MATHÉMATIQUES ET SCIENCES HUMAINES

FARID BENINEL
MICHEL GRUN-REHOMME

Un indice de concordance pour l'étude comparative d'opinions sur la conjoncture

Mathématiques et sciences humaines, tome 142 (1998), p. 17-25
<http://www.numdam.org/item?id=MSH_1998__142__17_0>

Math. Inf. Sci. hum., (36ᵉ année, n°142, *1998*, pp.17-25)

UN INDICE DE CONCORDANCE POUR L'ÉTUDE COMPARATIVE D'OPINIONS SUR LA CONJONCTURE

Farid BENINEL[1] Michel GRUN-REHOMME[2]

RÉSUMÉ. *Un modèle stochastique d'indices d'association sur données binaires d'un panel d'enquêtes conjoncturelles est proposé dans ce travail. Ce Modèle permet de quantifier la concordance d'opinions entre deux individus et de comparer cette concordance d'une paire d'individus à l'autre. Des propriétés des indices découlant de ce modèle sont également mises en évidence.*

SUMMARY. Concordance indices for the comparative study of opinions about economic climate.
In this paper, we propose a stochastic model of association indices based on binary data surveys. These indices calculate the opinion concordance between two individuals and allow to compare pairs. Some properties of such indices are derived.

1. INTRODUCTION

De nombreuses enquêtes conjoncturelles sont réalisées régulièrement auprès des entreprises, et principalement en France par l'INSEE et la Banque de France. L'analyse des réponses à ces enquêtes permet d'établir rapidement un diagnostic de l'économie à la fois sur le passé récent et le court terme. Les opinions émises par les chefs d'entreprise sur la production, les ventes, les prix, les effectifs, l'investissement..., contiennent de l'information macroéconomique sur le court terme. La plupart des questions sont qualitatives et concernent les évolutions de ces différentes grandeurs. En général, le répondant a le choix entre trois modalités : évolution à la hausse (amélioration, favorable, niveau supérieur à la normale), stabilité (normale) ou évolution à la baisse (diminution, défavorable, niveau inférieur à la normale). Ces réponses ne tiennent pas compte des effets saisonniers.

La première enquête conjoncturelle réalisée en france date de 1951 (1947 pour les Etats-Unis). Actuellement plus d'une dizaine d'enquêtes sont effectuées régulièrement dans différents secteurs d'activité (industrie, bâtiment, commerce, services). Les enquêtes sont réalisées sur un échantillon stratifié selon deux variables : le secteur d'activité et la taille (en termes d'effectifs ou de chiffre d'affaires) [5].

Pour les enquêtes d'opinion, les méthodes sont bien banalisées et utilisent à la fois l'approche multidimensionnelle et les séries temporelles. Les analyses sont toujours globales, que ce soit à travers les séries des soldes d'opinion (différence, exprimée en pourcentage, entre les réponses à la hausse et celles à la baisse) ou des comparaisons avec des séries

[1]Université de Poitiers, IUT, Dépt. STID, 8 rue Archimède, 79000 Niort
[2]Université Paris II, 92 rue d'Assas, 75006 Paris

d'indicateurs économiques (par régression linéaire). Ces réponses sont parfois pondérées par la taille de l'entreprise et une estimation de la précision est effectuée [3]. D'ailleurs les entreprises répondantes ne sont pas forcément les mêmes sur deux périodes consécutives. Les résultats sont agrégés par strates ou plus globalement. La confidentialité étant par ailleurs de rigueur.

La désaisonnalisation et la méthode de Box-Jenkins sont les principaux outils de l'approche temporelle. L'analyse des soldes d'opinion dans le temps nécessite de désaisonnaliser les séries [4]. L'utilisation des soldes d'opinion apporte des informations sur les tendances économiques à court terme et permet de faire des prévisions au niveau de la consommation ou de la production manufacturière.

