LES CAHIERS DE L'ANALYSE DES DONNÉES

D. BERGOUGNAN
C. COURAUD

Pratique de la discrimination barycentrique

Les cahiers de l'analyse des données, tome 7, nº 3 (1982), p. 341-354

<http://www.numdam.org/item?id=CAD_1982__7_3_341_0>

© Les cahiers de l'analyse des données, Dunod, 1982, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/

Les Cahiers de l'Analyse des Données
Vol. VII - 1982 - n° 3 - p. 341-354

PRATIQUE DE LA DISCRIMINATION BARYCENTRIQUE

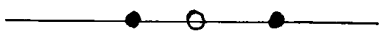
[GALETS II]

par D. Bergougnan (1)
et C. Couraud (2)

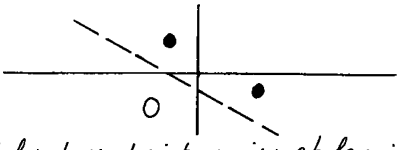
0 L'objet du présent article est l'application de la méthode d'analyse discriminante barycentrique, avec une variante nouvelle qui consiste à améliorer l'efficacité de la discrimination par une seconde analyse portant sur le sous-nuage des individus proches de la cloison mise en place lors d'une première analyse.

L'exemple auquel la méthode est appliquée est celui des galets peints aziliens, objet d'un précédent article ([*GALETS I*]) ; parmi les galets conservés dans les musées européens, 20% environ sont estimés faux. Au § 1 nous rappelons le problème général de l' analyse discriminante ; au § 2, nous résumons ce qui a été dit dans [*GALETS I*] sur le matériel étudié quant aux décors et aux formes des galets présumés vrais ; au § 3 on effectue une analyse globale portant à la fois sur les galets considérés comme vrais et ceux douteux ou certainement faux. Nous pensons que grâce aux §§ 2 et 3, le présent article peut être lu sans étudier en détail [*GALETS I*] par tout statisticien qui, sans se préoccuper d'archéologie préhistorique, pratique les méthodes de discrimination dont les résultats sont donnés au § 4.

1 L'objet de l'analyse discriminante est de séparer deux sous-ensembles de points d'un espace multidimensionnel par une cloison hyperplane placée dans l'espace ambiant. Il faut rappeler que plus les modalités descriptives sont nombreuses (i.e. plus la dimension de l'espace ambiant est élevée) plus il est aisé de déterminer une telle cloison : c'est ce qu'illustre la figure où l'on tente de séparer deux points noirs d'un point blanc sur la droite, puis dans le plan.



Il est impossible de trouver une coupure sur la droite séparant le point blanc des deux points noirs



Si les deux points noirs et le point blanc forment un vrai triangle, on peut toujours les séparer par une droite

(1) Docteur 3e cycle

(2) Docteur 3e cycle en archéologie préhistorique

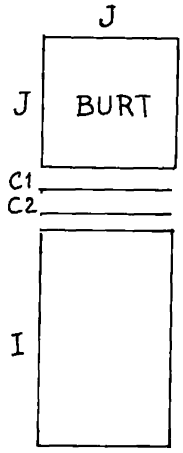
Dans le cas des galets, l'analyse factorielle montre qu'au moins sur les plans définis par les premiers axes ajustés à l'ensemble du nuage, les deux sous-ensembles empiètent notablement l'un sur l'autre; ce qui implique que quelle que soit la cloison hyperplane, il y ait un certain pourcentage de mal classés. On cherchera toutefois à réduire ce %, en tenant compte de ce que l'archéologue préfère sacrifier des vrais pour éliminer tous les faux : on s'attachera plus particulièrement à diminuer le nombre des faux retenus parmi les vrais, quitte à accroître celui des vrais éliminés avec les faux.

Quand, comme c'est le cas ici (cf. § 4) les deux sous-ensembles à discriminer sont d'effectifs très différents, il importe de noter qu'une procédure de discrimination singulièrement expéditive est nécessairement en concurrence avec tout calcul statistique : il suffit d'affecter systématiquement tout individu à la classe la plus nombreuse (ici celle des vrais) pour avoir un taux d'erreur exactement égal au % afférent à l'autre classe (ici 149/705 = 21%). Or, dans le cas présent, il ne semble pas possible de dépasser ce pourcentage par quelque variante que ce soit de l'analyse discriminante. De ce point de vue, nos efforts seront inutiles. Cependant, il reste possible de transformer le problème de discrimination en attribuant (d'une manière ou d'une autre, cf. *infra* §§ 4.2 et 4.3) le même poids total aux deux sous-ensembles à séparer. Dès lors, l'affectation systématique à l'une ou l'autre classe, comme l'affectation par pile ou face, conduit à un taux d'erreur de 50% ; au contraire, les calculs de discrimination proposés gardent dans ce cas leur efficacité relative, conduisant à un taux de 13 à 20% d'erreur.

Quant aux méthodes d'analyse discriminante, on utilise exclusivement ici la méthode d'analyse barycentrique (cf. ENS 2 II n° 7 pp 102-104). Dans cette méthode, on suppose qu'il s'agit de deux classes C1 et C2 d'individus décrits tous par le même ensemble J de modalités de variables logiques. Plaçant d'abord sur un **axe** *unique* deux points C1 et C2 figurant respectivement les classes C1 et C2, on met ensuite chaque modalité descriptive au barycentre des deux points C1 et C2 affectés chacun d'une masse égale au nombre de ses individus possédant la modalité considérée ; enfin, chaque individu est placé au centre de gravité des modalités descriptives qu'il possède.

