Evaluation des Performances des Systèmes Informatiques

Master Informatique

Alexandre Aussem

LIRIS UMR 5205 CNRS
Data Mining & Machine Learning Group (DM2L)
University of Lyon 1
Web: perso.univ-lyon1.fr/alexandre.aussem

February 1, 2019

Université Claude Bernard Lyon 1

Outline

Présentation du cours

- Total de 30 heures de cours :
  - 12h CM (3 séances de 4h),
  - 6h TD (2 séances de 3h)
  - 12h TP (3 séances de 4h)
- Travaux pratiques :
  - TP avec NS2. Modèles de trafics et d'applications (TCP/IP, FTP, Telnet etc.).
  - Simulation d'un réseau de files d'attente. Routage, queueing et QoS.

Analyse opérationnelle

- Technique **non probabiliste** basée sur des mesures élémentaires du système (compteurs de requêtes)
- Permet de **quantifier des critères de performance** du système en cours de fonctionnement (opérationnel)
- Exploite des relations fondamentales entre les critères de performance
- Système vu comme une **boîte noire.**
- Première ébauche d'analyse de performance

Les critères de performance

- Le **temps de réponse** $E[T]$ où T = temps séparant l'arrivée d'une requête de la fin de son traitement.
- Le **débit** $E[N]$ où N = nombre de requêtes traitées par unité de temps
- Le **taux d'occupation**, U = Probabilité qu'une ressource soit occupée
- Le **taux de perte/rejet** (de paquets) = Probabilité qu'un paquet soit perdu
- Probabilité = fréquence si ergodicité du processus.

Evaluation des performances

Cherche à **quantifier** les critères de performance.

- Guichet SNCF
  Temps d'attente moyen, nombre moyen de clients, débit d'un guichet etc.
- Réseaux de communication
  Débit en paquets/cellules, taux de perte, de retransmission etc.
- Atelier de production
  Taux d'utilisation d'une machine, temps de fabrication

Les méthodes

- La **résolution mathématique**
  Obtenir des formules, sous des hypothèses simplificatrices, pour les critères de performance de façon à les optimiser
- La **mesure** (métrologie)
  Ajuster le modèle à la réalité. Attention, l'outil de mesure peut influer sur le système.
- La **simulation**
  Estimer les probabilités par des moyennes dans le temps. Attention aux événements rares.

Formule de Little opérationnelle

Definition

Système = mécanisme recevant des requêtes et les restituant à l'issue d'un temps de traitement

- On ne connaît le système que via 2 compteurs A (arrivées) et D (départs)
- Aucune hypothèse (ordre de traitement, parallélisme etc.)

Formule de Little Opérationnelle

- Mesures élémentaires :
  T = durée de la mesure
  A = nombre total d'arrivées de requêtes
  D = nombre total de départs de requêtes
  $T(n)$ = durée cumulée pendant lequel le système a contenu n requêtes
- On cherche les critères :
  Λ = débit
  L = nombre moyen de requêtes dans le système
  R = temps de réponse moyen dans le système
- **Quel lien entre Λ, L et R ?**

Formule de Little Opérationnelle

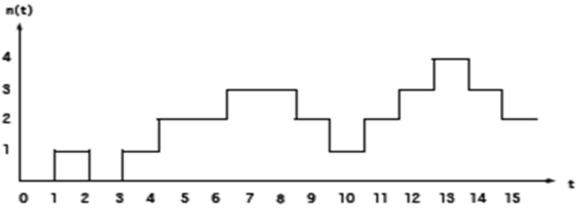
- Mesures élémentaires :
  T = durée de la mesure
  A = nombre total d'arrivées de requêtes
  D = nombre total de départs de requêtes
  $T(n)$ = durée cumulée pendant lequel le système a contenu n requêtes
- On cherche les critères :
  Λ = débit
  L = nombre moyen de requêtes dans le système
  R = temps de réponse moyen dans le système
- On en déduit :

Formule de Little

$$L = \Lambda \cdot R$$

Formule de Little Opérationnelle

Exemple d'évolution d'un système



T=15
A=7
D=5

T(0)=2
T(1)=3
T(2)=5
T(3)=4
T(4)=1

$$\Rightarrow \begin{cases} \Lambda = \dfrac{5}{15} \\ L = \dfrac{29}{15} \\ R = \dfrac{29}{5} \end{cases}$$

Temps de réponse d'un réseau local

- Des terminaux accèdent à un serveur. A chaque terminal, on associe un processus alternant entre 2 phases :
  Réflexion : l'utilisateur réfléchit ou frappe au clavier (avant le ENTER)
  Traitement : la requête est traitée par le serveur, attente de réponse.
- Mesures élémentaires :
  T = durée de la mesure
  N = nombre de terminaux
  A = nombre de requêtes envoyées depuis les terminaux
  D = nombre de requêtes traitées par le système
  $r(k)$ = durée cumulée passée en traitement par le processus k
  $z(k)$ = durée cumulée passée en réflexion par le processus k

Temps de réponse d'un réseau local

- On cherche les critères :
  Λ = Débit du système
  Z = temps de réflexion moyen sur les terminaux
  R = temps de réponse moyen dans le système
- **Quel lien entre R, Λ et Z ?**

Temps de réponse d'un réseau local

- On cherche les critères :
  Λ = Débit du système
  Z = temps de réflexion moyen sur les terminaux
  R = temps de réponse moyen dans le système
- Temps de réponse si A/D proche de 1 :

Temps de réponse

$$R = \frac{N}{\Lambda} - Z$$

Relation d'équilibre d'un système

- On considère un système constitué de **plusieurs stations mono-serveur** de traitement. Les requêtes envoyées au système (travaux) engendrent des requêtes élémentaires. Un **travail** peut engendre plusieurs **requêtes** et celles-ci peuvent être traitées simultanément sur différentes stations.
- Mesures élémentaires :
  T = durée de la mesure
  D = nombre total de requêtes globales traitées par le système
  D_i = nombre total de requêtes élémentaires traitées par la station i
  $T_i(n)$ = durée cumulée pendant laquelle la station i a contenu n requêtes

Relation d'équilibre d'un système

- On cerche les critères :
  Λ = Débit global du système
  Λ_i = Débit de la station i en requêtes élémentaires
  U_i = taux d'occupation de la station i
  e_i = nombre moyen de visites à la station i
  S_i = durée moyenne de service à la station i
  R_i = temps de réponse d'une requête élémentaire à la station i
  L_i = nombre moyen de requêtes élémentaires dans la station i

Relation d'équilibre d'un système

- Hypothèses : stations 'mono-serveur'

Théorème de *Chang-Lavenberg* opérationnel

$$\Lambda = \frac{\Lambda_i}{e_i} = \frac{U_i}{S_i e_i} = \frac{L_i}{R_i e_i}$$

- $S_i e_i$ = temps total de service demandé à la station i

Relation d'équilibre d'un système

- Si hypothèses supplémentaires : si une requête globale n'engendre pas simultanément plusieurs requêtes élémentaires alors :

$$
\begin{aligned}
L &= \sum_i Li \\
R &= \sum_i Ri \cdot e_i \\
L &= \Lambda \cdot R \\
L_i &= \Lambda_i \cdot R_i
\end{aligned}
$$

