

Université **M**Hamed **B**OUGARA de **B**oumerdès

Faculté des **H**ydrocarbures et de la **C**himie

Département d'Économie
et
Commercialisation des Hydrocarbures

**Les files
d'attente**

Razika TAHI

Année universitaire **2021-2022**

Table des matières

INTRODUCTION	4
I – Principales caractéristiques d’un processus d’attente.....	7
I – I – Processus d’arrivée	8
1 – 2 – Processus de service.....	13
1 – 3 – Discipline de la file d’attente.....	15
1 – 4 – Organisation des installations.....	17
1 – 5 – Conclusion.	17
2 – Modèle de file d’attente à station unique et temps de service exponentielle.....	19
2 – 1 – 1 – Calcul de la probabilité P_n qu’il y ait n unités dans le système.	19
2 – 1 – 2 – Nombre moyen d’unités dans le système n	22
2 – 1 – 3 – Temps moyen d’attente dans le système t	22
2 – 1 – 4 – Nombre moyen d’unités dans la file d’attente n'	23
2 – 2 – Modèle M/M/1 : avec priorité (deux files d’attente dont l’une est prioritaire).....	24
2 – 2 – 1 – Cas de priorité non absolue.	24
2 – 2 – 2 – Cas de priorité absolue.	26
2 – 3 – Modèle M/M/1 : (∞ , LIFO).....	26
2 – 4 – Modèle M/M/1 : (∞ , SIRO)	28
2 – 5 – M/M1 : (x , FIFO)	28
3 – Modèle de file d’attente à station unique et temps de service exponentielle.....	30
3 – 1 – Modèle M/E _k /1 : (∞ , FIFO) appelé aussi modèle d’Erlang.....	30
3 – 2 – Modèle M/D/1 : (∞ , FIFO).....	31
3 – 3 – Modèle M/G/1 : (∞ , FIFO)	32

4 - Modèle de file d'attente à plusieurs stations	33
4 - 1 - Modèle M/M/C : (∞ , FIFO)	33
4 - 2 - Modèle M/D/C : (∞ , FIFO)	37
4 - 3 - Modèle M/E _k /C : (∞ , FIFO)	38
4 - 4 - Modèle M/M/2 : (∞ , GD)	38
5 - Modèle de file d'attente à processus biphase.	42
5 - 1 - Modèle M/M/1 : (∞ , FIFO) alimentant M/M/1 : (∞ , FIFO)	42
5 - 2 - Modèles M/M/1 : (n_1 , FIFO) alimentant M/M/1 : (n_2 , FIFO)	44
6 - La simulation	45
CONCLUSION	49
Exercices	52
Solutions des exercices	56
Bibliographie	61

INTRODUCTION

L'objet de ce cours, concerne une partie déplaisante dans la vie de chacun : l'attente dans une file, plus connue sous le terme de 'chaîne'.

Dans la vie quotidienne, les files d'attente se présentent à nous à tout moment de notre existence, et il semble de plus en plus difficile de les éviter. Elles se présentent sous des formes diverses tels des cas très concrets comme les chaînes dans certains points de vente (supérette, guichet de postes, arrêt d'autobus, etc.). Il peut exister des cas moins visibles, mais tout autant néfastes ; c'est le cas d'avions attendant l'autorisation de la tour de contrôle avant d'atterrir, surcharge au standard téléphonique ou machine en panne dans un atelier de réparation.

Les répercussions de ces phénomènes entraînent des pertes qui ont amené les gestionnaires à la recherche de solutions pouvant y remédier. Parfois, pour un certain nombre de problèmes, des mesures pourraient venir immédiatement à l'esprit tels que l'augmentation des points de vente dans une grande surface de vente lors des heures de pointe, où la construction de plusieurs sorties à la fin d'une autoroute pour éviter les goulets d'étranglement. Cependant, l'essentiel dans les différentes situations étudiées et de déterminer à quelles conditions de telles solutions, entraînant souvent des frais supplémentaires élevés, peuvent être rentables, et dans quelle mesure elles réduiraient effectivement les files et les temps d'attente. C'est ici que la recherche opérationnelle intervient.

L'étude théorique des processus de files d'attente a pour origine les problèmes de liaisons téléphoniques, elle s'est ensuite appliquée à de très nombreux domaines¹.

La plupart des problèmes de file d'attente dans la recherche opérationnelle se ramènent à une série d'unités (appelé aussi client ou individus) qui se présentent devant une installation (stations aux guichets) assurant un service à chacune d'entre elles, puis la libérant. L'installation consiste souvent en plusieurs postes identiques. Des unités peuvent également avoir à subir plusieurs opérations successives, c'est le cas des chaînes de fabrication. S'il y a des variations dans

¹ Cette origine, associée aux multiples possibilités de résolution par l'analyse mathématique a beaucoup influencé le développement de la théorie des files d'attente. En effet, nous constaterons par la suite que nombreux sont les hypothèses couramment faites dans l'étude des files d'attente qui semblent raisonnables et valables quand on construit et qu'on étudie les modèles de file d'attente associés aux systèmes téléphoniques, ne sont pas toujours applicables ailleurs, et elles sont particulièrement inadaptées quand les serveurs et les clients sont des personnes.

la régularité des arrivées ou dans des délais de traitement de chaque unité, ou dans les deux cas, tout le système subi des fluctuations, et des fils de longueur variable se constitue pendant que parfois certains postes restent inactifs ; le débit à la sortie est irrégulier.

Dans une file d'attente est compris l'ensemble des clients qui attendent d'être servi, à l'exclusion de ce qui se font servir. En opposition, le système d'attente comprend l'ensemble des clients présents. Lorsque le nombre de clients est illimité nous parlerons de systèmes d'attente ouvert. C'est le cas d'arrivée de clients devant une station d'essence. Lorsque le nombre de clients est limité, nous parlerons de système d'attente fermé. C'est le cas dans une usine de machines en panne attendant d'être réparés (le nombre de machines en panne est donc inférieur, ou égal, au nombre de machines disponibles).

Pour l'étude des files d'attente, la recherche opérationnelle utilise des modèles mathématiques. Il est nécessaire de préciser que lorsqu'un mathématicien dit qu'il a résolu un problème de file d'attente, il veut seulement dire qu'il a bâti un modèle du processus dans lequel apparaît le problème. Si la différence entre modèle et réalité est parfois négligeable, elle peut par contre être très sensible, et l'optimisation d'un modèle dans ce cas n'aura plus aucun sens dans la pratique. La théorie des files d'attente est donc une théorie générale qui s'applique au comportement des modèles théoriques généralisés, et non pas au processus des files d'attente lui-même. Il faudra donc toujours garder en mémoire que le modèle n'est pas le processus. En recherche opérationnelle, 'solution' a pris le sens d'un modèle bien défini.

Les modèles mathématiques cherchent à analyser les différents états possibles dans lesquels le système peut se trouver, et calculent les probabilités pour qu'il s'y trouve. Pour cela, il faut d'abord exprimer en termes de probabilité la distribution des arrivées par unité de temps, et celle des services pendant la même unité de temps, le rapport représentant l'intensité du trafic. Suivant ce dernier, il y a allongement, ou non, de la file d'attente ;

Les calculs de probabilité permettent de déterminer, à un moment donné, le nombre d'unités dans la file d'attente, le nombre d'unité dans le système, le temps moyen d'attente dans la file, et d'étendre le problème au cas où il existe plusieurs files d'attente. Puis en fonction des temps d'attente, de déterminer les pertes qui en résultent. On peut alors envisager un poste de service supplémentaire si le coût de l'attente est supérieur au coût d'emploi du poste en question. Généralement, le type de questions qui se posent dans les processus de file d'attente et qui peuvent être résolus avec les calculs précédant sont les suivantes :

- Combien de stations supplémentaires doit-on employer si le nombre des clients, et le temps nécessaire pour servir chacun d'eux ne changent pas, pour réduire la proportion des clients qui doivent attendre plus de x unités de temps à moins de y % ?
- Quel effet aura la durée d'attente des clients l'introduction de priori
- Si l'on a deux catégories de clients, une qualité donnée de service té pour certaines catégories de clients ? (mesurée par la durée d'attente moyenne), sera-t-elle plus économique en établissant un groupe de guichets séparés pour chaque catégorie ?
- De combien, faut-il agrandir une salle d'attente pour réduire de x % le nombre de clients rejetés ?
- Quel serait le coût le plus avantageux si on devait choisir entre ajouter une station, ou agrandir la salle d'attente ?
- Quand est-ce que le coût d'attente dans une file d'attente est-il supérieur au coût d'achat d'une station supplémentaire ?

En matière économique, et notamment en économie d'entreprise, l'application pratique des files d'attente se situe dans la recherche des processus permettant d'optimiser les résultats en fonction des circonstances et des moyens mis en œuvre, tel que réduire les temps d'attente des clients, éviter la rupture des stocks, améliorer la vitesse de rotation de véhicules ou de navires, etc.

L'objet de la théorie des phénomènes d'attente et d'établir des relations entre la nature du flux de demandes et le nombre d'installations avec leur productivité, l'organisation du travail, ainsi que l'efficacité du service.

I – Principales caractéristiques d'un processus d'attente

Nous allons considérer dans ce qui suit les principales caractéristiques d'un phénomène de file d'attente, et plus particulièrement le processus d'entrée, le mécanisme de service, le nombre de stations, la longueur de la file et la discipline de la file d'attente. Pour cela, nous utiliserons la classification de Kendall, résumé sous la forme A/B/s : (d, e) où :

- A représente le processus d'arrivée, c'est-à-dire la loi de probabilité de l'intervalle de temps qui sépare deux arrivées consécutives. S'il s'agit d'une loi exponentielle, elle est notée E.
- B représente la distribution du temps de service.
- s représente le nombre de stations.
- d représente la longueur de la file, c'est-à-dire le nombre maximum de clients qui peuvent être contenu dans le système à n'importe quel instant (y compris les clients qui se font servir). Si le nombre de clients est illimité d est remplacé par le signe de l'infini (∞).
- e représente la discipline de la file d'attente.

Pour plus de clarté, considérons les exemples qui suivent.

- E/E/I : (∞ , FIFO). Il s'agit d'un phénomène d'attente avec des arrivées suivant une loi de Poisson (ou exponentielle), des temps des service distribués suivant une loi exponentielle, une seule station, une file d'attente illimitée et un discipline de file d'attente du type FIFO (les différentes disciplines font l'objet d'un autre paragraphe).
- M/M/I : (20, FIFO). Il s'agit d'un modèle d'attente où les arrivées sont aléatoires, ainsi que le temps de service. Il y a une station, une capacité maximale du système égale à 20 (donc quand 19 unités attendent alors qu'une autre se fait servir, tous les nouveaux arrivants sont refoulés ou perdus), et on applique la règle de service FIFO.
- M/E_k/5 : (∞ , SIRO). Il s'agit d'un modèle de file d'attente avec une introduction aléatoire, un temps de service suivant la loi d'Erlang², cinq stations, file d'attente de capacité infinie (donc

² Agner Krarup Erlang (1878-1929) : ingénieur et mathématicien danois qui établit en 1917 la théorie du trafic téléphonique. Il fut le premier à utiliser la loi qui porte son nom dans les études relatives aux files d'attente.

on ne refuse jamais une nouvelle arrivée) et qui sont servi dans un ordre aléatoire (**Served In Random Order, SIRO**).

I – I – Processus d'arrivée.

L'analyse de ces arrivées commence par une observation du phénomène. Il suffit donc de noter pendant un certain temps les heures d'arrivées de toutes les unités dans le système, d'en déduire les intervalles de temps séparant deux arrivées, de les classer, les dénombrer et de construire la courbe de fréquence. Ainsi, il est possible de trouver la densité de distribution des intervalles séparant deux arrivées consécutives.

Dans un grand nombre de cas pratiques, la courbe de fréquences ainsi mise en évidence a généralement la forme d'une expression mathématique relativement simple. C'est donc à partir d'une analyse statistique que nous étudierons le processus d'arrivée.

Les instants où les unités s'introduisent dans un système s'appellent les instants d'arrivées, et les intervalles entre ces instants s'appellent inter-arrivées.

Nous limiterons notre étude de cas où les arrivées sont indépendantes entre elles, ainsi que les inter-arrivées³. On considérera aussi que le phénomène étudié est stationnaire, c'est-à-dire qu'il ne varie pas au cours du temps⁴.

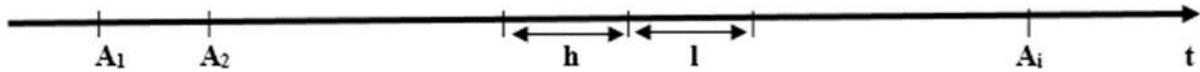
Ces limites ayant été précisées, nous allons à présent rechercher la densité de distribution, que nous noterons $g(\theta)$ des intervalles θ .

Désignons par $f(t)$ la probabilité pour qu'un évènement ne se produit pas dans un intervalle de temps de durée t , et soient deux intervalles consécutifs de durée h et l .

³ Nous faisons cette hypothèse pour des simplifications mathématiques. Cette hypothèse ne se vérifie pas toujours dans la réalité. Cependant, l'étude de cas où les arrivées sont dépendantes entre elles entraîne des calculs forts complexes et de peu d'utilité pratique, c'est-à-dire que l'approximation peut être acceptable dans la pratique.

⁴ Evidemment, c'est rarement le cas dans la réalité. Cependant, la différence entre la théorie et la réalité n'est pas très grande dans les applications pratiques, car la plupart des problèmes liés aux processus d'attentes concernent les heures de pointes. Si ce n'est pas le cas, on peut toujours décomposer la période étudiée en plusieurs périodes élémentaires, telles que chacune d'entre elles, la loi des inter-arrivées soit considérée comme stationnaires.

Figure 1 : Représentation schématique des arrivées au hasard.



On peut écrire que :

- La probabilité pour qu'un évènement ne se produise pas dans l'intervalle h est f (h),
- La probabilité pour qu'un évènement ne se produise pas dans l'intervalle l est f (l),
- La probabilité pour qu'un évènement ne se produise pas dans l'intervalle h + l est f (h + l).

Mais cette dernière probabilité est aussi f (h) x f (l), d'où la relation :

$$f (h) \times f (l) = f (h + l)$$

qui caractérise la fonction exponentielle⁵. Par conséquent la fonction f (t) est de la forme :

$$f (t) = e^{At}$$

Puisque f (t) est une probabilité, donc inférieur à 1, A doit être négatif. Il est donc pratique d'introduire une constante positive λ , et d'écrire :

$$f (t) = e^{-\lambda t}$$

On verra par la suite que λ désigne le taux moyen d'arrivées de clients par unité de temps.

Nous venons d'établir que la probabilité qu'il n'y ait pas d'évènement dans un intervalle quelconque t est $f (t) = e^{-\lambda t}$. Par conséquent, la probabilité du phénomène complémentaire, c'est-à-dire la probabilité qu'il y ait au moins un évènement dans l'intervalle t est :

$$1 - e^{-\lambda t}$$

⁵ Rappel sur la fonction et loi exponentielle.

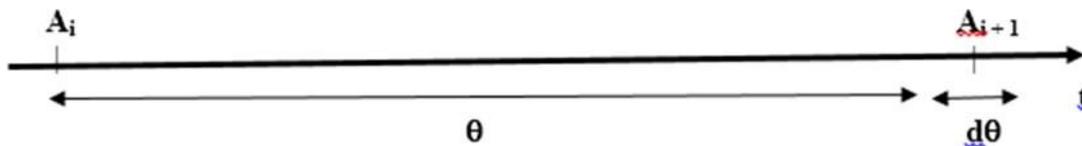
- Fonction de répartition : la probabilité des intervalles inférieur à θ est $G(\theta) = 1 - e^{-\lambda \theta}$.
- Fonction de répartition complémentaire : la probabilité des intervalles supérieurs à θ est $e^{-\lambda \theta}$.
- Valeur moyenne des intervalles : $\theta^- = 1/\lambda$.
- Nombre d'intervalles ou d'arrivées dans l'unité de temps : λ .
- Variance : $V = 1/\lambda^2$.
- Ecart-type : $\sigma = 1/\lambda$.

Après le calcul de ces probabilités, nous allons nous intéresser à la distribution des intervalles séparant les arrivées successives.

Remarquons que la probabilité $f(t)$ (aucune arrivée dans l'intervalle t) ne dépend pas de l'origine dans l'intervalle t , et que nous pouvons donc la confondre avec une arrivée A_i .

