

CLASSE EXTRÊME (F)

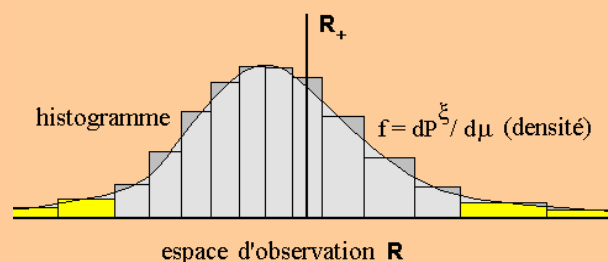
(23 / 12 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) On considère une **vars** ξ observée selon un N-**échantillon** $X = (X_1, \dots, X_N)$. On note Π_1 une partition de $\mathbf{R}$ en p classes (intervalles non vides et connexes de $\mathbf{R}$) notées $]a_0, a_1[, \dots,]a_{p-1}, a_p[$, et l'on classe les observations X_n dans ces classes en fonction de leur valeur. A chacune de ces classes est alors associée une **fréquence empirique absolue** n_j , et une fréquence empirique relative $f_j = n_j / \sum_h n_h$ ($j = 1, \dots, p$). La suite (finie) de couples $(]a_{j-1}, a_j[, f_j)_{j=1, \dots, p}$ définit ainsi un **histogramme** « scalaire » H_1 , ie :

$$(1) \quad n_j = \sum_{n=1}^N \mathbf{1}(X_n \in]a_{j-1}, a_j[), \quad (j = 1, \dots, p).$$

On appelle **classe extrême** de H soit la **classe de gauche** $]a_0, a_1[$, soit la **classe de droite** $]a_{p-1}, a_p[$, de cet histogramme (cf **valeur extrême**, **rapport des extrêmes**, **statistique des extrêmes**).

classes extrêmes d'un histogramme



Une acception plus large prendrait en compte un certain nombre de classes (contigües) situées le plus à gauche ou un certain nombre de classes (contigües) situées le plus à droite (deux classes latérales dans le graphique ci-dessus).

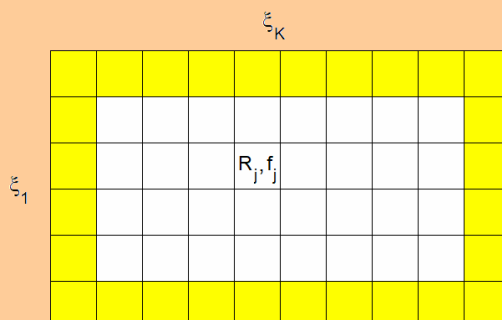
(ii) Dans le cas multidimensionnel (**stéréogramme**), on considère un **vecteur aléatoire** ξ , à valeurs dans $\mathbf{R}^K$, observé selon un N-**échantillon** $X = (X_1, \dots, X_N)$. On note Π_K une partition de $\mathbf{R}^K$ comportant p parties non vides connexes, notées $R_1, \dots, R_p$. Les observations X_n sont classées dans les p classes (parties de $\mathbf{R}^K$). A chacune de ces classes est encore associée une **fréquence absolue** empirique n_j , et une **fréquence relative** empirique $f_j = n_j / \sum_h n_h$ ($j = 1, \dots, p$). La suite (finie) de couples $(R_j, f_j)_{j=1, \dots, p}$ définit alors un **histogramme** « multidimensionnel » H_K , appelé **stéréogramme**, ie :

$$(2) \quad n_j = \sum_{n=1}^N \mathbf{1}(X_n \in R_j), \quad (j = 1, \dots, p).$$

On appelle alors, par extension, **classe extrême**, ou « **entourage extrême** », de H_K l'ensemble des classes de Π_K situées à la « périphérie » du support de H_K .

Classe extrême (partition en forme de quadrillage)

entourage extrême dans $\mathbf{R}^K$



(iii) Les notions empiriques précédentes correspondent à celle (théorique) de **queue d'une loi**.

Dans les deux **situations statistiques**, théorique aussi bien qu'empirique, la notion d'**extrémité** est une notion relative (ou corrélative), qui dépend de la notion de **centralité**. Elle peut faire référence à une conception de « **normalité** » (distincte de celle de **normalité**) ou à une conception d'**aberration**.

(iv) Il arrive souvent que les classes extrêmes soient non bornées, ie (cas scalaire) de la forme $]-\infty, a_0[$ ou $[a_k, +\infty[$ (**parties** de $\mathbf{R}$) : le calcul de certaines **caractéristiques légales** empiriques à partir de l'histogramme (ie calcul sur données regroupées) nécessite alors des hypothèses particulières, dans la mesure où le regroupement au **centre** des classes extrêmes peut constituer un problème (cf **agrégation**, **groupement de classes**, **correction de groupement**, **loi à queue épaisse**).

Ceci est notamment le cas pour les calculs des **moments**. Néanmoins, d'autres caractéristiques peuvent échapper à ce problème (eg **mode** ou **quantile**).