Journal de la société statistique de Paris

Fred Milhaud

Étude des relations logistiques entre les fréquences de deux phénomènes

Journal de la société statistique de Paris, tome 101 (1960), p. 282-291
<http://www.numdam.org/item?id=JSFS_1960__101__282_0>

Étude des relations logistiques entre les fréquences de deux phénomènes

PREMIÈRE PARTIE

Considérons deux événements A et B.

Soit un nombre N d'expériences au cours de chacune desquelles on vérifie leurs présences.

Soit P1 le nombre de cas où l'on a A présent, B présent.
P2 le nombre de cas où l'on a A présent, B absent.
P3 le nombre de cas où l'on a A absent, B présent.
P4 le nombre de cas où l'on a A absent, B absent.

Il est possible de construire un tableau.

	A PRESENT	A ABSENT
B présent	P1	P3
B absent	P2	P4
Pourcentage des présences de B	—	—

Un calcul de χ^2 permet de dire si la différence entre les pourcentages de B dans la 1re colonne et dans la 2e peut être observé de façon fortuite avec une certaine vraisemblance.

Si le χ^2 est trop élevé pour cela, s'il est significatif, A et B sont liés.

Ce renseignement est-il suffisant?

On peut souvent s'en contenter.

Si on dit par exemple que les garçons de 9, 10, 11 ans lisent plus souvent des histoires de gangsters que ceux de 6, 7, 8 ans et que les garçons de 12, 13, 14 ans en lisent encore plus, il est évident que c'est l'âge qui est la cause de leurs lectures et non pas les lectures qui sont causes de leur âge. Il semble d'ailleurs absolument vain de chercher à s'exprimer autrement qu'en disant que l'âge est l'un des facteurs en fonction desquels augmente, du moins durant l'enfance et l'adolescence, la fréquence des lectures ayant pour sujets des histoires de gangsters.

Cependant entre les fréquences de deux phénomènes, il peut y avoir conjonction, homologie, contradiction, implication, contrariété, disjonction.

Conjonction signifie que ces fréquences sont indépendantes.

Homologie signifie que si on considère des ensembles différents d'observations, on a une fréquence élevée de A et une fréquence élevée de B, on a une fréquence faible de A et une fréquence faible de B, on n'a pas une fréquence élevée de l'un et une fréquence faible de l'autre.

Contradiction signifie que l'on a toujours, soit une fréquence élevée de A et une fréquence faible de B, soit une fréquence faible de A et une fréquence élevée de B.

Implication signifie que la fréquence élevée de l'un d'eux s'observe dans des cas qui sont compris dans l'ensemble de ceux où la fréquence de l'autre est élevée. Par exemple si A implique B cela signifie que l'on peut avoir une fréquence élevée de A et une fréquence élevée de B, une fréquence faible de A et une fréquence faible de B, une fréquence élevée de B et une fréquence faible de A. On ne peut pas avoir une fréquence élevée de A et une fréquence faible de B, la fréquence élevée de A implique celle de B, par exemple la fréquence élevée des talons hauts dans une foule implique celle des femmes, alors qu'il existe par ailleurs de nombreux endroits où il y a beaucoup de femmes sans beaucoup de hauts talons.

Contrariété signifie que l'on ne peut pas avoir des fréquences élevées de A et de B dans le même ensemble. On peut avoir A fréquent et B rare, A rare et B fréquent, A rare et B rare.

Disjonction signifie au contraire que l'on peut avoir des fréquences élevées de A et de B, A fréquent et B rare, A rare et B fréquent. On ne peut pas avoir à la fois A et B rares dans la même population.

Considérons le tableau que nous avons présenté au début.

$$N = P1 + P2 + P3 + P4$$

Posons

$$F = P1 + P4$$
$$G = P2 + P3$$

F est la fréquence d'un événement α qui est la concordance des présences ou celle des absences.

G est celle d'un événement β contraire de α qui est leur discordance.