L'arrivée suivante sera A_{i+1} . Si elle est séparée de A_i par un intervalle θ , chercher la probabilité élémentaire de la durée θ revient à chercher la fonction $g(\theta)$ tel que la probabilité de l'arrivée A_{i+1} dans l'intervalle $]\theta, \theta + d\theta[$, soit $g(\theta)d\theta$.

Figure 2 : Représentation schématique des arrivées.



Or, dire que A_{i+1} se produit dans l'intervalle $]\theta, \theta + d\theta[$ compté à partir de A_i , revient à énoncer deux conditions :

- A_{i+1} ne se produit pas dans l'intervalle $]0, \theta[$.
- A_{i+1} ne se produit pas dans l'intervalle $]0, \theta + d\theta[$.

Nous connaissons la probabilité de ces événements :

$$e^{-\lambda\theta} \text{ pour la première condition}$$

$$\text{et } 1 - e^{-\lambda d\theta} \text{ pour la deuxième condition.}$$

Ainsi, la probabilité simultanée de ces événements que nous avons déjà noté $g(\theta)d\theta$ est :

$$g(\theta)d\theta = e^{-\lambda\theta} (1 - e^{-\lambda d\theta}).$$

Or, d'après l'égalité approximative $1 - e^{-\lambda d\theta} \approx 1 - \lambda d\theta$,

$$\text{on aura } g(\theta)d\theta = e^{-\lambda\theta} (1 - \lambda d\theta) = e^{-\lambda\theta} - \lambda d\theta e^{-\lambda\theta} = \lambda \cdot e^{-\lambda\theta} d\theta$$

$$\text{d'où : } g(\theta) = \lambda \cdot e^{-\lambda\theta}$$

Ainsi, si un phénomène d'arrivées aléatoires est tel que celles-ci soient indépendantes du temps et indépendantes les unes des autres, alors la fréquence des intervalles séparant deux arrivées successives est exponentielle. On pourra l'employer chaque fois qu'il sera empiriquement constaté que le temps moyen des intervalles séparant deux arrivées successives est égal à

l'écart-type, puisque pour la fonction exponentielle $E(\theta) = \sigma_0 = 1/\lambda$ où λ désigne le taux moyen d'arrivées de clients par unité de temps.

A partir de cette fréquence, nous demander aussi quelle est la probabilité $P_n(t)$ pour qu'il y ait exactement n arrivées dans cet intervalle t .

Pour cela, nous prenons tout d'abord pour hypothèses que des arrivées sont indépendantes les unes des autres, le phénomène est stationnaire et la probabilité qu'une arrivée se produise pendant l'intervalle de temps Δt est $\lambda\Delta t$ (la probabilité que deux arrivées ou plus se produisent est négligeable).

Supposons que n arrivées aient été enregistrées à la date $t + \Delta t$. Une telle situation ne peut parvenir que dans les deux cas suivants où $P_n(t)$ désigne la probabilité que :

- n arrivées avaient été enregistrées en t et aucune ne s'est produite pendant Δt (probabilité : $P_n(t) [1 - \lambda\Delta t]$).

- $n - 1$ arrivées avaient été enregistrées en t et une soit produite pendant Δt (probabilité : $P_{n-1}(t) [\lambda\Delta t]$).

Dans ces conditions, on a :

$$P_n(t + \Delta t) = P_n(t) [1 - \lambda\Delta t] + P_{n-1}(t) [\lambda\Delta t].$$

Cette formule est valable quel que soit n , sauf pour $n = 0$, puisque l'évènement $n - 1$ est alors impossible. Pour $n = 0$, la relation devient :

$$P_0(t + \Delta t) = P_0(t) (1 - \lambda\Delta t),$$

soit encore, en passant à la limite :

$P_0'(t) = -\lambda P_0(t)$, équation différentielle qui a pour solution $P_0(t) = C \cdot e^{-\lambda t}$ car $P_0(0) = 1$, $C = 1$ et

$$P_0(t) = e^{-\lambda t}.$$

Dans le cas où n est positif, on aura :

$$P_n'(t) + \lambda P_n(t) = \lambda P_{n-1}(t).$$

Par exemple, pour $n = 1$, on aura :

$$P_1'(t) + \lambda P_1(t) = \lambda P_0(t) = \lambda e^{-\lambda t}$$

équation différentielle dont la solution est :

$$P_1(t) = \lambda t \cdot e^{-\lambda t}.$$

Pour n quelconque, nous pouvons démontrer que la formule précédente se généralise, et que nous avons :

$$P_n(t) = \frac{(\lambda t)^n \cdot e^{-\lambda t}}{n!}$$

$P_n(t)$ représente la probabilité qu'il y ait exactement n arrivées dans cet intervalle, avec λ qui représente le taux moyen d'arrivées par unité de temps.

Cette démonstration peut être faite par la méthode de récurrence.

Nous supposons donc vraie que :

$$P_{n-1}(t) = \frac{(\lambda t)^{n-1} \cdot e^{-\lambda t}}{(n-1)!} \quad (1)$$

Nous avons :

$$P_n(t + \Delta t) = P_n(t) (1 - \lambda \Delta t) + P_{n-1}(t) (\lambda \Delta t),$$

Cette équation, après passage à la limite, s'écrit :

$$P'_n(t) = P_{n-1}(t) - P_n(t) \quad (2)$$

D'où, en reportant la relation (1) et (2) :

$$P'_n(t) = \frac{(\lambda t)^{n-1} \cdot e^{-\lambda t}}{(n-1)!} - P_n(t)$$

Ce qui implique :

$$P_n(t) + P'_n(t) = \frac{(\lambda t)^{n-1} \cdot e^{-\lambda t}}{(n-1)!} \quad (3)$$

Cette relation est une équation différentielle du premier ordre qui admet une solution de la forme $P_n(t) = Q(t) \cdot e^{-\lambda t}$, ce qui donne :

$$P'_Q(t) = Q'(t) \cdot e^{-\lambda t} - \lambda Q(t) \cdot e^{-\lambda t} = Q'(t) \cdot e^{-\lambda t} - \lambda P_n(t)$$

Et en reportant dans la formule (3) $Q'(t) = \frac{\lambda(\lambda t)^{n-1}}{(n-1)!}$, on obtient :

$$Q(t) = \frac{nt^n}{n(n-1)!} = \frac{(\lambda t)^n}{n!} \quad \text{et} \quad P(t) = \frac{(\lambda t)^n}{n!} \cdot e^{-\lambda t}.$$

La formule ayant été démontré pour $n = 1$, et la relation de récurrence vérifiée, la formule est donc établie pour toute valeur de n .

Il s'agit donc de la loi de Poisson⁶. En conséquence, arrivée poissonnienne et intervalle exponentielle⁷ sont deux aspects différents du même phénomène.

Les conditions nécessaires et suffisantes pour que les arrivées d'un phénomène d'attente obéissent à une loi de Poisson sont assez fréquentes. Cependant, il sera toujours indispensable de s'assurer qu'il en est bien ainsi, et cela à l'aide d'un test statistique. Il y aura des cas où la

⁶ Rappel. La loi de Poisson calcul la probabilité du nombre d'évènements dans un intervalle de temps.

⁷ Rappel. La loi exponentielle calcul la probabilité d'attendre plus de t minutes (ou une unité de temps) avant le premier évènement.

loi de Poisson et la loi exponentielle ne représentent pas correctement les arrivées et les intervalles de temps entre deux arrivées. D'autres lois statistiques sont alors utilisées.

1 – 2 – Processus de service.

L'intervalle séparant l'instant auquel une unité commence à être servie de celui où elle a fini d'être servie est son temps de service. Nous appellerons temps d'immobilisation l'intervalle qui s'écoule entre le moment où le temps de service d'une unité commence et celui où commence le temps de service de l'unité suivante lorsque celle-ci attendait son tour dans la file d'attente.

Evidemment, le temps de service et le temps d'immobilisation d'un client ne sont pas toujours égaux, particulièrement lorsque le service est assuré par une personne. Il arrive assez souvent qu'entre deux services, le serviteur marque une pause. Dans ce cas, le temps d'immobilisation est égal au temps de service plus le temps de la pause.

Dans l'étude des services que nous analyserons, nous tiendrons comptes des hypothèses suivantes :

- Le temps de service sera égal au temps d'immobilisation.
- Le temps de service de tous les clients suivent la même loi de probabilité, c'est-à-dire qu'il n'y a pas plusieurs catégories de clients.
- Les services de chaque station et identique, leur temps de service obéissent à la même loi de probabilité.
- Les temps de service sont indépendants (ils ne sont corrélés ni entre eux, ni avec des éléments du système).
- La loi de distribution des temps de services est stationnaire.

Ces hypothèses sont en général vérifiées dans les travaux théoriques, mais en pratique elles le sont moins souvent. C'est le cas des deux situations suivantes :

- Des bateaux minéraliers dans un port pour décharger leur cargaison de minerai de fer. Ils ont donc des tonnages, des structures et des capacités différents. Les temps nécessaires à les décharger (les temps de service) ne sont pas statistiquement homogènes ; ils proviennent de lois différentes.
- Dans un aéroport, les employés des lignes d'aviation en contact avec le public travaillent par poste de huit heures à enregistrer les passagers en partance. Des études ont montrées qu'ils se fatiguaient et que les temps de service ont à la fois plus longs et plus variables à la fin d'un poste.

Dans les nombreux cas, comme ceux cités ci-dessus, où les hypothèses ne sont pas rigoureusement vérifiées, ma précision demandée aux résultats justifie l'approximation ainsi faite qui a l'immense mérite de simplifier considérablement les calculs. Cependant, il ne faudrait pas négliger de vérifier l'exactitude de ces modèles.

Nous allons à présent examiner assez rapidement les types de lois du temps de service les plus fréquemment rencontrées et les plus étudiées. La première est la loi exponentielle négative, ou appelée tout simplement loi exponentielle.

On a déjà considérée la loi exponentielle à propos des arrivées. Dans le cas des services, la densité de probabilité d'une durée de service de longueur t est donnée par la formule suivante :

$$b(t) = \mu \cdot e^{-\mu t}$$

où μ représente le nombre moyen de clients servis par unité de temps.

Il s'agit là d'une forme mathématique tout à fait commode pour la loi du temps de service. La raison en est que le temps restant avant l'achèvement du service est indépendant du temps depuis lequel le service a déjà commencé. Cela signifie que si, par exemple, un client arrive et trouve seulement une personne devant lui en train de se faire servir, le temps pendant lequel il devra attendre suit exactement la même loi que le temps de service. Cela simplifie considérablement l'analyse.

Le temps de service exponentiel n'est pas général, mais il arrive souvent que les lois des durées observées ne diffèrent pas d'une façon notable de la loi exponentielle. Quand il se produit une différence notable, on peut, très souvent, utiliser les lois d'Erlang qui généralisent la loi exponentielle. Elles sont couramment employées dans le cadre de l'étude des phénomènes d'attente. Elles sont de la forme générale suivante :

$$f(t) = \frac{K\lambda (K\lambda t)^{K-1} \cdot e^{-K\lambda t}}{(K-1)!} = \frac{(\lambda K)^K \cdot e^{-K\lambda t} \cdot t^{K-1}}{(K-1)!}$$

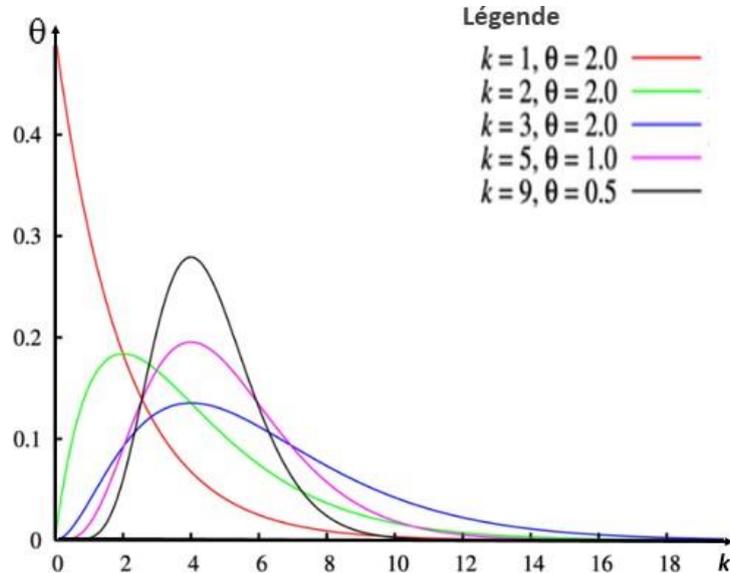
Avec pour espérance mathématique $E(t) = t = \frac{1}{\lambda}$ et pour variance $\sigma_t^2 = \frac{1}{K\lambda^2}$ pour la loi d'ordre K .

Ces lois peuvent couvrir une grande variété de situation. Elles sont très dissymétriques pour les faibles valeurs de K . Pour $K = 1$, on retrouve la loi exponentielle négative. Lorsque K augmente, la variance décroît, la loi s'approche d'une probabilité constante pour K très grand. Ainsi, pour $K = \infty$, cela correspond au cas d'intervalle constants entre deux évènements, la variance $\frac{1}{K\lambda^2}$ est

alors nulle.

Graphe 1 Distribution d'Erlang

$$f(t) = \frac{K\lambda (K\lambda t)^{K-1} \cdot e^{-K\lambda t}}{(K-1)!} \quad (6^8)$$



Quand aucune loi statistique connue ne permet de représenter correctement le phénomène étudié, on se contentera soit d'utiliser les valeurs empiriques de la moyenne et de la variance obtenue sur l'échantillon de données dont on dispose, ou soit faire une simulation.

1 – 3 – Discipline de la file d'attente.

La discipline d'une file d'attente a pour objet de préciser quelle unité parmi celle actuellement en attente sera traitée la première. Nous ne détaillerons pas ce point, mais cela ne signifie nullement qu'il est dénué d'importance.

Ainsi, dans les lignes de production puis dans des activités de répartition et d'entretien, l'assignation de priorités à certains articles ou composants est l'un des principaux moyen d'assurer une production efficace et de minimiser les stocks de pièces de rechange. Dans ce paragraphe, on se limitera à présenter seulement la discipline d'une file d'attente car leur formulation mathématique est très difficile et de peu d'utilité pratique.

- La discipline la plus courante est celle du 'premier arrivé, premier servi' (PAPS). Selon cette discipline, plus connue sous le nom de FIFO (First In, First Out), l'occupation des places dans

⁸ Sous cette forme, il s'agit de la loi suivie par la somme de K variables aléatoires indépendantes, suivant la même loi exponentielle de moyenne $1/\lambda K$. C'est un lien tout à fait utile entre la loi d'Erlang et la loi exponentielle. C'est aussi, après changement d'échelle, une loi du khi deux de moyenne $1/\mu$ à $2K$ degrés de liberté.

les files d'attente se fait selon l'ordre chronologique d'arrivée des demandes. Cette discipline est la plus équitable du point de vue des clients préoccupés de leur propre temps d'attente. Ce n'est pas précisément la meilleure du point de vue des serveurs ; et elle peut être sans intérêt dans les situations où les clients ne s'intéressent pas à leur temps d'attente.

- Lorsque le service est fait dans un ordre aléatoire, il s'agira d'une discipline connue sous le code SIRO (Service In Randomly Order). Cette situation se retrouve dans de nombreux cas tels que des personnes attendant un autobus à des heures de pointes, vente de marchandises soldées dans un marché, etc.

- Il est des cas où le dernier arrivé est le premier servi. C'est la discipline connue sous le code LIFO (Last In, First Out). C'est le cas où la file d'attente est constituée en pile, le premier arrivé est en bas de la pile et on commence à servir par le haut.

D'autres types de disciplines d'attente sont associés à des règles de priorités quand il y a plusieurs espèces de clients. Les clients appartenant, par exemple, à k catégories et chacune étant affectée d'un indice de priorité. Les clients faisant partie de la première catégorie arrivent suivant un schéma spécifié par la loi $A_i(t)$ des intervalles d'arrivées et leur temps de service obéit à une loi $B_i(t)$. Les clients appartenant à la même catégorie et servis suivant un règle FIFO, mais les clients d'indice prioritaire sont servis avant les clients d'indice prioritaire $i + 1$. Il existe deux types de schémas de priorité.

- Priorité absolue. Si un client d'indice prioritaire i est en train de se faire servir, on peut le remettre à la chaîne dès qu'un client prioritaire $i + 1$ pénètre dans le système. C'est ce qu'on appelle une priorité absolue connue sous le code PSPO⁹ (Premier Servi, Priorité Obligatoire).