Relation d'équilibre d'un système

- Si populations distinctes, se restreindre à chaque population.
- Au niveau de chaque station : additivité des débits et des taux d'occupation

$$\begin{aligned} \Lambda_s &= \sum_j \Lambda_s^j \\ U_s &= \sum_j U_s^j \end{aligned}$$

où j désigne une population et s la station

Saturation d'un système

- Si hypothèses supplémentaires : si une requête globale n'engendre pas simultanément plusieurs requêtes élémentaires alors :
- Comment se comporte le système **lorsque la charge augmente** ?

$$\Lambda_{max} = \frac{1}{\max_i\{S_i e_i\}}$$

- Hypothèse : les $S_i e_i$ sont insensibles à la charge La station i est le **goulot d'étranglement** du système Λ_{max} est le débit maximal théorique acceptable

Limites de l'analyse opérationnelle

- Basée sur des mesures élémentaires du système, exploite des relations d'équilibre fondamentales entre les critères de performance
- Dans le case d'une FA dont on connaît le débit d'arrivée L et le temps moyen de service S

$$
\begin{aligned}
U &= \Lambda S, \\
L &= \Lambda R
\end{aligned}
$$

- On en déduit : $\Lambda \leq \frac{1}{S}$
- Mais R et L ? Nécessiter d'analyser plus finement les interactions entre arrivées et services, formuler des hypothèses de nature statistique

Illustration

- Dans les locaux de l'UFR, il y a une salle en libre service où sont mises à la disposition des étudiants une photocopieuse 'noir & blanc' notée P1 et une photocopieuse 'couleur' notée P2.
- Deux types de populations étudiantes utilisent ces photocopieuses : les étudiants M1 et les étudiants M2.
- Les M1 effectuent en moyenne 30 photocopies noir et blanc et 5 photocopies couleur.
- Les M2 effectuent en moyenne 20 photocopies noir et blanc et 10 photocopies couleur.
- P1 est utilisée dans 70% du temps et P2 dans 60% du temps.
- Le temps pour effectuer une photocopie est de $1s$ sur P1 et de $2s$ sur P2.

Illustration

- Quels sont les débits des étudiants M1 et M2 ?
- Quels sont les débits des photocopieuses P1 et P2 ?
- Si le débit M1 augmente mais pas celui de M2, quelle photocopieuse va saturer ?
- Quel sera alors le débit max de M1 ?
- Si on change P1 par une autre photocopieuse 2 fois plus rapide, quel sera le débit max de M1 ?

Marches Aléatoires

Définition

- Les marches aléatoires sont une introduction aux **processus stochastiques** et aux **files d'attente.** Considérons le jeu suivant : un joueur lance successivement et indépendamment n fois une pièce de monnaie. Chaque fois qu'il obtient Pile, il gagne 1 point, chaque fois qu'il obtient Face, il perd un point.
- Soit $p \in]0,1[$ la probabilité d'obtenir Pile. $S_k = x_0 + X_1 + X_2 + \cdots + X_k$,
- S_k représente la fortune du joueur après le kième lancer, x_0 étant la fortune initiale, $X_k = \pm 1$.

Marches Aleatoires

Propriétés

Theorem

Le nombre de trajectoires de $M(k_1, y_1)$ à $N(k_2, y_2)$, noté $\mathcal{T}_{(k_1,y_1),(k_2,y_2)}$ est $C^g_{k_2-k_1}$ où $g = 1/2 \times (k2 - k1) + (y_2 - y_1)$.

Theorem

Principe de réflexion : Soit une trajectoire $\mathcal{T}$ de $M(k_1, y_1)$ à $N(k_2, y_2)$ touchant l'axe des abscisses pour la première fois en $k = p$. On lui associe la trajectoire T' de $M'(k_1, -y_1)$ à $N(k_2, y_2)$ telle que $s'_k = -s_k$ pour $k = 1 < p$ et $s'_k = s_k$ pour $k = 1 \geq p$. On a alors une bijection entre l'ensemble T et T'.

Marches Aleatoires

Propriétés

Corollary

La probabilité d'aller de M à N sans toucher l'axe des abscisses est $1 - C_{k_2-k_1}^{g'} / C_{k_2-k_1}^{g}$ *où* $g' = \frac{(k2-k1)+(y_2+y_1)}{2}$.

Marches Aleatoires

Exemples

Exemple

Au cours d'un scrutin opposant deux candidats X et Y, X a obtenu 600 voix et Y 400. Quelle est la probabailité que X ait été majoritaire tout au long du scrutin ?

Exemple

100 personnes font la queue à un guichet de cinéma; la place cote 5 euros. 60 personnes ont un billet de 5 euros en poche, les 40 autres n'ont chacune qu'un billet de 10 euros. Combien faut-il de billes de 5 euros dans la caisse pour que, avec une probabilité d'au moins 98%, chacun soit servi dès qu'il se présente ?

Processus stochastique

Definition

Un processus stochastique est une famille de variables aléatoires $\{X_t, t \in T\}$ définie sur un même espace probabilisé et indexée par $t \in T$ (t est la paramètre ”temps”).

- Les valeurs prises sont les états et leur ensemble est appelé espace d'état.
- L'espace d'état peut être discret ou continu.
- Le temps peut être discret ou continu: chaîne (à temps discret) et processus (à temps continu)
- Dans ce cours, on ne parlera que de chaînes et de processus à espace d'état discret.

Chaîne de Markov

Definition

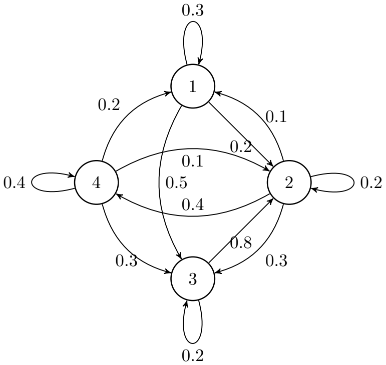
Un processus stochastique $\{X_t, t \in T\}$ est une chaîne de Markov si la probabilité d'être d'en l'état j à l'instant $t+1$ ne dépend que de l'état i à l'instant t.

$$P(X_{n+1} = j | X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(X_{n+1} = j | X_n = i_n)$$

- une chaîne de Markov est dite **sans mémoire.** Connaissant le présent, l'état futur ne dépend pas du passé.
- Si de plus $P(X_{n+1} = j | X_n = i_n) = p_{ij}$ ne dépend pas de n alors la chaîne de Markov est dite **homogène.**

Illustration

Chaîne de Markov



Matrice de transition associée

$$P = \begin{pmatrix} 0.3 & 0.2 & 0.5 & 0 \\ 0.1 & 0.2 & 0.3 & 0.4 \\ 0 & 0.8 & 0.2 & 0 \\ 0.2 & 0.1 & 0.3 & 0.4 \end{pmatrix}$$

Tel que $\forall i, \ \sum_j P_{ij} = 1$

Chaîne de Markov

- Les marches aléatiores $S_n = \sum_{k=0}^{n} X_k$ où X_k à valeur dans $\mathbb{Z}$, est un cas particulier de chaîne de Markov.
- Notons $p_{ij}^{(n)} = P(X_{k+n} = j | X_k = i)$, la probabilité d'aller de i à j en n étapes;
- On a $p_{ij}^{(n)} = \sum_{k \in E} p_{ik}^{(n-1)} p_{kj}$.
- Cette relation entraîne:

Matrice de transition après n itérations:

$$P^{(n)} = P^n.$$

Chaîne de Markov

- Posons $\pi_j(n)$ la probabilité d'être dans l'état j à l'intant n.
- Si $\pi(n) = (\pi_1(n), \ldots, \pi_m(n))$ avec $card(E) = m$ alors
- $\pi(0)$ est le vecteur d'état à $t = 0$.