Quand on fait N tirages, et qu'à chacun on a une probabilité $p = \frac{1}{2}$ de tirer un événement et $q = \frac{1}{2}$ de tirer son contraire, on se trouve dans une situation classique où l'on peut appliquer la théorie du schéma de Bernouilli. Le nombre modal de tirages de chacun des 2 événements est $\frac{N}{2}$, l'écart type est $\sqrt{N \frac{1}{2} \frac{1}{2}} = \frac{1}{2}\sqrt{N}$ et il est possible de dire qu'une différence de $\frac{3}{2}\sqrt{N}$, correspondant à deux différences chacune de (1,5) σ et de sens inverse, est peu vraisemblable :

Donc :

Si la différence constatée entre $F = P1 + P4$, et $G = P2 + P3$ dépasse en valeur absolue $\frac{3}{2}\sqrt{N}$, elle sera tenue comme significative, et on considérera son sens.

Si $F > G$, il y a homologie entre A et B.

Si $G > F$, il y a contradiction entre A et B.

Si la différence entre F et G n'est pas significative, on considère $F = P1 + P4$.

Il s'agit de savoir si en pratiquant F tirages d'un événement avec chaque fois une probabilité $p = \frac{1}{2}$ de tirer l'un, et $q = \frac{1}{2}$ de tirer son contraire, on aurait pu observer la différence rencontrée entre P1 et P4. Sinon ce n'est pas par hasard que l'on a trouvé surtout des concordances présence ou des concordances absence.

Si la différence constatée entre P1 et P4 dépasse en valeur absolue $\frac{3}{2}\sqrt{F}$, elle doit être tenue pour significative, et on considérera son sens.

Si $P1 > P4$, il y a *disjonction*, c'est-à-dire que les fréquences de A et de B peuvent être fortes ensemble, que chacune peut être forte sans l'autre, mais qu'elles ne peuvent pas être faibles ensemble.

Si $P4 > P1$, il y a *contrariété*, c'est-à-dire que les fréquences de A et de B ne peuvent pas être fortes ensemble, alors que chacune peut être forte sans l'autre, et qu'elles peuvent être faibles ensemble.

La disjonction prouve que le référentiel était, à l'insu peut-être de l'enquêteur, caractérisé par un *symptôme*, et que les phénomènes que l'on a trouvés en relation de disjonction indiquent des *diagnostics* possibles à l'occasion de ce symptôme.

Les termes de diagnostic et de symptôme pourraient d'ailleurs être employés aussi bien en Psychosociologie ou en Économie politique qu'en Médecine.

Remarquons que le même sujet peut faire l'objet de multiples diagnostics, selon que l'on porte un diagnostic de syndrome, de lésion, d'étiologie, etc... La logique formelle, aristotélicienne s'applique mal en médecine. Par contre la logistique peut s'appliquer et tout diagnostic représente l'un des ensembles dans lesquels on peut situer quiconque présente le symptôme ou les symptômes observés.

La relation de disjonction nous paraît devoir être surtout intéressante en Psychosociologie ou en Économie politique parce qu'elle peut révéler que le référentiel était, à l'insu de l'enquêteur, caractérisé par un symptôme. Elle doit amener à rechercher celui-ci et à reprendre l'enquête et ses conclusions.

La contrariété prouve que les phénomènes A et B comptent parmi leurs facteurs des alternatives contraires d'un même phénomène. Par exemple il y aura contrariété entre les accidents de la 1re dentition et ceux de la dent de sagesse, parce que leurs fréquences supposent une répartition différente des âges. Une population où l'une est fréquente, aura une répartition des âges telle que l'autre sera rare.

S'il n'y a ni homologie, ni contradiction, ni disjonction, ni contrariété, on considère P3 et P4.

Si la différence entre P2 et P3 dépasse en valeur absolue $\frac{3}{2}\sqrt{G}$, elle doit être tenue pour significative, et on considérera son sens.

Si $P2 > P3$, la fréquence de B implique celle de A.

Si $P3 > P2$, celle de A implique celle de B.

L'implication prouve que les conditions nécessaires pour observer le phénomène implicant contiennent celles nécessaires pour observer l'impliqué. La présence de bacille de Koch dans un crachat est un phénomène dont la fréquence implique celle d'une vitesse de sédimentation globulaire élevée. Cela signifie que la vitesse de sédimentation globulaire témoigne d'un processus infectieux, et que la présence de bacille de Koch dans les crachats témoigne d'un processus infectieux d'une certaine nature et d'une certaine localisation à un certain stade d'évolution.

