

REVUE DE STATISTIQUE APPLIQUÉE

CH. GILLOT

H. CAUSSINUS

Sur un modèle de comparaisons par paires avec une échelle de réponses à trois valeurs

Revue de statistique appliquée, tome 14, n° 3 (1966), p. 31-42

http://www.numdam.org/item?id=RSA_1966__14_3_31_0

© Société française de statistique, 1966, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

SUR UN MODÈLE DE COMPARAISONS PAR PAIRES AVEC UNE ÉCHELLE DE RÉPONSES A TROIS VALEURS

Ch. GILLOT et H. CAUSSINUS
(Laboratoire de Statistique
Faculté des Sciences de TOULOUSE)

I - INTRODUCTION

Un certain nombre de modèles de comparaisons par paires font appel à une échelle de réponses à deux valeurs. Ainsi, lors de la comparaison des objets A et B, le "juge" doit obligatoirement donner la préférence soit à l'objet A, soit à l'objet B. Mr Vessereau a présenté ces modèles dans un article récent de cette revue⁽¹⁾. Il a donné à cette occasion une importante bibliographie que nous ne reproduirons pas ici.

Dans la pratique, et sur la base d'un critère de choix donné, il se peut que l'on ne sache pas discerner lequel, de A ou de B, prévaut. Si un tel fait se produit, et si, d'autre part, le cas "d'équivalence" n'a pas été prévu, on peut penser que l'un des deux objets, sera classé supérieur à l'autre, au hasard.

II - PRESENTATION DU MODELE

Dans le modèle que nous exposons, nous avons introduit une échelle de réponses à trois valeurs, qui tient compte de ce cas critique.

On conçoit mieux l'utilité de cette dernière échelle si l'on raisonne non pas sur des objets à classer, mais en termes de compétition de joueurs, ce qui ne diminue pas la généralité de cette étude (cf. David, Tournaments and paired comparisons, *Biometrika*, 46).

On convient donc qu'un certain nombre de joueurs s'affrontent deux à deux, de telle façon qu'un joueur donné rencontre tous ses adversaires.

Pour la commodité du calcul, on considère $n + 1$ joueurs (que l'on appellera $C_1, C_2, \dots, C_{n+1}$), ce qui donne $\frac{n(n+1)}{2}$ parties effectivement jouées.

A l'issue de la confrontation (C_i, C_j), trois éventualités pourront se produire :

- 1/ C_i a battu C_j
- 2/ C_j a battu C_i
- 3/ C_i et C_j ont fait partie nulle.

(1) Les méthodes statistiques appliquées au test des caractères organoleptiques - Revue de Statistique Appliquée 1965. Vol. XIII n° 3.

Hypothèse nulle

Nous supposons a priori que tous les joueurs sont de même force, ce qui conduit à écrire, en notant " $C_i > C_j$ " le fait que C_i bat C_j et " $C_i = C_j$ " la partie nulle entre C_i et C_j :

$$\Pr(C_i > C_j) = p$$

$$\Pr(C_j > C_i) = p$$

$$\Pr(C_i = C_j) = q$$

pour $i \neq j$

avec : $0 \leq p \leq \frac{1}{2}, \quad 0 \leq q \leq 1$

p et q liés par la relation :

$$2p + q = 1 \quad (1)$$

Nous supposons en outre que les résultats des rencontres entre un joueur donné et ses divers adversaires sont indépendants entre eux.

Cotation des résultats

La variable aléatoire X_{ij} représente le résultat de la partie jouée entre C_i et C_j .

Nous avons adopté l'échelle de notes suivantes :

- Si C_i est battu, il reçoit la note -1 ($X_{ij} = -1$).
- Si C_i fait partie nulle avec C_j , tous deux ont la note 0. ($X_{ij} = 0$).

