INTERDÉPENDANCE (J1)
(03 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) De façon générale, une **famille** de **variables aléatoires** est une **famille dépendante** lorsque, à la fois (cf aussi **dépendance**, **dépendance stochastique**) :

(a) la **loi conjointe** de ces **va** n'est pas égale au produit des lois marginales (lois propres des variables considérées) (absence d'**indépendance stochastique**) ;

(b) cette même loi jointe est « définie » : ie aucune variable n'est dégénérée, ie aucune **loi conditionnelle** ni aucune **loi marginale** n'est une **loi de DIRAC**.

De façon plus spécifique, la notion d'**interdépendance** est souvent associée à une distinction, au sein d'une telle famille, entre deux catégories de variables : les **variables endogènes** et les **variables exogènes**.

(ii) Lorsque des **variables statistiques** sont reliées entre elles par une **relation fonctionnelle**, on dit que ce sont des **variables interdépendantes**. Si l'on distingue, parmi ces variables, des variables endogènes et des variables exogènes, on définit une **structure d'interdépendance** à partir de la notion de **fonction d'interdépendance** : cette fonction est au **modèle d'interdépendance** ce qu'est la **fonction de régression** à une **régression** (ou à un **modèle de régression**).

(iii) Soit $(\Omega, \mathcal{T}, \mathcal{P})$ un **modèle statistique**, $(\mathcal{X}, \mathcal{B})$ un **espace d'observation** de variables exogènes, $(\mathcal{Y}, \mathcal{C})$ un espace d'observation de variables endogènes et $(\xi, \eta) : \Omega \mapsto \mathcal{X} \times \mathcal{Y}$ un **couple aléatoire** dans lequel ξ désigne la « liste » des variables exogènes et η celle des variables endogènes. Le plus souvent, ces listes sont finies et l'on note $\xi = (\xi_1 ,..., \xi_K)$ et $\eta = (\eta_1 ,..., \eta_G)$.

Si la notion d'**espérance mathématique** (ou une notion quelconque de **centralité**) est définie, on appelle **fonction d'interdépendance** du modèle précédent toute fonction $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathcal{Y}$ tq :

(1) $(\xi, \eta) \mapsto f(\xi, \eta) = \varepsilon$,

où ε est une **va** à valeurs dans $\mathcal{Y}$, appelée **perturbation aléatoire du modèle**, ou encore **erreur sur l'équation** $f(\xi, \eta) = 0$.

On suppose notamment que $E\, f(\xi, \eta) = E\, \varepsilon = 0$, l'espérance E étant calculée à l'aide d'une **mesure de probabilité** quelconque $P \in \mathcal{P}$, ie $E\, f(\xi, \eta) = \int_\Omega f\{\xi(\omega), \eta(\omega)\}\, dP(\omega)$.

Alternativement, si c désigne un calcul de **caractéristique légale** définissant une notion de **centralité**, on suppose que : $c \circ f(\xi, \eta) = c(\varepsilon) = 0$.

La forme (1) est une **forme implicite** : l'**homme de l'art** fournit au **statisticien** une équation implicite f $(\xi, \eta) = 0$ exprimant une liaison entre η et ξ. Cette liaison peut s'écrire :

$$(2) \quad \begin{array}{l} f_1(\xi_1, \ldots, \xi_K, \eta_1, \ldots, \eta_G) = 0, \\ \ldots \\ f_G(\xi_1, \ldots, \xi_K, \eta_1, \ldots, \eta_G) = 0, \end{array}$$

ie sous une forme qui indique comment les variables η_g dépendent des variables ξ_k (avec g = 1 ,..., G et k = 1 ,..., K). Dans certains cas, l'homme de l'art spécifie les fonctions f_g qui interviennent dans les équations (2).

(iv) L'interprétation et le traitement statistique du « modèle » (2) sont fondés sur le **modèle image** $(\mathscr{X} \times \mathscr{Y}, \mathscr{B} \otimes \mathscr{C}, \mathscr{P}^{(\xi, \eta)})$, ie sur l'image par (ξ, η) du modèle initial. Une **loi** quelconque $P^{(\xi, \eta)} \in \mathscr{P}^{(\xi, \eta)}$ est tq les variables ξ et η sont « contraintes » sous la forme (en général conditionnelle), exprimée dans l'**espace des variables** :

$$(3) \qquad E\, f(\xi, \eta) = 0 \qquad\qquad (\text{resp } C\, f(\xi, \eta) = 0),$$

ce qui signifie que, en « moyenne » (ou en valeur centrale), la relation définie en (1) est nulle. Par suite, la va $\varepsilon = f(\xi, \eta)$, dont la loi est $P^{\varepsilon} \in \mathscr{P}^{f(\xi,\eta)}$, vérifie :

$$(4) \qquad E\, \varepsilon = 0 \qquad\qquad (\text{resp } C\, \varepsilon = 0).$$

(v) En particulier, si G = 1 et si f est partiellement inversible pr à η_1 en sorte que (cf **théorème de la fonction implicite**) :

$$(5) \qquad \eta_1 = f^*(\xi, \varepsilon),$$

on fait souvent l'hypothèse que f * est une **fonction de régression**, ie que :

$$(6) \qquad \eta_1 = f^{**}(\xi) + \varepsilon, \qquad \text{avec } E\, \varepsilon = 0.$$

(vi) En pratique, l'équation (1) est « **observable** », au sens où il existe des **observations** $(X_n, Y_n)_{n=1,\ldots,N}$ du **couple aléatoire** (ξ, η) tq (dans l'**espace des observations**) :

$$(7) \qquad f(X_n, Y_n) = U_n, \qquad \forall\, n \in N_N^*.$$

Dans cette expression, les U_n sont des **copies** (inobservables) de la variable ε. C'est à partir de la forme (7), dite **forme structurelle**, et d'**hypothèses stochastiques** (eg $E\, U_n = 0$ et $V\, U_n = \Sigma$, n = 1 ,..., N) que le modèle d'interdépendance ainsi défini peut être estimé ou testé. Ce modèle est souvent (approximé par) un **modèle linéaire** (ie tq $f(\xi, \eta) = A'\, \xi + B'\, \eta$).