

REMARQUES SUR LE MAUVAIS USAGE POSSIBLE ET SUR LES ERREURS
DES DIAGRAMMES DE FREQUENCES CUMULEE POUR LA COMPARAISON
DES ENSEMBLES INDUSTRIELS PREHISTORIQUES*

J.E. Kerrich et D.L. Clarke

On utilise depuis longtemps, pour l'étude et la comparaison des ensembles industriels du Paléolithique d'Afrique du Nord et d'Eurasie, les diagrammes de fréquence cumulée (1,2,3). Il est maintenant nécessaire de revoir soigneusement cette technique et son utilisation.

Des méthodes statistiques et des modèles mathématiques s'infiltrèrent lentement en archéologie, la remaniant et en augmentent la possibilité et le pouvoir d'analyse et de synthèse. Les progrès réalisés ainsi sont déjà évidents, comparés aux méthodes empiriques d'exposition des données, par leur rôle analytique accru, et leur pouvoir de prédiction. Parmi les plus anciennes de ces techniques, notons l'usage des graphiques, des polygones de fréquences et des histogrammes qui établissent un nombre strictement limité de pourcentages et de rapports en cause. Parmi les plus récentes, on voit apparaître les calculatrices, et l'exploration par l'archéologie de l'analyse factorielle, de l'analyse matricielle etc. La seule et la plus utile des contributions des techniques statistiques à la comparaison des ensembles industriels préhistoriques, a été l'étude de diagrammes de fréquence cumulée. Cependant, nous suggérons ici que cette technique ne suffit

* Il nous a semblé opportun de publier la traduction française, due à Mme D. AGHION, de l'important mémoire du Professeur J.E. KERRICH - Department of Statistic, University of Witwatersrand - et du Dr. D.L. CLARKE - Peterhouse, University of Cambridge - (Notes on the Possible Misuse and Errors of Cumulative Percentage Frequency Graphs for the Comparison of Prehistoric Artefact Assemblages. Proceedings of The Prehistoric Society Tome 33, 1967, N.4) démasquant sans appel le caractère illusoire des conclusions tirée des "diagrammes cumulatifs" par les tenants des "lystes-types" et autour duquel on a fait en France un pieux silence.

plus à l'analyse détaillée et à la comparaison des ensembles, car elle peut mener à des erreurs dangereuses qui en fausseraient l'interprétation. Cette mise en garde s'applique tout spécialement à des tentatives récentes d'hybridation entre l'emploi de la méthodologie ou de la typologie propres aux études de courbes cumulatives avec des techniques plus modernes (4,5).

Les dangers et les erreurs de cette sorte d'analyse ont été dénoncés avec de plus en plus de vigueur, au cours des dernières décennies, en particulier par les nombreux articles du Professeur J.E. Kerrich et du Docteur R.J. Manson (6, 7, 8).

Nous discuterons ici des dangers des courbes de fréquence cumulée, sous cinq aspects différents :

Erreurs liées aux prises d'échantillons

Erreurs liées aux pourcentages

Erreurs liées à la mise en ordre

Erreurs liées à la typologie

Erreurs liées à l'observateur

Certaines de ces erreurs, par exemple les erreurs liées aux prises d'échantillons, ne sont pas réservées aux courbes cumulatives. Les sources d'erreurs de loin les plus graves, les effets en étant cumulatifs, sont les erreurs liées à la mise en ordre. Pris point par point et erreur par erreur, ces cinq paragraphes peuvent paraître peu importants, mais, pris dans leur ensemble, ils confirment qu'on a déjà atteint les limites d'utilisation de la technique, et qu'un développement ultérieur, hybridé avec des techniques modernes, nuira sérieusement à l'interprétation des résultats.⁺

Erreurs liées aux prises d'échantillons.

De toute évidence, il faut commencer la discussion par ces erreurs, bien qu'elles ne soient pas uniquement le propre des études basées sur les courbes cumulatives. Nous les mentionnons au passage, comme sources possibles de risque.

Nature des échantillons - les ensembles qui figurent sur les courbes cumulatives proviennent de sites archéologiques, ce qui pose plusieurs questions importantes auxquelles il sera peut-être impossible de répondre, mais dont il faut tenir compte pour une interprétation comparative.

⁺ Le paragraphe sur les erreurs liées à la mise en ordre est dû à J.E. Kerrich, le reste du travail est presque entièrement dû à D.L. Clarke.

L'échantillon représente-t-il fidèlement la population du site ? Quelle inévitable marge d'erreur relie le pourcentage d'échantillons aux pourcentages de la population, et par quelle marge d'erreur se définit le champ de variation de l'ensemble culturel ? Certains ensembles proviennent de la fouille totale d'une couche de plusieurs mètres carrés, d'autres de sondages localisés, de peu d'étendue, bien que les fouilles modernes aient mis en évidence une forte tendance à la localisation des types spécifiques d'outils dans les sols d'occupation (9). Quelques échantillons proviennent de couches épaisses, mal définies, et peut-être très étendues. Les problèmes de cette nature, peuvent dans certains cas être résolus par un plan d'échantillonnage bien conçu, mais, très souvent, on se trouve en présence d'un éventail d'ensembles, où la simple présence-absence est la meilleure mesure d'affinité entre les ensembles, plutôt que la comparaison des pourcentages qui comporte une marge d'erreur non exprimée, de l'ordre des pourcentages eux-mêmes. On ne devrait jamais considérer des statistiques d'échantillons comme caractéristiques des paramètres de population, sans en avoir soigneusement apprécié les antécédents et les justifications.