- Si C_i bat C_j , C_i reçoit la note $+1$. ($X_{ij} = +1$)

On remarque que, quels que soient i et j , on a la relation :

$$X_{ij} + X_{ji} = 0 \quad (i, j = 1, 2, \dots, n+1) \quad (2)$$

L'hypothèse nulle s'écrit alors :

$$\Pr(X_{ij} = -1) = p$$

$$\Pr(X_{ij} = 0) = q \quad (3)$$

$$\Pr(X_{ij} = +1) = p$$

pour $i \neq j$

On conviendra en outre que X_{ii} prend la valeur zéro avec la probabilité 1. Le "score" du joueur C_i sera la variable aléatoire :

$$X_i = \sum_j X_{ij}$$

Le tableau (4) rend compte des résultats de la compétition et du score de chaque joueur :

$$\begin{array}{c}
C_1 \\
C_2 \\
C_3 \\
\vdots \\
C_{n+1}
\end{array}
\begin{bmatrix}
C_1 & C_2 & C_3 & \dots & C_{n+1} \\
0 & X_{12} & X_{13} & \dots & X_{1,n+1} \\
X_{21} & 0 & X_{23} & \dots & X_{2,n+1} \\
X_{31} & X_{32} & 0 & \dots & X_{3,n+1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
X_{n+1,1} & X_{n+1,2} & X_{n+1,3} & \dots & 0
\end{bmatrix}
\begin{array}{c}
\text{Scores} \\
X_1 = \sum_j X_{1j} \\
X_2 = \sum_j X_{2j} \\
X_3 = \sum_j X_{3j} \\
\vdots \\
X_{n+1} = \sum_j X_{n+1,j}
\end{array}
\quad (4)$$

En tenant compte de (2), on voit que la somme de tous les scores X_i est nulle,

$$\sum_{i=1}^{n+1} X_i = 0$$

III - LOI DE PROBABILITE DU VECTEUR SCORE

On appellera vecteur score l'ensemble des scores des $(n + 1)$ joueurs. Le calcul de la loi de probabilité d'un vecteur score peut s'effectuer par "remontée", en calculant la loi de probabilité marginale du score d'un joueur unique puis la loi de probabilité conjointe de deux scores, puis de trois, etc ...

Loi de probabilité du score d'un joueur

Notons que, d'après les hypothèses, toutes les variables aléatoires X_i ont même distribution, il suffira donc de considérer par exemple $X_1 = \sum_j X_{1j}$.

La fonction génératrice des probabilités de X_{1j} soit $E(t^{X_{1j}})$ est $q + p(t + t^{-1})$.

Les variables aléatoires X_{1j} étant indépendantes, d'après notre hypothèse, la fonction génératrice des probabilités de X_1 est :

$$E(t^{X_1}) = g_{X_1}(t) = [q + p(t + t^{-1})]^n = \sum_{k=0}^n C_n^k p^k (t + t^{-1})^k q^{n-k}$$

soit :

$$g_{X_1}(t) = \sum_{k=0}^n \sum_{i=0}^k C_n^k p^k q^{n-k} C_k^i t^{2i-k} \quad (5)$$

Or, on a aussi :

$$g_{X_1}(t) = \sum_{x=-n}^{+n} t^x \Pr(X_1 = x) \quad (6)$$

(il est évident que X_1 est compris entre $-n$ et $+n$).

En identifiant les coefficients des termes en t^x dans (5) et (6) on obtient la loi de probabilité du score X_1 du joueur C_1 :

$$\Pr(X_1 = x) = \sum_{k \equiv |x|}^n C_n^k C_k^{\frac{k+|x|}{2}} p^k q^{n-k} \quad (7)$$

le signe $\sum'$ indiquant que la sommation doit s'effectuer sur les valeurs de k ayant même parité que x .

MOMENTS DE CETTE LOI DE PROBABILITE

$E(X_{1j}) = 0$ et $\text{Var}(X_{1j}) = 2p$ d'où :

$$E(X_1) = \sum_{j=2}^{n+1} E(X_{1j}) = 0 \quad (8)$$

et puisque les variables aléatoires X_{1j} sont indépendantes :

$$\text{Var}(X_1) = \sum_{j=2}^{n+1} \text{Var}(X_{1j}) = 2np \quad (9)$$

LOI DE PROBABILITE CONJOINTE DES SCORES DE DEUX JOUEURS .