- Priorité non absolue. Si un client d'indice prioritaire $i + 1$ pénètre dans le système, le client de priorité d'indice i qui est en train de se faire servir peut terminer son service avant l'application de la règle de priorité. Il s'agit de priorité non absolue ou relative connue par modèle NPPS (Non Prioritaire, Premier Servi).

Il est aussi assez difficile de manipuler mathématiquement les priorités sauf dans le cas de modèles simples. Le problème des priorités soulève d'intéressantes questions concernant les priorités qui n'ont pas été incorporées dans les modèles d'attente autant qu'ils auraient dû l'être. Ainsi, on pourrait introduire des coûts pour comparer les pertes causées par le maintien d'un client de priorité inférieure déjà en service, à la perte partielle ou totale de service résultant de son rejet si on fait intervenir la règle de priorité absolue.

⁹ Il est rare de rencontrer une situation où les règles de priorités absolues soient en vigueur, ou bien où il soit possible de les mettre en application si cela avait été souhaité, en dehors des domaines de l'électronique et des télécommunications

Il existe encore plusieurs autres règles de discipline qu'on peut voir en service dans nombre de processus d'attente réels (priorité au temps de traitement le plus court, priorité selon la nature du traitement à réaliser, etc.).

La discipline d'une file d'attente présente un intérêt particulier pour le praticien en recherche opérationnelle car les règles qui régissent l'ordre du service sont parmi les aspects d'un processus d'attente les plus faciles à diriger et à modifier.

1 – 4 – Organisation des installations.

Lorsque la station n'est pas unique, et qu'il en existe plusieurs pour fournir un service, l'organisation présente un intérêt certain. L'organisation du service comporte plusieurs aspects. Lorsque plusieurs installations sont mises à la disposition des clients, des files d'attente parallèles vont coexister avec parfois la possibilité pour le client de choisir sa file d'attente, et éventuellement d'en changer. Quand tout client peut, en fait, choisir ou se faire désigner n'importe que guichet nous dirons que ces guichets fournissent une disponibilité totale. Il est très possible qu'il existe plusieurs guichets en service parallèle qui ne fournissent pas complètement une disponibilité totale, c'est-à-dire en d'autres termes qu'ils sont totalement disponibles pour certains clients mais pas pour d'autres. Quand il en aura ainsi, nous parlerons de disponibilité restreinte.

Il peut parfois exister plusieurs stations spécialisées pour des services différents. C'est le cas d'un bureau de poste avec un guichet pour les mandats, un pour les timbres, un pour les envois recommandés, etc.

Il y a aussi des cas où le traitement comporte plusieurs phases de services, c'est-à-dire que le processus de service peut être fractionné en plusieurs éléments. On dispose alors d'une série de stations ; chacune d'elles assurant un élément de service. Dans ce cas, le client devra s'intégrer dans plusieurs files successivement. Il peut avoir à attendre à chaque étape du processus.

Les files d'attentes des clients sont fortement affectés par l'orientation du service et il est important de décider correctement de la meilleure manière d'affecter l'effort de service.

1 – 5 – Conclusion.

Pour conclure ce chapitre nous allons résumer en quelques points essentiels l'algorithme pour l'étude d'un phénomène d'attente.

- Définir convenablement le problème à étudier,
- Observation par échantillonnage de la fréquence des arrivées et de la durée des services,

- Poser des hypothèses sur les distributions des intervalles d'arrivées et des temps de services qui ajustent le mieux les fréquences,
- Vérifier la légitimité de ces hypothèses par un test statistique (généralement le test du khi deux),
- Si les hypothèses sont confirmées, des représentations mathématiques serviront à l'étude analytique du problème. Dans le cas contraire, où les hypothèses sont réfutées, l'étude analytique sera presque toujours impossible et on pourra alors recourir à la simulation qui fera l'objet du chapitre 6.

2 – Modèle de file d’attente à station unique et temps de service exponentielle.

Dans ce chapitre nous allons exposer plusieurs modèles de files d’attente à guichet unique. Nous développons tout d’abord le modèle du processus $M/M/I : (\infty, \text{FIFO})$, ensuite nous nous limiterons à donner les principaux résultats pour les différents modèles $M/M/I : (\infty, \text{FIFO})$ avec priorité, $M/M/I : (\infty, \text{LIFO})$, $M/M/I : (\infty, \text{SIRO})$ et $M/M/I : (x, \text{FIFO})$.

2 – 1 – Modèle $M/M/I : (\infty, \text{FIFO})$.¹⁰

Lorsqu’une seule station fonctionne assez médiocrement, le problème consiste généralement alors d’évaluer l’intérêt qu’il y aurait à le remplacer par un modèle à deux stations. Un intérêt pratique de ce modèle est relativement faible. Si nous développons ce modèle en dépit de son intérêt pratique limité, c’est principalement pour des raisons pédagogiques car il s’agit du plus simple de tous les modèles théoriques.

2 – 1 – 1 – Calcul de la probabilité P_n qu’il y ait n unités dans le système.

Le point de départ de toute étude consiste à rechercher la probabilité qu’il y ait n unités dans le système.¹¹

Cette recherche peut être menée de la manière qui suit.

Désignons par $E(n)$ l’état du système lorsqu’il contient n unités, et associons à chaque état $E(n)$ une probabilité P_n . La correspondance est alors :

$$E(0), E(1), E(2), \dots, E(n-1), E(n), E(n+1)$$

¹⁰ Rappel.

- M : les arrivées s’effectuent au hasard, mais se produisent avec un taux constant λ par unité de temps.
- M : la loi du temps de service est exponentielle avec une valeur moyenne constante de $1/\mu$.
- I : il y a une station unique.
- ∞ : la longueur de la file n’est pas limitée, les clients ne sont pas influencés par cette longueur.
- FIFO : les clients sont servis strictement dans l’ordre d’arrivée.

¹¹ Nous supposons que le phénomène est permanent, c’est-à-dire que la probabilité P_n ne varie pas avec le temps. Ceci n’est vérifié, en général, que dans un certain intervalle de temps qu’il importe de déterminer. Deux causes peuvent provoquer la variation de P_n :

- Une modification du taux moyen des arrivées λ et (ou) celle du taux moyen de service μ .
- Le phénomène transitoire qui se produit entre le début du fonctionnement du service et l’instant où la situation s’est stabilisée. La durée de ce phénomène transitoire est fonction de λ et de μ .

Ainsi, la mesure des paramètres λ et μ doit être rapportée à un intervalle de temps où ces grandeurs peuvent être considérées comme constantes, lequel intervalle de temps doit être aussi sensiblement plus grand que celui nécessaire à la stabilisation. Il serait, pour cela, plus prudent de rendre pour intervalle de mesure une durée supérieure à trois ou quatre fois la durée de la stabilisation de la situation.

$$P_0, P_1, P_2, \dots, P_{n-1}, P_n, P_{n+1}$$

Dire que le phénomène est stationnaire impose que ces probabilités soient indépendante du temps. Or, le système est en constante évolution.

Autour de $E(n)$, par exemple, on constate les possibilités qui suivent.

- Une arrivée :

$E(n)$ devient $E(n + 1)$ ou $E(n + 1)$ devient $E(n)$.

- Une fin de service :

$E(n)$ devient $E(n - 1)$ ou $E(n - 1)$ devient $E(n)$.

- Pas d'arrivée, ni fin de service :

$E(n)$ reste $E(n)$.

La probabilité P_n ne peut être constante que si l'état correspondant $E(n)$ a autant de chances de disparaître que d'apparaître. Il faut donc que la somme des probabilités des deux premières transformations soit égale à la somme des deux dernières, c'est-à-dire qu'il faudrait avoir :

$$P[E(n) \rightarrow E(n + 1)] + P[E(n) \rightarrow E(n - 1)] = P[E(n + 1) \rightarrow E(n)] + P[E(n - 1) \rightarrow E(n)]$$

Nous avons vu dans le chapitre précédent que pour une loi exponentielle de paramètre λ , la probabilité d'une arrivée dans le temps Δt . Or la probabilité de l'état $E(n)$ était P_n , la probabilité que $E(n)$ devienne $E(n + 1)$ sera $P_n \lambda \Delta t$.

Appliquant ce raisonnement à tous les termes de l'équation, on aura :

$$P_n \lambda \Delta t + P_n \mu \Delta t = P_{n+1} \mu \Delta t + P_{n-1} \lambda \Delta t$$

D'où la relation, après simplification :

$$\mu \cdot P_{n+1} - (\lambda - \mu) P_n + \lambda \cdot P_{n-1} = 0$$

Ecrivant cette formule plusieurs fois, jusqu'à la dernière qui est :

$$\mu \cdot P_2 - (\lambda - \mu) P_1 + \lambda \cdot P_0 = 0$$

Si nous remarquons que les changements relatifs à l'état $E(0)$ ne peuvent se faire qu'avec $E(1)$, le même raisonnement donne :

$$P_0 \cdot \mu \Delta t = P_1 \cdot \mu \Delta t$$

Et finalement :

$$\mu \cdot P_{n+1} = \lambda \cdot P_n$$

Cette formule conduit alors à :

$$P_n = (\lambda/\mu)^n \cdot P_0$$

Il ne reste plus qu'à déterminer P_0 .

Comme $\sum_0^\infty P_n = 1$, on aura :

$$P_0 + (\lambda/\mu) \cdot P_0 + (\lambda/\mu)^2 \cdot P_0 + \dots + (\lambda/\mu)^n \cdot P_0 + \dots = \frac{P_0}{1 - \lambda/\mu} = 1 \text{ puisque } P_0 = 1 - \lambda/\mu$$

La probabilité cherchée est donc :

$$P_n = (\lambda/\mu)^n \cdot (1 - \lambda/\mu)$$

Ou en passant par $\rho = \lambda/\mu$:

$$P_n = \rho^n (1 - \rho)$$

Nous remarquons que le paramètre ρ caractérise le phénomène d'attente. Il est appelé intensité du trafic, ou facteur d'utilisation qu'on mesure en 'unité internationale de trafic' ou 'Erlang' (parfois même on ne lui attribue aucune unité).

Toutes les probabilités étant exprimées en fonction de ce paramètre, il devient possible de leur donner une valeur chiffrée. Si, par exemple, le système enregistre en moyenne huit arrivées à l'heure et si dix traitement par heure en moyenne sont réalisables, ce qui correspond à une intensité du trafic de $\rho = \lambda/\mu = 8/10 = 0,8$.

$$P_0 = 0,2 ; P_1 = 0,16 ; P_2 = 0,128 ; P_3 = 0,1024 ; \dots$$

Un tel système signifie qu'il :

- sera inutilisé 20% du temps, puisque $P_0 = 0,2$
- sera utilisé, c'est-à-dire caractérisé par la présence d'au moins un client, 80% du temps ($1 - P_0 = 0,8$)
- provoquera une file d'attente d'individus dont le traitement n'a pas encore commencé 64% du temps ($1 - P_0 - P_1 = 0,64$).

A partir de ces probabilités qui décrivent le système d'une façon un peu abstraite, il convient d'évaluer les caractéristiques de la file d'attente.

Concrètement, une file d'attente se traduit par un certain nombre d'unités qui perdent un temps plus ou moins long avant d'obtenir le service demandé.

Il est important pour le gestionnaire de la file d'attente de déterminer, en fonction de l'intensité du trafic, de connaître les grandeurs qui suivent.

- Le nombre moyen d'unités présents dans le système \bar{n}
- Le temps moyen d'attente dans le système \bar{t}
- Le nombre moyen d'unités dans la file d'attente, c'est-à-dire attendant un service \bar{h}
- Le temps moyen d'attente dans la file d'attente, c'est-à-dire avant d'être servis \bar{t}' .
- La probabilité pour qu'il y ait moins de $n + 1$ unités dans le système :

$$\sum_0^n P_n = \sum_0^n \rho^n (1 - \rho) = 1 - \rho^{n+1}$$

Nous allons considérer chacun de ces cas.

2 – 1 – 2 – Nombre moyen d'unités dans le système \bar{n}

Le nombre moyen d'unités présents dans le système s'évalue comme l'espérance mathématique de n :

$$\bar{n} = E(n) = \sum_{n=0}^{\infty} n \cdot P_n = \sum_{n=1}^{\infty} n \cdot P_n$$

Si l'on utilise les formules développées précédemment pour P_n , cette espérance mathématique s'exprime en fonction de l'intensité du trafic :

$$\bar{n} = \sum_{n=1}^{\infty} n \rho^n (1 - \rho)$$

Elle peut s'écrire également, après avoir mis $(1 - \rho)$ en facteur :

$$\bar{n} = \sum_{n=1}^{\infty} n \rho^{n-1} (1 - \rho) \quad \text{Voir la différence avec la formule précédente}$$

On remarque que le terme entre crochet (parenthèses ?) est la dérivée par rapport à ρ de

$$\sum_{n=1}^{\infty} \rho^n, \text{ soit encore de } \frac{1}{1 - \rho}.$$

Cette dérivée est égale à $\frac{1}{(1 - \rho)^2}$ donc $\bar{n} = \frac{1}{(1 - \rho)^2} \rho \cdot (1 - \rho)$, et après simplification :

$$\bar{n} = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda}$$

Dans le cas examiné précédemment où l'intensité du trafic s'élevait à 0,8, il y a en moyenne quatre unités dans le système.

2 – 1 – 3 – Temps moyen d'attente dans le système \bar{t}

La file d'attente est alimentée par des arrivées qui se produisent au taux λ . Le fait qu'il y ait en

moyenne \bar{n} unités dans le système implique que chacun y reste en moyenne un temps $\bar{t} = \frac{\bar{n}}{\lambda}$

\bar{t} peut être exprimé en fonction de l'intensité du trafic, avec $\bar{t} = \frac{\rho}{\lambda (1 - \rho)}$, et aussi sous la

forme :

$$\bar{t} = \frac{1}{\mu - \lambda}$$

Ainsi, huit arrivées à l'heure et un nombre moyen \bar{n} de quatre unités doivent être associées à un séjour d'une durée moyenne de 0,5 heure par personne, soit 30 minutes.

2 – 1 – 4 – Nombre moyen d'unités dans la file d'attente \bar{h}

Parmi les unités présentes dans le système, il convient de faire la distinction entre ceux qui sont en cours de traitement et ceux qui constituent la file d'attente proprement dite. Seuls ces derniers perdent réellement du temps.

Lorsqu'une seule station fonctionne, une seule unité peut être servie à la fois. Sur n personnes séjournant dans le système, il y a ainsi $(n - 1)$ qui attendent (sauf évidemment si $n = 0$). Le nombre moyen d'unités qui attendent est alors donné par :

$$\bar{h} = \sum_{n=1}^{\infty} (n - 1) \cdot P_n = \sum_{n=1}^{\infty} n \cdot P_n - \sum_{n=1}^{\infty} P_n \quad \text{et} \quad \bar{h} = n' - \rho.$$

Le nombre moyen d'unités qui attendent est ainsi égal au nombre moyen d'unités dans le système moins le nombre moyen d'unités en cours de traitement.

n' peut être obtenu directement à partir de l'intensité du trafic avec la formule qui suit.

$$\bar{h} = \bar{n} \cdot \rho = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho}$$

$$\text{Et qui s'écrit également : } n' = \frac{\lambda^2}{\mu - \lambda}$$

Avec les chiffres précédents ($\rho = 0,8$ et $\bar{n} = 4$), nous arrivons ainsi à $\bar{h} = 3,2$ personnes en situation d'attente pure.

Le temps d'attente avant d'être servi est calculé de la même manière que le temps de séjour dans le système. La file des unités en attente pure est alimentée par les arrivées au taux λ .

Comme cette file s'élève en moyenne à \bar{h} ceci implique que l'attente moyenne s'élève à :

$$\bar{t}' = \frac{\rho^2}{\lambda(1 - \rho)} = \frac{\rho}{\mu(1 - \rho)}$$

La dernière expression nous donne une autre façon d'interpréter t' . Quand un nouveau client arrive, il se trouve en moyenne derrière \bar{h} unités déjà dans le système. Le temps de traitement

moyen s'élevait à $\frac{1}{\mu}$, il faut un temps total $\frac{\bar{h}}{\mu}$ pour que les individus présents dans le système

soient servis puis sortent et donc pour que le nouveau client atteigne le poste de traitement.

On vérifie également à partir des formules développées par t' et \bar{t}' que le temps de séjour dans le système est égal au temps d'attente ajouté au temps de traitement : $t' = \bar{t}' + \frac{1}{\mu}$

Si $\lambda = 0,8$ et $\mu = 10$ (nous savons que $\bar{n} = 4$, $\bar{t} = 0,5$ et $\bar{h} = 3,2$), le temps moyen d'attente est égal à :

$$\bar{t}' = \frac{\bar{h}}{\lambda} = 3,2/8 = 0,4.$$

$$\bar{t}' = n/\mu = 4/10 = 0,4.$$

$$\bar{t}' = t' - 1/\mu = 0,5 - 0,1 = 0,4.$$

Nous pouvons résumer les principaux résultats calculés dans le tableau qui suit.