Effectif des échantillons - les ensembles figurés sur les diagrammes cumulatifs ont des effectifs très variables, allant d'environ une centaine, dans certains cas, à plusieurs milliers d'outils, dans d'autres. Disons tout de suite que la plupart des études du Paléolithique par les courbes cumulatives, ont très soigneusement évité les échantillons trop restreints, mais nul n'est sans reproche, et parfois le désir d'avoir des échantillons nombreux a conduit l'analyste à utiliser des échantillons peu crédibles.

Erreurs liées aux pourcentages.

La méthode des pourcentages, en statistique, est des plus dangereuses. Les pourcentages se réfèrent à la structure interne d'un ensemble donné, d'une dimension déterminée, avec une erreur déterminée en relation avec un site et un ensemble culturel. Il en résulte que les pourcentages sont particulièrement peu crédibles lorsqu'il s'agit d'étudier et d'établir les affinités entre des ensembles.

Les ensembles portés sur des courbes cumulatives étant de tailles variées, les pourcentages ne sont en pratique jamais équivalents, bien que la méthode les fasse apparaître tels, et les utilise pour établir des analogies. 10 % de plus dans ensemble de 1.500 outils représentent 150 outils ; 10 % dans un ensemble de 500 représentent 50 outils, et une de ces valeurs de 10 % est triple de l'autre. En général ce facteur d'erreur est moins apparent,

mais il y a toujours une erreur de cette nature à chaque entrée dans les courbes cumulatives, et ces erreurs s'ajoutent l'une à l'autre, d'un composant à l'autre, comme d'ailleurs toute erreur qui apparaît dans cette sorte de graphique.

Les pourcentages sont des données statistiques interdépendantes, et leurs variations ou leurs différences ne peuvent pas aussi facilement être mises en relation qu'on pourrait le croire. La variation du pourcentage d'un type d'outil peut fausser tous les autres pourcentages. Supposons un ensemble simple contenant 50 % d'outils d'un type A et 50 % d'un type B : une diminution de 1 % du type A entraîne automatiquement une augmentation de 1 % du type B. La variation réelle de chaque type est seulement de 1 %, mais les variations étant de sens inverse, il en résulte un écart de 2 % entre A et B. Cela peut se produire même si le nombre réel d'outils B n'est pas modifié.

Le rôle joué par 0 % dans les courbes cumulatives est très curieux. Si un type d'outil, ou une catégorie d'outils est représenté sur le diagramme par 0 %, nous savons - avec une certaine marge d'erreur - que ce type d'outil manque à la place où il aurait dû figurer. Des valeurs de 0 % pour 4 ou 5 types d'outils peuvent être dispersées dans les courbes cumulatives, et pourtant il n'y aura que peu ou pas de différence entre la forme apparente de la courbe et celle de la courbe où ces types figurent en faibles pourcentages. Sur des bases archéologiques, les ensembles qui diffèrent en nombre de types d'outils non départagés doivent être séparés, même s'ils se partagent entre 13,2 % du type X et 17,5 % du type Y (quand ces pourcentages, comme nous l'avons vu, représentent des nombres très différents d'outils). Si le but des études des courbes cumulatives est l'évaluation de l'analogie ou de la différence entre les ensembles, il est évident que la présence ou l'absence d'un certain nombre de types d'outils ne doit pas être masquée par des rencontres fortuites dans quelques pourcentages. On doit aussi insister sur le fait que la différence entre 3 % et 2 % n'est pas du tout de la même nature que la différence entre 1 % et 0 % ; 0 % n'est pas 1 % de moins que 1 %, et de plus 0 % est la relation la moins discernable et la plus importante dans les courbes cumulatives, alors même que 0 % n'introduit qu'un changement minime. Une classification numérique évalue les ressemblances et les différences des ensembles sur des bases entièrement opposées, et fournirait visiblement une classification tout à fait différente des ensembles réunis jusqu'ici par comparaison des pourcentages (10,11).

La méthode des pourcentages tend à cacher les aspects significatifs des données et ne remplace pas les véritables données comme bases d'analyse et de synthèse.

Erreurs liées à la mise en ordre.

La mise en ordre, ou arrangement arbitraire de la séquence de types d'outils sur l'axe horizontal du diagramme de fréquence, influe beaucoup sur la forme des courbes cumulatives, et donc sur notre estimation des analogies entre les ensembles industriels. Quelques exemples vont illustrer cet aspect du problème.

Une mise en ordre des données statistique consiste d'abord dans la distribution de fréquence des observations d'une variable statistique. Exemple dans le Tableau I.