On considère les joueurs C_1 et C_2 , et l'on cherche à évaluer

$$\Pr[X_1 = x_1, X_2 = x_2]$$

x_1 et x_2 étant des entiers compris entre $-n$ et $+n$.

On peut décomposer les deux premières lignes du tableau (4) de la façon suivante :

	C_1	C_2	C_3	C_4	C_5	...	C_{n+1}	Scores
C_1	0	X_{12}	X_{13}	X_{14}	X_{15}	...	$X_{1,n+1}$	$X_1 = \sum_j X_{1j}$
C_2	X_{21}	0	X_{23}	X_{24}	X_{25}	...	$X_{2,n+1}$	$X_2 = \sum_j X_{2j}$
C_3	X_{31}	X_{32}	0	X_{34}	X_{35}	...	$X_{3,n+1}$	...
....								...
C_{n+1}	$X_{n+1,1}$						0	...

D'une part l'élément X_{12} , auquel est associé d'après (2) l'élément X_{21} .
D'autre part, les ensembles :

$$(X_{13}, X_{14}, \dots, X_{1,n+1}) \text{ et } (X_{23}, X_{24}, \dots, X_{2,n+1})$$

Cette façon de partager les X_{1j} présente l'avantage de fournir des ensembles de variables aléatoires indépendantes entre elles, les scores X_1 et X_2 n'étant reliés que par l'intermédiaire de X_{12} et X_{21} .

On peut donc obtenir aisément la probabilité cherchée :

$$\begin{aligned} \Pr[X_1 = x_1, X_2 = x_2] &= \\ &= \sum_{k=-1}^{+1} \Pr[X_{12} = k]. \Pr[X_1 - X_{12} = x_1 - k]. \Pr[X_2 - X_{21} = x_2 + k] \end{aligned}$$

En effet, x_1 étant la valeur du score X_1 , si X_{12} prend la valeur k , la quantité $X_1 - X_{12}$, qui n'est autre que la somme des variables :

$$X_{13}, X_{14}, \dots, X_{1,n+1}$$

(encadrées dans la première ligne du tableau précédent) vaudra $x_1 - k$.

De même, x_2 étant la valeur de X_2 , puisque $X_{21} = -X_{12} = -k$, la quantité $X_2 - X_{21}$ vaudra $x_2 + k$.

On remarque que les probabilités du type

$$\Pr[X_1 - X_{12} = x_1 - k]$$

sont de la forme (7), mais avec $(n - 1)$ compétitions au lieu de n .

Introduisons la fonction π définie de la façon suivante

$$\pi(k) = \begin{cases} p & \text{si } k = \pm 1 \\ q & \text{si } k = 0 \\ 0 & \text{pour les autres valeurs de } k \end{cases}$$

L'expression (10) s'écrit alors, sous forme explicite :

$$\begin{aligned} \Pr[X_1 = x_1, X_2 = x_2] &= \\ &= \sum_{k=-1}^{+1} \pi(k) \left[\sum_{i=|x_1-k|}^{n-1} C_{n-1}^i C_1^{\frac{i+|x_1-k|}{2}} p^i q^{n-i-1} \right] \left[\sum_{j=|x_2+k|}^{n-1} C_{n-1}^j C_j^{\frac{j+|x_2+k|}{2}} p^j q^{n-1-j} \right] \end{aligned}$$

Le signe $\sum'$ ayant la même signification que dans (7).

Corrélation entre 2 scores

$$\begin{aligned} E[X_1 X_2] &= E \left[\left(X_{12} + \sum_{j=3}^{n+1} X_{2j} \right) \left(X_{21} + \sum_{j=3}^{n+1} X_{2j} \right) \right] \\ &= E(X_{12} X_{21}) + E \left(X_{12} \sum_{j=3}^{n+1} X_{2j} \right) + E \left(X_{21} \sum_{j=3}^{n+1} X_{1j} \right) + \\ &+ E \left(\sum_{j=3}^{n+1} X_{1j} \sum_{j=3}^{n+1} X_{2j} \right) \end{aligned}$$

