MDI230
Files d'attentes

I. Introduction

Les files d'attente permettent de modéliser des phénomènes courants :

- Queue à la caisse, à la Poste, à l'aéroport…
- Lignes de production, chaines de montage, réparation…
- Services en ligne avec accès limités, standards, jeux….
- Bufferisation des routeurs Internet
- Appels voix ou données dans un réseau mobile
- …

Vocabulaire :

- Serveurs / service
- Clients

Dans l'étude des files d'attente, on cherche à mesurer l'occupation des serveurs. On considère N files identiques, et on souhaite connaitre l'occupation moyenne sur les N files à l'instant T. Par ergodicité, cette moyenne est égale à la moyenne d'occupation sur une durée T pour un seul serveur.

On regarde donc ce qu'il se passe pour un seul serveur. On définit la charge comme étant :

charge = nombre moyen de clients arrivant par unité de temps x temps de service moyen

On la note généralement ρ. C'est une quantité sans dimension dont l'unité est l'Erlang. Par exemple, si on a une charge de 3 Erlangs, cela signifie qu'en moyenne il y a un trafic de 3 clients. Autre exemple, si il y a en moyenne 2,4 appels par heure, et que chaque appel dure en moyenne 3 minutes, alors la charge du réseau de téléphonie est de 2,4x3/60 = 0,12 Erlangs.

Mais la charge ne suffit pas à décrire un modèle. Par exemple les 2 modèles suivants ont la même charge :

- Un système déterministe avec exactement un appel de 30 secondes toutes les minutes. La charge est de 1x1/2 = 0,5 Erlangs. Le système est occupé exactement 50% du temps, il n'y a besoin que d'un seul serveur.
- Un système aléatoire avec en moyenne un appel toutes les minutes qui dure en moyenne 30 secondes. La charge est identique 0,5 Erlangs. Mais la probabilité de perte est différente, 2 appels peuvent être simultanés. Le nombre de serveurs à mettre est différent.

C'est la différence entre le dimensionnement en moyenne, et la prise en compte du trafic. On s'intéresse donc à des modèles aléatoires qui permettent de prendre en compte le trafic.

La notation de Kendall permet de nommer les différents modèles de files d'attente.
On écrit A/B/C/D/E :

- A : loi des temps d'inter-arrivées
  - exemples : M = Memoryless + Markov, D = Déterministe, GI = General Independent (intér-arrivées iid), G = General
- B : loi des temps de service
  - exemples : M, D, GI, G
- C : nombre de serveurs
- D : nombre de places dans le système (serveurs + file)
  - si non renseigné alors D = infini
- E : discipline de service
  - exemples : FIFO (file), LIFO (pile), SRPT (Shortest Remaining Processins Time), RR (Round Robin)
  - si non renseigné : E = FIFO

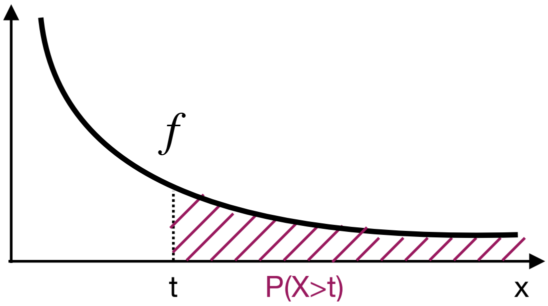
II. Processus de Poisson

1. Loi exponentielle

Définition : Loi exponentielle

Soit X une variable aléatoire réelle. X suit une loi exponentielle de paramètre μ>0 si et seulement si sa densité de probabilité est :

$$f(x) = \mu e^{-\mu x} \text{ pour } x \in \mathbb{R}^+.$$



On peut alors calculer :

$\mathrm{P}(X \leq t) = \int_0^t \mu e^{-\mu x} \mathrm{d}x = 1 - e^{-\mu t}$

$P(X > t) = e^{-\mu t}$

Espérance : $\mathrm{E}[X] = \int_{\mathbb{R}^+} x f(x) \mathrm{d}x = \frac{1}{\mu}$

Variance : $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_0^{+\infty} x^2 \mu e^{-\mu x} \mathrm{d}x - \frac{1}{\mu^2} = \frac{1}{\mu^2}$

Fonction génératrice des moments : $\mathbb{E}[e^{tX}] = \int_{\mathbb{R}^+} e^{tx} \mu e^{-\mu x} \mathrm{d}x = \frac{\mu}{\mu - t}$

Propriété : Soient X_1 et X_2 deux variables aléatoires réelles indépendantes avec $X_1 \sim$ exp(μ_1) et $X_2 \sim$ exp(μ_2). On pose $X = \min(X_1, X_2)$.

Alors X suit une loi exponentielle de paramètre $\mu_1 + \mu_2$, et $\mathrm{P}(X = X_i) = \frac{\mu_i}{\mu_1 + \mu_2}$.

Démonstration :

$\mathrm{P}(X > t) = \mathrm{P}(X_1 > t, X_2 > t) = \mathrm{P}(X_1 > t)\mathrm{P}(X_2 > t) = e^{-(\mu_1 + \mu_2)t}$.

