TEST DES MULTIPLICATEURS DE LAGRANGE (A11, H3, I4)
(27 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **test des multiplicateurs** de LAGRANGE est général en **Statistique** : il peut être défini dès que les notions de **vraisemblance** (cf **famille de lois dominée**, **modèle dominé**) et de **différentiabilité** ont un sens. En effet, les méthodes d'**optimisation** statistiques comportent, très généralement, des **contraintes** (égalités, inégalités ou contraintes mixtes) : ceci conduit à l'utilisation de la méthode des muliplicateurs de J.L. LAGRANGE (cf **lagrangien**, **équations de EULER-LAGRANGE**), ou encore de la méthode de H.W. KUHN - A.W. TUCKER (cf **conditions de KUHN-TUCKER**, **théorème de KUHN-TUCKER**).

Ses domaines d'applications sont ainsi très variés.

(i) Soit ($\mathscr{X}$, $\mathscr{B}$, $(P_\theta^X)_{\theta \in \Theta}$) un **modèle image** paramétrique dans lequel $\mathscr{X}$ = **R**, X = $(X_1, ..., X_N)$ est l'**échantillon** observé et $\Theta \subset \mathbf{R}^Q$. On suppose la **famille** $\mathscr{P}^X = (P_\theta^X)_{\theta \in \Theta}$ uniformément dominée par une **mesure positive** σ-finie μ et l'on note :

(1) f_θ ou $f(. , \theta) = dP_\theta^X / d\mu$

la **dérivée de NIKODYM-RADON** (ie **fonction de vraisemblance**) associée à toute **loi** P_θ^ξ (où $\theta \in \Theta$).

Etant donné une fonction mesurable $h : \Theta \mapsto \mathbf{R}^K$ (avec $K < Q$) (cf **application mesurable**), on définit l'**ensemble** $\Theta_0 \subset \Theta$ selon :

(2)
$$\Theta_0 = \{\theta \in \Theta : h(\theta) = 0\},$$
$$(\text{resp } \Theta_0 = \{\theta \in \Theta : h(\theta) = h(\theta_0)\},$$

où $\theta_0 \in \Theta$ est donné, et l'on suppose que $\Theta_0 \neq \varnothing$ et $\Theta_0 \neq \Theta$.

On considère alors un **problème de test** portant sur l'**hypothèse de base** :

(3) $H_0 : \theta \in \Theta_0$.

On appelle **test des multiplicateurs de J.L. LAGRANGE** le test de H_0 fondé sur la **statistique de test** suivante, dite **statistique de J. L. LAGRANGE** :

(4) $L_N = S(X, \tilde{\theta})' \{I(\tilde{\theta})\}^{-1} S(x, \tilde{\theta})$,

dans laquelle :

(a) $\tilde{\theta}$ est l'**estimateur du maximum de vraisemblance** de θ estimé sous la **contrainte** $\theta \in \Theta_0$ (estimateur contraint par l'hypothèse H_0) ;

(b) $I(\theta)$ est la matrice d'**information de FISHER** (évaluée au « point » $\tilde{\theta}$) (cf **matrice d'information**) ;

(c) $S(X, \theta)$ est le **score** associé à la vraisemblance f_θ et défini par (cf aussi **score**) :

(5) $$S(X, \theta) = D_2 \operatorname{Log} f_\theta (X).$$

C'est donc le vecteur **gradient** de la fonction vectorielle $\theta \mapsto \operatorname{Log} f_\theta$, qui est aussi évalué au point $\theta^{\sim}$.

(ii) Sous l'hypothèse H_0, ce test (asymptotique) est basé sur la **propriété asymptotique** suivante (**convergence en loi**) :

(6) $$\mathscr{L}(L_N) \to^{H0}{}_{N \to +\infty} \mathscr{X}_K{}^2$$ (**loi du chi-deux** à K **degrés de liberté**).

On peut alors tester H_0 contre une **alternative** tq $H_1 = H_0{}^c$ (ie $H_1 : \theta \in \Theta_1 = \Theta_0{}^c$) puisque (6) permet de définir une **région critique** asymptotique pour tout seuil $\alpha \in]0, 1[$ donné.

(iii) L'origine du nom attribué à ce test vient du fait que l'**estimateur du mv** sous contrainte $\theta^{\sim}$ est solution du problème de **programmation mathématique** suivant :

(7) $$\sup_{\theta \in \Theta_0} \operatorname{Log} f_\theta (x),$$

dont le **lagrangien** (ou fonction de LAGRANGE) associé(e) :

(8) $$L(\theta, \lambda) = \operatorname{Log} f_\theta (X) - \lambda' h(\theta), \quad \text{avec } \lambda \in \mathbf{R}^K,$$

conduit, sous des **conditions de régularité** classiques, aux équations (ie aux conditions du premier ordre) :

(9) $$\begin{cases} D_2 \operatorname{Log} f_\theta (x) = \lambda' D h(\theta), \\ h(\theta) = 0, \end{cases}$$

dans lesquelles λ est le multiplicateur (vectoriel) de J.L LAGRANGE.

(iv) Un exemple classique d'application est celui du **modèle de régression linéaire** multiple $y = X b + u$, dans lequel on suppose que $u \sim \mathcal{N}_N (0, \Sigma)$ (où Σ est connue), que $\Theta_0 = \{b \in \mathbf{R}^K : R b = r\}$, où $R \in M_{JK} (\mathbf{R})$ (de rang rg $R = J < K$) et $r \in \mathbf{R}^J$ sont connus, et ou $\Theta_1 = \Theta_0{}^c$.

Le lagrangien s'écrit :

(10) $$L(b, \lambda) = - 2^{-1} \{N \operatorname{Log} (2 \pi) + \operatorname{Log} |\Sigma| + (y - X b)' \Sigma^{-1} (y - X b)\} - \lambda' (R b - r),$$

et conduit, après **dérivation** pr à θ, à la **statistique des multiplicateurs de J.L. LAGRANGE** suivante :

(11) $$L_N = (R b^\# - r)' \{R (X' \Sigma^{-1} X)^{-1} R'\}^{-1} (R b^\# - r),$$

dans laquelle $b^{\#} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} y$ est l'estimateur du maximum de vraisemblance (gaussienne) non contraint de b (ie son **estimateur des moindres carrés généralisés**).

On remarque que L_N est égale à la **statistique** de WALD (cf **test de WALD**) et à la statistique du **rapport des vraisemblances** R_N .

(v) Le test précédent, défini à partir de la statistique (4), est aussi appelé **test des multiplicateurs de H.W. KUHN - A.W. TUCKER** lorsque Θ_0 est défini par des inégalités :

$$(2)' \quad \begin{array}{l} \Theta_0 = \{\theta \in \Theta : h(\theta) \geq 0\}, \\ (\text{resp } \Theta_0 = \{\theta \in \Theta : h(\theta) \geq h(\theta_0)\}, \end{array}$$

Ainsi, dans le modèle de régression linéaire précédent, on aurait, par exemple, $R b \geq r$. Le test serait celui de l'hypothèse H_0 définie par : $E b = r$.