

## L'ÉTUDE DES DISTRIBUTIONS DE FRÉQUENCES EN VUE DE DÉCELER LA PRÉSENCE DE GÈNES A EFFET IMPORTANT

P. MÉRAT

Station centrale de Génétique animale,  
Centre national de Recherches zootechniques, 78 - Jouy-en-Josas  
Institut national de la Recherche agronomique

---

### INTRODUCTION

Dans plusieurs cas, des chercheurs ont montré l'existence d'un ou de quelques gènes à effet prépondérant, dans des caractères auparavant considérés comme polygéniques (WRIGHT, 1950 ; THODAY et BOAM, 1961 ; SPICKETT, 1963). Ceci a pu être réalisé chez la drosophile grâce aux gènes marqueurs répartis sur tous les chromosomes.

Pour des espèces moins bien connues, on ne dispose pas de méthodes permettant de déceler une telle situation. CASTLE (1921), DEMPSTER et SNYDER (1950), WRIGHT (1950) ont estimé dans des hypothèses restrictives un nombre *minimum* de loci en cause, ce qui ne répond donc pas à la question que nous nous posons. Seules, à notre connaissance, les considérations émises par ROBERTSON (1961, 1964) suggèrent une telle méthode, reposant sur la mise en œuvre d'expériences de sélection sur un certain nombre de générations.

Aussi nous a-t-il paru intéressant de chercher dans une autre direction, basée sur l'étude des distributions de fréquences à l'intérieur de familles ou en général de groupes génétiquement distincts.

### I. — L'EFFET DE LA SÉGRÉGATION DE GÈNES MAJEURS SUR LA FORME DES DISTRIBUTIONS

Nous raisonnerons sur les paramètres des distributions théoriques, dans l'hypothèse où il ne se pose pas de questions d'échelle liées à la nature même du caractère enregistré, telles que limite supérieure ou inférieure imposées ou corrélation entre

moyenne et variance, et où, d'autre part, les facteurs non génétiques de variation sont nombreux et répartis au hasard entre les différentes familles ou groupes comparés.

a) Une hétérogénéité des variances intra-familles dans une population panmictique peut provenir de la ségrégation de gènes dans certaines familles et pas dans d'autres, mais elle peut aussi avoir d'autres causes (répartition variable, d'une famille à l'autre, d'allèles de « sensibilité » à des fluctuations d'origine non génique, ou effets maternels influant sur la variance des descendants).

De toute façon, nous avons vérifié sur l'exemple de gènes à effet égal et additif que le test de cette hétérogénéité ne permet pas de discriminer entre l'hypothèse d'un petit nombre de gènes en cause et celle d'un nombre modérément grand. Son intérêt est donc limité.

b) L'étude de la forme des distributions intra-groupes paraît susceptible d'apporter des indices plus intéressants.

En l'absence de questions d'échelle, une hétérogénéité de la forme des distributions doit recevoir son explication propre. Nous nous plaçons dans le cas où ces questions ne se posent pas à première vue, et où, de plus, les distributions observées sont sensiblement normales dans certaines familles ou groupes génétiques ayant, dans leur ensemble, même intervalle de variation que les autres. Un problème d'échelle peut alors, selon toute vraisemblance, être éliminé.

Le coefficient de symétrie (KENDALL, 1959) peut être rendu non nul par la ségrégation, par exemple, d'un gène dominant à effet important (ségrégation 3/1) ; mais le même résultat pourrait se concevoir avec des polygènes influant sur le « tamponnement » d'une performance dans une direction seulement.

Par contre, le coefficient  $\gamma_2$  de kurtosis (KENDALL, 1959) ne devrait pas s'écarter de la valeur 0 dans des groupes particuliers si des polygènes sont partout en cause, des facteurs de variation multiples, dont aucun n'est prépondérant, produisant une distribution normale. Pour des différences *polygéniques* de sensibilité au milieu, elles tendraient à rendre  $\gamma_2$  positif (MÉRAT, 1967).