$\mathrm{P}(X = X_1) = \mathrm{P}(X_2 > X_1) = \int \mathrm{P}(X_2 > t) \mathrm{d}\mathrm{P}_{X_1}(t) = \int_0^{+\infty} \mu_1 e^{-\mu_1 t} \mathrm{P}(X_2 > t) \mathrm{d}t = \int_0^{+\infty} \mu_1 e^{-\mu_1 t} e^{-\mu_2 t} \mathrm{d}t = \frac{\mu_1}{\mu_1 + \mu_2}$.

Propriété : La loi exponentielle est sans mémoire, et c'est la seule. Soit X une variable aléatoire réelle qui suit une loi exponentielle de paramètre $\mu > 0$.

Soient $t_1, t_2 > 0$, alors $\mathrm{P}(X > t_1 + t_2 \,|\, X > t_1) = \mathrm{P}(X > t_2)$.

Démonstration :

$$\mathrm{P}(X > t_1 + t_2 \,|\, X > t_1) = \frac{\mathrm{P}(X > t_1 + t_2, X > t_1)}{\mathrm{P}(X > t_1)} = \frac{e^{-\mu(t_1 + t_2)}}{e^{-\mu t_1}} = e^{-\mu t_2}.$$

2. Définitions d'un processus de Poisson

Définition : Loi de Poisson
Une variable aléatoire discrète N suit une loi de Poisson de paramètre $\lambda > 0$ si et seulement si

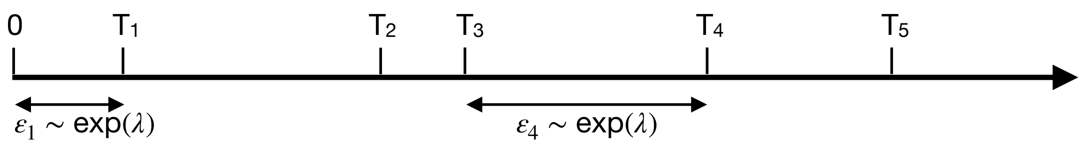
$$\mathrm{P}(N = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Espérance : $\mathrm{E}[N] = \sum_{k=1}^{+\infty} k\mathrm{P}(N = k) = e^{-\lambda}\lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$

Variance : $\mathrm{Var}[N] = \mathrm{E}[N^2] - \mathrm{E}[N]^2 = \sum_{k=1}^{+\infty} k^2\mathrm{P}(N = k) - \lambda^2 = e^{-\lambda}\lambda(\lambda+1)e^{\lambda} - \lambda^2 = \lambda$

Définition : Processus ponctuel sur $\mathbb{R}^+$
On appelle processus ponctuel une famille de variables aléatoires indexée par $\mathbb{N}$ et à valeurs dans $\mathbb{R}^+$. C'est une suite strictement croissante de variables aléatoires positives $(T_1, T_2, T_3, \ldots)$ telle que $(T_n) \underset{x\to+\infty}{\longrightarrow} +\infty$ presque sûrement. Par convention, $T_0 = 0$.

Définition 1 : Processus de Poisson
$N = (T_1, T_2, \ldots)$ est un processus de Poisson d'intensité $\lambda > 0$ si et seulement si les variables aléatoires $\varepsilon_n = T_n - T_{n-1}$ sont indépendantes et de même loi exponentielle de paramètre λ.



Soit $n \in \mathbb{N}$, on cherche à caractériser la loi de $T_n = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$.
Soit f quelconque,

$$\begin{aligned}\mathrm{E}[f(T_n)] &= \int_0^{+\infty} \ldots \int_0^{+\infty} f(x_1 + x_2 + \ldots + x_n)\mathrm{dP}_{\varepsilon_1}(x_1)\ldots\mathrm{dP}_{\varepsilon_n}(x_n) \\ &= \int_0^{+\infty} \ldots \int_0^{+\infty} f(x_1 + x_2 + \ldots + x_n)\lambda^n e^{-\lambda(x_1+\ldots+x_n)}\mathrm{d}x_1\ldots\mathrm{d}x_n\end{aligned}$$

On effectue le changement de variable $t_1 = x_1, t_2 = x_1 + x_2, \ldots, t_n = x_1 + \ldots + x_n$

$$\begin{aligned}\mathrm{E}[f(T_n)] &= \int_0^{+\infty}\int_0^{t_n}\int_0^{t_{n-1}} \ldots \int_0^{t_2} f(t_n)\lambda^n e^{-\lambda t_n}\mathrm{d}t_1\ldots\mathrm{d}t_n \\ &= \int_0^{+\infty} f(t_n)\lambda^n e^{-\lambda t_n}\left(\int_0^{t_n}\ldots\int_0^{t_2}\mathrm{d}t_1\ldots\mathrm{d}t_{n-1}\right)\mathrm{d}t_n \\ &= \int_0^{+\infty} f(t_n)\lambda^n e^{-\lambda t_n}\frac{t_n^{n-1}}{(n-1)!}\mathrm{d}t_n\end{aligned}$$

La densité de probabilité de T_n est donc $\lambda^n e^{-\lambda t}\frac{t^{n-1}}{(n-1)!}$, ainsi T_n suit une loi gamma de paramètres n et λ.

