

# REVUE DE STATISTIQUE APPLIQUÉE

E. LABORIE

A. KESSOUS

## **Essai d'analyse multidimensionnelle d'un caryotype**

*Revue de statistique appliquée*, tome 20, n° 2 (1972), p. 71-77

[http://www.numdam.org/item?id=RSA\\_1972\\_\\_20\\_2\\_71\\_0](http://www.numdam.org/item?id=RSA_1972__20_2_71_0)

© Société française de statistique, 1972, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

# ESSAI D'ANALYSE MULTIDIMENSIONNELLE D'UN CARYOTYPE

E. LABORIE et A. KESSOUS  
Faculté des Sciences Exactes de Pau

Le caryotype somatique de *Pleurodèles waltlii* Michah (Amphibien urodèle) a fait l'objet de plusieurs publications.

Gallien et ses collaborateurs [5] ont effectué le classement des chromosomes en tenant compte des deux principaux critères d'identification qui sont la longueur relative et l'indice centromérique (système de Denver 1960).

Il nous a paru intéressant de reprendre l'étude du caryotype de cette espèce pour essayer d'apprécier par une méthode statistique multidimensionnelle, la valeur d'un tel classement.

## ETABLISSEMENT DU CARYOTYPE

Les 19 mitoses étudiées ont été obtenues à partir de gonades mâles adultes.

Nous avons apparié les chromosomes et rangé les paires par ordre de longueur décroissante dans les 3 groupes établis par Gallien et collaborateurs.

Les chromatides de chaque bras ont été mesurées séparément du centromère à l'extrémité libre. Nous avons ensuite calculé la longueur moyenne des 2 chromatides pour obtenir la longueur des bras de chacun des 24 chromosomes et la longueur totale du lot diploïde. Nous avons alors constaté que la longueur d'un même chromosome variait beaucoup selon les mitoses; alors nous avons exprimé cette longueur en pourcentage relativement à la longueur totale du lot diploïde. C'est ce que nous appelons la longueur relative.

De ces mesures, nous avons tiré l'indice centromérique, rapport entre la longueur du petit bras et la longueur totale du chromosome. Chaque chromosome se trouve donc caractérisé par son indice centromérique et sa longueur relative. Avec ces données, nous avons vérifié si l'appariement initial et le placement sont corrects.

Deux problèmes sont posés : l'appréciation de la validité de l'appariement et de la validité du classement.

Les techniques d'analyse multidimensionnelle permettent de considérer le vecteur des caractéristiques de chaque chromosome (vecteur à deux coordonnées).

Par analogie avec l'analyse de variance à une dimension, on construit un critère de test pour éprouver l'hypothèse d'homogénéité

- 1/ entre membres d'une même paire ;
- 2/ entre membres de paires "voisines".

Un test de comparaison des moyennes est fait dans un cas de populations indiscernables.

Formalisation du problème : les observations conduisent à l'étude de  $k \times 19$  vecteurs aléatoires  $X_{ij}$  (la première coordonnée étant la longueur relative, la deuxième l'indice centromérique) où  $i = 1, 2, \dots, 19$  est l'indice de la population.

Ces  $X_{ij}$  sont supposés normaux indépendants, de même matrice de covariance  $\Sigma$  et notés  $X_{ij} : N_2(\mu_i, \Sigma)$ .

Les valeurs observées de ces  $k \times 19$  v.a. sont les mesures faites lors de 19 épreuves indépendantes (les 19 mitoses) sur des individus issus de  $k$  populations. A chaque épreuve, on obtient  $k$  valeurs vectorielles  $x_{1(1)}, \dots, x_{j(k)}$  mais sans savoir laquelle est issue de la population 1, laquelle est issue de la population 2, etc... On ne connaît les valeurs observées des  $k$  populations qu'à une permutation près.

C'est cette situation que nous appelons : cas de populations indiscernables. Le critère de test doit être indépendant de l'affectation des valeurs observées c'est-à-dire invariant par permutation de l'indice  $j$  (indice de la population).

## I - TEST D'HOMOGENEITE ENTRE MEMBRES D'UNE MEME PAIRE

Si l'on admet que les chromosomes, que l'expérimentateur a groupé en une seule paire, sont issus de deux populations différentes (alors  $k = 2$ ), on éprouve l'hypothèse  $H_0 : \mu_1 = \mu_2 = \mu$ .