Le vecteur d'état après n itérations:

$$\pi(n) = \pi(n-1) \cdot P = \pi(0) \cdot P^n$$

Etats récurrents et transitoires

- Deux types de problèmes se posent dans l'étude des chaînes de Markov.
  - Comportement asymptotique ($n \to \infty$) de la chaîne (X_n) et de leurs lois P^n,
  - Passage éventuel de la chaîne (X_n) par un état donné.
- On note $P_x(\cdot) = P(\cdot|X_0 = x)$. Un état est dit récurrent si la probabilité $\pi_x = P_x(\exists n \geq 1 : X_n = x) = 1$. Si $\pi_x < 1$ alors l'état est dit transitoire.
- Si N_x est le nombre de passages en x :

$$N_x = \sum_{n \in \mathbb{N}} \mathbf{1}_{\{X_n = x\}}$$

Etats récurrents et transitoires

- On démontre que
  - si x est **récurrent** ($\pi_x = 1$) alors $P_x(N_x = \infty) = 1$
  - si x est **transitoire** ($\pi_x < 1$) alors $P_x(N_x < \infty) = 1$ et N_x suit une loi géométrique $P_x(N_x = k) = (1 - \pi_x)\pi_x^k$.
- Si x est **transitoire**, $E_x(N_x) = \sum_{k \geq 1} k(1 - \pi_x)\pi_x^{k-1} = \frac{1}{1-\pi_x} < \infty$.
- Si x est **récurrent**, $E_x(N_x) = \infty$
- Si E est fini, alors toute chaîne de Markov possède au moins un état récurrent.

Etats récurrents et transitoires

- Posons π_x^n = P(le premier retour en x, partant de x, se produise au bout de n pas).
- On a clairement $\pi_x = \sum_{n=1}^{\infty} \pi_x^n$
- Si x est récurrent, on définit $R_x = \sum_{n=1}^{\infty} n\pi_x^n$ = **temps moyen de retour en** x.
  - si $R_x = \infty$, alors l'état x est **récurrent nul**
  - si $R_x < \infty$, alors l'état x est **récurrent non nul** (le cas qui nous intéresse)

Etat d'équilibre stationnaire

Definition

Si tous les états d'une chaîne sont récurrents non nuls et apériodiques, alors la chaîne est dite **ergodique**.

- Autre définition équivalente :

Definition

Si P^n tend vers une limite P^∞ et tous ses éléments de P^∞ sont strictement positifs, alors la chaîne est dite **ergodique**. On dit aussi que P est **régulière**.

Etat d'équilibre stationnaire

Theorem

Si P est une matrice de transition régulière (ou si la chaîne de Markov est ergodique)

- $\pi(n)$ *admet un point fixe unique* $\pi^\star$ *qui est un vecteur de probabilité à composantes strictement positive*
- $P^\infty = lim_{n\to\infty} P^n$ *a pour lignes les éléments de* $\pi^\star$
- $\pi(0)$ *étant un vecteur de probabilité quelconque,*

$$lim_{n\to\infty}\pi(0)\cdot P^n = \pi^\star \quad \text{ou encore} \quad \pi^\star\cdot P = \pi^\star$$

- $\pi_i^\star$ *est la proportion de temps passé dans l'état* i *quand le* $t\to\infty$

Exercice de cours

On se donne la matrice de transition :

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

1. Montrer qu'il existe un équilibre stationnaire $\pi^\star$ et calculer-le.
2. En déduire l'expression de P^∞.
3. Montrer qu'ont atteint $\pi^\star$ indépendamment de $\pi(0)$.
4. Quelle est la proportion de temps passé dans l'état 1 ?
5. Quel est la loi du temps de séjour dans un état ?
6. Quel est le temps de séjour moyen dans un état?

Processsus de Markov

Definition

$\{X_t, t \in \mathbb{R}\}$ est une chaîne de Markov à temps continu si

- $\forall (t_1, \ldots, t_n) \in \mathbb{R}^n$ tels que $t_1 < t_2 < \ldots < t_n$
- $\forall (i_1, \ldots, i_n) \in \mathbb{E}^n$
- $P(X_{t_n} = j | X_{t_1} = i_1, X_{t_2} = i_2, \ldots, X_{t_{n-1}} = i_{n-1}) = P(X_{t_n} = j | X_{t_{n-1}} = i_{n-1})$

- On note $p_{s,t}(i,j) = P(X_t = j | X_s = i)$
- Si de plus $p_{s,t}(i,j) = p_{t-s}(i,j)$ alors le processus est **homogène** en temps.

Equation de Chapman-Kolmogorov

- En écriture matricielle, $P_{t-s} = \{p_{t-s}(i,j), i,j \in E\}$

Equation de Chapman-Kolmogorov

$$P_{t-s}(i,j) = \sum_{k \in E} P_{u-s}(i,k) P_{t-u}(k,j)$$

- $P_{v+w} = P_v \cdot P_w$
- P_s est la matrice de transition entre t et $t+s$.
- On appelle Q la matrice dont les termes sont :
  - $q_{ij} = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t}$
  - $q_{ii} = \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - 1}{\Delta t}$

Temps de séjour

- Supposons $X_0 = x$. Soit T le premier instant où X_t n'est plus égal à x (donc le temps de transition vers de x vers $y \neq x$).
- On montre que $P_x(T > t) = \exp(-\lambda_x t)$ où λ_x est un paramètre strictement positif.
- La loi exponentielle joue le même rÙle que la loi géométrique, ce sont des lois sans mémoire.

Propriété sans mémoire

- Cas discret : loi géométrique.

$$
\begin{aligned}
P(T = n + k | T > n) &= \frac{P(T = n + k)}{P(T > n)} \\
&= \frac{(1-p)^{n+k} \cdot p}{(1-p)^n} \\
&= (1-p)^k \cdot p = P(T = k)
\end{aligned}
$$

Propriété sans mémoire

- Cas continu : loi exponentielle

$$
\begin{aligned}
P(T > t_1 + t_2 | T > t_1) &= \frac{P(T > t_1 + t_2)}{P(T > t_1)} \\
&= \frac{\exp\left(-\lambda(t_1 + t_2)\right)}{\exp\left(-\lambda t_1\right)} \\
&= \exp\left(-\lambda t_2\right) \\
&= = P(T > t_2)
\end{aligned}
$$

Régime permanent

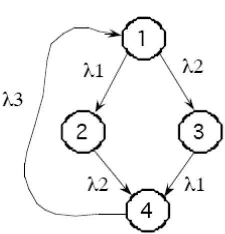
Theorem

Si la chaîne de Markov à temps continu, homogène, est irréductible (partant d'un état, on peut rejoindre tout autre état en un temps fini avec une proba non nulle) alors

