L'Actualité économique

# Files d'attente

Claude Tricot

Volume 38, Number 4, January–March 1963

URI: https://id.erudit.org/iderudit/1001823ar
DOI: https://doi.org/10.7202/1001823ar

See table of contents

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)
1710-3991 (digital)

Explore this journal

Cite this article

Tricot, C. (1963). Files d'attente. *L'Actualité économique*, *38*(4), 626–642. https://doi.org/10.7202/1001823ar

# Analyses

## Recherche opérationnelle - 2

### *Files d'attente*

La théorie des files d'attente existait avant que ne fût imaginé, sous le titre «recherche opérationnelle», un groupement des différentes études faites par des mathématiciens, de problèmes posés par l'économie ou la gestion des entreprises.

C'est, en effet, en 1917 qu'Erlang publia pour la première fois ses études concernant l'encombrement des lignes téléphoniques.

Les résultats fournis par la théorie ont des applications dans des domaines très divers; cependant son origine est encore marquée, dans le vocabulaire, chez beaucoup d'auteurs: pour ceux-ci, un poste de service est une «ligne», l'arrivée d'un élément à ce poste est un «appel» et l'occupation d'un poste de service est une «conversation».

On sait, cependant, que *the theory of queuing*, comme disent certains, peut fournir des renseignements utiles chaque fois qu'un «élément» se présente à un «poste» pour recevoir un «service».

Citons quelques exemples bien connus ; nous terminerons l'article par l'étude d'une application plus nouvelle:

1) les pannes de machines: l'élément est la machine qui tombe en panne, le poste est le mécanicien, le service est la réparation;

2) les arrivées de véhicules à un poste de chargement (arrivées de bateaux dans un port, d'avions au-dessus d'un aérodrome, de bétonnières se faisant charger en ciment, etc.);

3) les arrivées de clients dans un magasin;
4) le règlement de la circulation au moyen des feux rouges et verts, etc.

Ces problèmes se rattachent à la recherche opérationnelle en ce sens qu'un optimum économique peut y être cherché. Par exemple, un compromis entre les différents coûts dus aux attentes des clients face à des postes de services trop peu nombreux, et les attentes des préposés aux services en surnombre.

Il est difficile, à moins de tout lire, de naviguer sur l'océan de la littérature engendrée par les files d'attente. On y rencontre l'amoureux des mathématiques pures pour lequel un tel problème est l'occasion de se mesurer avec les équations intégrales, et l'amateur de petits cas dont l'imagination construit des phénomènes jamais vus dans le but de varier les hypothèses; on y admire aussi des solutions numériques ou graphiques qui suppléent à des solutions théoriques incomplètes.

La théorie est ancienne, mais les ouvrages généraux relativement récents. Avant d'aborder les revues nous commencerons par une comparaison entre deux ouvrages généraux écrits de deux points de vue différents, celui du mathématicien pur et celui du mathématicien appliqué: *Mathematical Methods in the theory of queuing* traduction anglaise de A.-Y. Khintchine[1], et *Initiation aux processus aléatoires* de M. Girault[2]; ce dernier est publié dans une collection destinée aux ingénieurs mais rédigée par des universitaires.

Dans chacun de ces ouvrages, la première partie est consacrée à l'étude des «arrivées» (arrivées de clients, de bateaux, appels téléphoniques) et les hypothèses sont dès le début assez restrictives: les mobiles n'ont pas le droit d'arriver n'importe comment, et l'on finira par comprendre que l'élégance des solutions est causée par ces hypothèses simplificatrices, ce qui ne veut pas dire que la construction est gratuite.

Quelles sont ces hypothèses ? Tout d'abord, l'hypothèse de stationnarité qui est énoncée par Khintchine d'une façon élémentaire au premier chapitre et plus générale au deuxième: on suppose que la probabilité de $k$ arrivées dans un intervalle de temps de

1. Hafner Publishing, N.Y., 1960.
2. Dunod, Paris, 1959.

durée $t$ est indépendante de l'époque; en d'autres termes, cette probabilité est la même entre 8 et 9 heures qu'entre 11 et 12, et, par conséquent, on ne considère que des phénomènes homogènes dans le temps. Une question se pose alors, que M. Girault pose effectivement à un autre propos: pour qu'une telle hypothèse soit acceptable, il faudra souvent considérer de courts intervalles de temps. Or nous verrons que les lois découvertes ne sont applicables qu'au bout d'un temps assez long de fonctionnement du système; il y aura donc une critique à faire quant à l'application de ces lois.