Notre démarche est différente puisqu'elle envisage des comparaisons individuelles ou plus exactement la modélisation d'un indice de concordance d'opinion entre deux individus, deux dirigeants d'entreprise. On considère un panel d'entreprises de tailles et de secteurs d'activité différents. A intervalle de temps régulier (par exemple, chaque trimestre), les responsables de ces entreprises expriment leur opinion sur la conjoncture économique. Le choix est forcé entre un avis favorable (en hausse) ou défavorable (en baisse). Pour une entreprise donnée, le même responsable répond à l'enquête. La non-réponse à une date donnée, exclut le participant du panel. D'autres enquêtés peuvent intégrer le panel à d'autres dates.

Dans ce travail, on s'intéresse à l'élaboration et à l'étude d'un modèle stochastique d'indice d'association ayant pour objet de mesurer le niveau d'adéquation entre différents individus de cette population. Des études similaires ont déjà été réalisées en sciences du comportement [2]. On peut aussi citer les travaux de Snijders *et al.* (1990), qui portent sur l'étude des propriétés statistiques d'indices déjà existant dans la littératures, ceux dûs aux taxonomistes (Jaccard, Dice, Michener...). Une autre approche développée par Roberts & Evans (1993) considère l'association entre deux individus comme étant la probabilité d'apparition simultanée. Il en découle comme indices d'association les estimateurs de cette probabilité; plus précisément, ceux obtenus par la méthode du maximum de vraisemblance.

La construction des indices d'association présentés ici est nouvelle et possède les propriétés suivantes :

- Le calcul ne se fait que sur la période où les individus sont présents dans le panel.
- Dans la concordance d'opinion entre deux individus, il est tenu compte du poids de l'ensemble des individus du panel ayant le même avis. Il est clair que cette concordance entre deux individus est plus significative lorsqu'ils émettent le même avis, fortement minoritaire dans le panel.

2. LA METHODOLOGIE

Soient E l'ensemble des individus du panel où $N = card(E)$ et p le nombre de périodes de recueil des données durant l'étude. Les données se présentent donc sous la forme d'une matrice de terme général (X_{it}), d'ordre (N, p), où $X_{it} = 1$ (resp. 0) si l'individu ω_i du panel répond favorablement (resp. défavorablement) à la date t.

Précisons que cette matrice peut présenter des données manquantes du fait des entrées et des sorties dans le panel. Plus formellement, à tout individu $\omega_i \in E$, on associe le couple (X_i, π_i) avec $X_i \in \{0,1\}^{p_i}$ et π_i l'ensemble des dates de participation de l'individu ω_i à l'étude. Dans cette expression p_i désigne le nombre de périodes où ω_i est présent dans le panel. Si l'individu est présent pendant toute l'étude, $p_i = p$.

Il s'agit donc de construire un indice qui, pour toute paire d'individus, mesure le niveau de leur association. Notons M un tel indice; Il doit vérifier les axiomes de symétrie (a_1)

et de maximalité (a_2) ci-dessous:

$$\forall\ \omega_i, \omega_j \in E \qquad \begin{array}{ll} (a_1) & M(\omega_i, \omega_j) = M(\omega_j, \omega_i). \\ (a_2) & M(\omega_i, \omega_i) \geq M(\omega_i, \omega_j). \end{array}$$

De plus cet indice doit être d'autant plus élevé que deux individus expriment le même avis aux mêmes dates et que l'effectif des autres individus ayant le même avis est faible. Le modèle d'indice, préconisé dans cette étude, est celui évaluant la concordance d'opinions entre ω_i et ω_j par

$$M(\omega_i, \omega_j) = \sum_{t \in \pi_i \cap \pi_j} (\alpha_t X_{it} X_{jt} + \beta_t (1 - X_{it})(1 - X_{jt})),$$

où α_t (resp. β_t) est une pondération qui a pour objet de faire intervenir la taille F_t(resp. D_t) du groupe des individus ayant un avis favorable (resp. défavorable) à la date t. Plus précisément, α_t(resp. β_t) est une fonction décroissante de F_t(resp. D_t). Ainsi, α_t et β_t sont fonction décroissante l'une de l'autre.

