Étude statistique d'un problème de sociométrie

par M. M. P. SCHÜTZENBERGER.

L'article est consacré à la formulation du traitement statistique de certains aspects d'une enquête sociométrique « par choix ». Un exemple d'application est donné, qui conduit à rejeter l'hypothèse nulle.

Le travail suivant est issu d'une enquête du groupe de Psychométrie Pédagogique (G.P.P.) sous la direction de M. P. H. MAUCORPS au cours de l'année 1945-1946 dans quatre classes de préparation de l'Ecole Navale. L'une des séries de questions posées aux élèves était la suivante :

« Parmi les élèves qui appartiennent à la même classe de préparation que vous, désignez :

A. - Votre meilleur ami ;
B. - Vos quatre meilleurs amis (y compris le précédent) ;
C. - Les trois élèves qui vous sont les plus antipathiques ».

Parmi les multiples problèmes qui peuvent se poser au cours du dépouillement de semblables matériaux, on a choisi de traiter ici la question suivante :

Dans quelle mesure peut-on affirmer que le nombre de

réciprocités — c'est-à-dire de désignations (choix) simultanés du sujet *i* par le sujet *j* et du sujet *j* par le sujet *i* est supérieur à celui qui résulterait d'une répartition aléatoire de ces mêmes choix.

C'est ce qui constitue « l'hypothèse nulle » (Ho) en fonction de laquelle seront calculées toutes les statistiques suivantes.

On établira des tests permettant de savoir quelle est la probabilité pour que dans « l'hypothèse nulle » les valeurs observées diffèrent d'une certaine mesure des valeurs théoriques. Quand cette probabilité tombera au-dessous d'un certain niveau (en général 5 %, 1 % ou 1 °/00) on sera amené à écarter l'hypothèse nulle. Ceci constitue le mécanisme habituel du raisonnement statistique.

Les formules approchées précédemment calculées par MORENO et JENNINGS (Statistics of Social configuration - Sociometry 1938 pp. 342-374), ne s'appliquant que difficilement au cas actuel où il n'a pas été possible de recueillir toutes les réponses, on a été conduit à formuler le problème d'une manière rigoureuse et générale permettant un test plus sûr.

Notations.

Soient 1,2....., n les n sujets dont on a recueilli les choix (sujets « actifs ») et n+1 ..... N les N — n sujets du groupe qui ont pu être désignés mais dont les choix sont inconnus (sujets « passifs », — soit k le nombre fixe des choix émis par chaque sujet (ici : $k_A = 1$, $k_B = 4$, $k_C = 3$).

On aura avantage à représenter l'ensemble de l'enquête par une matrice U d'ordre nxU où les éléments a_j^i (à l'intersection de la colonne i et de la ligne j) sont égaux à 1 ou nuls suivant que le j ème sujet a ou non choisi le i ème.

On appellera réciprocité C_{ij} l'événement

$$a_1^j \times a_j^i = 1$$

On a évidemment :

$a_i^i = o$ (aucun sujet ne peut se choisir lui-même)

$\sum a_j^i = k$

$o \leqslant i < N + 1$

La probabilité a priori de a $\frac{i}{j} = 1$ ($i =/= j$) sera $p\frac{i}{j}$ et l'hypothèse nulle Ho sera :

1° $p\frac{i}{j} = \frac{k}{N-1} = p$

2° Prob. C i j $= p\frac{i}{j}\ p\frac{j}{i}$

On posera en outre : $q = 1 - p = \frac{N-k-1}{N-1}$

$$d = n/N$$

Par abréviation on appellera :

1° « cas carré » le cas où $d = 1$
(Tous les sujets sont « actifs »).

2° « cas de Poisson » le cas infini où, k restant fixe, d a une limite non nulle δ_1

Formules préliminaires.

Soit fx le nombre de sujets ayant reçu x choix ; il convient d'abord de tester l'accord de la repartition des fx avec l'hypothèse nulle Ho.

