CENTRALITÉ (C5, F3)
(15 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Notion importante qui peut se relier :

(a) en termes ponctuels, à celles de **paramètre de position** ou de **caractéristique** de centralité ;

(b) en termes ensemblistes, à celles de **région de confiance**, ou de **queue d'une loi** ou d'**aberration**.

Définir la centralité d'une **variable aléatoire** ou de sa **loi de probabilité** équivaut à définir une valeur, ou un ensemble de valeurs, qui caractérise(nt), d'une façon typique, cette variable ou cette loi : la centralité est considérée comme représentative de l'ensemble des valeurs de la va considérée. C'est autour d'elle que se « concentre » la plupart de ces valeurs (cf **valeur centrale**). La notion est cependant distincte de celle de **concentration**.

(i) Soit $(\Omega, \mathscr{T}, P)$ un **espace probabilisé**, $(\mathscr{X}, \mathscr{B})$ un **espace d'observation** dans lequel $\mathscr{X}$ est un **espace topologique** muni de sa **tribu borélienne** $\mathscr{B}$. On considère la **va** $\xi : \Omega \mapsto \mathscr{X}$ dont la **lp** est notée P^{ξ}.

Le concept de **centralité** pour une **partie mesurable** $C \in \mathscr{B}$ combine généralement un ou plusieurs critères parmi les suivants.

(a) un **critère de nature probabiliste** : l'ensemble C ne doit pas être trop « rare », ie ne doit pas porter (ou être chargé par) une faible **probabilité**. On dit que C est une **partie centrale de niveau p** $\in]0,1[$ (avec $p >> 0$) ssi (cf **partie centrale**) :

(1) $\quad P^{\xi}(C) \geq p.$

Il existe donc, en général, de nombreuses parties de niveau p donné ;

(b) un **critère de nature topologique** : l'ensemble C ne doit pas être, globalement, « proche » de la frontière de Im ξ (image de l'**application mesurable** ξ). Si l'on note Supp P^{ξ} le **support de la loi P^{ξ}** (cf **support d'une probabilité**), on doit avoir :

$$C \cap \mathrm{Im}\,\xi \neq \mathrm{Supp}\,P^{\xi},$$
(2)
$$C \cap \partial\,\mathrm{Supp}\,P^{\xi} = \varnothing,$$
$$C \cap \mathrm{Im}\,\xi \subset \mathrm{Supp}\,P^{\xi},$$

ie C appartient à l'image de ξ, son intersection avec elle est incluse dans le support de sa loi, sans être égale à ce support ($C \subset \mathrm{Supp}\,P^{\xi}$), et C n'a pas de point commun avec la **frontière** de ce support ;

(c) un **critère lié à la notion de mesure** : l'ensemble C doit être de **mesure** aussi faible que possible (« **précision** »). Autrement dit, si P^ξ est dominée par une **mesure positive** μ (définie sur $\mathscr{B}$), alors μ (C) << ∞ (ie la mesure de C est faible) (cf aussi **partie floue**).

L'ensemble Supp $P^\xi \setminus C$, ou l'ensemble Im $\xi \setminus C$, est appelé **queue**, ou **extrémité**, de la loi P^ξ ou de la variable aléatoire ξ (cf **queue d'une loi**). On considère parfois que la valeur $x = \xi(\omega)$ est une **aberration** (ou « valeur atypique ») de la variable ξ ssi x appartient à une queue ainsi définie (cf aussi **valeur extrême**).

Si $C' \subset C$, alors C' est une partie centrale de niveau $p' \leq p$.

(ii) Très souvent, une partie centrale C est définie à partir d'un seul point α, aussi appelé **caractéristique de centralité** ou **paramètre de centralité**. La centralité d'une loi peut alors directement se définir à partir de α. En effet, si α est un paramètre de centralité pour ξ, et si Supp $P^\xi \subset$ Im ξ (Supp $P^\xi \neq$ Im ξ), alors toute partie $C \subset$ Supp P^ξ tq $\alpha \in C$ est centrale pour P^ξ (ou pour la va ξ).

Les notions courantes de centralité, définies à partir de parties probabilisables C contenant l'espérance, le mode ou la médiane d'une va, ne satisfont pas toujours à l'ensemble des trois critères (infra). Parfois même, ces critères sont trop restrictifs : ainsi en est-il du cas où le **mode** d'une va se trouve sur la **frontière** du support de sa loi (eg **loi de PARETO**).

Aussi définit-on généralement ce concept pour des classes de lois particulières (cf aussi **forme légale**) :

(a) **lois symétriques** unimodales (cf **loi unimodale**) : toute partie $C \in \mathscr{B}$ (éventuellement symétrique) contenant le mode est admissible comme partie centrale ;

(b) **mélange de lois** : toute partie qui est la réunion de parties centrales des lois composantes peut être centrale (cette réunion n'est alors pas nécessairement connexe).

(iii) Soit $\alpha \in]0,1[$ (α << 1) un nombre donné. Si l'on pose $p = 1 - \alpha$, on est souvent conduit à résoudre le **problème d'optimisation** suivant (qui combine les critères (a) et (c) précédents) :

(3) $\min \mu(C) \quad \text{sous} \quad P^\xi(C) = p,$

où $P^\xi \ll \mu$ (mesure dominante) Cette situation intervient notamment dans le cas d'un **estimateur ensembliste** : on doit choisir une partie centrale relative à une **statistique** donnée (cf aussi **région de confiance la plus précise**).

En **théorie des tests** (ou en théorie des régions de confiance), une **région d'acceptation** $w^c \subset \Omega$ (ou une **région de confiance** S (α)) est généralement une partie centrale permettant de tester l'hypothèse, alors appelée **hypothèse de centralité**, $H_0 : \alpha = \alpha_0$.