'ordre de λ , et par suite que :

$$\Delta = 2\lambda pT\Sigma'p'T'\rho' + 0(\lambda^2)$$

est de l'ordre de λ^2 , et par suite négligeable.

On a donc :

$$\begin{cases} (3') & P\mu + Q\nu = Pi + Qj - (Pi + Qj) = 0(\lambda^2) \\ (3'') & Q\nu + R\theta = Qj + Rk - (Qj + Rk) = 0(\lambda^2) \end{cases}$$

ce qui donne en ajoutant puis retranchant :

$$(4) \quad \begin{cases} Pi+2Qj+Rk=Pi+2Qj+Rk+0(\lambda^2)=0(\lambda^2) \\ Pi-Rk=Pi-Rk+0(\lambda^2) \end{cases}$$

la première de ces 2 équations donne : $p\mathbf{i}=-q\mathbf{k}$ (au terme en λ^2 près) d'où la solution du système (3) :

$$(5) \quad \mathbf{i}=q(Pi-Rk)/(pR+qP) \quad \mathbf{k}=-p(Pi-Rk)/(pR+qP)$$

à l'ordre de λ^2 près, et de même pour les autres couples.

Un autre ajustement équivalent; analyse de la variance de chaque couple

Au lieu de faire un ajustement global en considérant l'ensemble des couples de gènes, nous pouvons aussi considérer chacun d'eux séparément, par exemple ajuster \mathbf{H} à H indépendamment de sa liaison avec les autres couples. Alors, au lieu de la condition $\Gamma(M(D^2)$ minimum) nous imposons à chaque couple la condition :

$$(C) \quad M(d^2)=P(i-\mathbf{i})^2+2Q(j-\mathbf{j})^2+R(k-\mathbf{k})^2 \quad \text{minimum}$$

ce qui donne, en dérivant par rapport à s et t :

$$(4') \quad \begin{cases} Pi+Qj=Pi+Qj \\ Qj+Rk=Qj+Rk \end{cases}$$

On obtient donc, à l'ordre de λ^2 près, les mêmes équations que dans le cas de l'ajustement Γ . La solution de (4') est fournie rigoureusement par les équations (5). Les 2 ajustements sont donc équivalents dans l'hypothèse que λ^2 est négligeable. Il nous sera avantageux d'employer l'un ou l'autre, suivant les propriétés à mettre en évidence.

Par exemple nous pouvons interpréter, comme au chapitre II, l'ajustement C , en disant que $M(d)=0$ quand \mathcal{H} (ou \mathcal{H}') est fixé à la valeur s , ou t . On en déduit :

$$M(\mathbf{H}d)=M(\mathcal{H}d)+M(\mathcal{H}'d)=0$$

et

$$M(H^2)=M(\mathbf{H}+d)^2=M(\mathbf{H}^2)+M(d^2)$$

La variance de chaque couple se décompose donc en 2. Exprimons ces différentes variances :

$$M(H^2) = Pi^2 + 2Qj^2 + Rk^2$$

$$\beta^2 = M(\mathbf{H}^2) = Pi^2 + 2Qj^2 + Rk^2 = (P+Q/2)i^2 + Qik + (R+Q/2)k^2$$

$$= [(P+Q/2)q^2 - pqQ + (R+Q/2)p^2](Pi - Rk)^2 / (pR + qP)^2$$

Pour simplifier cette expression, remarquons que, d'après le chapitre I [équation (16')] :

$$P = p^2 + z \quad 2Q = 2pq - 2z \quad R = q^2 + z \quad \text{avec} \quad z = \lambda p^2 T^2$$

On a alors $pR + qP = pq + z$, et la relation devient :

$$M(\mathbf{H}^2) = (Pi - Rk)^2 / 2(pq + z)$$

Quant à d , ses valeurs $\mu = i - \mathbf{i}$, $\theta = k - \mathbf{k}$, sont d'après (5) :

$$\mu = R(pi + qk) / (pq + z) \quad \theta = P(pi + qk) / (pq + z)$$

d'où, d'après (4) :

$$\nu = j - \mathbf{j} = -(P\mu + R\theta) / 2Q = -PR(pi + qk) / Q(pq + z)$$

donc :

$$\varepsilon^2 = M(d^2) = P\nu^2 + 2Q\nu^2 + R\theta^2 =$$

$$PQR(QR + PQ + 2PR)(pi + qk)^2 / Q^2(pq + z)^2 =$$

$$PR(pi + qk)^2 / Q(pq + z)$$

On vérifie aisément que cela peut s'écrire ainsi :

$$M(d^2) = PQR(i - 2j + k)^2 / (pq + z)$$

Analyse de la variance totale

Revenons maintenant à l'ajustement Γ , et interprétons la condition (1). Elle signifie que, quand H est fixé à la valeur l , $M(D) = 0$. L'ensemble des équations (3) peut donc se traduire en disant que D a une valeur moyenne nulle chaque fois qu'on fixe la valeur d'un des termes de la somme $z = H + H' + K + K' + \dots$

Comme précédemment on en déduit :

$$M(zD) = M(HD) + M(H'D) + \dots = 0$$

$$\text{donc} \quad M(x^2) = M(z + D)^2 = M(z^2) + M(D^2)$$

On retrouve donc encore la décomposition de la variance globale en variance génétique et variance résiduelle.

