

TEST DE COX (G2, I2, I3, J6)

(10 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **test de COX** est un test d'hypothèses « séparées » (ie non emboîtées), notamment utilisé comme **test de sélection de modèles**.

(i) Soit $(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta})$ un **modèle image** dominé par une **mesure positive** σ -finie μ , où $X = (X_1, \dots, X_N)$ est un **échantillon iid** selon une des **loi** P_θ^X . Soit Θ_0 et Θ_1 deux **parties** non vides de Θ . On note :

$$(1) \quad f(x, \theta) = (dP_\theta^X / d\mu)(x), \quad \forall (x, \theta) \in \mathcal{X} \times \Theta,$$

la **fonction de vraisemblance** du modèle et :

$$(2) \quad f_i(x, \theta) = (dP_\theta^X / d\mu)(x), \quad \forall (x, \theta) \in \mathcal{X} \times \Theta_i,$$

la **restriction** de f à Θ_i ($i = 0, 1$), ie la fonction de vraisemblance du **sous-modèle** $((\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta(i)}))$, où $\Theta(i)$ désigne par commodité Θ_i ($i = 0, 1$).

On veut tester l'**hypothèse de base** :

$$(3) \quad H_0 : \theta \in \Theta_0,$$

contre l'**hypothèse alternative** :

$$(4) \quad H_1 : \theta \in \Theta_1,$$

où l'on suppose que H_0 et H_1 sont :

(a) des **hypothèses non emboîtées**, au sens où l'on a ni $\Theta_0 \subset \Theta_1$, ni $\Theta_1 \subset \Theta_0$;

(b) des **hypothèses « séparées »**, ie aucune loi P_θ^X tq $\theta \in \Theta_0$ ne peut être obtenue comme « limite » (quelconque) d'une **famille** de lois P_θ^X tq $\theta \in \Theta_1$.

Soit S (resp T) l'**estimateur** de θ obtenu par maximisation de la **vraisemblance** f_0 (resp f_1). Cet estimateur est donc contraint par $\theta \in \Theta_0$ (resp $\theta \in \Theta_1$) (cf **contrainte**).

Le **test de D.R. COX** est fondé sur la **statistique de D.R. COX** suivante (cf aussi **test du rapport des vraisemblances**) :

$$(5) \quad C_N = N^{-1} \cdot \sum_{n=1}^N \text{Log} \{f_{0,1}^{\sim}(X_n, T) / f_{0,0}^{\sim}(X_n, S)\},$$

dans laquelle $f_{0,i}^{\sim}$ ($i = 0, 1$) est la densité élémentaire composant $f_i^{\sim}$, ie $f_i^{\sim}(x_1, \dots, x_N, \theta) = \prod_{n=1}^N f_{0,i}^{\sim}(x_n, \theta)$.

(ii) On montre que, sous l'hypothèse H_0 et sous certaines **conditions de régularité**, il existe $\sigma(\theta) > 0$ tq la **convergence en loi (normalité asymptotique)** suivante :

$$(6) \quad N^{1/2} \cdot (C_N - E_\theta C_N) \xrightarrow{\mathcal{L}_{N \rightarrow +\infty}} \mathcal{N}_1(0, \sigma^2(\theta))$$

est vérifiée.

Sous l'hypothèse H_1 , $P_\theta\text{-}\lim_N C_N = +\infty$ ($\forall \theta \in \Theta_1$) : le test possède donc une **puissance** élevée et permet de rejeter H_0 lorsque $C_N \gg 0$. Si $\alpha \in]0, 1[$ est le seuil retenu pour le test (cf **risque de première espèce**), la **région critique** est de la forme :

$$(7) \quad w = \{C_N / \sigma(S) > q_{1-\alpha}\},$$

où $q_{1-\alpha}$ est le **quantile** d'ordre $1 - \alpha$ de la **loi normale réduite** $\mathcal{N}(0, 1)$.

Cette région est asymptotique et dépend de certaines conditions tq :

$$(8) \quad \{\sigma(S)\}^{-1} \cdot C_N \xrightarrow{\mathcal{L}_{N \rightarrow +\infty}} \mathcal{N}_1(0, 1).$$

Si l'on pose, parmi d'autres formulations possibles :

$$(9) \quad f_\lambda(x, \theta) = [(f_0(x, \theta))^\lambda \cdot (f_1(x, \theta))^{1-\lambda}] / c(\theta, \lambda), \quad \forall \lambda \in [0, 1],$$

où $c(\theta, \lambda)$ est une constante de **normalisation** tq f soit une **densité de probabilité** (somme à 1), le test précédent est souvent présenté comme test de l'hypothèse :

$$(10) \quad H_0' : \lambda = 1$$

contre l'hypothèse :

$$(11) \quad H_1' : \lambda = 0.$$