Dans cette application , on choisit $\alpha_t = \dfrac{N_t + 1}{F_t + \frac{1}{2}}$ et de façon analogue $\beta_t = \dfrac{N_t + 1}{D_t + \frac{1}{2}}$ où N_t désigne l'effectif du panel à la date t.

Ainsi α_t et β_t sont liés par les relations $\alpha_t = \dfrac{\beta_t}{\beta_t - 1}$ et $\beta_t = \dfrac{\alpha_t}{\alpha_t - 1}$. On peut donc écrire l'association entre ω_i et $\omega_j$sous la forme

$$(1)\ \ M(\omega_i, \omega_j) = \sum_{t \in \pi_i \cap \pi_j} (\alpha_t X_{it} X_{jt} + \frac{\alpha_t}{\alpha_t - 1} (1 - X_{it})(1 - X_{jt})).$$

Remarque 1.
La recherche d'une relation symétrique entre les pondérations α et β (ie. $\alpha_t = T(\beta_t)$ et $\beta_t = T(\alpha_t)$ avec T décroissante) revient à choisir des fonctions idempotentes d'ordre 2. Différentes classes de telles fonctions sont envisageables. Dans notre application, on s'intéresse particulièrement à la classe des fonctions homographiques, forcément de la forme $T(x) = \dfrac{ax + b}{cx - a}$, avec $a^2 + bc > 0$. Les autres choix envisageables pour α_t et β_t via la relation $\beta_t = \dfrac{\alpha_t}{\alpha_t - 1}$ introduisent une dissymétrie dans la pondération des avis favorables et défavorables simultanés.

Remarque 2.
L'association $M(\omega_i, \omega_j)$ est positive et majorée par $q_{ij} \max_i (\alpha_i, \beta_i)$ où $q_{ij} = card(\pi_i \cap \pi_j)$.

3. ÉTUDE DE LA VRAISEMBLANCE DE LA CONCORDANCE

Deux paires d'individus peuvent avoir la même valeur observée pour l'indice d'association avec une longévité dans le panel très différente. D'où la nécessité d'adjoindre, à la valeur observée, sa vraisemblance. Cette approche est très usitée chez les concepteurs d'indices, on citera en particulier les travaux de I.C. Lerman [6].

Pour étudier la vraisemblance d'une association, on compare la valeur observée $M(\omega_i, \omega_j)$ à toutes les valeurs $M(Y_i, Y_j)$ où Y_i, Y_j constituent des profils de réponse, de même type, que ceux qu'auraient pu avoir les individus ω_i et ω_j pendant la période commune de participation au panel. Il convient donc, de définir des profils de même type ou plus exactementt les paires de profils de même type.

Notons G_p l'ensemble des permutations de $\{1, 2, ..., p\}$ et considérons A_{ij} la partie de $G_p \times G_p$ définie par

$$A_{ij} = \left\{ \left(\Gamma, \Gamma'\right) \in G_p \times G_p :\ (\Gamma X_i)_t = X_{it}\ et\ \left(\Gamma' X_j\right)_t = X_{jt},\ \forall t \notin \pi_i \cap \pi_j \right\}.$$

Les paires (X_i, X_j) et (Y_i, Y_j) seront dites du même type s'il existe $\left(\Gamma, \Gamma'\right) \in A_{ij}$ tel que $(Y_i, Y_j) = \left(\Gamma X_i, \Gamma' X_j\right)$. Ainsi donc, les paires de même type de (X_i, X_j) sont celles composées d'un profil à même potentiel de présence que X_i sur $\pi_i \cap \pi_j$ et d'un profil à même potentiel de présence que X_j sur le même intervalle.

Remarquons que le type ainsi défini permet d'établir une relation d'équivalence sur l'ensemble $\{0,1\}^{2p}$.