Tableau 1. Modèle M/M/1 : (∞ , FIFO)

- Loi des arrivées : loi de Poisson de moyenne λ
- Loi de service : loi exponentielle de moyenne $1/\mu$
- Nombre de station : 1
- Règle de priorité : FIFO
- Intensité du trafic = $\rho = \lambda/\mu$
- Probabilité qu'il n'y ait aucun client dans la file d'attente : $P_0 = 1 - \rho$
- Probabilité qu'il y ait n unités dans le système : $P_n = \rho^n (1 - \rho)$
- Nombre moyen de clients n dans le système : $n = \frac{\rho}{1 - \rho}$
- Nombre moyen de clients n' dans la file d'attente : $n' = \frac{\rho^2}{1 - \rho}$
- Temps moyen d'attente dans le système : $\bar{t} = \frac{\rho}{\lambda (1 - \rho)}$
- Temps moyen d'attente dans la file d'attente : $\bar{t}' = \frac{\rho}{\mu (1 - \rho)}$

2 – 2 – Modèle M/M/1 : avec priorité (deux files d'attente dont l'une est prioritaire).

Ce type de modèle où la discipline des files d'attente est régie par des règles de priorité se rencontre le plus souvent dans des lignes de production, ainsi que dans des activités de réparation et d'entretien.

Il existe des cas de priorité relative (ou priorité non absolue) et de priorité absolue.

2 – 2 – 1 – Cas de priorité non absolue.

Considérons deux catégories de clients C_1 et C_2 tel que tous les clients C_1 qui arrivent, passent avant tous les clients C_2 présents, mais passent seulement après les autres clients C_1 qui seraient déjà dans la station en train de se faire servir, c'est-à-dire ceux qui sont arrivés avant lui.

Posons α pour la proportion de clients 1 et $(1 - \alpha)$ pour la proportion de clients C_2 . On a alors les différentes expressions qui suivent.

- Nombre moyen d'unités dans le système \bar{n} et \bar{n}_2

$$\bar{n}_1 = \alpha \cdot \rho \cdot \left(\frac{1 + \rho - \lambda \rho}{1 - \alpha \rho} \right) \quad \text{et} \quad \bar{n}_2 = (1 - \alpha) \cdot \rho \cdot \left(\frac{1 - \alpha \rho + \alpha \rho^2}{(1 - \rho)(1 - \alpha \rho)} \right)$$

- Nombre moyen d'unités dans les files d'attente \bar{n}_1^f et \bar{n}_2^f

$$\bar{n}_1^f = \frac{\alpha \rho^2}{1 - \alpha \rho} \quad \text{et} \quad \bar{n}_2^f = \frac{(1 - \alpha) \rho^2}{(1 - \rho)(1 - \alpha \rho)}$$

- Temps moyen d'attente dans les files d'attente \bar{t}_1^f et \bar{t}_2^f

$$\bar{t}_1^f = \frac{\rho}{\mu - \alpha \rho} \quad \text{et} \quad \bar{t}_2^f = \frac{\rho}{(1 - \rho)(\mu - \alpha \rho)}$$

On remarquera que le rapport $\frac{\bar{t}_1^f}{\bar{t}_2^f} = \frac{1}{\mu} = 1 - \rho$ (rappel : $\rho = \frac{\lambda}{\mu}$) des durées moyennes d'attente

pour les deux catégories de clients est indépendant de α . Présenté sous la forme $(1 - \rho)$ on constate que si l'intensité du trafic approche l'unité, la file d'attente de la catégorie C_2 est complètement écartée alors que pour les clients C_1 on tend vers une sorte d'état limite.

A partir de ces résultats, il est simple de généraliser la priorité à deux catégories aux cas de priorité à catégorie multiple. Ainsi, on aura :

$$\bar{t}_{qj} = \frac{\frac{1}{2} \cdot \sum_{i=1}^k \alpha_i \cdot C_i}{(1 - \sum_{i=1}^j \frac{\alpha_i}{\mu_i}) \cdot (1 - \sum_{i=1}^j \frac{\alpha_i}{\mu_i})}$$

Avec :

- α_i est la proportion de clients appartenant à la catégorie C_i ,
- μ_i est le taux de service moyen de ces mêmes clients,
- C_i est le moment d'ordre deux de la loi de distribution du temps de service pour les membres de C_i .

Ces expressions peuvent être utilisées si l'on peut affecter aux temps d'attente des pénalités de valeurs proportionnelles aux durées moyennes de queues des différentes catégories de priorité, alors l'attribution des priorités devra se faire dans l'ordre des rapports

$$\frac{\text{Temps de service moyen}}{\text{Pénalité par unité de temps d'attente}}$$

où à la plus faible valeur correspondra la plus forte priorité.

2 – 2 – 2 – Cas de priorité absolue.

C'est un cas qui se rencontre très rarement lorsque les clients de la file d'attente sont des êtres humains. Il est surtout employé dans les grands ensembles de calculs électroniques digitaux travaillant à distance sur ligne. Il n'est pas rare pour des systèmes travaillant sur lignes d'avoir à faire face à une grande variété de travaux. Certains sont plus vitaux que d'autre et ont une priorité absolue de façon que s'il est besoin d'en affecter un, l'unité centrale enregistre sur disque le programme sur lequel elle travaillait et continue immédiatement avec la haute priorité numéro 1.

Considérons dans ce cas aussi deux catégories de clients C_1 avec une proportion α et C_2 avec une proportion $(1 - \alpha)$. Les clients de la classe C_1 se placent avant tout client de la classe C_2 , même si celui-ci est en cours de traitement.

Dans les systèmes à priorité absolue, les clients de priorité la plus élevée ne sont absolument pas influencés par ceux qui ont une priorité plus basse. Nous aurons dans ce cas avec pour les temps de service μ_1 et μ_2 qui suivent une loi exponentielle les expressions qui suivent.

- Probabilité qu'il y ait n clients de C_1 dans le système :

$$P_{1n} = \rho_1^n (1 - \rho_1) \quad \text{avec } n = 0, 1, 2, \dots \text{etc. et } \rho_1 = \alpha \cdot \frac{\lambda}{\mu_1}$$

- Nombre moyen de clients de C_2 présent dans le système :

$$\bar{n} = \frac{\rho_2}{1 - \rho_1 - \rho_2} \left[1 + \frac{\rho_1 \mu_2}{(1 - \rho_1) \mu_1} \right]$$

- Temps d'attente moyen dans la file \bar{t} et \bar{t}' :

$$\bar{t} = 1 + \frac{1}{(1 - \rho) \mu_1} \quad \text{et} \quad \bar{t}' = \frac{1}{\mu_2 (1 - \rho_1 - \rho_2)} \left[1 + \frac{\rho_1 \mu_2}{(1 - \rho_1) \mu_1} \right]$$

- Temps moyen d'attente dans les systèmes \bar{t}_1 et \bar{t}_2 :

$$\bar{t}_1 = \frac{\rho_1}{\mu_1 (1 - \rho_1)} \quad \text{et} \quad \bar{t}_2 = \frac{1}{\mu_2 (1 - \rho_1 - \rho_2)} \left[\rho_1 + \rho_2 + \frac{\rho_1 \mu_2}{(1 - \rho_1) \mu_1} \right]$$

2 – 3 – *Modèle M/M/1 : (∞ , LIFO).*

La loi de probabilité du nombre de clients dans la file est la même pour ce processus que pour le cas précédent (le cas M/M/1 : ∞ , FIFO), c'est-à-dire en état d'équilibre $P_n = \rho^n (1 - \rho)$.

La loi de distribution des temps d'attente est par contre différente. Si nous pouvons juger de file d'attente FIFO comme la plus correcte vis-à-vis des clients, la moins correcte est la discipline LIFO (dernier entré, dernier sorti). La loi des temps d'attente dans ce cas nécessite une analyse assez élaborés.

Supposons qu'un client arrive et prenne la file. Son temps d'attente ne sera pas affecté par le nombre de personnes qui y sont déjà, puisque selon la règle FIFO, il sera servi avant qu'ils ne le soient. Ce temps dépend seulement du nombre de clients qui arrivent après lui et avant que son service commence.

En se rappelant la loi de distribution des périodes d'activité que nous avons trouvée dans le cas du modèles M/M/1 : (∞ , FIFO), et après une courte réflexion on constatera que la loi de distribution des périodes d'activité est la même dans le cas du modèle avec discipline LIFO. La démonstration en est la suivante ; un client, le sujet, trouve en arrivant n personnes en train d'attendre, y compris celle qui se fait servir. Il va au début de la file.

Pendant le temps qui s'écoule entre son arrivée et l'achèvement du service en cours, d'autres clients peuvent arriver ; si c'est le cas, ils le précèdent dans la file d'attente. En d'autres termes, notre sujet commence à être servi au moment où un client partira en laissant n clients dans la file d'attente. Ainsi, la loi de distribution des temps d'attente dans le modèle M/M/1 : (∞ , FIFO) est la loi de distribution d'une période d'activité qui commencerait quand il y a n personnes dans la file d'attente et qui finirait quand la longueur de la file retombe à nouveau à n. Des périodes de ce genre sont indépendantes de n, donc les périodes d'activité sont les mêmes pour les règles FIFO et LIFO. Du fait de la distribution exponentielle des temps de service, le temps qui s'écoule entre l'arrivée de notre sujet et la fin du service en cours doit aussi avoir une distribution exponentielle. Donc, l'instant réel auquel notre sujet arrive n'a pas d'importance. Le temps d'attente est précisément égal à une période d'activité produite par un temps de service complet ; c'est naturellement la même chose qu'une période d'activité.

Dans ce cas la durée moyenne d'une période d'activité est :

$$\bar{t} = \frac{1}{\mu(1-\rho)}$$

Nous remarquons que le temps d'attente moyen d'un modèle M/M/1 est indépendant de la discipline de la file d'attente.

2 – 4 – Modèle M/M/1 : (∞ , SIRO).

Il s'agit de cas où les clients sont impatients et quittent la file lorsque leur temps d'attente dépasse une certaine période t_0 . Les clients sont choisis au hasard dans la file d'attente au lieu de servir le premier. Il n'est pas simple dans ce cas de déduire la loi des temps d'attente de la loi des longueurs de file. Comme exemple de cette situation nous pouvons considérer des clients dans une station d'autobus aux heures de pointe. Souvent, pour ne pas dire toujours, les files ne sont pas respectées et ce n'est malheureusement à toujours le premier arrivé qui monte dans l'autobus le premier lors de son arrivé dans la station.

La probabilité qu'un client arbitraire doive attendre au moins un temps t est donné par la formule qui suit.

$$(1 - \rho) \sum_{n=0}^{\infty} \rho^{n+1} \cdot q_n(t)$$

Où $q_n(t)$ est la probabilité qu'un client donnée qui a trouvé en arrivant n autres clients en train d'attendre, soit encore dans la file un temps t plus tard. Nous supposons cette probabilité connu.

La probabilité P_n qu'il y ait n unités dans le système est :

$$P_n = \frac{\rho^n P_0}{\mu \prod_{k=1}^n (1 - e^{-\mu t_k})}$$

Il peut être démontré que le temps moyen d'attente dans la file d'attente est :

$$\bar{t} = \frac{\rho}{\mu(1 - \rho)}$$

Ce temps est le même dans le modèle de règle FIFO ou LIFO. Donc, on le constat encore une fois que la discipline de la file d'attente n'affecte pas la valeur moyenne de la durée d'attente.

La probabilité qu'une unité qui doit de toute façon attendre, l'attente dans la file plus longtemps que t est donné par l'approximation de Riordon :

$$\frac{1}{2} \alpha e^{-(\alpha\mu - \alpha\lambda)t} - \frac{1}{2} \beta \cdot e^{-\beta(\mu - \lambda)t} \quad \text{avec} \quad \alpha = 1 + \sqrt{\frac{1-\rho}{1-\rho}} \quad \text{et} \quad \beta = \sqrt{\frac{\rho}{1-\rho}}$$

La démonstration de ces différentes formules étant longue et difficile, nous avons jugé inutile de les décrire dans ce cours.

2 – 5 – M/MI : (x , FIFO).

Il s'agit d'une situation où le nombre de clients limité. C'est le cas d'un parc de machines où le nombre de ces dernières en panne est limité au nombre de machines.

Si nous considérons la capacité limitée à x (y compris celui en cours de traitement) et $\lambda \neq \mu$, on aura les principaux résultats qui suivent.

- Evaluation de la probabilité qu'il n'y ait aucun client dans le système, c'est-à-dire qu'il soit vide P_0 :

$$P_0 = \frac{1-\rho}{1-\rho^{x+1}}$$

- Evaluation de la probabilité qu'il y ait n unités dans le système P_n :

$$P_n = P_0 \rho^n$$

- Evaluation du nombre de clients présent dans le système \bar{n}

$$\bar{n} = \frac{\rho}{1-\rho} \left[\frac{1-(x+1)\rho^x + x\rho^{x+1}}{1-\rho^{x+1}} \right]$$

- Evaluation du nombre moyen de clients en attendant un service \bar{n}' :

$$\bar{n}' = \frac{\rho^2}{1-\rho} \left[\frac{1-x\rho^{x+1} + (x+1)\rho^x}{1-\rho^x} \right]$$

- Evaluation de la durée moyenne dans la file \bar{t} :

$$\bar{t} = \frac{\rho}{\mu(1-\rho)} \left[\frac{1-x\rho^{x+1} + (x+1)\rho^x}{1-\rho^x} \right]$$

- Evaluation de la durée moyenne d'attente dans le système \bar{t}' :

$$\bar{t}' = \bar{t} + \frac{1}{\mu}$$

Dans le cas où $\lambda = \mu$, on aura :

$$P_n = \frac{1}{x+1} \quad \bar{n} = \frac{x}{2} \quad \text{et} \quad \bar{t}' = \frac{x+1}{2\mu}$$

3 – Modèle de file d'attente à station unique et temps de service exponentielle.

On considère en général que les lois des arrivées sont poissonniennes et les lois de service sont exponentielles car on a pu vérifier par expérience que ces lois sont les plus fréquemment vérifiées par les phénomènes d'attente. Cependant, il existe toutefois de nombreux modèles où la loi des services n'est pas exponentielle. Nous allons présenter quelques cas, en nous limitant à donner les formules générales sans démonstration.

3 – 1 – Modèle $M/E_k/1$: (∞ , FIFO) appelé aussi modèle d'Erlang.

Plusieurs lois non exponentielle de temps de service peuvent être ajusté par la loi de distribution qui, selon le lieu, le moment et la forme qu'elle prend est connue sous le nom de loi d'Euler, de Pearson type III, de khi-deux, d'Erlang ou de loi Gamma.

Nous utiliserons dans notre cas la distribution d'Erlang.

Comme nous l'avons vu (chapitre 1, page 20), c'est la distribution de la somme K variables obéissant chacune à une loi exponentielle de moyenne $1/\lambda K$. Erlang s'est rendu compte qu'un mécanisme de service qui obéit à une loi de ce type peut être envisagé comme s'il comportait K étapes en série, chaque étape étant un service exponentielle. Les équations à résoudre pour un modèle qui peut être élaboré sur la base d'éléments exponentielle, sont de beaucoup simplifié.

Les règles qui gouvernent le déroulement d'un mécanisme de service à K phases sont les suivantes.

- Le service ne peut être fourni qu'à un seul client à la fois par station.
- Chaque client doit passer par toutes les phases de service en commençant par la première.
- Dès l'achèvement d'une phase, le client se déplace immédiatement vers la suivante.
- Quand un client quitte la $K^{\text{ième}}$ phase, le client suivant peut aborder la première, mais pas avant.

Il est à remarquer que ce cas est différent de K stations en série, car un client seulement peut être servi à un instant donné, alors que pour K guichets en série l'on peut voir jusqu'à K clients concernés en même temps par le mécanisme de service. C'est précisément à cause de la première règle que nous pouvons considérer les K phases comme équivalentes à une seule opération.

La loi d'Erlang est donnée par la densité de probabilité :

$$F(t) = \frac{(\lambda k)^k}{(k-1)!} \cdot e^{-\lambda k t} \cdot t^{k-1} = \frac{k\lambda(k\lambda t)^{k-1} \cdot e^{-k\lambda t}}{(k-1)!} \quad \text{où } k \text{ est un entier positif.}$$

Elle correspond au cas où, à partir d'arrivées poissonniennes on accepte dans le service les unités 0, k, 2k, 3k, ..., etc. Tapez une équation ici, c'est-à-dire que les intervalles entre deux événements sont constants. La variance $\frac{1}{k\lambda^2}$ est alors nulle.