Définition 2 : Processus de Poisson
$N = (T_1, T_2, \ldots)$ est un processus de Poisson d'intensité $\lambda > 0$ si est seulement si :
- $N(t)$ le nombre de points T_i dans $[0,t]$ suit une loi de Poisson de paramètre λt
- Conditionnellement à $N(t) = k$, la suite $(T_1, T_2, \ldots, T_k)$ est uniformément distribuée sur $[0,t]$.

$N(t) \sim$ Poisson(λt), on obtient k $= 5$

$T_i \sim$ Unif$[0,t] \forall i \in \{1,\ldots,k\}$

Définition 3 : Processus de Poisson
$N = (T_1, T_2, \ldots)$ est un processus de Poisson d'intensité $\lambda > 0$ si est seulement si sa transformée de Laplace vaut :
$\mathscr{L}_N(f) = \mathrm{E}[e^{-\sum_{n\geq 1} f(T_n)}] =$ exp$(\int_0^{+\infty} (e^{-f(x)} - 1)\lambda \mathrm{d}x)$ pour toute fonction $f : \mathbb{R}^+ \to \mathbb{R}^+$.

3. Simulation d'un processus de Poisson

Il existe deux manières de simuler un processus de Poisson sur $\mathbb{R}^+$ selon les deux premières définitions données.
On cherche à simuler un processus de Poisson d'intensité λ sur $[0,t_{\max}]$.

1ère méthode :

```
Données : lambda, t_max
T <= [] %tableau de points
t <= 0 %curseur
n <= 0 %nombre de points
Tant que t ≤ t_max
      x <= réalisation d'une v.a. exp(lambda) %Ecart entre 2 points
      t <= t+x %calcul de Tn+1 à partir de Tn et de l'écart tiré
      n <= n+1
      T[n] <= t
Retourner n, T
```

2ème méthode :

```
Données : lambda, t_max
T <= [] %tableau de points
n <= 0 %nombre de points
n <= réalisation d'une v.a. Poisson (lambda*t_max)
Pour i=1:n
      T[i] <= réalisation d'une v.a. Unif[0,t_max]
Tri croissant sur T
Retourner n, T
```

4. Propriétés

Théorème : La superposition de deux processus de Poisson indépendants d'intensités respectives λ_1 et λ_2 est un processus de Poisson d'intensité $\lambda_1 + \lambda_2$.

Définition : Amincissement
Soit N un processus ponctuel. A chaque point de N, on décide s'il appartient au processus N' avec une probabilité $p \in [0,1]$ indépendamment des autres points.

Théorème : Le processus résultant de l'amincissement d'un processus de Poisson d'intensité λ avec la probabilité $p \in [0,1]$ est un processus de Poisson d'intensité λp.
Ce processus est indépendant du processus aminci "restant" d'intensité $\lambda(1-p)$.

5. Paradoxe de l'autobus

On considère une ligne de bus dont les instants d'arrivées de bus à un arrêt donné sont les points d'un processus de Poisson $N = (T_1, T_2, \dots)$:

- A l'instant t, $N(t)$ bus sont déjà passés.
- T_n est l'instant d'arrivée du n-ième bus
- $\varepsilon_n = T_n - T_{n-1}$ est le temps d'attente entre le $(n-1)$-ième bus et le n-ième bus

On s'intéresse à $A(t) = T_{N(t)+1} - t$ le temps d'attente d'un usager qui arrive à l'arrêt de bus à l'instant t.
Soit $x \geq 0$ on a alors,

$$\begin{aligned}
\mathrm{P}(A(t) > x) &= \sum_{n=0}^{+\infty} \mathrm{P}(A(t) > x, N(t) = n) \\
&= \sum_{n=0}^{+\infty} \mathrm{P}(T_{n+1} - t > x, T_n < t \leq T_{n+1}) \\
&= \sum_{n=0}^{+\infty} \mathrm{P}(T_n + \varepsilon_{n+1} - t > x, T_n < t \leq T_n + \varepsilon_{n+1}) \\
&= \sum_{n=0}^{+\infty} \int_0^{+\infty} \int_0^{+\infty} 1_{u+v-t>x} 1_{u<t<u+v} \mathrm{dP}_{T_n}(u) \mathrm{dP}_{\varepsilon_{n+1}}(v) \\
&= \sum_{n=0}^{+\infty} \int_0^{+\infty} \int_0^{+\infty} 1_{u+v-t>x} 1_{u<t} \mathrm{dP}_{T_n}(u) \mathrm{dP}_{\varepsilon_{n+1}}(v) \\
&= \sum_{n=0}^{+\infty} \int_0^{t} \left(\int_{x+t-u}^{+\infty} \lambda e^{-\lambda v} \mathrm{d}v \right) \lambda^n e^{-\lambda u} \frac{u^{n-1}}{(n-1)!} \mathrm{d}u \\
&= \sum_{n=0}^{+\infty} \int_0^{t} e^{-\lambda(x+t-u)} \lambda^n e^{-\lambda u} \frac{u^{n-1}}{(n-1)!} \mathrm{d}u \\
&= \sum_{n=0}^{+\infty} \lambda^n e^{-\lambda(x+t)} \int_0^{t} \frac{u^{n-1}}{(n-1)!} \mathrm{d}u \\
&= e^{-\lambda(x+t)} \sum_{n=0}^{+\infty} \frac{\lambda^n t^n}{n!}
\end{aligned}$$

$$\mathrm{P}(A(t) > x) = e^{-\lambda x}$$

Théorème :