Les valeurs théoriques de fx sont :

$$f_x = \left[\begin{matrix} n \\ x \end{matrix}\right] p^x q^{n-x-1}(n-x+q^{x-n}) \quad *$$

Ce qui conduit à une répartition dont les valeurs typiques sont

$$m = p\,d\,(N-1) \quad ; \quad \sigma^2 = (N-1)\,p\,d\,(1-p\,d)$$

Un autre moyen de tester Ho serait de comparer le nombre de sujets n'ayant jamais été choisi à sa valeur a priori fo

En posant :

$$\theta_r = \frac{(N-p)!\,(N-k-1)!}{(N-1)!\,(N-k-r)!}$$

On emploiera la notation $\left[\begin{matrix} a \\ b \end{matrix}\right]$ pour les coefficients binomaux.

* $\left[\begin{matrix} A \\ a \end{matrix}\right] = \frac{A!}{(A-a)!\,a!} = C_A^a$

on trouve les moments factoriels de la répartition de fo

$$B_r = (\theta_{r+1})^n \sum_{f=0}^{n+1} \left[\begin{matrix} n \\ f \end{matrix}\right] \left[\begin{matrix} N-n \\ r-f \end{matrix}\right] \left(\frac{N-r}{N-r-K}\right)^f$$

$$= (\theta_{r+1})^n \sum_{f=0}^{r+1} \left[\begin{matrix} n \\ f \end{matrix}\right] \left[\begin{matrix} N-f \\ r-f \end{matrix}\right] \left(\frac{K}{N-r-K}\right)^f$$

Ces formules ne deviennent maniables que dans le cas carré où :

$$B_r = \left[\begin{matrix} N \\ r \end{matrix}\right] (\theta_{r+1})^{n-r} (\theta_r)^r$$

Cependant on a toujours pour valeur moyenne de fo :

$$B_1 = \bar{f}_o = N q^{n-1}(q + dp)$$

Et dans le cas de Poisson $\overline{fo}$ tend vers $e^{-k\delta}$

Réciprocités.

On formulera d'abord un théorème général relatif à la répartition *a priori* de la répétition d'une configuration contenue dans une sous-matrice diagonale U' de U.

Théorème. - Pour un U' formé de l_1 lignes à 1 choix.
l_2 lignes à 2 choix.
l_3 lignes à 3 choix
et dont les lignes sont elles-mêmes fixées, la probabilité a priori est :

$$P_{U'} = \left[\begin{matrix} N-1 \\ K \end{matrix}\right]^{-n} \prod_{i=0}^{K+1} \left[\begin{matrix} N-i-1 \\ K-1 \end{matrix}\right]^{l_i} = \prod_{i} \left\{\frac{K!}{(K-i)!} \times \frac{(N-i-1)!}{(N-1)!}\right\}^{l_i}$$

Par conséquent si g est l'ordre du groupe de substitution de U' la valeur moyenne du nombre de répétition de U_i dans l'hypothèse H_o est

$$\bar{r}_{U'} = P_{U'} \times \frac{n!}{(n-\Sigma l_i)!} \times g^{-1}$$

Calculs des moments factoriels.

On considère successivement la matrice U' d'ordre 2×2 que constitue une réciprocité puis les matrices U' constituant une double réciprocité et l'on obtient les deux premiers moments factoriels d'où les valeurs typiques

011	0100
100	1000
106	0001
	0010

$$\bar{r} = p^2 \frac{n(n-1)}{2} \quad ; \quad \sigma^2 = \bar{r}\, q \left(q + 2 \frac{N-n}{N-2} p \right)$$

d'où l'on déduit que σ_r^2 est *toujours* compris entre $\bar{r}$ et $\bar{r}q^2$.

