Faculté Abdelmalek Essaadi
Tétouan

1 année DCESS

Année université 2008-2009

Rapport d'exposé sous le thème :

FILE D'ATTENTE

Réalisé par :

- SOUADKA

SOMMAIRE

I - GENERALITES.

La théorie des files d'attente s’attache à modéliser et à analyser de nombreuses situations en apparence très diverses, mais qui relèvent néanmoins toutes du schéma descriptif général suivant. Des clients arrivent à intervalles aléatoires dans un système comportant plusieurs serveurs auxquels ils vont adresser une requête. La durée du service auprès de chaque serveur est elle-même aléatoire. Après avoir été servis (ce qui suppose un arrêt chez un ou plusieurs serveurs selon le cas), les clients quittent le système.

Le but de l'analyse est de caractériser le degré de performance du système en répondant à des questions du type suivant:

• En moyenne, combien de temps attend un client avant d'être servi?
• Quel est le nombre moyen de clients dans le système?
• Quel est le taux d'utilisation moyen des serveurs?
• ...

Illustrons cette description générale par quelques exemples spécifiques.

Agence bancaire

Ici, les serveurs sont les guichets de l'agence. Typiquement, tous les guichets offrent le même service, et chaque client ne devra donc visiter qu'un seul guichet.

Atelier de production

Les 'clients' sont les ordres de fabrication à exécuter, et les 'serveurs' sont les machines nécessaires à l'exécution de chaque ordre de fabrication.

<u>Parking</u>

Les 'clients' sont les véhicules qui cherchent à stationner (plutôt que les occupants, en nombre variable, de ces véhicules). Les 'serveurs' sont les emplacements de parking, et la 'durée de service' est la durée pendant laquelle chaque véhicule reste stationné.

1- Définitions des termes

<u>File d'attente</u> : l'ensemble des clients qui attendent d'être servis, à l'exclusion de celui qui est en train de se faire servir.

<u>Système d'attente</u> : l'ensemble des clients qui font la queue, y compris celui qui se fait servir.

Le **phénomène d'attente** s'étend à tous les clients possibles (dans le cas de systèmes bouclés, où les mêmes clients reviennent plus tard à l'entrée – par exemple les machines qui tombent en panne dans un atelier -, le nombre des clients est, en général, fini). Ces appellations se généralisent et prennent surtout leur intérêt dans les situations où coexistent plusieurs stations et plusieurs files d'attente.

2- Caractéristiques

Une *file d'attente* est constituée de :

a) Un flux d'arrivée qui représente les instants où arrivent les « clients » (terme générique représentant aussi bien des pièces de fabrication, des véhicules, des appels à la mémoire d'un ordinateur, sollicitations de l'organe central d'un réseau téléinformatique, …).

En première approximation on considère souvent que les délais entre les arrivées sont des variables aléatoires indépendantes.

Le flux est alors ce qu’on appelle un *processus de renouvellement stationnaire*. Le cas le plus simple est celui où la loi commune est la loi exponentielle, le flux d’arrivée est alors un processus de Poisson.

b) Un organe de service qui est caractérisé par :

1. Le temps de service : un client qui commence à être servi sera immobilisé pendant un temps aléatoire dont on supposera connue la loi.

2. Le nombre de guichets.

c) Une règle de service ou discipline qui indique comment fonctionne le système:

1. La capacité du système, c’est-à-dire le nombre de clients pouvant être simultanément présents dans le système, est limitée ou non. Dans le premier cas, on suppose que les clients qui arrivent lorsque le système est déjà saturé le quittent immédiatement sans obtenir le service désiré. On dit que ces clients sont *perdus*. Dans le cas d'un système à capacité illimitée, bien sûr, aucun client n'est perdu (mais la longueur des files d'attente peut grandir indéfiniment !).