L'implicant est un symptôme plus spécifique que l'impliqué, et cette notion pourrait valoir en Psycho sociologie ou en Économie politique aussi bien qu'en Médecine.

Lors même que nous trouvons une homologie entre la fréquence de A et celle de B, nous ne pouvons pas énoncer sans précaution que cette homologie soit un fait d'ordre général.

Il se peut que dans une population variée la fréquence de B implique celle de A. Soit β la fréquence de B et α celle de A, on a :

$$\beta = f(\alpha)$$

Mais nos échantillons sont constitués de telle façon qu'ils ne diffèrent nullement en ce qui concerne les facteurs de β autres que α, α et β dépendent des mêmes facteurs et ont une relation d'homologie.

En effet, on a :

$$\beta = f_1'(u, v, w...)$$

et, si

$$\alpha = f_1(u, v, w...)$$

Il en résulte :

$$\alpha = f(\beta) \qquad \beta = f(\alpha)$$

Par exemple, la fréquence élevée du jeu de la poupée dans une population enfantine implique la fréquence du sexe féminin. Mais si les fréquences sont établies sur des échantillons de population tous du même âge et de même éducation, c'est une homologie que l'on trouvera.

Mais en pareil cas $\beta = f(\alpha)$ est une fonction monotone, et la corrélation entre α et β par la méthode des rangs sera égale à $+1$. C'est-à-dire que l'on aura :

$$P2 = P3 = 0$$

Le même raisonnement peut être repris en cas de contradiction, avec un coefficient de corrélation $= -1$ et aussi une égalité :

$$P_3 = P_4 = 0$$

Il sera impossible d'affirmer d'une homologie avant d'avoir éliminé celle d'une implication d'après la relation qui existe entre P2 et P3, et celle-ci ne peut pas être interprétée si $P2 = P3 = 0$.

Il en sera de même en cas de contradiction.

D'ailleurs toute relation d'homologie est équivalente à une relation de contradiction en modifiant un terme.

Si dans une population il y a homologie entre le jeu de la poupée et le sexe féminin, il y aura contradiction entre le jeu de la poupée et le sexe masculin, et cette contradiction éveille les mêmes critiques.

Il y a cependant une complication supplémentaire.

En cas de contradiction absolue, $P1 = P4 = 0$.

Il n'est pas exclu que l'on ait entre P2 et P3 des inégalités qui nous ont amené à parler d'implication. L'implication peut être reconnue. Par exemple si même dans une population on appelle A le sexe masculin, non A le sexe féminin, B le fait de jouer à la poupée et non B le fait de ne pas y jouer, on a :

$$P1 = 0 \text{ et } P4 = 0$$

mais si $P2 < P3$, cela signifie qu'il y a, dans la population de l'âge considéré moins de garçons que de filles, et ne permet de conclure à aucune implication.

En résumé on parlera d'homologie ou de contradiction quand on ne pourra démontrer ni contrariété, ni disjonction, ni implication.

Cependant les conclusions ne seraient valables que dans les conditions d'expérience considérées.

Il y aura toujours une part d'induction nécessaire avant de formuler comme une loi d'ordre général la relation trouvée, et on devra se méfier fortement de ce genre d'induction quand la corrélation sera voisine de l'unité. On pourra dire, en termes familiers, que « la mariée est trop belle ».

DEUXIÈME PARTIE

On interprète l'Homologie entre les fréquences de 2 phénomènes A et B en disant que l'une et l'autre sont influencées dans le même sens par un ou plusieurs facteurs communs.

On interprête habituellement la contradiction en disant que les fréquences des phénomènes A et B sont influencées en sens opposés par un ou plusieurs facteurs communs.

Dire que les fréquences de A et B sont influencées par un facteur f, signifie que l'un et l'autre ont leur probabilité d'apparition influencée par celle d'un autre phénomène C.

Si C augmente les probabilités de A et de B, la fréquence de C est le facteur qui agit dans le même sens sur les fréquences de A et de B.

