PROBLÈME DE L'INDÉPENDANCE (D1, I2)
(12 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **problème de l'indépendance consiste**, dans sa forme élémentaire, à s'assurer que les coordonnées d'un couple aléatoire sont des variables indépendantes (cf **indépendance stochastique**). Ceci est réalisé à partir d'un échantillon supposé engendré par ce couple.

(i) **Indépendance dans un couple aléatoire**. Soit $(\Omega, \mathcal{T}, \mathcal{P})$ un **modèle statistique**, $(\mathcal{X}, \mathcal{B})$ et $(\mathcal{Y}, \mathcal{C})$ deux **espaces mesurables** et $\zeta : (\xi, \eta) : \Omega \mapsto \mathcal{X} \times \mathcal{Y}$ un **couple aléatoire** dont la **loi** $P^{\zeta} = P^{(\xi,\ \eta)}$ est l'image par ζ de l'une des **mesures de probabilité** $P \in \mathcal{P}$.

On observe un **échantillon** $Z = (Z_1, ..., Z_N)$ de ζ, avec $Z_n = (X_n, Y_n)$, $\forall\ n \in N_N{}^*$. On suppose que Z est un **échantillon iid** selon P^{ζ} et l'on note $X = (X_1, ..., X_N)$ et $Y = (Y_1, ..., Y_N)$ les échantillons composants.

Dans ce cadre, le **problème de l'indépendance** est le **problème de test** de l'**hypothèse d'indépendance** (ou **test d'indépendance**), ie de l'**hypothèse de base** suivante :

(1) $H_0 : \xi \perp \eta$ (ie ξ et η sont indépendantes),

ou encore :

(2) $H_0 : P^{\zeta} = P^{\xi} \otimes P^{\eta}$,

où P^{ξ} (resp P^{η}) est la **loi marginale** (ie **loi propre**) de ξ (resp de η).

Le test en question peut s'effectuer de plusieurs façons, selon que l'on met en oeuvre des **lois de probabilité**, des **fonctions de répartition**, des **fonctions caractéristiques** ou encore des **caractéristiques légales** (**coefficient de corrélation**, etc).

Ainsi, dans le cas des fr relatives à la **mesure de LEBESGUE** λ_2 définie sur la **tribu** $\mathcal{B}(\mathbf{R}^2)$, le test de H_0 revient au test de l'hypothèse :

(3) $H_0' : H(x, y) = F(x) . G(y), \quad \forall\ (x, y) \in \mathbf{R}^2$,

où F (resp G, resp H) désigne la fr de ξ (resp de η, resp de ζ), avec $F(x) = H(x, +\infty)$, $\forall\ x \in \mathbf{R}$, et $G(y) = H(+\infty, y)$, $\forall\ y \in \mathbf{R}$ (ie presque partout pour λ_1).

Dans un contexte « non paramétrique », on désigne resp par $X^{(.)}$ et $Y^{(.)}$ les **statistiques d'ordre** de X et Y, et par R et S leurs **statistiques de rang**. On pose :

(4) $\rho(\alpha) = \beta \Leftrightarrow R_{\beta} = \alpha, \quad \forall\ \alpha \in N_N{}^*$,

puis :

(5) $R_\alpha^* = S_{\rho(\alpha)}$ et $R^* = (R_1^*, \ldots, R_N^*)$.

On montre alors que, quelle que soit l'**hypothèse alternative** simple :

(6) $H_a : H = H^*$ (où H^* est une fr donnée),

pour laquelle la **vraisemblance** s'écrit (en notant h la **densité** de P^ζ pr à λ_2 et h* celle associée à H*) :

(7) $l(z) = l(z_1, \ldots, z_N) = \Pi_{n=1}^{N} h^*(z_n)$, avec $z_n = (x_n, y_n)$, $\forall n \in N_N^*$,

il existe un **test de permutation** de **niveau** $\alpha \in]0,1[$ le plus puissant (cf **test le plus puissant**) tq (en notations usuelles) :

$$(8) \quad l(z) = \begin{cases} 1 & \text{ssi } h^*(z) > k(x^{(.)}, y^{(.)}, r^*), \\ \gamma \text{ (quelconque)} & \text{ssi } h^*(z) = k(x^{(.)}, y^{(.)}, r^*), \\ 0 & \text{ssi } h^*(z) < k(x^{(.)}, y^{(.)}, r^*). \end{cases}$$

(ii) **Indépendance interne à un échantillon**. Dans le cadre du **problème à un échantillon** $X = (X_1, \ldots, X_N)$, on appelle aussi **problème de l'indépendance**, ou **problème de l'échantillon aléatoire**, le **problème de test** portant sur l'**hypothèse d'indépendance** des coordonnées X_n de X, ie :

(9) $H_0 : P^X = P^{X(1)} \otimes \ldots \otimes P^{X(N)}$,

où $P^{X(n)}$ désigne la loi marginale (ou loi propre) de X_n, $\forall n \in N_N^*$, et les X(n) désignent les Xn (n = 1 ,..., N).

Divers tests de H_0 sont disponibles (cf **test d'indépendance**).