PROBLÈME À PLUSIEURS ECHANTILLONS (I2)
(05 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Il arrive souvent que le **statisticien** possède plusieurs **échantillons** relatifs à un **phénomène** donné, ou à des phénomènes comparables. Un problème usuel consiste alors à comparer ces échantillons (cf aussi **comparaison multiple**) : ce problème revient donc à définir divers **tests d'hypothèses** les concernant dans leur ensemble. Généralement, ces tests sont fondés sur des **statistique de test** déduites des échantillons (eg **estimateurs** divers).

(i) Soit $(\Omega, \mathcal{T}, \mathcal{P})$ un **modèle statistique**, $(\mathcal{X}, \mathcal{B})$ un **espace mesurable** (eg un **espace d'observation**) et $\xi_i : \Omega \mapsto \mathcal{X}$ ($i \in N_k^*$) une **suite** (finie) de **va** (indépendantes entre elles) dont les **lp** resp $P^{\xi(i)}$ admettent les **densités** $f_i = dP^{\xi(i)} / d\mu$ pr à une **mesure positive** μ donnée et les **fonctions de répartition** F_i , $\forall i \in N_k^*$.

Soit $X^i = (X^i_1 ,..., X^i_{N(i)})$ un N_i-**échantillon iid** généré par la **variable parente** ξ_i , $\forall i \in N_k^*$ (en désignant $\xi(i)$ par ξ_i et $n(i)$ par n_i).

On peut disposer l'ensemble X de ces va selon le **tableau statistique** (non « équilibré ») :

$$X_{11} ,..., X_{1,N(1)}$$
$$........$$
$$(1) \qquad X_{i1} ,..., X_{i,N(i)}$$
$$........$$
$$X_{k1} ,..., X_{k,N(k)} ,$$

dans lequel on suppose que $k \geq 2$ et $N_i \geq 1$, $\forall i \in N_k^*$. On pose enfin $\Sigma_{i=1}^k N_i = N$ (nombre total d'**observations**).

La **vraisemblance** associée à l'échantillon partiel X^i, s'écrit (échantillon iid) :

$$(2) \qquad p_i (x_i) = \Pi_{n(i)=1}^{N(i)} f_i (x_{i,n(i)}), \forall x_i \in \mathbf{R}^{N(i)}$$

(où les N(i) désignent les N_i).

Celle du N-échantillon d'ensemble X s'écrit (**indépendance** des X^i entre eux) :

$$(3) \qquad p (x) = \Pi_{i=1}^k p_i (x_i) = \Pi_{i=1}^k \Pi_{n(i)=1}^{N(i)} f_i (x_{i,n(i)}), \qquad \forall x \in \mathbf{R}^N.$$

(ii) On appelle **problème à plusieurs échantillons**, ou **problème à k échantillons**, un **problème de test** dont l'**hypothèse de base** est, le plus souvent, la suivante :

$$(4) \qquad H_0 : P^{\xi(1)} = ... = P^{\xi(k)} = P^{\xi(0)} \qquad \text{(identité entre elles des lois des } \xi_i\text{)},$$

ou encore, en termes de fonctions de répartition (resp de densités) :

$$(5) \qquad H_0 : F_i = F_0 \quad (\text{resp } f_i = f_0), \ \forall i \in N_k^*.$$

On appelle souvent H_0 l'**hypothèse de l'échantillon aléatoire**.

D'autres hypothèses de base sont aussi concevables (cf eg **test de BARTLETT**).

L'alternative peut être :

(a) une **alternative de position**, où l'on suppose l'existence d'une fr F fixe tq :

(6) $F_i(x_{i,n(i)}) = F(x_{i,n(i)} - \alpha_i), \quad \forall i \in N_k^*,$

où $\alpha_i \in \mathbf{R}$ est, $\forall i \in N_k^*$, un **paramètre de position** ;

(b) une **alternative d'échelle**, où l'on suppose l'existence d'une fr F fixe tq :

(7) $F_i(x_{i,n(i)}) = F(x_{i,n(i)} / \beta_i), \quad \forall i \in N_k^*,$

où $\beta_i \in \mathbf{R}^*_+$ est, $\forall i \in N_k^*$, un **paramètre d'échelle** ;

(c) une **alternative d'échelle et de position**, dans laquelle on suppose l'existence d'une fr F fixe tq :

(8) $F_i(x_{i,n(i)}) = F\{(x_{i,n(i)} - \alpha_i) / \beta_i\}, \quad \forall i \in N_k^*,$

où $(\alpha_i, \beta_i) \in \mathbf{R} \times \mathbf{R}_+^*$ est, $\forall i \in N_k^*$, un **paramètre d'échelle et de position** ;

(d) une **alternative ordonnée** ;

(e) ou encore une **alternative en parapluie**.

(iii) Divers tests permettent de résoudre le problème à k échantillons : **test de KRUSKAL-WALLIS** (des **rangs**), **test de MOOD** (de la **médiane**) (cf **méthode de BROWN-MOOD**), **test de J.C. KIEFER** (analogue, pour k échantillons, du **test de KOLMOGOROV-SMIRNOV**), etc.

(iv) En pratique, le cas où k = 2 est fréquent : **problème à deux échantillons**.

L'exemple-type est le **test de comparaison** entre :

(a) un **échantillon expérimental**, dont les **unités expérimentales** sont soumises à des **traitements** : **facteurs expérimentaux**, **stimuli**, etc ;

(b) un **échantillon témoin**, dont les unités ne sont pas soumises aux stimuli, ou sont soumises à d'autres traitements, ou encore à un traitement « fictif » (placebo).

Le but est de vérifier si les traitements exercent des « effets » différents de ceux observés sans traitements.