On notera $A(X_i, X_j)$ la classe des paires de même type que (X_i, X_j). Ce type de classe peut être caractérisé par les fréquences d'avis favorables de chacun des individus ω_i et ω_j sur la période $\pi_i \cap \pi_j$.

3.1. La variable aléatoire *Indice de concordance d'opinions*

Dans le but de mesurer la vraisemblance de la concordance observée $M(\omega_i, \omega_j)$, on s'intéresse à la variable aléatoire M définie sur $A(X_i, X_j)$ par

$$\forall \ (Y_i, Y_j) \in A(X_i, X_j) \qquad M(Y_i, Y_j) = \sum_{t \in \pi_i \cap \pi_j} (\alpha_t Y_{it} Y_{jt} + \frac{\alpha_t}{\alpha_t - 1}(1 - Y_{it})(1 - Y_{jt})).$$

Ainsi, $M(\omega_i, \omega_j)$ constitue la valeur observée de cette statistique; Ou encore ,

$$M(\omega_i, \omega_j) = M(X_i, X_j).$$

Considérons l'application f_{X_i, X_j}définie sur A_{ij} et à valeurs dans $A(X_i, X_j)$par

$$f_{X_i, X_j}\left(\Gamma, \Gamma'\right) = \left(\Gamma X_i, \Gamma' X_j\right).$$

Si l'on munit A_{ij} de la mesure uniforme , $A(X_i, X_j)$ de la mesure image et si l'on note Φ_M et $\Phi_{M \circ f_{X_i, X_j}}$ les fonctions de répartition des variables aléatoires M et $M \circ f_{X_i, X_j}$, on a $\Phi_M = \Phi_{M \circ f_{X_i, X_j}}$.

Etant donnée cette relation et pour des commodités de présentation, l'étude de l'indice de concordance se fera donc en considérant comme univers l'ensemble A_{ij}.

3.2. Les variables aléatoires *Nombres de concordances*

N_F désigne la variable aléatoire qui, à tout couple de permutations $\left(\Gamma, \Gamma'\right) \in A_{ij}$, associe le nombre d'avis favorables simultanés exprimés par ω_i et ω_j sur la période commune de présence dans le panel. Elle est définie par

$$N_F\left(\Gamma, \Gamma'\right) = \sum_{\pi_i \cap \pi_j} (\Gamma X_i)_t \left(\Gamma' X_j\right)_t.$$

De façon analogue, on définit N_D la variable aléatoire qui associe le nombre d'avis défavorables simultanés. n_{ij} (resp.n_{ji}) correspond au nombre d'avis favorables, exprimés par ω_i (resp. ω_j) sur la période $\pi_i \cap \pi_j$. Il est aisé de vérifier la relation

$$(2) \ \forall \ \left(\Gamma, \Gamma'\right) \in A_{ij}, \ N_F(\Gamma, \Gamma') = q_{ij} - (n_{ij} + n_{ji} - N_D(\Gamma, \Gamma')).$$

Le domaine de réalisation de N_F est donné par l'ensemble $\{m, ..., M\}$, où

$$m = \max(0, n_{ij} + n_{ji} - q_{ij}) \text{ et } M = \min(n_{ij}, n_{ji}).$$

Opérer une permutation des avis de l'un et l'autre des deux individus, revient au même que de laisser fixe les avis exprimés par l'un et appliquer une certaine permutation sur les avis exprimés par l'autre. Ce résultat est formalisé par le théorème ci-après.