2. Ordre dans lequel les clients sont servis :

Les clients forment une ou plusieurs files d'attente, éventuellement caractérisées par des *priorités* différentes (par exemple, dans un atelier, commandes urgentes ou non). Au sein de chaque file, le prochain client à servir est sélectionné sur base d'une règle prédéterminée, appelée *discipline de service*. Les disciplines de service les plus courantes sont: premier arrivé premier servi (First In First Out, First Come First Served) (surtout utilisée dans les services, par souci d'équité), temps de service le plus court d'abord (utilisé dans les ateliers de production), dernier arrivé premier servi, sélection aléatoire, etc.

3. Plusieurs classes de clients chacune prioritaire par rapport aux classes suivantes.

4. Le client à son arrivée, si la queue est trop longue, quitte le système avec une certaine probabilité dépendant de la longueur de la queue.

Dans ce bref exposé, nous nous limiterons à l'étude de systèmes

- à serveurs parallèles
- à arrivées et service individuels.

Nous supposerons également que tous les clients demandent le même service et que les serveurs sont identiques.

II – THEORIE ELEMENTAIRE DES FILES D'ATTENTE

1- Mesures de performance - Comportement à long terme.

Nous appelons *état* d'un système à l'instant t le nombre n(t) de clients présents dans le système à cet instant (un client est « présent dans le système » si il est en file d'attente ou en cours de service). Les quantités fondamentales auxquelles s'intéresse l'analyste dans le cadre des modèles de files d'attente sont les *probabilités d'état*, que nous définissons de la façon suivante:

Pour n = 0, 1, 2, ... et t ≥ 0,

pn(t) = probabilité de l'état n à l'instant t (1)

= probabilité que n clients soient présents dans le système à l'instant t

= Pr[n(t) = n].

Sous certaines conditions, les *probabilités à long terme* ou *probabilités stationnaires*

$$Pt= \lim p\ (t). \qquad (2)$$

existent pour n = 0, 1, 2, ... et définissent effectivement une mesure de probabilité :

$$\sum_{n=0}^{\infty} p_n = 1.$$

On interprète alors pn comme la probabilité que, à un instant quelconque dans le long terme, exactement n clients soient présents dans le système. Sous des hypothèses assez faibles, on montre que pn représente également la proportion du temps (à long terme) où le système contient n clients. L'existence des limites (2) implique donc que, dans le long terme, le système atteint un état d'équilibre stable, indépendant de son état initial. Lorsque c'est le cas, on se contentera souvent de la description à long terme du système fournie par (2) plutôt que de calculer les distributions de probabilité (1) qui caractérisent le système dans sa phase 'non stabilisée'.

Lorsque les probabilités (1) et/ou (2) sont connues, il est possible de calculer de nombreuses mesures de performance du système de files d’attente à étudier. Parmi les plus importantes, on considérera les mesures à long terme suivantes:

Ls = nombre espéré de clients dans le système (à un instant quelconque dans le long terme).

Lq = nombre espéré de clients dans la file d'attente (à un instant quelconque dans le long terme).

Ws = temps espéré passé par chaque client dans le système (dans le long terme)

Wq = temps espéré passé par chaque client dans la file d'attente (dans le long terme).

Par définition, on a:

$$\mathbf{Ls} = \sum^{\infty} n p_n \qquad (3)$$

De plus, si le système comporte c serveurs parallèles, alors :

$$\mathbf{Lq} = \sum \mathbf{L\ (n\text{-}c)p_n} \qquad \mathbf{(4)}$$

Puisqu'il y a (n-c) clients dans la file si et seulement si il y a n clients dans le système.

La relation entre les paramètres L et W est plus subtile, mais joue un rôle primordial dans la théorie des files d’attente. Pour la décrire, notons:

λeff = taux d'entrée moyen des clients dans le système (nombre moyen de clients entrant dans le système par unité de temps).

L’équation suivante, désignée sous le nom de ***loi de Little***, lie les paramètres LS, WS et λeff et permet donc de calculer l’une de ces quantités lorsque les deux autres sont connues:

Loi de Little : Pour tout système de files d'attente :

$$\mathbf{Ls} = \boldsymbol{\lambda}_{\mathbf{eff}} \times \mathbf{Ws.} \qquad (5)$$

<u>Exemple</u>

Le gestionnaire d'un petit atelier enregistre en moyenne 5 commandes par jour. Les 6 ouvriers employés dans l'atelier sont très polyvalents, si bien que chaque commande peut être réalisée par n'importe lequel d'entre eux. Néanmoins, le gestionnaire est inquiet car il constate que les ouvriers sont occupés en permanence et que son carnet de commandes contient, en moyenne, une vingtaine de commandes en cours (enregistrées, mais non satisfaites).