Une deuxième hypothèse est la suivante: le phénomène est «*without after effects*» ce qui signifie que la probabilité de $k$ arrivées durant la période $(a, a + t)$ est indépendante de toute supposition faite sur ce qui s'est passé avant le début de cette période (en ce qui concerne les arrivées, naturellement).

Enfin, une troisième hypothèse est mentionnée par Khintchine et sous-entendue par M. Girault, c'est l'hypothèse de la régularité: la probabilité de deux arrivées ou plus dans une période de durée $t$ est un infiniment petit avec $t$, d'ordre supérieur à $t$. La suppression de cette hypothèse entraînerait, comme le montre Khintchine au chapitre 2, que la probabilité de plus d'une arrivée, à un instant où il s'en produit une, n'est pas nulle: une arrivée peut en entraîner d'autres, c'est un processus en «grappe».

Ces trois hypothèses permettent aux deux auteurs d'établir que la loi de probabilité du nombre d'arrivées dans un temps donné est une loi de Poisson:

$$P\,(K = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

Dans sa démonstration fort simple, M. Girault établit tout d'abord la loi de probabilité du temps qui sépare deux arrivées consécutives. Le raisonnement est le suivant: la première hypothèse implique que la probabilité qu'aucune arrivée n'ait lieu entre les instants $t_0$ et $t_0 + t$ ne dépende que de $t$: c'est une fonction de $t$ : $g(t)$.

Entre les instants $t_0$ et $t_0 + t + t'$, ce sera donc $g(t + t')$, mais l'hypothèse 2 permet d'affirmer que cette probabilité est aussi $g(t) \times g(t')$.

On a donc:

$$g\,(t + t') = g(t) \times g(t')$$

ce qui est une relation caractéristique de la fonction exponentielle. Conclusion: $g(t) = e^{-ct}$, $c$ étant positif puisqu'une probabilité est un nombre inférieur à 1.

Si $t_i$ et $t_{i+1}$ sont les instants de deux arrivées consécutives, il y a équivalence entre les deux affirmations:

a) $t_{i+1} - t_i < t$

b) au cours de la durée $t$, une arrivée au moins se produit.

Cette dernière probabilité étant complémentaire de $g(t)$, on a donc:

$$P\{t_{i+1} - t_i < t\} = 1 - e^{-ct},$$

ce qui est la fonction de répartition du temps qui sépare deux arrivées consécutives: c'est celle d'une loi exponentielle.

On en déduira la loi de probabilité du temps qui sépare une arrivée de la $k^{\text{ième}}$ qui la suit. Elle aura pour fonction de densité:

$$c\,\frac{e^{-ct}\,(ct)^{k-1}}{(k-1)!}$$

d'où la loi de probabilité du nombre d'arrivées dans un temps donné.

Il convient de remarquer que ce procédé conduit à considérer non pas *des arrivées d'éléments* dans le système, mais *les époques où se produit un événement*. Implicitement, on n'exclut donc pas que cet événement puisse être une arrivée en grappe. La loi trouvée est donc moins celle du nombre d'arrivées (de clients à un poste par exemple), que celle du nombre d'instants où se produisent des arrivées; ceci explique pourquoi cette démonstration ne fait pas appel à l'hypothèse de la régularité, qui est nécessaire à Khintchine, lequel compte les éléments.

À ce point se marque la différence de point de vue des deux ouvrages: M. Girault, écrivant son livre en vue des applications, consacre le chapitré 4, tout d'abord, à donner les considérations à priori qui permettent au bon sens d'admettre les hypothèses précédentes dans de nombreux cas, en précisant qu'elles ne valent que pour des périodes de temps réduites, puis, donnant deux exemples numériques, un cas d'arrivées de bateaux dans un port et un cas d'arrivées de camions à un poste de chargement, il teste à posteriori la validité de ces hypothèses. Ainsi, le livre descend-t-il du nuage des symboles pour toucher le terrain prosaïque des applications.