Théorème 1. *(Théorème de dualité)*

$$\forall\ \left(\Gamma,\Gamma'\right)\in A_{ij},\ N_F(\Gamma,\Gamma')=N_F(Id_p,{}^t\Gamma\Gamma')=N_F({}^t\Gamma\Gamma',Id_p).$$

Preuve
Si $t\notin\pi_i\cap\pi_j$ alors $t\notin\pi_i$ ou $t\notin\pi_j$ et par suite, on a

$$\sum_{t\notin\pi_i\cap\pi_j}(\Gamma X_i)_t\left(\Gamma' X_j\right)_t=0;$$

Ce qui permet d'obtenir

$$N_F(\Gamma,\Gamma')=\sum_t(\Gamma X_i)_t\left(\Gamma' X_j\right)_t=\left\langle\Gamma X_i,\Gamma' X_j\right\rangle.$$

On peut donc écrire

$$N_F(\Gamma,\Gamma')=\left\langle X_i,{}^t\Gamma\Gamma' X_j\right\rangle=\left\langle{}^t\Gamma'\Gamma X_i,X_j\right\rangle.\ \blacksquare$$

Notons $A_{i/j}=\{\Gamma\in G_p\ :\ (\Gamma X_i)_t=X_{it},\ \forall\ t\notin\pi_i\cap\pi_j\}$. Il s'agit de l'ensemble des permutations laissant invariant les positions de l'individu ω_i sur les périodes exterieurs à l'intervalle $\pi_i\cap\pi_j$.Remarquons que $A_{ij}=A_{i/j}\times A_{j/i}$.

Lemme 1. *Soit k une réalisation possible de N_F; On a*

$$Card\left\{\left(\Gamma,\Gamma'\right)\in A_{ij}\ :\ N_F(\Gamma,\Gamma')=k\right\}=q_{ij}!Card\left\{\Gamma\in A_{j/i}\ :\ N_F(Id,\Gamma)=k\right\}$$

Preuve : Pour tout Γ_0 de $A_{i/j}$
$Card\left\{\Gamma\in A_{j/i}\ :\ N_F(Id,\Gamma)=k\right\}=Card\left\{\Gamma\in A_{j/i}\ :\ N_F(\Gamma_0,\Gamma)=k\right\}$
Il s'ensuit que :
$\sum_{\Gamma_0}Card\left\{\Gamma\in A_{j/i}\ :\ N_F(Id,\Gamma)=k\right\}=\sum_{\Gamma_0}Card\left\{\Gamma\in A_{j/i}\ :\ N_F(\Gamma_0,\Gamma)=k\right\}$
Le terme argument de la somme de gauche, étant indépendant de Γ_0, on a
$CardA_{j/i}\times Card\left\{\Gamma\in A_{j/i}\ :\ N_F(Id,\Gamma)=k\right\}$
ou encore,
$q_{ij}!Card\left\{\Gamma\in A_{j/i}\ :N_F(Id,\Gamma)=k\right\}$
Ceci établit le résultat. $\blacksquare$

Lemme 2. *La distribution de probabilité de la variable N_F est l'hypergéometrique $H(q_{ij},n_{ij},n_{ji})$*

Preuve:
Soit k une réalisation possible de N_F,
$P(N_F=k)=\frac{1}{(q_{ij}!)^2}Card\left\{\left(\Gamma,\Gamma'\right)\in A_{ij}\ :\ N_F(\Gamma,\Gamma')=k\right\}$
En vertu du lemme 1, l'on établit :
$P(N_F=k)=\frac{1}{q_{ij}!}Card\left\{\Gamma\in A_{j/i}\ :\ N_F(Id,\Gamma)=k\right\}$
Ce qui donne,

$$P(N_F=k)=\frac{1}{q_{ij}!}\sum_{\Gamma\in A_{j/i}}\mathbb{I}_{\{k\}}\left(\sum_{\pi_i\cap\pi_j}X_{it}.\,(\Gamma X_j)_t\right)$$

Ici $\mathbb{1}_{\{k\}}$ désigne la fonction indicatrice de k.

Posons pour simplifier $r = n_{ij} - k$ et notons P_k l'ensemble des parties à k éléments de $\pi_i \cap \pi_j$. Pour $\tau \in P_k$ on notera $P_{r,\tau}$ l'ensemble des parties à r éléments pris dans le complémentaire de τ dans $\pi_i \cap \pi_j$.