TABLEAU I

Nombre de garçons dans des familles de 5 enfants x	Nombre de familles f	Pourcentage de familles $\% f$	Pourcentage de familles avec 5 garçons ou moins $c \ \% f$
0	5	2	2
1	35	14	16
2	75	30	46
3	80	32	78
4	45	18	96
5	10	4	100
	<hr/> n = 250	<hr/> 100	<hr/>

x = nombre de garçons dans des familles de 5 enfants. C'est une quantité observable qui varie d'une famille à l'autre, et constitue une variable statistique. C'est une caractéristique naturelle, calculable de chaque famille individuelle.

Dans cet exemple, on a examiné un échantillon pris au hasard de 250 familles d'une population bien définie, et on a noté la fréquence avec laquelle chaque valeur de x apparaît dans l'échantillon. Ainsi, 5 familles ont 0 garçons (et 5 filles), 35 familles ont 1 garçon (et 4 filles) etc...

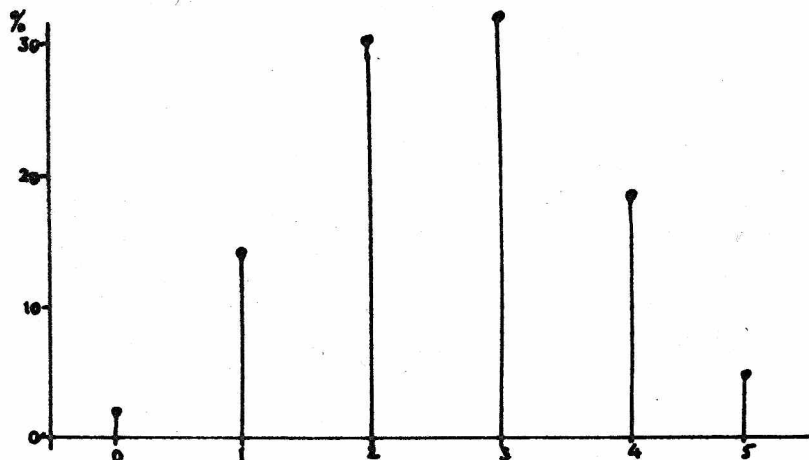
Ces deux colonnes constituent ensemble la fréquence dite de distribution observée de x .

Comme il est plus facile de comparer 2 pourcentages que de comparer 2 rapports, on a ajouté une troisième colonne qui donne le pourcentage de familles, $\% f$, qui correspond à chaque valeur de x .

Figure 1

 $Y = \% f$

= % des familles
avec X garçons

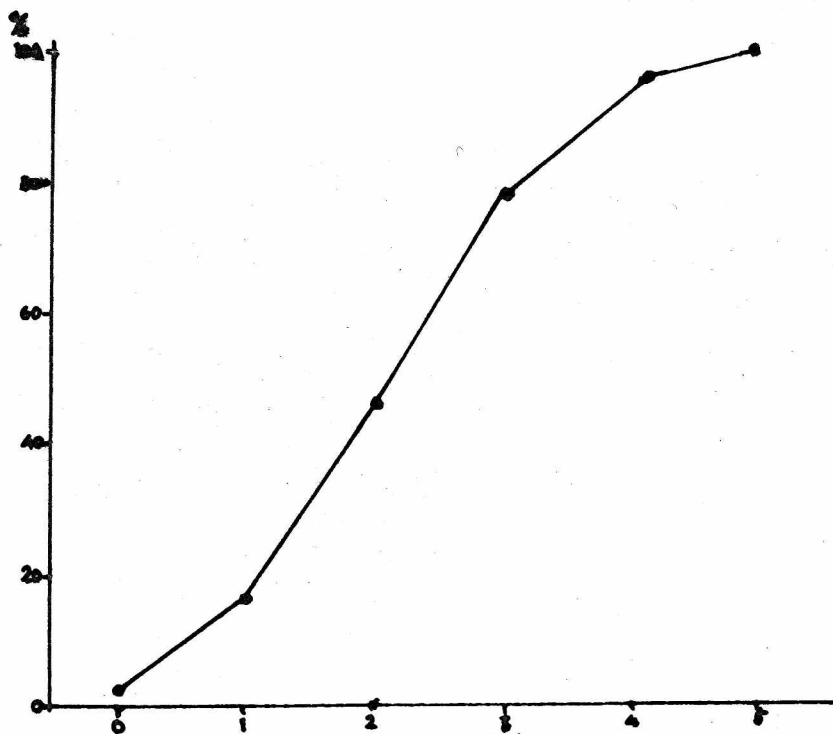


X = nombre de garçons dans des familles de 5 enfants

Figure 2

 $Y = c \% f$

= % des familles
avec X garçons
au moins



X = nombre de garçons dans des familles de 5 enfants

Enfin, par additions successives des fréquences %, on a encore ajouté une quatrième colonne des fréquences cumulées, $c \%$ f , qui donne le pourcentage des familles qui ont \underline{x} garçons, ou moins, pour chaque valeur de \underline{x} .

On représente graphiquement un tableau de fréquences % par la méthode fondamentale du diagramme en bâtons.

