QUANTILE (C5, F3)
(31 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **quantile** est une **caractéristique de position** associée à une **loi de probabilité**. L'ensemble des quantiles d'une loi fournit une gamme étendue de **caractéristiques** de cette loi (**caractéristiques de forme**, notamment). Par ailleurs, cette notion est à l'origine de certaines méthodes d'estimation (cf eg **régression quantilaire**).

(i) Soit $(\Omega, \mathcal{T}, P)$ un **espace probabilisé**, $\xi : \Omega \mapsto \mathbf{R}$ une **vars**, P^{ξ} sa **lp**, F sa **fr** et $f = dP^{\xi} / d\mu$ la **densité** de P^{ξ} pr à une mesure positive μ.

On appelle **quantile**, ou **fractile (A.HALD)**, ou encore **p-quantile** ou **p-fractile**, **théorique d'ordre p** $\in]0,1[$ de P^{ξ} (ou de ξ, ou de F, ou encore de f) tout nombre réel, noté $Q_p\,\xi \in \mathbf{R}$, tq :

(1) $\quad P([\xi \leq Q_p\,\xi]) \;\geq\; p \quad$ et $\quad P([\xi \geq Q_p\,\xi]) \;\geq\; 1 - p.$

Lorqu'il ne se réduit pas à un seul point de **R**, l'ensemble $\mathcal{Q}_p$ des valeurs $Q_p\,\xi \in \mathbf{R}$ vérifiant (1) est appelé **ensemble des quantiles**, ou parfois **intervalle quantilaire**, d'ordre p de dP^{ξ}.

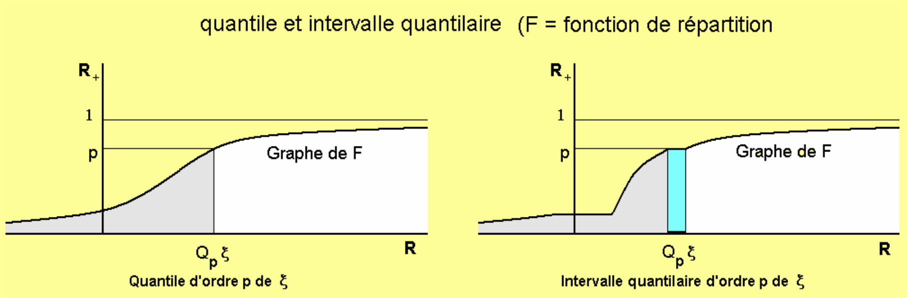
On peut exprimer (1) à l'aide de F selon :

(2) $\quad F(Q_p\,\xi) \;\geq\; p \quad$ et $\quad 1 - F(Q_p\,\xi) \;\geq\; 1 - p.$

Si F est continue et strictement croissante, le p-quantile est unique (ie $\mathcal{Q}_p = \{Q_p\,\xi\}$). Dans ce cas, les conditions (2) sont remplacées par l'équation implicite de définition (cf aussi **fonction quantile**) :

(3) $\quad F(Q_p\,\xi) \;=\; p \in]0,1[.$

Représentations d'un quantile et d'un intervalle quantilaire



(ii) D'un point de vue terminologique, $Q_{1/2}\,\xi$ est appelé **médiane**, $Q_{1/4}\,\xi$ **quartile (inférieur)**, $Q_{3/4}\,\xi$ **quartile (supérieur)**, $Q_{1/5}\,\xi$ **quintile (inférieur)**, $Q_{1/10}\,\xi$ **décile (inférieur)**, etc.

Plus systématiquement, en notant $d = 10^{-n}$ (n donné), $Q_d\ \xi$ est appelé **10^{-n}-quantile**, **quantile d'ordre 10^{-n}** ou encore **10^{n}-ile** (**centile** si n = 2, **millile** si n = 3, etc) (**inférieur**). Par suite, on peut partitionner **R** en n - 1 classes à partir des quantiles d'ordre 10^{-n}.

(iii) On établit les importantes propriétés de monotonie suivantes :

(a) si $f : \mathbf{R} \mapsto \mathbf{R}$ est monotone croissante et si $p \in]0, 1[$, alors $Q_p\ f(\xi) = f(Q_p\ \xi)$, ou bien $Q_p\ \xi = f^{-1}(Q_p\ f(\xi))$;

(b) si $g : \mathbf{R} \mapsto \mathbf{R}$ est décroissante, alors $Q_p\ g(\xi) = g(Q_{1-p}\ \xi)$, ou encore $Q_{1-p}\ \xi = g^{-1}(Q_p\ g(\xi))$.

(iv) De même que l'on associe à certaines classes de va leur **espérance mathématique** (ce qui définit une **application linéaire**), de même on peut associer à certaines classes de va leur p-quantile (ce qui définit une application plus complexe, notée aussi Q_p).

Si l'on définit une **inverse de F** selon (**fonction quantile**) :

(4) $\quad F^{-1}(y) = \inf \{x \in \mathbf{R} : F(x) \geq y\}$,

on appelle **quantile théorique d'ordre p** toute valeur $Q_p\ \xi$ tq $F^{-1}(Q_p\ \xi) = p$. Pour en assurer l'unicité, on peut donc poser :

(5) $\quad Q_p\ \xi = \inf \{x \in \mathbf{R} : F(x) \geq p\}$.