Quant à l'existence de gènes dont l'effet propre serait de produire des distributions aplaties ou au contraire leptokurtiques, il semble difficile à concevoir qu'elle soit un phénomène répandu.

Au contraire, la ségrégation de gènes à effet important peut, elle, produire des distributions de forme *aplatie* (platykurtiques).

On est tenté de penser qu'une telle ségrégation doit se refléter par une distribution à plusieurs modes. Or, il est facile de montrer que, dans le cas d'un mélange en proportion égale de deux populations normales de même variance, la distribution résultante ne devient bimodale que si la différence des moyennes de ces deux populations atteint une valeur supérieure à 2 écarts-types. Sinon, la distribution résultante est seulement « trop aplatie » (coefficient  $\gamma_2$  négatif).

Dans le cas général, considérons plusieurs populations normales de variance unité, de moyennes  $m_1, m_2, \dots$  mélangées dans des proportions  $p_1, p_2, \dots$

En prenant pour origine la moyenne de la population résultante, le coefficient d'aplatissement  $\gamma_2$  de cette population sera du signe de

$$E(x^4) - 3 [E(x^2)]^2$$

Si l'on détaille d'après les distributions composantes et pose  $x'_i = x_i - m_i$ ,  $i$  étant l'indice de l'une de ces distributions, on a

$$E(x^4) = \sum_i p_i E(x'_i + m_i)^4 \text{ et } E(x^2) = \sum_i p_i E(x'_i + m_i)^2$$

Compte tenu de ce que les distributions élémentaires sont normales de variance unité, donc que  $E(x'_i) = 0$ ,  $E(x'^2_i) = 1$ ,  $E(x'^3_i) = 0$ ,  $E(x'^4_i) = 3$ , la quantité étudiée équivaut à

$$\sum_i p_i m_i^4 - 3 \left( \sum_i p_i m_i^2 \right)^2,$$

et l'on montre qu'elle doit généralement être négative.

On le vérifie, par exemple, pour deux distributions mélangées en proportion égale, ou en proportion 3/1, ou pour trois distributions mélangées dans la proportion 1/2/1.

La valeur absolue du coefficient  $\gamma_2$  diminue assez vite lorsque croît le nombre de gènes en cause.

En fin de compte, la présence, seulement dans certaines familles, ou groupes génétiquement individualisés, d'une distribution aplatie suggère avec une forte vraisemblance la ségrégation dans ces groupes d'un petit nombre de loci à effet important. On ne voit guère, sous les hypothèses que nous avons définies, d'autres facteurs susceptibles d'aboutir à ce résultat.

Dans ces mêmes groupes, on devrait observer une variance phénotypique plus élevée.

## II. — TESTS PROPOSÉS POUR LA PRÉSENCE DE GÈNES

### A EFFET PRÉPONDÉRANT SUR UN CARACTÈRE QUANTITATIF

a) Dans une comparaison entre deux populations distinctes A et B et leurs croisements  $F_1$  et  $F_2$ , si ces populations diffèrent par un petit nombre de gènes « importants » pour un caractère, le croisement de retour  $F_1 \times A$  ou  $F_1 \times B$  présentera une ségrégation de ces gènes. En conséquence, un échantillon suffisant de ce croisement, élevé dans les mêmes conditions que les lignées parentes, devra présenter, outre une variance phénotypique plus élevée, un coefficient d'aplatissement  $g_2$  négatif.

Le test que nous suggérons est donc la comparaison des coefficients  $g_2$  (SNEDECOR, 1959) dans ces croisements de retour et les lignées parentes.

b) Lors de la comparaison entre familles à l'intérieur d'une population panmictique, le test de l'homogénéité des coefficients  $g_2$  observés sera souvent impossible, car leur distribution n'est normale que pour des effectifs relativement élevés (KENDALL, 1959).