Si k tend vers l'infini, on retrouve la distribution uniforme, et si k = 1 on retrouve la distribution exponentielle.

Nous aurons :

- La moyenne des temps de service : $S = \frac{1}{\mu}$

- L'écart-type des temps de service : $\sigma_s = \frac{1}{\mu\sqrt{k}}$

- Le nombre moyen de clients attendant un service : $\bar{n} = \frac{k+1}{2k} \frac{\rho^2}{1-\rho}$

- Le nombre moyen de clients présents dans le système : $\bar{n} = \bar{n} + \rho$

- La durée moyenne d'attente dans la file : $\bar{t} = \frac{k+1}{2k} \frac{\rho}{\mu(1-\rho)}$

- La durée moyenne d'attente dans le système : $\bar{t} = \bar{t} + \frac{1}{\mu}$

3-2 - Modèle M/D/1 : (∞ , FIFO).

Ce modèle est assez utile, car il y a beaucoup de situations réelles dans lesquelles le coefficient de variation de la distribution des temps de service est assez petit, si bien que les temps de service peuvent effectivement être considérés comme uniforme, c'est-à-dire que le service dure exactement $1/\mu$ de temps.

Nous aurons les formules qui suivent.

- Un temps de service constant qui est égal à $1/\mu$.

- La probabilité qu'il y ait aucun client dans le système : $P_0 = 1 - \rho$

- La probabilité qu'il y ait n client dans le système :

$$P_n = (1 - \rho) \sum_{k=1}^n (-1)^{n-k} \cdot e^{k\rho} \left[\frac{(k\rho)^{n-k}}{(n-k)!} + \frac{(k\rho)^{n-k-1}}{(n-k-1)!} \right]$$

Il s'agit du cas où n est supérieur ou égal à deux.

- Le nombre moyen de clients en attente dans la file : $\bar{n} = \frac{\rho^2}{2(1-\rho)}$

- Le nombre moyen de clients présents dans le système : $\bar{n} = \bar{n} + \rho$

- La durée moyenne d'attente dans la file : $\bar{t} = \frac{\rho}{2\mu(1-\rho)}$

On peut remarquer que la durée est deux fois plus petite que dans le cas où les arrivées sont aléatoires (discipline SIRO) et le service distribué selon la loi exponentielle.

- La durée moyenne d'attente dans le système : $\bar{t} = \bar{t} + \frac{1}{\mu}$

On remarquera que ce modèle est équivalent à celui de M/Ek/I : (∞ , FIFO) pour $k = \infty$.

3 – 3 – *Modèle M/G/I : (∞ , FIFO).*

Ce modèle correspond à une distribution des arrivées de type Poisson, celle du temps de service quelconque, la capacité du système illimité et une règle de priorité FIFO.

Dans ce cas l'écart-type des temps de service σ_s sera déterminé empiriquement.

On obtient les formules qui suivent.

- La moyenne des temps de service : $S + \frac{1}{\mu}$

- La probabilité qu'il y ait aucun client dans le système : $P_0 = 1 - \rho$

- Le nombre moyen de clients en attente dans la file : $\bar{n} = \frac{\rho^2}{2(1-\rho)} [1 + \mu^2 \sigma_s^2]$

- Le nombre moyen de clients présents dans le système : $\bar{n} = \bar{n} + \rho$

- La durée moyenne d'attente dans la file : $\bar{t} = \frac{1}{2\mu(1-\rho)} \times [1 + \mu^2 \sigma_s^2]$

- La durée moyenne d'attente dans le système : $\bar{t} = \bar{t} + \frac{1}{\mu}$

Ce cas est particulièrement intéressant dans la mesure où il existe de nombreuses situations où la distribution des temps de service ne s'ajuste pas à une loi connue.

4 - Modèle de file d'attente à plusieurs stations.

Contrairement aux modèles précédents où l'intérêt pratique était relativement faible, les modèles à plusieurs stations présentent une plus grande application dans la pratique.

Nous considérerons tout d'abord dans ce chapitre le modèle $M/M/C : (\infty, \text{FIFO})$ (appelé modèle d'Erlang construit par lui) qui constitue l'un des modèles à guichets multiples le plus simple. Ensuite, nous présenterons très succinctement le modèle $M/D/C : (\infty, \text{FIFO})$, $M/E_k/C : (\infty, \text{FIFO})$ et $M/M/2 : (\infty, \text{GD})$.

4 - 1 - Modèle $M/M/C : (\infty, \text{FIFO})$.

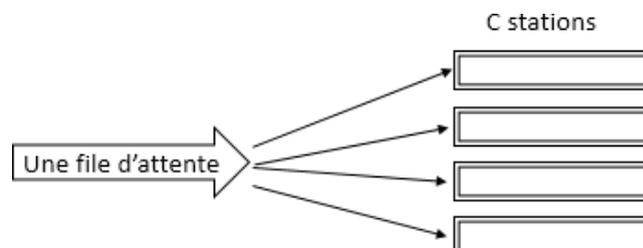
Mis à part le nombre de stations qui diffère, toutes les données du modèle $M/M/1 : (\infty, \text{FIFO})$ sont conservées, c'est-à-dire :

- Les arrivées se font selon un processus de Poisson de taux λ .
- Les durées des services suivent une loi exponentielle de taux μ .
- Il y a une seule file d'attente commune aux stations, ou ce qui est équivalent, s'il y a plusieurs files, les unités changent de file de manière à se trouver toujours dans la file la plus courte.
- Il n'y a pas de priorité entre les clients. Les stations ont la même loi de service et il n'y a pas de préférence de la part des clients : ils se présentent toujours devant la station qui se libère.
- La longueur de la file d'attente est illimitée.

Aussi la condition d'un taux d'arrivée inférieur au taux de service dans le cas précédent ($M/M/1 : (\infty, \text{FIFO})$) est remplacée par la condition que le taux d'arrivée soit inférieur au produit du taux de service par le nombre de station c'est-à-dire $\lambda < \mu C$.

Ce modèle se présente schématiquement comme ci-dessous.

Figure 3. Représentation schématique du modèle d'Erlang.



Bien que cette situation présente beaucoup de similitude avec le cas où il y a une seule station, il est nécessaire de préciser qu'il en fait très différent. En effet, il serait absolument faux de supposer que ce modèle peut se ramener à une seule station débitant C fois plus vite, ou que

chaque station reçoit un débit équivalent à celui d'une station divisée par C ; Le temps de traitement d'un client n'est pas accéléré du fait que plusieurs stations sont disponibles ; il est toujours égal en moyenne à $\frac{1}{\mu}$.

Dans ce cas, nous procéderons comme précédemment, en évaluant tout d'abord la probabilité qu'il y ait n personnes dans le système, puis les différentes caractéristiques de la file d'attente. Pour le calcul de la probabilité P_n qu'il y ait n personnes dans le système, il faudrait distinguer deux cas ; celui où n est supérieur à C (c'est-à-dire le nombre de station), dans ce cas toutes les stations sont occupées et la probabilité pour qu'un service s'achève pendant l'intervalle Δt est $C \cdot \mu \Delta t$; et le cas où n est inférieur ou égal à C , c'est-à-dire que certaines stations vont rester inoccupées ($C - n$) ; la probabilité qu'une sortie soit enregistrés pendant l'intervalle Δt est dans ce cas $n \cdot \mu \Delta t$.

Nous allons calculer les probabilités de ces deux situations.

- $n \geq C$

On peut aussi dans ce cas, en admettant que le phénomène est stationnaire, que les chances de disparition de l'état $E(n)$ sont égale à ses chances d'apparition d'où la relation de base qui suivante :

$$\lambda P_n + C \mu P_n = \lambda P_{n-1} + C \mu P_{n+1}.$$

- $n \leq C$

Pour $N = 0$, les modifications d'état ne peuvent se faire qu'entre $E(0)$ et $E(1)$ puis, pour $E(1)$ une seule station est occupée, d'où la relation suivante :

$$\mu P_1 = \lambda P_0.$$

Pour $n = 1$, on trouve l'égalité :

$$\lambda P_1 + \mu P_1 = \lambda P_0 + 2\mu P_2$$

Qui se ramène, en considérant la précédente à :

$$\lambda P_1 = 2\mu P_2$$

Et ainsi, de proche en proche nous aurons :

$$n = 2 \rightarrow \lambda P_2 = 3\mu P_3$$

.....

$$n = i - 1 \rightarrow \lambda P_{i-1} = i\mu P_i$$

.....

$$n = C - 1 \rightarrow \lambda P_{C-1} = C\mu P_C$$

Et finalement :

$$P_i = \frac{\rho^i}{i!} \times P_0 \quad \text{avec } i < C \quad \text{et} \quad P_C = \frac{\rho^C}{C!} \times P_0$$

En considérant les formules de base pour $n \geq C$, faisant la somme de ces équations jusqu'à $n = C$ et remarquant que $\lambda P_{C-1} = C\mu P_C$, on aura :

$$P_n = \frac{\rho^C}{C!} \cdot P_0 \quad \text{et} \quad P_n = \left[\frac{\lambda}{C\mu} \right]^{n-C} \cdot P_C = \frac{\rho^{n-C}}{C^{n-C}} \cdot P_C \quad \text{mais} \quad P_C = \frac{\rho^C}{C!} \cdot P_0$$

Donc on aura :

$$P_n = \frac{\rho^n}{C^{n-C} \cdot C!} \cdot P_0 \quad \text{avec } n \geq C.$$

Quant à P_0 , on l'obtient comme précédemment en écrivant :

$$\sum_0^\infty P_n = 1$$

$$\sum_0^\infty P_n = \sum_0^{C-1} P_n + \sum_C^\infty P_n = P_0 \sum_0^{C-1} \frac{\rho^i}{i!} + P_C \sum_0^\infty \frac{\rho^{n-C}}{C^{n-C}} = P_0 \sum_0^{C-1} \frac{\rho^i}{i!} + P_0 \frac{\rho^C}{C!} \sum_0^\infty \frac{1}{C^{n-C}}$$

D'où :

$$P_0 = \frac{1}{1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^{C-1}}{(C-1)!} + \frac{\rho^C}{C!} \sum_{n=C}^\infty \frac{1}{C^{n-C}}}$$

Ou bien en résumant :

$$P_0 = \frac{1}{\sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C}{C!(1-\frac{\rho}{C})}}$$

Ayant calculé la probabilité P_n , nous pouvons décrire les principales caractéristiques de ce modèle comme suit.

- Nombre moyen de clients en attente dans la file d'attente \bar{n} :

Lorsque C stations de service sont ouvertes, pour qu'un client attende, il faut que n soit supérieur à C ; ainsi le nombre moyen de clients en attente est donné par l'espérance mathématique de $n - C$.

$$\bar{n} = E(n - C) = \sum_{n=C}^\infty (n - C) P_n$$

En remplaçant $(n - C)$ par k , on obtient :

$$\bar{n} = E(k) = \sum_{k=0}^{\infty} k \cdot P_{c+k} \quad k \cdot \frac{\rho^{c+k}}{k! C^k} \times P = \frac{\rho^{c+1}}{C! C} \times P \sum_{k=0}^{\infty} k \cdot \frac{\rho^{k-1}}{k! C^{k-1}}$$

En remplaçant P_{c+k} par l'expression développée précédemment (numéroter les formules pour simplifier l'écriture), il vient :

$$\bar{n} = \sum_{k=0}^{\infty} \frac{C! C^k}{C! C} \cdot 0 \quad 0 \quad k=0 \quad C$$

L'expression est la dérivée de la somme d'une progression géométrique de raison $\frac{\rho}{C}$, que l'on remplacera par $\frac{1}{(1-\frac{\rho}{C})^2}$, et l'on obtient :

$$\bar{n} = \frac{\rho^{c+1}}{C! C (1-\frac{\rho}{C})^2} \times P_0$$

- Temps d'attente moyen dans la file.

Avant tout traitement, il est donné par $\frac{n^f}{\lambda}$

Soit
$$\bar{t} = \frac{\rho^c}{\mu C! C (1-\frac{\rho}{C})^2} \times P_0$$

- Temps d'attente moyen dans le système.

Chaque client passant un temps moyen \bar{t} à attendre et recevant un service dont la durée

moyenne est $1/\mu$, aura un temps moyen de présence dans le système de : $\bar{t} = \bar{t}' + 1/\mu$ d'où :

$$\bar{t} = \frac{\rho^c}{\mu C! C (1-\frac{\rho}{C})^2} \times P_0 + \frac{1}{\mu}$$

- Nombre moyen de client présents dans le système.

Les arrivées s'effectuent au taux λ , d'où un nombre moyen de clients dans le système de :

$$\bar{n} = \bar{t} \cdot \lambda \quad \text{soit} \quad \bar{n} = \left[\frac{\rho^c}{\mu C! C (1-\frac{\rho}{C})^2} \cdot P_0 + \frac{1}{\mu} \right] \cdot \lambda$$

Il y a aussi, en moyenne, ρ clients en cours de traitement. Dans la mesure où C stations de traitement sont disponibles, $C - \rho$ restent en moyenne inutilisées.

Nous allons résumer dans un tableau les principales caractéristiques concernant la file d'attente de type $M/M/C : (\infty, \text{FIFO})$.

Tableau 2. Synthèse des caractéristiques du modèle $M/M/C : (\infty, \text{FIFO})$.

- Loi des arrivées : loi de Poisson de moyenne λ
- Loi de service : loi exponentielle de moyenne $1/\mu$
- Nombre de stations : C
- Règle de priorité : FIFO
- Intensité du trafic = $\rho_C = \rho/C$
- Probabilité qu'il n'y ait aucun client dans la file d'attente : $P_0 = 1 - \rho$
- Probabilité qu'il y ait n unités dans le système : $P_n = \rho^n (1 - \rho)$
- Nombre moyen de clients n dans le système : $\bar{n} = \frac{\rho}{1 - \rho}$
- Nombre moyen de clients n' dans la file d'attente : $\bar{n}' = \frac{\rho^2}{1 - \rho}$
- Temps moyen d'attente dans le système : $\bar{t} = \frac{\rho}{\lambda (1 - \rho)}$
- Temps moyen d'attente dans la file d'attente : $\bar{t}' = \frac{\rho}{\mu (1 - \rho)}$

4 - 2 - Modèle $M/D/C : (\infty, \text{FIFO})$.

La distribution des temps de service est considérée comme uniforme.

Soient la probabilité P_n de se faire service et A_n la probabilité cumulée, on aura :

$$A_n = \sum_{i=0}^n P_i$$

Nous pouvons mesurer le temps en multiples de la durée constante de service, qui elle-même peut être prise pour unité, et de là on peut relier le nombre de clients présents au début d'un intervalle à ceux présents à la fin.

On aura la probabilité P_0 qui est égale à : $P_0 = A_C \cdot e^{-\lambda}$

Evidemment s'il y a moins de clients que de stations au début de l'intervalle et aucune arrivée

pendant l'intervalle, il n'y aura plus de clients à la fin de l'intervalle.

$$P_1 = A_C \cdot \lambda e^{-\lambda} + P_{C+1} \cdot e^{-\lambda}$$

.....

$$P_n = A_C \frac{\lambda^n}{n!} \cdot e^{-\lambda} + P_{C+1} \frac{\lambda^{n-1}}{(n-1)!} \cdot e^{-\lambda} + \dots + P_{n+c} \cdot e^{-\lambda}$$

Cette formule ne doit pas trop nous inquiéter car il existe des abaques pour \bar{t}' quand le nombre de stations varie de 1 à 100 et l'intensité du trafic ρ entre 0,1 et 0,96 ($\rho = \lambda/C\mu$).

Dans beaucoup d'applications on a besoin de savoir quelle est la probabilité d'attendre dans la file un temps supérieur à t : $P(>t)$. C'est une expression très complexe et presque inutilisable, sauf travail démesuré d'un ordinateur. Ce calcul peut être évité par l'utilisation d'abaques qui sont disponibles pour $C = 1$ à 20 et $\rho \leq 0,8$ qui sont suffisants dans la plupart des utilisations pratiques.

Ce modèle s'applique surtout dans le domaine des télécommunications et dans la conception des systèmes de calcul en temps réel.

4 – 3 – Modèle $M/E_k/C : (\infty, FIFO)$.