La figure 1 représente le diagramme en bâtons pour les données du Tableau I. Pour chaque valeur de \underline{x} , la hauteur du bâton représente le pourcentage de familles correspondant à cette valeur de \underline{x} .

Une autre méthode pour représenter l'information donnée par le Tableau I, est le diagramme cumulatif de la figure 2, où la hauteur du point au-dessus de l'axe des x donne le $c \%$ f associé à la valeur correspondante de \underline{x} . Les droites qui joignent les points servent à suivre visuellement l'information, mais n'ont aucune signification physique.

Les deux figures 1 et 2 contiennent exactement la même information, exprimée de deux façons différentes. L'une étant donnée, on peut immédiatement en déduire l'autre, et d'une seule manière.

Il arrive souvent qu'une information non évidente sur la figure 1 puisse être obtenue plus vite à partir de la figure 2, spécialement quand il y a beaucoup plus de valeurs de \underline{x} que dans l'exemple donné.

L'échantillon de la figure 1 montre une tendance vers des familles qui ont plus de garçons que des filles. Supposons maintenant un échantillon de même effectif, choisi dans une population différente où la tendance soit inverse (plus de filles que de garçons). Aussitôt les sociologues, les généticiens, les anthropologues et les autres, seront alertés et demanderont à savoir pourquoi et dans quelle mesure les deux populations diffèrent.

Le premier point à noter est que les deux communautés ont exactement les mêmes caractéristiques, et que les différences observées dans les échantillons peuvent être dûes au hasard. (C'est ce qu'on appelle l'Hypothèse nulle). La première mesure à prendre est de tester cette hypothèse nulle. Il existe pour cela un test classique (le test du Chi carré). Le résultat de ce test vous dira soit d'accepter l'hypothèse nulle, soit de la rejeter.

Si vous acceptez l'hypothèse nulle, il faut savoir que l'information partielle contenue dans ces deux échantillons ne suffit pas à décider en toute connaissance de cause, si oui ou non les deux communautés ont des caractères différents.

Si le résultat du Chi carré vous conseille de rejeter l'hypothèse nulle, vous saurez alors que les échantillons montrent à l'évidence que les communautés ont des caractères différents, pourvus que l'on soit sûr que les

échantillons aient été choisis au hasard.

Dans ce cas, la démarche suivante consiste à estimer la valeur numérique de certains aspects de ces différences. Il faut rappeler pourtant que, au moyen des échantillons, on peut seulement estimer l'information concernant les populations, mais jamais la déterminer exactement.

Il faut rappeler aussi que les résultats de ces tests et de ces estimations dépendent de l'effectif des échantillons, autant que des pourcentages dans ces échantillons. Comme toujours, la totalité de l'information disponible influe sur la qualité des opinions qu'on peut fonder sur cette information.

Il est impossible, dans un tel article, de développer plus avant les notions ci-dessus exposées. Mais ce sont des notions fondamentales, basées sur le simple bon sens, et il faut encourager les archéologues à consulter des traités élémentaires de statistique et à se familiariser avec eux. Le résultat obtenu pour certains types de données devrait inciter les archéologues à plus de prudence dans leurs jugements, et à se débarrasser de tout dogmatisme et de toute acrimonie dans leurs écrits. Pourquoi les archéologues devraient-ils conserver un retard de quarante ans sur leurs collègues médecins, sociologues, psychologues, etc., dans l'emploi des méthodes statistiques ?

Examinons maintenant les données archéologiques du Tableau II, qui se rapportent à deux échantillons, l'un provenant d'un site A, l'autre d'un site B.