On pourrait, à défaut, utiliser le test de GEARY (1935).

Mais comme, dans l'hypothèse de gènes majeurs, les familles présentant une ségrégation pour ces gènes auront une distribution plus variable en même temps qu'aplatie, il paraît naturel de comparer les coefficients  $g_2$  obtenus dans l'ensemble

des familles à plus grande et dans l'ensemble à celles à plus petite variance observée <sup>(1)</sup> (à supposer que cette dernière se soit révélée hétérogène), ces deux groupes étant délimités arbitrairement.

Par cette séparation des familles à variance *observée* grande ou petite, on risque, ou d'avoir dans chacune des deux catégories un mélange de populations à variance « vraie » différente, ou, inversement, un échantillon de familles à variance observée plus homogène qu'un échantillon « au hasard ».

Cependant, l'une ou l'autre de ces éventualités ne risque pas de rendre négatif le coefficient de « kurtosis », car, d'une part, on peut montrer qu'un mélange de distributions de variance « vraie » différente doit le rendre positif, de l'autre, la covariance entre ce coefficient et la variance observée dans un échantillon est nulle.

Ce test permettra donc de conclure si, dans les familles à variance *vraie* la plus grande, l'espérance mathématique de  $g_2$  ne diffère pas de celle des autres familles.

*En conclusion*, les tests proposés nous paraissent pouvoir apporter, sur l'existence de gènes « à effet important », des présomptions sérieuses.

Leur sensibilité est *a priori* relativement faible, mais la contrepartie est la généralité de leur emploi.

On peut penser à les utiliser préalablement avant qu'une confirmation puisse être donnée par une expérience de sélection s'inspirant des considérations de ROBERTSON.

Un exposé comportant des démonstrations plus détaillées de nos conclusions, et leur application à des données expérimentales sur volailles, sera publié prochainement.

*Reçu pour publication en octobre 1967.*

## SUMMARY

### A STUDY ON FREQUENCY DISTRIBUTIONS IN ORDER TO TEST FOR THE PRESENCE OF GENES WITH AN IMPORTANT EFFECT

A method for the detection of genes with an important effect for a quantitative trait in a random population is suggested as follows. It is possible to compare the coefficients of « kurtosis » between families in which the variance observed was the higher and families in which it was the lower, after having tested for the heterogeneity of these variances.

This comparison of « kurtosis » can also be used with « back crosses » compared to the parental populations bred in the same conditions.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- CASTLE W., 1921. An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Science*, **54**, 93.
- DEMPSTER E. R., SNYDER L. A., 1950. A correction for linkage in the computation of number of gene differences. *Science*, **111**, 283-285.

<sup>(1)</sup> Ces coefficients  $g_2$  étant toujours calculés sur la base des écarts aux moyennes de familles.

- GEARY R. C., 1935. The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrics*, **27**, 310.
- KENDALL M. G., 1959. *The advanced theory of statistics*. II. Griffin and Co, London.
- MÉRAT P., 1967. Moyens possibles de déceler des gènes influant sur la variance phénotypique. *Ann. Biol. anim. Bioch. Biophys.* (sous presse).
- ROBERTSON A., 1960. A theory of limits in artificial selection. *Proc. Roy. Soc. B*, **153**, 234-249.
- ROBERTSON A., 1964. The effect of initial reverse selection upon total selection response. *Genet. Res.*, **5**, 68-79.
- SNEDECOR G. W., 1959. *Statistical methods*. Iowa State College Press.
- SPICKETT S. G., 1963. Genetic and developmental studies of a quantitative character. *Nature*, **199**, 870-873
- THODAY J. M., BOAM T. B., 1961. Regular responses to selection. I. Description of response. *Genet. Res.*, **2**, 161-176.
- WRIGHT S., 1950. *The genetics of quantitative variability*. Quantitative inheritance. London, H. M. S. O., p. 5-41.
-