Il est intéressant de remarquer que cette décomposition subsisterait encore s'il y avait un degré quelconque d'épistacy, car alors la condition (1) ne serait pas modifiée, quoique x_{lmn} ne soit plus alors la somme des contributions des couples.

Naturellement, du fait de la liaison entre les différents couples introduite par l'homogamie, $M(z^2)$ n'est plus égal à $\Sigma M(\mathbf{H}^2)$ ni $M(D^2)$ à $\Sigma M(d^2)$; un premier problème est donc de calculer ces variances.

Calcul des variances

On a : $M(x^2) = M(Hx) + M(Kx) + \dots = \Sigma PiI + 2QjJ + RkK$
 puisque par définition $x = I$ quand $H = i$, $x = J$ quand $H = j$, etc.
 de même $M(z^2) = M(\mathbf{H}z) + M(\mathbf{K}z) + \dots = \Sigma P\mathbf{i}I + 2Q\mathbf{j}J + R\mathbf{k}K$,
 ce qui s'écrit encore, puisque $2\mathbf{j} = \mathbf{i} + \mathbf{k}$, et d'après (3) :

$$M(z^2) = \Sigma P\mathbf{i}I + 2Q\mathbf{j}J + R\mathbf{k}K$$

relation d'ailleurs immédiate en remarquant que, puisque $M(zD) = 0$:

$$M(z^2) = M(zx) = M(\mathbf{H}z) + M(\mathbf{K}z) + \dots$$

De même :

$$M(D^2) = M(Dx) = \Sigma M(dx) = \Sigma P_{\mu}I + 2Q_{\nu}J + R_{\theta}K$$

Calculons en premier lieu $V = M(z^2)$, car cela peut se faire sans approximation. Remplaçons $2\mathbf{j}$ par $\mathbf{i} + \mathbf{k}$.

$$M(z^2) = \Sigma \mathbf{v}$$

en posant $\mathbf{v} = \mathbf{i}(PI + QJ) + \mathbf{k}(QJ + Rk) = pT\mathbf{i} + qS\mathbf{k} = pT(\mathbf{i} - \mathbf{k})$
 Or, d'après (2) :

$$PI + QJ = P\mathbf{i} + Q\mathbf{j} + \Sigma'(\mathbf{11})\mathbf{i}' + (\mathbf{13})\mathbf{k}'$$

ou encore, d'après l'équation (21) du chapitre I :

$$\begin{aligned} PI + QJ &= P\mathbf{i} + Q\mathbf{j} + p\Sigma'(p'\mathbf{i}' + q'\mathbf{k}') + \lambda pT\Sigma'(p'T'\mathbf{i}' + q'S'\mathbf{k}') \\ &= P\mathbf{i} + Q\mathbf{j} - (p^2\mathbf{i} + pq\mathbf{k}) - \lambda p^2T^2(\mathbf{i} - \mathbf{k}) + \lambda pT\Sigma p'T'(\mathbf{i}' - \mathbf{k}') \end{aligned}$$

en tenant compte de (4), remplaçant $q'S'$ par $-p'T'$, et introduisant la somme Σ étendue à tous les couples de gènes y compris H .

Nous avons donc :

$$\begin{aligned} \mathbf{v} &= M(\mathbf{H}^2) - (p\mathbf{i} + q\mathbf{k})^2 - \lambda\mathbf{v}^2 + \lambda\mathbf{v}\Sigma\mathbf{v} \\ \text{d'où } \mathbf{V} &= \Sigma M(\mathbf{H}^2) - \Sigma(p\mathbf{i} + q\mathbf{k})^2 - \lambda\Sigma\mathbf{v}^2 + \lambda\mathbf{V}^2 \end{aligned}$$

Remarquons maintenant que nous pouvons négliger le 2^e et le 3^e terme, car d'après (4) $p\mathbf{i} + q\mathbf{k}$ est de l'ordre de λ^2 ; et \mathbf{v} est de l'ordre de \mathbf{V}/n , et par suite $\Sigma\mathbf{v}^2$ de l'ordre de \mathbf{V}^2/n .
Donc :

$$(5) \quad M(x^2) = \mathbf{V} = \Sigma M(\mathbf{H}^2) + \lambda\mathbf{V}^2 + 0(\lambda^4) + 0(\lambda/n)$$

L'effet de l'homogamie est donc d'augmenter la somme des variances génétiques partielles de la quantité $\lambda\mathbf{V}^2$, sensiblement.

Calculons maintenant $V = M(x^2)$.

$$M(x^2) = \Sigma v \text{ en posant } v = PiI + 2QjJ + RkK$$

Or on peut éliminer J par $2QJ = -PI - RK$, et on a, en combinant les équations (18) et (1) du chapitre I :

$$\begin{aligned} I &= i + \Sigma' P'(i' - j')f_{11} - R'(j' - k')f_{13} \\ K &= k + \Sigma' P'(i' - j')f_{31} - R'(j' - k')f_{33} \end{aligned}$$