Toute autre est la manœuvre de Khintchine; son étude des arrivées (des «appels», suivant sa terminologie) occupe la moitié du livre. Il s'agira d'éprouver les trois hypothèses précédentes en étudiant l'effet que produit sur le flux des appels la suppression de l'une d'elles, pour terminer par cette constatation que, dans la pratique, la loi étudiée tout d'abord est celle que l'on rencontre le plus fréquemment.

Cependant, ce jeu avec les hypothèses n'est nullement stérile: il permet au contraire de se faire une juste idée de l'importance de chacune, résultat que l'on n'obtient pas par la lecture du seul livre de M. Girault.

Par exemple, A.-Y. Khintchine montre que la seule hypothèse concernant la stationnarité permet d'établir que, $W(t)$ étant la probabilité d'au moins une arrivée au cours de la durée $t$, le rapport $\frac{W(t)}{t}$ tend vers une limite. Si, de plus, on ajoute l'hypothèse 2, cette limite est finie, de sorte que l'on peut écrire:

$$W(t) = \lambda t + o(t),$$

ce qui est pris à tort comme hypothèse du processus simple (arrivées poissonniennes) dans bien des ouvrages (en particulier dans le livre très répandu de Kaufmann publié chez Dunod, lequel, en revanche, utilise la deuxième hypothèse sans la donner).

De même, la suppression de la troisième hypothèse (régularité) conduit, comme nous l'avons dit, au processus en grappe; or, il n'est pas difficile d'imaginer des cas pratiques de processus en grappe (par exemple, l'arrivée d'un camion sur une petite route détermine l'arrivée des voitures qui n'ont pu le dépasser).

Notons enfin que, dans la démonstration de M. Girault dont nous avons parlé, les hypothèses de départ servent à montrer l'indépendance des intervalles de temps successifs séparant deux arrivées consécutives sur quoi est fondée la démonstration. On peut se demander si la réciproque n'est pas vraie: un flux d'appels stationnaire, régulier, tel que les intervalles de temps dont nous avons parlé soient indépendants, est-il poissonnien ? Khintchine montre qu'il n'en est rien et donne à ce propos un apparent paradoxe qui est le suivant. Il s'agit du problème de Palm: les appels (ou arrivées) sont poissonniens; ils se présentent à une suite de lignes ordonnées: $L_1$ $L_2$ . . . $L_r$ . . .; si la ligne $L_1$ est occupée, l'appel est envoyé sur la ligne $L_2$, puis sur $L_3$ si $L_2$ est occupée, etc. On suppose que la durée des conversations est exponentielle, ce qui permet d'appliquer la formule d'Erlang donnant la probabilité pour un groupe de 2 lignes d'être occupées à un instant quelconque; c'est:

$$\frac{\dfrac{\lambda^r}{2!}}{\displaystyle\sum_{i=0}^{r}\frac{\lambda^i}{i!}}$$

$\lambda$ étant le nombre moyen d'arrivées par unité de temps. La probabilité, pour l'ensemble des lignes $L_1$ et $L_2$ de perdre l'appel est donc:

$$\frac{\dfrac{\lambda^2}{2!}}{\displaystyle\sum_{i=0}^{2}\frac{\lambda^i}{i!}} = \frac{\dfrac{\lambda^2}{2}}{1+\lambda+\dfrac{\lambda^2}{2}} = \frac{\lambda^2}{\lambda^2+2\lambda+2}\,.$$

Mais c'est encore une probabilité composée: probabilité que chacune des deux lignes $L_1$ et $L_2$ perdent l'appel, soit

$$\frac{\lambda}{1+\lambda} \times \frac{\lambda'}{1+\lambda'}\,,$$

d'après la formule d'Erlang pour $r = 1$, et $\lambda'$ étant le nombre moyen d'appels sur la seconde ligne par unité de temps, nombre évidemment plus petit que $\lambda$. Comme la proportion des pertes sur la ligne $L_1$ est $\frac{\lambda}{1+\lambda}$, on a:

$$\lambda' = \lambda \times \frac{\lambda}{1+\lambda} = \frac{\lambda^2}{1+\lambda}$$

d'où:

$$\frac{\lambda}{1+\lambda} \times \frac{\lambda'}{1+\lambda'} = \frac{\lambda}{1+\lambda} \times \frac{\dfrac{\lambda^2}{1+\lambda}}{1+\dfrac{\lambda^2}{1+\lambda}} = \frac{\lambda^3}{(1+\lambda)(1+\lambda+\lambda^2)}$$

nombre qui diffère de la valeur $\dfrac{\lambda^2}{\lambda^2+2\lambda+2}$

précédemment trouvée.