En fait, cette construction dont le principe justifie le terme même de barycentrique, équivaut à analyser un tableau de correspondance {C1 ,C2} × J dont les deux lignes principales ne sont autres que les descriptions cumulées des individus qui composent respectivement chaque classe C1 et C2 ; les individus eux-mêmes sont adjoints à l'analyse en lignes supplémentaires. L'axe unique issu de cette analyse est (à un changement d'échelle près) l'axe construit par les calculs barycentriques.

On peut encore dire que l'on est dans l'espace R_J, espace qui apparaît aussi par l'analyse du tableau de Burt J × J mais avec d'autres axes propres, ajustés au nuage N(J) (axes qui d'ailleurs ne diffèrent pas de ceux ajustés au nuage N(I) dans le même espace).. Les deux points C1 et C2 sont les centres de gravité des sous-nuages afférents à ces classes et finalement, l'axe "barycentrique" n'est autre que la droite qui les joint ; droite sur laquelle on peut projeter les individus i considérés comme des lignes supplémentaires. L'unique facteur (équivalent à la coordonnée barycentrique) est la coordonnée sur cet axe. A la discrimination qui se fait sur l'axe barycentrique en fixant un point de séparation entre les deux demi-droites d'affectation à chacune des deux classes, il correspond dans l'espace R_J une discrimination par une cloison hyperplane perpendiculaire à l'axe.



2 *Le matériel étudié*

Sur l'ensemble des galets décorés dit "aziliens" et issus du Mas d'Azil que l'on peut trouver dans les divers musées d'Europe , l'archéologue C. Couraud évalue à 20% le pourcentage de faux. Sur les 705 galets peints sur une face dont il nous a confié les données, C. Couraud juge que 149 sont faux et sur les 215 galets décorés sur les deux faces, 44.

Les décors de ces galets sont faits de peinture (oxyde de fer additionné d'un liant) appliquée au doigt ou plus rarement au pinceau : points, traits transversaux, signes divers... Les faussaires ont aussi usé de cette méthode et il semble qu'ils aient choisi leurs galets à décorer dans le site même du Mas d'Azil, en sorte que le support d'un faux peut avoir séjourné des milliers d'années en compagnie de vrais galets aziliens. Mais le décor n'a pu recevoir la patine du temps: calcite et moisissure. Ce n'est toutefois pas là un moyen de reconnaître les faux car un vrai peut avoir été conservé dans des conditions telles qu'il ne comporte ni calcite ni moisissure. D'autres indices entrent en jeu : les traces récentes (les faussaires ont pu aussi imiter des traces), l'aspect de neuf, le bon état de conservation, sont autant de raisons, insuffisantes il est vrai, de penser qu'un galet peut être faux.

Aux variables descriptives utilisées dans l'étude morphothématique des galets (cf. [*GALETS I*]§§ 1.1 et 1.2) : longueur, largeur etc. qui sont précisées ci-dessous, scindées en leurs modalités :

ETAT DE CONSERVATION : E1, bon et très bon ; E2, moyen ; E3, mauvais et très mauvais.

ASPECT : A1, sans aspect particulier ; A2, ancien ; A3, neuf

TRACES D'UTILISATION : TU1, absence ; TU2, présence.

TRACES RECENTES : TR1, absence ; TR2 présence.

CALCITE (sur le décor) : CA1, absence ; CA2, présence.

MOISISSURES (sur le décor) : MO1, absence ; MO2, présence.

Les deux dernières variables (calcite et moisissures) ont été mises en éléments supplémentaires dans les analyses à cause de leur caractère dissymétrique quant au vrai et au faux : leur présence sur un galet implique l'authenticité de celui-ci, mais leur absence ne signifie rien de certain.

Au total, les galets sont décrits par 19 variables découpées en modalités. Ces modalités diffèrent selon qu'il s'agit de galets peints sur une face ou de galets peints sur les deux faces. Les 19 variables comprennent, outre les six variables d'authenticité énumérées ci-dessus, 5 variables de forme, 1 variable de matière, 9 variables de peinture et décor. Le nombre des modalités s'élève à 76 (dont 72 principales) pour les galets peints sur les deux faces.

3 *Analyse de correspondances globale (galets vrais et faux)*

Nous traitons ici brièvement des galets peints sur deux faces, en même temps que des galets peints sur une face, bien que l'analyse discriminante porte seulement sur ces derniers.

Qu'il s'agisse de galets peints sur une seule face ou de galets peints sur les deux faces, on analyse le tableau de Burt construit sur les modalités des 17 variables principales (les variables calcite et moisissures étant mises en éléments supplémentaires) ; à ce tableau les galets eux-mêmes sont ajoutés en éléments supplémentaires.

Dans l'une et l'autre analyse (galets peints sur une face et galets peints sur deux faces) la suite des valeurs propres présente un saut entre la 1° valeur propre et les suivantes qui décroissent plus lentement. Le 1° axe représente 51,5% de l'inertie pour les galets peints sur une seule face et 38% pour les galets peints sur deux faces. Et sur cet axe, s'opposent :

- d'un côté, grande largeur, grande épaisseur, forme ronde, décors complexes, aspect de neuf, couleur sortant du code, matière autre que schiste ;

- de l'autre côté, petite largeur, petite épaisseur, forme oblongue décors formés de points (deux points tangents à chaque extrémité ou à chaque bord, points en lignes dont un tangent pour les galets peints sur une face, points sur les deux faces pour les galets peints sur les deux faces).

