NORMALISATION (C2, C7)
(09 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le terme de **normalisation** peut, selon le contexte, comporter quatre significations usuelles.

(i) **Normalité** et **loi normale**. Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé**, $\xi : \Omega \mapsto$ **R** une **vars** de **loi** P^{ξ} et X un **échantillon** iid selon P^{ξ} (cf **échantillon iid**).

On appelle **normalisation**, ou **transformation normalisante**, toute fonction ϕ : **R** $\mapsto$ **R** tq que la loi de la **va** $\phi(\xi)$ soit normale. Autrement dit, il existe un couple $(\mu, \sigma^2) \in$ **R** x **R***$_+$ tq :

(1) $\quad P^{\phi \circ \xi} = \mathcal{N}_1(\mu, \sigma^2)$.

Souvent, en pratique, $\mu = 0$ et $\sigma^2 = 1$.

L'extension au cas multidimensionnel $\xi : \Omega \mapsto$ **R**^{K} est directe.

En général, ϕ dépend d'un nombre (fini) de paramètres. Ainsi, on choisit parfois une **transformation des données** tq l'une des suivantes :

(2) $\quad \phi(\xi) = \alpha + \beta \,.\, g\{a(\xi + b)\}$,

ou encore :

(2) $\quad \phi(\xi) = \alpha + \beta \,.\, g\{(\xi - \mu) / \lambda\}$,

dans lesquelles g : I $\mapsto$ **R** est une **application bijective** d'un intervalle I $\in \mathcal{T}$(**R**) dans **R**, $\beta > 0$, $a > 0$ et $\lambda > 0$.

A titre d'exemple, si g = Log, b = 0, a = 1, I = **R**$_+$* et si ξ est une variable qui suit une **loi Log-normale** de **paramètres** (α, β), alors $\phi(\xi)$ est une variable normale de paramètres (μ, λ).

(ii) **Changement de variable aléatoire**. Le terme de **normalisation** est aussi utilisé pour désigner la transformation d'une va $\xi \in L_{\mathbf{R}}^2(\Omega, \mathcal{T}, P)$ par **centrage** (pr à sa **moyenne**) et « **réduction** » (pr à son **écart-type**), ie (cf **variable centrée**, **variable réduite**, **variable normée**) :

(3) $\quad \xi \mapsto \upsilon = (\xi - E\,\xi) / \sigma(\xi)$,

ou, plus généralement, par centrage pr à un **paramètre de position** $\alpha \in$ **R** et réduction pr à un **paramètre d'échelle** $\beta \in$ **R**$_+$*) :

(4) $\xi \mapsto \upsilon = (\xi - \alpha) / \beta$.

On peut alors définir une notion empirique analogue à celle exprimée par (3), ie (cf **statistique naturelle**) :

(5) $X_n \mapsto U_n = (X_n - \bar{X}_N) / S_N$,

où $\bar{X}_N = e_N' X / e_N' e_N$ (**moyenne empirique**) et $S_N^2 = N^{-1} X' P X$ (**dispersion** empirique), P est la **matrice de centrage par rapport à la moyenne** (empirique) et X désigne l'**échantillon** disposé en vecteur colonne.

Dans ce sens, la va η, ou sa **copie** T_n , est appelée **variable normalisée**, ou **variable normée**, ou **variable centrée-réduite**, ou encore **variable standardisée**.

Une telle variable est de **dimension zéro** pr aux unités de mesures de ξ (ou de U_n). Ceci explique notamment son intérêt pratique : ainsi, l'**analyse en composantes principales** normées est effectuée sur une **matrice d'observation** relative à des variables centrées-réduites.

(iii) **Contrainte** (propriété spécifique) imposée à une fonction. Une **normalisation** consiste à « modifier » cette fonction, le plus souvent à l'aide d'une **constante** (ou parfois d'une autre fonction), afin de respecter une **condition de régularité**.

Ainsi, dans l'étude d'une **loi** à **densité**, une fonction g, non négative et intégrable, mais ne « sommant » pas à l'unité peut être transformée μ-p.p. en densité de probabilité : elle permet notamment de définir une **loi de probabilité** P^{ξ} (ou une famille de telles lois). En effet, il est possible de déduire de g une fonction f satisfaisant la propriété de sommation unitaire comme suit :

(a) calcul de la « **constante** » de **normalisation** :

(6) $c(g) = \int g(x)\, d\mu(x) \neq 1, \qquad \mu\text{-p.p.},$

où μ est une **mesure positive** dominant g ;

(b) calcul du rapport :

(7) $f = g / c(g), \qquad \mu\text{-p.p.}.$

La classe des fonctions f ainsi définie est donc constituée de densités de probabilité.

Si la fonction initiale g dépend d'un paramètre, la constante c (g) en dépend aussi généralement.

(iv) **Règle de normalisation.** La notion de normalisation intervient enfin dans l'étude du **modèle d'interdépendance linéaire**.