

1972. MARTOJA (R.) et SEUREAU (C.). — Répartition des accumulations de métaux et de déchets puriques chez *Calliphora erythrocephala* (Diptère, Brachycère). *C.R. Acad. Sc.*, Paris, D, 274, 1534-1537.
1970. PETIT (J.). — Sur la nature et l'accumulation de substances minérales dans les ovocytes de *Polydesmus complanatus* (L.) (Myriapode Diplopode). *C.R. Acad. Sc.*, Paris, D, 270, 2107-2110.
1961. RAZET (P.). — Recherches sur l'uricolyse chez les Insectes. *Thèse*, Imp. bretonne éd., Rennes.

**ÉTUDE BIOMÉTRIQUE DES POPULATIONS
DE *PYRGUS CARLINAE* Rbr.
ET DE *P. CIRSI* Rbr. (LÉPID. HESPÉRIIDAE)**

**I. — MÉTHODE SIMPLE D'ÉVALUATION
DU TAUX DE RECOUVREMENT
DES DISTRIBUTIONS STATISTIQUES DE DEUX POPULATIONS
EN RELATION AVEC LA NOTION DE DISTANCE**

par

Michel GUILLAUMIN

Laboratoire de Zoologie, Université de Paris-VI

Manuscrit reçu le 20 juin 1972

RÉSUMÉ

Le calcul de la distance géométrique d entre les centres de distribution de deux populations permet, si l'on part des données centrées réduites, de définir

- d'une part le taux d'individus mal classés en séparant arbitrairement les distributions à la valeur $d/2$;

- d'autre part le taux de non recouvrement si l'on fixe des limites aux distributions comparées.

Deux tables donnent respectivement, en fonction de la valeur à la distance

I. Le taux d'individus mal classés.

II. Le taux de non recouvrement pour trois seuils : 95; 99; 99,9 p. cent.

Une troisième table donne l'intervalle de confiance et le seuil de signification de d en fonction des effectifs des populations pour différents seuils, 95; 99; 99,9 p. cent.

Il est montré que ces tables peuvent être utilisées, aussi bien si la distance est calculée à partir d'un seul caractère qu'à partir de plusieurs caractères indépendants, à partir de la relation

$$d = \sqrt{\sum_{i=1}^n d^2 i}$$

Elles peuvent également être utilisées dans le cas de plusieurs caractères intercorrélés à condition que la distance soit calculée par la méthode du D^2 de Mahalanobis.

SUMMARY

The calculation of the geometrical distance between the central points of two populations by mean of standardized data allows us to definite :

— first, the rate of misclassified individuals when the distributions are arbitrarily separated at $d/2$;

— and the percentage of symmetrical nonoverlap, if end points are defined in the compared distributions.

The percentage of misclassified individuals and the percentage of nonoverlap at three levels (95; 99; 99,9 p. cent) are stated in function of d value and summarized in two tables.

Table III shows the discriminant score and d level of significance in terms of importance of the populations at several rates (95; 99; 99,9 p. cent).

The formula

$$d = \sqrt{\sum_{i=1}^n d^2_i}$$

shows out that the tables may be used when the distance is calculated from a single character as well as it is when calculated from several characters.

It is verified too when several intercorrelated characters are used, if the distance is evaluated by mean of Mahalanobis's D^2 . So Mahalanobis's method is given a new application.

In the case of independent characters, a simple discriminative function allows us to situate every individual as regards to the overlapping.

INTRODUCTION

Il peut paraître anachronique, à une époque où les ordinateurs, de plus en plus puissants et rapides, permettent l'utilisation des techniques statistiques les plus élaborées, de proposer une méthode simple et, par là même, restreinte dans ses applications, pour comparer deux populations, deux échantillons. L'utilisation d'un ordinateur suppose la constitution d'un fichier, sa perforation; il faut ensuite avoir accès à un ordinateur, disposer d'un programme adapté au problème traité, utilisable par le calculateur. Une telle mise en œuvre, qui nécessite souvent de laborieuses mises au point, se justifie

pleinement si la masse de données à exploiter est très importante, le nombre des variables élevé et les problèmes envisagés particuliers.

Or il arrive souvent, au cours d'un travail de systématique, que l'on se trouve en face d'un problème de spéciation, de distinction entre sous-espèces un peu délicat, pour lequel aucun des caractères envisagés, pris isolément, ne donne de réponse satisfaisante, alors que cette réponse serait donnée par une combinaison judicieuse d'un certain nombre d'entre eux; le problème relève alors de l'analyse multivariate et de nombreux chercheurs, faute de moyens matériels ou rebutés par la difficulté des méthodes, ne vont pas plus avant.

La méthode proposée ici ne prétend pas apporter la solution à tous les problèmes; elle ne s'applique que dans des conditions précises.

— Comparaison de populations deux à deux.

— pour un caractère;

— pour un ensemble de n caractères, à condition que chacun d'eux soit distribué normalement et que toutes les intercorrélations entre eux soient nulles (sinon numériquement du moins statistiquement).

Le problème, en systématique évolutive, revient le plus souvent à définir la proportion d'individus de chaque population qui peut être identifiée et classée sans ambiguïté, puisque la définition quantitative du niveau subsppécifique est basée sur ce critère.

Cela nous amènera donc à envisager les définitions quantitatives du niveau subsppécifique d'une part et, d'autre part, à préciser les manières d'envisager le chevauchement de deux distributions.

Enfin il nous a paru utile de donner un moyen de localiser chaque individu et en particulier de le situer par rapport au chevauchement; fait-il partie des individus mal classés ou au contraire des individus identifiables? Dans le cas d'un seul caractère, la solution est immédiate; mais, lorsque plusieurs variables sont envisagées, il faut définir une nouvelle variable, combinaisons des variables de départ, qui est, en quelque sorte, une fonction discriminante, elle aussi simplifiée, car, nous le verrons, il n'y a pas lieu de chercher à optimiser la discrimination.

Il ne suffit pas en effet de définir des coupures taxonomiques; encore faut-il avoir un moyen de détermination d'identification qui

permette de ranger un individu quelconque dans l'une ou l'autre des catégories définies.

Les calculs des distances et des fonctions λ peuvent s'exécuter très facilement sur des machines électromécaniques de bureau.

I. — DÉFINITION QUANTITATIVE DU NIVEAU SUBSPÉCIFIQUE

Deux conceptions sont couramment adoptées.

- La sous-espèce à 75 p. cent (AMADON, 1949; MAYR, LINSLEY, USINGER, 1953). Deux populations sont considérées comme appartenant à deux sous-espèces différentes si 75 p. cent des individus de chacune est distinct de l'ensemble de l'autre.
- La sous-espèce à 100 p. cent (WARREN, 1936). Seules sont considérées comme appartenant à deux sous-espèces distinctes des populations dont tous les individus de l'une sont distinguables de l'ensemble des individus de l'autre.

En fait la définition originale de la règle des 75 p. cent est la suivante :

« Deux populations sont considérées comme appartenant à deux sous-espèces différentes si 75 p. cent de l'une sont distinctes de la totalité de l'autre. La situation n'étant pas réciproque. » Cette position ne semble pas logique; il est difficile d'admettre que l'on puisse être conduit à adopter une distinction subsppécifique en comparant A à B et à la rejeter en comparant B à A ! De plus, une certaine ambiguïté dans la formulation des pourcentages relatifs a entraîné une certaine confusion dans les modalités d'application de cette règle (GÉRY, 1963). Actuellement elle est relativement peu employée. Les Lépidoptéristes, en particulier, lui préfèrent la définition à 100 p. cent.

II. — ESTIMATION DU TAUX DE CHEVAUCHEMENT PAR LA « DISTANCE »

A) Notion de distance.

La difficulté dans l'évaluation du taux de recouvrement réside dans la dissymétrie de la figure d'intersection des deux courbes de Gauss (différentes par leur hauteur, leur étalement). En toute rigueur, il faudrait déterminer avec précision le point d'intersection des deux courbes (MAYR, LINSLEY, USINGER), ce qui conduit ces auteurs à donner une solution approchée sous forme du « coefficient de différence » (CD) (1) et une table de correspondance entre la valeur du « C.D. » et le taux de chevauchement.