- $A(t)$ suit une loi exponentielle de paramètre λ (et ne dépend donc pas de t).
- Le temps d'attente et le temps écoulé depuis le dernier passage sont indépendants.
- Quelque soit le temps écoulé depuis le passage du dernier bus, le temps d'attente est identique et vaut en moyenne $1/\lambda$.

Analogue discret :
L'analogue discret de la loi exponentielle est la loi géométrique qui peut par exemple représenter le nombre de lancers pour atteindre le premier pile dans un jeu de “pile ou face”. Le nombre de lancers avant le prochain pile (= temps d'attente) est indépendant du nombre de lancers déjà effectués (=temps déjà écoulé).

III. Processus de Markov

1. Chaines de Markov

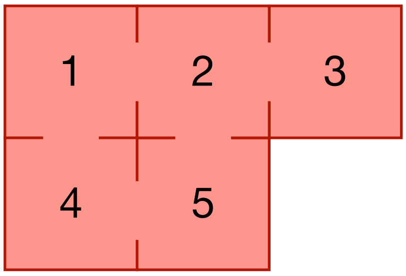
Définition : Chaine de Markov
Une suite $(X_n, n \in \mathbb{N})$ de variables aléatoires discrètes (à valeurs dans E dénombrable) est une chaine de Markov si :

$$\forall n \in \mathbb{N},\ \mathrm{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = \mathrm{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

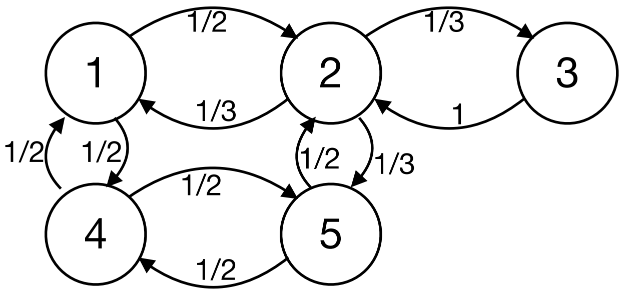
Une chaine de Markov est définie sur un espace discret ($\mathbb{N}$) et à valeurs discrètes (E).
Une chaine de Markov est dite sans mémoire, car seul son dernier état compte pour déterminer l'état suivant.

Propriété : On dit qu'une chaine de Markov est homogène si $\forall x, y \in E,\ \mathrm{P}(X_{n+1} = y | X_n = x)$ ne dépend pas de n, on peut alors écrire $p(x,y) = \mathrm{P}(X_{n+1} = y | X_n = x)$.

Exemple : On considère le labyrinthe suivant à 5 cases. A chaque étape, on change de case. Si une case a n portes, on choisit la porte pour aller à la case suivante avec la probabilité 1/n. Chaque déplacement ne dépend que de la situation courante.



On peut modéliser les déplacements dans le labyrinthe par une chaine de Markov. On peut la représenter par son diagramme de transition : un état est représenté par une bulle, une transition possible est représentée par une flèche avec sa probabilité de transition associée.



La somme des probabilités sur les transitions sortantes doit être égale à 1.
On peut aussi écrire la matrice de transition :

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

La somme sur les lignes est égale à 1.

Propriétés : Soit X une chaine de Markov,

- X est dite irréductible lorsque tous ses états conduisent les uns aux autres, i.e. le graphe de transition est fortement connexe.
- Un état de X est dit récurrent lorsque la probabilité que la chaine passe par cet état est égale à 1. La chaine de Markov y passera alors un nombre infini de fois. Sinon l'état est dit transient.
- La chaine X est dite récurrente si tous ses états le sont.

Si X est irréductible et a un état récurrent alors elle est récurrente.

Définition : Probabilité invariante

Une mesure ν sur l'espace d'états E discret est dite invariante par rapport à la matrice de transition P si et seulement si pour tout $y \in E,\ \nu = \nu P \Leftrightarrow \nu(y) = \sum_{x \in E} \nu(x) p(x, y)$.

Si de plus $\sum_{x \in E} \nu(x) = 1$ alors ν est une probabilité invariante.

Théorème : Soit X une chaine de Markov homogène, irréductible, et récurrente, alors il existe une unique probabilité invariante π telle que $\forall x \in E, 0 < \pi(x) < +\infty$.

On l'appelle aussi probabilité stationnaire.

$\pi(x)$ est la probabilité d'être dans l'état x, $\pi(x) = \mathrm{P}(X = x)$.

