

COMPLEXITÉ (A1, A15, B4, C8, C11, G2, G11, K, N8)

(06 / 12 / 2020, © Monfort, Dicostat2005, 2005-2020)

La notion de **complexité** fait souvent référence aux notions courantes de dénombrement, de **structure**, de **variabilité**, etc. Initialement associée à certains **domaines de connaissance**, elle a été ensuite précisée et étudiée selon une analyse mathématique.

(i) La notion de **complexité** est une notion double :

(a) la **complexité descriptive** vise à comparer des « entités » du point de vue de leur **structure** interne (A.A. MARKOV, N. KOLMOGOROV, L. LEVIN, etc).

Ainsi, on peut admettre que :

(a₁) à cardinalité identique, la suite {a, b, c, d, e} (de cardinalité 5) est plus « complexe » que la suite **constante** {c, c, c, c, c}, dès que deux éléments de cette suite sont distincts ;

(a₂) la suite {a, a, a, a} peut paraître plus simple que la suite {c, d} dès que $d \neq c$, bien qu'étant de cardinalité supérieure ;

(a₃) la suite $u = (u_n)_n$ définie par l'**équation de récurrence** $u_n = f_n(u_{n-1})$ ($n \in \mathbf{N}^*$, u_0 donné) peut sembler moins simple qu'une suite $v = (v_n)_n$ définie par $v_n = g_n(v_{n-1})$ ($n \in \mathbf{N}^*$, v_0 donné) si, pour au moins un indice n , la forme analytique de f_n est plus « élaborée » que celle de g_n (eg quadratique au lieu de linéaire) ;

(b) la **complexité calculatoire** vise à évaluer l'« **espace - temps** » nécessaire à un « algorithme » (notion admise ici) afin de « construire » des entités tq les précédentes. Ces moyens (eg nombre minimal de « mots » ou de « phrases » nécessaires) peuvent, en effet, impliquer des coûts plus ou moins importants (cf **fonction de coût**). De plus, la construction n'est pas toujours possible, comme le montrent la théorie de la calculabilité et l'algorithmique.

Ainsi, un algorithme (évident) consistant à calculer une suite tq la suite $u = (u_n)_n$ précédente est de même complexité que celui destiné à calculer la suite $(v_n)_n$ dès que les f_n et les g_n sont identiques ($\forall n \in \mathbf{N}^*$), quelles que soient les valeurs initiales ; par contre, il sera de complexité supérieure si les f_n sont quadratiques (trois paramètres) et les g_n linéaires-affines (deux paramètres).

La notion de complexité est donc un concept relatif. De plus, on doit distinguer entre le concept et sa **mesure**, généralement effectuée à l'aide d'un **indice de complexité**. On se limite ci-après à la notion descriptive.

(ii) De façon générale, au sein d'un **domaine de connaissance**, un **phénomène** donné peut se représenter à l'aide d'un système, lequel peut, à son tour, se décrire à l'aide d'un **processus stochastique**.

Un **système** A est constitué essentiellement :

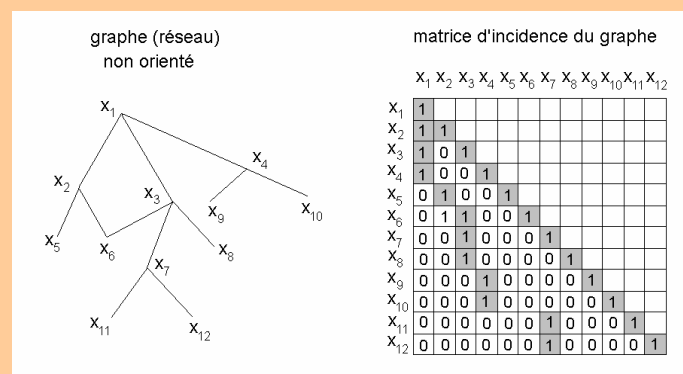
(a) d'un **ensemble d'éléments** $A = \{a_1, \dots, a_N\}$, souvent appelés « **unités** », ou « **composants** », du système. Dans un **système « fermé »**, A est fixe au cours du temps (eg système physique : matériel mobile, système planétaire, etc) ; dans un **système « ouvert »**, A peut varier (entrées / sorties d'unités) (eg système écologique : démographie d'une espèce végétale ou animale ; système biologique : échanges intercellulaires, mouvements toxiques, bactériologiques, viraux) ;

(b) d'un **ensemble de relations**. En effet, les composants sont généralement reliés entre eux par diverses relations qui possèdent des « fonctionnalités ». Dans la situation la plus « simple », il existe au plus deux fonctionnalités $\phi_{\alpha\beta} : a_\alpha \leftrightarrow a_\beta$ reliant deux éléments (une fonctionnalité dans chaque sens : eg **signal** - réponse, ou **action** - résultat, etc). On note alors $\Psi = (\phi_{\alpha\beta})_{(\alpha,\beta)}$ l'ensemble de ces relations et $K = \text{Card } \Psi$ leur nombre.