La difficulté de la dissymétrie n'est qu'apparente et peut être très facilement levée. En pratique, ce que l'on compare, ce sont des courbes ajustées à des histogrammes expérimentaux. L'étalement des courbes traduit la variabilité du caractère dans chaque population. La première chose à faire est de tester si les variabilités sont significativement différentes dans les deux échantillons (2); si oui, la comparaison des chevauchements n'a pas de sens; sinon, on peut admettre, malgré la petite différence, qu'elle est la même dans les deux lots et calculer la variabilité moyenne.

Les ordonnées traduisent des fréquences et les différences de hauteurs des courbes des différences d'effectifs.

Si, au lieu de considérer les données brutes, on envisage les données centrées réduites (par rapport à la moyenne générale et l'écart-type moyen (3)), les courbes comparées sont les courbes normales réduites ayant même étalement caractérisé par $\sigma = 1$. Les

(1) Le coefficient de différence est :

$$C.D. = \frac{\bar{X} A - \bar{X} B}{\sigma A + \sigma B}$$

(2) Par un test d'homogénéité des variances.

(3) La définition et le mode de calcul de l'écart-type moyen sont rappelés plus loin.

ordonnées qui représentent des densités de probabilité, indépendantes de l'effectif, sont donc égales dans les deux courbes. La figure d'intersection devient alors symétrique et la comparaison réciproque possible.

L'ampleur du chevauchement dépend directement de l'écart entre les moyennes, c'est-à-dire de la distance linéaire qui sépare les points moyens sur le graphique. Cette distance d est égale à la différence entre les moyennes exprimées en écart-type (1).

$$d = \frac{\bar{X}_A - \bar{X}_B}{\sigma_G}$$

ou, expression équivalente :

$$d = \bar{x}_A - \bar{x}_B$$

Cette notion de distance ainsi définie a déjà été utilisée par TEISSIER (1958) dans la comparaison de souches de Drosophiles françaises

(1) Par convention, nous désignerons, dans la suite du texte, les données originales par des majuscules et les données centrées réduites par des minuscules

$$x_i = \frac{X_i - \bar{X}_G}{\sigma_G}$$

La distance, telle que nous la calculons ici, est exprimée en écarts-types; elle est donc désignée par un d minuscule; cependant, dans les tableaux, la distance est désignée par un D majuscule qui exprime par convention la distance généralisée de MAHALANOBIS.

Evaluation de l'écart-type moyen intrapopulation σ_G .

A l'intérieur de chaque population, l'écart-type est donné par les relations : dans la population A,

$$\sigma_A = \sqrt{\frac{\sum (X_i - \bar{X}_A)^2}{N_A - 1}}$$

dans la population B,

$$\sigma_B = \sqrt{\frac{\sum (X_i - \bar{X}_B)^2}{N_B - 1}}$$

L'évaluation la plus rigoureuse de l'écart-type moyen intrapopulation σ_G est :

$$\sigma_G = \sqrt{\frac{\sum (X_i - \bar{X}_A)^2 + \sum (X_i - \bar{X}_B)^2}{N_A + N_B - 2}}$$

Dans le cas où les deux populations ont le même effectif, l'expression de σ_G se simplifie. σ_G est égal à la racine carrée de la demi-somme des variances. Si $N_A = N_B = N$

$$\sigma_G = \sqrt{\frac{\sum (X_i - \bar{X}_A)^2}{2(N-1)} + \frac{\sum (X_i - \bar{X}_B)^2}{2(N-1)}} = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$$

et japonaises. Le degré de recouvrement de deux distributions peut être formulé de deux manières.

- Soit par la proportion d'individus mal classés.
- Soit par la proportion d'individus de chaque population extérieure au chevauchement, donc classés avec « certitude ».

B) Proportion d'individus mal classés.

Si la séparation entre les deux populations est fixée au niveau de l'intersection des courbes (à la demi-distance : $d/2$ (Fig. 1)), une partie des individus de chaque population (zone hachurée) sera l'objet d'une erreur de classement; elle est égale à l'aire située sous la courbe extérieure à l'abscisse de la coupure.

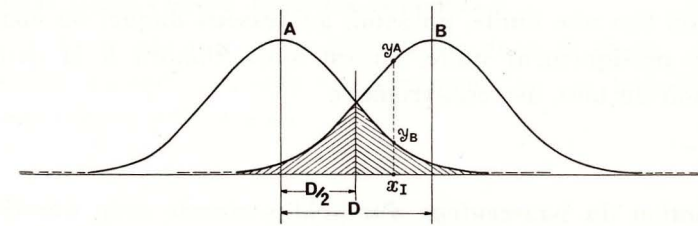


Fig. 1. — Chevauchement de deux distributions centrées normées. La zone hachurée représente la proportion d'individus mal classés en fixant la séparation entre les populations à la valeur $D/2$.

La table I donne les valeurs du taux d'erreur de classement pour différentes valeurs de d .

TABLE I

Pourcentage d'individus mal classés dans chaque population (aires hachurées de la figure 1) quand la coupure entre les deux distributions est faite à la valeur $D/2$.

D	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.	50,00	48,00	46,61	44,04	42,07	40,13	38,21	36,32	34,46	32,64
1.	30,85	29,12	27,42	25,78	24,20	22,66	21,18	19,77	18,40	17,11
2.	15,87	14,67	13,56	12,51	11,51	10,56	9,68	8,85	8,07	7,35
3.	6,68	6,05	5,48	4,95	4,46	4,00	3,59	3,21	2,87	2,56
4.	2,27	2,02	1,79	1,58	1,39	1,22	1,07	0,93	0,82	0,71
5.	0,62	0,54	0,46	0,40	0,35	0,30	0,26	0,22	0,19	0,16
6.	0,13	0,11	0,09	0,08	0,07	0,06	0,05	0,04	0,033	0,028

Il est bien évident, puisque les distributions s'étendent théoriquement jusqu'à l'infini, qu'aucun individu ne peut être affecté avec certitude à l'une ou à l'autre des deux populations.

A chaque individu I est lié une probabilité d'erreur de classement. Ainsi la probabilité P pour qu'un individu I, d'abscisse x_i , classé dans B du fait de la coupure arbitraire, appartienne à la population A est égale à :

$$P = \frac{y_A}{y_A + y_B}$$

où y_A et y_B représentent les densités de probabilité correspondant au point x_i dans les distributions A et B (1).

P est égal à 1/2 au point d'intersection des deux courbes et tend vers zéro à mesure que x_i s'accroît sans, théoriquement, être jamais nulle. Si on fixe une limite, un seuil, au-dessous duquel on considère P comme pratiquement nulle, on est alors conduit à la deuxième formulation du taux de recouvrement.

C) Evaluation du pourcentage d'individus classés avec « certitude ».

En pratique, les populations naturelles et expérimentales sont limitées et ne montrent jamais toute l'étendue de la variation théorique. On peut donc, en admettant *a priori* une marge d'erreur définie, fixer les limites de telles populations en tronquant les extrémités de la distribution. Deux types de coupure sont couramment admis.

— La population standard de SIMPSON englobe 99,88 p. cent de la population théorique; sa distribution s'étend à 3,24 σ de part et d'autre de la moyenne. Dans ce cas 0,12 p. cent des individus sont ignorés, ce qui entraîne, dans une comparaison avec une autre popu-

(1) Rappelons que, dans la distribution normale centrée réduite, la densité de probabilité, à l'abscisse x , est égale à l'ordonnée correspondante :

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

lation, une erreur systématique de classement de 0,06 p. cent, soit 6 individus sur dix mille.

— La population de MAYR, LINSLEY, USINGER englobe 94 p. cent de la population théorique; sa distribution s'étend à 1,88 σ de part et d'autre de la moyenne; le taux d'erreur systématique de classement est alors de 3 p. cent.

Dans ces comparaisons, MAYR et al. ne tronquent que l'extrémité intéressée par le chevauchement, de sorte que leur population contient 97 p. cent de la population théorique totale (99,94 p. cent dans le cas de la population standard). Comme l'on peut être amené à comparer une même population avec une plus petite et avec une plus grande, il n'est pas logique de n'amputer qu'une seule extrémité de la distribution et il faut admettre la même erreur systématique aux deux extrémités.

Quels seuils choisir pour les coupures ?

Afin de ne pas introduire d'arbitraire dans ce choix, je propose de prendre les seuils utilisés classiquement par les biologistes dans les comparaisons statistiques.

— *Seuil à 95 p. cent.*

Coupures situées à 1,960 σ .

Taux systématique d'erreur de classement : 5 p. cent.