Pour mieux comprendre la situation, le gestionnaire aimerait estimer le temps moyen consacré par les ouvriers à chaque commande. Il voudrait également pouvoir annoncer à ses clients, au moment de la passation de commande, un délai de livraison attendu.

On a: $\lambda eff = 5$ et $Ls = 20$. Donc, par la loi de Little, $Ws = 20/5 = 4$: le gestionnaire peut annoncer un délai de livraison de 4 jours.

Par ailleurs, $Ls - Lq = 6$ (puisque les ouvriers sont occupés en permanence). Donc,

$$\mathbf{W_S - W_Q = (L_S - L_Q)/ \lambda_{eff} = 6/5}$$

Et chaque commande requiert en moyenne 1,2 jours de travail effectif.

En pratique, les modèles de files d'attente sont souvent utilisés dans des situations où les seules caractéristiques directement observables ou prédictibles du système concernent le processus d'arrivée des clients et la distribution du temps de service (c'est par exemple le cas dans la phase de conception d'un système). Le but de l'analyse est de calculer les paramètres Ls, Lq, Ws, Wq, ... sur base de ces données.

Dans de nombreux cas, il est impossible de dériver des expressions analytiques pour ces mesures de performance et il faut alors recourir à des procédures de calcul numérique ou à des simulations du système étudié pour estimer leur valeur. Dans d'autres cas, par contre, les processus d'arrivée et de service possèdent des propriétés autorisant une analyse très complète du système. Nous allons maintenant présenter les plus typiques de ces propriétés.

2 - Processus de Poisson.

Nous notons N(t) le nombre d'arrivées de clients survenues dans l'intervalle de temps [0, t), pour $t \geq 0$. La quantité $N(a+t) - N(a)$ représente alors le nombre d'arrivées enregistrées entre les instants a et a+t, pour tout $a \geq 0$ et $t \geq 0$.
En règle générale, pour chaque $t \geq 0$, N(t) est une variable aléatoire. L'ensemble de ces variables aléatoires fournit une représentation mathématique, c'est-à-dire un modèle, des arrivées de clients dans le système. On baptise *processus (stochastique) d'arrivée* cet ensemble $\{ N(t) : t \geq 0 \}$.
Le processus d'arrivée peut bien sûr présenter des caractéristiques variées en fonction de la situation modélisée. Mais il est fréquent dans la pratique que ce processus soit (du moins en première approximation) un *processus de Poisson*, ce qui signifie qu'il existe un paramètre
$\lambda > 0$ (appelé *taux du processus*) tel que:
(i) le nombre d'arrivées dans tout intervalle [a,a+t) de longueur t suit une *loi de Poisson* de moyenne λt, c'est-à-dire: pour a, $t \geq 0$ et n = 0, 1, 2, ...

$$\mathbf{p_n(t)=(l\ t)^n.e^{-lt}/n\ !}$$

(ii) si [a,b) et [c,d) sont des intervalles de temps disjoints, alors le nombre d'arrivées dans [a,b) est indépendant du nombre d'arrivées dans [c,d).
(iii) $N(0) = 0$.