En utilisant l'indépendance des variables telles que X_{12} et $\sum_{j=3}^{n+1} X_{2j}$, et la relation $E(X_{12} X_{21}) = -2p$:

$$E(X_1 X_2) = \text{Cov}(X_1 X_2) = -2p. \quad (11)$$

le coefficient de corrélation entre X_1 et X_2 est donc :

$$\frac{-2p}{2np} = \frac{-1}{n} \quad (12)$$

De (11) et (9) on déduit la matrice Σ_x des moments centrés du deuxième ordre du vecteur score :

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{n+1} \end{bmatrix}$$

$$\Sigma_x = 2p \mid \mid (n+1) \delta_{ij} - 1 \mid \mid \quad (\delta_{ij} \text{ est le symbole de Kronecker}) \quad (13)$$

Généralisation

Loi de probabilité marginale des scores de trois joueurs

L'expression de cette loi se déduit de celle de la loi précédente, en employant un procédé analogue de séparation des variables aléatoires.

En considérant les joueurs C_1 , C_2 et C_3 , on forme le tableau suivant :

	C_1	C_2	C_3	C_4	C_5		C_{n+1}	Scores
C_1	0	X_{12}	X_{13}	X_{14}	X_{15}		$X_{1,n+1}$	X_1
C_2	X_{21}	0	X_{23}	X_{24}	X_{25}		$X_{2,n+1}$	X_2
C_3	X_{31}	X_{32}	0	X_{34}	X_{35}		$X_{3,n+1}$	X_3
C_4	X_{41}	X_{42}	X_{43}	0	X_{45}		$X_{4,n+1}$	X_4
.	.	.	.	.	.		.	.
.								.
C_{n+1}	$X_{n+1,1}$	$X_{n+1,2}$	.	.	.		0	X_{n+1}

On effectue la décomposition suivante des scores :

- Pour le joueur C_1 : d'une part l'élément X_{13} , d'autre part, l'ensemble $(X_{12}, X_{14}, X_{15}, \dots, X_{1,n+1})$
- Pour le joueur C_2 : d'une part, l'élément X_{23} , d'autre part, l'ensemble $(X_{21}, X_{24}, X_{25}, \dots, X_{2,n+1})$.
- Pour le joueur C_3 : l'ensemble (X_{31}, X_{32}) , puis l'ensemble $(X_{34}, X_{35}, \dots, X_{3,n+1})$.

On obtient la probabilité cherchée :

$$\begin{aligned}
 \Pr[X_1 = x_1, X_2 = x_2, X_3 = x_3] &= \\
 &= \sum_{u=-1}^{+1} \sum_{v=-1}^{+1} \left\{ \Pr[X_{13} = u] \cdot \Pr[X_{23} = v] \cdot \Pr[X_1 - X_{13} = x_1 - u, X_2 - X_{23} = x_2 - v] \cdot \Pr[X_3 - (X_{31} + X_{32}) = x_3 + u + v] \right\} \quad (14)
 \end{aligned}$$

Le calcul de la probabilité (14) nécessite la connaissance de l'expression du troisième terme entre accolades. Ce terme est analogue à celui donné par la relation (10), le quatrième terme étant analogue à celui donné par (7).

On voit donc que le calcul de la probabilité d'un vecteur score de $(n + 1)$ joueurs pourra se faire par "remontée", en partant de la loi de probabilité marginale du score d'un seul joueur (C_1), puis à l'aide de la loi de probabilité des scores de deux joueurs (C_1 et C_2), puis de trois joueurs (C_1 , C_2 et C_3), en utilisant chaque fois les résultats précédents, jusqu'à ce que l'on ait atteint l'ensemble des joueurs.