— *Seuil à 99 p. cent.*

Coupures situées à 2,575 σ .

Taux systématiques d'erreur de classement : 1 p. cent.

— *Seuil à 99,9 p. cent.*

Coupures situées à 3,290 σ .

Taux systématique d'erreur de classement : 1 p. mille.

On remarquera que la population standard de SIMPSON (99,88 p. cent) est très proche de ce dernier seuil, on peut donc assimiler la population standard et le seuil à 99,9 p. cent sans risque d'erreur appréciable.

La table II, établie à partir de la table de la distribution normale de PEARSON, donne, pour chacun des trois seuils retenus, le taux d'individus de chaque population classés avec certitude en fonction de la distance d séparant les deux moyennes.

TABLE II

Evaluation du taux de non-recouvrement (pourcentage d'individus classés avec certitude en fonction de la distance entre deux populations).

Les distributions sont centrées sur la moyenne générale des deux populations.

$$\bar{X}_m = \frac{\bar{X}_A + \bar{X}_B}{2}$$

et réduites par l'écart-type moyen commun intra-population : la distance est exprimée en écart-type moyen σ_m

$$D = \frac{\bar{X}_A - \bar{X}_B}{\sigma_m}$$

Seuil à 95 p. cent. On admet que la population expérimentale représente 95 p. cent de la population théorique. Sa distribution s'étend à $S = 1,96 \sigma$ de part et d'autre de la moyenne. Les chiffres du tableau indiquent, pour chaque valeur de D (de $1/10^e$ en $1/10^e$), la fraction de chaque population extérieure au chevauchement (aire hachurée, Fig. 2 A). Pour une distance $D = 2.S = 2 \times 1,96 = 3,92$, l'aire sous la courbe correspond à 95 p. cent de la population théorique, soit à la totalité de la population expérimentale. Au seuil à 95 p. cent, deux populations seront entièrement distinctes dès qu'une distance, $D = 3,92 \sigma$ les sépare.

Seuil à 99 p. cent. La population expérimentale englobe 99 p. cent de la population théorique, dont les extrémités sont tronquées à $S = \pm 2,575 \sigma$.

On remarquera que deux populations expérimentales sont distinctes à 100 p. cent dès que la distance qui les sépare atteint $D = 5,150$. (L'aire sous la courbe représente 99 p. cent de la population théorique).

Seuil à 99,9 p. cent. Population standard de SIMPSON. La coupure des « queues » de distribution sont situées à $S = \pm 3,24$. (L'aire sous la courbe représente alors 99,88 p. cent, # 99,9 p. cent, de la population théorique). La discrimination à 100 p. cent des populations expérimentales est assurée dès que la distance atteint $D = 2 \times 3,24 = 6,48 \sigma$.

Seuil à 95 p. cent.

D	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.	0	0,64	1,42	2,34	3,44	4,71	6,19	7,88	9,80	11,96
1.	14,35	16,99	19,86	22,96	26,27	29,77	33,44	37,24	41,14	45,11
2.	49,10	53,07	56,98	61,41	64,51	68,04	71,39	74,54	77,46	80,14
3.	82,58	84,79	86,75	88,49	90,01	91,33	92,45	93,41	94,22	94,88
4.	95,44	100	-	-	-	-	-	-	-	-

Seuil à 99 p. cent.

D	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.	0,0	0,16	0,38	0,64	0,98	1,40	1,91	2,54	3,29	4,19
1.	5,26	6,51	7,95	9,61	11,50	13,61	15,97	18,57	21,41	24,48
2.	27,76	31,24	34,88	38,66	42,55	46,51	50,50	54,48	58,41	62,24
3.	65,96	69,52	72,91	76,08	79,03	82,75	84,24	86,47	88,48	90,24
4.	91,79	93,14	94,29	95,27	96,10	96,79	97,36	97,82	98,20	98,50
5.	98,74	98,92	99,07	100	-	-	-	-	-	-

Seuil à 99,9 p. cent.

D	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0.	0	0,02	0,06	0,16	0,25	0,35	0,49	0,67	0,90	1,19
1.	1,56	2,00	2,55	3,23	4,03	4,99	6,12	7,43	8,95	10,69
2.	12,65	14,86	17,30	19,98	22,90	26,05	29,40	32,93	36,63	40,45
3.	44,37	48,34	52,33	56,30	60,20	64,00	67,07	71,17	74,48	77,58
4.	80,46	83,09	85,49	87,64	89,56	91,25	92,73	93,95	95,10	95,46
5.	96,02	96,80	97,44	97,97	98,41	98,75	99,03	99,25	99,42	99,55
6.	99,65	99,73	99,79	99,83	99,86	99,88	100	-	-	-

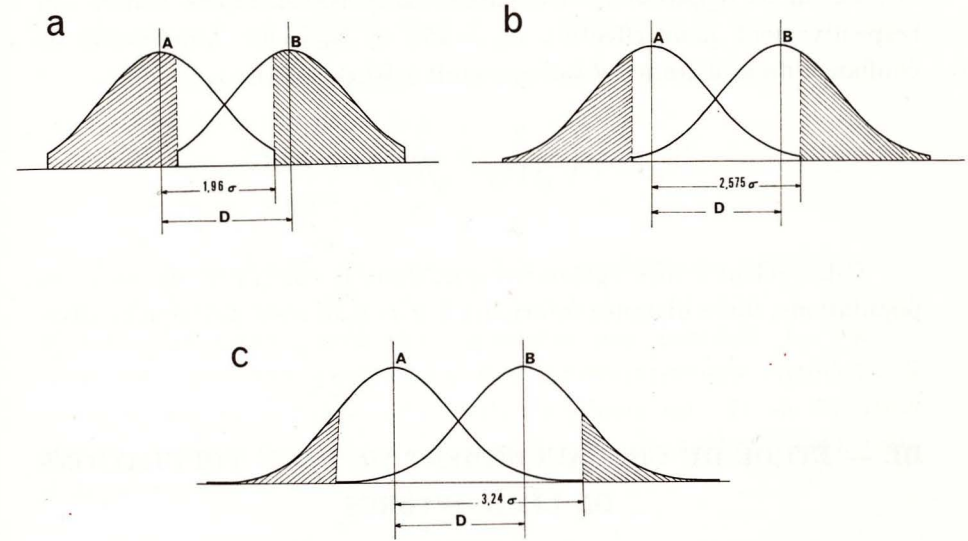


FIG. 2. — Les trois modalités de troncature des queues de distribution. A, seuil à 95 p. cent. B, seuil à 99 p. cent. C, seuil à 99,9 p. cent.

D) Intervalle de confiance de la distance.

La confiance à accorder à l'estimation d'une distance d dépend de la précision avec laquelle on connaît les moyennes de chacune des deux populations comparées; elle dépend donc en dernier ressort de l'effectif des échantillons.

Une estimation facile et rapide de l'intervalle de confiance peut être obtenue par les relations :

$$d_{max} = d + \left(\frac{t_A}{\sqrt{N_A}} + \frac{t_B}{\sqrt{N_B}} \right)$$

$$d_{min} = d - \left(\frac{t_A}{\sqrt{N_A}} + \frac{t_B}{\sqrt{N_B}} \right)$$

déduite de la définition de l'intervalle de confiance d'une moyenne : $\pm t \cdot e_x$; où : « t » est la statistique de FISHER et e_x l'erreur sur la moyenne (pour les échantillons dont l'effectif est supérieur à 30, $t \neq 2$.)

Les deux populations comparées dans les exemples traités ont respectivement pour effectif : $N_A = 153$ et $N_B = 86$. L'intervalle de confiance de la distance d est (au seuil à 95 p. cent) :

$$I_{cd} = \pm \left(\frac{2}{\sqrt{153}} + \frac{2}{\sqrt{86}} \right) = \pm 0,38$$

Cela revient à dire également que, dans le cas précis de ces deux populations, toute distance inférieure à $d = 0,38$ n'est pas significative.

III. — ÉTUDE DU CHEVAUCHEMENT DE DEUX POPULATIONS DE LÉPIDOPTÈRES

L'étude de la spéciation dans le genre *Pyrgus* et plus particulièrement l'étude des populations de la zone de contact entre espèces étroitement apparentées (*P. carlinae* et *P. cirsii*) nous ont amené à comparer les distributions d'un certain nombre de caractères dans deux populations considérées comme représentatives des deux espèces.