D'où en remplaçant, à l'ordre de λ^2 près, f_{11} par $4\lambda TT'$, f_{13} par $4\lambda TS'$, etc., et en remarquant que $P' = p'^2 + 0(\lambda)$, $R' = q'^2 + 0(\lambda)$:

$$(6) \quad \begin{cases} I = i + 4\lambda T \Sigma' p' T' [p'(i' - j') + q'(j' - k')] + 0(\lambda^2) \\ K = k + 4\lambda S \Sigma' p' T' [p'(i' - j') + q'(j' - k')] + 0(\lambda^2) \end{cases}$$

$$\begin{aligned} \text{Or } pT[p(i-j) + q(j-k)] &= pT[p(\mathbf{i}-\mathbf{j}) + q(\mathbf{j}-\mathbf{k})] + pT\epsilon \\ &= pT(\mathbf{i}-\mathbf{k})/2 + 0(\lambda) \quad \text{d'après p. 81} \\ &= \mathbf{v}/2 + 0(\lambda) \end{aligned}$$

donc la quantité Σ' qui figure dans I et K peut s'écrire :

$$(\mathbf{V} - \mathbf{v})/2 + 0(\lambda)$$

et nous pouvons alors exprimer :

$$\begin{aligned} v &= PI(i-j) - RK(j-k) \\ &= Pi(i-j) - Rk(j-k) + 4\lambda[PT(i-j) - RS(j-k)](\mathbf{V} - \mathbf{v})/2 + 0(\lambda)^2 \\ &= M(H^2) + \lambda \mathbf{vV} - \lambda \mathbf{v}^2 + 0(\lambda^2) \end{aligned}$$

D'où :

$$\begin{aligned} M(x^2) = V = \Sigma v &= \Sigma M(H^2) + \lambda \mathbf{V}^2 - \lambda \Sigma \mathbf{v}^2 + 0(\lambda^2) \\ &= \Sigma M(H^2) + \lambda \mathbf{V}^2 + 0(\lambda^2) + 0(\lambda/n) \end{aligned}$$

La variance totale globale est donc augmentée de la même quantité $\lambda \mathbf{V}^2$ que la variance génétique globale.

Calculons alors $M(D^2)$.

$$\begin{aligned} (8) \quad M(D^2) = M(x^2) - M(z^2) &= \Sigma M(H^2) - \Sigma M(\mathbf{H}^2) + 0(\lambda^2) \\ &= \Sigma M(d^2) + 0(\lambda^2) \end{aligned}$$

La variance résiduelle totale est donc la somme des variances résiduelles partielles (à l'ordre de λ^2 près).

Ce résultat peut naturellement se retrouver directement.

On a :

$$M(D^2) = \Sigma PI(\mu - \nu) - RK(\nu - \theta)$$

et d'après (6) :

$$= \Sigma M(d^2) + 4\lambda[PT(\mu - \nu) - RS(\nu - \theta)]\Sigma'(\dots) + 0(\lambda^2)$$

or la quantité entre crochets est équivalente à $p\ell\rho$, et est donc de l'ordre de λ .

Ou, d'une autre façon, à l'aide de (3') et (3'') :

$$M(D^2) = \Sigma(-Q\nu)(I - 2J + K) + 0(\lambda^2) = \Sigma(-\nu)(pI + qK) + 0(\lambda^2)$$

c'est-à-dire, d'après (6) :

$$M(D^2) = \Sigma(-\nu)(pi + qk) + 0(\lambda^2) = \Sigma M(d^2)$$

Corrélations entre conjoints

Ces corrélations se déterminent aisément à l'aide de la formule du choix.

Nous désignerons toujours par des lettres surlignées

(\bar{H} , \bar{x} , etc.) les aléatoires relatives au conjoint de l'individu considéré.

$$\begin{aligned} M(H\bar{K}) &= PP'ii'(1+\lambda II') + 2PQ'ij'(1+\lambda IJ') + \dots \\ &\quad + 2QP'ji'(1+\lambda JI') + \dots \\ &= \lambda vv' \end{aligned}$$

de même $M(H\bar{H}) = \lambda v^2$, d'où :

$$M(x\bar{x}) = \Sigma\Sigma M(H\bar{K}) = \lambda(\Sigma v)^2 = \lambda V^2$$

Le coefficient de corrélation entre x et \bar{x} , que nous désignerons par m , est donc :

$$m = M(x\bar{x})/V = \lambda V$$

(c'est ce que Fisher désigne par ρ).

Nous pouvons d'ailleurs remarquer que la formule du choix entraîne une régression linéaire de \bar{H} et de \bar{x} par rapport à x , car on a : si x est fixé :

$$M(\bar{H}) = Pi(1+\lambda Ix) + 2Qj(1+\lambda Jx) + Rk(1+\lambda Kx) = \lambda vx$$

et
$$M(x) = \Sigma M(\bar{H}) = \lambda Vx$$

De même $M(\mathbf{H}\bar{\mathbf{K}}) = PP'ii'(1+\lambda II') + \dots = \lambda \mathbf{v}\mathbf{v}'$

puisque $\mathbf{v} = PiI + 2QjJ + RkK$

donc $M(z\bar{z}) = \Sigma\Sigma M(\mathbf{H}\bar{\mathbf{K}}) = \lambda(\Sigma \mathbf{v})^2 = \lambda \mathbf{V}^2$

et le coefficient de corrélation entre z et \bar{z} , que nous désignerons avec Fisher par A , est :

$$A = M(z\bar{z})/\mathbf{V} = \lambda \mathbf{V} = \xi^2 m$$

ξ^2 désignant toujours le rapport $\mathbf{V}/V = M(z^2)/M(x^2)$. (C'est ce que Fisher désigne par C_2 .)