Quelques commentaires s'imposent au sujet de cette définition. La propriété (iii) est évidemment requise si { N(t) : t ≥ 0 } doit pouvoir être interprété comme un processus d’arrivée. Par contre, la propriété (ii) n’est pas satisfaite par tous les processus d’arrivée: intuitivement, elle implique que le nombre d'arrivées enregistrées dans un intervalle [a,b)
(par exemple, [0,5)) ne fournit aucune indication sur le nombre d'arrivées à attendre dans un intervalle disjoint [c,d) (par exemple l’intervalle [5,8)).
La propriété (i), quant à elle, est encore plus restrictive et beaucoup moins intuitive. Elle implique en particulier que le processus d’arrivée est *stationnaire*, ce qui veut dire que les nombres d’arrivées enregistrés dans deux intervalles distincts de même durée suivent la même loi de probabilité (ceci, en dépit de leur indépendance mutuelle, exprimée par la propriété (ii)). Pour un intervalle de longueur t fixée, disons [0,t), le nombre espéré d'arrivées se calcule comme suit:

$$\mathbf{E[N(t)]= \lambda t.e^{-\lambda t}\sum \lambda t^{n-1}/(n -1)!}$$

3 - Processus de service.

Nous avons centré la discussion précédente autour de la modélisation du processus d’arrivée des clients, mais les mêmes concepts s’appliquent à la modélisation du processus de service. Plus précisément, on décrit généralement le processus de service par une suite de variables aléatoires X1, X2, ..., où Xi est la *durée de service* du client i (i = 1, 2, ...). Si l’on suppose que chaque serveur fournit le même service, il est assez naturel d’admettre que les variables X1, X2, ... sont indépendantes et identiquement distribuées. Ainsi, dans la pratique des modèles de files d'attente, on posera souvent l’hypothèse que les durées de service des différents clients sont des variables indépendantes suivant toutes la même distribution exponentielle de paramètre μ (dans toute application spécifique, cette hypothèse devrait être infirmée ou validée par un test d'hypothèse classique, comme par exemple le test du chi-carré). On dit que μ est le *taux de service moyen* des serveurs.

Pour compléter cette discussion, supposons que le processus de service d'un serveur unique satisfait aux hypothèses précédentes: en particulier, les durées X1, X2, ... sont indépendantes et suivent toutes une loi exponentielle de même paramètre μ. Remarquons que, *si le serveur est occupé en permanence*, alors ces durées peuvent être vues comme des 'durées entre sorties successives' du système (par analogie avec les 'durées entre arrivées successives' dans le système). Notons S(t) le nombre de clients servis par le serveur dans l'intervalle de temps [0,t) ou, de façon équivalente, le nombre de sorties de clients enregistrées dans l'intervalle [0,t). Si le serveur est occupé en permanence, la Proposition 2 implique donc que { S(t) : t ≥ 0 } est un processus de Poisson de taux μ, et μ représente l'espérance du nombre de clients servis par unité de temps.

En général, cependant, *le serveur n'est pas occupé en permanence* (en raison du caractère aléatoire du processus d'arrivée, ou par suite d'un taux d'arrivée relativement faible). Dans ce cas, le processus de service diffère du processus de sortie et, en particulier, le taux de sortie du système est inférieur à μ. Il est utile de bien comprendre cette distinction entre taux de service et taux de sortie; nous y reviendrons dans notre discussion des processus de naissance et de mort.

4 - Notations de Kendall.

Suivant une suggestion de Kendall, on décrit souvent les caractéristiques essentielles d'un système de files d'attente par une notation abrégée de la forme a/b/c/N, où

• $a \in \{M, G, ...\}$ précise le processus d'arrivée: M (Markovien) pour un processus de Poisson, G (général) pour un processus quelconque non précisé, ...;
• $b \in \{M, G, ...\}$ précise la distribution des durées de service: M (Markovien) pour des durées indépendantes et distribuées exponentiellement avec moyenne identique pour chaque serveur, G pour des durées quelconques, ...;
• $c \in \{1, 2, ..., \infty\}$ indique le nombre de serveurs;
• $N \in \{1, 2, ..., \infty\}$ précise la capacité du système, c'est-à-dire le nombre de clients qui peuvent simultanément s'y trouver.
Le paramètre N est souvent omis lorsqu'il vaut $+\infty$. Ainsi, on parlera de systèmes M/M/1,
M/M/5/8, M/G/1, etc. Par ailleurs, la notation a/b/c/N est parfois complétée par des champs supplémentaires indiquant par exemple la discipline de service (FIFO, aléatoire, ...), le nombre de clients existant dans l'univers considéré, etc.
Dans la suite de cet exposé, nous nous limiterons à l'étude des systèmes de type M/M/c/N (systèmes de file d'attente à arrivées poissonniennes et à durées de service exponentielles) pour lesquels les processus de naissance et de mort fournissent un cadre d'analyse uniforme.