Un certain nombre de ces caractères sont utilisés en systématique traditionnelle (WARREN, 1926). Certains sont qualitatifs comme l'habitus (le faciès général) ou l'état libre ou non de la plaque centrale de la valve des genitalia. Ce dernier caractère est certainement le plus discriminant, mais il n'est pas absolu; il peut être envisagé sous une forme quantitative, mais, étant donné la non normalité de sa distribution, nous ne l'introduisons pas ici. L'expression de l'habitus, combinaison de huit caractères externes élémentaires, a également été quantifiée sous forme d'une cote individuelle établie de manière à obtenir la meilleure discrimination possible. La manière dont elle est établie sera exposée dans un autre travail.

D'autres caractères sont quantitatifs par essence : le nombre de dents aux apophyses des subunci (éléments des genitalia) est un caractère à variation discrète ; la longueur du pénis, sa largeur ou la largeur de la plaque centrale de la valve sont des caractères à variation continue.

Parmi l'ensemble des caractères étudiés, nous en avons retenu quatre.

- la largeur de la plaque centrale de la valve (LPL),
- la largeur du pénis (LIP),
- le nombre des dents des subunci (DTS),
- l'habitus (CEXT).

Ils ont été retenus en raison de leur distribution normale et de l'absence d'intercorrélation entre eux (tableau I) puisque la méthode proposée ici s'applique à des caractères non corrélés. Le cas de caractères présentant entre eux des intercorrélations significatives est traité par la méthode de la distance généralisée (D^2 de MAHALANOBIS); il sera exposé dans un travail en préparation (GUILLAUMIN et LEFÈVRE).

TABLEAU I
Intercorrélations entre les caractères LPL, LIP, DTS et CEXT.

	LPL	LIP	DTS	CEXT
LPL	I	-0,054	0,053	-0,061
LIP		I	0,020	-0,019
DTS			I	0,006
CEXT				I

A) Etude individuelle de chaque caractère.

Le pouvoir discriminant varie beaucoup d'un caractère à l'autre et l'aire de recouvrement des histogrammes de distribution des deux populations est plus ou moins importante (Fig. 3). Il est possible de déduire de l'observation directe de ces figures le pourcentage d'individus de chaque population extérieur au chevauchement (les deux dernières colonnes du tableau II) et de les comparer au taux de non recouvrement théorique (4^e colonne du tableau II) déduit de la Distance D entre les populations (3^e colonne). On remarquera que le taux de non recouvrement théorique est presque toujours très voisin de la moyenne des valeurs observées dans les deux populations. Dans l'ensemble, on constate une bonne concordance entre les taux de non recouvrement théorique et observé.

TABLEAU II

Taux de recouvrement pour quatre caractères considérés individuellement.

Pour chaque caractère, la valeur de la distance D est accompagnée des limites de l'intervalle de confiance; et le taux de non recouvrement théorique par les valeurs correspondant aux limites.

Caract.	Moyennes de population		Moyenne générale		d	Taux de non recouvrement au seuil à 99 p. cent			
	\bar{x}	σ	\bar{x}_G	σ_G		Théorique	Observé		
							carlinae	cirsii	
LPL	ca	33,46	2,60	28,625	2,53	4,38	96,1	94	82
	ci	23,79	2,47						
LIP	ca	20,90	2,14	19,235	2,04	2,12	31,3	26,4	11,2
	ci	17,57	1,95						
DTS	ca	8,38	1,84	10,22	1,82	2,51	46,5	17,6	40,6
	ci	12,06	1,81						
CEXT	ca	3,33	1,39	6,705	1,33	5,56	100,0	99,4	82,6
	ci	10,08	1,27						

B) Combinaison de deux caractères; cas d'une distribution bivariate.

Graphiquement une telle distribution se représente dans un espace plan défini par deux axes rectangulaires. Les caractères étant non corrélés et les données de base centrées, réduites, les individus de chaque population se distribuent en nuages pratiquement circulaires. La limite de dispersion, correspondant au seuil à 99 p. cent est un cercle de $2,575 \sigma$ de rayon, centré sur le point moyen de chaque population (Fig. 4).

Si on porte sur une troisième dimension, perpendiculaire au plan, les densités de probabilité combinées des deux caractères, la distribution bivariate est un volume en forme de cloche, dont 99 p. 100 se situe au-dessus du cercle de base de $2,575 \sigma$ de rayon.

La distance entre les centres des deux distributions est, dans le cas des données brutes originales :

$$d_{xy} = \sqrt{d_x^2 + d_y^2} = \sqrt{\frac{(\bar{X}_A - \bar{X}_B)^2}{\sigma_{G_x}} + \frac{(\bar{Y}_A - \bar{Y}_B)^2}{\sigma_{G_y}}}$$

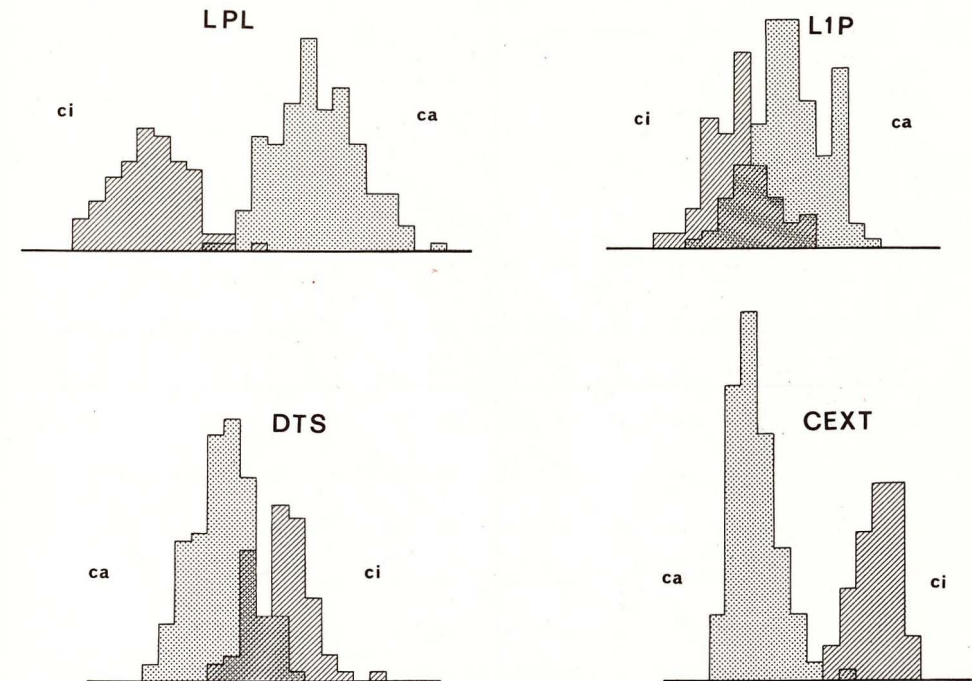


Fig. 3. — Histogrammes correspondant aux quatre caractères étudiés.

- LPL : Largeur de la plaque centrale de la valve.
- LIP : Largeur du coecum pénis.
- DTS : Nombre de dents des sub-unci.
- CEXT. : Caractères externes
- ca. : Distribution de la population de *Pyrgus carlinae*.
- ci. : Distribution de la population de *Pyrgus cirsii*.

ou dans le cas des données réduites :

$$d_{xy} = \sqrt{(2\bar{x}_A)^2 + (2\bar{y}_A)^2} = \sqrt{(2\bar{x}_B)^2 + (2\bar{y}_B)^2}$$

Ces deux quantités sont d'ailleurs égales.

C) Evaluation du taux de chevauchement de deux distributions bivariates.

