IDENTITÉ (C2, J1)
(03 / 11 / 2019, © Monfort, Dicostat2005, 2005-2019)

Outre son sens mathématique usuel (ie **application identique**), la notion statistique d'identité se rencontre souvent dans l'étude du **modèle d'interdépendance**.

(i) On appelle alors **identité** toute équation reliant les variables (**endogènes** ou **exogènes**, du modèle et tq :

(a) l'équation ne contient pas de **paramètres**, ou bien ceux-ci sont connus ;

(b) l'équation ne contient pas de **perturbation aléatoire**, ou bien celle-ci admet pour **lp** une **loi de DIRAC** placée à l'origine δ_0 (ie est presque sûrement nulle).

(ii) Ainsi, lorsqu'un **modèle d'interdépendance linéaire** est spécifié dans l'**espace des variables** :

(1) $B\ \eta + C\ \xi\ =\ \varepsilon$,

l'équation d'indice g $\in$ N^*_G s'écrit :

(2) $B_g\ \eta + C_g\ \xi\ =\ \varepsilon_g$,

où B_g est la g-ième ligne de B et C_g la g-ième ligne de C.

On dit que l'équation g est une **identité** ssi :

(3) $B_g\ \eta + C_g\ \xi\ =\ 0$ (ou alors $\varepsilon_g = 0$, P-p.s.).

Souvent, dans ce cas, (B_g , C_g) est connu. Le modèle (1) est alors un modèle avec **contrainte sur les observations**, ou un modèle dont certains paramètres sont fixés (cf **contrainte sur les paramètres**).

(iii) On distingue deux principaux types d'identités :

(a) les **identités « comptables »**. Dans ce cas, les coefficients de (B_g , C_g) prennent souvent leurs valeurs dans l'ensemble {-1, 0, +1}.

Ainsi, en sociologie (économie), l'équation suivante :

(4) $R\ =\ C + E$,

dans laquelle R représente le revenu disponible brut, C la consommation finale et E l'épargne brute des ménages, est une identité comptable ;

(b) les **identités de définition**. Les variables (ξ, η) interviennent généralement de façon non additive entre elles dans l'identité formant l'équation n° g, leurs coefficients étant soit connus, soit inconnus. Ainsi :

$(b)_1$ en physique, la *« loi de E. MARIOTTE »* d'équation :

(5) $(P\ V) / T = (P_0\ V_0) / T_0 =$ constante,

dans laquelle P représente la pression d'un gaz parfait, V son volume et T sa température, résulte d'une propriété des gaz parfaits. Elle est parfois aussi considérée comme une identité de définition (ou parfois comme une contrainte dans un système).

Il en va de même de la « loi d'attraction » de I. NEWTON :

(6) $A = (1/2) . d^{-2} . M_1 . M_2$,

dans laquelle d représente une distance entre deux corps physiques 1 et 2 de masses respectives M_1 et M_2 .

$(b)_2$ en sociologie (économie), les équations :

(7) $I = \tau * B$,

ou :

(8) $v = p' q$,

dans lesquelles I représente l'impôt sur le bénéfice des sociétés, τ le taux d'imposition et B le bénéfice imposable, v une valeur, p un vecteur de prix et q un vecteur de quantités, sont des identités de définition.

(iii) Du point de vue de l'**inférence statistique**, l'existence d'identités ne pose pas, en général, de difficultés :

(a) on peut souvent résoudre les identités pr aux variables endogènes, et les expressions ainsi obtenues pour ces variables sont ensuite substituées dans les « véritables » équations du modèle, souvent appelées **équations de comportement** ;

(b) on peut aussi utiliser les **procédures statistiques** usuelles (estimation, tests, prévision) avec contraintes : ces **contrainte sur les variables** (donc sur leurs observations) sont donc représentées par les identités elles-mêmes.

(iv) Dans un modèle d'interdépendance, on distingue ainsi entre **identité** (au sens précédent) et **équation de comportement**. Ces dernières décrivent des propriétés caractérisant le **phénomène** représenté par le modèle : ie elles constituent les « **lois** » (au sens de l'**homme de l'art**) gouvernant le phénomène en question.

Ainsi, l'équation de MARIOTTE (5) est, en fait, une propriété caractéristique des gaz parfaits (du moins dans la théorie élémentaire). Une équation de comportement peut donc aussi jouer le rôle d'identité dans le contexte d'un modèle différent (cf aussi **analyse des modèles**).