A est inférieur à m , s'il y a dominance (cela tient à ce qu'alors le choix des conjoints ne dépend que de leur « phénotype » x , et retentit plus faiblement sur leur « génotype » z .)

On peut aussi chercher la corrélation entre x et \bar{z} :

Quand x est fixé :

$$M(\bar{\mathbf{H}}) = Pi(1+\lambda Ix) + 2Qj(1+\lambda Jx) + Rk(1+\lambda Kx) = \lambda \mathbf{v}x$$

et
$$M(\bar{z}) = \Sigma M(\bar{\mathbf{H}}) = \lambda \mathbf{V}x$$

La régression de \bar{z} en x est donc linéaire, et le coefficient de corrélation est :

$$M(x\bar{z})/\sqrt{V\bar{V}}=\lambda\sqrt{V\bar{V}}$$

Corrélation père-fils (en ne considérant chez le fils que des contributions génétiques).

Si nous supposons fixées les constitutions génétiques de 2 conjoints, c'est-à-dire l'ensemble des aléatoires H, \bar{H}, K, \bar{K} , etc., le raisonnement fait au début du chapitre II est encore valable et nous fournit pour les contributions génétiques chez un de leurs enfants, les valeurs moyennes :

$$M(\mathbf{H}_1)=\mathbf{H}/2+\bar{\mathbf{H}}/2 \quad M(z_1)=z/2+\bar{z}/2$$

(en affectant de l'indice 1 les aléatoires relatives à l'enfant).

On a alors : $M(xz_1)=xz/2+x\bar{z}/2$ d'où, pour des constitutions génétiques quelconques des parents :

$$M(xz_1)=M(xz/2)+M(x\bar{z}/2)$$

La corrélation entre x et \bar{z}_1 est donc immédiate, car on a là 2 termes égaux, l'un à $M(z^2)/2$ d'après p. 83, l'autre à $\lambda V\bar{V}$ d'après ci-dessus :

$$M(xz_1)=\mathbf{V}(1+\lambda V)/2=\mathbf{V}(1+m)/2$$

On a aussi :

$$M(zz_1)=M(z^2/2)+M(z\bar{z}/2)=\mathbf{V}(1+\lambda\bar{\mathbf{V}})/2=\mathbf{V}(1+A)/2$$

Mais cette méthode simple n'est plus applicable si on considère chez l'enfant le caractère « phénotypique » x_1 , et non plus le caractère « génotypique » z_1 . On a :

$$M(xx_1)=M(xz_1)+M(xD_1),$$

et il faut étudier la corrélation de D_1 avec x , ce que nous allons faire maintenant. Nous montrerons qu'elle est négligeable.

Tables d'association des facteurs du père et du fils

On a $M(xx_1) = \Sigma\Sigma M(HK_1)$

Il faut donc déterminer les probabilités d'association entre couples de gènes du père et du fils ; nous les obtenons grâce à la formule du choix. Par exemple la probabilité [11] pour que $H=i$, et $K_1=i'$ s'obtient ainsi :

Les valeurs i et i' pour H et K_1 sont obtenues nécessairement dans le cas où : $H=i$, $K=i'$, $\bar{K}=i'$; elles sont obtenues avec la probabilité $1/2$ dans les cas où $H=i$, $K=j'$, $\bar{K}=i'$, et où $H=i$, $K=i'$, $\bar{K}=j'$; avec probabilité $1/4$ dans le cas où $H=i$, $K=j'$, $\bar{K}=j'$; et ce sont là les seuls cas où elles soient possibles ; en exprimant les probabilités de ces cas par la formule du choix, on a :

$$[11] = (11)\star P' + (1/2)(12)\star P' + (1/2)(11)\star 2Q' + \\ (1/4)(12)\star 2Q' = (11)\star p'$$

cette relation, à rapprocher de celles du chapitre I, p. 26, exprime que la probabilité cherchée est le produit symbolique des probabilités des associations de gènes de chaque conjoint qui fournissent le résultat cherché.

On trouve des expressions analogues pour les quantités [12], [13], etc. Nous les transcrivons dans la table à 9 cases suivante, où chaque case contient la probabilité pour que soient réalisées à la fois la tête de ligne et la tête de colonne correspondantes :

	$K_1 =$	i'	j'	k'
$H =$	i	$(11)\star p'$	$(11)\star q' + (13)\star p'$	$(13)\star q'$
	j	$(21)\star p'$	$(21)\star q' + (23)\star p'$	$(23)\star q'$
	k	$(31)\star p'$	$(31)\star q' + (33)\star p'$	$(33)\star q'$

Il nous faut également la table d'association entre H et H_1 qui s'obtient très simplement : la probabilité pour que $H=i$ et $H_1=i$ est $P \star P + (1/2)P \star 2Q = P \star p$, etc., ce qui fournit :

		$H_1 =$		
		i	j	k
$H =$	i	$P \star p$	$P \star q$	0
	j	$Q \star p$	Q	$Q \star q$
	k	0	$R \star p$	$R \star q$

Ces tables nous permettent naturellement de retrouver les résultats du paragraphe précédent. On vérifie par exemple que

$$M(x\mathbf{H}_1) = M(H\mathbf{H}_1) + \Sigma' M(K\mathbf{H}_1) \text{ est égal à } \mathbf{v}(1 + \lambda V) / 2$$

Elles vont nous servir en outre à étudier le résidu de dominance $D_1 = d_1 + d'_1 + \dots$ du fils.