L'erreur provient de l'évaluation par la formule d'Erlang de la probabilité de perdre l'appel pour la ligne $L_2$; appliquer cette formule, c'est supposer implicitement les arrivées poissonniennes. Le paradoxe précédent prouve qu'elles ne le sont pas sur $L_2$, bien que les hypothèses de stationnarité, de régularité et d'indépendance entre les intervalles de temps successifs séparant deux appels consécutifs, soient réalisées.

Cet exemple montre, en outre, que l'utilisation de l'outil mathématique est une affaire délicate.

Cependant, l'investigation de Khintchine parmi les différents flux d'appels se termine par la conclusion que, malgré les trois hypothèses très restrictives qui le conditionnent, le processus poissonnien est encore celui qu'on rencontre le plus fréquemment, et le chapitre 5 est destiné à donner une explication de ce phénomène.

L'explication est due à Palm et elle n'est pas sans rappeler celle qu'on avance pour rendre raison de l'application fréquente de la loi normale en statistique: un processus donné représente la superposition d'un grand nombre de processus indépendants entre eux, de faible intensité, stationnaires et réguliers, mais auxquels il n'est pas nécessaire de supposer l'absence de quelque «*after-effect*» (deuxième hypothèse) pour que le processus-somme soit poissonnien.

Cette assertion repose sur le théorème de Palm: $\lambda_1 \lambda_2 \ldots \lambda_n$ étant les intensités des processus élémentaires, avec $\Lambda = \lambda_1 + \ldots + \lambda_n$, si, lorsque $n \to \infty$, les $\lambda_i$ tendent uniformément vers 0, $\Lambda$ restant constant; si, d'autre part, la probabilité pour chaque processus élémentaire d'au moins un appel dans une période de durée $t$ donnée suivant un appel tend, elle-même, uniformément vers 0, le processus-somme est alors poissonnien.

* * *

L'étude des arrivées est, comparativement, la plus simple et les lois trouvées sont générales. Le problème se complique lorsqu'on examine le système, point d'aboutissement des divers éléments.

Cette complication provient, soit de la diversité des systèmes étudiés, soit de la loi de probabilité de la durée des conversations (ou services).

On peut en effet imaginer une ou plusieurs lignes de lignes également accessibles ou non, les éléments refusés s'en retournant ou faisant la queue. Ces lignes peuvent être ordonnées, ou leur occupation peut se faire au hasard, l'ordre dans lequel sont servis les éléments a son importance, on peut considérer différentes sortes de priorité, etc.

D'autre part, les durées de conversation peuvent être déterminées ou aléatoires, et dans ce cas obéir à des lois très diverses. Le cas de la loi exponentielle est le plus fréquemment considéré, moins peut être à cause de son application fréquente que parce que l'étude est plus simple dans ce cas que dans un autre.

Enfin, et c'est par là que nous commencerons, le régime peut être transitoire ou permanent.

Dans tous les cas, un nombre aléatoire définira l'état du système. Ce peut être, par exemple, le nombre $k$ de lignes occupées, si le système est sans attente, ou le nombre d'éléments dans le système dans le cas contraire.

Ce nombre, à un instant $t$, a une probabilité qui dépend de $t$; on pourra désigner par $P_k(t)$ cette probabilité. D'autres probabilités interviennent, qui sont les probabilités de passage d'un état $i$ à un état $j$ dans l'intervalle de temps $t$, dans le cas, généralement réalisé, où le processus est Markovien (le futur n'y dépend pas du passé si le présent est connu); soient $P_{ij}(t)$ ces probabilités.

L'état du système sera dit permanent si les probabilités $P_k(t)$ ne dépendent plus de $t$. On voit qu'un système qui a atteint un tel état est beaucoup plus simple à étudier; il est défini par la loi de probabilité des $P_k(t)$ devenus des $p_k$.