Le problème revient à évaluer, dans l'une quelconque des populations, la portion de volume campaniforme située au-dessus de la zone hachurée de la figure 5. Ce volume est assez laborieux à calculer; on peut cependant l'obtenir à partir des Tables de la distribution biva-

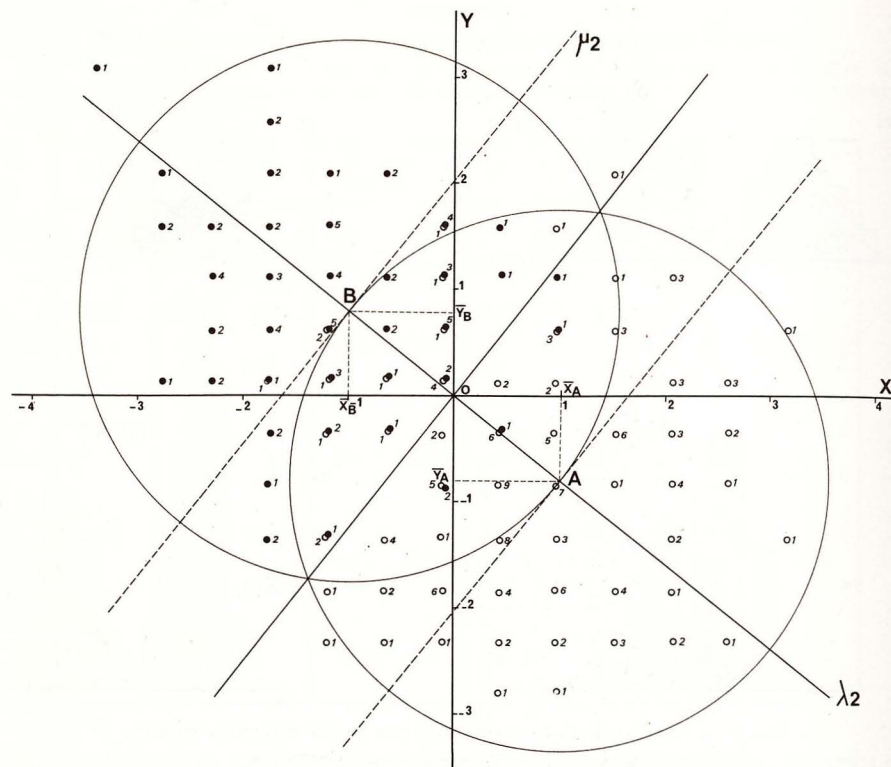


Fig. 4. — Distribution bivariate correspondant à la combinaison de deux caractères : X = LIP et Y = DTS.
 ca. : Distribution de *carlinae*, en cercles évidés.
 ci. : Distribution de *cirsiit*, en cercles pleins.
 Les chiffres situés à côté de chaque point indiquent le nombre d'individus correspondant. Les individus intéressés par le chevauchement sont ceux situés dans la partie commune aux deux cercles, pour la distribution bivariate, et entre les deux parallèles en traits tirets pour la distribution univariante de λ_2 .

riate ($\rho = 0$) de PEARSON, en le découpant en un très grand nombre de parallélépipèdes élémentaires. Dans le cas d'une distribution trivariante, le chevauchement est un volume à quatre dimensions, et son estimation, d'un point de vue pratique, devient inextricable.

Revenons au cas bivariate et envisageons la portion de la distribution correspondant à la zone ombrée (Fig. 5). Son volume peut être facilement obtenu à partir des Tables de la distribution bivariate en prenant comme système d'axes de référence λ_2 et μ ; il est égal au double du volume extérieur aux coordonnées :

$$\lambda_{2c} = d - 2,575 \sigma \text{ (seuil à 99 p. cent)}$$

et $\mu = 0$

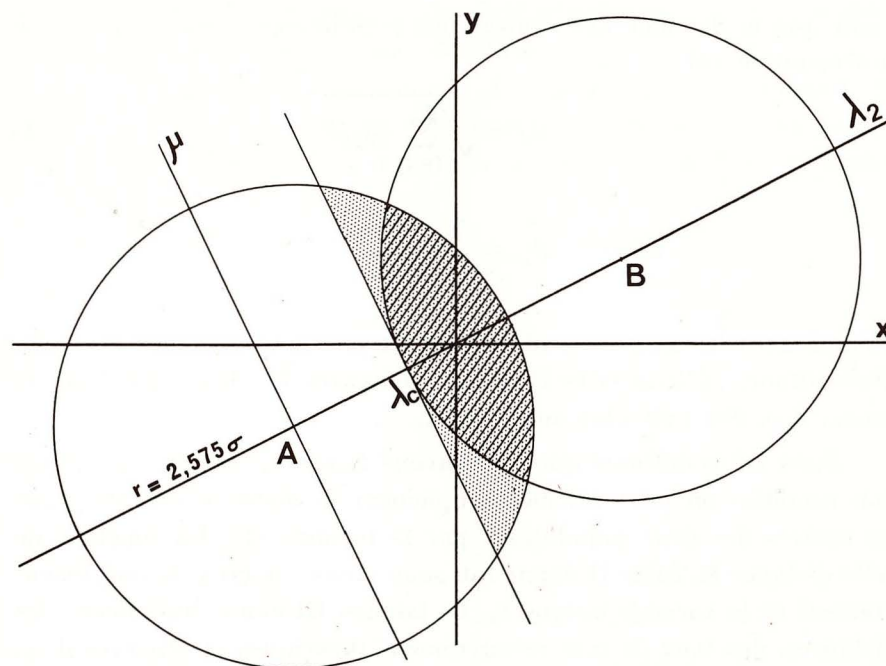


Fig. 5. — Comparaison des taux de chevauchement de deux distributions bivariates (aire hachurée) à celui de deux distributions univariates (aire ombrée).
 La : variable unique λ_2 résulte de la combinaison des variables d'origine, x et y .

On se trouve donc ramené à un cas univariate puisque l'une des variables μ est toujours nulle. Le volume ainsi défini exprime le chevauchement des distributions de la variable unique λ_2 , qui représente l'abscisse des points observation sur l'axe λ .

On montre aisément que la variance moyenne intra-population de la variable λ_2 est aussi égale à l'unité. λ_2 se trouve donc également centrée normée.

Si l'on considère que, quelle que soit l'importance du recouvrement, le chevauchement réel des distributions bivariates est toujours inférieur à celui de la variable unique λ_2 , on a intérêt, pour la simplicité des comparaisons et malgré la perte d'un peu d'information, à se ramener au cas univariate.

D) Combinaison de plusieurs caractères.

Si on envisage maintenant n caractères, non intercorrélés, et représentés dans un système de n axes orthogonaux, on montre aisément

ment que la distance des centres des populations dans cet espace à n dimensions est :

$$d_n = \sqrt{\sum_{i=1}^n (d_{xi})^2} \quad (1)$$

avec
$$d_{xi} = \frac{\bar{X}_{iA} - \bar{X}_{iB}}{\sigma_{G_{xi}}}$$

Comme dans le cas précédent, le taux de chevauchement de ces deux distributions multivariates revient à comparer les deux distributions univariates des individus sur l'axe λ_n .

Dans les conditions que nous avons fixées au départ (caractères non corrélés) on peut calculer simplement la distance linéaire entre les centres des deux populations par la formule (1). En fonction de cette distance la Table II donne (au seuil choisi) le taux de non recouvrement de la variable unique λ_n . Le tableau III donne les valeurs des distances, des taux de non recouvrement théoriques et observés dans le cas de deux, trois et quatre variables. Comme dans le cas des caractères

TABLEAU III
Taux de recouvrement après combinaison de deux, trois ou quatre caractères
(même remarque que pour le tableau II).

Caract.	Moyennes des populations		Moyenne générale		D	Taux de non recouvrement au seuil à 99 p. cent			
	x	σ	x_G	σ_G		Théorique	Observé		
							carlinae	cirsii	
LIP	ca	+ 1,295	1,00	0,00	1,00	3,08	69,4	43,1	55,8
DTS	ci	- 1,295	1,00						
LIP	ca	+ 2,310	1,00	0,00	1,00	5,11	98,9	99,3	97,6*
DTS	ci	- 2,310	1,00						
LIP	ca	+ 3,425	1,00	0,00	1,00	7,34	100,0	100	100
DTS	ci	- 3,425	1,00						
LPL	ci			0,00	1,00	6,85	100	100	100
CEXT	ci			0,00	1,00	6,36	100,0	100	100

* Deux individus mal classés sur les 86 de la population totale mais ils sont en dehors des limites à 99 p. cent. Ils font donc partie de la marge d'erreur systématique et d'après les conventions adoptées on pourrait mettre 100 p. cent de la population « expérimentale ».

tères pris individuellement, on constate une bonne concordance entre les taux théoriques et les taux observés.