Les variables calcite et moisissures sont mal corrélées avec ce 1° axe ; notons toutefois qu'elles se projettent du côté des petites dimensions.

Quant aux galets, dans la 1° analyse, 81% des galets jugés faux par l'archéologue, se projettent du côté des grandes dimensions.

Les faux galets sont en effet pour la plupart de grosses pièces, de forme ronde, au profil lourd , la couleur sortant souvent du code habituel, et nombre d'entre eux ne sont pas en schiste. Leur aspect est généralement neuf et leurs décors complexes. "Ils sont trop beaux pour être vrais", telle est l'impression qui découle de ce premier examen. Cependant plusieurs galets tenus pour vrais par C. Couraud présentent les mêmes caractéristiques. Dans la 2° analyse, 38 faux galets sur 44 se projettent du côté des grandes dimensions.

Grosso modo, on peut dire que le 1° axe oppose les vrais galets aux faux. Mais ce n'est pas un bon facteur de discrimination , à en juger par le pourcentage de mal classés :

- pour les galets peints sur une seule face,

19% de faux et 33% de vrais sont mal classés, soit, au total, 26% de mal classés ; or les 705 galets comportent 149 reconnus faux et 556 reconnus vrais ; en les classant tous vrais, on aurait 149/705 ≈ 21% de mal classés ;

- pour les galets peints sur deux faces,

6/44 ≈ 13% de faux galets et 50/171 ≈ 29% de vrais sont mal classés; or ici encore, les galets reconnus faux constituent 44/215 ≈ 21% de la population totale.

En conclusion, l'analyse du tableau de Burt issu du tableau de description logique des galets permet, sinon de discriminer les vrais et les faux galets, du moins de mettre en évidence, parmi les variables descriptives, celles qui différencient le plus les vrais et les faux galets : outre l'aspect neuf, les traces récentes et le bon état de conservation, les faux ont de grandes dimensions, des formes rondes, une allure lourde des décors en général complexes et souvent leur support n'est pas en schiste.

Signalons encore que les axes suivant différencient les diverses collections qui constituent le matériel étudié : dans telle collection les faux présentent telles modalités. Nous n'en parlerons pas dans cet article (cf. thèse).

4 *Discrimination barycentrique des galets peints sur une face*

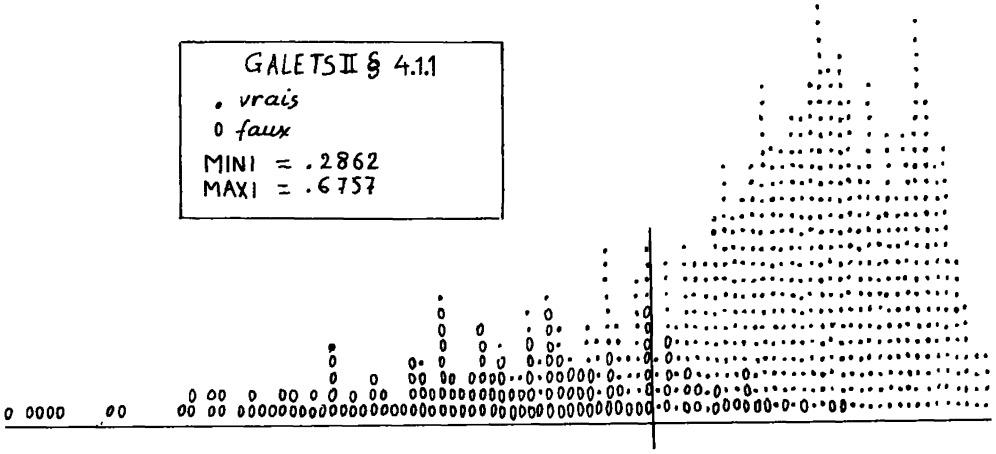
Comme nous l'avons annoncé, nous nous restreignons désormais aux galets peints sur une seule face.

La méthode de d. barycentrique (cf. § 1) consiste à prendre comme direction de l'hyperplan séparateur la direction orthogonale à la droite joignant les centres de gravité des deux sous-ensembles à séparer. Les centres de gravité s'obtiennent très facilement à partir du tableau de description des galets, par cumul des lignes de galets composant chaque sous-ensemble. Soit V la ligne cumulant les 556 lignes décrivant les 556 vrais galets ; F celle cumulant les 149 lignes décrivant les faux. L'analyse du tableau de correspondance à deux lignes : V et F, fournit un seul axe dans l'espace à 72 dimensions (les 72 modalités principales servant à décrire les galets), et cet axe passe nécessairement par les deux centres de gravité ; le facteur sur cet axe, ou coordonnée barycentrique, va servir de fonction discriminante, ξ, dite barycentrique : les galets étant mis en éléments supplémentaires, au tableau à deux lignes V et F, l'analyse de correspondance fournit la valeur de ξ pour chacun des galets. On dresse alors l'histogramme des tranches de valeurs prises par ξ sur l'ensemble des galets en notant, pour chaque tranche, combien de vrais galets et combien de faux ont une valeur de ξ tombant dans cette tranche. On cherche alors la position de l'hyperplan perpendiculaire à l'axe séparant au mieux les faux et les vrais.