Exemple : Labyrinthe

On résout l'équation $\pi = \pi P$:

- $\pi(1) = \frac{1}{3}\pi(2) + \frac{1}{2}\pi(4)$
- $\pi(2) = \frac{1}{2}\pi(1) + \pi(3) + \frac{1}{2}\pi(5)$
- $\pi(3) = \frac{1}{3}\pi(2)$
- $\pi(4) = \frac{1}{2}\pi(1) + \frac{1}{2}\pi(5)$
- $\pi(5) = \frac{1}{3}\pi(2) + \frac{1}{2}\pi(4)$
- $\pi(1) + \pi(2) + \pi(3) + \pi(4) + \pi(5) = 1$

On obtient alors $\pi(1) = \pi(4) = \pi(5) = \frac{2}{10}$, $\pi(2) = \frac{3}{10}$, et $\pi(3) = \frac{1}{10}$.

2. Construction trajectorielle

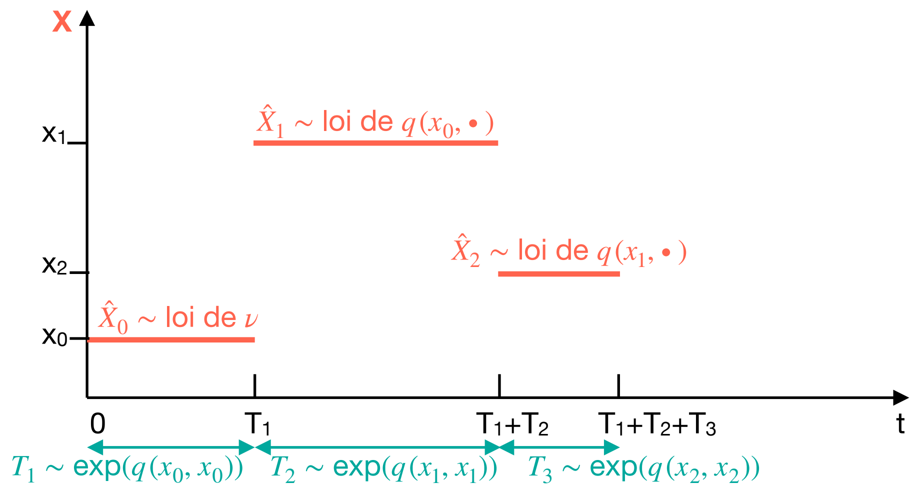
Soit $(X(t), t \in \mathbb{R}^+)$ défini sur $\mathbb{R}^+$ à valeurs dans l'espace d'états E dénombrable. Un processus est défini sur un espace continu ($\mathbb{R}^+$) et à valeurs discrètes (E).

Soit ν une mesure de probabilité sur E.

Soit $Q = \{q(x, y), \forall x, y \in E\}$ telle que $q(x, y) \geq 0\ \forall x, y \in E$ et $\sum_{y \neq x} q(x, y) = 1 \forall x \in E$.

- Soit $\hat{X}_0 = X(0)$ variable aléatoire de loi ν.
- Si $\hat{X}_0 = X(0) = x_0$ alors X reste dans l'état x_0 pendant un temps qui suit une loi exponentielle de paramètre $q(x_0, x_0)$. On note T_1 cette durée, donc $T_1 \sim \exp(q(x_0, x_0))$. Ainsi pour $0 \leq t < T_1$, on a $X(t) = x_0$.
- Soit $\hat{X}_1$ une variable aléatoire indépendante de $\hat{X}_0$ et T_1, dont la loi est la loi de transition depuis $x_0 : \mathrm{P}(\hat{X}_1 = y) = q(x_0, y),\ \forall y \in E$.

- Si $\hat{X}_1 = x_1$ alors X reste dans l'état x_1 pendant un temps qui suit une loi exponentielle de paramètre $q(x_1, x_1)$. On note T_2 cette durée, donc $T_2 \sim \exp(q(x_1, x_1))$. Ainsi pour $T_1 \leq t < T_1 + T_2$, on a $X(t) = x_1$.
- Soit $\hat{X}_2$ une variable aléatoire indépendante de $\hat{X}_0, \hat{X}_1, T_1$ et T_2, dont la loi est la loi de transition depuis x_1 : $\mathrm{P}(\hat{X}_2 = y) = q(x_1, y), \ \ \forall y \in E$.
- Si $\hat{X}_2 = x_2$ alors X reste dans l'état x_2 pendant un temps qui suit une loi exponentielle de paramètre $q(x_2, x_2)$. On note T_3 cette durée, donc $T_3 \sim \exp(q(x_2, x_2))$. Ainsi pour $T_1 + T_2 \leq t < T_1 + T_2 + T_3$, on a $X(t) = x_2$.
- …



On a ainsi construit :

- $\hat{X} = (\hat{X}_0, \hat{X}_1, \ldots)$ une chaine de Markov de matrice de transition $\hat{Q}$ telle que $\hat{Q}(x, y) = q(x, y)$ si $x \neq y$, et $\hat{Q}(x, x) = 0$ (une transition ne dépend que du dernier état).
- $X = (X(t), t \in \mathbb{R}^+)$ un processus de Markov de paramètres (ν, Q).