Il en est de même si C diminue les probabilités de A et de B.

Par contre si C augmente la probabilité de A et diminue celle de B, ou s'il augmente la probabilité de B et diminue celle de A, la fréquence de C est le facteur qui agit en sens opposés sur les fréquences de A et de B.

Mais il y a une hypothèse qu'il faut envisager.

Le phénomène A, tant que B n'est pas apparu, dépend de conditions qui n'interviennent nullement dans l'apparition de B. De même le phénomène B, tant que A n'est pas apparu, dépend de conditions qui n'interviennent nullement dans celle de A. Mais A, une fois apparu, favorise B, et B, une fois apparu, favorise A.

De nombreux exemples de ce genre de « réaction en chaîne » viennent à l'esprit en matière de Psychologie de la Vie sociale, mais aucun ne nous a paru pouvoir être énoncé de façon précise antérieurement à toute vérification expérimentale.

Remarquons que si on appelle y la fréquence de A et x celle de B, on a 2 fonctions linéaires :

$$y = f(x) \quad \text{et} \quad x = f(y)$$

Dans un tableau comme ceux que nous avons envisagés pour la détermination des « complexions » homologie, contradiction, contrariété, disjonction, implication, on trouvera indiscutablement qu'il y a homologie.

Nous y arriverons en utilisant le schéma qui nous a servi à établir les « coefficients de proximité ».

Considérons un espace ayant autant de dimensions qu'il y a de variables dont dépendent les fréquences de A et de B.

Soit O le point qui correspond au minimum de probabilité pour A, et P celui qui correspond au maximum de probabilité pour A.

Le segment $\overline{\mathrm{OP}}$ se situe dans un espace à n dimensions, où n est le nombre de variables dont dépend la fréquence de A.

Soit O' le point correspondant au minimum de probabilité pour B, et P′ celui correspondant au maximum de probabilité.

Le segment $\overline{\mathrm{O'P'}}$ se situe dans un espace à n' dimensions, où n' est le nombre de variables dont dépend la fréquence de B.

L'originalité de la situation est que le support de $\overline{\mathrm{OP}}$ trouve être l'une des dimensions du 2e espace, et que celui de OP′ se trouve être l'une des dimensions du 1re.

Donc le sous-espace formant intersection des deux espaces considérés contient $\overline{\mathrm{OP}}$ et $\overline{\mathrm{O'P'}}$. Il contient donc $\overline{\mathrm{PP'}}$ et $\overline{\mathrm{OO'}}$. Et dire que l'apparition de l'un quelconque des deux phénomènes favorise celle de l'autre signifie que $\overline{\mathrm{PP'}} < \overline{\mathrm{OO'}}$ ou, pour employer le langage que nous avons employé en exposant notre coefficient de proximité, $\overline{\mathrm{V}} < \overline{\mathrm{W}}$ et $\overline{\mathrm{W}} > \overline{\mathrm{V}}$.

On peut envisager une autre hypothèse.

Deux phénomènes A et B dépendent d'un facteur commun, mais la réalisation de A diminue les risques d'apparition de B, et réciproquement.

Une telle situation paraît vraisemblable dans certains cas. Je pense aux diverses maladies mentales dont les diagnostics s'excluent, même si cette exclusion est un peu artificielle. Je pense aux divers comportements qui sont considérés comme diverses manières de résoudre un même conflit. La représentation peut être étendue comme hypothèse à vérifier, en Médecine Somatique, devant des manifestations différentes d'allergie ou devant tous les symptômes que l'on doit être des équivalents de telle ou telle autre manifestation.

On peut la représenter de la façon suivante.

On a un ensemble E formé par la réunion de deux sous-ensembles E′ et E″ qui n'ont entre eux aucune intersection.

E′ est lui-même formé par la réunion de deux sous-ensembles plus petits S′ et S″ qui n'ont aucune intersection.

Le phénomène A est l'appartenance à S′.

Le phénomène B est l'appartenance à S″.

On voit que A et B ont un facteur commun qui est l'appartenance à E′.