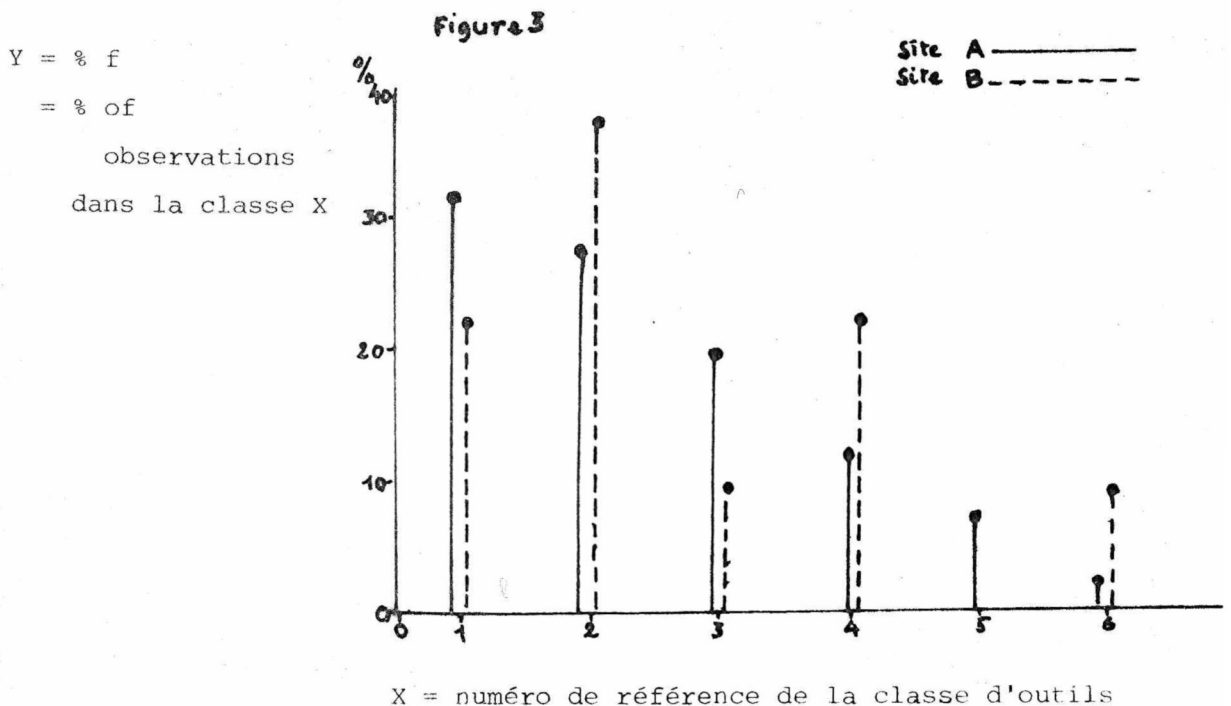
TABLEAU II

Numéro de référence x	Classe d'outils	Nombre d'outils		Pourcentage d'outils		Pourcentages cumulés	
		A	B	A	B	A	B
1	Handaxes	96	44	32	22	32	22
2	Sidescrapers	84	76	28	38	60	60
3	Cleavers	57	18	19	9	79	69
4	Endscrapers	36	44	12	22	91	91
5	Long flakes	21	0	7	0	98	91
6	Short flakes	6	18	2	9	100	100
		300	200	100	100		

L'information fondamentale contenue dans les deux échantillons est donnée par les pourcentages. On peut les représenter facilement et simplement par des diagrammes en bâtons (voir figure 3), les bâtons correspondant à chaque classe d'outils dessinés l'un à côté de l'autre, ou bien on peut établir un diagramme cumulatif (voir figure 4), selon la méthode suivie par (1,2,3). Mais quelle information utile pouvons-nous obtenir de la figure 4 qui ne puisse être obtenue plus facilement du diagramme en bâtons ou du tableau des pourcentages ?

Il n'y a aucune ironie dans cette question. On peut remarquer certains avantages immédiats de la figure 4 sur la figure 1, tels que "la figure 4 montre au premier coup d'oeil que pour les deux échantillons 91 % des observations se trouvent dans les quatre premières classes".

Il est plus difficile d'obtenir cette information à partir du diagramme de la figure 3.



Mais, pour que la phrase entre guillemets soit compréhensible, le lecteur doit savoir ce que signifie l'expression "les quatre premières classes". Opposez ceci à l'énoncé "dans un échantillon de familles de 5 enfants, 96 % des familles ont 4 garçons, ou moins", énoncé qui, tel quel, est tout à fait clair (voir fig. 1).

TABLEAU III

Numéro de référence x	Classes d'outils	Pourcentage d'outils		Pourcentage cumulé	
		A	B	A	B
1	Sidescrapers	28	38	28	38
2	Endscrapers	12	22	40	60
3	Short flakes	2	9	42	69
4	Long flakes	7	0	49	69
5	Cleavers	19	9	68	78
6	Handaxes	32	22	100	100
		100	100		

Les classes d'outils données dans le Tableau II, peuvent être réarrangées de plusieurs façons judicieuses. Il est par exemple aussi logique de les ordonner conformément au Tableau III.

Apartir de ces pourcentages cumulatifs du Tableau III, nous obtenons la courbe cumulative de la figure 5.

Etant donné que les figures 4 et 5 sont des graphiques de même type et représentant exactement la même information, lequel est préférable, et pourquoi ?

Il y a pire encore, car les classes d'outils peuvent être ordonnées de $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ manières différentes, de sorte qu'il y a 720 diagrammes possibles, dont les figures 4 et 5 ne sont que deux exemples. Lequel des 720 allez-vous choisir ?

