MODÈLE DE COX (F8, J8, N)
(08 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **modèle de COX** se relie à l'étude des **durées de vie** (ou de survie) d'**unités statistiques** sujettes à une **fin de vie** ou à une **disparition** eg : par désagrégation ou fin de fonctionnement (pannes ou « défaillances » d'unités physiques, décès d'unités biologiques), par transformation (d'unités sociologiques), etc.

(i) Soit (Ω, $\mathcal{T}$, $\mathcal{P}$) un **modèle statistique** de base, ($\mathcal{Z}$, $\mathcal{C}$) un **espace mesurable** auxiliaire et (τ, ζ) : $\Omega \mapsto \mathbf{R}_+ \times \mathcal{Z}$ un **couple aléatoire** dans lequel τ représente le « **temps** » (cf **durée de vie**) et ζ une **variable exogène** appelée **variable « concomitante »**.

On suppose que la **loi conditionnelle** de τ pr à ζ possède une fonction de **taux d'échec** r : $\mathbf{R}_+ \mapsto \mathbf{R}_+$ de la forme :

(1) $\quad r\,(t\,/\,z) = r_0\,(t)\,.\,w\,(z, b), \quad \forall\, b \in \mathbf{R}^Q,$

dans laquelle r_0 (t) est un **hasard de référence** (calculé lorsque t = 0) et w (z, b) un **facteur de risque** relatif tq w (t, 0) = 1. Le **paramètre d'intérêt** est alors b.

On appelle alors **modèle à hasard proportionnel de D.R. COX**, ou **modèle à taux d'échec proportionnel de D.R. COX**, le modèle défini par l'équation (1).

Lorsque b = 0, z n'exerce aucun effet sur le hasard r (cf aussi **hasard**, **fonction de hasard**).

Lorsque ζ est un vecteur de **vars** exogènes ζ_1 ,..., ζ_L et que Q = L, alors :

(a) si w (z, b) = v (z' b) (ie ne dépend que du **produit scalaire**), on définit le **modèle de D.R. COX linéaire** ;

(b) si, de plus, Log v (z' b) = z ' b, on obtient le modèle de COX original.

(ii) Etant donné un **ensemble** (eg **échantillon**) A (Card A = N > L) d'unités statistiques a_n (où $n \in N_N^*$), on observe les instants (aléatoires) de disparition t_n de ces unités, ainsi que les valeurs z_n correspondantes de ζ. On suppose que ces instants sont ordonnés selon $t_1 < ... < t_N$ et l'on note Z_n la n-ième **observation** de ζ (ie celle effectuée sur a_n).

On appelle **ensemble à risque** à l'instant t_n l'ensemble R_n des unités a_n qui survivent jusqu'en t_n . Le couple (t_n , Z_n) est donc l'observation de (τ, ζ) associée à l'unité a_n qui risque de disparaître à l'instant t_n . La variable τ « censure » ainsi la variable ζ, laquelle est observée selon Z_n tant que $t < t_n$ (cf **censure**). On note enfin p_n (z) la **proportion** des unités $a_n \in R_n$ pour lesquelles on observe ζ et Z (n) l'ensemble des valeurs Z_n des unités $a_n \in R_n$.

Pour estimer b, la **méthode du maximum de vraisemblance**, adaptée au cadre précédent, conduit à définir la **vraisemblance** conditionnelle de b (pr aux instants de disparition antérieurs) selon (contribution de R_n) :

(2) $L_n (Z_n , b) = p_n (Z_n) . w (Z_n , b) / \Pi_{z \in Z (n)} p_n (z) w (z, b), \qquad \forall n \in N_N*.$

On appelle **(fonction de) vraisemblance « partielle »** (ie pr à b) du modèle le produit des vraisemblances précédentes, ie :

(3) $L (Z, b) = \Pi_{n=1}^{N} L_n (Z_n , b).$

L'estimateur $b^{\#}$ recherché est alors solution du problème de **programmation mathématique** :

(4) $\max_{b \in \mathbf{R}Q} L (Z, b)$ (où **R**Q désigne $\mathbf{R}^Q$).

Ainsi, dans le cas eg du modèle original, on peut écrire :

(5) $L_n (Z_n , b) = r (t, Z_n) / \Pi_{\{n : a(n) \in Rn\}} r (t, Z_n) = Z_n' b / \Pi_{\{n : an \in Rn\}} Z_n' b.$

et :

(6) $L (Z, b) = \Pi_{\{n : n \in Dn\}} L_n (Z_n , b),$

où D_n (aussi noté Dn) désigne l'ensemble des unités décédées à l'instant t_n et Rn désigne R_n .