Les cas étudiés de ce modèle se limitent au cas où $C = 2$ et $k = 2$. Les calculs sont extrêmement difficiles, et il semble presque impossible d'engager des cas plus compliqués. Il est alors nécessaire d'utiliser l'une des trois méthodes qui suivent.

- Simuler le processus, surtout pour avoir des renseignements sur les états transitoires.
- Utiliser un ordinateur pour résoudre les équations du régime permanent.
- Utiliser un modèle plus facile et connu qui présente plusieurs analogies au problème étudié.

Il n'existe pas à notre connaissance de table ou d'abaque commode qui puisse fournir des renseignements concernant le temps d'attente dans le système ou la file, si ce n'est dans de rares cas où $C = 2$ ou $C = 3$.

4 – 4 – Modèle $M/M/2 : (\infty, GD)$.

Dans les trois modèles à plusieurs stations présentés ci-dessus, nous avons supposés que plusieurs guichets sont alimentés en clients par une seule file d'attente. Dans le premier chapitre, nous avons brièvement abordé le cas d'une seule station alimentés par plusieurs files d'attente, chaque file correspond à une priorité différente.

Des modèles théoriques de tels processus sont rares et mal adaptée, et les problèmes mettant en jeu ces processus sont mieux résolus par les méthodes de simulation. Cependant, nous allons présenter ces modèles en supposant que les arrivées suivent une loi de Poisson, et qu'à son arrivée chaque client a le choix de prendre une file d'attente devant n'importe laquelle des C stations.

Il est clair que si une règle quelconque peut être établie qui détermine quelle file le client va prendre, et si une fois dans une file particulière, il doit y rester, alors le modèle, n'est autre que la juxtaposition de stations indépendantes.

Nous nous intéresserons aux situations qui autorisent un client à changer de file à sa guise et dans le but de réduire son temps d'attente.

On prendra le modèle $M/M/2 : (\infty, GD)$. où GD représente les différentes stratégies possibles que peut prendre un client dans un système à deux guichets (évidemment le cas peut se généraliser à C stations).

Les stratégies possibles qui peuvent être affectées au client sont très nombreuses mais les principales sont les suivantes.

- Le principe du maître d'hôtel ou principe de la file unique contrôlée. Le client procède ainsi : si les deux stations sont occupées, il attend sur une seule file et avance en respectant la discipline FIFO. Si les deux stations sont libres, il choisit l'une ou l'autre au hasard avec la même probabilité.

- Le principe de Krishnamoorthi. Si les deux stations sont libres, le client choisit la première avec la probabilité P_1 et la seconde avec la probabilité P_2 , mais dans les autres cas, la règle du maître d'hôtel est applicable. Selon son auteur, le principe de Krishnamoorthi doit être utilisé quand les taux de service des stations sont différents, et le client qui, en arrivant, trouve deux stations libres, va toujours vers la plus rapide. L'idée semble séduisante et Krishnamoorthi a montré que cela minimise le nombre moyen de clients dans le système. Cependant, il est presque impraticable dans beaucoup de situations réelles car le client n'a aucun moyen de connaître le guichet le plus rapide. Dans certains cas, il est possible aux serveurs eux-mêmes d'indiquer au nouvel arrivant un guichet dont ils savent eux-mêmes qu'il est le plus avantageux, et quand il en est ainsi, la règle apporte les avantages annoncés par le calcul. Une telle situation peut être observée en marche dans des opérations de production où les serveurs sont en fait des machines de cadence différente.

- Les guichets de caissiers. Dans ce cas chaque guichet a sa propre file d'attente ; un client qui arrive prend la file la plus courte. Si deux files sont vides, il choisit l'une ou l'autre avec la même probabilité. Une fois dans la file particulière il y reste. Cette règle est notée $E_k - 1$, ce qui signifie règle de Koenigsberg. Il est pratique de prendre pour notation les initiales de la première personne à avoir étudié et classé la règle.

Dans le modèle M/M/2 : (∞ , EK-1) où les taux de service des deux guichets sont respectivement μ_1 et μ_2 , il est intéressant de remarquer que la longueur moyenne des deux files d'attente est la même et qu'elle est indépendante des taux de service.

Ainsi, les préférences du client n'affectent pas les longueurs moyennes de files quand elles sont de la forme 'choisir la file la plus courte et y rester'. Les clients qui prennent la file la plus lente doivent par conséquent, en moyenne, attendre plus longtemps que les autres. La proportion de clients dans chaque file est alors $\frac{\mu_1}{\mu_1 + \mu_2}$.

Un changement de file probabiliste dans le cas des guichets de caissiers, n'a pas d'objet pratique. Le modèle de file d'attente conduit aux mêmes caractéristiques de longueur de file que si chaque client prenait la file la plus courte et y restait, et ce résultat est complètement indépendant de la valeur des constantes de proportionnalité mesurant les taux de changement de file. Ce sont deux processus de changement de file assez peu utilisés, et tous les deux conduisent à une perte du potentiel des serveurs, et corrélativement une perte de capacité de service.

- Les guichets de caissiers avec changement de file ou EK- 2. Comme EK – 1, chaque guichet a sa propre file d'attente, mais dans ce cas les clients peuvent sauter d'une file à une autre plus courte. Le changement de file peut être probabiliste ou instantané. Il y a ainsi deux cas : EK- 2a et EK- 2b. Pour l'un et l'autre les principes généraux de fonctionnement sont les mêmes que pour EK -1, le cas des guichets de caissiers, avec une modification spécifique. Dans EK-2a les clients abandonnent la file la plus longue pour la plus courte à un taux proportionnel à la différence de longueur entre les files. Dans EK-2b, dès l'instant où la différence de longueur excède un, le dernier client de la file la plus longue va instantanément à la fin de la plus courte. Ce modèle a exactement les mêmes caractéristiques de longueur moyennes de files que le modèle M/M/2 : (∞ , FIFO), et ceci est vrai même si les taux de service des guichets sont différents ; c'est-à-dire que les résultats subsistent pour des taux moyens de service μ_1 et μ_2 .

- Le changement d'allée ou EK – 3. Les arrivants choisissent une file initiale, mais sautent de la plus longue à la plus courte à un taux proportionnel à la différence de longueur. Un client donné de la file la plus longue peut changer ou non, car d'autres peuvent le faire avant lui et égaliser ainsi les files avant qu'il ne puisse bouger, c'est-à-dire que les transferts d'une file à l'autre forment une séquence et ne sont pas simultanés.

- Le changement de route ou EK – 4. Les clients choisissent une file initialement, mais les changements de files dans les deux sens sont autorisés. La probabilité de quitter une file pour l'autre dépend seulement

de la longueur de la file. Le client n'a aucune information sur la longueur des files autres que celle dans laquelle il se trouve.

Koenigsberg a examiné l'application de chacune de ces stratégies dans la situation $M/M/2 : (\infty, GD)$, le GD étant en fait remplacé tour à tour par chacune des règles spécifiques. En cela, il met en évidence le fait que les changements de files n'améliorent les performances d'un système de file d'attente que si le temps d'inactivité auxquels les serveurs sont contraints se trouve diminué : c'est le seul cas où les changements de files sont payants.

Koenigsberg a limité son étude au cas de deux stations. Quand il y a plusieurs stations (plus de cinq), d'autres effets du comportement des clients commencent à apparaître. Les modèles mathématiques dont on dispose dans des cas semblables sont très peu nombreux, si tant est qu'il y en ait. C'est peut-être parce qu'il est difficile de déterminer ce que font exactement les clients quand on observe de tel processus.

5 – Modèle de file d'attente à processus biphasé.

Dans les paragraphes précédents, nous avons considéré des modèles où il existait une seule étape de service. Dans la pratique, tous les phénomènes ne se limitent pas toujours à un seul centre de service. Il peut y avoir des cas où après un premier processus de file d'attente, les clients sortant de ces derniers après un service, constituent le flot d'entrée d'un deuxième processus attendant un service. Il existe ainsi deux phases dans ce système. Il peut y en avoir plusieurs. Lorsque le processus est constitué de deux phases, il s'agit d'un système biphasé. Lorsque dans le système chaque demande est servie successivement par plusieurs unités, il s'agit d'un processus multi phase ; on l'appelle aussi processus sériels de file d'attente, ou file d'attente en tandem ou bien en série.

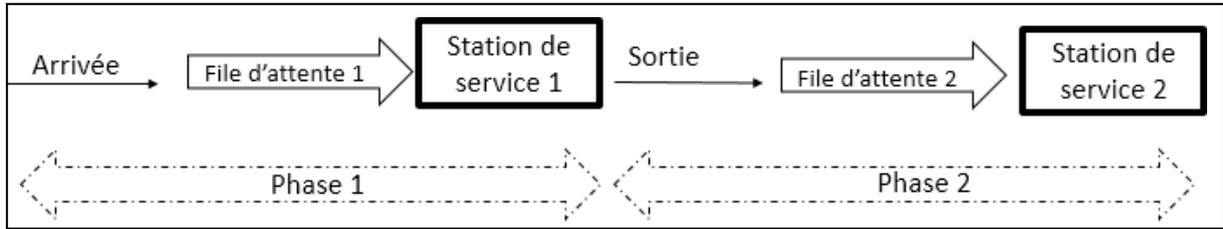
Il est difficile d'utiliser des modèles multiphases, sauf si les distributions de temps de service sont exponentielles. La raison en est qu'en dehors du cas de ces distributions exponentielles, les intervalles séparant les départs ne sont naturellement pas distribués de façon exponentielle. Ainsi, les différentes étapes régissent les unes sur les autres. Dans les cas de file d'attente multiphase assez complexe, l'usage des techniques de simulation est plus approprié.

Il est possible de progresser dans de nombreux cas pratiques en supposant seulement en dépit des faits que la seconde étape, dans un système à deux phases par exemple, est indépendante de la première. Si la première agit comme un filtre du fait d'une capacité de file limitée, il faut dûment modifier le taux moyen d'alimentation de la seconde étape.

5 – 1 – Modèle $M/M/1 : (\infty, FIFO)$ alimentant $M/M/1 : (\infty, FIFO)$.

Les modèles de processus biphasé auxquels les mathématiciens se sont le plus intéressés comportent un système $M/M/1 : (\infty, FIFO)$ alimentant $M/M/1 : (\infty, FIFO)$. Pour notre cas, nous nous limiterons à considérer le cas particulier où $C_1 = C_2 = 1$ et à donner les formules sans démonstration car ces dernières demanderaient un travail considérable qui d'un point de vue pratique serait peu judicieux.

Figure 4. Schéma d'un système biphasé.



Le flux de sortie de la première phase constitue le flux d'entrée de la deuxième phase.

Notons $\rho_1 = \lambda/\mu_1$ pour la première phase et $\rho_2 = \lambda/\mu_2$ pour la deuxième phase.

Posons aussi $\rho_1 < 1$ et $\rho_2 < 1$ qui sont la condition d'existence d'un régime permanent.

Nous aurons les caractéristiques qui suivent.

- Nombre moyen de clients dans les deux files :

$$\bar{n} = \frac{\rho_1^2}{1-\rho_1} + \frac{\rho_2^2}{1-\rho_2}$$

Avec $\bar{n}_1 = \frac{\rho_1^2}{1-\rho_1}$ le nombre moyen de clients dans la file 1

et $\bar{n}_2 = \frac{\rho_2^2}{1-\rho_2}$ le nombre moyen de clients dans la file 2.

- Nombre moyen de clients dans le système :

$$\bar{n} = \frac{\rho_1}{1-\rho_1} + \frac{\rho_2}{1-\rho_2}$$

Avec $\bar{n}_1 = \frac{\rho_1}{1-\rho_1}$ le nombre moyen de clients dans la phase 1

et $\bar{n}_2 = \frac{\rho_2}{1-\rho_2}$ le nombre moyen de clients dans la phase 2.

- Nombre moyen de clients en cours de service :

$$\bar{j} = \rho_1 + \rho_2 \quad \text{avec } \rho_1 \text{ pour la station 1, et } \rho_2 \text{ pour la station 2.}$$

- Durée moyenne dans les files 1 et 2.

$$\bar{t}_1 = \frac{1}{\mu_1} \times \frac{\rho_1}{1-\rho_1} \quad \text{représente la durée moyenne d'attente dans la file 1.}$$

$$\bar{t}_2 = \frac{1}{\mu_2} \times \frac{\rho_2}{1-\rho_2} \quad \text{représente la durée moyenne d'attente dans la file 2.}$$

- Durée moyenne dans le système :

$$\bar{t} = \frac{1}{\mu_1} \times \frac{\rho_1}{1-\rho_1} + \frac{1}{\mu_2} \times \frac{\rho_2}{1-\rho_2}$$

Avec $\frac{1}{\mu_1} \times \frac{\rho_1}{1-\rho_1}$ la durée moyenne d'attente dans la phase 1

et $\frac{1}{\mu_2} \times \frac{\rho_2}{1-\rho_2}$ la durée moyenne d'attente dans la phase 2.

- Probabilité qu'il y ait n_1 demande dans la première phase et n_2 demandes dans la deuxième phase :

$$P_{n_1} P_{n_2} = (1 - \rho_1) \cdot (1 - \rho_2) \cdot \rho_1^{n_1} \cdot \rho_2^{n_2}$$

- Probabilité que les deux stations soient libres :

$$P_{01} P_{02} = (1 - \rho_1) \cdot (1 - \rho_2)$$

5 - 2 - Modèles M/M/1 : (n_1 , FIFO) alimentant M/M/1 : (n_2 , FIFO).

Nous aurons les caractéristiques qui suivent avec $n_1 + n_2 = M$.

- Probabilité que les deux stations soient libres :

$$P_{00} = \frac{(\rho_1 - \rho_2)(1 - \rho_1) \cdot (1 - \rho_2)}{\rho_1 - \rho_2 (\rho_1^{n_1 + n_2 + 2} - \rho_2^{n_1 + n_2 + 2}) + \rho_1 \rho_2 (\rho_1^{n_1 + n_2 + 2} - \rho_2^{n_1 + n_2 + 2})}$$

- Probabilité qu'il y ait n_1 demandes dans la première phase, et n_2 demandes dans la deuxième phase.

$$P_{n_1} P_{n_2} = P_{00} \times \rho_1^{n_1} \cdot \rho_2^{n_2}$$

- Nombre moyen de clients dans les deux files :

$$\bar{n} = \frac{\rho_1^2(1 + \rho_1)[(1-M)\rho^{M-1} + (M-1)\rho^M]}{(1 - \rho_1)^2} + \frac{\rho_2^2(1 + \rho_2)[(1-M)\rho^{M-1} + (M-1)\rho^M]}{(1 - \rho_2)^2} + \frac{\rho_1 \rho_2 P_{00}}{\rho_1 - \rho_2} \left[\left(\frac{1 - (M-1)\rho^{M-2} + (M-2)\rho^{M-1}}{(1 - \rho_1)^2} \right) - \left(\frac{1 - (M-1)\rho^{M-2} + (M-2)\rho^{M-1}}{(1 - \rho_2)^2} \right) \right]$$

- Nombre moyen de clients dans le système :

$$\bar{n} = \frac{P_{00}}{\rho_1 - \rho_2} \left[\frac{\rho^2[1 - (M+1)\rho^M + M\rho^{M+1}]}{(1 - \rho_1)^2} - \frac{\rho^2[1 - (M+1)\rho^M + M\rho^{M+1}]}{(1 - \rho_2)^2} \right]$$

- Nombre moyen de clients en cours de service :

$$\bar{j} = P_{00} (\rho_1 + \rho_2) \left[\frac{(\rho_1 + \rho_2) - (\rho_1^{M+1} - \rho_2^{M+1}) + \rho_1 \rho_2 (\rho_1^M - \rho_2^M)}{(1 - \rho_1) \cdot (1 - \rho_2) \cdot (\rho_1 - \rho_2)} \right]$$

6 – La simulation

Si l'importance des modèles mathématiques pour l'étude des phénomènes économiques n'est plus à démontrer, il s'avère cependant que pour l'optimisation de certains processus complexe, l'utilisation de modèle est insuffisante. En effet, un algorithme utilisé en recherche opérationnelle ne peut être obtenu que si le type de problème à traiter est parfaitement déterminé, en d'autres termes, s'il a pu être entièrement analysé et formulé. Mais bien souvent la complexité des phénomènes étudiés n'autorise pas une telle modélisation, et de nombreux cas pratiques peuvent être difficilement ramenés à ces modèles mathématiques type, à moins de faire des simplifications abusives ou des calculs inextricables qui pourraient fausser le résultat final. Effectivement, il est difficile de mettre tout élément économique en équation et analyser leur évolution dans le temps. Cette évolution devient encore plus difficile lorsqu'elle est soumise à des variations aléatoires. Ainsi, pour le phénomène des files d'attente, où beaucoup de problèmes concrets sont bien plus compliqués que les situations décrites dans les paragraphes précédents, il arrive souvent qu'il soit formé d'opérations très complexes où les arrivées et les services ne se font pas suivant une loi classique ; le cas des modèles à priorité assez complexe pose aussi fréquemment des problèmes, surtout lorsque les distributions sont non stationnaires.