On peut écrire :

$$P(N_F = k) = \frac{1}{q_{ij}!} \sum_{\Gamma} \sum_{\tau_1 \in P_k} \mathbb{1}_{\{k\}} \left(\sum_{\tau_1} X_{it} \right) \mathbb{1}_{\{r\}} \left(\sum_{\tau_2 \in P_{r,\tau}} X_{it} \right) \mathbb{1}_{\{k\}} \left(\sum_{\tau_1} (\Gamma X_i)_t \right) \mathbb{1}_{\{0\}} \left(\sum_{\tau_2} (\Gamma X_j)_t \right)$$

En intervertissant les sommes, l'on obtient :

$$(3)\ P(N_F = k) = \frac{1}{q_{ij}!} \sum_{\tau_1} \mathbb{1}_{\{k\}} \left(\sum_{\tau_1} X_{it} \right) \mathbb{1}_{\{r\}} \left(\sum_{\tau_2} X_{it} \right) \sum_{\Gamma} \mathbb{1}_{\{k\}} \left(\sum_{\tau_1} (\Gamma X_j)_t \right) \mathbb{1}_{\{0\}} \left(\sum_{\tau_2} (\Gamma X_j)_t \right)$$

On montre aisément les relations qui suivent:

$$\sum_{\tau_1} \mathbb{1}_{\{k\}} \left(\sum_{\tau_1} X_{it} \right) \mathbb{1}_{\{r\}} \left(\sum_{\tau_2} X_{it} \right) = C^k_{n_{ij}}$$

$$\sum_{\Gamma} \mathbb{1}_{\{k\}} \left(\sum_{\tau_1} (\Gamma X_j)_t \right) \mathbb{1}_{\{0\}} \left(\sum_{\tau_2} (\Gamma X_j)_t \right) = C^k_{n_{ji}} \times C^r_{q_{ij}-n_{ji}} \times (k!)\,(r!)\,(q_{ij} - k - r)!$$

Par substitution dans l'expression (2), on obtient

$$P(N_F = k) = \frac{1}{q_{ij}!} C^k_{n_{ji}} \times C^r_{q_{ij}-n_{ji}} \times (k!)\,(r!)\,(q_{ij} - k - r)!$$

et par suite,

$$P(N_F = k) = \frac{C^k_{n_{ji}} \times C^{n_{ij}-k}_{q_{ij}-n_{ji}}}{C^{n_{ij}}_{q_{ij}}}$$

Ce qui établit le résultat. ■

3.3. Distribution de l'indice de concordance

Paramètres *Min et Max*. On donne dans ce paragraphe l'expression, pour un couple d'individus quelconque, de la valeur du minimum et du maximum de l'indice, relativement aux couples de permutations de l'ensemble A_{ij} défini au paragraphe 3.

Etant donné un couple d'individus $(\omega_i, \omega_j) \in E^2$, on note $\alpha_{(1)}, \alpha_{(2)}, ..., \alpha_{(q)}$ le rangement par ordre croissant des poids $\alpha_1, \alpha_2, ..., \alpha_q$ correspondants aux occurrences $(1, 1)$ éventuelles sur les dates de $\pi_i \cap \pi_j$. L'ordre induit sur les valeurs β_t, via la relation $\beta_t = \dfrac{\alpha_t}{\alpha_t - 1}$, à savoir $\beta_{(1)}, \beta_{(2)}, ..., \beta_{(q)}$ est un ordre décroissant. D'où les expressions,

$$(4)\ Max = \max_{(\Gamma,\Gamma') \in A_{ij}} M\left(\Gamma X_i, \Gamma' X_j\right) = \sum_{i=1}^{M} \alpha_{(q-i+1)} + \sum_{i=1}^{L} \beta_{(i)}$$

$$(5)\ Min = \min_{(\Gamma,\Gamma') \in A_{ij}} M\left(\Gamma X_i, \Gamma' X_j\right) = \sum_{i=1}^{m} \alpha_{(i)} + \sum_{i=1}^{l} \beta_{(q-i+1)}$$

où M (resp. m) désigne le nombre maximum (resp. minimum) d'avis favorables simultanés et L (resp. l) le nombre maximum (resp. minimum) d'avis défavorables simultanés. Ainsi en divisant l'indice de concordance, donné en (1), par le maximum donné par la relation (4), on obtient un indice standardisé.