Khintchine prend le problème sous l'angle mathématique. Il montre, tout d'abord, que, pour que les $P_k(t)$ tendent vers des quantités $p_k$ indépendantes des données initiales lorsque $t$ tend vers l'infini, il faut et il suffit que les probabilités de passage $P_{ik}(t)$ tendent vers la même limite pour toutes valeurs de $i$ (le nombre $k$ étant d'ailleurs supposé fini, ce qui restreint la portée de la démonstration).

Puis, il établit le théorème de Markov: pour tout processus transitif de Markov,

$$\lim_{t \to \infty} P_{ik}(t) = p_k$$

existe et ne dépend pas de $i$.

Montrons en quoi l'introduction d'un régime permanent au bout d'un temps infini simplifie les calculs, lorsqu'il s'agit, par exemple, de déterminer la probabilité, pour le système, d'avoir $k$ lignes occupées.

On suppose les arrivées poissonniennes d'intensité $\lambda$ et la durée de conversation, exponentielle, l'unité de temps étant la durée moyenne d'une conversation.

La structure particulière de ces lois permet, tout d'abord, une simplification: si $\Delta t$ est une durée infiniment petite, $P_{ik}(\Delta t) = 0$ dès que $i$ et $k$ diffèrent de plus d'une unité. Autrement dit, il n'y a pas de probabilité de deux arrivées ou de deux départs dans le même intervalle de temps $\Delta t$ suffisamment petit. De plus, $P_{k,\, k+1}(\Delta t) = \lambda \Delta t$, comme nous l'avons vu à propos des arrivées.

De même, la loi des durées de conversations étant exponentielle, la probabilité de l'arrêt d'une conversation sur une ligne dans l'intervalle de temps $\Delta t$ (qui est la même que la probabilité d'une durée de conversation inférieure à $\Delta t$ d'après la structure particulière de la loi exponentielle) est:

$$1 - e^{-\Delta t} \backsimeq \Delta t.$$

Si $k$ lignes sont occupées, la probabilité d'arrêt sur l'une quelconque de ces $k$ lignes est $k\Delta t$.

Par suite,

$$P_{k,k-1}(\Delta t) = k\Delta t.$$

Enfin,

$$\begin{aligned} P_{k,k}(\Delta t) &= (1 - \lambda\Delta t)(1 - k\Delta t) \\ &\simeq 1 - \lambda\Delta t - k\Delta t\,. \end{aligned}$$

Puisque:

$P_k(t+\Delta t) = P_{k-1}(t)P_{k-1,k}(\Delta t) + P_k(t)P_{k,k}(\Delta t) + P_{k+1}(t)P_{k+1,k}(\Delta t)$,

on a donc:

$P_k(t+\Delta t) = P_{k-1}(t)\,\lambda\Delta t + (1 - \lambda\Delta t - k\Delta t)\,P_k(t) + (k+1)\Delta t\; P_{k+1}(t)$,

d'où il résulte que:

$$P_k\frac{(t+\Delta t) - P_k(t)}{\Delta t} = \lambda P_{k-1}(t) - (\lambda + k)\,P_k(t) + (k+1)\,P_{k+1}(t),$$

et, lorsque $\Delta t \to 0$,

$P_k'(t) = \lambda P_{k-1}(t) - (\lambda + k)\,P_k(t) + (k+1)\,P_{k+1}(t)$.

Le théorème de Markov introduit la simplification suivante: si $t \to \infty$, $P_k(t) \to p_k$, donc $P'_k \to 0$,

et la relation devient:

$\lambda p_{k-1} - (\lambda + k)\,P_k + (k+1)\,p_{k+1} = 0$

qui est une relation de récurrence permettant l'évaluation des $p_k$.

Finalement, on trouve la formule d'Erlang dont nous avons parlé précédemment:

$$P_k = \frac{\dfrac{\lambda^k}{k!}}{\displaystyle\sum_{i=0}^{n} \frac{\lambda^l}{l!}},$$

$n$ étant le nombre de lignes dont on dispose.

Mais la question suivante se pose: dans la pratique, qu'est-ce qu'un temps infini, et atteint-on jamais un régime permanent ?