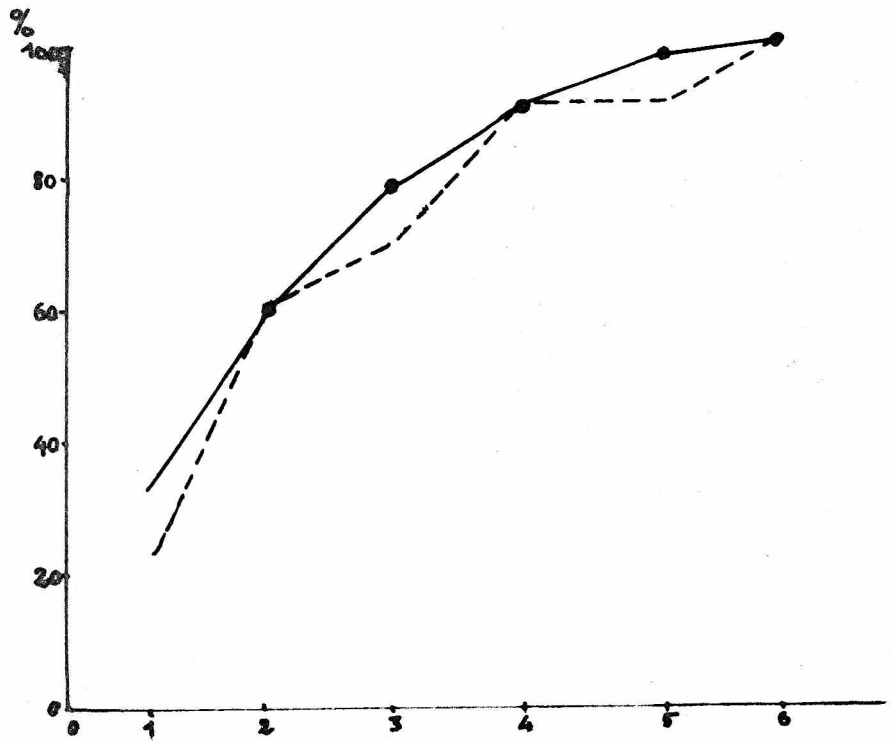
N'est-il pas dangereux de choisir la figure 4, si on juge subjectivement que les deux échantillons "se ressemblent beaucoup", ou la figure 5, si on considère que les deux échantillons sont "très différents" ?

Si, comme cela se présente en pratique, un tableau d'outils contient 112 éléments, on peut établir $112! = 1.972 \times 10^{182}$ diagrammes cumulatifs différents, à partir des mêmes données, toutes contenant la même information.

On comprend ce qui s'est passé en comparant le Tableau I et le Tableau II. Dans le Tableau I, où x est le nombre de garçons dans des familles de 5 enfants, chaque valeur de x est associé à un phénomène naturel particulier

Figure 4

Y = c % f
 pourcentage
 d'observations
 dans les x
 premières
 classes

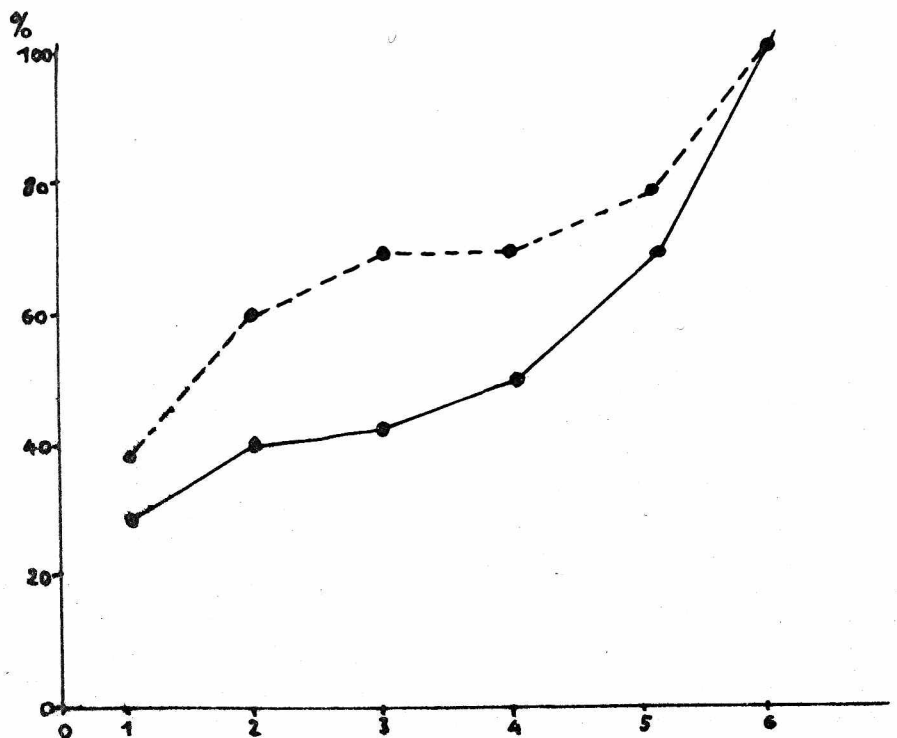


X = numéro de référence des classes d'outils

Figure 5

Site A ———
 Site B - - - -

Y = c % f
 pourcentage
 d'observations
 dans les x
 premières
 classes



X = numéro de référence des classes d'outils

tandis que dans le Tableau II, où x est le numéro de référence d'une classe d'outils, il peut être arbitrairement associé à n'importe lequel des autres parmi un grand nombre de classes d'outils. Ainsi, dans ce cas, x est lui-même un outil.

Tant que les archéologues ne se seront pas mis d'accord sur une méthode unifiée de classement, ils devront considérer avec une certaine méfiance les diagrammes cumulatifs.