Les résultats de cette analyse ainsi que ceux d'analyses complémentaires sont exposés au § 4.1. Comme les faux galets sont beaucoup moins nombreux que les vrais (149 faux et 556 vrais) on compense ce déséquilibre, dans les §§ suivants, soit en multipliant, avant de faire l'analyse, la masse de chaque faux galet par 556/149 (effectif des vrais rapporté à l'effectif des faux) : § 4.2, soit en ne prenant que 149 vrais tirés au hasard parmi les 556 : § 4.3. L'origine des abscisses sur l'axe barycentrique est alors le milieu du segment joignant les deux centres de gravité (V et F).

4.1 Analyse du tableau brut

4.1.1 La figure montre l'histogramme des 705 galets rangés selon leur facteur barycentrique. Les vrais et les faux sont représentés par des sigles différents : on remarque que les faux sont plus dispersés que les vrais. Sur cet histogramme, on peut placer une cloison laissant mal classés 23 faux et 47 vrais, soit 70/705 = 10% de la population totale.



Parmi les vrais mal classés 19 ont leur décor recouvert de calcite ou de moisissures ; les 28 autres ne présentant pas ce caractère peuvent être qualifiés de "douteux", rien ne permettant de prouver de façon certaine qu'ils sont vrais. Ceci nous a amené à refaire l'analyse en mettant tous les douteux en supplémentaires dans le tableau de description des galets : ils ne sont pas pris en compte dans le calcul des centres de gravité des deux sous-ensembles de vrais et de faux. Mais on les met en éléments supplémentaires, comme tous les autres galets dans l'analyse barycentrique (cf. § 4.4).

Pour parfaire la discrimination faite ici (sur les 705 galets) et juger de sa qualité, nous avons effectué trois analyses barycentriques annexes :

- la première portant sur les 106 galets les plus proches de la cloison hyperplane (§ 4.1.2) ;

- la seconde portant sur un échantillon de base composé de 70% de la population des vrais et de 70% de la population des faux, les 30% restants constituant un échantillon d'épreuves (§ 4.1.3) ;

- la troisième portant sur la population constituée de deux groupes : l'un comprenant la moitié des vrais, l'autre la moitié des faux, le choix des galets se faisant par tirage au hasard (§ 4.1.4).

4.1.2 Les 106 galets les plus proches de la cloison comprennent 77 vrais et 29 faux, soit 15% de la population totale. L'examen des trois centres de gravité respectifs des 77 vrais, des 29 faux et des 705 galets permet de dégager les faits suivants :

- les décors complexes sont fréquents sur l'ensemble des 106 galets proches de la cloison :

19 signes linéaires, soit 27% du total des galets ayant ce décor ;

16 traits perpendiculaires (29%) ;

11 décors complexes divers (29%) ;

- les traits transversaux et surtout les points sont beaucoup plus rares sur l'ensemble des 106 galets que sur la totalité des 705 galets :

21 traits transversaux (16% du total des galets ayant ce décor ;

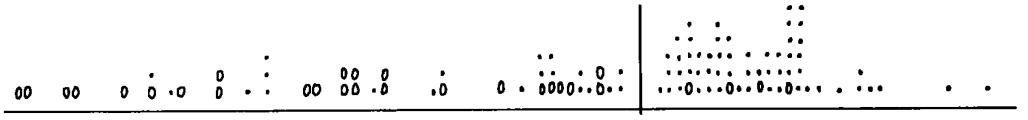
38 points (9%) ;

les 29 faux ont, comme sur la totalité des galets, les modalités : aspect de neuf, bon et très bon état de conservation, absence de bordure ; par contre, ils possèdent des modalités qui sont généralement propres aux vrais galets : petite largeur, longueur moyenne, faible épaisseur, forme oblongue, profil fin, présence de points;

- les 77 vrais présentent, eux, les modalités : grande longueur, grande largeur, grande épaisseur, forme ronde, profil lourd, signes complexes, ce qui les rapproche des faux de la population totale des 705 galets ; mais ils ont, ce qui est en général propre aux vrais , les modalités : aspect ancien, absence d'aspect particulier, mauvais état de conservation, présence de bordure.

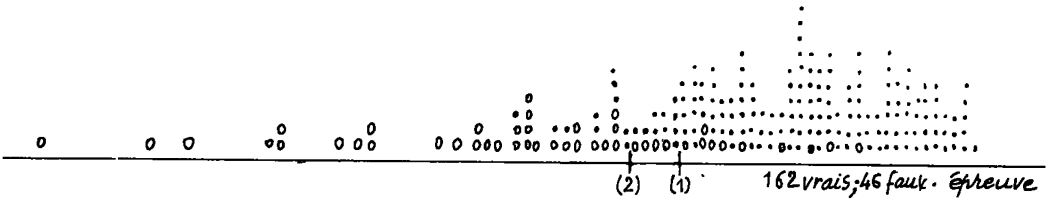
L'analyse barycentrique de ces 106 galets permet de placer au sein du sous-nuage qu'ils constituent une cloison ne laissant mal classés que 4 faux et 20 vrais, soit 23% des 106 galets, alors que l'analyse portant sur les 705 galets laissait mal classés près de 50% de ces 106 galets.