- $\pi(t) = \{P(X(t) = 1), \ldots, P(X(t) = m)\}$ *admet une distribution stationnaire* $\pi^\star$ *telle que* $\lim_{t\to\infty} \pi(t) = \pi^\star$.
- $\pi^\star$ *vérifie* $\pi^\star \cdot Q = 0$ *et* $\sum_i \pi_i^\star = 1$.
- $\pi_i^\star$ *est la proportion de temps passé dans l'état* i *quand* $t \to \infty$

On sait que $\pi(t+u) = \pi(t)P(u)$ donc en régime stationnaire il vient : $\pi^\star = \pi^\star P(u)$ et donc $\forall u, \pi^\star(\frac{P(u)-I}{u}) = 0$. Quand $u \to 0$ on retrouve bien $\pi^\star \cdot Q = 0$

Illustration et notations



Chaîne de Markov

Matrice des taux de transitions

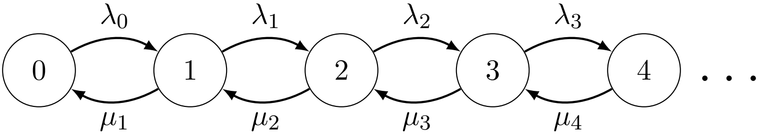
$$Q = \begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_1 & \lambda_2 & 0 \\ 0 & -\lambda_2 & 0 & \lambda_2 \\ 0 & 0 & -\lambda_1 & \lambda_1 \\ \lambda_3 & 0 & 0 & -\lambda_3 \end{pmatrix}$$

$$avec \ \sum_i Q_{ij} = 0$$

Processus de naissance et de mort

- Ne peut passer de l'état n que dans l'état :
  - $n+1$ (une arrivée)
  - $n-1$ (un départ)
- La probabilité d'une arrivée entre t et $t+dt$ est $\lambda(n)dt$
- La probabilité d'un départ entre t et $t+dt$ est $\mu(n)dt$
- La probabilité qu'il n'y ait ni arrivée ni départ entre t et $t+dt$ est $1-(\lambda(n)+\mu(n))dt$

Processus de naissance et de mort



Processus de naissance et de mort

$$
\begin{aligned}
P(n,t) &= P(X(t)=n) \\
P(n,t+dt) &= [1-(\lambda(n)dt+\mu(n)dt)]P(n,t) \\
&\quad +\lambda(n-1)dtP(n-1,t)+\mu(n+1)dtP(n+1,t) \\
\frac{dP(n,t))}{dt} &= -(\lambda(n)+\mu(n))P(n,t) \\
&\quad +\lambda(n-1)P(n-1,t)+\mu(n+1)P(n+1,t)
\end{aligned}
$$

et pour $n=0$, $\frac{dP(0,t))}{dt} = -\lambda(0)P(0,t)+\mu(1)P(1,t)$

Loi stationnaire

- Il faut poser $\frac{dP(n,t)}{dt} = 0$

$$\begin{aligned} \lambda(0)P(0) &= \mu(1)P(1) \\ (\lambda(n)+\mu(n))P(n) &= \lambda(n-1)P(n-1)+\mu(n+1)P(n+1) \end{aligned}$$

$$\Rightarrow P(n) = \frac{\lambda(n-1)\lambda(n-2)\ldots\lambda(0)}{\mu(n)\mu(n-1)\ldots\mu(1)}P(0)$$

- Pour éliminer $P(0)$, on effectue $\sum_{n=0}^{\infty} P(n) = 1$

Loi stationnaire

- On trouve $1/P(0) = 1 + \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} \frac{\lambda(j)}{\mu(j+1)}$ et donc

$$P(n) = \frac{\prod_{j=0}^{n-1} \frac{\lambda(j)}{\mu(j+1)}}{1 + \sum_{k=1}^{\infty} \prod_{j=0}^{k-1} \frac{\lambda(j)}{\mu(j+1)}}$$

- L'existence de la loi de probabilité invariante $(P(n))$ est assurée si $P(0)$ est non nulle.
- Pour cela, il la condition d'érgocité est satisfaite si : $\exists k_0$ tel que $\forall k > k_0, \frac{\lambda(k)}{\mu(k+1)} < 1$

Les coupes

Theorem

Si on fait des coupes dans l'ensemble des états en deux régions R_1 et R_2, alors à l'état stationnaire, il y a conservation des flux :

$$\sum_{i \in R_1, j \in R_2} \lambda_{ij} P(i) = \sum_{i \in R_2, j \in R_1} \lambda_{ij} P(i)$$

Processus de Poisson

- Le processus de Poisson un cas particulier de processus de naissance et de mort. C'est un processus de naissance pur pour lequel l'espace d'états est $\mathbb{N}$ et les seules transitions possibles sont de n à $n+1$.

$$\begin{cases} \lambda_n = & \lambda \\ \mu_n = & 0 \end{cases}$$

- Attention : tous les états sont transitoires donc pas de régime permanent !

Processus de Poisson

- Il faut revenir aux équations

$$\begin{aligned} P(n,t+dt) &= (1-\lambda dt)P(n,t) + \lambda dt P(n-1,t) \\ \frac{dP(n,t))}{dt} &= -\lambda P(n,t) + \lambda P(n-1,t) \\ \frac{dP(0,t))}{dt} &= -\lambda P(0,t) \end{aligned}$$

- La dernière équation est facile à résoudre : $P(0,t) = \exp(-\lambda t)$
- Par récurrence, on montre que $P(n,t) = \frac{(\lambda t)^n}{n!}\exp(-\lambda t)$.

Processus de Poisson

- La loi de **Poisson** et la loi **exponentielle décrivent le même phénomène** mais sous deux angles différents.
- Posons N_{t_1,t_2} = nombres d'arrivées pendant t_1 et t_2. $P(N_{0,t} = k) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t)$ et donc $E[N_{0,t}] = \lambda$ et $Var(N_{0,t}) = \lambda$.
- Les événements $\{T_1 \geq t\}$ et $\{N_{0,t} = 0\}$ sont **identiques** donc $P(T_1 \leq t) = 1 - P(N_{0,t} = 0) = 1 - \exp(-\lambda t)$ pour $t \geq 0$.
- Si le processus de comptage est décrit par une loi de Poisson alors le temps des inter-arrivées T est décrit par une loi exponentielle, $f_T(t) = \lambda \exp(-\lambda t)$, $E[T] = \frac{1}{\lambda}$ et $Var(T) = \frac{1}{\lambda^2}$.