Soit la matrice :

$$M_1 = \sum_{i=1}^{19} 2 (\bar{X}_{i.} - \bar{X}_{..}) (\bar{X}_{i.} - \bar{X}_{..})'$$

Les  $\bar{X}_{i.}$  étant indépendants et de même loi, cette matrice suit une loi de Wishart à  $n - 1 = 18$  degrés de liberté.  $M_1 : W(18, \Sigma)$ .

Soit la matrice :

$$M_2 = \sum_{i=1}^{19} \sum_{j=1}^2 (X_{ij} - \bar{X}_{i.}) (X_{ij} - \bar{X}_{i.})' = \sum_{i=1}^{19} M_{2i}$$

Si l'hypothèse  $H_0$  est vérifiée,  $X_{ij} : N_2(\mu, \Sigma)$  et  $M_{2i}$  suit une loi de Wishart à  $k - 1 = 1$  degré de liberté  $M_2 : W(1, \Sigma)/H_0$ .

L'indépendance des épreuves implique celle des matrices  $M_{21}$  et d'après le théorème d'additivité des matrices de Wishart indépendantes :

$$M_2 : W(n(k-1), \Sigma) / H_0$$

Les matrices aléatoires  $M_1$  et  $M_2$  étant indépendantes en probabilité, et ne faisant pas intervenir l'indice  $j$ , elles permettent de construire une statistique de test pour  $H_0$ .

$$\text{Formons } U = \frac{|M_1|}{|M_1 + M_2|} \text{ notée } U_{2, 19, 18}.$$

On montre (voir (7)) que, si  $H_0$  est vraie, les moments de la variable aléatoire  $U$  peuvent s'exprimer sous forme d'un produit de moments :

$$E[U^h] = E[Y_1^h] \times E[Y_2^h]$$

où les variables aléatoires  $Y_i$  sont indépendantes et de densité bêta. A l'aide de la formule de duplication de Legendre, on peut encore écrire (voir (1)) :  $E[U^h] = E[(Y^2)^h]$  où  $Y$  suit une loi  $\beta(n-2=17, n(k-1)=19)$ .

La variable aléatoire  $U$  est distribuée comme le carré d'une variable aléatoire bêta et par changement de variable  $V = \frac{1 - \sqrt{U}}{\sqrt{U}}$ . On vérifie que  $\frac{n-2}{n(k-1)} V$  suit une loi de Fisher.

$$\frac{17}{19} \frac{1 - \sqrt{U_{2, 19, 17}}}{\sqrt{U_{2, 19, 17}}} \text{ suit une loi de Fisher à 38 et 34 degrés de liberté.}$$

C'est ce critère que nous allons utiliser pour tester l'hypothèse  $H_0$  contre son contraire.

La valeur observée de cette variable aléatoire sera calculée et comparée à la valeur  $f_\alpha(38, 34)$ . L'hypothèse sera rejetée si cette valeur est supérieure à  $f_\alpha$ .  $\text{Pr}[F > f_\alpha] = \alpha$ .

Pour la paire 1 par exemple, les résultats sont les suivants :

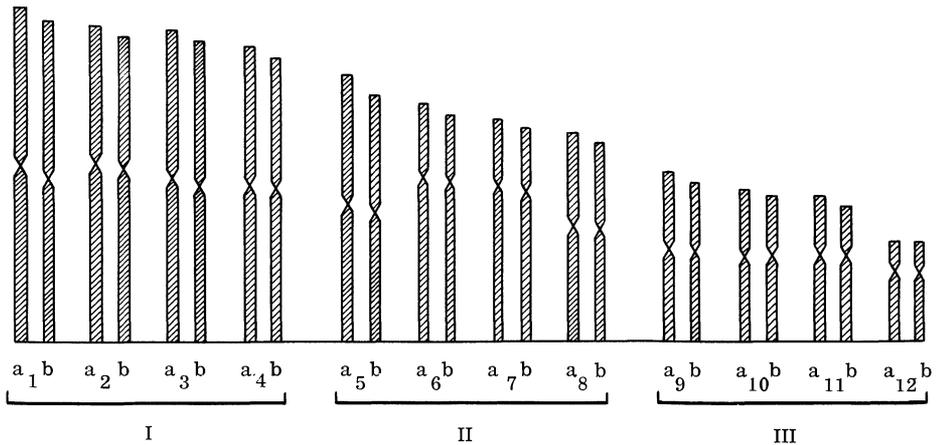
- la valeur observée de  $U_{2, 19, 18}$  est : 0,259.