Un processus de Markov est un processus dont les trajectoires sont faites de sauts régulés par une chaine de Markov, et dont les temps de séjour dans chaque état suivent une loi exponentielle dépendant de l'état (de paramètre $q(x, x)$).

Définition : Processus de Markov

$(X(t), t \in \mathbb{R}^+)$ défini sur $\mathbb{R}^+$ à valeurs dans l'espace d'états E dénombrable est un processus de Markov si et seulement si :

$$\mathrm{P}(X(t_1 + t_2) = y \,|\, X(t_1) = x, \{X(t), t < t_1\}) = \mathrm{P}(X(t_1 + t_2) = y \,|\, X(t_1) = x)$$

pour tout $t_1, t_2 > 0$ et pour tout $x, y \in E$.

Seul le dernier état connu compte pour déterminer la suite, un processus de Markov est sans mémoire.

Propriété : On dit qu'un processus de Markov est homogène si $\forall x, y \in E, \forall t_1, t_2 > 0$, $\mathrm{P}(X(t_1 + t_2) = y \,|\, X(t_1) = x)$ ne dépend pas de t_1.

On peut alors écrire $p_{t_2}(x, y) = \mathrm{P}(X(t_2) = y \,|\, X(0) = x)$ la probabilité de passer de x à y en un temps t_2.

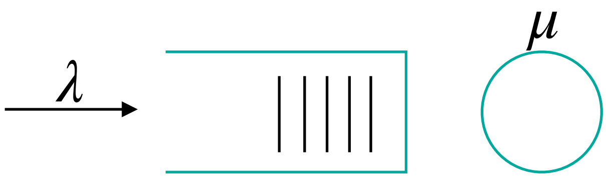
Exemple : le processus de Poisson est un processus de Markov homogène :

- défini sur $\mathbb{R}^+$ et à valeurs dans $\mathbb{N}$
- $Q = \begin{pmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \lambda & 1 & \\ & & & \lambda & 1 \\ 0 & & & & \ddots \end{pmatrix}$
- $\nu \,|\, \mathrm{P}(N(0) = 0) = 1$

3. M/M/1

On s'intéresse à la file M/M/1 :

- Processus d'arrivées : M = les inter-arrivées suivent une loi exponentielle = Processus de Poisson. On note λ son intensité.
- Temps de services : M = les temps de services suivent une loi exponentielle. On note μ son paramètre.
- 1 serveur
- Salle d'attente infinie
- Politique de service FIFO



On note $X = (X(t), t \geq 0)$ le processus qui représente le nombre de clients dans le système (file + serveur) à chaque instant t.

- L'espace d'état est $E = \mathbb{N}$
- Au départ ($t = 0$), la file est vide : $X(0) = 0$. (i.e. $\nu \,|\, \mathrm{P}(X(0) = 0) = 1$)
- A l'instant $t \geq 0$,
  - Si $X(t) = i$ avec $i > 0$, il y a i clients dans le système dont 1 en train d'être servi, et $i - 1$ dans la file d'attente. Les transitions possibles sont alors :
    - Une arrivée d'un nouveau client dans la file : on passe alors à $i + 1$ clients.Cet événement se produit au bout d'un temps suivant une loi exponentielle de paramètre λ.
    - Le départ du client servi : on passe alors à $i - 1$ clients. Cet événement se produit au bout d'un temps suivant une loi exponentielle de paramètre μ.
    - La probabilité qu'il y ait deux ou plus événements simultanés est négligeable.

    Le processus reste dans l'état i jusqu'au prochain événement. Le temps jusqu'au prochain événement est le minimum entre le temps jusqu'à la prochaine arrivée et le temps jusqu'au prochain départ : le temps jusqu'au prochain événement suit une loi exponentielle de paramètre $\lambda + \mu$.

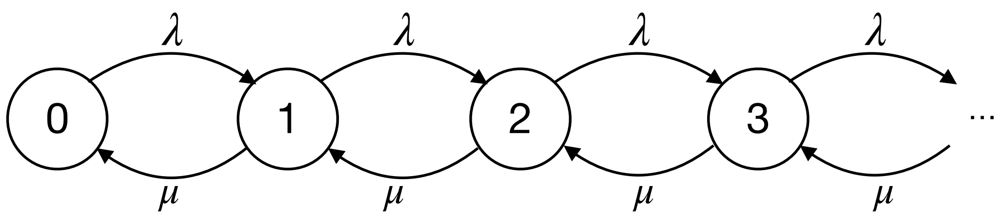
    Ainsi, la probabilité de transition de i à $i + 1$ est $\dfrac{\lambda}{\lambda + \mu}$, et celle de i à $i - 1$ est $\dfrac{\mu}{\lambda + \mu}$.
  - Si $X(t) = 0$, la seule transition possible est une arrivée, qui se produit après un temps suivant une loi exponentielle de paramètre λ.