N'ayant plus la possibilité d'utiliser des modèles mathématiques, on se servira des méthodes de simulation où l'on s'efforce de faire paraître comme vrai (ou du moins possible) ce qui ne l'est pas. Ainsi, les simulations sont des expériences artificielles qui ne modifient pas le fonctionnement du système réel. Simuler, c'est substituer à un système 'réel' complexe, un système 'artificiel' plus simple qui lui soit équivalent à certains points de vue, afin qu'il lui soit possible de procéder à des expériences sur celui-ci et de tirer des renseignements sur le système 'réel' par analogie. Le chercheur opérationnel adopte donc alors la même attitude que son confrère du laboratoire, il expérimente. Ne pouvant expérimenter sur des faits réels, il expérimentera dans l'abstrait, mais de telle manière que les conditions de son essai soient aussi conformes que possible à la réalité et sans que la situation réelle en soit affectée ; on dit qu'on procède par simulation. L'opérateur économique dispose là d'une méthode d'une grande efficacité. Grâce à elle, il peut procéder à des essais tout comme son confrère des sciences expérimentales. La simulation est le pont qui relie ces dernières à la recherche opérationnelle.

La simulation est un instrument puissant de travail qui permet d'étudier à loisir le comportement d'un processus après avoir décomposé les variables de commandes en opération élémentaires, en faisant varier les paramètres et les grandeurs aléatoires.

La simulation présente de nombreux avantages qui tiennent le plus souvent à son aspect expérimental. L'expérience est d'un apport plus concret que le calcul théorique ; la simulation est donc plus facile à utiliser et ses résultats, qui sont mieux compris, sont utilisés avec plus de confiance. Ses principaux avantages sont :

- d'expérimenter un très grand nombre de comportement ou de politiques différentes,
- de prendre en compte des évènements exceptionnels,
- de reproduire l'aspect dynamique des situations étudiées, c'est-à-dire leur évolution,
- de permettre de lever rationnellement l'incertitude dans les cas où la trop grande complexité du problème exclut des méthodes mathématiques. Et c'est là, la raison principale de la simulation.

Pour le processus de files d'attente, on aura recours aux méthodes de simulation où on simulera un certain nombre de situations possibles pour pouvoir appliquer les théories de l'échantillonnage statistique et en déduire l'estimation des grandeurs recherchées (nombre de clients dans la file d'attente, durée moyenne d'attente, etc.).

Le développement important de l'informatique ces dernières décennies a donné une impulsion considérable à l'utilisation des méthodes de simulation, car seul l'ordinateur permet de répéter la simulation des milliers de fois pour obtenir une bonne précision et traiter des problèmes de structure très complexe, comprenant de nombreuses variables. Les modèles de simulation prennent le plus souvent la forme de programmes d'ordinateur. Les techniques interactives et graphiques permettent à l'utilisateur de dialoguer avec l'ordinateur en réagissant immédiatement aux résultats qui lui sont présentés.

Cependant si simuler signifie reproduire et l'idée de reproduire pour mieux comprendre a le mérite de la simplicité, celle-ci n'est pas synonyme de faciliter et il ne faudrait pas croire que l'on dispose là d'un instrument d'efficacité absolue. En d'autres termes si l'on néglige de s'entourer de multiples précautions tels que l'exactitude et la vérification des éléments considérés, la mise en œuvre de la méthode de simulation est vouée à l'échec ; sans doute conduit-elle à un résultat, mais celui-ci n'a aucune signification et il serait alors dangereux d'en user comme d'un outil précis de décision. Par exemple les durées d'attente des clients successifs dans une file ne sont pas des variables aléatoires indépendantes, et la variance de cette série de valeur serait une estimation biaisée de la variance du temps d'attente ou bien le calcul toujours

possible d'une moyenne statistique peut masquer la non existence d'une espérance mathématique infinie.

Ainsi, la simulation n'est pas la méthode universelle. En effet l'analyse, la programmation, les tests et la mise au point de programmes de simulation sont des opérations très longues et coûteuses. La longueur des opérations et leur coût dépend du nombre d'expériences à effectuer, c'est-à-dire le nombre d'itérations. Si dans le modèle mathématique tel que $M/M/1 : (\infty, \text{FIFO})$ qui commence à fonctionner à partir d'un état définie à l'instant 0 doit passer par une phase transitoire avant d'atteindre un régime permanent, il en est de même lorsqu'on étudie un phénomène de file d'attente par simulation. Quand on veut évaluer un modèle de file d'attente au moyen de la simulation, on doit supposer certaines valeurs initiales aux paramètres du système. Par exemple, il faut dire combien de clients il y a dans la file (ou les files) ou combien dure un service. Ces conditions initiales doivent inévitablement avoir une influence sur les résultats produits par la simulation pour une période. La question est de savoir au bout de combien de temps la simulation obtient un état d'équilibre statistique et sous quelles conditions. Il n'y a pas de règles précises en la matière, cependant plus le nombre d'expérience est grand et plus le résultat est meilleur, d'où un coût élevé.

En plus, si la simulation permet de tester une politique, elle n'indiquera pas si elle est meilleure et comment la modifier pour obtenir les résultats escomptés. Cependant, si un processus de file d'attente ne peut être traité analytiquement vu sa complexité, une solution par voie de simulation vaut mieux que pas de solution du tout

Les techniques de simulation ont leur principal domaine d'utilité dans les comparaisons entre des formes d'organisation des facilités de service pour des files d'attente identique quand il n'existe pas de modèle théorique

Ces quelques limites ne doivent pas nous détourner de son utilisation mais au contraire nous persuader de la nécessité de soumettre tout modèle d'échantillonnage fictif à une étude de validation et d'analyse théorique. Il sera souvent bon d'insérer des parties complètement traitées mathématiquement ou de décomposer les chroniques étudiées en suites indépendantes correspondant aux passages par des points de régénération. La bonne connaissance des modèles théoriques 'élémentaires' semblent contribuer beaucoup à la qualité des résultats d'une simulation de phénomènes réels et complexes.

Nous ne donnerons pas dans ce chapitre la méthodologie pour l'étude d'une simulation, car cela n'est pas l'objet de notre travail. Elle sera étudiée en détail dans le cours de la simulation.

En conclusion, nous rappelons que la méthode de simulation, aussi simple puisse-t-elle paraître est une affaire de spécialistes qui savent exploiter au maximum les méthodes de calculs

analytiques et ne confier à la méthode expérimentale que la partie irréductible des problèmes. C'est un moyen d'obtenir des résultats valables, d'éviter aussi des modèles trop chargés car l'encombrement n'est pas toujours garant de la qualité.

CONCLUSION

L'ensemble des modèles présentés ont été des plus simples et consacrés à des processus en équilibre statique. Aucun renseignement n'a été indiqué sur le processus transitoire. On peut constater que pour des modèles simples les expressions mathématiques sont parfois complexes. On peut alors se demander à quoi ressemblent les expressions de modèles plus complexes et en période transitoire. Elles sont généralement ou inaccessibles ou inextricables. C'est une des raisons pour laquelle l'étude des phénomènes de files d'attente ne s'étudie le plus souvent qu'après que la période de démarrage soit terminée. Si l'étude de cette période est nécessaire, on utilise alors une méthode de simulation car une analyse mathématique précise est pratiquement impossible.

Nous pouvons considérer que seuls les modèles les plus simples sont susceptibles d'une analyse mathématique précise, mais qu'ils étaient assez peu rencontrés dans le monde réel. Cela ne doit pas nous conduire à refuser l'utilisation de tels modèles car ils peuvent rapporter des solutions opérationnelles dans de nombreuses situations. Il suffira d'accepter certaines approximations. L'une de ces approximations consistera à obtenir des solutions opérationnelles en appliquant des modèles a priori faux, mais simples à manier. Si cette approximation est jugée inacceptable, on peut utiliser alors la simulation qui est une forme d'approximation.

Des modèles approximatifs, élaborés toutefois avec une compréhension claire et précise de ce qu'ils représentent approximativement sont en général adéquats.

Evidemment, il serait préférable d'utiliser des modèles rigoureux que des modèles approximatifs, mais il vaut mieux avoir des modèles approximatifs que pas de modèle du tout. Toute théorie ayant une base concrète et logique, même incomplète et embryonnaire, vaut mieux que l'absence de théorie.

Dans la pratique, les phénomènes de file d'attente se présentent sous trois formes principales qui sont les suivantes.

- La prévision quantitative et qualitative de nouveaux équipements de service dans le fonctionnement desquels les files d'attente sont vraisemblablement des facteurs importants d'engorgement. Les problèmes d'élaboration et d'analyse des modèles de file d'attente, pour prévoir la capacité qu'il faut donner à de nouveaux équipements peuvent être considérables. Cependant les modèles de file d'attente ne sont pas la seule source de difficultés. C'est au moins autant, si ce n'est plus, qu'il faut attacher à la prévision des paramètres de ce modèle. Or, la prévision est une entreprise plutôt hasardeuse, surtout lorsqu'il s'agit du long terme. De plus, il

sera très difficile de prévoir ce que sera juste le temps de service moyen d'une action, surtout si les opérations de service qui la constituent sont plus ou moins nouvelles ou organisées d'une façon nouvelle. Il sera donc toujours nécessaire pour un économiste de savoir quelle précision il veut atteindre et peut-il atteindre ?

- L'élaboration de stratégies pour la répartition de la main d'œuvre dans des équipements de service existants, avec pour objectif de satisfaire certaines normes de service et de coût de fonctionnement. On voit donc que cette situation est toute différente de la précédente qui était basée essentiellement sur la prévision. Ici, les stations existent déjà, et la question est de savoir combien il faut mettre en service des stations de façon à fournir une certaine norme de service. Dans ce cas un modèle très fidèle sera souvent compliqué, tout à fait précis, pas toujours utile. Par contre, un modèle moins fidèle, mais plus simple et pas trop précis sera peut-être utile. La décision du choix du modèle reviendra au chercheur opérationnel qui tiendra compte de tous les facteurs quantitatifs et qualitatifs. Le facteur coût sera l'un des plus importants. Il s'agira d'un critère d'efficacité économique du phénomène de file d'attente. En fonction du problème posé par le gestionnaire, il sera fait un choix pour évaluer 'efficacité économique 'un processus de file d'attente. Ce n'est qu'après identification d'un système que le responsable pourra procéder à l'optimisation des moyens.

Nous pouvons retenir, parmi tant d'autres, pour un système à file limitée le critère économique relatif aux pertes subies par le système. Celui-ci est le suivant :

$$X = D_{lib} \cdot C_{lib} + D_{oc} \cdot C_{oc} + D_f \cdot \bar{n} + D_p \cdot \lambda \cdot P_r$$

Où D_{lib} désigne les dépenses pour une unité de service non occupée par unité de temps,

C_{lib} désigne le nombre de stations libres,

D_{oc} désigne les dépenses pour une unité de service occupée par unité de temps,

C_{oo} désigne le nombre de stations occupées,

D_f désigne les dépenses liées à l'attente d'une demande dans le système par unité de temps,

\bar{n} désigne le nombre moyen de clients dans la file d'attente,

D_p désigne les pertes dues au refus d'une demande par unité de temps,

P_r désigne la probabilité de refus d'une unité.

Les critères suivants peuvent être retenus respectivement pour un système d'attente avec file illimitée ($s \rightarrow \infty$) et pour un système à demande refusée ($s = 0$).

$$X_i = D_{lib} \cdot C_{lib} + D_{oc} \cdot C_{oc} + D_f \cdot \bar{n} \quad \text{car } P_r = 0 \text{ puisque la file d'attente est illimitée}$$

$$X_r = D_{lib} \cdot C_{lib} + D_{oc} \cdot C_{oc} + D_p \cdot P_r \quad \text{car } \bar{n} = 0 \text{ puisque les clients sont refusés lorsque la station est occupée, puis } \lambda = 1 \text{ puisque le client se fait servir dès qu'il arrive.}$$

- Comparaison des différentes formes d'organisation d'équipements de service. Prenons le cas où nous supposons qu'il y ait trois arrangements possibles et que le trafic d'alimentation soit constant et indépendant de l'arrangement. Considérons le problème de savoir quel arrangement conduira, pour un effectif donné de serveurs, au temps d'attente moyen le plus bas **minimal** pour le client.

En connaissant approximativement les résultats généraux de la théorie des files d'attente, il sera souvent possible de répondre à ce genre de problème sans faire absolument aucun calcul. Par exemple, en supposant que la distribution du temps de service ne soit pas affecté par la nature de l'arrangement, il est clair qu'un arrangement qui fournirait une disponibilité totale serait préférable à un qui ne le ferait pas. Ou qu'un arrangement qui arrangerait la mobilité entre les stations serait meilleur qu'un qui ne le ferait pas, bien que moins bon qu'un système du type 'maître d'hôtel' avec disponibilité complète. C'est dans des cas semblables que l'on apprécie les résultats généraux, même s'ils découlent de modèles passibles d'équations intégrales-différentielles. Le problème en pareil cas est d'identifier les distributions du temps de service, le type d'arrivée et les disciplines réelles des files d'attente plutôt que de calculer avec précision les paramètres d'un modèle le plus adéquat pour son application. Parfois, il ne suffit pas de savoir seulement quel est l'arrangement le plus efficace, mais d'être en mesure de le quantifier c'est-à-dire d'être capable de dire de combien.

En conclusion nous pouvons définir une solution opérationnelle à un processus de files d'attente qui discrimine suffisamment pour conduire à une bonne décision. Cette solution n'est pas obligatoirement basée sur un modèle très complexe. Elle peut être déduite de quelques hypothèses, qui parfois même semblent erronées, sur la nature des lois de distribution des temps de service.

L'essentiel en recherche opérationnelle n'est pas d'obtenir des solutions au sens mathématique du terme, mais des solutions opérationnelles qui peuvent mener à de décisions pratiques et rationnelles.

Exercices

Exercice 1

Pour analyser le flux d'entrée de véhicules dans une station d'essence, on a mesuré la durée de temps (en secondes) des intervalles entre les arrivées des véhicules à la station.

L'observation a donné les résultats suivants :

48 – 2 – 12 – 82 – 159 – 24 – 63 – 116 – 238 – 1 – 3 – 61 – 100 – 1 – 414 – 99 – 11 – 20 – 7 –
87 – 64 – 5 – 177 – 10 – 236 – 201 – 110 – 8 – 18 – 134 – 156 – 131 – 81 – 17 – 9 – 13 – 15 –
47 – 90 – 190 – 53 – 7 – 4 – 194 – 206 – 116 – 64 – 91 – 22 – 12 – 47 – 6 – 66 – 81 – 324 –
191 – 1 – 42 – 28 – 70 – 119 – 8 – 59 – 26 – 100 – 224 – 208 – 101 – 104 – 14 – 3 – 39 – 27 –
75 – 55 – 21 – 56 – 5 – 94 – 389 – 25 – 33 – 74 – 142 – 81 – 108 – 41 – 28 – 87 – 121 – 29 –
474 – 61 – 50 – 187 – 166 – 27 – 26 – 150 – 263 – 175 – 13 – 62 – 76 – 43 – 182 – 212 – 40 –
17 – 126 – 107 – 88 – 160 – 30 – 144 – 56 – 22 – 112 – 156 – 219 – 147 – 59 – 74 – 80 – 117
– 162 – 110.

Déterminez la loi de distribution caractérisant le flux d'entrée des véhicules dans la station et testez le résultat obtenu en utilisant le critère de Pearson avec un seuil de confiance égal à 0,05.

Exercice 2

La direction d'un magasin d'alimentation se propose de modifier l'installation de celui-ci afin d'appliquer le self-service. Le problème qu'elle se pose se limite au choix du nombre de points de sortie à prévoir pour payer.

Dans ce but, il a été procédé à une série de 3 000 observations, portant sur le flux d'arrivées, pendant 50 jours, sur l'heure la plus chargée (60 minutes durant 50 jours : $60 \cdot 50 = 3\,000$ observations). Les résultats sont récapitulés dans le tableau qui suit.

Nombre de clients arrivés dans la minute n_i	Nombre de cas observés N_i
0	702
1	977
2	710
3	402
4	153
5	48
6	8
	3000

Sachant qu'une caissière peut servir, en moyenne, un maximum de 24 clients par heure, la direction recherche combien il faut installer de caisses pour qu'un client attende, en moyenne, dans la file, moins d'une minute.