- la valeur observée de  $F_{38, 34}$  est : 0,862 donc très inférieure aux valeurs de  $F$  aux niveaux usuels ( $f_{38, 34}$  au niveau 5 % = 1,75) l'hypothèse d'homogénéité est acceptée.

## II - TEST DE COMPARAISON DES MOYENNES

Les expérimentateurs rangent les chromosomes par ordre de grandeur décroissante en 12 paires, ces 12 paires formant 3 groupes (5).

Les paires, numérotées 1, 2, 3, 4 forment le groupe 1, les paires 5, 6, 7, 8 forment le groupe 2 et les paires 9, 10, 11, 12 dans le groupe 3. Après avoir calculé la longueur relative moyenne et l'indice centrométrique moyen de chaque paire, le caryogramme moyen a été établi.



Caryogramme relatif au Pleurodèle (x 2000)

On constate d'une part que des chromosomes de longueur et d'indice centromérique très voisins sont considérés comme appartenant à des paires différentes (c'est le cas des paires du groupe 1).

On constate d'autre part que des chromosomes de longueur voisine ont des indices centromériques très différents (paires 5 et 6).

En introduisant une covariance entre les deux coordonnées de chaque chromosome, l'analyse multidimensionnelle nous permet d'apprécier la validité d'un tel classement. Un test des comparaisons des moyennes significatif justifiera une séparation en deux paires de chromosomes empiriquement classés ainsi; inversement, un test non significatif nous amènera à préconiser un groupement plus large.

Les hypothèses de travail sont les mêmes qu'au paragraphe I :  $X_{1j}$  :  $N_2(\mu_1, \Sigma)$  et  $X_{1j}$  indépendant de  $X_{1j'}$ , dès que  $i$  est différent de  $i'$  ou  $j \neq j'$ . Considérons que les chromosomes rangés en deux paires voisines appartiennent à quatre populations différentes (ici  $k = 4$ ) et éprouvons l'hypothèse  $H_0$  : ces populations ont même moyenne ou  $X_{1j}$  :  $N_2(\mu, \Sigma)$ .

Comme précédemment, on forme la matrice

$$N_1 = \sum_{i=1}^4 (\bar{X}_{i.} - X_{..}) (\bar{X}_{i.} - \bar{X}_{..})'$$

et la matrice

$$N_2 = \sum_{i=1}^{19} \sum_{j=1}^4 (\bar{X}_{ij} - \bar{X}_{i.}) (\bar{X}_{ij} - \bar{X}_{i.})'$$

La matrice  $N_1$  suit toujours une loi de Wishart à 18 degrés de liberté.

Si l'hypothèse  $H_0$  est vraie, la matrice  $N_2$  suit une loi de Wishart à  $n(k - 1) = 57$  degrés de liberté.

Ces deux matrices sont indépendantes en probabilité et toutes les deux indépendantes de l'indice  $j$ .

$$N_1 : W(18, \Sigma) \quad \text{et} \quad N_2 : W(n(k-1) = 17 ; \Sigma)/H_0$$

Si nous formons  $V = \frac{|N_1|}{|N_1 + N_2|}$  notée  $V_{2,57,18}$ , nous pouvons montrer (7) que sous  $H_0$  la variable aléatoire  $V$  est distribuée comme le carré d'une variable aléatoire bêta donc la variable aléatoire

$$\frac{n-2}{n(k-1)} V = \frac{17}{57} \frac{1 - \sqrt{U_{2,57,18}}}{\sqrt{U_{2,57,18}}}$$

suit une loi de Fisher

$$(2n(k-1) = 114 \quad \text{et} \quad 2(n-2) = 34 \text{ degrés de liberté.})$$

De même qu'au paragraphe I, les tables numériques de Snedecor permettent de tester l'hypothèse  $H_0$  contre son contraire.

Remarque :

Ce test reste évidemment valable pour la comparaison de moyennes de  $k$  populations ou  $k$  serait supérieur à 4 ( $k$  pair).