M. Girault, ici, n'établit pas le théorème de Markov, mais, tout d'abord, traite deux exemples de systèmes, l'un évoluant vers un régime permanent et l'autre non. Ce dernier est repris de G. Kendall: des taxis viennent en stationnement attendre des clients et leurs arrivées forment un processus de Poisson d'intensité $a$; les clients qui arrivent au poste de stationnement forment eux-mêmes un processus de Poisson d'intensité $b$; initialement il n'y a ni taxi ni clients; au bout d'un temps $t$ on a $X$ taxis et $Y$ clients; le système est caractérisé par le nombre aléatoire $Z = X - Y$, qui est la longueur de la file d'attente, soit de taxis, soit de clients suivant que $Z$ est positif ou négatif.

La moyenne de $Z$, au bout du temps $t$, est $(a - b)\,t$ et son écart-type $\sqrt{(a + b)\,t}$ qui augmentent indéfiniment avec $t$: le système n'évolue pas vers un régime stable.

En revanche, dans le second exemple donné, l'évolution d'une population, il y a convergence vers un état permanent, et, de plus, il est possible de lier les fonctions caractéristiques du régime transitoire et du régime permanent, ce qui permet d'apprécier la rapidité de la convergence.

Enfin, dans le cas d'arrivées poissonniennes avec services constants, M. Girault donne, en appendice, plusieurs tableaux permettant de comparer les états permanents et transitoires; par exemple, pour un coefficient d'emploi égal à 0.6, au bout d'un temps égal à la durée de 6 ou 7 services, l'infini est atteint, le régime peut être considéré comme permanent.

Cette façon de présenter les choses est conforme au plan de l'auteur dont le but est d'initier et non pas de publier un traité exhaustif. Il conclut même (p. 79) «Notre objectif serait atteint si le lecteur, fatigué de nos explications, fermait ce livre pour prendre un papier, un crayon, et continuait seul à étudier les problèmes qui le préoccupent».

Nous venons de faire une distinction entre régimes permanents et régimes transitoires. Les régimes transitoires ne sont pas abordés par nos deux auteurs.

Signalons la solution donnée au problème du régime transitoire dans les systèmes à 2 stations à service exponentiel, dans l'article de Dhandt des cahiers du Centre d'étude de recherche opérationnelle (volume 2, no 1, 1961) sur lequel nous reviendrons.

D'autres distinctions peuvent être faites: Khintchine qui garde le langage des téléphones, fait la distinction entre système avec attente ou non.

Dans le cas de l'attente, un nombre aussi grand que l'on veut d'éléments dans le système (nombre qui, ici, caractérise l'état) est possible: la démonstration du théorème de Markov donnée précédemment n'est alors pas applicable et on est amené à un acte de foi quant à l'existence d'un régime permanent dans les cas étudiés.

Les lois de durée de service (ou conversation), ici comme dans M. Girault, sont principalement la loi exponentielle et la durée constante. L'étude est complète en ce qui concerne la loi exponentielle (étude plus simple, étant due au fait que la loi de la durée restante d'une conversation déjà commencée est la même que la loi de la durée de cette conversation) et l'on donne, outre la loi du nombre d'éléments dans le système (qui définit l'état), la loi de la durée des attentes; M. Girault distingue, pédagogiquement, un poste de service, deux postes, $r$ postes. Il ajoute à l'étude probabiliste une considération qui intéresse plus spécialement la recherche opérationnelle: la question du temps perdu qui sera, pour beaucoup d'applications, le point où il conviendra de trouver un optimum économique. Par exemple, dans le cas d'un seul poste de service, $c$ étant le coefficient d'emploi (produit du nombre moyen d'arrivées par unité de temps par la durée moyenne d'un service), la loi du nombre d'éléments dans le service est la loi géométrique:

$$P(\mathcal{N} = n) = (1 - c)c^n$$

Le nombre moyen d'individus en attente est alors:

$$\begin{aligned} &(1 - c)c^2 + 2(1 - c)c^3 + 3(1 - c)c^4 + \dots \\ = \; & c^2(1 - c)[1 + 2c + 3c^2 + \dots] \\ = \; & \frac{c^2}{1 - c} \end{aligned}$$

ce qui fait, au cours d'une durée $\Delta t$, un temps perdu pour l'ensemble des «clients»:

$$\frac{c^2}{1 - c}\Delta t.$$

On comparera à la durée d'inoccupation du service. La probabilité qu'aucun élément ne se trouve dans le système est $1 - c$; donc, en moyenne, au cours de la durée $\Delta t$, le temps d'inoccupation du service sera:

$$(1 - c)\Delta t$$

et le coût global dû au temps perdu, de la forme:

$$\left[A(1-c) + B\frac{c^2}{1-c}\right]\Delta t.$$

Pour les temps de service constant, l'étude est difficile si l'on calcule préalablement les $P_k(t)$ quitte à passer ensuite à la limite. C'est la méthode de Khintchine qui n'étudie ce genre de service que pour un service unique. M. Girault, qui se place directement en régime permanent, donne les moyens d'obtenir les probabilités d'états, lorsqu'on dispose de $r$ services; une étude très intéressante est faite alors de la loi de la durée des attentes.