Voici à titre indicatif l'expression développée de la loi de probabilité conjointe des scores de quatre joueurs parmi un ensemble de $(n + 1)$ joueurs ($n \geq 3$).

$$\begin{aligned} \Pr[X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4] = \\ = \sum_{u=-1}^{+1} \sum_{v=-1}^{+1} \sum_{w=-1}^{+1} \left\{ \pi(u) \pi(v) \pi(w) \left[\sum_{k=|x_4+u+v+w|}^{n-3} C_{n-3}^k C_k^{\frac{k+|x_4+u+v+w|}{2}} p^k q^{n-3-k} \right] \right. \\ \cdot \left\{ \sum_{\alpha=-1}^{+1} \sum_{\beta=-1}^{+1} \left\{ \pi(\alpha) \pi(\beta) \left[\sum_{k=|x_3-w+\alpha+\beta|}^{n-3} C_{n-3}^k C_k^{\frac{k+|x_3-w+\alpha+\beta|}{2}} p^k q^{n-3-k} \right] \right\} \right. \\ \cdot \left\{ \sum_{\gamma=-1}^{+1} \left\{ \pi(\gamma) \left[\sum_{k=|x_1-u-\alpha-\gamma|}^{n-3} C_{n-3}^k C_k^{\frac{k+|x_1-u-\alpha-\gamma|}{2}} p^k q^{n-3-k} \right] \right\} \right. \\ \cdot \left. \left[\sum_{k=|x_2-v-\beta+\gamma|}^{n-3} C_{n-3}^k C_k^{\frac{k+|x_2-v-\beta+\gamma|}{2}} p^k q^{n-3-k} \right] \right\} \right\} \right\} . \end{aligned}$$

Les fonctions π ayant été définies plus haut, ainsi que le signe Σ' .

Cette expression n'est pas simple : dans la pratique, on aura intérêt à partir de la formule suivante :

$$\begin{aligned} \Pr[X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4] = \\ = \sum_{u=-1}^{+1} \sum_{v=-1}^{+1} \sum_{w=-1}^{+1} \Pr[X_{14} = u] \Pr[X_{24} = v] \Pr[X_{34} = w] . \\ \cdot \Pr[X_1 - X_{14} = x_1 - u, X_2 - X_{24} = x_2 - v, X_3 - X_{34} = x_3 - w] \\ \cdot \Pr[X_4 - (X_{41} + X_{42} + X_{43}) = x_4 + u + v + w] \end{aligned}$$

et à utiliser les résultats obtenus pour 3 joueurs, lesquels font eux-mêmes appel aux résultats obtenus pour 2 joueurs.

TABLEAU (A)

Vecteurs scores $X_1 \ X_2 \ X_3 \ X_4$	Nombre de vecteurs scores d'égale probabilité	Probabilité du vecteur
		score $X_1 \ X_2 \ X_3 \ X_4$
-3 -1 1 3	24	p^6
-3 0 0 3	12	$p^5 q$
-3 -1 2 2	12	$p^5 q$
-3 0 1 2	24	$p^5 q + p^4 q^2$
-3 1 1 1	4	$2p^6 + p^3 q^3$
-2 -2 1 3	12	$p^5 q$
-2 -2 2 2	6	$p^4 q^2$
-2 -1 0 3	24	$p^5 q + p^4 q^2$
-2 -1 1 2	24	$p^5 q + 2p^4 q^2 + p^3 q^3$
-2 0 0 2	12	$3p^4 q^2 + 2p^3 q^3$
-2 0 1 1	12	$2p^5 q + 2p^4 q^2 + 2p^3 q^3 + p^2 q^4$
-1 -1 -1 3	4	$2p^6 + p^3 q^3$
-1 -1 0 2	12	$2p^5 q + 2p^4 q^2 + 2p^3 q^3 + p^2 q^4$
-1 -1 1 1	6	$4p^6 + 4p^3 q^3 + 2p^2 q^4$
-1 0 0 1	12	$3p^5 q + 4p^4 q^2 + 2p^3 q^3 + 2p^2 q^4 + p q^5$
0 0 0 0	1	$6p^4 q^2 + 8p^3 q^3 + q^6$

Le tableau (A) donne, pour un ensemble de 4 joueurs, les probabilités du vecteur score (X_1, X_2, X_3, X_4) ainsi que le nombre de vecteurs scores équiprobables.

