PROBLÈME DE FILE D'ATTENTE (B2, B3, B5, N11)
(01 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) Dans divers **domaines de connaissance**, certaines **unités** concernées par un **phénomène** peuvent manifester des tendances à l'accumulation ou au regroupement, dans l'**espace** aussi bien que dans le **temps** : occupation de l'espace avec agglutination, circulation contrainte, alignement croissant, etc. On parle alors généralement d'**encombrement**, d'**engorgement**, d'**embouteillage**, etc.

En effet, ces regroupements augmentent les contraintes qui peuvent s'exercer « spontanément » sur chaque unité : limitations du mouvement, génération de délais.

Un **problème de temps d'attente** peut ainsi intervenir dans des contexte tq les suivants :

(a) physique : **écoulement de fluide** (dans un milieu, entre milieux), ou encore (technologie) circulation d'information dans un **réseau d'échanges** (téléphoniques, informatiques, traitements dans un micro-processeur, etc). ;

(b) biologie : **circuits fermés** (eg réseau sanguin, système lymphatique, réseau neuronal) ;

(c) écologie : **accès à un nutriment** (point d'eau, dépouille animale) lorsqu'il existe une préséance entre animaux ;

(d) sociologie : **queue** pour l'accès à un service (caisse d'une entreprise commerciale, entrées-sorties d'une autoroute à péage, guichet d'institution : administration, banque).

(ii) Un **problème de file d'attente**, parfois appelé **problème de congestion**, comporte les données de base suivantes (cf aussi **temps d'attente**) :

(a) un ensemble de **serveurs**, ou **guichets**, généralement fini. Chaque serveur peut rendre un **service** pendant une certaine durée de **temps** : on suppose que cette durée est une **vars** positive, appelée **temps d'attente** ou **temps de service** (cf aussi **temps de séjour**, **temps d'occupation**) ;

(b) un ensemble de **clients**, a priori non fini, le plus souvent « discret » (ie dénombrable). Chaque client arrive à un instant donné, appelé **temps d'arrivée** (cf aussi **temps d'entrée**) : cet instant peut s'analyser comme une **variable discrète** positive, dont la loi est en général commune à tous les clients (eg il n'existe pas de « groupe » de clients, ou alors celui-ci ne demande qu'un service identique à celui demandé par un client représentatif). Par ailleurs, chaque client peut se rendre au **hasard** à l'un des guichets.

(iii) Chaque guichet est ainsi caractérisé par trois processus :

(a) un **processus d'arrivée** des clients, ou **processus d'entrée** ;

(b) un **processus de temps de service**, ou **processus de durée de service** ;

(c) un **processus de temps d'attente**, qui résulte des précédents.

Il s'ensuit un **processus de départ**, ou **processus de sortie**.

Les deux processus (a) et (b) dépendent du guichet considéré (cf **hétérogénéité**) : en effet, les serveurs peuvent avoir des aptitudes différentes, et le « service » offert (ie le problème à traiter) d'un guichet à l'autre peut être de difficulté ou de nature variables, etc.

Le principal objectif consiste à éviter une **congestion**, ie que les temps d'attente soit les plus courts : faible probabilité de dépasser des valeurs jugées « intolérables » par les clients (aspects psychologiques).

(iv) Le schéma de base précédent peut être amélioré pour tenir comptes de diverses particularités :

(a) certains guichets peuvent rendre le même service. Par suite, la **modélisation** peut aussi porter sur les possibilités de « transferts » des clients entre les files de ces guichets, après comparaison de leur longueur (transferts depuis les files « longues » vers les files « courtes ») ;

(b) certains services peuvent nécessiter un temps « extra-ordinaire » (service spécialisé, etc). On doit donc déterminer la loi du processus qui génère ainsi une **valeur extrême** (cf **aberration**).

Un problème de temps d'attente met donc en jeu les outils de l'**analyse combinatoire**, du **calcul des probabilités** et de la **théorie des processus**.

(v) Un problème de file d'attente est parfois associé à un **problème d'occupation** et analysé à l'aide d'un **schéma d'urne**. Ainsi :

(a) on considère H **urnes** et l'on affecte un à un les **éléments** d'un **ensemble** donné E, jusqu'à ce que l'un d'eux aille dans une urne déjà occupée une première fois. Alors :

$(a)_1$ la probabilité p_n pour que cette **procédure d'affectation** s'arrête au n-ième instant est :

(1) $p_n = (n - 1)\, H^{n-1} / H^n = (n - 1)\, H^{-1} . \Sigma_{i=1}^{n-2} (1 - (i / H)), \quad \forall\, n \geq 3,$

avec $p_1 = 0$ et $p_2 = H^{-1}$;

$(a)_2$ de même, la probabilité q_n pour qu'elle dure plus de n instants est :

(2) $q_n = (H)_n / H^n = \Sigma_{i=1}^{n-1} (1 - (i / H)), \quad \forall\, n \geq 2,$

avec $q_1 = 1$, où $(x)_j = x\, (x - 1) \ldots (x - j + 1)$ désigne le **polynôme factoriel** d'ordre j ;

(b) on considère une urne particulière $h \in N_H^*$ et l'on affecte au hasard (uniforme) les éléments de E dans l'ensemble des cases tant que la case h considérée demeure vide. Alors :

$(b)_1$ la probabilité p_n pour que ce processus d'affectation s'arrête à l'instant n est :

$$(3) \qquad p_n = H^{-1} (H - 1) / H^{n-1}, \qquad \forall n \in \mathbf{N}^* ;$$

$(b)_2$ de même, la probabilité q_n pour qu'il dure plus de n instants est :

$$(4) \qquad q_n = (1 - H^{-1})^n, \qquad \forall n \in \mathbf{N}^*.$$