Processus de Poisson

- Propriétés du processus de Poisson :

$$
\begin{aligned}
P(N_{0,t} = 1) &= \lambda t + o(t) \\
P(N_{0,t} > 1) &= o(t) \\
\lim_{t \to 0} \frac{o(t)}{t} &= 0
\end{aligned}
$$

Processus de Poisson

- Autre façon de voir : considérons le nombre d'appels reçus $S_n = \sum_{j=1}^{n} X_j$ où $P(X_j = 1) = p$ est la probabilité qu'un appel soit reçu par l'individu j.
- $P(S_n = k) = C_n^k p^k (1-p)^{1-k}$ est une loi binomiale avec $E[S_n] = np$ et $Var(S_n) = np(1-p)$.
- Posons $\lambda_n = np$ avec $p = f(n)$ et supposons que $\lim_{n\to\infty} \lambda_n = \lambda$.
- Etant donné que $\lim_{n\to\infty}(1-\lambda/n)^n = e^{-\lambda}$, on montre que $\lim_{n\to\infty} P(S_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Processus de Poisson

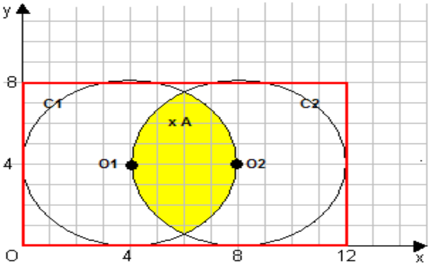
- Le processus de Poisson rend compte de nombreux phénomènes physiques et naturels, lorsque les arrivées sont produites par un **grand nombre de sources indépendantes** (entrée d'un calculateur, central téléphonique, ...)
- La somme de deux processus de Poisson de paramètres λ_1 et λ_2 est encore un processus de Poisson de paramètre $\lambda_1 + \lambda_2$

Processus de Poisson

- Deux théorèmes aux limites :
  - **Théorème d'échantillonnage** : la distribution limite obtenue par échantillonnage avec la probabilité $p \to 0$ parmi des arrivées de taux $\lambda \to \infty$ est une loi de Poisson si $\lambda p \to constante$
  - **Théorème de superposition** : la distribution limite obtenue par superposition de N arrivées indépendantes de même loi, de distribution quelconque lorsque $N \to \infty$, $\lambda_i \to 0$ et $\sum_i^N \lambda_i \to constante$ est une loi de Poisson.

Principes de simulation

- Que vaut l'aire de la surface jaune ?



Principes de simulation

- La simulation est basée sur l'utilisation de nombres pseudo-aléatoires pour la fabrication d'échantillons artificiels
- On tire au hasard N points dans le rectangle rouge. Si le point est dans les 2 cercles, on incrémente une variable. Après 5000 tirages, la proportion de points dans l'intersection est d'environ 1/5 $\Rightarrow$ l'aire jaune est environ 1/5 de l'aire rouge.

Principes de simulation

Cet exemple illustre le principe d'une simulation :

- On ne fait pas une expérience réelle (cela prendrait beaucoup de temps), mais on utilise l'ordinateur pour constituer un échantillon aléatoire.
- Si on veut un résultat significatif, il faut un nombre de tirages assez élevé. Normalement, on s'attend à ce que le résultat converge vers la "bonne valeur". On peut arrêter le tirage lorsque la différence entre deux valeurs successives est inférieure à un seuil (10^{-3} par ex).

Principes de simulation

Deux types de simulations :

- **Simulation stationnaire** : processus sous-jacent est stationnaire (à partir d'un certain temps), l'espérance mathématique est estimée par une moyenne temporelle. Une seule simulation (longue) suffit.
- **Simulation de Monte Carlo** : si le taux de panne d'une machine évolue en fonction du temps (vieillissement). Il faut fournir une estimation du taux de panne par tranche d'ge. Plusieurs simulations (courtes) sont nécessaires.

Intervalle de confiance pour simulation de Monte-Carlo

- Soit X un résultat de simulation et $(X_1, \ldots, X_n)$ l'échantillon obtenue en effectuant n réplications. Généralement, les calculs d'intervalles de confiance sont réalisés sur des moyennes, l'estimateur utilisé étant la moyenne arithmétique

$$\bar{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- En supposant que les X_i ont des distributions identiques, indépendantes, et **gaussiennes**, on calcule :

$$\bar{X}(n) - t_{n-1,\alpha/2}\frac{s(n)}{\sqrt{n}} \leq E[X] \leq \bar{X}(n) + t_{n-1,1-\alpha/2}\frac{s(n)}{\sqrt{n}}$$

avec $s^2(n) = \frac{\sum_{i=1}^{n}(X_i - \bar{X}(n))^2}{n-1}$ (estimateur sans biais de la

Intervalle de confiance pour simulation de Monte-Carlo

- $t_{n-1,\alpha/2}$ et $t_{n-1,1-\alpha/2}$ sont les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une **loi de Student** à $n-1$ degré de liberté
- Par parité, on a $t_{n-1,\alpha/2} = -t_{n-1,1-\alpha/2} \leq 0$.
- Le rayon de l'intervalle de confiance , au niveau de confiance $1-\alpha$, est donc : $R = t_{n-1,1-\alpha/2}\frac{s(n)}{\sqrt{n}}$.
- En général, on veut une précision $p = R/\bar{X}$ souhaitée, le nombre n de réplications prévisible si $s^2(n) \simeq s^2$ et $\bar{X}(n) \simeq \bar{X}$ doit respecter l'inégalité :

$$\frac{n}{(t_{n-1,1-\alpha/2})^2} > \frac{s^2}{(p\bar{X})^2}$$

Intervalle de confiance pour simulation de Monte-Carlo

- Les X_i ne sont pas nécessairement gaussiens, toutefois le calcul précédent est robuste à cette hypothèse. La robustesse dépend en particulier de son coefficient d'assymétrie.
- En vertu du théorème central limite, $\frac{\bar{X}-E[X]}{s(n)/\sqrt{n}}$ converge vers une loi normale lorsque $n \to \infty$.
- En pratique, l'approximation s'applique dès que $n \geq 30$, sinon il faut supposer la normalité des X_i, hypothèse impossible à tester avec $n \geq 30, \cdots$

Est-ce une loi de Poisson ?

Dans central téléphonique, on étudie X, le nombre d'appels reçus par minute. Pour cela, on fait 100 expériences d'une minute. On obtient les résultats suivants :

nombre d'appels x_i	0	1	2	3	4	5
fréquence n_i	21	38	22	10	4	4

- Il faut calculer la moyenne et la variance

$$
\begin{aligned}
m &= \frac{\sum n_i x_i}{\sum n_i} = 1,52, \\
s^2 &= \frac{1}{n-1}\left(\sum n_i x_i^2 - \frac{1}{n}(\sum n_i x_i)^2\right) = 1,65
\end{aligned}
$$

- On remarque d'emblée que $m \simeq s^2$. Peut-on admettre que X suit une loi de Poisson ?

Est-ce une loi de Poisson ?

Pour admettre que X suit une loi de Poisson de paramètre m, il faut employer le test du χ^2. On sait que si $X \sim \mathcal{P}(\lambda)$ alors $E[X] = Var(X) = \lambda$. Donc λ est estimé par m.

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

nombre d'appels x_i	0	1	2	3	4	5
n_i observés	21	38	22	10	4	4
n_i théorique	21,9	33,3	25,3	12,8	4,9	1,5

Les deux dernières classes de valeurs comportent moins de 5 observations (i.e., $n_{th} > 5$) $\Rightarrow$ on les regroupe (4,9+1,5=6,4) avant d'effectuer le test.

Est-ce un loi de Poisson ?

$$\chi^2 = \sum_{i=0}^{5} \frac{(n_i^{obs} - n_i^{th})^2}{n_i^{th}} = \frac{(21 - 21,9)^2}{21,9} + \frac{(38 - 33,3)^2}{33,3} + \dots$$

- On trouve $\chi^2 = 2,798$
- Le nombre ν de degrés de liberté dépend du nombre k de modalités retenues et du nombre r de paramètres qu'il a fallu estimé, on a $\nu = k - r - 1$
- ici, $\nu = 5 - 1 - 1 = 3$ car on a estimé λ uniquement.
- $\chi^2_{5\%} = 7,81$ et $\chi^2_{1\%} = 11,34$. On peut donc admettre que X est distribuée selon un loi de Poisson de paramètre 1,52.