Et même si cet heureux accord se réalisait entre de tels individualistes, ne serait-il pas opportun, pour éviter les impressions subjectives suggérées par tout diagramme, d'utiliser des tests et des procédés de mesures que fournissent les méthodes objectives de certaines techniques statistiques unifiées modernes ?

Erreurs liées à la typologie.

Avec les problèmes de mise en ordre que nous venons de discuter, les erreurs liées à la typologie constituent l'autre source principale d'erreurs dans les diagrammes cumulatifs utilisés pour comparer des ensembles industriels.

Bien plus, la base de cette sorte d'incertitude est dès aujourd'hui reportée dans la seconde génération de techniques statistiques, ce qui entraîne de graves conséquences, pour leur pouvoir, par ailleurs considérable, d'analyse des données archéologiques.

Le problème commence par la définition adéquate des types d'outils en tant qu'ensemble d'outils selon un champ exact de dispersion des caractères, de variabilité des caractères, et le problème dépend, de plus, du traitement propre de l'analyse de l'ensemble en fonction de ces types.

On ne peut trop estimer l'importance d'une définition exacte des types d'outils, puis de l'application de cette définition avec toute la rigueur possible. Beaucoup d'études paléolithiques jouent avec la variation, d'un ensemble à l'autre, des pourcentages relatifs d'une catégorie donnée d'outils. On a récemment appliqué à cette sorte de données, des techniques statistiques modernes des plus élégantes et des plus efficaces (4,5). Pourtant, à peine quelques unes de ces études définissent leurs types d'outils autrement que sur des bases intuitives et arbitraires, de justesse contestable, et ces catégories ne sont certainement pas assez complètes ni assez bien définies, quant aux caractères et aux variations de ces caractères.

Il serait dangereux que des définitions contradictoires ou incompatibles des types d'outils dans tel ensemble puissent modifier radicalement les pourcentages relatifs, déjà très discutés, et puissent par suite modifier aussi leur interprétation signifiante. Une telle manière de faire anéantit immédiatement l'efficacité et l'élégance des techniques utilisées dans les analyses récentes, et presque toutes les qualités objectives des diagrammes et des courbes cumulatives de tels types intuitifs. C'est un anachronisme absurde d'user de techniques modernes efficaces pour construire en préhistoire des hauts niveaux apparents, quand les bas niveaux et les fondations en sont inadéquats et peu sûrs ; on ne peut construire des structures solides sur des bases douteuses.

Si des techniques élaborées doivent remplacer les courbes cumulatives aux plus hauts niveaux d'interprétation, alors les concepts du niveau inférieur doivent être définis pareillement. On doit, malgré une hâte injustifiée, surmonter l'ennui d'avoir à bien définir chacune des catégories d'outils, et la poursuite des résultats rapides ne doit pas altérer la qualité de résultats acquis par beaucoup de temps et des analyses coûteuses

Une typologie inadéquate conduit à des analyses contradictoires d'un même ensemble, et rend tout à fait impossible le recoupement des diagrammes cumulatifs ou de leurs analogues plus récents. Comme exemple de la confusion typologique qui se dissimule derrière des désignations de types admises sans discussion, notons les typologies aussi diverses que celles des grattoirs frontaux du Paléolithique Supérieur, pour lesquels, depuis 1954, trois études proposent, les unes quatre types, les autres onze (12). Il devient impossible, sur de telles bases, de comparer valablement les courbes cumulatives du Moustérien de Hongrie par exemple, avec le Moustérien de France ; les types d'outils n'étant pas définis de la même façon ne peuvent pas être comparés. Au contraire, la réponse à ce dilemme dépend de la définition plus précise des types d'outils selon le champ de distribution de leurs caractères, au moyen précisément de ces méthodes de regroupement multifactorielles qui se sont lancées à l'assaut de ces typologies arbitraires tristement peu satisfaisantes. Quelques archéologues ont heureusement bien évalué ce facteur, et se sont mis à cette tâche essentielle, avec les méthodes les plus récentes.

Une source majeure d'incertitude dans la comparaison des diagrammes cumulatifs réside donc dans les résultats contradictoires qu'obtiennent des chercheurs différents analysant le même échantillon. Un bon enseignement magistral ne fera que rassembler les contradictions individuelles en une

contradiction collective, et ne résoudra pas le problème - montre que les résultats des assitants d'un même maître sont tous compris dans une marge de 4 à 10 %, ce qui démontre seulement la cohérence de l'enseignement. Des essais sur de vraies données paléolithiques reportées sur des courbes cumulatives ont montré des écarts individuels de l'ordre de 4 à 50 %, dans l'attribution de certaines catégories - et ces écarts sont le fait de bons spécialistes (13). De telles erreurs ou différences sont bien sûr cumulatives et vont s'ajoutant de type en type, dans le diagramme.

Enfin, il faut noter deux autres sources d'erreur ou d'incertitude typologiques. La première est le chevauchement des types, lorsque deux types empiriques fusionnent et ont un domaine commun. Là encore, la définition de l'outil est décidée arbitrairement et peut différer incroyablement d'un analyste à l'autre, ce qui modifie une fois de plus les pourcentages cumulatifs. Il reste enfin les "types divers", fourre tout résiduel, naturellement un risque de plus pour l'interprétation des pourcentages réels.