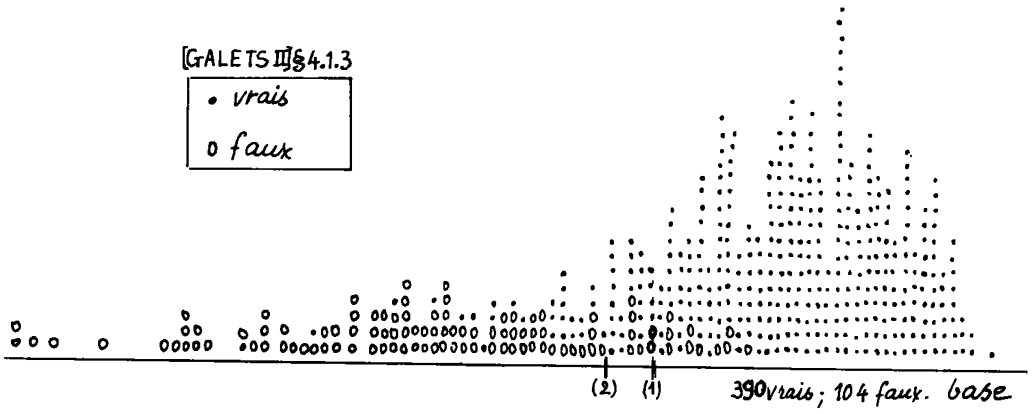
Le classement partiel par analyse barycentrique des individus proches de la cloison placée au sein de la population totale est donc un perfectionnement de la méthode de discrimination par une seule analyse barycentrique.



[GALETS II] § 4.1.2 ; • vrais ; o faux.

4.1.3 Pour juger de l'aptitude du facteur à classer des individus nouveaux, on a prélevé sur les 705 galets un échantillon d'épreuves constitué par 30% des vrais et 30% des faux. On a fait l'analyse barycentrique sur les 493 galets restants dont 389 vrais et 104 faux, l'échantillon d'épreuves étant mis en éléments supplémentaires. La figure 4.1.13 montre les histogrammes des deux échantillons et une cloison laissant mal classés 12% des 493 galets de base et 14 % de l'échantillon d'épreuves. En plaçant la cloison plus à droite, on obtient une proportion moindre de mal classés mais celle des faux mal classés est plus élevée qu'avec la 1° cloison :





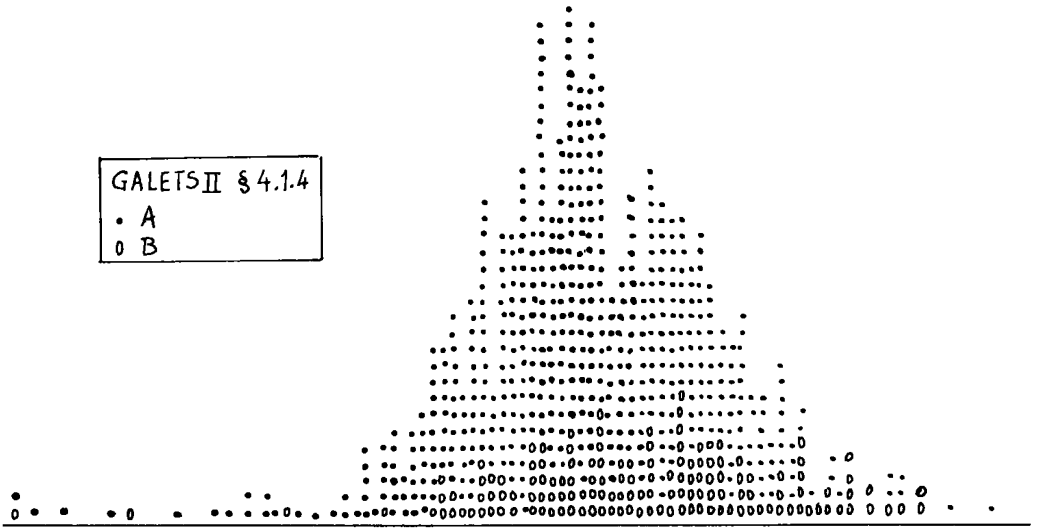
	% de mal classés		nombre de faux classés comme vrais		nombre de vrais classés comme faux	
	base (494 gal.)	épreuve (208 gal.)	base	épreuve	base	épreuve
1° cloison	12 %	14 %	10	10	51	20
2° cloison	9 %	11 %	17	14	26	10

[GALETS II] §4.1.3

4.1.4 Devant l'imperfection des résultats obtenus, on peut se demander si ceux-ci ne sont pas dus à de simples fluctuations statistiques. Pour répondre à cette question, on soumet des données à l'épreuve suivante : on constitue une partition des 705 galets en deux groupes A et B d'effectifs équivalents à ceux des classes de "vrais" et "faux" galets ; mais dans A et B les "vrais" et "faux" sont mêlés : sur l'ensemble des 705 galets, on extrait 144 galets en tirant au hasard 72 galets parmi les vrais et 72 parmi les faux ; les 144 galets ainsi extraits constituent le sous-ensemble B, A étant formé des 561 galets restants.

Dans l'espace des modalités, les deux sous-nuages représentant les sous-ensembles A et B se recouvrent, et il en va de même en projection sur l'axe barycentrique joignant les centres de gravité de A et B . On le vérifie en considérant la figure résultant de l'analyse

barycentrique de la partition des 705 galets en les sous-ensembles A et B. Il est impossible de placer une cloison séparant B de A . En comparant les figures des §§ 4.1.1 et 4.1.4, on constate que dans la première, l'analyse a opéré une véritable séparation, quoique imparfaite, entre les vrais et faux galets, et que dans la seconde, elle n'a fait apparaître aucune distinction entre les groupes A et B.