Ordre de grandeur de $M(xD_1)$

La première table nous fournit :

$$M(Hd'_1) = [(11)i + (21)j + (31)k] \star (p'p' + q'v') + [(13)i + (23)j + (33)k] \star (p'v' + q'b')$$

En prenant d'abord dans cette expression le produit ordinaire des facteurs, il s'écrit, en remarquant que la somme des 2 crochets est $M(H) = 0$:

$$[(11)i + (21)j + (31)k] \times [p'(p' - v') + q'(v' - \theta')] = 0(\lambda) \times \rho' = 0(\lambda^2)$$

Calculons maintenant à l'ordre de λ^2 près le terme complémentaire du produit symbolique. Il s'obtient en mettant λ en

facteur devant l'expression, et en y remplaçant (11) par (11) $x_{11}=Pp'(I+T')+0(\lambda)$, etc., et $p'p'$ par $p'T'p'$, $q'q'$ par $q'S'q'=-p'T'q'$, etc. Il s'écrit donc :

$$\lambda p'v \times p'T'(p' \rightarrow q') + \lambda q'v \times p'T'(q' \rightarrow q') = \lambda v p'T'p' = 0(\lambda^2)$$

Nous voyons donc que $M(Hd'_1) = 0(\lambda^2)$ est négligeable. On peut dire aussi que $M(Kd_1)$ est négligeable.

La 2^e table nous fournit :

$$\begin{aligned} M(Hd_1) &= (Pi + Qj) \star (p\psi + q\psi) + (Qj + Rk) \star (p\psi + q\psi) \\ &= (Pi + Qj)\zeta + \lambda pT[(PIi + QJj)(\psi \rightarrow \psi) + (QJj + RKk)(\psi \rightarrow \theta)] \end{aligned}$$

Calculons cela à λ^2 près, en remplaçant ζ par sa partie principale obtenue p. 81, I par $i + 0(\lambda)$, P par $p^2 + 0(\lambda)$, T par $pi + qj + 0(\lambda)$, etc.

On trouve :

$$M(Hd_1) = \lambda pT p q [p(i-j)^2(\psi \rightarrow \psi) + q(j-k)^2(\psi \rightarrow \theta)] + 0(\lambda^2)$$

On a donc :

$$M(xd_1) = M(Hd_1) + \Sigma' M(Kd_1) = M(Hd_1) + 0(\lambda^2)$$

Or nous avons vu que $M(xH_1) = \mathbf{v}/2 + \lambda \mathbf{v}V/2$. Comparons ces quantités : $M(Hd_1)$ est de l'ordre de λv^2 . Si nous supposons que \mathbf{v} n'est pas petit par rapport à v , c'est-à-dire que les contributions ajustées sont de l'ordre de grandeur des contributions réelles, $M(Hd_1)$ est par rapport à $\lambda \mathbf{v}V/2$, de l'ordre de v/V , c'est-à-dire de l'ordre de $1/n$, et par suite négligeable.

Valeur approchée de la corrélation père-fils

Donc :

$$\begin{aligned} M(xH_1) &= M(xH_1) + M(xd_1) = \mathbf{v}/2 + \lambda(1 + 0(1/n))\mathbf{v}V/2 + 0(\lambda^2) \\ M(xz_1) &= \Sigma M(xH_1) = \mathbf{V}(1 + \lambda V)/2 + 0(\lambda/n) + 0(\lambda^2) \end{aligned}$$

Ainsi il résulte des approximations faites que $M(xz_1)$ se réduit à $M(xz_1)$.

Le coefficient de corrélation père-fils est alors :

$$M(x_1)/V \sim \mathbf{V}(1+\lambda V)/2V = \lambda^2(1+m)/2$$

C'est l'expression fournie, sans démonstration effective, par Fisher.

Corrélation entre frères ; formule générale

Nous désignerons par les indices 1 et 2 les aléatoires relatives à deux frères. Nous avons à calculer $M(x_1x_2)$, que nous obtiendrons à partir de $M(H_1K_2)$ et $M(H_1H_2)$.

Pour calculer $M(H_1K_2)$ nous remarquerons que, si on suppose fixées les constitutions génétiques des 2 parents, la loi de probabilité de l'ensemble H_1K_2 et celle de l'ensemble H_1K_1 sont alors les mêmes (c'est le produit de la probabilité de H_1 par celle de K_1 ou K_2). On a donc $M(H_1K_2) = M(H_1K_1)$ dès que l'on fixe les constitutions génétiques des parents, et par suite aussi quels que soient ceux-ci dans la population. On a donc, pour 2 frères quelconques :

$$M(H_1K_2) = M(H_1K_1) = (11)ii' + (12)ij' + \dots$$

On voit que la table d'association de H_1 et K_2 chez 2 frères n'est autre que la table d'association des couples de gènes chez un même individu, c'est-à-dire la table des (lm) .

