VARIABLE LATENTE (G, J, K, K12)
(19 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) De façon générale, on appelle **variable latente** une **variable statistique** ou une **variable aléatoire**, par nature **inobservable**, sensée intervenir au sein d'un **phénomène** donné. On parle aussi de **variable cachée**, de **variable non apparente** ou encore de **variable sous-jacente**.

Ce type de variable est souvent pris en compte dans la **modélisation** ou la **spécification** d'un **modèle statistique** associé au phénomène.

On suppose que (ξ, η, υ) est un triplet aléatoire dont la composante (ξ, η) est observable et la composante υ inobservable. La loi jointe du triplet est notée $P^{(\xi, \eta, \upsilon)}$ ou $\mathscr{L}(\xi, \eta, \upsilon)$: elle fait partie d'une **famille de lois** $\mathscr{P}^{(\xi, \eta, \upsilon)}$ susceptibles de gouverner ce triplet ou d'engendrer les **« observations »** de ce triplet (ie les observations « vraies » de (ξ, η) et les valeurs inobservables correspondantes de υ).

Des situations tq les suivantes peuvent notamment se présenter :

(a) **marginalisation**, ou non prise en compte explicite de υ dans le modèle. Bien que le phénomène considéré dépende aussi de l'**action** de υ, la relation d'intérêt est cependant une **relation fonctionnelle** tq $\eta = \rho(\xi)$ dérivée de la **loi conjointe** $P^{(\xi, \eta)}$ (**loi marginale** de la précédente), laquelle fait partie d'une **famille** $\mathscr{P}^{(\xi, \eta)}$ décrivant le modèle ;

(b) **inobservable** → **observable**, ou prise en compte de υ sous la forme $\xi = f(\upsilon)$ et $\eta = g(\upsilon)$ (ie existence d'influences, inobservables, de υ sur ξ et η). Alors :

(b_1) si f et g ne sont pas connues (situation a priori), l'observable (ξ, η) ne peut être modélisé qu'à partir de leur **loi jointe** $P^{(\xi, \eta)}$, cette loi incorporant implicitement les effets de υ. Cette modélisation peut prendre une forme tq la relation $\eta = \rho(\xi)$ précédente ;

(b_2) si f et g sont connues (ie de formes analytiques explicitées), et si f est inversible, il est possible d'exprimer η en fonction de ξ (ou inversement) en éliminant υ, ie : $\upsilon = f^{-1}(\xi)$, d'où $\eta = g(f^{-1}(\xi))$. L'observable (ξ, η) ne peut donc être modélisé qu'à partir de leur loi jointe $P^{(\xi, \eta)}$, cette loi incorporant implicitement les effets de υ ;

(c) prise en compte de υ sous la forme $\eta = r(\xi, \upsilon)$ (ie existence d'une influence directe, mais inobservable, de υ sur η). La relation r peut être connue ou non. Si υ est considérée comme exerçant une influence négligeable sur η, un **modèle de régression** ou un **modèle d'interdépendance** peut être mis en oeuvre.

(ii) Les variables latentes se rencontrent, en particulier :

(a) en **analyse des données** : cf eg **analyse en facteurs communs et spécifiques**, **composante principale** ;

(b) dans des **relations fonctionnelles** usuelles : **modèle de régression**, **modèle d'interdépendance** (cf aussi **perturbation aléatoire**) ;

(c) dans l'étude plus spécifique des **structures latentes** : cf **modèle de structures latentes**, **modèle des classes latentes**.