On peut donc écrire la matrice Q :

$$Q = \begin{pmatrix} \lambda & 1 & & & & & 0 \\ \frac{\mu}{\lambda+\mu} & \lambda+\mu & \frac{\lambda}{\lambda+\mu} & & & & \\ & \frac{\mu}{\lambda+\mu} & \lambda+\mu & \frac{\lambda}{\lambda+\mu} & & & \\ & & \frac{\mu}{\lambda+\mu} & \lambda+\mu & \frac{\lambda}{\lambda+\mu} & & \\ & 0 & & \frac{\mu}{\lambda+\mu} & \lambda+\mu & \frac{\lambda}{\lambda+\mu} & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

On peut représenter X par un diagramme de transition :

- Chaque état est représenté par une bulle.
- Chaque transition est représentée par une flèche avec le taux de transition associé (paramètre de la loi exponentielle)



4. Générateur infinitésimal

Définition : Générateur infinitésimal
Soit X un processus de Markov homogène sur E de paramètres (ν, Q). Son générateur infinitésimal est la matrice A définie par :

- $A(x,x) = -q(x,x)$ pour tout $x \in E$
- $A(x,y) = q(x,x)q(x,y)$ pour tout $x, y \in E$ tels que $x \neq y$

Remarques :

- Le paramètre Q n'est pas utilisable, le générateur A est l'outil mathématique qui le remplace.
- $A(x,x)$ est l'opposé du paramètre de la loi exponentielle qui régit le temps de séjour dans l'état x.
- $\dfrac{A(x,y)}{|A(x,x)|}$ est la probabilité qu'un processus aille dans l'état y quand il quitte l'état x.
- $A(x,y)$ est le taux de transition de x à y.
- La somme de A sur les lignes est nulle : $\displaystyle\sum_{y \in E} A(x,y) = -q(x,x) + \sum_{y \neq x} q(x,x)q(x,y) = 0.$

Théorème : Le générateur infinitésimal caractérise entièrement le processus.
Soient X et Y deux processus de générateurs infinitésimaux respectifs A_X et A_Y,

$$A_X = A_Y \Leftrightarrow X \text{ et } Y \text{ ont même loi.}$$

Exemple : M/M/1

- $A(i,i) = -q(i,i) = \begin{cases} -\lambda & \text{si } i = 0 \\ -(\lambda+\mu) & \text{si } i > 0 \end{cases}$
- $A(i,i+1) = q(i,i)q(i,i+1) = \lambda$
- $A(i,i-1) = q(i,i)q(i,i-1) = \mu$ si $i > 0$
- $A(i,j) = 0$ si $j \notin \{i-1, i, i+1\}$

$$A = \begin{pmatrix} -\lambda & \lambda & & & & & 0 \\ \mu & -(\lambda+\mu) & \lambda & & & & \\ & \mu & -(\lambda+\mu) & \lambda & & & \\ & & \mu & -(\lambda+\mu) & \lambda & & \\ & 0 & & \mu & -(\lambda+\mu) & \lambda & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

5. Probabilité stationnaire

Théorème : Soit X un processus de Markov, et soit $\hat{X}$ sa chaine de Markov incluse. On a alors :

- $\hat{X}$ est irréductible $\Leftrightarrow$ X est irréductible
- Un état de $\hat{X}$ est récurrent (resp. transient) $\Leftrightarrow$ l'état correspondant de X est récurrent (resp. transient)
- X est récurrent si tous ses états le sont.

Théorème : Soit X un processus de Markov homogène, irréductible, récurrent, alors il existe une unique probabilité stationnaire π.
Elle s'obtient en résolvant $\pi A = 0$, où A est le générateur infinitésimal de X.
L'unicité s'obtient avec la condition de normalisation $\sum_{x\in E} \pi(x) = 1$.

Remarques :

- $$\begin{aligned} \pi A = 0 &\Leftrightarrow \forall y \in E, \pi A(y) = 0 \\ &\Leftrightarrow \forall y \in E, \sum_{x\in E} \pi(x)A(x,y) = 0 \\ &\Leftrightarrow \forall y \in E, \sum_{x\in E, x\neq y} \pi(x)q(x,x)q(x,y) - \pi(y)q(y,y) = 0 \\ &\Leftrightarrow \forall y \in E, \pi(y)q(y,y) = \sum_{x\in E, x\neq y} \pi(x)q(x,x)q(x,y) \end{aligned}$$

  On retrouve une équation d'équilibre en y.
- Si on note $\hat{\pi}$ la probabilité stationnaire de la chaine de Markov incluse, par définition on a $\hat{\pi}(y) = \sum_{x\in E} \hat{\pi}(x)q(x,y)$. Ainsi, $\hat{\pi}(x) = \pi(x)q(x,x)$ pour tout $x \in E$.

Théorème : Soit $X = (X(t), t \geq 0)$ un processus de Markov homogène, irréductible et récurrent. Alors il est ergodique :

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T f(X(t))\mathrm{d}t = \sum_{x\in E} f(x)\pi(x)$$

pour toute fonction $f : E \to \mathbb{R}$ intégrable.

Ainsi pour obtenir la probabilité stationnaire d'un processus de Markov homogène, irréductible et récurrent, il suffit de l'observer pendant un temps long.

En particulier si $f = 1_x$, alors l'équation devient $\mathrm{P}(X = x) = \pi(x)$.
Et si $f = 1_J$ pour $J \subset E$, alors l'équation devient $\mathrm{P}(X \in J) = \sum_{x \in J} \pi(x)$.