Une fois encore, jusqu'à ce que les archéologues se mettent d'accord sur une procédure et une définition valable des types d'outils, ils devront se méfier de leur formulation cumulative, aussi sophistiquée que puisse être la méthode de ressemblément des ensembles de types arbitraires.

Erreurs liées à l'observateur.

Des erreurs de perception apparaissent chaque fois que des individus différents doivent comparer et opposer des lignes ou des courbes, et en déduire des analogies. C'est un facteur d'erreur tout à fait mineur, mais qui ajoute néanmoins une difficulté pour le recouplement des courbes cumulatives. La psychologie différentielle nous dit que les individus forment une suite continue, allant des "percepteurs synthétiques" aux "percepteurs analytiques" (14). En présence de lignes complexes, l'individu peut percevoir un ensemble total et comparer ainsi courbe entière avec courbe entière ou bien il peut s'attacher à la perception fine d'un détail isolé, et comparer détail local avec détail local. La comparaison des courbes cumulatives constitue une bonne étude de perceptions comparées. Que faut-il comparer, la ressemblance des formes dans leur ensemble, le dessin détaillé des paliers de la courbe ascendante, ou la différence entre les mesures des pourcentages ? Une chose est certaine : étant donnée la même paire de courbes, les uns les verront tout à fait comparables, les autres tout à fait différentes.

Conclusion

Le développement et l'utilisation des diagrammes de fréquence cumulée, constituent la seule contribution, et la plus encourageante, de la première génération des techniques statistiques en archéologie. Il faut féliciter ces archéologues qui ont contribué à cette importante recherche d'objectivité, dans une tentative d'importance primordiale, pour faire de l'archéologie une discipline.

Pourtant le développement constant des techniques statistiques, et la large utilisation des calculatrices annoncent une nouvelle génération de méthodes et de modèles, de pénétration et de possibilités très améliorées. La nouvelle étape a déjà suscité un flot croissant de tentatives neuves et audacieuses d'utilisation des techniques les plus récentes pour l'analyse archéologique. Elle n'est pas sans dangers, et très instructives seront les erreurs commises par ceux qui voudront affronter ses difficultés dans la recherche de la connaissance, et ne se satisferont pas de leur "esprit de clocher". Dans cette perspective, nous devons insister beaucoup sur les dangers de l'extension de la technique des pourcentages cumulatifs et de la typologie correspondante, qui peuvent amener à des conclusions dangeureuses.

Une remarque finale. Les archéologues doivent savoir que, même si certaines techniques statistiques leur fournissent des instruments utiles pour les aider à penser clairement et à se faire un jugement sain, ces techniques ne vont pas automatiquement penser pour eux. La responsabilité finale de la validité de leur recherche reste aux archéologues eux-mêmes.

BIBLIOGRAPHIE

- 1) BORDES F. 1954 - Les limons quaternaires du Bassin de la Seine, Arch. Inst. Pal. Hum., vol. 26.
- 2) SONNEVILLE-BORDES D. 1960 - Le Paléolithique supérieur en Périgord, I et II, Bordeaux.
- 3) TIXIER J. 1963 - Typologie de l'Epipaléolithique du Magreb, Mem. C.R.A.P.E., Alger.
- 4) BINFORD L. and BINFORD S.R. 1966 - A preliminary analysis of functional variability in the Mousterian of Levallois Facies, American Anthropologist, vol. 68, n.2, part. 2, 238-96.

- 5) DORAN J.E. and HODSON F.R. 1966 - A digital computer analysis of Paleolithic flint assemblages, Nature, vol. 210, n. 5037, May 14th, 688-9.
- 6) KERRICH J.E. 1957 - Statistical Note, South African Archaeological Bulletin, n. 48, vol. XII, 137.
- 7) MASON R.J. 1957 - The Transvaal Middle Stone Age and Statistical Analysis, South African Archaeological Bulletin, n. 48, vol. XII, 119-43.
- 8) MASON R.J. 1962 - Prehistory of the Transvaal, p. 224.
- 9) MOVIUS H.L. 1966 - The ...Aurignacian horizons at the Abri Pataud, Les Eyzies, American Anthropologist, vol. 68, n. 2, part. 2, 216-325.
- 10) SOKAL R.R. and SNEATH P.H.A. 1963 - Principles of numerical taxonomy.
- 11) ESCALON DE FONTON M. et D. LUMLEY H. 1955 - Quelques civilisations de la Méditerranée septentrionale et leurs interurrences, Bull. Soc. Préhist. Française, vol. 52, 379-90 Table.
- 12) SACKETT J.R. 1966 - Quantitative analysis of Upper Paleolithic stone tools, American Anthropologist, vol. 68, n. 2, part. 2, 356-90.
- 13) MELLARS P. - Unpublished Thesis material, PhD (Cantab).
- 14) ANASTASI A. 1965 - Differential psychology, individual and group differences in behaviour, reference to p. 354 of 3rd Edition.