Le calcul des moments du 3e ordre entraînerait l'application du théorème général aux matrices suivantes :

0111	0110	00010	010000	011
1000	1001	00001	100000	101
1000	1000	10000	000100	110
1000	0100	10000	001001	
		01100	000001	

d'où la valeur suivante du moment factoriel de troisième ordre :

$$B_3 = \bar{r}\, p^4 \frac{(n-2)(n-3)(n-4)(n-5)}{24} + \bar{r}\, p^3 \frac{(n-2)(n-3)(n-4)}{2} \frac{K-1}{N-2}$$
$$+ \bar{r}\, p^2 (n-2)(n-3) \left(\frac{K-1}{N-2} \right)^2 + \bar{r}\, p^2 \frac{(n-2)(n-3)}{3} \frac{(K-1)(K-2)}{(N-2)(N-3)}$$
$$+ \bar{r}\, p \frac{(n-2)}{3} \left(\frac{K-1}{N-2} \right)^3$$

L'expression du moment du 4e ordre serait encore plus compliquée (11 termes). Cependant, pour k = 1, on a les formules simples :

$$B_r = \frac{n!}{(n-2r)!\, r!}\, 2^{-r} (N-1)^{-2r}$$

car une seule matrice U' est seulement à considérer et l'emploi des inégalités de Loeve ou de Bonferati est extrêmement facile.

Cas de Poisson.

Il en est ici de même car la probabilité des U' ayant deux choix dans une ligne est infiniment petite par rapport aux U' n'ayant qu'un choix par ligne.

Un calcul classique montre la répétition des réciprocités, suit une loi de Poisson de paramètre $k^2 \delta^2/2$

Il est vraisemblable que dans la pratique et dès que $k/N-1$ est suffisamment petit, cette approximation est applicable au cas fini.

Données incomplètes.

Ce cas est celui où les sujets n'ont pas tous effectué le même nombre de κ de choix ; il est encore possible, si l'on connaît pour chaque sujet actif les nombres k_i et h_i de choix *émis* et reçus, d'obtenir une *valeur rigoureuse* de $\bar{r}$:

$$\bar{r} = \frac{1}{2} \sum_{i=0}^{n+1} \frac{k_i K_i}{N-1} .$$

que l'on pourra utiliser en 1re approximation comme paramètre d'une loi de Poisson (si N est petit).

Il faut signaler que dans le cas carré et quand tous les k_i sont égaux à k l'on retrouve la valeur précédemment calculée dans l'hypothèse Ho.

Ainsi donc, dans ce cas, la connaissance complète de la distribution marginale des choix ne *change en rien* la valeur moyenne a priori du nombre des réciprocités.

Extension.

Le théorème général permet également d'étudier des configurations plus complexes : par exemple les chaînes d'ordre λ c'est-à-dire des événements C (i, j, k...) constitués par le choix simultané de i par j, de j par k, de k par I etc...(λ sujets) : on trouve encore pour valeur moyenne du nombre des répétitions de chaîne d'ordre λ :

$$\overline{r}_\lambda = \frac{n!}{(n-\lambda)!} \frac{(N-1)^{-\lambda}}{\lambda} K^\lambda$$

et dans le cas de Poisson une distribution de Poisson de paramètre :

$$K^\lambda \hat{o}^\lambda \lambda^{-1}$$

On peut également (toujours par application du théorème général) chercher la valeur moyenne du nombre de « blocs compacts d'ordre μ » c'est-à-dire d'ensemble de μ ($\mu \leq k + 1$) sujets se choisissant tous entre eux; les matrices U' sont alors :

$\mu = 2$	$\mu = 3$	$\mu = 4$
01	011	0111
10	101	1011
(réciprocité)	110	1101 etc...
	(double chaîne d'ordre 3)	1110

on a sans peine :

$$\overline{r}_\mu = \binom{n}{\mu} \times \left\{ \frac{K!}{(K-\mu+1)!} \times \frac{(N-\mu)!}{(N-1)!} \right\}^\mu$$

$\overline{r}_\mu$ tend vers zéro pour $\mu \geq 3$ 3 lorsque, k étant fixe, tend vers l'infini.

Application à l'enquête G.P.P.

On utilisera les résultats précédents pour tester l'hypothèse nulle relativement à chacun des trois types de choix indiqués.

Le tableau I donne pour chacune des quatre classes le nombre N d'élèves dans la classe, le nombre de sujets ayant répondu au questionnaire et les nombres AA, BB et CC de réciprocités pour les choix A. B. et C.

