CARACTÉRISTIQUE CONDITIONNELLE, CARACTÉRISTIQUE MARGINALE (C5, F3)
(17 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'analyse d'un **phénomène** particulier prend en compte un **ensemble** (ou une **famille**) de **variables** décrivant ce phénomène : ces variables peuvent donc être de divers types (cf **type de variable aléatoire**). Or, certaines de ces variables peuvent, le cas échéant :

(a) soit sortir du cadre du phénomène considéré, ie concerner d'autres phénomènes appartenant toutefois au même **domaine de connaissance** ;

(b) soit encore concerner des phénomènes relevant d'autres **domaines de connaissance**.

De plus, pour l'**homme de l'art** ou le **statisticien**, une **variable externe** (au phénomène) peut être, sans exclusivité :

(a) une inconnue, ie être une variable absente ou manquante (cf **erreur de spécification**) : cf **variable cachée**, **variable latente** ;

(b) une **inobservable**, ie le **système statistique** qui dessert le scientifique ne peut procéder à son **observation** (cf **observabilité**, **production statistique**) : ceci peut être le cas eg d'une variable tronquée (cf **troncature**) ou d'une variable censurée (cf **censure**), lesquelles ne sont pas observables sur toute l'étendue de leurs valeurs théoriques (cf aussi **étendue**, **support d'une probabilité**).

Dans certains cas, une **variable proche** (« proxy ») peut être associée à une variable externe, ce qui contribue à « compléter » la « liste » des variables prises en compte. Cette solution implique l'existence d'une « relation forte » (**association**, **corrélation**, **relation fonctionnelle** ou encore **régression**) entre ces deux variables, relation dont la qualité doit donc être faire l'objet d'un **test**.

(i) Dans un **contexte** de « carence » du type précédent, l'analyse peut suivre deux démarches, en général alternatives, ou parfois conjuguées (cf aussi **inférence conditionnelle**) :

(a) **marginalisation** : le raisonnement se réalise quelles que soient les valeurs des variables externes. Le concept sous-jacent est donc celui de **loi marginale** ;

(b) **conditionnement** : le raisonnement se réalise en supposant que ces variables externes prennent certaines valeurs, même si ces dernières sont inconnues. Le concept sous-jacent est donc celui de **loi conditionnelle** (cf aussi **probabilité de transition**).

Dans les deux cas, c'est la loi retenue (marginale ou conditionnelle) qui sera considérée comme gouvernant la famille des variables prises en compte. C'est, en général, une **loi multivariée** (eg une **loi multidimensionnelle**).

Par suite, une **caractéristique légale** est appelée :

(a) **caractéristique conditionnelle** ssi elle est calculée à l'aide d'une **loi conditionnelle** ;

(b) **caractéristique marginale** ssi elle est calculée à l'aide d'une **loi marginale**.

(i) **Caractéristique conditionnelle théorique**. Soit $(\Omega, \mathscr{T}, P)$ un **espace probabilisé**, $(\mathscr{X}, \mathscr{B})$ et $(\mathscr{Y}, \mathscr{C})$ deux **espaces mesurables** (eg **espaces d'observation**) et $\zeta = (\xi, \eta) : \Omega \mapsto \mathscr{X} \times \mathscr{Y}$ un **couple aléatoire** dont la **lp** est notée $P^{\zeta} = P^{(\xi, \eta)}$ (cf **loi multivariée**). On suppose définie la **loi conditionnelle** $P(. / \xi)$ de η sachant ξ et l'on note $\gamma_{\eta} : P(. / \xi) \mapsto \mathscr{Y}$ une application **caractéristique** donnée.

On appelle alors **caractéristique conditionnelle** de η relativement à ξ soit l'application γ_{η} précédente, soit, le plus souvent, la valeur même de cette application.

(ii) Ainsi, lorsque $\mathscr{X} = \mathbf{R}^K$, $\mathscr{Y} = \mathbf{R}^G$ et que h est la densité du couple $(\xi, \eta) : \Omega \mapsto \mathbf{R}^K \times \mathbf{R}^G$, supposée absolument continue pr à la **mesure de LEBESGUE** λ_{K+G} (cf **mesure absolument continue**), alors la **densité conditionnelle** de η sachant ξ est donnée par la relation :

(3) $\quad g(y / x) = h(x, y) / f(x) \quad (\lambda_{K+G}\text{-p.p.}),$

avec $f(x) = \int_{\mathbf{R}G} h(x, y)\, dy$ (**densité marginale** de ξ) (où **R**G désigne $\mathbf{R}^G$).

Si $\varphi : \mathbf{R}^G \mapsto \mathbf{R}^H$ (avec $H > 1$) est une fonction λ_G-intégrable, l'**espérance mathématique conditionnelle** de la va $\varphi(\eta)$ sachant ξ est :

(4) $\quad E(\varphi(\eta) / x) = E(\varphi(\eta) / \xi = x) = \int_{\mathbf{R}G} \varphi(y)\, g(y / x)\, dy.$

En particulier, si $\varphi(\eta) = \eta_1^{j(1)} \ldots \eta_G^{j(G)} = \Pi_{g=1}^{G} \eta_g^{j(g)}$ (en notant j(g) pour désigner un j_g, avec $\Sigma_{g=1}^{G} j_g = j$), on définit le **moment conditionnel** d'ordre j de η sachant $\xi = x$.

(iii) **Caractéristique conditionnelle empirique**. Des considérations analogues permettent de définir une **caractéristique conditionnelle empirique**. à l'aide d'un N-échantillon $(X, Y) = (X_1, \ldots, X_N, Y_1, \ldots, Y_N)$, encore écrit $((X_1, Y_1), \ldots, (X_N, Y_N))$, du couple $\zeta = (\xi, \eta)$:

(1) $\quad \xi \mapsto \gamma_{\eta\xi}(N) \qquad$ (ou encore $\gamma_N(Y / X)$ ou $\gamma_N(\eta / \xi)$).

Cette caractéristique est calculée à partir de la **loi empirique conditionnelle** $P_N(. / X)$ qui se déduit de P_N (**loi empirique** associée au couple (X, Y)) (cf **statistique naturelle**).

(iv) **Caractéristique marginale théorique**. Dans le même cadre qu'initialement, on note P^{ξ} (resp P^{η}) la **loi marginale** (ie la **loi propre**) de ξ (resp de η), définie par sa **densité** :

(5) $\quad f(x) = \int h(x, y)\, d\mu_y(y), \quad$ (resp $g(y) = \int h(x, y)\, d\mu_x(x)$),

où μ_x et μ_y sont les composantes d'une **mesure positive** $\mu = \mu_x \otimes \mu_y$ supposée dominer la loi $P^{(\xi, \eta)}$. Si l'on note γ_x (resp γ_y) une caractéristique de la loi P^{ξ} (resp P^{η}), on dit que l'application :

(6) $\quad P^{\xi} \mapsto \gamma_x \qquad$ (resp $P^{\eta} \mapsto \gamma_y$)

est une **(application) caractéristique marginale** (théorique) de P^{ξ} (resp P^{η}).

(v) **Caractéristique marginale empirique**. On peut alors définir l'analogue empirique de la caractéristique théorique précédente, de façon parallèle à la définition d'une caractéristique conditionnelle empirique.