(Tout vecteur dont les "composantes" se déduiraient de celles d'un vecteur donné par une permutation de ces composantes a la même probabilité d'apparition que le vecteur initial).

En effet cela revient simplement à changer les numéros des joueurs. Il en résulte une importante économie de calcul.

Il faut remarquer que tout ensemble de nombres compris entre -3 et $+3$ et dont la somme est nulle ne forme pas obligatoirement un vecteur score.

Par exemple, il n'est pas possible d'avoir un vecteur tel que $(-3, -3, 3, 3)$, comme on pourra s'en rendre compte en écrivant le tableau des variables X_{ij} .

IV - APPROXIMATION DES LOIS DE PROBABILITE

Théoriquement, ce modèle probabiliste permet de connaître les probabilités des différents vecteurs scores pour un nombre quelconque de joueurs. Mais pratiquement, le calcul est beaucoup trop long.

On a donc cherché à approcher les lois de probabilité marginales des scores par une loi plus maniable.

D'après le théorème central limite, on peut approcher la distribution de $\frac{X_1}{\sqrt{2np}}$ par la distribution normale réduite.

Pour étudier la valeur de cette approximation pour un n donné on a tracé, (figures I), les histogrammes correspondant à la loi exacte et à la loi approchée, loi normale de moyenne nulle et de variance $2np$.

Les figures 1a et 1b représentent :

- en trait plein l'histogramme de la loi exacte
- en pointillé l'histogramme de la loi normale $N(0, 2np)$.

Sur la figure 1a ($n = 6, p = 1/3$), le raccordement est assez bon, avec des écarts inférieurs à 0,005.

Sur la figure 1b ($n = 10, p = 0,2$), ces écarts sont tous inférieurs au millièème.

On peut montrer aussi que la distribution conjointe des scores de k joueurs est proche de la distribution de k variables normales indépendantes si k est fixé et n tend vers l'infini (dans la pratique k doit être petit devant n). Mais ce résultat ne peut évidemment pas être généralisé au cas de la distribution des scores de k joueurs si k est grand et en particulier si $k = n$.

V - UTILISATION PRATIQUE

Dans la pratique l'étude précédente peut servir à réaliser un test de l'hypothèse nulle présentée plus haut contre l'alternative qu'un joueur donné est "plus fort" que les autres : dans ces conditions son score aura tendance à être plus grand que sous l'hypothèse nulle. La distribution