$M(H_1H_2)$ s'obtiendra en considérant les différents génotypes des 2 parents relativement au couple de gènes H . Pour des génotypes fixés, les lois de probabilité de ce couple chez 2 de leurs enfants deviennent indépendantes, on a donc :

$$M(H_1H_2) = M(H_1) \times M(H_2) = [M(H_1)]^2$$

Par exemple, si les 2 parents sont hétérozygotes (ce qui se produit avec la probabilité $2Q \star 2Q$), H_1 prendra les valeurs i, j , ou k , avec les probabilités $1/4, 1/2$, ou $1/4$; et il en sera de même, indépendamment, pour H_2 ; on aura donc alors :

$$M(H_1H_2) = (i+2j+k)^2/16$$

quantité qui multipliée par : $2Q \star 2Q$, fournira un des termes de la moyenne générale ; en déterminant de même les autres, on voit que cette moyenne générale s'écrit :

$$\begin{aligned} M(H_1H_2) = & P \star P i^2 + 2Q \star P(i+j)^2 / 4 + R \star P j^2 \\ & + P \star 2Q(i+j)^2 / 4 + 2Q \star 2Q(i+2j+k)^2 / 16 + R \star 2Q(j+k)^2 / 4 \\ & + P \star R j^2 + 2Q \star R(j+k)^2 / 4 + R \star R k^2 \end{aligned}$$

Elle est différente cette fois de la moyenne générale de $M(H^2_1)$.

En effet, si les 2 parents sont hétérozygotes,

$$M(H^2_1) = i^2 / 4 + j^2 / 2 + k^2 / 4$$

(au lieu de $(i+2j+k)^2 / 16$) ; si un seul est hétérozygote, $M(H^2_1) = i^2 / 2 + j^2 / 2$ par exemple (au lieu de $(i+j)^2 / 4$) ; ce n'est que si les 2 parents sont homozygotes qu'on obtient le même terme que dans $M(H_1H_2)$.

La différence, que nous désignerons par g , s'écrit donc :

$$\begin{aligned} g = M(H^2_1) - M(H_1H_2) = & 4Q \star P [i^2 / 2 + j^2 / 2 - (i+j)^2 / 4] + \\ & 4Q \star R [j^2 / 2 + k^2 / 2 - (j+k)^2 / 4] \\ & + 2Q \star 2Q [i^2 / 4 + j^2 / 2 + k^2 / 4 - (i+2j+k)^2 / 16] \end{aligned}$$

$$(9) \quad g = Q \star [p(i-j)^2 + q(j-k)^2 - (Q/4)(i-2j+k)^2]$$

Considérons maintenant

$$\begin{aligned} M(H_1x_2) = & M(H_1H_2) + \Sigma' M(H_1K_2) \\ \text{on a} \quad \Sigma' M(H_1K_2) = & \Sigma' M(H_1K_1) = M(H_1x_1) - M(H^2_1) \\ \text{d'où} \quad M(H_1x_2) = & M(H_1x_1) - g \end{aligned}$$

$$\text{et} \quad (10) \quad M(x_1x_2) = \Sigma M(H_1x_2) = M(x^2_1) - \Sigma g$$

Il nous suffit donc pour avoir une valeur approchée de la corrélation de chercher une valeur approchée de g , toujours sous les mêmes hypothèses, puis de sommer pour tous les couples de gènes qui interviennent.

Corrélation entre frères ; valeur approchée

Pour cela nous allons exprimer g d'une autre façon, en distinguant dans les contributions i, j, k , qui y figurent leurs parties génétiques $\mathbf{i}, \mathbf{j}, \mathbf{k}$, et leurs résidus de dominance μ, ν, θ . Si nous posons : $\mathbf{i} - \mathbf{j} = \mathbf{j} - \mathbf{k} = h$, et $P\mu = -Q\nu = R\theta = PQRC$ [à $0(\lambda^2)$ près, d'après (3') et (3'')], il suffit de faire dans g :

$$(11) \quad \left\{ \begin{array}{l} i - j = h + \mu - \nu = h + CpR \\ j - k = h + \nu - \theta = h - CqP \end{array} \right.$$

Donnons quelques précisions sur ces quantités h et C . Les résidus de dominance μ, ν, θ sont proportionnels à C , et de même les contributions génétiques sont proportionnelles à h (car, en vertu de $p\mathbf{i} + q\mathbf{k} = 0$ et $\mathbf{i} - \mathbf{k} = 2h$, on a $\mathbf{i} = 2qh$, $\mathbf{k} = -2ph$, $\mathbf{j} = (q - p)h$).

La variance génétique et la variance de dominance, calculées p. 83, doivent donc être proportionnelles respectivement à h^2 et C^2 ; on trouve pour ces variances :

$$\zeta^2 = 2h^2(pq + z) \quad \varepsilon^2 = C^2(pq + z)PQR$$

Si les couples de gènes sont nombreux et de même ordre de grandeur, la variance génétique et la variance de dominance globales, qui sont voisines de $\Sigma \zeta^2$ et $\Sigma \varepsilon^2$ d'après (5) et (8), sont de l'ordre de $n\zeta^2$ et $n\varepsilon^2$. Pour que ces variances globales soient finies, il faut que les ζ et les ε , c'est-à-dire les C et les h , soient petits, de l'ordre de $1/\sqrt{n}$.