5 - Processus de naissance et de mort: le cas général.

Rappelons que l'*état* d'un système de files d'attente à l'instant t, noté n(t), est simplement le nombre de clients présents dans le système à l'instant t. L'ensemble des variables aléatoires d'état décrit un processus stochastique $\{n(t): t \geq 0\}$.
Considérons à présent un système très général, dans lequel nous ferons abstraction (du moins, à première vue) des caractéristiques telles que nombre de serveurs, capacité, etc. On peut utilement visualiser un tel système comme une 'boîte noire', simplement caractérisée par un processus d'arrivée, un processus de sortie et un processus d'état résultant de la combinaison des arrivées et des départs : à chaque

instant $t+\Delta t$, l’état $n(t+\Delta t)$ du système résulte des arrivées et sorties enregistrées entre t et $t+\Delta t$. Sans être parfaitement rigoureuse, la définition suivante permet d’introduire les caractéristiques d’un tel système auxquelles nous allons nous intéresser.

Définition. Le processus d’état stochastique $\{n(t): t \geq 0\}$ est un *processus de naissance et de mort* si, pour chaque n = 0, 1, 2, ..., il existe des paramètres λ_n et μ_n (avec $\mu_o = 0$) tels que, lorsque le système est dans l'état n, le processus d'arrivée est poissonnien de taux λ_n et le processus de sortie est poissonnien de taux μ_n.

Dans un processus de naissance et de mort, les taux d’arrivée et de service sont donc variables en fonction de l’état du système. Ceci semble nous éloigner des modèles de files d’attente M/M/c/N, puisque le taux d’arrivée des clients et le taux de sortie sont apparemment constants dans ces modèles (égaux à λ et μ, respectivement). Mais en réalité, les processus de naissance et de mort fournissent un cadre d’analyse idéal pour les modèles
M/M/c/N, dans lequel les notions de « nombre de serveurs » et de « capacité du système » peuvent être traitées à un niveau d'abstraction élevé, à travers des variations des taux d'arrivée et de sortie. Plus précisément, nous montrerons ci-dessous que le processus $\{n(t): t \geq 0\}$ associé à un système M/M/c/N est toujours un processus de naissance et de mort.

III – ETUDE DE CAS

Enoncé du problème :

Les ouvriers peuvent se rendre chez le médecin d'entreprise pendant les heures de travail. Le chef du personnel demande une étude relative au fonctionnement du cabinet du médecin.

On recherchera donc le temps moyen d'attente des ouvriers et la durée moyenne de chaque consultation.

Résolution :

Il s'agit d'un problème d'attente à une station. Il faut tout d'abord restreindre l'étude à la période où le phénomène est stationnaire : durant la première demi-heure et la dernière de la journée, ainsi que l'heure du repas, l'affluence est très fluctuante et faible. Au contraire, durant les autres périodes de la journée, le phénomène stationnaire est établi. On peut donc étudier le phénomène d'un point de vue statistique dans cette période horaire.

a) Décompte des arrivées :

Pendant 100 intervalles de 5 minutes (successifs ou non) tous situés dans la période de stationnarité, le nombre de patients arrivant durant chaque période de 5 minutes est décompté.

Supposons que les résultats soient les suivants :

Nombre d'ouvriers arrivant pendant une période de 5 minutes	Fréquences observées f_n
0	29
1	34
2	24
3	9
4	3
5	1
6	0
	Total = 100

La moyenne de cette loi de distribution est facile à calculer :

SOMME($n.f_n$/100) = 1/100.(0 x 29 + 1 x 34 + 2 x 24 + 3 x 9 + 4 x 3 + 5 x 1) =1,26

Optimisation du nombre de guichetiers

L'emploi d'un test statistique va permettre de vérifier si la loi observée se rapproche d'une loi théorique classique (la loi de Poisson).