Dans certains cas, les fonctionnalités peuvent relier, non pas les unités elles-mêmes, mais des « mesures » (ou **observations**) $x_n = \xi(a_n)$ ou $y_n = \eta(a_n)$ effectuées sur ces unités.

Le nombre $N = \text{Card } A \in \mathbf{N}^*$ est appelé la **dimension du système** et K le **nombre de liens**.

Une façon commode de représenter A consiste à lui associer un **graphe** (orienté), constitué de sommets (les composants) et d'arcs (les fonctionnalités). Par suite, à chaque système peut être associée une (N,N)-**matrice d'incidence** $M \in M_N(\mathbf{R})$ (cf schéma ci-dessous) qui décrit, à l'aide de 1 et de 0, la structure en « **réseau** » de A ainsi que ses propriétés (eg fonctions dynamiques) : ie $M = (m_{\alpha\beta})_{(\alpha,\beta)} \in M_N(\{0, 1\})$.



S'il existe au moins deux relations entre unités, on peut « décomposer » les éléments a_n de A en autant d'éléments distincts et continuer l'analyse comme suit.

(iii) La définition de la complexité, se base d'abord sur le quadruplet (A, N, Ψ , K) ou, alternativement, sur la structure M des incidences. Alors :

(a) si le couple (A, N) est donné (ie fixe), on considère qu'un système (A, N, Ψ_2, K_2) est plus complexe qu'un système (A, N, Ψ_1, K_1) ssi la cardinalité de Ψ_2 est supérieure à celle de Ψ_1 , ie $K_2 = \text{Card } \Psi_2 > \text{Card } \Psi_1 = K_1$. En effet, dans ce cas, c'est la densité des liens internes qui discrimine les deux systèmes du point de vue de leur complexité.

En termes de graphes et de matrices d'incidence associées, les sommets et la dimension sont ici donnés, mais leur « remplissage » diffère (ie M_2 est plus remplie que M_1). Autrement dit, si $K_i = \sum_{\alpha=1}^N \sum_{\beta=1}^N m_{\alpha\beta}^i$ désigne le nombre de 1 dans $M_i = (m_{\alpha\beta}^i)_{(\alpha,\beta)}$ ($i = 1, 2$), on a $K_2 > K_1$;

(b) l'analyse est moins simple lorsque (A, N) peut varier, car il s'agit alors de comparer des matrices d'incidence différentes (dimensions, liens). La mesure de la complexité, ie la définition d'un **indice de complexité**, doit alors tenir compte à la fois des matrices M_i , de leurs dimensions N_i et du réseau constitué des liens entre les unités.

Un indice élémentaire doit donc dépendre de **rapports** tq :

$$(1) \quad R_i = K_i / N_i^2,$$

avec $K_i = \sum_{\alpha=1}^{N(i)} \sum_{\beta=1}^{N(i)} m_{\alpha\beta}^i$, où $N(i)$ désigne N_i .

(iv) En **théorie de la complexité**, on considère souvent un « système contrôlé », ou « système piloté », $S(X, U)$ dont les variables de **contrôle** (ou variables de commande) U doivent conduire le système d'un **état** initial $X_{t(0)}$ vers un état final (souhaitable) $X_{t(1)}$ (où $t(i)$ désigne t_i , $i = 1, 2$), sachant que le système est astreint à respecter une équation d'état tq $f(X, u) = 0$. Soit U_{01}^* une solution du **problème d'optimisation** ainsi défini.

La théorie suit alors généralement les étapes suivantes :

(a) définition de la complexité $C_t(S)$ de S à une date quelconque $t \in [t_0, t_1]$;

(b) étude des variations de l'optimum U_{01}^* lorsque $C_t(S)$ varie ;

(c) étude des coûts induits par les variations de $C_t(S)$: eg variation « extrême » de U_{01}^* entre t_0 et t_1 lorsque $C_t(S)$ varie entre $C_{t(0)}(S)$ et $C_{t(1)}(S)$.

(v) La complexité est ainsi une notion qui peut combiner plusieurs concepts :

(a) ceux de la logique mathématique, de la **théorie des graphes** ou de l'**analyse des données** ;

(b) les concepts de **probabilité** (dont celui de **probabilité géométrique**) et de **mélange légal** ;

(c) ceux des méthodes de **modélisation** et de **représentation statistique**, ce qui conduit aux notions importantes d'**information** (cf **information statistique**) et de **causalité** ;

(d) le principe de **principe de parcimonie**, ainsi que les méthodes de réduction d'un **problème statistique** : cf **parcimonie**, **principe de réduction**, **réduction des données** ;

(e) l'**inférence statistique** multivariée (eg multidimensionnelle), dont la complexité augmente avec le nombre de variables ou les dimensions considérées : cf **analyse multidimensionnelle**, **loi multidimensionnelle**, **tableau statistique multidimensionnel**, **variable multidimensionnelle** ;

(f) ceux de la théorie des systèmes et de la **théorie des processus** : cf **système aléatoire**, **processus multidimensionnel**.