On a alors en faisant dans (9) les substitutions (11) :

$$(12) \quad g = Q[h^2 + 2zhC(p - q) + C^2(p^3R^2 + q^3P^2) - (QC^2/4)(pq + z)^2] \\ + \lambda QJ[pTC(pq + z)[2h + C(pR - qP)] - C^2(pq + z)^2QJ/4]$$

rappelons que $z = \lambda p^2 T^2$ (p. 83), et remarquons que :

$$p^3R^2 + q^3P^2 = p^3q^3 + 2\lambda q^2 p^4 T^2 + 0(\lambda^2) \\ Q(pq + z)^2 = p^3q^3 + \lambda q^2 p^4 T^2 + 0(\lambda^2)$$

On voit que (12) peut s'écrire :

$$(13) \quad g = Qh^2 + (3/4)Qp^3q^3C^2 + \lambda \zeta(h, C) + 0(\lambda^2)$$

$\varepsilon(h, C)$ étant un polynôme dont tous les termes contiennent 4 facteurs tels que h, C , ou J, T , qui en sont des combinaisons linéaires [à $0(\lambda)$ près, $J=j=(q-p)h-PRC$, $T=pi+qj=qh$]; il est donc homogène et du 4^e degré en h et C , et son ordre de grandeur est par suite $1/n^2$. Remarquons qu'avec ce même ordre d'approximation on peut remplacer dans (13) Qh^2 par $\beta^2/2$ et $Qp^2q^2C^2$ par ε^2 . On a donc :

$$g = \beta^2/2 + (3/4)\varepsilon^2 + 0(\lambda/n^2) + 0(\lambda^2)$$

D'où en faisant la somme pour les n couples, et remarquant que d'après (5) et (8) $\Sigma\beta^2 = \mathbf{V}(1-A)$ et $\Sigma\varepsilon^2 = M(D^2)$ [à $0(\lambda/n)$ et $0(\lambda^2)$ près] :

$$\Sigma g = ((1-A)/2)M(z^2_1) + (3/4)M(D^2_1) + 0(\lambda/n) + 0(\lambda^2)$$

En portant alors dans (10) :

$$M(x_1x_2) = ((1+A)/2)M(z^2_1) + (1/4)M(D^2_1) + 0(\lambda/n) + 0(\lambda^2)$$

ce qui fournit pour le *coefficient de corrélation entre frères* la valeur approchée :

$$f = M(x_1x_2)/M(x^2_1) \approx \beta^2(1+A)/2 + (1/4)(1-\beta^2)$$

C'est bien là encore la valeur prévue par Fisher.

Remarquons que dans les cas du « random-mating » ($\lambda=0$), nous retrouvons le résultat du chapitre III.

Dans le cas où il n'y a pas dominance ($\beta^2=1$), nous avons $f=(1+A)/2$, résultat qui peut se retrouver rigoureusement : il suffit de remarquer que si on fixe les constitutions génétiques des parents celles de 2 enfants deviennent stochastiquement indépendantes, on a alors $M(z_1z_2) = [M(z_1)]^2 = (z+\bar{z})^2/4$, ce qui donne, pour tous les parents possibles :

$$M(z_1z_2) = M[(z+\bar{z})^2/4] = M(z^2)/2 + M(z\bar{z})/2 = M(z^2)(1+A)/2$$

C. Q. F. D.

Influence du milieu

Nous n'avons pas tenu compte jusqu'ici de l'influence sur le caractère considéré des conditions de milieu, variables d'un individu à l'autre.

Introduisons-les en faisant l'hypothèse suivante :

Le caractère mesuré, que nous représentons maintenant par y , est la somme d'une aléatoire x déterminée par la constitution héréditaire de l'individu, et que nous avons étudiée en détail dans tout ce qui précède ; et d'une aléatoire e qui représente l'effet sur l'individu des conditions de milieu ; nous supposons cette dernière indépendante de la constitution génétique (cela revient à dire que les conditions de milieu introduisent dans le caractère mesuré un terme analogue à une erreur accidentelle).

Nous supposerons pour simplifier que l'origine de e est choisie de telle façon que sa moyenne soit nulle. Alors la moyenne de $y = x + e$ est aussi nulle.

En vertu de l'indépendance, la variance de y , que nous désignerons par W , est : $W = M(x^2) + M(e^2) = V + M(e^2)$. Nous désignerons le rapport V/W par $C^2 (< 1)$.

Examinons les corrélations entre conjoints. Nous avons supposé l'homogamie basée sur la ressemblance des caractères, le choix des conjoints se fait donc d'après les valeurs de y . La fréquence d'accouplement, entre 2 groupes d'individus de valeurs y et \bar{y} et de fréquences Π et $\bar{\Pi}$ sera, d'après l'hypothèse de la page 21 : $\Pi \bar{\Pi} (1 + \lambda y \bar{y})$.

On aura donc pour la population totale :

$$M(y\bar{y}) = \Sigma \Sigma \Pi \bar{\Pi} y \bar{y} (1 + \lambda y \bar{y}) = \lambda (\Sigma \Pi y^2)^2 = \lambda W^2$$

Nous désignerons par $\rho = \lambda W$ le coefficient de corrélation entre y et \bar{y} ; c'est là le coefficient de corrélation que l'on mesure entre conjoints.