La formule $p_n(t)$ = (l.t).e-l.t/n! avec l.t = 1,26 et e-l.t = 0,2837, permet de calculer les fréquences théoriques d'une loi de poisson de moyenne 1,26 ; ce sont 100 fois les valeurs :

p_0 = 0,28 ; p_1 = 0,36 ; p_2 = 0,23 ; p_3 = 0,09 ; p_4 = 0,03 ; p_5 = 0,01 ;

(Cette dernière probabilité est en fait la différence entre 1 et la somme des autres probabilités jusqu'à 4).

Employons le test du χ^2 de Pearson qui s'applique à des classes dans lesquelles le nombre d'évènements attendus est de l'ordre de 5 (il nous suffira de regrouper les deux dernières classes pour satisfaire grossièrement à cette exigence).

Ce test consiste à calculer les carrés des différences entre les fréquences théoriques de la loi à vérifier et les fréquences observées, puis à diviser chacun de ces carrés par la fréquence théorique de la classe à laquelle les fréquences sont relatives.

n	**Fréquences observées (f_n)**	**Fréquences observées (f_{th})**	**|d|**	**D²**	**D²/ f_{th}**
0	29	28	1	1	0.0357
1	34	36	2	4	0.1111
2	24	23	1	1	0.435
3	9	9	0	0	0
> 4	4	4	0	0	0
					0,1903 (somme)

Le nombre de degrés de liberté du système est : d = c - 1 - p, c étant le nombre de classes (ici 5) et p le nombre de paramètres de la loi théorique (ici 1 : la moyenne). On a donc : d = 5 - 1 - 1 = 3

Or, en se reportant à une table des χ^2, on trouve, pour 3 degrés de liberté :

$\chi^2(0,95) = 0,352$

Comme le χ^2 calculé (0,19) est inférieur àcette valeur, on décide d'admettre que le phénomène suit une loi de Poisson (en réalité, le test signifie que, si l'hypothèse était exacte, pour 95% des échantillons, la valeur χ^2 serait inférieure à 0,352 ; le risque de rejeter l'hypothèse alors qu'elle est exacte, si le χ^2 calculé est supérieur au

$\chi^2(0,95)$ relevé dans la table, n'est que 5%).

La conclusion à tirer de l'éude des arrivées est qu'elles se font "à la Poisson", avec un taux de 1,26/5 = 0,25 arrivée par minute.

c) Décompte des services

A 100 reprises, consécutives ou non, mais prises dans la période stationnaire, on relève la durée du service, c'est-à-dire le temps passé par un client au guichet.

Supposons que les résultats obtenus soient les suivants :

Durée des services	Nombre de services observés
< 1 min.	23
de 1 à 2 min.	20
- 2 - 3 -	14
- 3 - 4 -	12
- 4 - 5 -	9
- 5 - 6 -	5
- 6 - 7 -	4
- 7 - 8 -	5
- 8 - 9 -	3
- 9 - 10 -	2
- 10 - 11 -	2
- 11 - 12 -	1
> 12 min.	0

La moyenne est ici :

1/100 (0,5x23 + 1,5x20 + 2,5x14 + 3,5x12 + 4,5x9 + 5,5x5 + 6,5x4 + 7,5x5 + 8,5x3 + 9,5x2 + 10,5x2 + 11,5x1) = 3,27 mn

Le nombre de clients servis par minute est donc : 1/3,27 = 0,30

On fait alors l'hypothèse que la loi des services est une loi exponentielle de taux μ =0,30.

En réduisant à neuf classes et 7 degrés de liberté, on a χ^2 calculé = 0,9609 ;

Le χ^2 (0,95), lu dans la table, s'élève à 2,167 et l'analyste décide d'admettre qu'il a bien une loi exponentielle de taux :

μ =0,30.