L'étude, dans l'un et l'autre ouvrage, se termine par la considération d'une loi quelconque, mais donnée, de durée de service, les arrivées étant poissonniennes.

Dans tous les exemples cités, les arrivées sont poissonniennes, cas très fréquent avons-nous dit mais cependant pas obligatoire. Saaty dans son ouvrage *Elements of queuing theory* cite un cas bien simple: dans une usine, des pièces fabriquées sont portées à l'inspection toutes les cinq minutes. La durée du service est exponentielle, mais on voit que les arrivées n'ont rien d'aléatoire. On peut alors utiliser des méthodes empiriques; par exemple effectuer une simulation du phénomène au moyen de la méthode de Monte Carlo par exemple, ce qui permettra d'obtenir des valeurs numériques pour les durées moyennes d'attente des clients, ou du préposé au service, ou pour tout autre nombre qu'il peut être intéressant de connaître.

*
* *

Pour terminer, montrons que les résultats obtenus dans la théorie des files d'attente peuvent être utilisés dans des domaines où l'on ne trouve ni poste de service ni queue.

Supposons qu'un grossiste reçoive des commandes aléatoires à des époques elles-mêmes aléatoires; si l'on se trouve, cas fréquent, dans le cadre de nos trois hypothèses, on peut considérer que les époques de commandes sont poissonniennes. Le nombre moyen de commandes par mois, par exemple, est facile à évaluer; désignons-le par $c$.

Au cours d'une durée $t$, le nombre $\mathcal{N}$ de commandes obtenu a donc pour la loi de probabilité:

$$\text{Prob}\left\{ \mathcal{N} = n \right\} = \frac{e^{-ct}(ct)^n}{n!}.$$

Le nombre $K$ d'unités à livrer à chaque commande ($K$ est nécessairement supérieur ou égal à 1) sera caractérisé par sa fonction génératrice $g_k(u)$: rappelons que

$$g_k(u) = \sum_{k=1}^{\infty} u^k \text{ Prob } \left\{ K = k \right\}$$

On voit qu'on a, ici, un exemple de processus en grappe; on peut s'intéresser au nombre total d'unités commandées au cours de la durée $t$, par exemple, pour éviter une rupture de stock.

Soit $S$ ce nombre, lui-même aléatoire.

Soit enfin $S_n$ le nombre total d'unités commandées au moyen de $n$ commandes. On a:

$$S_n = K_1 + K_2 + \ldots + K_n$$

$K_1, K_2, \ldots, K_n$ étant les quantités respectivement commandées au moyen de la 1ère, de la 2ième, . . ., de la $n$ième commande.

Ces $K_i$ sont généralement indépendants. La fonction génératrice de leur somme est le produit des fonctions génératrices des $K_i$; donc:

$$g_{S_n}(u) = [g_k(u)]^n.$$

S sera égal à $s$ si l'un des événements suivants est réalisé:

a) une commande a été passée pendant la durée $t$ et cette commande a été de $s$ unités.

   Probabilité correspondante:

$$e^{-ct}ct \times \text{Prob}\,[S_1 = s];$$

b) deux commandes ont été passées et, globalement, ces commandes ont été de $s$ unités.

   Probabilité correspondante:

$$\frac{e^{-ct}(ct)^2}{2!} \times \text{Prob}\,[S_2 = s];$$

etc.

On voit que

$$\text{Prob}\left\{ S = s \right\} = \sum_{n=1}^{\infty} \frac{e^{-ct}(ct)^n}{n!} \times \text{Prob}\left\{ S_n = s \right\}.$$

De plus, il est clair que:

$\text{Prob}\{ S = 0 \} = \text{Prob}\{ \mathcal{N} = 0 \} = e^{-ct}$.