Le taux de non recouvrement observé se déduit à l'examen des figures de dispersion (figure 4 pour la combinaison des deux caractères LIP et DTS et figure 6 pour la combinaison des trois caractères

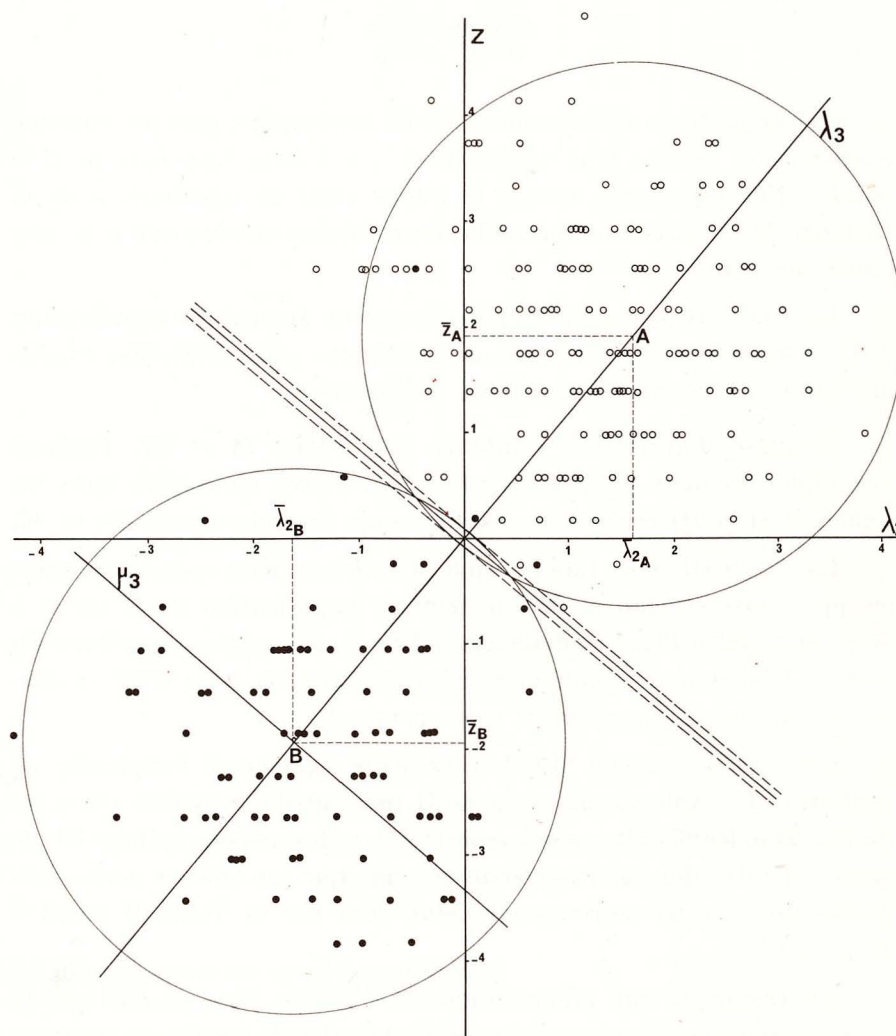


Fig. 6. — Distribution construite à partir de trois caractères : LIP et DTS = λ_{23} et LPL = z.
ca. : Distribution de *carlinae* en cercles évidés.
ci. : distribution de *cirsii* en cercles pleins.
Les zones de chevauchement sont définies comme dans le cas de la figure 4.

LIP, DTS et LPL) en dénombrant les individus de chaque population effectivement extérieurs au chevauchement.

L'intervalle de confiance à accorder à une distance, calculée avec plusieurs caractères, est la même que dans le cas d'un seul caractère.

En effet, l'expression :

$$\frac{t_A}{\sqrt{N_A}} + \frac{t_B}{\sqrt{N_B}}$$

est indépendante du nombre de variables envisagées; puisque chacune étant normée réduite, leur combinaison λ_n se trouve également normée réduite. Elle représente donc à la fois le seuil de signification de la Distance généralisée et l'intervalle de confiance à accorder à la distance calculée.

La Table III permet de calculer la valeur du seuil de signification de la Distance (ou son intervalle de confiance) pour différentes combinaisons d'effectifs des populations comparées.

Exemple d'utilisation combinée des Tables II et III. Prenons l'exemple des deux populations de Lépidoptères envisagées dans les Tableaux II et III et dont les effectifs sont respectivement 153 et 86.

La Table III nous indique que pour les valeurs correspondantes les plus voisines (150 et 90), le seuil de signification de D (seuil à 99 p. cent) est 0,492. Les distances relatives aux quatre caractères du tableau II sont toutes hautement significatives. On peut donc envisager le taux de chevauchement correspondant.

Cette même valeur, 0,492, représente également l'intervalle de confiance. Les valeurs calculées de D peuvent donc osciller entre les limites $D \pm 0,492$; elles sont reportées sur les tableaux II et III de part et d'autre des valeurs calculées, ainsi que les valeurs correspondantes du taux de non recouvrement, lues sur la Table II (seuil à 99 p. cent).

On remarque que l'imprécision est d'autant plus grande que la distance calculée est plus petite, et l'intervalle de confiance d'autant plus grand que les effectifs des échantillons sont plus faibles. Constatation qui doit mettre en garde et inciter à la prudence dans l'interprétation des résultats.

TABLE III

Valeur de la quantité t/\sqrt{N} représentant le demi-intervalle de confiance du point moyen d'une population d'effectif N.

Eff.	95 %	99 %	99,9 %	Eff.	95 %	99 %	99,9 %
2	8,985	45,019	450,225	60	0,258	0,343	0,446
3	2,484	5,730	18,209	70	0,239	0,317	0,411
4	1,591	2,920	6,462	80	0,223	0,257	0,382
5	1,241	2,059	3,850	90	0,210	0,279	0,359
6	1,049	1,646	2,804	100	0,200	0,264	0,340
7	0,925	1,401	2,252	150	0,161	0,213	0,275
8	0,836	1,237	1,912	200	0,140	0,185	0,238
9	0,768	1,118	1,680	250	0,125	0,165	0,213
10	0,715	1,027	1,512	300	0,114	0,151	0,194
12	0,635	0,896	1,280	350	0,105	0,139	0,180
14	0,577	0,805	1,128	400	0,099	0,130	0,168
16	0,532	0,736	1,018	450	0,093	0,123	0,158
18	0,497	0,683	0,934	500	0,088	0,116	0,149
20	0,468	0,639	0,868	600	0,080	0,106	0,136
25	0,412	0,559	0,749	700	0,074	0,098	0,125
30	0,373	0,503	0,668	800	0,070	0,091	0,117
35	0,343	0,461	0,608	900	0,066	0,086	0,110
40	0,319	0,427	0,561	1 000	0,062	0,081	0,104
45	0,300	0,401	0,524	5 000	0,028	0,036	0,046
50	0,284	0,379	0,493	10 000	0,019	0,025	0,032

L'intervalle de confiance, I_{c_d} de la distance d entre deux populations P_1 et P_2 , d'effectif respectifs N_1 et N_2 s'obtient par la relation :

$$I_{c_d} = \pm (t_1/\sqrt{N_1} + t_2/\sqrt{N_2})$$

Exemple : l'intervalle de confiance de la distance d entre deux populations P_1 et P_2 , d'effectif respectif 150 et 90, est, au seuil de 99 p. cent

$$I_{c_d} = \pm (0,213 + 0,279) = \pm 0,492.$$

IV. — CALCUL DE LA FONCTION λ , PROJECTION SUR L'AXE JOIGNANT LES CENTRES DE DISTRIBUTION

A) Combinaison de deux caractères.

Le calcul de l'abscisse centrée réduite λ_{2i} de la projection de l'individu I sur l'axe λ_2 est simple (fig. 7).

On montre aisément que

$$\lambda_{2i} = x_i \cos.\alpha. + y_i \sin \alpha.$$

l'angle α traduisant l'inclinaison de l'axe λ_2 sur l'axe des x est défini par le rapport des coordonnées centrées réduites de chaque population.

$$\text{tg. } \alpha = \frac{\bar{y}_A}{\bar{x}_A} = \frac{\bar{y}_B}{\bar{x}_B}$$

$\text{tg } \alpha$ peut être positif ou négatif suivant l'orientation de l'axe des centres; dans le cas où $\text{tg} \alpha$ est négatif, ne pas oublier que $\sin \alpha$ est également négatif.