Toutes les paires peuvent être ainsi comparées 2 à 2 (ou 3 ensemble, etc...). Le tableau suivant résume les résultats obtenus dans la comparaison des paires du groupe 1 et de certaines paires du groupe 3.

Paires comparées	$V_{(2,57,18)}$	$F_{(114,34)}$	au seuil de 5 %
1 et 2	0,0300	1,42	non significatif
2 et 3	0,0365	1,26	non significatif
1 et 3	0,0580	0,94	non significatif
2 et 4	0,0240	1,63	significatif
3 et 4	0,3339	0,22	non significatif
9 et 10	0,0304	1,41	non significatif
10 et 11	0,0340	1,32	non significatif
11 et 12	0,0109	2,55	très significatif

Valeur critique de  $f$  au seuil 5 % : 1,5 environ.

La séparation en deux paires des chromosomes classés en 11 et 12 est justifiée par ces résultats puisque dans ce cas le test est très significatif. Par contre, pour les paires du groupe 1, il n'apparaît de différences significatives que pour les paires 2 et 4.

Il convient de faire les remarques suivantes sur la méthode adoptée; elle nécessite l'hypothèse de normalité et l'hypothèse d'égalité des matrices

de covariance mais elle peut permettre un classement aussi objectif que possible (si l'on choisit d'éloigner les paires dont la différence est la plus significative) et ne favorise par l'une des coordonnées par rapport à l'autre.

## CONCLUSION

A l'heure actuelle, le problème de l'identification des paires chromosomiques chez le pleurodèle est presque résolu par l'emploi de la technique de marquage de Callan (4). Cependant, le problème reste posé pour toutes les espèces animales où un tel marquage ne peut être effectué.

On peut alors, dans de tels cas, préconiser la méthode statistique précédente d'analyse multidimensionnelle pour apprécier la validité du classement.

L'appariement objectif des chromosomes pourrait être également précisé grâce à l'introduction d'une troisième coordonnée qui serait le dosage de la teneur des acides nucléiques de ces chromosomes selon la technique de Bailly (2). Dans ce cas, on ne pourrait plus utiliser une distribution de Fisher, mais une étude asymptotique (pour  $n$  grand) serait possible. On pourrait, par exemple, prendre comme variable aléatoire :

$$S = \frac{-\sqrt{n} \text{ Log } U - \sqrt{n} \text{ Log } K}{\sqrt{\frac{2p(K-1)}{K}}}$$

qui converge vers une variable aléatoire  $N(0,1)$  sous l'hypothèse nulle. Si le nombre d'épreuves ( $n$ ) est grand, la valeur observée de cette variable aléatoire sera donc comparée à la valeur correspondant à la loi normale réduite au seuil utilisé.

## REFERENCES

- [1] ANDERSON, 1958 - An introduction to multivariate statistical analysis John Wiley.
- [2] BAILLY S., 1967 - Etude cytophotométrique de la teneur en acides nucléiques des Chromosomes métaphasiques de l'Amphibien Urodèle *Pleurodeles waltlii* Michah. Exp. Cell. Res., t. 46, n° 3, p. 549-556.
- [3] BEETSCHEN J.C. et JAYLET A., 1961 - Caryotype somatique de l'Amphibien Urodèle *Pleurodeles waltlii* Michah. C. R. Acad. Sc. Paris, t. 262, p. 122-125.
- [4] CALLAN M.G., 1966 - *Ambystoma mexicanum*. J. Cell. Sci., t. 1, n° 1, p. 85-108.
- [5] GALLIEN C.L., 1966 - Comparaison des caryotypes de *Pleurodeles poireti* et *waltlii* C.K. Acad. Sc. Paris, T. 262, n° 1, p. 122-125.

- [6] KESSOUS et CAUSSINUS, 1968 - Essai d'analyse biométrique du caryotype de l'Amphibien Urodèle *Salamandra salamandra* L. *Chromosoma*.
- [7] LABORIE, 1971 - Recherches sur un problème d'analyse multidimensionnelle. Thèse 3ème cycle Toulouse.
- [8] SCHEFFE, 1961 - The analysis of variance. John Wiley.
- [9] WILKS, 1962 - Mathematical statistics. John Wiley.