CONTINGENCE (D2, D3, K1)
(26 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **contingence** est une notion de base dans l'analyse d'un **tableau de contingence**.

(i) Soit (Ω, $\mathcal{T}$, P) un **espace probabilisé**, A' = $(A_i')_{i \in I}$ et A" = $(A_j")_{j \in J}$ deux **suites** finies d'**événements** définis sur Ω.

Le triplet (A' x A", P $(A_i')_{i \in I}$, P $(A_i")_{j \in J}$) associe de façon probabiliste les évènements de chaque type A_j' et $A_j"$ (cf aussi **structure produit**).

On appelle :

(a) **contingence (théorique)** entre A' et A" le nombre :

(1) $\kappa^2 (A', A") = \Sigma_{i \in I} \Sigma_{j \in J} \{P(A_i') P(A_j")\}^{-1} . \{P (A_i' \cap A_j") - P(A_i') P(A_j")\}$

(ou aussi sa racine carrée), généralement appelé « chi-deux ».

Ce nombre est une somme de Card I x Card J **quotients**. Chaque « case » (i, j) associée contient le **rapport** entre :

$(a)_1$ l'**écart** P $(A_i' \cap A_j")$ - $P(A_i') P(A_j")$, qui exprime une « distance probabiliste » pr à l'**indépendance** ;

$(a)_2$ l'expression probabiliste $P(A_i') P(A_j")$ de indépendance elle-même.

Par suite, si les **variables qualitatives** $\eta' : \Omega \mapsto A'$ et $\eta" : \Omega \mapsto A"$ définissant le tableau considéré sont indépendantes, alors κ^2 (A', A") = 0. Il est équivalent de dire que, si A' et A" sont indépendantes (en probabilité), alors κ^2 (A', A") = 0.

(b) **matrice de contingence (théorique)** associée au scalaire κ^2 (A', A") la matrice (certaine) d'élément général (cf aussi **loi multivariée**) :

(2) $\kappa^2_{ij} (A', A") = \{P(A_i') P(A_j")\}^{-1} . \{P (A_i' \cap A_j") - P(A_i') P(A_j")\}, \quad \forall (i, j) \in I \times J.$

(i) Soit N = $(n_{ij})_{(i,j) \in I \times J}$ un tableau constitué d'entiers $n_{ij} \in \mathbf{N}$, et dont au moins une case (i, j) comporte un **effectif** (**fréquence absolue**) non nul(le). Ce tableau peut provenir de dénombrements divers.

Pour tout (i, j) $\in$ I x J, on appelle :

(a) **contingence** **(empirique)** associée à la case (i, j) de N la **statistique** :

(3) $c_{ij} = n_{ij} - (n_{..})^{-1} (n_{i.} \ n_{.j})$;

(b) **matrice de contingence (empirique)** associée à N la matrice (aléatoire) :

(4) $C = (c_{ij})_{(i,\, j) \in I \times J} = N - (e_m' \, N \, e_n)^{-1} \, (N \, e_n \;\; e_m' \, N)$,

avec m = Card I et n = Card J.

La matrice des **fréquences relatives** :

(5) $F = (f_{ij})_{(i,\, j) \in I \times J} = (e_m' \, N \, e_n)^{-1} \, N$, avec $f_{ij} = n_{ij} / n_{..}$,

conduit aux écritures équivalentes suivantes :

(6)
$$c_{ij} = n_{..} \, (f_{ij} - f_{i.} \;\; f_{.j}),$$
$$C = (e_m' \, N \, e_n) \, (F - F \, e_n \, e_m' \, F).$$

La **statistique du chi-deux** (à deux critères) X^2 (C, F), aussi notée X^2, d'une matrice de contingence est définie par la formule :

(7) $X^2 = \Sigma_{i \in I} \, \Sigma_{j \in J} \, (n_{ij} - c_{ij})^{-1} \, c_{ij}^2$.

On appelle alors **coefficient de contingence** (empirique) associé à N toute **statistique** dépendant de X^2, eg :

(a) le **coefficient de K. PEARSON** :

(8) $C_P = \{X^2 / (n_{..} + X^2)\}^{1/2}$;

(b) le **coefficient de A.A. CHUPROV** :

(9) $C_T = \{X^2 / (n_{..} \, (m - 1)^{1/2} \, (n - 1)^{1/2})\}^{1/2}$.

Ces coefficients mesurent le degré de la liaison (ou du **lien**) entre les deux **critères** (ou **caractères**) qui constituent le tableau N.

(iii) Les notions précédentes s'étendent directement à des tableaux de contingence à plus de deux critères. Ainsi, dans le cas de trois critères croisés, on transpose (1) selon :

(1)' $c_{ijk} = (n_{ijk} - n_{i..} \, n_{.j.} \, n_{..k}) / n_{...}$.