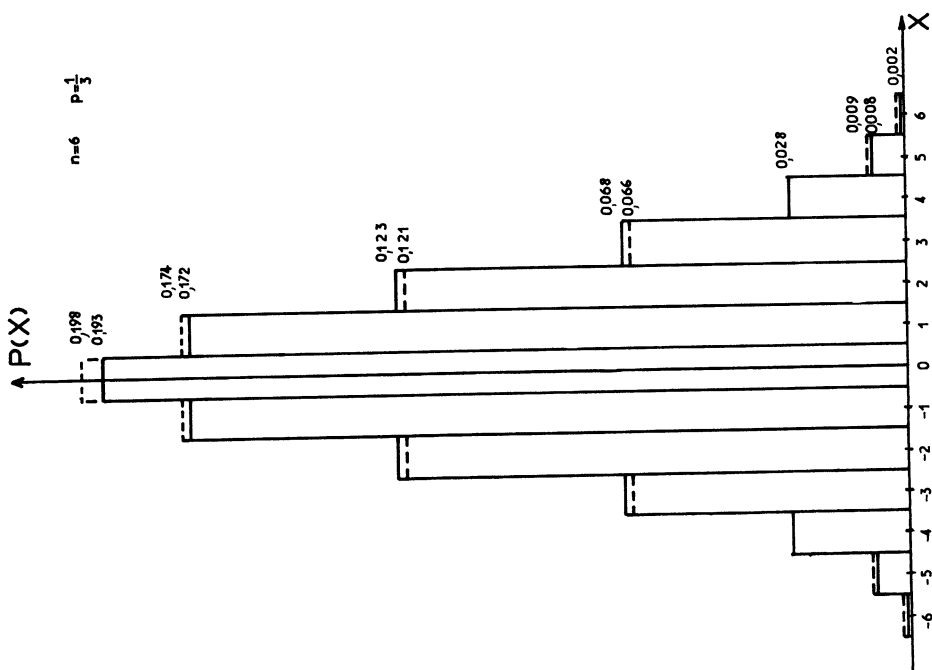


Figure 1a

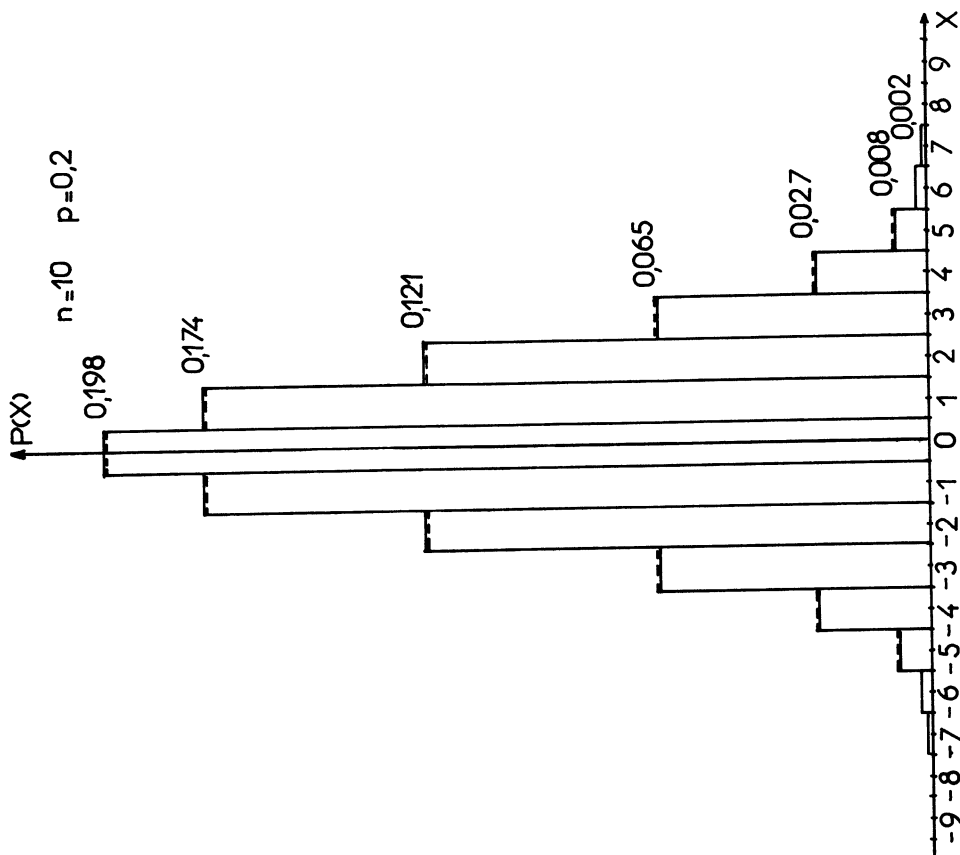


Figure 1b

des scores donnés ci-dessus permet de construire une région critique de seuil donné pour ce test, si la probabilité q de match nul est connue. Cependant, dans la pratique, il n'en sera pas ainsi, et l'on devra se contenter d'estimer cette probabilité à partir des données, la meilleure estimation de q étant : $\hat{q} = \frac{2r}{n(n+1)}$ où r est le nombre total de parties

nulles parmi les $\frac{n(n+1)}{2}$ parties jouées. On en déduit bien sûr l'estimation de p :

$$\hat{p} = \frac{1}{2} - \frac{r}{n(n+1)}$$

VI - CAS DE s COMPETITIONS SUCCESSIVES INDEPENDANTES

On considère maintenant le cas où une compétition du type décrit plus haut est répétée s fois, les s répétitions étant supposées stochastiquement indépendantes. Nous utiliserons les mêmes notations que précédemment, en ajoutant l'indice supérieur (k) pour la k ème compétition ($k = 1, 2, \dots, s$). Ainsi :

$$X^{(k)} = \begin{bmatrix} X_1^{(k)} \\ X_2^{(k)} \\ \cdot \\ \cdot \\ X_{n+1}^{(k)} \end{bmatrix} \quad \text{est le vecteur score attaché à la } k\text{ème compétition.}$$

Tous les vecteurs $X^{(k)}$ sont indépendants en probabilité et identiquement distribués.

