

RANG MULTIPLE (F6)

(21 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

Lorsque des **observations** doivent être ordonnées (cf **relation d'ordre**), et que certaines prennent une même valeur, la définition de la **statistique d'ordre** ou de la **statistique de rang** doit être modifiée (cf **modification**).

Soit $(\Omega, \mathcal{F}, P)$ un **espace probabilisé**, $(\mathcal{X}, \leq)$ un **ensemble** ordonné et $X : \Omega \mapsto \mathcal{X}^N$ un **échantillon aléatoire** (non nécessairement indépendant) dont les n coordonnées X_n sont à valeurs dans $\mathcal{X}$. On note $X^{(\cdot)} = (X^{(1)}, \dots, X^{(N)})$ l'**échantillon** ordonné associé à X .

(i) On dit qu'une coordonnée $X^{(n)}$ de $X^{(\cdot)}$ présente un **rang multiple** ssi il existe (au moins) une autre coordonnée de $X^{(\cdot)}$ qui est égale à $X^{(n)}$. Soit M le nombre de coordonnées de $X^{(\cdot)}$ égales à $X^{(n)}$ (y compris $X^{(n)}$), comptées à partir de la première, notée $X^{(L)}$. On a donc :

$$(1) \quad X^{(L)} = \dots = X^{(n)} = \dots = X^{(L+M-1)},$$

avec $L \in \mathbb{N}_N^*$, $0 \leq M \leq N - L + 1$, $X^{(\alpha)} < X^{(n)}$, $\forall \alpha < L$, et $X^{(\beta)} > X^{(n)}$, $\forall \beta > L + M - 1$. De façon explicite, on a :

$$(1) \quad X^{(1)} \leq \dots \leq X^{(L-1)} < X^{(L)} = \dots = X^{(n)} = \dots = X^{(L+M-1)} < X^{(L+M)} \leq X^{(L+M+1)} \leq \dots \leq X^{(N)}.$$

S'il existe G groupes de coordonnées égales entre elles, et distinctes entre groupes, on peut écrire (en notant de façon équivalente N_g ou $N(g)$) :

$$(2) \quad X^{(N(1) + \dots + N(g-1) + 1)} = \dots = X^{(N(1) + \dots + N(g))}, \quad \forall g \in \mathbb{N}_G^* \setminus \{1\},$$

avec :

$$(3) \quad \begin{aligned} &X^{(N(1) + \dots + N(g-1))} < X^{(N(1) + \dots + N(g-1) + 1)}, \quad \forall g \in \mathbb{N}_G^* \setminus \{1\}, \\ &\sum_{g=1}^G N_g = N \text{ et } N_g \geq 1, \quad \forall g \in \mathbb{N}_G^*. \end{aligned}$$

(ii) Dans cette **situation statistique**, il existe plusieurs façons de définir une statistique de rang $R = (R_1, \dots, R_N)$ associée à X . Ainsi :

(a) pour un groupe donné $X^{(N(1) + \dots + N(g-1) + 1)}, \dots, X^{(N(1) + \dots + N(g))}$, le rangement des valeurs peut se faire par **tirage au hasard** (uniforme) des nombres $N_1 + \dots + N_{g-1} + 1, \dots, N_1 + \dots + N_g$ avec la probabilité $1 / N_g$. On attribue alors aux **coordonnées multiples** du groupe g les rangs qui résultent de ce tirage ;

(b) la **méthode des rangs moyens** consiste à associer à chacune des coordonnées multiples un même rang, égal à la moyenne des rangs qu'elles ont dans $X^{(\cdot)}$;

(c) la **méthode des rangs extrêmes** consiste à poser, $\forall n \in \{\sum_{h=1}^{g-1} N_h + 1, \dots, \sum_{h=1}^g N_h\}$, $R_n = \sum_{h=1}^{g-1} N_h + 1$ (rang inférieur de la série), ou, alternativement, $R_n = \sum_{h=1}^g N_h$ (rang supérieur de la série) ;

(d) enfin, dans certains cas, on peut remplacer (temporairement), dans les calculs, les coordonnées multiples du g-ième groupe par les coordonnées « perturbées » :

$$\begin{aligned} Y^{(N(1)+\dots+N(g-1)+1)} &= X^{(N(1)+\dots+N(g-1)+1)} + \varepsilon_{g,1}, \\ (3) \quad Y^{(N(1)+\dots+N(g))} &= X^{(N(1)+\dots+N(g))} + \varepsilon_{g,N(g)} \end{aligned}$$

(où $n(g)$ désigne n_g), puis faire tendre indépendamment vers zéro les nombres $\varepsilon_{g,n(g)}$, où $n_g \in \{1, \dots, N_g\}$, $\forall g \in N_G^*$.

(iii) En pratique, des rangs multiples apparaissent souvent lorsque :

(a) les **lp** sont des **lois discrètes** ;

(b) ou encore lorsqu'elle sont continues et que les coordonnées X_n de X sont calculées de façon approximativement discrète, eg $X^{(n)}$ est à valeurs :

(b)₁ soit dans **Q** (ensemble des nombres rationnels) ;

(b)₂ soit, le plus souvent, dans une partie de **D** : il s'agit des nombres décimaux dont le développement est fini (précision numérique).

Dans ces situations, les **statistiques** $S = s(X)$, ou $T = t(X^{(\cdot)})$, ou $U = u(R)$, calculées sous l'hypothèse de continuité de la loi P^X de X , doivent être « corrigées » pour tenir compte des rangs multiples (multiplicité des rangements possibles).