Il y aura un lien entre les faibles fréquences de A et B, car elles se trouvent réalisées si une grande partie des tirages est effectuée en E″.

Mais il n'y aura pas de lien entre les très fortes fréquences de A et de B. En effet pour que A soit très fréquent, il ne suffit pas que beaucoup de tirages soient faits en E′, mais il faut encore qu'ils soient effectués en une partie déterminée de E′, à savoir en S′ et non en S″. De même pour que B soit très fréquent, il faut qu'un très grand nombre de tirages soit effectué en S″.

Tout se passera comme s'il y avait un facteur commun agissant de la même façon sur la fréquence de A et de B, et l'ensemble F est la zone où $\varpi_\alpha = 1$, l'ensemble G celle où $\varpi_\beta = 1$, l'ensemble E′ leur réunion, et l'ensemble E″ le reste de l'espace.

La zone où $\varpi_\beta = 1$ et celle où $\varpi_b = 1$, ont une partie commune qui est leur frontière, segment d'une droite $\overline{\mathrm{SS'}}$.

Soit M le milieu de ce segment.

Soit une droite $\overline{TT'} \perp \overline{SS'}$ et coupant $\overline{SS'}$ en M.

Sur TT′ nous prenons un point L très éloigné de M.

Déplaçons-nous de L vers M et, tout le long de ce déplacement, considérons les points considérés sur une droite perpendiculaire à T T′ et parallèle à S S′ dont elle se rapproche.

Nous voyons que, au cours de ce déplacement, nous avons très longtemps

$$\varpi_\alpha = 0$$
$$\varpi_\beta = 0$$

un autre facteur qui agit de façon inverse sur les fréquences de A et de B, mais seulement quand le 1re a atteint un seuil très élevé.

En réalité, nous pouvons reprendre le schéma des nuages que nous avons évoqué dans notre article sur une nouvelle manière d'étudier les relations entre aptitudes, et que l'on trouve dans le numéro du *Journal de la Société de Statistique de Paris,* d'octobre-novembre-décembre 1957, à la page 275 et par lequel nous expliquons pourquoi on a $\overline{V} > \overline{W}$.

L'ensemble E′ est une portion d'espace qui contient les points que nous avons appelés I et J dans notre schéma. Le point I correspondra à la conjoncture optimale pour observer A et le point J à la conjoncture optimale pour observer B. La probabilité ϖ_a d'observer A décroît régulièrement quand on s'éloigne de I pour tomber à zéro quand on sort de E′, et la probabilité ϖ_b d'observer B décroît quand on s'éloigne de J pour tomber à zéro quand on sort de E'. C'est la situation que nous avons déjà décrite, et telle que $\overline{V} > \overline{W}$, que nous ayons :

$$\overline{V} > \overline{W} > 0{,}5 \text{ (contradiction)}$$
$$0{,}5 > \overline{V} > \overline{W} \text{ (homologie)}$$
$$\overline{V} > 0{,}5 > \overline{W} \text{ (contrariété)}$$

Calculer $\overline{V}$ ou $\overline{W}$ est impossible. Il faudrait avoir 10 lots composés chacun de 100 échantillons de population et déterminer dans chaque échantillon la fréquence de A et de B. On classerait dans chaque lot les échantillons du point de vue de la fréquence de A et de celle de B. On déterminerait ensuite V_{ab}, V_{ba}, W_{ab}, W_{ba}. Une telle étude demanderait des centaines de milliers de sujets et il faut y renoncer.

On utilise le lien qui existe entre $\overline{V}$ et $\overline{W}$ et le coefficient de corrélation en cas de fonction linéaire (voir notre article), et la possibilité d'identifier le coefficient de corrélation au cœfficient de concordance.