Le score définitif S_i du joueur C_i est :

$$S_i = \sum_{k=1}^s X_i^{(k)} \quad (i = 1, 2, \dots, n+1)$$

On notera S le vecteur

$$\begin{bmatrix} S_1 \\ S_2 \\ \cdot \\ \cdot \\ S_{n+1} \end{bmatrix}$$

Nous nous bornerons ici à l'étude des résultats asymptotiques lorsque s devient très grand, résultats qui sont très simples et faciles à utiliser. Dans ce cas la distribution de :

$$\frac{S}{\sqrt{s}} = \sum_{k=1}^s \frac{X^{(k)}}{\sqrt{s}}$$

tend vers la loi normale de matrice des moyennes nulle et de matrice des variances et covariances Σ_x calculée plus haut.

Posons :

$$D = \frac{1}{\sqrt{2(n+1)p}} \frac{S}{\sqrt{s}}$$

La distribution asymptotique du vecteur D est normale de moyenne nulle de matrice des variances et covariances

$$\Sigma_D = \left\| \delta_{ij} - \frac{1}{n+1} \right\|$$

On vérifie facilement que Σ_D a une valeur propre nulle et n valeurs propres égales à un. Il existe donc une transformation orthogonale

$$Y = C D$$

telle que la matrice des variances de Y soit :

$$\Sigma_Y = C \Sigma_D C' = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ & & 1 & \\ & & & \ddots \\ & & & & 1 \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} \quad (\text{où } C' \text{ est la transposée de } C)$$

Dans ces conditions Y'Y suit asymptotiquement une loi de χ^2 à n degrés de liberté ; il en sera de même de :

$$\sum_{i=1}^{n+1} D_i^2 = \frac{1}{2(n+1)s p} \sum_{i=1}^{n+1} S_i^2$$

puisque $Y'Y = D'C'CD = D'D = \sum_{i=1}^{n+1} D_i^2$

Dans la pratique p est inconnu, mais si $\hat{p}$ est un estimateur convergent de p, la distribution asymptotique de

$$\sum_{i=1}^{n+1} \frac{S_i^2}{2(n+1)s \hat{p}} \text{ sera encore un } \chi^2 \text{ à n ddl.}$$

Il est facile de vérifier qu'un bon estimateur de p sera :

$$p = \frac{T}{s n(n+1)} \quad \text{où } T \text{ est le nombre total de parties non nulles dans les } s \frac{n(n+1)}{2} \text{ parties jouées.}$$

Un test de l'hypothèse nulle selon laquelle les joueurs sont de forces égales pourra être ainsi effectué à partir de :

$$\chi^2 = \frac{n}{2T} \sum_{i=1}^{n+1} S_i^2$$

la région critique étant formée des grandes valeurs du χ^2 .

VII - CONCLUSION

Certains auteurs, lorsqu'ils sont en présence d'expériences de comparaisons par paires où figurent des égalités, décident de classer arbitrairement un objet comme étant supérieur à l'autre pour pouvoir continuer à employer le modèle à deux catégories.

L'emploi d'un modèle à trois catégories pourrait alors se justifier dès que le nombre de cas où l'on n'a pas su discerner entre deux objets n'est pas petit devant le nombre total de comparaisons effectuées.

Nous avons présenté un tel modèle et avons donné quelques indications sur son étude. Ce modèle est plus lourd à manier que le modèle plus simple où l'équivalence n'est pas considérée, et nécessite en outre l'introduction d'un paramètre inconnu même sous l'hypothèse nulle.

Cependant, comme il a été vu dans la dernière partie, il est possible de trouver au moins des résultats asymptotiques simples.