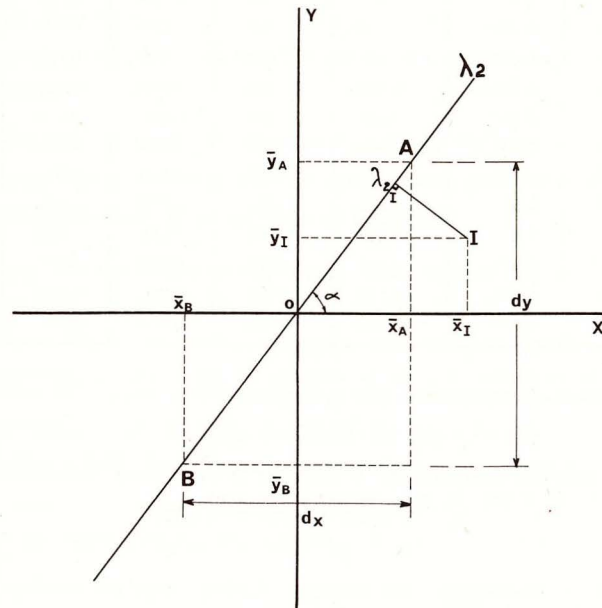


Fig. 7. — Projection d'un individu I sur l'axe λ_2 .
L'abscisse $0\lambda_{2i}$ est égale à $x_i \cos \alpha + y_i \sin \alpha$.

B) Combinaison de trois caractères et généralisation à n caractères.

La figure 8 montre que dans le cas de trois variables x, y, z , il est très facile de se ramener au cas bivariante en envisageant comme nouveau couple de variables, d'une part, la composante λ_2 établie à partir de x et y et, d'autre part, la variable z .

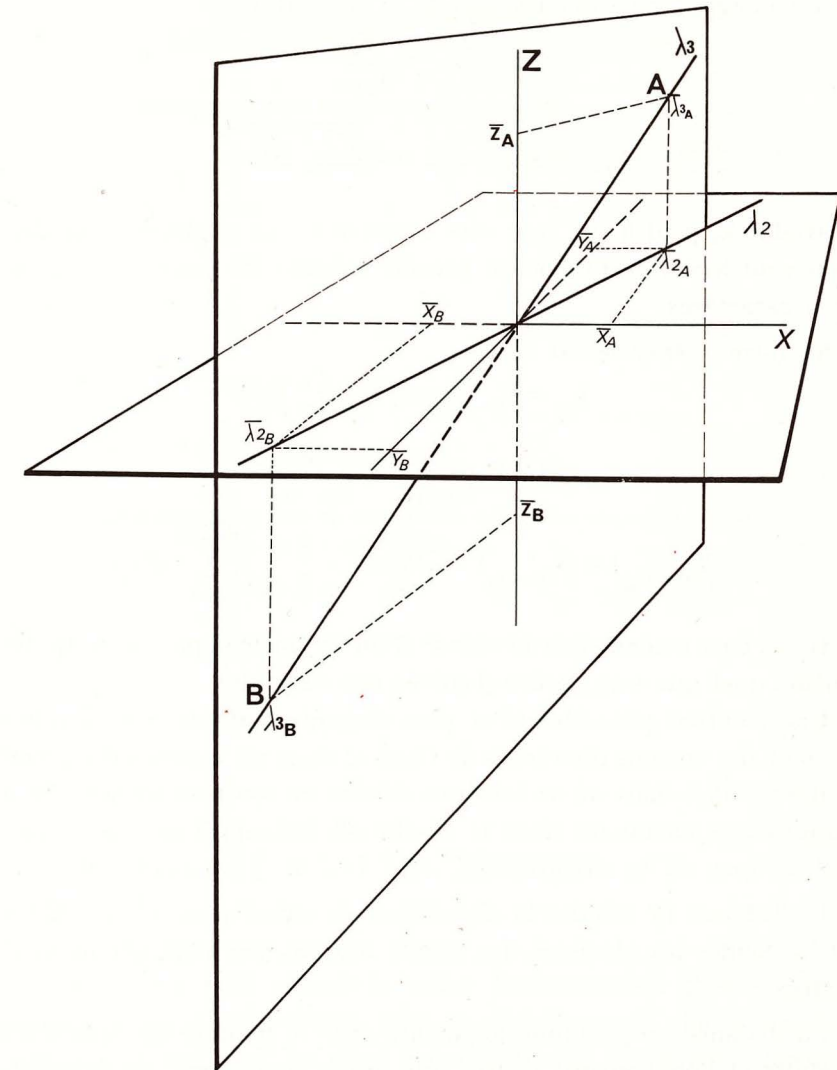


Fig. 8. — Combinaison trivariante faisant intervenir trois caractères x, y et z .
A et B sont les centres des distributions des deux populations.
Dans le plan horizontal, combinaison bivariante avec les caractères x et y . Dans le plan vertical, combinaison bivariante faisant intervenir la fonction λ_2 (combinaison de x et y) et la troisième variable z .

La composante λ_3 est obtenue par la relation

$$\lambda_{zi} = \lambda_{2i} \cos \beta + z_i \sin \beta$$

où β est l'angle formé par les axes λ_3 et λ_2 , défini par

$$\operatorname{tg} \beta = \frac{\bar{z}_A}{\bar{\lambda}_{2A}} = \frac{\bar{z}_B}{\bar{\lambda}_{2B}}$$

$$\bar{\lambda}_{2A} = \bar{\lambda}_{2B} = \frac{d_{xy}}{2} = \frac{\sqrt{d_x^2 + d_y^2}}{2}$$

La variable λ_3 peut à son tour être associée à une quatrième variable et l'on peut ainsi, de proche en proche, calculer la fonction λ_n groupant n caractères.

Sa forme générale est :

$$\lambda_{ni} = \lambda_{(n-1)i} \cos \theta + x_{ni} \sin \theta$$

avec :

$$\operatorname{tg} \theta = \frac{\bar{x}_A}{\bar{\lambda}_{(n-1)A}} = \frac{\bar{x}_B}{\bar{\lambda}_{(n-1)B}}$$

et

$$|\bar{\lambda}_{(n-1)A}| = |\bar{\lambda}_{(n-1)B}| = \frac{d_{\lambda(n-1)}}{2} = \frac{1}{2} \sqrt{\sum_{j=1}^{(n-1)} (dx_j)^2}$$

Il est bien évident que la valeur d'un λ_n quelconque est toujours la même quel que soit l'ordre d'entrée des variables.

Une solution plus directe et plus élégante consiste à rechercher le vecteur des cosinus directeurs de l'axe λ_n dans un espace orthogonal à n dimensions; mais on se retrouve devant un système linéaire de n équations à n inconnues dont la résolution fait appel au calcul matriciel, ce que, en la circonstance, nous voulons justement éviter.

Le Tableau IV montre la disposition de calcul pour les fonctions λ_2 et λ_3 . Seules les abscisses des points moyens des populations sont calculées.

La distance ainsi établie augmente avec le nombre de caractères introduits et l'on pourrait penser que, quels que soient les échantillons comparés on pourrait toujours, en introduisant un nombre suffisant de caractères, arriver à obtenir une discrimination totale entre les deux. C'est un fait; mais si le pouvoir discriminant de chaque caractère est faible, le gain sur la distance globale reste très faible. A titre d'exemple nous citerons deux cas théoriques.

Il suffit de quatre caractères discriminant chacun à 50 p. cent ($d = 2,57$) pour que leur combinaison donne une discrimination à 100 p. cent (au seuil à 99 p. cent; $d = 5,2$).

TABLEAU IV
Calcul des fonctions discriminantes λ .
Disposition des calculs pour les points moyens des populations.

Fonction d'ordre 2 avec les deux caractères LIP et DTS

x	LIP	ca : $\bar{x}_A = +0,8161$	$\operatorname{tg} \alpha = \frac{\bar{y}_A}{\bar{x}_A} = \frac{\bar{y}_B}{\bar{x}_B} = \frac{-1,0082}{0,8161}$
		ci : $\bar{x}_B = -0,8161$	$\operatorname{tg} \alpha = -1,2354$
y	DTS	ca : $\bar{y}_A = -1,0082$	$\cos \alpha = \frac{1}{\sqrt{1 + \operatorname{tg}^2 \alpha}} = 0,6293$
		ci : $\bar{y}_B = +1,0082$	$\sin \alpha = \frac{\operatorname{tg} \alpha}{\sqrt{1 + \operatorname{tg}^2 \alpha}} = -0,7771$

$$\bar{\lambda}_{2A} = \bar{x}_A \cos \alpha + \bar{y}_A \sin \alpha = 0,8161 \times 0,6293 + (-1,0082 \times -0,7771) = +1,2921$$

$$\bar{\lambda}_{2B} = \bar{x}_B \cos \alpha + \bar{y}_B \sin \alpha = -0,8161 \times 0,6293 + (-1,0082 \times -0,7771) = -1,2921$$

Fonction d'ordre 3 avec les trois caractères LIP, DTS et LPL. Les deux premiers caractères interviennent par leur composante λ_2 .