Soit N le nombre total des cas;

R′ celui des cas où l'on observe à la fois A et B (P1 sur notre premier tableau);

R″ celui des cas où l'on n'observe ni A ni B (P4 sur notre premier tableau);

N′ = (N — R″) celui des cas où se trouvent réalisés soit A, soit B, soit les deux à la fois (P1 + P2 + P3 sur notre premier tableau);

N″ = (N — R′) celui des cas où manquent, soit A, soit B, soit les deux à la fois (P2 + P3 + P4 sur notre premier tableau).

$$\rho = \frac{2\,R' - (N' - R')}{N'} = \frac{3\,R' - N'}{N'}$$

$$1 - \rho = \frac{N' - (3\,R' - N')}{N'} = \frac{2\,N' - 3\,R'}{N'}$$

Et nous remplacerons :

$$\overline{V} = \frac{1 - \rho}{2}$$

par :

$$\gamma = \frac{2\,N' - 3\,R'}{2\,N'}$$

Et nous remplacerons $\overline{W}$ par :

$$\delta = \frac{2\,N'' - 3\,R''}{2\,N'}$$

Mais nous avons ainsi deux coefficients faciles à déterminer et dont les inégalités ont les mêmes significations que celles des coefficients $\overline{V}$ et $\overline{W}$.

La distribution des valeurs possibles est par ailleurs facile à déterminer.

$$\gamma = \frac{2\,N' - 3\,R'}{2\,N'}$$

a le même écart type que $\frac{3\,R'}{2\,N'}$.

Mais celui-ci est facile à calculer.

Sur N' tirages, nous avons chaque fois une probabilité $\frac{R'}{N'}$ de concordance et $1 - \frac{R'}{N'}$ de discordance.

L'écart type des concordances ou des discordances est donc :

$$\sqrt{N'\,\frac{R'}{N'}\left(1 - \frac{R'}{N'}\right)} = \sqrt{R'\left(1 - \frac{R'}{N'}\right)} = \sqrt{\frac{R'\,N' - R'^2}{N'}} = \frac{\sqrt{R'\,(N' - R')}}{\sqrt{N'}}$$

Tel est l'écart type de R'

Celui de $\frac{3\,R'}{2\,N'}$ est $\frac{\sqrt{R'\,(N' - R')}}{\sqrt{N'}}\;\frac{\sqrt{3}}{\sqrt{2\,N'}} = \frac{\sqrt{R'\,(N' - R')}}{N}\sqrt{\frac{3}{2}}$

Tel est celui de γ

$$\sigma_\gamma = \frac{\sqrt{3\,R'\,(N' - R')}}{N'\sqrt{2}}$$

Et de même

$$\sigma_\delta = \frac{\sqrt{3\,R''\,(N'' - R'')}}{N''\sqrt{2}}$$

On déterminera alors si l'écart entre γ et δ est significatif, en calculant le τ de Student :

$$s = N'\,\sigma_\gamma^2 + N''\,\sigma_\delta^2$$

$$\tau = \frac{\gamma - \delta}{s\sqrt{\frac{1}{N'} + \frac{1}{N''}}}$$

En résumé

1) Après détermination d'un χ^2 qui prouve l'existence d'un lien entre les fréquences de deux phénomènes, une série de calculs très simples permettent de dire s'il y a homologie, contradiction, implication, contrariété, ou disjonction entre ces phénomènes.

2) Si on trouve une homologie, celle-ci, valable dans la population étudiée, ne peut être énoncée qu'avec beaucoup de prudence comme une loi générale. Il faut surtout se méfier des homologies les plus fortes. La même méfiance s'impose pour l'interprétation des contradictions qui paraissent absolues. Dans les deux cas nous pouvons être en présence d'un cas particulier d'une loi qui est en réalité une loi d'implication.

3) L'homologie peut être due à un facteur commun à la fréquence des deux phénomènes, mais elle peut être due aussi à ce que la présence du phénomène A favorise l'apparition du phénomène B et à ce que la présence du phénomène B favorise celle de A. Une vérification mathématique permet de retenir la 2e hypothèse et, sans exclure la vérité possible de la 1re hypothèse, car les 2 hypothèses ne sont pas contraires, la rend inutile. Ce genre de problème aurait sans doute beaucoup d'applications en Médecine comme dans la Psychologie et la Vie sociale, et probablement dans d'autres domaines.

4) L'hypothèse que l'apparition du phénomène A s'oppose à celle du phénomène B et celle du phénomène B au phénomène A est compatible avec des relations d'homologie, contradiction ou contrariété. Il est nécessaire et possible de le vérifier en cas d'homologie.