Le tableau II, donnant les valeurs typiques $\overline{r}$ et σ^2 pour chacun de ces $4 \times 3 = 12$ cas, permet de voir qu'il serait fort légitime d'utiliser une répartition de Poisson pour tester Ho. Cependant dans le cas AA, pour la 1re et la 2e classe, k étant égal à 1, on a appliqué les formules classiques :

$$P_r = B_r - \binom{r}{1} B_{r+1} + \binom{r+1}{2} B_{r+2} - \ldots$$

TABLEAU I.

Classe	N	n	AA k=1	BB k=4	CC k=3
1	44	36	3	18	5
2	36	33	6	22	10
3	39	27	3	17	2
4	37	20	7	17	2

Données de l'enquête :

N = nombre de sujets dans la classe.

n = nombre de sujets dont les réponses ont été recueillies.

AA = nombre de réciprocités pour le 1er choix.

BB = nombre de réciprocités pour le 2e choix.

CC = nombre de réciprocités pour le 3e choix.

TABLEAU II.

Valeur théorique dans l'hypothèse nulle.

Classe	AA Réciprocités	Réciprocités BB	Réciprocités CC
1	$\bar{r} = 0{,}341$ $\sigma^2 = 0{,}328$ $\sigma = 0{,}573$	= 5,45 = 4,66 = 2,16	= 3,07 = 2,73 = 1,65
2	$\bar{r} = 0{,}431$ $\sigma^2 = 0{,}404$ $\sigma = 0{,}639$	= 6,90 = 5,54 = 2,35	= 3,88 = 2,62 = 1,62
3	$\bar{r} = 0{,}243$ $\sigma^2 = 0{,}234$ $\sigma = 0{,}436$	= 3,89 = 3,35 = 1,83	= 2,19 = 1,96 = 1,40
4	$\bar{r} = 0{,}147$ $\sigma^2 = 0{,}142$ $\sigma = 0{,}377$	= 2,34 = 2,07 = 1,44	= 1,32 = 1,21 = 1,10

$\bar{r}$ = valeur moyenne du nombre de réciprocités.

σ^2 = écart type » » »

qui permettent d'affirmer rigoureusement que l'hypothèse nulle (évidemment rejetée dans la 2e et la 4e classe a beaucoup plus de 1 0/00) doit également l'être dans la 1re et la 3e classe, avec des probabilités de 13 o/oo et de 14 o/oo respectivement.

La signification des écarts entre les valeurs théoriques et les valeurs observées pour BB est hors de contestation (plus d'une chance sur mille), soit que l'on emploie une approximation de Poisson (classes 3 et 4) soit que l'on utilise la règle des 3σ (classes 1 et 2).

En ce qui concerne CC il semble difficile de rejeter Ho pour les classes 1, 3 et 4 où la valeur empirique est très voisine de la moyenne. L'étude de la classe 2 est relativement facile : en effet le théorème général donne l'expression de chacun des termes qui contribuent à former les moments factoriels et il est aisé de voir que si une matrice U' (pour l'ordre ρ) est une matrice $(2\rho - x)$ $(2\rho - x)$ l'ordre en n et N du terme correspondant est — x ; l'on est ainsi amené à ne prendre en considération que les U' d'ordre 2ρ, $2\rho - 1$ et $2\rho - 2$, les autres U' n'introduisent que des différences de l'ordre de 1°/10000e chacune. On trouve aussi que la probabilité d'atteindre par le seul jeu des fluctuations 10 réciprocités est inférieure à 4 ‰.

RESUME ET CONCLUSIONS

Diverses statistiques ont été établies pour analyser une « enquête par choix » dans quatre classes de préparation aux grandes écoles.

La valeur exacte de la moyenne et de l'écart type de la répétition du nombre de choix réciproques ont été calculées dans l'hypothèse nulle de l'indépendance et de l'équiprobabilité des choix.

Une application en a été faite d'où il semble ressortir que les antipathies ne manifestent le phénomène de réciprocités qu'à un degré beaucoup moindre que les sympathies pour lesquelles l'hypothèse de la répartition indépendante des choix est absolument à rejeter.

H. MARCHAND & Cie, imprimeurs, PARIS. Dépôt légal : 1er trimestre 1948.
Le Directeur-Gérant : P. GAVAUDAN.