(v) Si une vars ξ est de classe L^2, de **moyenne** μ et de **variance** σ^2, on établit, $\forall\ p \in]0,1[$, les **inégalités de H. HOTELLING - L.M. SOLOMONS** :

(6) $\quad |Q_p\ \xi - \mu| \leq \sigma \,.\, \max \{(q / p)^{1/2}, (p / q)^{1/2}\}, \qquad$ avec $q = 1 - p$.

En particulier (avec p = 1/2), la médiane vérifie :

(7) $\quad (Q_{1/2}\ \xi - \mu)^2 \leq \sigma^2$.

(vi) Un estimateur « naturel » de $Q_p\ \xi$ (supposé unique) est le **quantile empirique**, calculé à partir de la **fr empirique** F_N fondée sur un N-**échantillon** $X = (X_1, ..., X_N)$ (cf **statistique naturelle**). Ce quantile peut s'écrire :

(8) $\quad q_N\ X$ (ou $s_N\ \xi$) $= (1 - \alpha) \,.\, X^{([(N+1)p])} + \alpha \,.\, X^{([(N+1)p] + 1)}$,

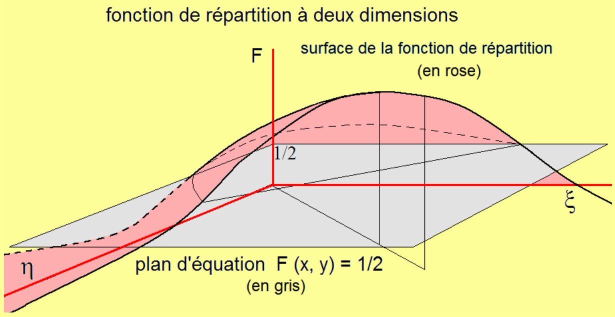
expression dans la quelle $(X^{(1)}, ..., X^{(N)}) = X^{(.)}$ est l'échantillon ordonné (cf **statistique d'ordre**), $p \in]0, 1[$, $(N+1)\,p = [(N+1)\,p] + \alpha$ et [.] désigne la fonction **partie entière**.

(vii) Un **quantile** peut, plus généralement, être défini pour une va $\xi : \Omega \mapsto \mathscr{X}$ à valeurs dans un ensemble ordonné $(\mathscr{X}, \leq)$ (cf **relation d'ordre**). En effet, dans ce cas, $(\mathscr{X}, \mathscr{O})$ est un **espace topologique** (pour la **topologie** de l'ordre $\mathscr{O}$) et $(\mathscr{X}, \mathscr{B})$

est un **espace mesurable** pour la **tribu borélienne** $\mathscr{B} = \sigma(\mathscr{O})$. Autrement dit, ξ est supposée $(\mathscr{T}, \mathscr{B})$-mesurable.

C'est ainsi le cas pour une **variable ordinale**, ie une **variable qualitative** κ à valeurs dans un ensemble (pré)ordonné $(\mathscr{K}, \leq)$, ou encore pour une qualitative « valuée ».

(viii) La notion de quantile peut s'étendre au cas d'un **vecteur aléatoire** $\xi : \Omega \mapsto \mathbf{R}^K$, mais elle ne possède pas, sauf restrictions supplémentaires, de propriété d'unicité (cf schéma ci-après, dans le cas d'une médiane multi).



Si F est la fr de ξ, on appelle **partie iso-quantilaire**, ou **partie p-quantilaire, muldidimensionnelle** de $Q_p\, \xi$ la **partie** $\mathscr{Q}_p{}^* \subset \mathbf{R}^K$ définie par l'image inverse :

(9) $\mathscr{Q}_p{}^* = \{Q_p\, \xi \in \mathbf{R}^K : F\,(Q_p\, \xi) = p\} = \{Q_p\, \xi \in \mathbf{R}^K : Q_p\, \xi = F^{-1}\,(p)\}$,

où l'on note $F^{-1}\,(p) = \inf \{x \in \mathbf{R}^K : F\,(x) \geq p\}$ la fonction quantile.

Si F est continue, on obtient :

(10) $\mathscr{Q}_p{}^* = \{Q_p\, \xi \in \mathbf{R}^K : Q_p\, \xi = F^{-1}\,(p)\} \subset \mathbf{R}^K$,

où $F^{-1}\,(p) = \{x \in \mathbf{R}^K : F\,(x) = p\}$.

En pratique, en notant F_k la **fonction de répartition marginale** d'indice $k \in N_K{}^*$ (ie la fr de la va $\xi_k = pr_k\, \xi$), on définit parfois simplement $\mathscr{Q}_p\, \xi$ comme le vecteur $(Q_p\, \xi_1, ..., Q_p\, \xi_K)$ de $\mathbf{R}^K$ dont les coordonnées sont les p-quantiles resp des marginales F_k . Dans ce cas, $Q_p\, \xi$ est unique (ie Card $\mathscr{Q}_p{}^* = 1$) ssi toutes les marginales admettent un quantile unique (et non « ensembliste »).

(ix) On montre enfin que le quantile (ordinaire) est le scalaire qui minimise l'**écart absolu moyen** $E\,|\xi - a|$ pr à a (propriété comparable à celle de l'espérance pr à la variance).

On fonde parfois la définition de **quantile vectoriel** sur une propriété analogue dans $\mathbf{R}^K$.