Exercice 3

Soit un phénomène d'attente à plusieurs stations. Le taux moyen des arrivées par dizaine de minutes est égal de $a = 8$ et la durée moyenne de service est de cinq minutes.

Calculez pour un nombre de stations $n = 5, 6$ et 7 , le nombre moyen d'unités dans la file d'attente \bar{u} et le temps moyen d'attente \bar{t} dans la file.

Exercice 4

Soit un atelier dans lequel les machines tombent en panne selon une loi de Poisson de taux (par heure) $\lambda = 3$. Le coût d'arrêt d'une machine peut être évalué à 10 U.M par heure. On a le choix entre d'une part :

- un mécanicien expérimenté, réparant les machines selon un service exponentiel de taux égal à 4 par heure, qui demande 3 UM de l'heure ou 5 UM, y compris les charges.

Et d'autre part :

- un mécanicien très expérimenté réparant les machines selon un service exponentiel de taux 5 par heure. Il demande 4 UM de l'heure ou soit 6,5 UM y compris les charges.

1. Lequel embaucher ?

2. Lorsqu'on a le choix entre embaucher deux ouvriers expérimentés et deux ouvriers très expérimentés, lesquels embaucher ?

Exercice 5

On veut construire des installations de déchargement de minerais. On a le choix entre trois types a, b et c dont les capacités moyennes journalières (quotidiennes) sont respectivement de 3 000, 5 000 et 7 000 tonnes. Les coûts fixes (amortissement, entretien, etc.) sont parallèlement de 5 000, 10 000 et 15 000 UM par jour et les coûts de fonctionnement respectifs, qui ne courent que lorsque des opérations de déchargement ont lieu, s'établissent à 6 000, 9 000 et 12 000 UM par jour.

Les navires, chacun portant 9 000 tonnes, arrivent, selon un processus poissonien, à la moyenne de deux tous les trois jours. Le coût d'attente d'un navire est de 20 000 UM par jour.

Sachant que pour des raisons d'homogénéité de matériel, on doit choisir des installations d'un seul type, déterminez :

1. L'installation à réaliser actuellement si l'on ne prévoit pas une sensible augmentation du trafic à l'avenir ;
2. La même installation, dans le cas où l'on prévoirait une **autre** augmentation sensible de ce trafic (100% par exemple).

On admettra que le service est exponentiel.

Exercice 6

Il existe au nouveau d'une gare donnée, un service chargé de l'enregistrement des commandes de billets par téléphone. L'analyse des moments d'appels montre qu'ils obéissent à une loi de Poisson de moyenne 2,5 commandes par minute. Le temps d'enregistrement d'une commande constitue une variable aléatoire dépendants de plusieurs facteurs et égal à en moyenne à 0,8 minutes.

Si une commande arrive alors que toutes les lignes sont occupées, elle n'est pas enregistrée.

Définir le nombre de lignes téléphoniques dont le service a besoin pour pouvoir servir 90% des commandes.

Exercice 7

Afin de veiller à maintenir un certain niveau de la qualité d'un produit, les responsables de l'entreprise ont procédé à l'installation d'une série de stands d'essais et de contrôle de la production. L'analyse du flux d'entrée du produit au niveau des stands a montrée qu'il suit une loi de Poisson d'une moyenne de deux pièces par heure.

1. Trouvez l nombre de stands dont il faut disposer pour que toute la production soit contrôlée sachant qu'il n'existe pas de refus de contrôle et sachant aussi les coûts qui suivent.

- Coûts liés à l'activité d'un stand : 15 UM par heure,
- Coûts liés à l'inactivité d'un stand : 10 UM par heure.
- Coût lié à l'attente d'une unité d'un stand : 10 UM par heure.

2. Analysez le travail du nombre de stand.

Exercice 8

Dans un bureau de poste une des caisses est réservée aux chèques postaux. Avant de se présenter à la caisse pour k retraits de la somme demandée, chaque client doit déposer son chèque au niveau d'un guichet qui contrôle son avoir. La densité de population dans la région est telle que le flux d'entrée des clients obéit à une loi de Poisson. Une enquête a permis d'estimer l'intensité de ce flux qui est de 20 clients par heure. Le temps de service du contrôleur et celui du caissier sont deux variables aléatoires indépendantes et ont comme moyenne respective 1,5 minutes et 2 minutes.

Calculez les caractéristiques du système formé par les deux guichets.

Solutions des exercices

Solution de l'exercice 1

Avant de déterminer la loi de distribution, il faut tout d'abord regrouper les données de l'échantillon en classes égales. L'amplitude sera déterminée au moyen de la formule de Sterges, qui est :

$$a = \frac{V_{\text{Max}} - V_{\text{min}}}{1 + 3,32 \cdot \text{Log}_{10}n}$$

Avec a : amplitude de classe,

V_{Max} : valeur maximale de l'échantillon,

V_{min} : valeur minimale de l'échantillon,

n : taille de l'échantillon.

Dans notre cas, nous avons $n = 127$, $V_{\text{Max}} = 474$ et $V_{\text{min}} = 1$.

D'où le calcul de a :

$$a = \frac{474 - 1}{1 + 3,32 \cdot 2,10} = 59,3 \approx 60 \text{ secondes, soit près d'une minute.}$$

Dans le tableau de calculs qui suit V'_i désigne le centre de la classe i, k le nombre de classes et n_i la fréquence absolue attachée à la classe i.

Tableau des calculs.

Classes en mn	V'_i	n_i	$n_i \cdot V'_i$	$n_i \cdot V'_i \cdot V'_i$
0 – 1	0,5	56	28	14
1 – 2	1,5	37	55,5	83,5
2 – 3	2,5	16	40	100
3 – 4	3,5	13	45,5	159,5
4 – 5	4,5	1	4,5	20,25
5 – 6	5,5	1	5,5	30,25
6 – 7	6,5	2	13	84,5
7 – 8	7,5	1	7,5	56,25
		127	199,5	547,75

Nous savons que si la moyenne et l'écart type ne sont pas très différents¹², nous pouvons alors supposer que l'intervalle séparant l'apparition de deux phénomènes consécutifs suit approximativement une loi exponentielle de paramètre $\lambda = \frac{1}{\bar{m}}$.

Dans notre cas, nous aurons :

$$\bar{m} = \frac{199,5}{127} = 1,57 \text{ et } \text{Var} = \frac{547,75}{127} - 2,47 = 1,84 \Rightarrow \sigma = \sqrt{\text{Var}} = \sqrt{1,84} = 1,36$$

Ainsi, la différence entre \bar{m} et σ n'est pas importante (< 20%). Nous pouvons supposer que les valeurs de l'échantillon obéissent à une loi exponentielle de paramètre $\lambda = \frac{1}{1,57}$ véhicules par

$$\text{minute} : \lambda = \frac{1}{1,57} = 0,637.$$

La fonction cumulative des données de l'échantillon est donc $F(t) = 1 - e^{-0,637t}$.

En utilisant le critère de Pearson test du khi-deux, vérifions l'hypothèse H_0 selon laquelle l'intervalle entre les arrivées des véhicules suit une loi exponentielle. Pour cela, nous procéderons d'abord au calcul des probabilités P_i attachées à chaque classe i .

$$P_i = \Pr (V_i < V < V_{i+1})$$

$$P_i = F(V_{i+1}) - F(V_i) = (1 - e^{-0,637V_{i+1}}) - (1 - e^{-0,637V_i}) = e^{-\lambda V_i} - e^{-\lambda V_{i+1}} \text{ avec } \lambda = 0,637.$$

D'où les probabilités qui suivent.

$$P_1 = 0,4706$$

$$P_2 = 0,2493$$

$$P_3 = 0,1317$$

$$P_4 = 0,0698$$

$$P_5 = 0,0369$$

$$P_6 = 0,0196$$

$$P_7 = 0,0104$$

$$P_8 = 0,0055$$

Les fréquences absolues théoriques sont données par $n'_i = (n \cdot P_i)$, d'où :

$$n'_1 = 0,4706 \cdot 127 = 59,77$$

$$n'_2 = 0,2493 \cdot 127 = 31,66$$

$$n'_3 = 0,1317 \cdot 127 = 16,72$$

$$n'_4 = 8,86$$

$$n'_5 = 4,69$$

¹² Une variation inférieure à 20% est considérée comme une variation non significative.

$$n'_6 = 2,49$$

$$n'_7 = 1,32$$

$$n'_8 = 0,70$$

Les hypothèses du test du khi-deux sont données ci-dessous.

H_0 : l'intervalle séparant l'arrivée de deux véhicules suit une loi exponentielle.

H_1 : l'intervalle séparant l'arrivée de deux véhicules ne suit pas une loi exponentielle.

Nous présentons dans un tableau les calculs intermédiaires pour la statistique du khi-deux ou le

critère de Pearson observé qui est égal à : $\chi^2_{obs} = \sum_{i=1}^5 \frac{(n_i - n'_i)^2}{n'_i}$

Comme une classe doit être supérieure à 5 dans un test du khi-deux, nous regroupons les quatre dernières classes en une seule égale à 9,2 (0,70 + 1,32 + 2,49 + 4,69 = 9,2).

Tableau des calculs.

N	n_i	n'_i	$(n_i - n'_i)$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	56	59,77	3,77	14,2129	0,238
2	37	31,67	5,34	28,5156	0,901
3	16	16,92	0,92	0,8464	0,050
4	13	8,86	4,14	17,1396	1,934
5	5	9,20	4,20	17,6400	1,917
					5,040

$$\chi^2_{obs} = 5,040$$

Le nombre de degré de liberté ddl = n - 2 où n est le nombre de classes. Ainsi, ddl = 5 - 2 = 3.

Le seuil de confiance retenu est 0,05, d'où :

$$\chi^2_t(0,05 ; 3) = 7,8 \quad \text{et} \quad \chi^2_{obs} = 5,8 \text{ ou } 5,040 \quad \Rightarrow \quad \chi^2_{obs} < \chi^2_t$$

L'hypothèse H_0 étant acceptée, on peut donc admettre que l'intervalle séparant l'arrivée de deux véhicules successifs obéit à une loi exponentielle de paramètre $\lambda = 0,637$.

De même, on peut conclure que le flux d'entrée des véhicules à la station suit une loi de Poisson est simple :

$$P_k(t) = \frac{(0,637)^n}{n!} \cdot e^{-0,637n}$$

$P_k(t)$ désigne la probabilité d'apparition de k véhicules dans un intervalle de temps t.

Solution de l'exercice 2

On commence par l'étude des arrivées.

On pense à une loi de Poisson, mais il faut le vérifier. Sa moyenne est égale à :

$$\lambda = \frac{0.702 + 1.977 + 2.710 + 3.402 + 4.153 + 5.48 + 6.8}{3\,000} = \frac{4\,503}{3\,000} = 1,501 \text{ arrivées/minute.}$$

Son équation générale pour le calcul des probabilités est :

$$P_n(t) = \frac{(\lambda t)^n}{n!} \cdot e^{-\lambda t}$$

où pour $t = 1$ minute, on a $P_n(1) = \frac{(\lambda)^n}{n!} \cdot e^{-\lambda} = \frac{(1,5)^n}{n!} \cdot e^{-1,5}$

Il est donc possible de calculer le nombre théorique d'observation qui correspondraient à cette loi. Soit $n'_i = n \cdot P_n(1)$ d'où les calculs dans le tableau qui suit.

n clients	$P_n(1)$	n'_i
0	0,224	672
1	0,335	1005
2	0,251	756
3	0,125	375
4	0,047	141
5	0,014	42
6	0,003	9
		3000

Comparons alors les effectifs théoriques n'_i aux effectifs réels n_i à l'aide du test de khi-deux.

Cela nous amène à établir le tableau de calcul qui suit.

n	n_i	n'_i	$(n_i - n'_i)$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
0	702	672	30	900	1,33
1	977	1005	- 28	784	0,78
2	710	756	- 46	2116	2,79
3	402	375	27	729	1,94
4	153	141	12	144	1,02
5	48	42	6	36	0,85
6	8	9	- 1	1	0,11
					8,82

Hypothèses du test concernant les arrivées des clients.

- H_0 : l'intervalle séparant l'arrivée de deux clients suit une loi de Poisson.
- H_1 : l'intervalle séparant l'arrivée de deux clients ne suit pas une loi de Poisson.

La table du khi-deux, au seuil de 5 %, avec cinq degrés de liberté donne 11,0. Le khi-deux calculé (8,82) étant inférieur à 11, nous n'avons pas de raison suffisante pour rejeter l'hypothèse H_0 . Nous admettrons par la suite que la loi des arrivées est assez proche d'une loi de Poisson de paramètre $\lambda = 1,5$.

Calculons à présent le temps moyen d'attente dans la file t , et pour cela, le nombre moyen d'unité n' dans la file en fonction du nombre c de station (caisses). Pour cela, nous remarquons que, s'il y a n clients dans le système (clients en attente aux caisses et en cours de service), il y en aura $n - c$ en attente dans les files. P_n étant la fréquence de ce cas le nombre moyen de clients dans les files d'attente sera :

$$\bar{n} = \sum_{n=c}^{\infty} (n - c) P_n \quad \text{avec } P_n = \frac{\rho^n}{C^{n-c} \cdot C!} \cdot P_0 \quad (n \geq C)$$

$$\text{Et } \bar{n} = \frac{P_0 \rho^{c+1}}{c!(c - \rho^2)}$$

Bibliographie

1. **M. Aidene** et **B. Oukacha**. *Recherche opérationnelle*. Edition Pages bleues, Alger, 2005.
2. **S. Aïvazian, I. Enikov** et **L. Mechalkine**. *Éléments de modélisation et traitement primaire des données*. Edition Mir, Moscou, 1986.
3. **J. P. Ancot** et **J. P. Paelinck**. *Modèles et choix. Une initiation à la modélisation pour pays en développement*. Edition Economica, 1990.
4. **M. Angers**. *Initiation pratique à la méthodologie des sciences humaines*. C T R. Casbah Université, Alger 1997.
5. **J. P. Aubin, P. Neponiastchy** et **A.M. Charles**. *Méthodes explicites de l'optimisation*. Editeur Dunod, 1982.
6. **P. Azoulay** et **P. Dassonville**. *Recherche opérationnelle de gestion*. Tome 1, Presse Universitaire de France, 1976.
7. **R. Benayoun**. *La pratique de l'optimisation dans l'entreprise*. Presse Universitaire de France, 1974.
8. **A. M. Benghezal**. *Programmation linéaire*. Office des Publications Universitaires, Alger, 2000.
9. **B. Benmazouz**. *Recherche opérationnelle de gestion*. Atlas Editions, Alger, 1995
10. **J. L. Brissard** et **M. Polizzi**. *Gérer la production industrielle : outils et méthodes*. Editions Mare Nostrum, 1996.
11. **D. Brunel** et **J. C. Goaër**. *La théorie des jeux et les tests statistiques*. Editeur Dunod Economica, 1972.
12. **P. Caron, A. Juhel** et **F. Vandeveld**. *Programmation linéaire, méthodes et applications*. Editeur Dunod, 1988.
13. **P. Chretienne, Y. Pesqueux** et **J. C. Grandjean**. *Algorithme et pratique de programmation linéaire*. Edition Technip, Paris, 1980.
14. **A. David, A. hatchuel** et **R. Laufer**. *Les nouvelles fondations des sciences de gestion*. Editeur Vuibert, 2000.
15. **G. Desbazzelle**. *Exercices et problèmes de recherche opérationnelle*. Editeur Dunod, 1976.
16. **M. Drescher**. *Jeux de stratégie. Théorie et application*. Editeur Dunod, 1965.
17. **F. Drosbeke, M. Hallin** et **C. L. Lefevre**. *Programmation linéaire par l'exemple*. Ellipses, Edition Marketing, Paris, 1986.
18. **E. Galdstein**. *Problèmes de programmation linéaire*. Edition Mir, Moscou, 1973.
19. **C. Guéret, C. Prins** et **M. Sevaux**. *Programmation linéaire*. Editions Eyrolles, Paris, 2000.
20. **J. M. Helary** et **R. Pedrono**. *Recherche opérationnelle*. Collection Méthodes, Paris, 1983.
21. **A. Henry**. *Recherche opérationnelle. Programmation linéaire combinatoire. Exercices et problèmes avancés*. Edition Masson, 1976.
22. **A. Kaufman**. *Méthodes et modèles de la recherche opérationnelle*. Tome 1, Editeur Dunod, 1992.