Quel intervalle de confiance pour λ ?

- m est la moyenne de $n = 100$ échantillons, donc de grands echantillons. D'après la loi des grands nombres, m est approximée par une normale de paramètres :

$$
\begin{aligned}
m &= 1,52 \\
\sigma^2 &= \frac{s^2}{n} = 1,65 \cdot 10^{-2}
\end{aligned}
$$

- L'intervalle de confiance, au seuil de 5%, a pour bornes $m \pm 1,96 \cdot \sqrt{1,65} \cdot 10^{-1} = 1,52 \pm 0,252$.
- Il faut multiplier n par 4 pour diviser par 2 la taille de l'intervalle de confiance

Est-ce une loi exponentielle ?

- Il s'agit d'une distribution continue, le test du χ^2 ne s'applique pas ici. On appliquera le test non-parametrique d'ajustement de Kolmogorov basé sur $D_n = \sup |F_n^\star(x) - F(x)|$ où $F_n^\star$ est la distribution empirique.
- Exemple : test du caractère exponentiel d'un loi de service (ou de durée de vie). On dispose de $n = 5$ durées de service en minutes :
  $x_1 = 133, x_2 = 169, x_3 = 8, x_4 = 122, x_5 = 58$
  Le paramètre λ est estimé par $\lambda = \sum_i x_i/n = 98$, la fonction de répartition estimée est $F(x) = 1 - e^{-x/98}$, d'où le tableau :

x_i	8	58	122	133	169
$F(x_i)$	0.079	0.447	0.711	0.743	0.821

Est-ce une loi exponentielle ?

- On montre que

$$
\begin{aligned}
D_n &= \sup_x |F_n^\star(x) - F(x)| \\
&= \sup_i \{|F(x_i) - \frac{i}{n}|, |F(x_i) - \frac{i-1}{n}|\} \\
&= 0.311
\end{aligned}
$$

- D_n est à comparer à la valeur critique d_n lue dans la table du test de Kolmogorov, au seuil de $\alpha = 0.05$: $d_5 = 0.563$. L'hypothèse H_0 est donc retenue.

Confection d'échantillons

Imaginons un guichet devant lequel arrive des clients. On mesure tous les 15 minutes le nombre de clients qui arrivent et on obtient après 100 mesures les résultats statistiques suivants :

arrivées	0	1	2	3	4	5	6	7	8	9	10
fréq.	0,00	0,01	0,12	0,23	0,22	0,11	0,08	0,08	0,07	0,05	0,04
nombres	-	00	01-12	13-35	36-57	58- 68	69-76	77-84	85-91	92-95	96-99

- A partir d'une suite de nombres pseudo aléatoires $90\ 27\ 14\ 39\ldots$, on lit le nombre d'arrivées à simuler grce à la dernière ligne du tableau.
- Pour le 1er quart d'heure, on tire 90, cela signifie que 8 clients arrivent pendant ce quart d'heure. Pour le second quart d'heure, on tire 27 soit 3 arrivées etc.

Simulation suivant une loi de probabilité

- Il arrive que l'on connaisse les lois de probabilité. Par exemple, supposons que la loi des arrivées soit une loi de Poisson de paramètre $\lambda = 4,77$
- Il faudra donc utiliser un programme générateur de nombres aléatoires suivant la distribution de Poisson.
- Pour trouver un nombre aléatoire suivant la loi de Poisson, on détermine un nombre entier n tel que

$$\sum_{k=0}^{n} \frac{\lambda^k}{k!} e^{-\lambda} \leq z < \sum_{k=0}^{n+1} \frac{\lambda^k}{k!} e^{-\lambda}$$

où z est un nombre aléatoire compris entre 0 et 1 suivant la loi uniforme.

Comparaison de deux simulations

- Il arrive que l'on modifie un paramètre de la simulation, ou que l'on compare deux simulations. On suppose qu'on cherche à maximiser une **fonction score** qui reflête la justesse de la simulation, et qu'on teste les programmes sur de multiples jeux de tests.
- Y a-t-il une différence siginificative entre les deux au vu des scores ?
- **Test de rang de Wilcoxon** : test non paramétrique réputé plus robuste qu'un test de Student. Hypothèse nulle H_0 : les différences observées entre les scores ne sont pas statistiquement significatives et peuvent être attribuées au hasard.

Test de rang de Wilcoxon

- On pose d_i la différence absolue de score sur le jeu i

$$R^+ = \sum_{d>0} rank(d_i) + \frac{1}{2}\sum_{d=0} rank(d_i),$$
$$R^- = \sum_{d<0} rank(d_i) + \frac{1}{2}\sum_{d=0} rank(d_i)$$

- On pose $T = \min\{R^-, R^+\}$. Pour $N > 25$,

$$z = \frac{T - \frac{1}{4}N(N+1)}{\sqrt{\frac{1}{24}N(N+1)(2N+1)}}$$

suis approximativement une loi normale.

Illustration

	Simu1	Simu2	difference	rank
adult	0.763	0.768	+0.005	3.5
breast cancer	0.599	0.591	-0.008	7
breast cancer wisconsin	0.954	0.971	+0.017	9
cmc	0.628	0.661	+0.033	12
ionosphere	0.882	0.888	+0.006	5
iris	0.936	0.931	-0.005	3.5
liver disorders	0.661	0.668	+0.007	6
lung cancer	0.583	0.583	0.000	1.5
lymphography	0.775	0.838	+0.063	14
mushrooms	1.000	1.000	0.000	1.5
primary tumor	0.940	0.962	+0.022	11
rhum	0.619	0.666	+0.047	13
voting	0.972	0.981	+0.009	8
wine	0.957	0.978	+0.021	10

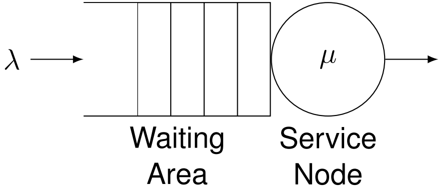
$R^+ = 3.5 + 9 + 12 + 5 + 6 + 14 + 11 + 13 + 8 + 10 + 1.5 = 93$, $R^- = 7 + 3.5 + 1.5 = 12$. Au seuil $\alpha = 0.05$ et $N = 14$ la différence est significative car $12 < 21$ (21 est la valeur critique exacte lue dans une table).

Comparaisons entre de multiples simulations

- Il arrive que l'on modifie plusieurs paramètre de la simulation, ou que l'on compare plusieurs simulations. On suppose qu'on dispose d'une fonction score qui reflête la justesse de la simulation, et qu'on teste les simulateurs sur de multiples jeux de tests.
- Le **problème des tests multiple** est connu : il faut contrôler la *family-wise error* $\rightleftharpoons$ la probabilité de faire *au moins* une erreur de type 1 dans l'ensemble des test réalisés.
- Y a-t-il une différence siginificative entre les simulations au vu des scores ?
- **Test de Friedman**: test non paramétrique réputé plus robuste qu'une ANOVA. **Hypothèse nulle** H_0 : les différences observées entre les scores ne sont pas statistiquement significatives et peuvent être attribuées au

File d'attente

- Les phénomènes d'attente sont des phénomènes courants : attente à un guichet, formation d'embouteillages ou de bouchons, etc... Pour maîtriser ces phénomènes on est conduit à les modéliser.
- On appellera système d'attente un ensemble composé d'une file d'attente et d'une ou plusieurs stations.