Les corrélations entre x et \bar{x} , z et \bar{z} , sont toujours fournies par la page 87, dont tous les calculs restent valables. En effet, pour les individus chez lesquels $H = i$, on a par

définition $M(x)=I$ et par suite aussi $M(y)=I$. La formule du choix fournit donc bien :

$$M(H\bar{K})=\lambda vv' ; M(x\bar{x})=\lambda V^2 ; m=\lambda V=C^2\bar{p}$$

et de même

$$M(\bar{z}z)=\lambda \mathbf{V}^2 ; A=\lambda \mathbf{V}=\varepsilon^2 m=\varepsilon^2 C^2\bar{p}$$

Remarquons maintenant que, si y est fixé, on a (cf. p. 87) :

$$M(\bar{\mathbf{H}})=\lambda \mathbf{v}y \quad \text{et} \quad M(\bar{z})=\lambda \mathbf{V}y, \quad \text{d'où} \quad M(y\bar{z})=\lambda \mathbf{V}W$$

Pour la *corrélation parentale*, on a :

$$M(yy_1)=M(yz_1)+M(yD_1)+M(ye_1)$$

Or $M(ye_1)=0$, puisque e_1 est indépendant de y ; et on peut montrer que $M(yD_1)$ est négligeable comme $M(xD_1)$.

On a donc encore :

$$\begin{aligned} M(yy_1) \cap M(yz_1) &= M(yz) / 2 + M(\bar{y}z) / 2 \\ &= \mathbf{V} / 2 + \lambda \mathbf{V}W / 2 \end{aligned}$$

puisque $M(ez)=0$ à cause de l'indépendance, et $M(Dz)=0$ (p. 83).

La *corrélation parentale* est donc :

$$p = M(yy_1) / W \cap C^2 \varepsilon^2 (1 + \mu) / 2$$

La *corrélation fraternelle* s'obtient immédiatement en remarquant que le calcul fait pour $M(x_1x_2)$ n'est pas modifié, et que e_1 et e_2 sont indépendantes de x_1 et x_2 , et sans liaison entre elles. On a donc $M(y_1y_2)=M(x_1x_2)$, ce qui donne :

$$f = M(y_1y_2) / W \cap C^2 \varepsilon^2 [(1 + C^2 \varepsilon^2 \mu) / 2 + (1 - \varepsilon^2) / 4 \varepsilon^2]$$

Nous voyons donc que, dans la *corrélation parentale*, la dominance et l'influence du milieu jouent le même rôle, qui est de réduire la *corrélation* dans le rapport des variances (facteurs C^2 et ε^2). Il en serait de même d'ailleurs dans le cas de la *corrélation ancestrale*, qui ne contiendrait aussi que le produit $C^2 \varepsilon^2$. Mais dans le cas des frères il en est autre-

ment, à cause du terme en $(1-\xi^2)$. La corrélation fraternelle permet ainsi, comme Fisher l'a souligné, de séparer ce qui revient à la dominance et ce qui revient à l'influence du milieu.

Concordance avec l'expérience

Fisher a appliqué ces formules à l'interprétation des résultats publiés par Pearson et Lee dans le tome II de *Biometrika*, au sujet de l'hérédité de certains caractères mesurables chez l'homme. Par exemple, entre les tailles, les corrélations maritale, paternelle, et fraternelle, sont :

$$\mu=0,280 \qquad p=0,506 \qquad f=0,543$$

Des 2 premiers coefficients on tire $C^2\xi^2=2p/(1+\mu)=0,79$ et f fournit alors :

$$\xi^2=0,76.$$

On trouve donc que C^2 est très voisin de 1, ce qui concorde bien avec les observations des biométriciens sur la faible influence des conditions de milieu sur la taille. Le taux de dominance ξ^2 est de l'ordre de 0,75, il joue donc un rôle très appréciable, ce qui justifie l'étude détaillée que nous avons faite des répercussions de la dominance.

BIBLIOGRAPHIE

L. BLARINGHEM. — Principes et formules de l'hérédité mendélienne (Gauthier-Villars, 1928).

R. A. FISHER. — Transactions of the R. S. of Edinburgh, 52, 1918, p. 399.

R. A. FISHER. — The Genetical Theory of Natural Selection (Oxford, 1930).

F. GALTON. — Natural Inheritance (London, 1889).

P. LEVY. — Théorie de l'addition des variables aléatoires (Gauthier-Villars, 1938).

K. PEARSON. — Philosophical Transactions, 203 A, 1903, p. 53.

K. PEARSON. — Proceedings of the R. S. of London, B 81, 1909, p. 225.

SNOW. — Proc. of the R. S. of London, B 83, 1911, p. 37.

S. WRIGHT. — Journal of Agricultural Research, 20, p. 560.

S. WRIGHT. — Genetics, 1921 et 1931.

TABLE DES MATIÈRES

AVANT-PROPOS	1
CHAPITRE I. — Les variables aléatoires qui traduisent l'action des facteurs mendéliens.	3
CHAPITRE II. — Les corrélations entre apparentés dans le cas du " random mating ", en l'absence de dominance.	31
CHAPITRE III. — Les corrélations dans le cas du " random mating " avec dominance.	43
CHAPITRE IV. — Conséquences de la non-additivité des gènes, dans le cas du " random mating "	61
CHAPITRE V. — Les corrélations usuelles, dans le cas de l'homogamie	77
BIBLIOGRAPHIE	99

DEUXIÈME THÈSE

PROPOSITIONS DONNÉES PAR LA FACULTÉ

**Les théorèmes généraux de la théorie
des fonctions méromorphes**

Vu et approuvé :

Paris, le 3 Mars 1939.

LE DOYEN DE LA FACULTÉ DES SCIENCES,
CH. MAURAIN

Vu et Permis d'imprimer :

LE RECTEUR DE L'ACADÉMIE DE PARIS,
G. ROUSSY

IMPRIMERIE J. GUILHOT

52, Boulevard Malesherbes

PARIS