La fonction génératrice de $S$ est

$$g_S(u) = \sum_{s=0}^{\infty} u^s \times \text{Prob}\left\{ S = s \right\}$$

$$= e^{-ct} + \sum_{s=1}^{\infty} u^s \left[ \sum_{n=1}^{\infty} \frac{e^{-ct}(ct)^n}{n!} \times \text{Prob}\{ S_n = s \} \right]$$

ce qui peut s'écrire:

$$g_S(u) = e^{-ct} + \sum_{n=1}^{\infty} \frac{e^{-ct}(ct)^n}{n!} \left[ \sum_{s=1}^{\infty} u^s \times \text{Prob}\{ S_n = s \} \right].$$

La quantité entre crochets est la fonction génératrice de $S_n$: c'est $[g_K(u)]^n$, d'où

$$g_S(u) = e^{-ct} + \sum_{n=1}^{\infty} \frac{e^{-ct}(ct)^n [\, g_K(u) \,]^n}{n!}$$

$$g_S(u) = \sum_{n=0}^{\infty} \frac{e^{-ct}[\, ctg_k(u) \,]^n}{n!}$$

$$= e^{-ct} \sum_{n=0}^{\infty} \frac{[\, ctg_k(u) \,]^n}{n!}$$

$$\boxed{g_S(u) = e^{ct}(g_k(u) - 1)}$$

Ce qui détermine la loi de $S$. Pratiquement, il sera suffisant de connaître la moyenne et la variance. On peut les obtenir à partir de la fonction $g_S(u)$.

On a:

$$g_S(1 + h) = e^{ct}[\, g_k(1 + h) - 1 \,]$$

$$= 1 + ct[\, g_k(1 + h) - 1 \,] + \frac{c^2t^2}{2}[\, g_k(1 + h) - 1 \,]^2 + \ldots$$

mais, $g_k(1 + h) = 1 + E(K)h + E[K(K-1)]\frac{h^2}{2} + \ldots$

c'est-à-dire, $m$ et $v$ désignant moyenne et variance de K:

$$g_K(1 + h) = 1 + mh + (v + m^2 - m)\frac{h^2}{2} + \ldots$$

d'où

$$g_S(1 + h) = 1 + ct\left[mh + (v + m^2 - m)\frac{h^2}{2} + \ldots\right] + \frac{c^2t^2}{2}\left[mh + (v + m^2 - m)\frac{h^2}{2} + \ldots\right]^2$$

On en déduit la moyenne de $S$:

$$M = ctm,$$

et son moment factoriel du 2[ième] ordre:

$$ct[v + m^2 - m + ctm^2],$$

d'où l'on déduit la variance:

$$V = ct(v + m^2).$$

On peut d'ailleurs l'écrire:

$$V = \frac{c^2t^2m^2}{ct}\left(\frac{v}{m^2} + 1\right) = \frac{M^2}{ct}\left(\frac{v}{m^2} + 1\right).$$

Pratiquement, le rapport $\frac{v}{m^2}$, suivant une remarque de M. Ferrier[3], est toujours inférieur ou égal à 1, donc on obtient une borne supérieure de la variance de $S$ qui sera: $\frac{2M^2}{ct}$.

La connaissance de cette variance sera nécessaire si l'on veut prévoir, par exemple, le stock minimum de sécurité nécessaire. On voit que les seules données demandées à la statistique seront: le nombre moyen de commandes par unité de temps et le nombre moyen d'unités commandées chaque fois.

3. *Revue française de la Recherche Opérationnelle*, no 14, 1[er] trimestre 1960, pp. 74-78.

En résumé, constatons que le champ des applications de la théorie des files d'attentes s'étend, mais qu'une étude complète de l'évolution d'un système donné n'a été obtenue que dans des cas particuliers; chaque cas pratique étudié ou chaque modification des hypothèses initiales entraîne une publication dans une revue spécialisée; ce qui forme un ensemble de problèmes et non pas une théorie. Cependant, l'outil mathématique utilisé est le même et donne à l'ensemble une manière d'unité. C'est en ce sens qu'on peut parler de théorie des files d'attentes.

Claude TRICOT,
*professeur à l'École des*
*Hautes Études commerciales (Montréal).*