2	(LIP-DTS)	ca : $\bar{\lambda}_{2A} = +1,2921$	$\operatorname{tg} \beta = \frac{\bar{z}_A}{\bar{\lambda}_{2A}} = \frac{\bar{z}_B}{\bar{\lambda}_{2B}} = \frac{1,9110}{1,2921}$
		ci : $\bar{\lambda}_{2B} = -1,2921$	$\operatorname{tg} \beta = 1,4789$
z	LPL	ca : $\bar{z}_A = +1,9110$	$\cos \beta = 0,5625$
		ci : $\bar{z}_B = -1,9110$	$\sin \beta = 0,8268$

$$\bar{\lambda}_{3A} = \bar{\lambda}_{2A} \cos \beta + \bar{z}_A \sin \beta = +1,2921 \times 0,5625 + (1,9110 \times 0,8268) = +2,3068$$

$$\bar{\lambda}_{3B} = \bar{\lambda}_{2B} \cos \beta + \bar{z}_B \sin \beta = -1,2921 \times 0,5625 + (-1,9110 \times 0,8268) = -2,3068$$

Par contre, pour obtenir la même discrimination globale avec des caractères peu discriminants, dont la distance pour chacun serait de l'ordre de $d = 0,5$ (si les deux échantillons se composent chacun de 50 individus cette distance est juste significative), il faudrait en combiner un peu plus de 100 !

Nous verrons par ailleurs que la méthode du D^2 de MAHALANOBIS revient à faire un changement de variable de telle sorte que les nouvelles variables sont indépendantes, ce qui nous ramène au cas étudié ici; dans ces conditions la participation de 31 caractères conduit à des distances moyennes de $D = 2,50$ entre populations conspécifiques et $D = 11$ entre populations d'espèces différentes.

V. — DISCUSSION ET CONCLUSION

Sans entrer dans une comparaison détaillée, qui a déjà été faite par SOKAL et SNEATH (1962), des différentes estimations possibles de la distance, je me bornerai à envisager les estimations les plus voisines.

La distance taxonomique proposée par SOKAL :

$$\Delta_{jk} = \sqrt{\sum_{i=1}^n (\bar{x}_{ij} - \bar{x}_{ik})^2}$$

entre deux populations j et k est établie à partir de données originales et sans tenir compte des intercorrélations possibles entre deux caractères. Elle peut conduire de ce fait à une surestimation des différences.

La distance « d », utilisée par TEISSIER à propos des *Drosophiles*, est calculée sur des données centrées réduites (les données ayant été préalablement transformées en logarithmes de façon à homogénéiser les variances et à réduire l'effet de la métrique), mais sans tenir compte des intercorrélations possibles entre les différents caractères; ces derniers ayant été mesurés sur des individus différents, il n'était donc pas possible, en l'occurrence, d'évaluer leur liaison. En outre, TEISSIER a proposé une fonction discriminante très simplifiée combinant ces différents caractères, en prenant « la moyenne des écarts réduits pondérés par les différences entre les valeurs moyennes des deux séries comparées ». L'efficacité des différentes fonctions calculées est chiffrée par la valeur du taux de chevauchement.

PENROSE (1954) a proposé une estimation des différences par le calcul du coefficient C_H^2 :

$$C_H^2 = \frac{1}{n} \sum_{i=1}^n d_i^2$$

d_i étant l'écart réduit de chaque caractère. On voit immédiatement que la relation entre notre d et C_H^2 est :

$$d^2 = n C_H^2$$

PENROSE divise ce coefficient en deux composantes : une de forme et une de taille, et recherche, comme fonction discriminante, une combinaison linéaire de ces deux composantes qui maximise les différences inter-populations. Dans le même travail, il donne la relation existant entre la distance généralisée de MAHALANOBIS (D^2) et le C_H^2 , dans le cas particulier où tous les caractères envisagés ont entre eux la même intercorrélation moyenne R :

$$D^2 = \frac{\sum_{i=1}^n d_i^2 - \frac{R \left(\sum_{i=1}^n d_i \right)^2}{1 - R + Rm}}{1 - R}$$

On constate que, en se ramenant aux prémices du présent travail : $R = 0$, on a

$$D^2 = \sum_{i=1}^n d_i^2 = d^2$$

Dans le cas où tous les caractères sont indépendants, notre estimation de la distance d et la distance généralisée ($\sqrt{D^2}$) sont équivalentes. Cette constatation est importante, car tous les développements que nous avons faits sur le taux de chevauchement dans le cas de caractères indépendants peuvent être appliqués à la distance généralisée quelles que soient les intercorrélations réunissant les caractères. En effet la méthode de calcul de D^2 (RAO, 1952) consiste à faire un changement de variable qui ne modifie pas l'espace vectoriel original, mais qui transforme les variables intercorrélées d'origine en variables indépendantes. On se trouve donc ramené au cas traité ci-dessus et les tables proposées peuvent également servir à estimer le taux de chevauchement (ou de non recouvrement) de deux populations à partir de la Distance généralisée $D = \sqrt{D^2}$. Cette interprétation apporte plus d'informations que l'interprétation classique du D^2 comme test statistique assimilable à un χ^2 , en particulier dans les comparaisons de populations naturelles. Un travail en préparation (GUILLAUMIN et LEFÈVRE) utilise cette interprétation de la distance généralisée, complétée par la méthode des projections pour comparer et analyser les populations naturelles.

BIBLIOGRAPHIE

1949. AMADON (D.). — The seventy-five per cent rule for subspecies. *Condor*, 51, 250-258.
1962. GÉRY (J.). — Le problème de la sous-espèce et de sa définition statistique à propos du coefficient de Mayr-Linsley-Usinger. *Vie et Milieu*, XIII, 521-542.
1936. MAHALANOBIS (P.C.). — On the generalized distance in statistics. *Proc. Nat. Inst. Sc. India*, 12, 49.
1933. MAYR (E.), LINSLEY (E.R.) et USINGER (R.). — Methods and principles of systematic zoology. Mc Graw Hill Books, New York, 328 pp.
1954. PEARSON (K.). — Tables for statisticians and biometricians. Part II. pp. I-CCI et 1-262. Cambridge Univ. Press.
1952. RAO (C.R.). — Advanced statistical methods in biometric research. J. Wiley and Sons, New York, 390 pp.
1958. TEISSIER (G.). — Distinction biométrique des *Drosophila melanogaster* françaises et japonaises. *Ann. Gén.*, I, 2-11.
1926. WARREN (B.C.S.). — Monograph of the tribe Hesperidi (European species) (Lep. Hesp.). *Trans. Ent. Soc. London*, 74, 1-170, 60 pl. hors-texte.
1936. WARREN (B.C.S.). — Monograph of the genus *Erebia*. Londres, VII, 1-407, 104 planches.

DESCRIPTION DE L'ENDOSQUELETTE CÉPHALIQUE
CHEZ L'ISOPODE EPICARIDE
HEMIONISCUS BALANI BUCHHOLZ

par

Marie GOUDEAU

Laboratoire de Zoologie, Université de Paris VI
Station biologique de Roscoff

Manuscrit reçu le 1^{er} juillet 1972

RÉSUMÉ

Des formations endosquelettiques ont été mises en évidence au niveau de la capsule céphalique de l'Isopode Epicaride *Hemioniscus balani*. La musculature associée à ces structures particulières est étudiée.

SUMMARY

The presence of an endoskeleton has been shown in the head of *Hemioniscus balani*, Isopoda Epicaridea. The muscles associated to the peculiar framework are described.

INTRODUCTION

Le mode de vie parasitaire des Isopodes Epicarides s'est accompagné d'importantes transformations morphologiques, étroitement liées à l'éthologie souvent très particulière de ces Crustacés.

Dans une note précédente (GOUDEAU, 1969b), j'ai décrit, chez le mâle et la femelle d'*Hemioniscus balani*, la région buccale profondément modifiée en un complexe piqueur-suceur. J'ai également tenté