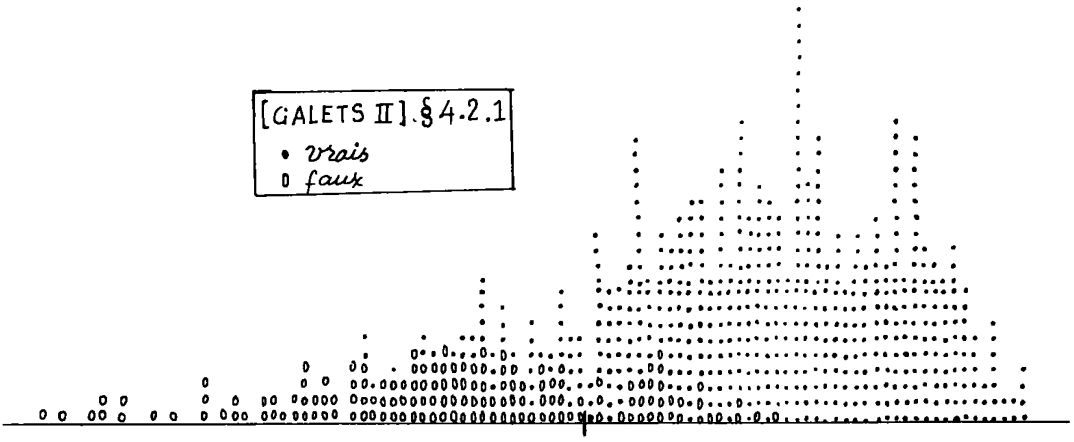


4.2 *Analyses sur le tableau pondéré* : Comme nous l'avons annoncé, on compense ici l'inégalité d'effectif des deux sous-ensembles de vrais et faux galets en pondérant le tableau de description : comme il y a 556 vrais et 149 faux, chaque ligne du tableau correspondant à un faux est multipliée par 556/149. En sorte que les deux lignes cumulant, l'une les vrais galets, l'autre les faux ont même masse.

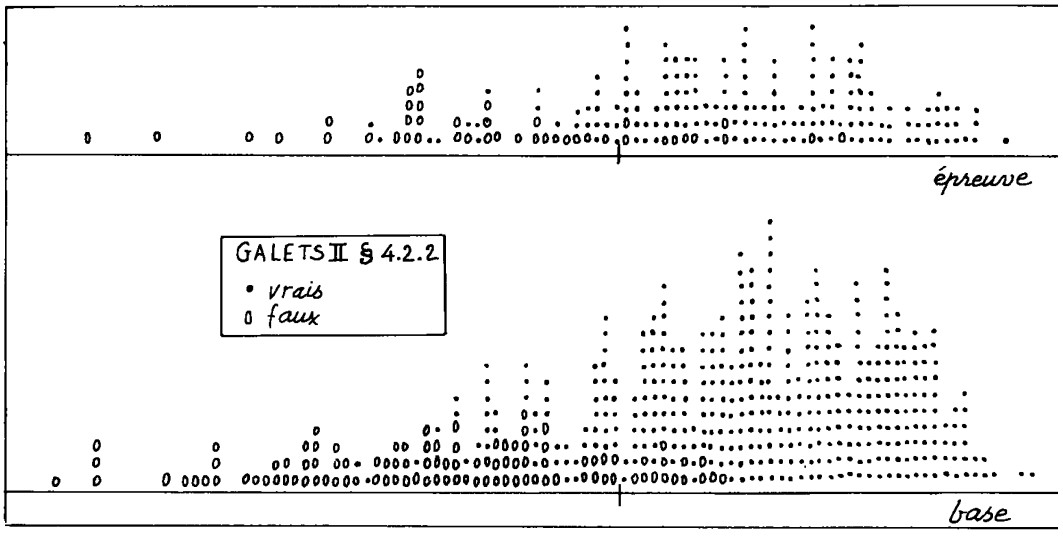
On refait à partir du tableau pondéré, les mêmes analyses que celles décrites au § 4.1 ci-dessus.

4.2.1 *Analyse barycentrique portant sur la totalité de la population:*

On peut placer une cloison laissant mal classés 47 vrais et 32 faux, soit, puisque chaque faux est compté plus de 3 fois (556/149): 47 vrais et 119 faux, ce qui représente 15% de la population totale (556 × 2). Rappelons qu'il faut ici comparer ce taux d'erreur à celui, 50%, que donnerait un classement au hasard. On est donc maintenant assuré d'avoir vraiment effectué une discrimination.



4.2.2 *Analyse barycentrique fondée sur un échantillon de base comportant 70% de la population ; et projection des 30% restants sur l'axe barycentrique*



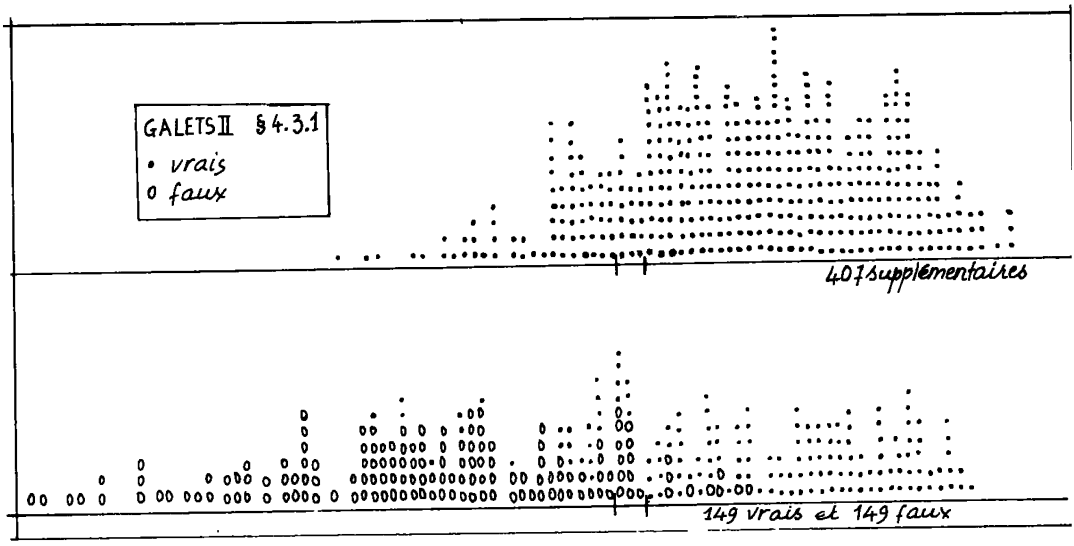
L'analyse barycentrique de l'échantillon de base répartit les galets de telle sorte qu'on peut placer une cloison laissant mal classés 52 faux (14 × (556/149)) et 55 vrais, soit, au total, 14% de mal classés.