TROISIÈME PARTIE

Application a l'analyse hiérarchique

On sait que l'analyse hiérarchique oblige à considérer un grand nombre de « trajets ». Soit, par exemple, à hiérarchiser 30 réponses, le nombre de trajets à étudier est :

$$\frac{30\,!}{2\,!\;28\,!} = 435$$

S'il faut les étudier tous selon nos critères, la tâche devient très compliquée.

Nous trouverons la possibilité de le simplifier en expliquant pourquoi certaines relations peuvent être intransitives.

Soit P_a une fréquence élevée ($> 1/2$) de réponses favorables à la question a.

Soit P_b une fréquence élevée ($> 1/2$) de réponses favorables à la question b. etc.

$$P_a \longrightarrow P_b$$
$$P_b \longrightarrow P_c$$

On peut ne pas avoir $\qquad P_a \longrightarrow P_c$

Comment cela est-il possible?

Soit A l'ensemble des sujets qui répondront favorablement à *a*.

« B « « « *b*.

« C « « « *c*.

α l'intersection de A et B;

β l'intersection de B et C.

On peut avoir :

$$\frac{A}{2} < \alpha < B$$

$$\frac{B}{2} < \beta < C$$

ce qui donne $A \longrightarrow B$ et $B \longrightarrow C$ alors que

$$\frac{B}{2} + B' = \beta \quad \text{et} \quad \frac{B}{2} + B' + B'' = B$$

Soit γ l'intersection de A et C, on a $\gamma = \varnothing$ si $A < \frac{C}{2}$.

Donc, prenons les sujets qui présentent tous un caractère donné i. Cherchons les autres caractères rencontrés parmi eux et classons-les par ordre de fréquence décroissante parmi les possesseurs du caractère i jusqu'à ce que la fréquence trouvée soit $= \frac{1}{2}$. Nous avons une suite transitive.

Par exemple on a $(a\,b\,c\,d) - (c\,e\,f\,g)$ et $(d\,h\,i\,j)$.

On aura une suite $a\,b\,c\,f\,g$ et une suite $a\,b\,c\,d\,h\,i\,j$.

Chacune forme un sous-ensemble H où la hiérarchie est strictement transitive.

A chaque sous-ensemble H de caractère hiérarchisé d'une certaine façon, nous pouvons faire correspondre le sous-ensemble S de sujets dont les caractères ne sont pas contraires à cette hiérarchie. Si, par exemple, un sujet présente le caractère classé 8e en H_i sans présenter celui classé 5e, il n'appartient pas à S_i.

On a donc :

D'une part, des sous-ensembles H de caractères appartenant à une même hiérarchie; d'autre part des sous-ensembles S de sujets dont les caractères correspondent à une hiérarchie.

Deux éléments ne peuvent se trouver à des places différentes sur les trajets représentatifs de 2 ensembles H. Donc le nombre maximum d'ensembles H est réalisé quand chacun ne comporte que 2 éléments et qu'il n'y a aucune intersection entre eux.

Donc, avec 30 caractères, en supposant que ceux-ci n'aient aucun lien entre eux, les ensembles H, si leur constitution est possible, ne seront pas en nombre supérieur à 15. En pratique, si tous ces caractères ont été choisis pour exprimer quelque chose, ils auront des liens, et leur nombre sera beaucoup plus faible.

Le nombre des sous-ensembles S est nécessairement égal à celui des sous-ensembles H, donc faible.

Mais les sous-ensembles S ont entre eux des intersections qui ne correspondent pas à celles des ensembles H.

Par exemple, un sous-ensemble H_a est défini par la hiérarchie (a, b, c, d, e, f), et un sous-ensemble H_i par la hiérarchie (i, j, k, l, m, n). Ils n'ont aucune intersection. Mais un sujet qui présente les caractères (a, b, c, i, j, k, l) appartient à l'intersection de S_a et de S_i.

L'analyse hiérarchique se fera alors en 2 temps :

1) la détermination des ensembles H peu nombreux;

2) l'étude des relations logistiques entre les fréquences des appartenances aux ensembles S qui leur correspondent.

Fred MILHAUD.