La fonction de répartition. Ce paragraphe a pour objectif d'expliciter la fonction de répartition Φ de l'indice de concordance.

Soit $x \in I\!R$;

$$\Phi(x) = \sum_k \sum_l P\left((N_F, N_D) = (k,l)\right) \times P\left((M \leq x \mid (N_F, N_D) = (k,l)\right)).$$

Or, pour k fixé, on a

$$(6)\ P\left((N_F, N_D) = (k,l)\right) = P(N_F = k).$$

Pour $k \in [m, M]$ et $l = q_{ij} - (n_{ij} + n_{ji} - k)$, on obtient

$$(7)\ P((M \leq x \ / \ (N_F, N_D) = (k,l)) = \sum_{\tau_1 \in P_k} \sum_{\tau_3 \in P_{l,\tau_1}} \frac{1\!\!1_{[0,x]}\left(\sum\limits_{t \in \tau_1} \alpha_t + \sum\limits_{t \in \tau_3} \frac{\alpha_t}{\alpha_t - 1}\right)}{C^k_{q_{ji}} \times C^l_{q_{ji}-k}}.$$

En utilisant les relations (6) et (7), on obtient

$$\Phi(x) = \sum_{k=m}^{M} P(N_F = k) \sum_{\tau_1 \in P_k} \sum_{\tau_3 \in P_{l,\tau_1}} \frac{1\!\!1_{[0,x]}\left(\sum\limits_{t \in \tau_1} \alpha_t + \sum\limits_{t \in \tau_3} \frac{\alpha_t}{\alpha_t - 1}\right)}{C^k_{q_{ji}} \times C^l_{q_{ji}-k}}$$

En remplaçant $P(N_F = k)$ par son expression donnée dans le lemme 2, l'on obtient

$$\Phi(x) = \sum_{k=m}^{M} \frac{C^k_{n_{ji}} \times C^{n_{ij}-k}_{q_{ij}-n_{ji}}}{C^{n_{ij}}_{q_{ij}}} \sum_{\tau_1 \in P_k} \sum_{\tau_3 \in P_{l,\tau_1}} \frac{1\!\!1_{[0,x]}\left(\sum\limits_{t \in \tau_1} \alpha_t + \sum\limits_{t \in \tau_3} \frac{\alpha_t}{\alpha_t - 1}\right)}{C^k_{q_{ji}} \times C^l_{q_{ji}-k}}$$

Un algorithme, extension de celui développé en [1], permet le calcul des valeurs de $\Phi(x)$ pour des séquences courtes (ie. $q_{ij} \leq 40$).

4. EXEMPLE

On considère à titre illustratif les données mensuelles suivantes :

individu/date	1	2	3	4	5	6	7	8	9
ω_1	0	0	0	1	1	1	1	0	0
ω_2	1	0	0	1	1	1	1	0	0
ω_3	.	.	.	1	1	1	1	1	0
ω_4	1	0	1	0	1	1	0	1	.
ω_5	0	0	0	1	0	0	1	0	0

Taille du panel	20	20	18	21	20	19	19	20	22
Favorable (N_F)	11	8	10	14	15	17	15	10	8

On calcule la valeur observée de l'indice de concordance de tout couple pris parmi ces 5 individus du panel, ainsi que la vraisemblance associée, donnée par la fonction de répartition.