Exemples

Médecin

Des patients arrivent chez le médecin. Ils attendent leur tour dans la salle d'attente. Le médecin passe un temps aléatoire avec chaque patient.

Réseau

Le problème de transmission de données se modélise par des files d'attente. Chaque paquet de longueur variable attend dans un buffeur avant d'être transmis sur la ligne de communication.

Le modèle

- On va essayer de lever les contraintes de l'analyse operationnelle avec les files d'attente.
- On peut representer un systeme par un ensemble de files d'attente, chaque file modelisant une ressource par exemple.
- Une file d'attente est definit par :
  - La suite des instants d'arrivees des clients
  - La suite des temps de service des clients
  - La disciple de service qui donne l'ordre dans lequel seront servi les clients.
  - La capacite de la file
  - Le nombre de serveurs
  - Population totale de clients (rare)

Quantités étudiées

- Les quantités étudiées sont des moyennes en probabilité.
  - Le débit $\mathbf{\Lambda}$: nombre moyen de clients traversant le système.
  - Le temps de réponse moyen **R** d'un client (temps d'attente **W** + temps de service **S**)
  - La longueur moyenne **L** de la file
- On définit l'**intensité du trafic** par :

$$\rho = \frac{\text{temps moyen de service}}{\text{temps moyen entre 2 arrivées}} = \lambda S$$

- On définit aussi le **taux d'occupation du serveur** : $U = \lambda' S$ où λ' est le taux d'arrivée effectif des clients au serveur.
- U=ρ pour les files à un seul serveur

Formule de Little

- Elle est valable pour toute arrivée, toute loi de priorité, tout temps de service à condition qu'il existe un régime stationnaire limite :

$$L = \Lambda R$$

- On peut l'appliquer séparément à l'ensemble des clients en attente et à ceux en train d'être servis. On obtient

$$\begin{aligned} L_W &= W\Lambda \\ L_S &= S\Lambda \end{aligned}$$

- $L_S = 1 \cdot U + 0 \cdot (1 - U)$ car U est la probabilités que le serveur soit occupé. D'où $L_S = U$
- On retrouve pour les files à un seul serveur $U = \Lambda S$

Notations de Kendall

- **Une file d'attente se note :**

A/S/C (DS/K/L)
A/S/C/K/L/ (DS)
A/S/C/K/L/DS

- **Avec :**

A : processus d'arrivée
S : processus de sortie
C : nombre de serveurs
K : capacité maximale de la file
L : population de clients
DS : discipline de service

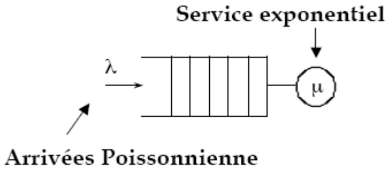
- **Symbole pour les arrivées et les services**
  - M : loi exponentielle (Markovienne)
  - D : loi constante
  - E_k : loi Erlang-k
  - H_k : loi hyper-exponentielle ordre k
  - GI : loi générale indépendante
  - G : loi générale

- **Symbole pour les discipline de service**
  - FCFS : First Come First Serve (Preempt)
  - LCFS : Last Come First Serve (Preempt)
  - QUANTUM : Round Robin
  - PS : Processor Sharing
  - RANDOM
  - PRIORITY

M/M/1/∞/∞

- **Résultats**

$$\begin{cases} \lambda_k = \lambda \\ \mu_k = \mu \end{cases} \qquad \rho = \frac{\lambda}{\mu}$$



$$P_k = \rho^k P_0$$
$$P_0 = 1 - \rho$$
$$U = 1 - P_0 = \rho$$
$$P[n \geq N] = \rho^N$$

$$N = L = \sum_{k=0}^{\infty} k P_k = \frac{\rho}{1-\rho}$$
$$R = \frac{1}{\mu(1-\rho)}$$
$$S = \frac{1}{\mu}$$

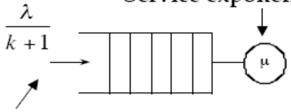
Arrivées découragées

- On peut introduire un facteur d'impatience dans ce sytème, c'est-à-dire qu'un client arrivant dans la FA a une certaine probabilité de quitter le système immédiatement, sans se faire servir, en fonction du nombre de clients présents devant lui.
- Prenons l'exemple où cette probabilité d'impatience vaut $\frac{n}{n+1}$ si n clients sont présents dans le système.

Arrivées découragées

- **Résultats**

$$\begin{cases} \lambda_k = \dfrac{\lambda}{k+1} \\ \mu_k = \mu \end{cases} \qquad \rho = \frac{\lambda}{\mu}$$



$$P_k = \frac{\rho^k}{k!} P_0$$

$$P_0 = \frac{1}{e^\rho}$$

$$U = 1 - P_0$$

$$P[n > m] = 1 - P_0 \sum_0^{N-1} \frac{\rho^k}{k!}$$

$$N = \rho$$

$$R = \frac{\rho}{\mu(1 - P_0)}$$

M/M/∞/∞/∞

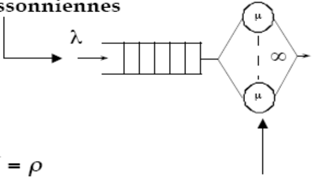
- La population est infinie est aucun client n'est rejeté car nous avons une infinité de serveurs !
- Un client qui arrive est immédiatement servi. Donc pas de file d'attente avant service.

M/M/∞/∞/∞

- **Résultats**

$$\begin{cases} \lambda_k = \lambda \\ \mu_k = k\mu \end{cases} \qquad \rho = \frac{\lambda}{\mu}$$

Arrivées Poissonniennes



Services exponentiels

$$P_k = \frac{\rho^k}{k!} P_0$$

$$P_0 = \frac{1}{e^\rho}$$

$$U = 1 - P_0$$

$$P[n > m] = 1 - P_0 \sum_{0}^{N-1} \frac{\rho^k}{k!}$$

$$N = \rho$$

$$R = \frac{1}{\mu}$$

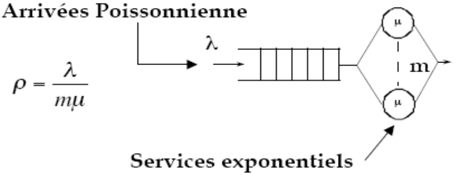
✎ **Même résultat que pour les arrivées découragées ! Sauf que l'expression de R est plus simple.**

M/M/m/m/∞

- La population est infinie, il y a m serveurs et aucune place disponible dans la FA c'est-à-dire qu'un client arrivant, alors que la file est pleine, doit quitter le système.
- La loi de cette perte est très importante en téléphonie. Elle est connue sous le nom de **formule de perte d'Erlang**.