Cette même cloison sépare, au sein de l'échantillon d'épreuves, les vrais des faux avec 17% de mal classés : 37 faux (10 × (556/149)) et 24 vrais.

4.3 Le tableau principal est un sous-tableau du tableau brut comprenant les 149 faux et 149 vrais galets tirés au hasard parmi les 556 vrais.

4.3.1 On revient au tableau de description brut, non pondéré, mais, pour déterminer l'axe barycentrique, on prélève deux échantillons de même effectif : l'un formé des 149 faux, l'autre de 149 vrais tirés au hasard parmi les 556 vrais.

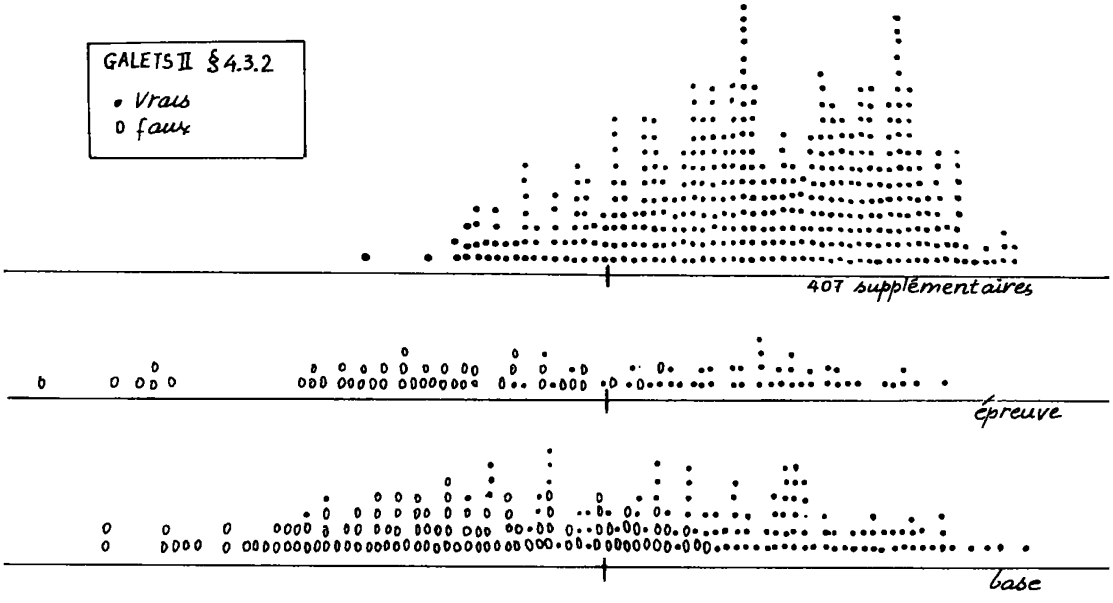
Les 407 vrais restants sont mis en éléments supplémentaires. On place la cloison de séparation sur l'histogramme relatif aux éléments principaux ; puis on regarde comment se projettent les éléments supplémentaires par rapport à cette cloison. Comme on le voit sur la figure, on peut placer deux cloisons.



	% de mal classés		nombre de faux classés comme vrais		nombre de vrais classés comme faux	
	149 vrais et 149 faux principaux	407 vrais suppl.	principaux	supplém.	principaux	supplém.
1° cloison	13 %	22 %	9		29	90
2° cloison	14 %	17 %	21		22	71

4.3.2 L'analyse barycentrique est fondée sur un échantillon de base comprenant 17% des éléments traités en éléments principaux au § 4.3.1: soit 104 vrais tirés au hasard parmi les 149 faux. Les 30% restants (45 vrais et 45 faux) constituent un échantillon d'épreuve que l'on projette sur l'axe barycentrique de l'échantillon de base. En outre, on place en supplémentaire les 407 vrais galets qui n'ont pas été pris lors du tirage des 149 (§ 4.3.1) : ce troisième groupe est projeté, comme l'échantillon d'épreuve , sur l'axe barycentrique de l'échantillon de base.