Le tableau ci-dessous, qui prend en compte la période commune entre deux individus donnés, présente ces résultats :

Paire	$M(\omega_i, \omega_j)$	$M(\omega_i, \omega_j)/Max$	$\Phi(M)$
$(1,2)$	12.799	0.454	0.82
$(1,3)$	6.891	0.442	0.77
$(1,4)$	4.171	0.171	0.38
$(1,5)$	12.512	0.470	0.64
$(2,3)$	6.891	0.443	0.77
$(2,4)$	5.997	0.232	0.29
$(2,5)$	10.301	0.423	0.61
$(3,4)$	4.498	0.923	0.95
$(3,5)$	4.393	0.346	0.47
$(4,5)$	1.673	0.081	0.10

La valeur de l'indice dépend bien sûr de la comparaison des opinions des deux individus, de la taille du panel et du nombre d'avis favorables, mais aussi de la longueur de la période commune. On peut avoir des indices proches (paires $(2,3)$ et $(2,4)$) et des vraisemblances différentes ou des indices différents (paires$(2,3)$ et $(1,2)$) avec des fonctions de répartition proches. L'observation des résultats montre l'importance de l'usage de la fonction de répartition. De plus, l'ordre induit sur les couples par la valeur observée de l'indice n'est pas forcément le même que celui induit par le niveau de significativité associé ($(2,5)$ et $(3,4)$). Par contre les paires $(1,5)$ et $(2,5)$ présentes sur toutes le périodes ont des indices voisins et des vraisemblances pratiquement identiques.

Pour une période commune écourtée, le calcul de l'indice correspond à un traitement de la non réponse avec des avis contraires : cet indice est le même que celui calculé sur la période totale, en imputant les données par des valeurs d'avis contraires. Mais la valeur de la fonction de répartition est la même pour des avis totalement concordants ou discordants quelque soit la longueur de la période commune.

Comme on peut le constater sur les données brutes, l'indice confère un niveau de significativité presque nul pour la paire en désaccord presque maximal $(4,5)$.

5. CONCLUSION

Cette démarche a l'avantage de proposer un modèle d'indice de concordance facile calculer et dont on connaît la distribution exacte, ce qui permet l'étude sur des périodes courtes. Les résultats obtenus diffèrent de ceux fournis par l'approximation gaussienne habituellement utilisée dans ce type de problème.

Il resterait à expliciter la loi limite de l'indice considéré, d'étendre le modèle au cas de plus de deux réponses possible et au cas à n voies.

Remerciements

Les auteurs remercient Bruno Leclerc, Georges Le Calvé et les referees anonymes pour leurs suggestions et conseils. Ils tiennent également à remercier Mohammed Benayade pour l'aide informatique.

BIBLIOGRAPHIE

[1] Beninel F., Husson F., 1998. Calculation for small sequences of the cumulative distribution function for particular family of discretes statistics. Computational statistics, (à paraître).

[2] Bretagnolle V., Beninel F., 1994. Un index d'association basé sur des variables de présence-absences: Application à l'étude des structures sociales. Proceedings of the international Meeting"Ecology and statistical methods", Niort.

[3] Caron N., Ravalet P., Sautory O., 1996. Estimation de la précision d'un solde d'opinion dans les enquêtes de conjoncture auprès des entreprises. INSEE, Coll. méthodes statistiques, n^o 9602.

[4] Grun-Réhomme M., Ladiray D., 1994. Moyennes mobiles centrées et non-centrées: construction et comparaison. Revue de Statistiques Appliquées (RSA), vol. XLII, $n^o 3, 1994$.

[5] INSEE, 1992. La rénovation des enquêtes de conjoncture, coll. Méthodes, n^o 32.

[6] Lerman I.C., 1981. Etude de la notion de ressemblance. Classification et analyse ordinale des données, Dunod.

[7] Snijders T.A.B., and al., 1990. Distribution of some similarity coefficients for dyadic binary data in the case of associated attibutes, J. of Classification, $17, 1$, $p.5 - 31$.

[8] Roberts G., Evans P.R., 1993. A method for the detection of non random associations, Behavioral Ecology and Sociobiology, $15, p.349 - 354$.