LOI SCIENTIFIQUE (A10, A16, C, J, O)
(12 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) L'expression générale, souvent vague, de « **loi scientifique** » peut cependant être entendue selon deux sens plus précis :

(a) soit comme la **loi de probabilité** (théorique) qui « gouverne » un **phénomène** donné, étudié dans le cadre d'un **domaine de connaissance**. Ce premier sens correspond à la notion la plus générale : connaître la loi en question équivaut à connaître entièrement le phénomène considéré. Cette acception n'implique pas de causalité particulière ;

(b) soit comme **relation fonctionnelle** (théorique) reliant une **variable endogène** η (simple ou multiple) à une **variable exogène** ξ (simple ou multiple), chacune de ces variables pouvant comporter des composantes de type numérique ou de type qualitatif (cf **loi multivariée**). Ce deuxième sens suggère souvent une idée de **causalité** et semble le plus couramment perçu, quoique souvent implicitement (voire même inconsciemment). Deux exemples typiques de ce sens sont la **loi multidimensionnelle** (variables numériques) et le **tableau de contingence** (variables qualitatives).

(ii) Diagramme non commutatif

Sauf dans le cas d'une **simulation**, dans laquelle la loi est connue (donnée) a priori, les deux notions précédentes doivent généralement faire l'objet d'une **estimation** : l'objectif est d'approcher au maximum la **lp** théorique ou le lien fonctionnel par l'**observation** du phénomène (cf aussi **estimation d'une loi**).

Soit $\mathscr{L}(\zeta)$ la **lp** (théorique) du phénomène, dont la liste des variables ζ est partitionnée en variables exogènes ξ et variables endogènes η selon $\zeta = (\xi, \eta)$. On note $\rho : \xi \mapsto \eta$ une relation fonctionnelle faisant l'objet d'une hypothèse (eg une **fonction de régression** $x \mapsto y = \rho(x)$).

Les N observations disponibles sont un N-**échantillon** Z = (X, Y) de ζ. On note alors $\mathscr{L}(Z)$ la loi de Z et $L_N(Z)$ la **loi empirique** associée à Z. Par suite, deux approches d'estimation sont possibles :

(a) une approche courante consiste à estimer directement ρ, calculée à partir de $\mathscr{L}(\zeta)$, à l'aide d'un premier **estimateur** $\rho_N^{(1)}$ (eg $r_N^{(1)}$ si ρ est une fonction de régression). D'où le chemin $\mathscr{L}(\zeta) \rightarrow \rho \rightarrow \rho_N^{(1)}$;

(b) une autre approche, indirecte, consiste à déterminer un estimateur L_N (z) de $\mathscr{L}$ (z) (eg la **loi empirique** P_N), dont se déduit la relation fonctionnelle $\rho_N^{(2)}$, considérée comme estimateur de ρ (eg $y^{(2)} = r_N^{(2)}$ (x) si ρ est une fonction de régression). D'où le chemin $\mathscr{L}(\zeta) \to L_N(z) \to \rho_N^{(2)}$.

Les deux estimateurs $\rho_N^{(1)}$ et $\rho_N^{(2)}$ obtenus en bout de chaîne ne sont pas, en général, identiques. La distance entre eux peut donc être plus ou moins éloignée de zéro.