M/M/m/m/∞

- **Résultats**



$$\begin{cases} \lambda_k = \lambda \\ \mu_k = \begin{cases} k\mu & 1 \le k \le m \\ m\mu & k \ge m \end{cases} \end{cases} \qquad \rho = \frac{\lambda}{m\mu}$$

$$P_k = \begin{cases} \dfrac{1}{k!}(m\rho)^k P_0 & k \le m \\ \dfrac{1}{m!} m^m \rho^k P_0 & k \ge m \end{cases} \qquad P_0 = \frac{1}{\displaystyle\sum_0^{m-1} \frac{1}{k!}(m\rho)^k + \frac{(m\rho)^m}{m!(1-\rho)}}$$

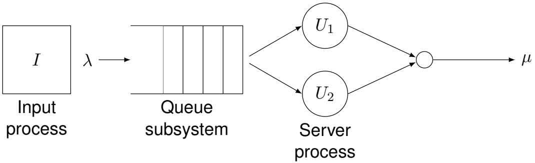
$$U = 1 - P_0$$

$$P[n > m] = \frac{1}{P_0} \frac{(m\rho)^m}{m!(1-\rho)}$$

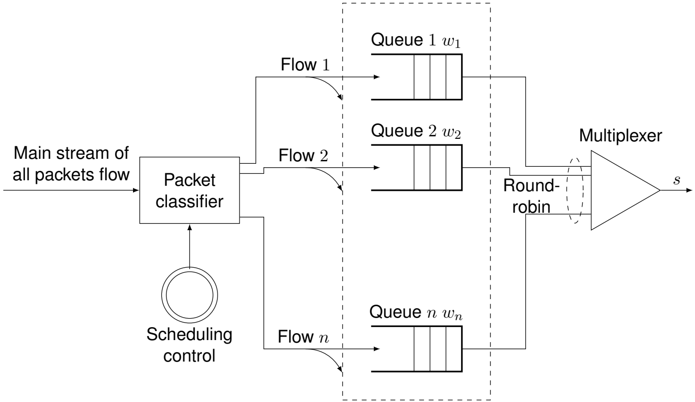
M/M/m/∞/∞

- La population est infinie. Il y a m serveurs et aucun client n'est rejeté car la capacité de stockage des clients en attente est infinie.
- La probabilité de mise en attente (avant service) constitue la **deuxième formule d'Erlang**.

M/M/2/∞/∞


I
λ
U_1
U_2
μ
Input process
Queue subsystem
Server process

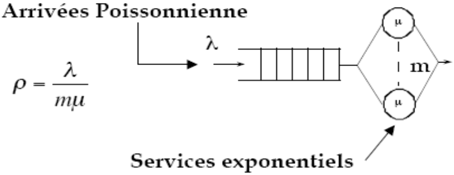
M/M/m/∞/∞


Queue 1 w_1
Flow 1
Queue 2 w_2
Multiplexer
Main stream of all packets flow
Packet classifier
Flow 2
Round-robin
s
Queue n w_n
Scheduling control
Flow n

M/M/m/∞/∞

- **Résultats**

$$\begin{cases} \lambda_k = \lambda \\ \mu_k = \begin{cases} k\mu & 1 \le k \le m \\ m\mu & k \ge m \end{cases} \end{cases} \qquad \rho = \frac{\lambda}{m\mu}$$



$$P_k = \begin{cases} \frac{1}{k!}(m\rho)^k P_0 & k \le m \\ \frac{1}{m!} m^m \rho^k P_0 & k \ge m \end{cases} \qquad P_0 = \frac{1}{\sum_{0}^{m-1} \frac{1}{k!}(m\rho)^k + \frac{(m\rho)^m}{m!(1-\rho)}}$$

$$U = 1 - P_0$$

$$P[n > m] = \frac{1}{P_0} \frac{(m\rho)^m}{m!(1-\rho)}$$

M/M/1/K/∞

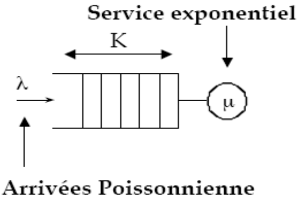
- La population est infinie. Un client trouvant toutes les K places disponibles occupées est rejeté ou “perdu” et rejoint à nouveau l'ensemble des clients potentiels.
- C'est le cas réaliste d'un buffeur de taille limité.

M/M/1/K/∞

- **Résultats**

$$\begin{cases} \lambda_k = \begin{cases} \lambda & k < K \\ 0 & k \geq K \end{cases} \\ \mu_k = \mu \quad 1 \leq k \leq K \end{cases}$$

$$\rho = \frac{\lambda}{m\mu}$$



$$P_k = \rho^k P_0$$

$$P_0 = \frac{1-\rho}{(1-\rho^{K+1})}$$

$$U = 1 - P_0$$

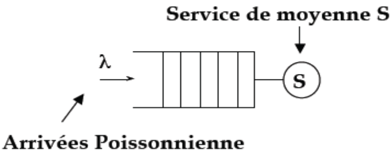
$$P_K = \text{Probabilité de rejet}$$

$$N = \frac{\rho}{1-\rho} - (K+1)\frac{\rho^{K+1}}{1-\rho^{K+1}}$$

$$R = \frac{N}{\lambda(1-P_K)}$$

M/G/1/∞/∞

- **Résultats**



Soit h(t) le pdf du temps de service,

$$b_n = \int_0^{\infty} t^n h(t)dt$$

On démontre alors que $\begin{cases} S = b_1 \\ Var[S] = b_2 - b_1^2 \\ \rho = \lambda S \end{cases}$

En utilisant $c^2 = CCV = \frac{Var[S]}{S^2} = \lambda^2 \frac{b_2}{\rho^2} - 1$ $\Rightarrow$

$$\boxed{\begin{aligned} N &= \rho\left(1 + \frac{\rho}{1-\rho}\left(\frac{1+c^2}{2}\right)\right) \\ \frac{R}{S} &= 1 + \frac{\rho}{1-\rho}\left(\frac{1+c^2}{2}\right) \\ \frac{W}{S} &= \frac{\rho}{1-\rho}\left(\frac{1+c^2}{2}\right) \end{aligned}}$$

M/M/1/∞/N

- A faire ...
- La population est ici finie (source finie). Une capacité de stockage de N aurait suffit !
- Pour garder le caractère markovien des arrivées, on suppose que chaque client qui n'est pas dans le système arrive selon une loi exponentielle de paramètre λ, indépendamment des autres.

M/M/∞/∞/N

A faire . . .

References I

Christopher M. Bishop.
Pattern Recognition and Machine Learning.
Springer, 2006.

Gianluca Bontempi and Souhaib Ben Taieb.
Statistical foundations of machine learning
OTexts: Melbourne, Australia, 2016.

Maxime Gasse.
Probabilistic Graphical Model Structure Learning: Application to Multi-Label Classification.
Ph.D. thesis, Université de Lyon, 2017.

References II

- Lawrence R. Rabiner.
  A tutorial on Hidden Markov Models and selected applications in speech recognition, Proceedings of the IEEE, vol. 77, no 2, 1989.
- Guillaume Obozinski - Francis Bach
  Introduction to Graphical Models, Master "Mathématiques Appliquées", Parcours "Mathematiques, Vision et Apprentissage", ENS Cachan 2017.