Nous donnons le résultat de ce partage dans le tableau suivant:

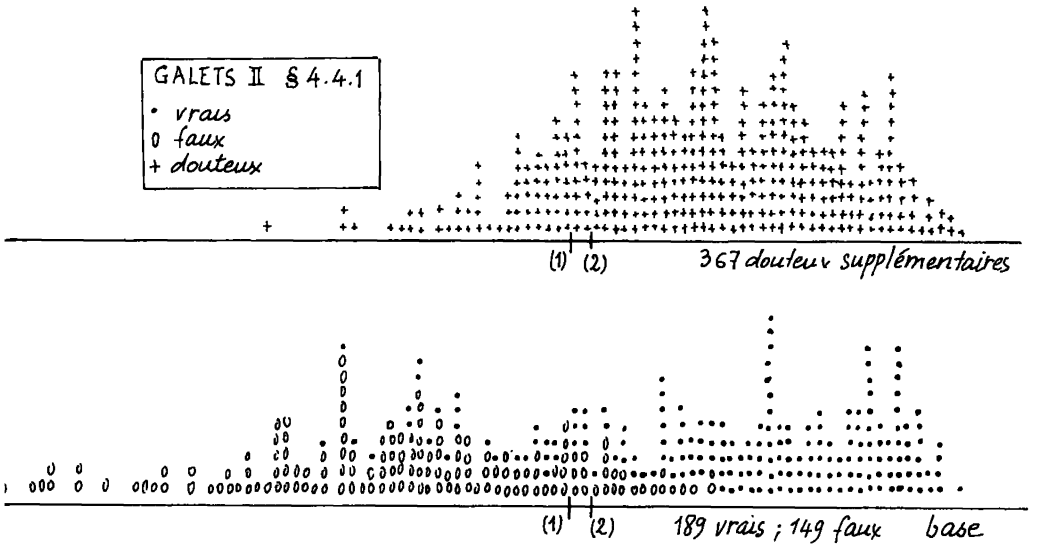


% de mal classés			nb de faux cl. vrais			nb de vrais cl. faux		
base	épreuve	suppl.	base	épreuve	suppl.	base	épreuve	suppl
17 %	9 %	14 %	18	3	/	18	6	59

4.4 Mise en supplémentaires des douteux

Parmi les 556 galets jugés "vrais" par l'archéologue, on isole 367 galets, dits douteux, dont le décor ne porte ni moisissures, ni calcite. Les 705 galets sont ainsi séparés en 149 faux, 189 vrais , et 367 douteux.

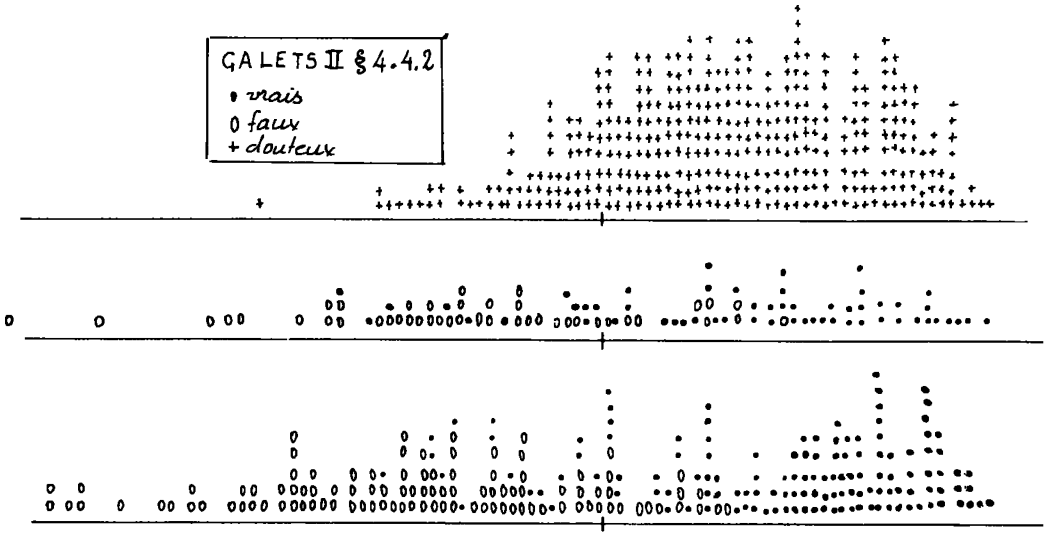
4.4.1 Les douteux étant mis en éléments supplémentaires, on refait l'analyse factorielle et l'analyse barycentrique.



Dans toutes ces analyses, les douteux se projettent avec les vrais. Dans l'analyse barycentrique, la mise en supplémentaire des douteux entraine un étalement de l'histogramme rendant plus diffile le choix d'une cloison. On a placé deux cloisons partageant les galets comme l'indique le tableau ci-dessous.

	% de mal classés			nombre de mal classés		
	vrais (sur 189)	faux (sur 149)	douteux (sur 367)	vrais classés faux	faux classés vrais	douteux cl. faux
1° cloison	14 %	17 %	18 %	26	25	66
2° cloison	16 %	13 %	22 %	30	19	82

4.4.2 Enfin, pour éprouver la discrimination, on a, comme plus haut § 4.1.3, constitué un échantillon d'épreuves comprenant 30% des vrais galets et 30% des faux ; les 70% de vrais et 70% de faux complémentaires servant de base à l'analyse barycentrique les douteux sont mis en éléments supplémentaires. L'échantillon de base contient donc 132 vrais galets et 104 faux ; l'échantillon d'épreuve, 57 vrais et 45 faux ; l'ensemble supplémentaire, 367 douteux.



% de mal classés			nombre de faux classés vrais		nombre de vrais classés faux		nombre de douteux classés faux
base	épreuve	douteux	base	épreuve	base	épreuve	
(sur 236) 13 %	(sur 102) 20 %	(sur 367) 18,5 %	15	9	16	11	68