Les formules données précédemment permettent de calculer :

- le nombre moyen de clients dans la file :

  $n = (\lambda /\mu) / (1 - \lambda /\mu) = (0,25/0,30) / (1 - 0,25/0,30) = 0,25 / (0,30 – 0,25) =5;$

- le temps moyen d'attente dans la file :

  $t_f = n /\mu = 5 / 0,30 = 16$ mn 2/3

On se rend compte que le temps perdu par les "clients" (les ouvriers de l'entreprise qui, pendant ce temps, ne prennent pas part à la production) est considérable ; à 8 heures par jour et 0,25 client par minute, attendant chacun en moyenne 16 2/3 mn, cela fait :

8 x 60 x 0,25 x 3,27 mn = 6 h ½

e) Optimisation du nombre de guichetiers

La " solution par les moyennes " consisterait pratiquement à adopter un nombre de guichetiers permettant d'assurer le service et l'on trouverait ainsi que l'unique préposé suffit à sa tâche.

La solution économique, fondée sur la recherche opérationnelle, a pour but d'établir un bilan total de l'opération qui consisterait à embaucher 1,2, … autres préposés (cette décision accroissant le coût de fonctionnement du service), en vue de diminuer le temps perdu par les ouvriers (ce résultat faisant réaliser un gain sur le temps total de travail, donc sur la production). Définir le nombre d'employés correspondant au bilan optimal est justement l'objet du problème.

Il ne faudrait surtout pas imaginer que si, par exemple, on double le nombre de guichetiers, le nombre moyen d'ouvriers dans la file diminuera de moitié et qu'en conséquence le temps d'attente sera également diminué de moitié.

L'établissement des formules, par des méthodes analogues à celles qui ont été utilisées plus haut, est quelque peu laborieux ; aussi nous contenterons nous de les donner directement ici.

Appelant S le nombre de stations (ici : le nombre de guichetiers), et Ψ le rapport λ / μ , on a :

- probabilité d'une attente nulle (probabilité de l'absence de tout client dans la file) :

$$P_0 = \frac{1}{\frac{\Psi^S}{S!(1-\Psi/S)^2} + \sum_{0}^{S-1} \frac{\Psi^n}{n!}}$$

- attente moyenne dans la file :

$t_f = (\Psi^s \times P_0) / [\ \mu\, SS\,!\ (1 - \Psi / S)^2\ ]$

- nombre moyen de clients dans la file :

$\underline{n} = \Psi\ [\mu\, t_f + 1\ ]$

On désigne, de plus, par φ le taux moyen d'inactivité des stations.

Dans le cas qui nous occupe, on calcule facilement :

- pour S = 1,

$p_0 = 1 - \Psi = 1 - 0,833 = 0,167$

- pour S = 2,

$p_0 = 1/[\ (O{,}833^2 / 2\,!\ (1 - 0{,}833 / 2)) + 1 + 0{,}833 / 1\,!\ ] = 0{,}411$

- pour S = 3,

$p_0 = 1/[\ (O{,}833^3 / 3\,!\ (1 - 0{,}833 / 3)) + 1 + 0{,}833 / 1\,! + 0{,}833^2 / 2\,!] = 0{,}432$

On peut alors calculer aussi :

- pour S =1, $\underline{t_f}$ = 16 mn 2/3 φ = 0,167 ;
- pour S =2, $\underline{t_f}$ = 0,70 mn φ = 1,167 ;
- pour S =3, $\underline{t_f}$ = 0,09 mn φ = 2,167 ;

Supposant alors que le coût de l'inactivité des employés représente le salaire (et les charges) correspondant à la durée de cette inactivité, soit, par exemple, 5 F de l'heure et que le temps perdu en attente par les ouvriers peut être évalué à 12 F de l'heure (compte tenu, cette fois, de la perte à la production), on obtient les coûts totaux suivants :

Pour S = 1, Γ (1) = 120 x 16x2/3 x 12/60 + 480 x 0,167 x 5/60 = 406,7 F

Pour S = 2, Γ (2) = 120 x 0,70 x 12/60 + 480 x 2,167 x 5/60 = 88,80 F

On constate qu'il suffit de doubler le guichetier pour atteindre la perte minimale (il faudrait connaître le taux d'absentéisme des employés pour être à même d'examiner s'il ne serait pas plus rationnel, finalement, de prévoir trois préposés).