Exemple : M/M/1

$\pi A = 0 \Rightarrow -\lambda\pi(0) + \mu\pi(1) = 0$ et $\lambda\pi(i-1) - (\lambda+\mu)\pi(i) + \mu\pi(i+1) = 0 \ \forall i > 0$

$$\begin{aligned}
\Rightarrow \pi(1) &= \frac{\lambda}{\mu}\pi(0) \\
\pi(2) &= \frac{1}{\mu}((\lambda+\mu)\pi(1) - \lambda\pi(0)) = \frac{\lambda^2}{\mu^2}\pi(0) \\
\pi(3) &= \frac{1}{\mu}((\lambda+\mu)\pi(2) - \lambda\pi(1)) = \frac{\lambda^3}{\mu^3}\pi(0) \\
&\dots
\end{aligned}$$

On a donc pour tout $i > 0$, $\pi(i) = \frac{\lambda^i}{\mu^i}\pi(0) = \rho^i\pi(0)$ où ρ est la charge en Erlang.

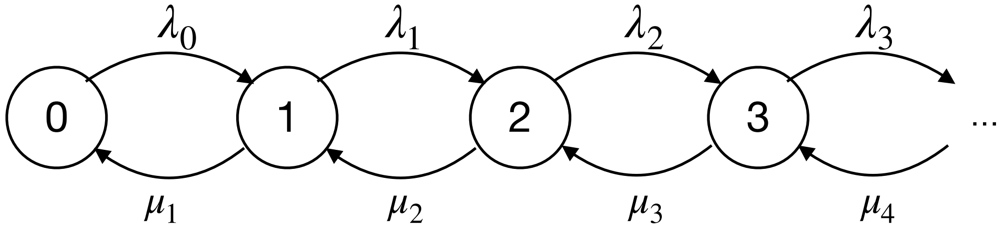
La condition de normalisation donne $\sum_{i=0}^{+\infty} \pi(i) = 1 \Leftrightarrow \sum_{i=0}^{+\infty} \rho^i\pi(0) = 1$.

- Si $\rho > 1$, la file M/M/1 n'est pas stable, les états sont transferts, il n'existe pas de probabilité stationnaire.
- Si $\rho < 1$, la file M/M/1 est homogène, récurrente et irréductible. La probabilité stationnaire existe et vaut $\pi(i) = \rho^i(1-\rho)$ pour tout $i \geq 0$. On reconnait une loi géométrique de paramètre $(1-\rho)$.

On peut calculer le nombre moyen de clients dans la file :

$$\mathrm{E}[X] = \sum_{i=0}^{+\infty} i\pi(i) = \sum_{i=0}^{+\infty} i\rho^i(1-\rho) = \frac{\rho}{1-\rho}.$$

Cas des processus de naissance et de mort :
Un processus de naissance et de mort est un processus où les seules transitions possibles sont depuis un état i vers les états $i-1$ ou $i+1$. Le diagramme de transition d'un tel processus est linéaire, les taux de transition ne sont pas forcément égaux.



Le générateur infinitésimal d'un processus de ce type est toujours de forme tri-diagonale :

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & & & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & & & 0 \\ & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & & & \\ & & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & & \\ & 0 & & \mu_4 & -(\lambda_4+\mu_4) & \lambda_4 & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

Et par récurrence, sa probabilité stationnaire, si elle existe, est, pour tout $i > 0$, de la forme $\pi(i) = \dfrac{\lambda_0 \dots \lambda_{i-1}}{\mu_1 \dots \mu_i}\pi(0)$.

Théorème : Soit X un processus de Markov homogène, irréductible et récurrent. On note π sa probabilité stationnaire. Alors, pour tout $x, y \in E$,

$$p_t(x,y) = \mathrm{P}(X(t) = y \,|\, X(0) = x) \underset{t\to+\infty}{\longrightarrow} \pi(y).$$

Cela signifie que quand t tend vers l'infini, la probabilité de transition de x vers y en un temps t ne dépend plus de l'état de départ x, mais tend vers la probabilité d'être dans l'état y.

6. Réversibilité

Soit $X = (X(t), t \geq 0)$ un processus de Markov homogène, irréductible et récurrent.
On pose $\bar{X}^T(t) = X(T-t)$, c'est aussi un processus de Markov homogène, irréductible et récurrent. On a :

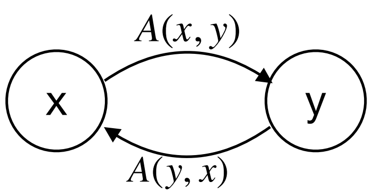
$$\begin{aligned}
\bar{p}_t(x,y) &= \mathrm{P}(\bar{X}^T(t) = y \,|\, \bar{X}^T(0) = x) \\
&= \mathrm{P}(X(T-t) = y \,|\, X(T) = x) \\
&= \mathrm{P}(X(T) = x \,|\, X(T-t) = y)\frac{\mathrm{P}(X(T-t) = y)}{\mathrm{P}(X(T) = x)}
\end{aligned}$$

$$\bar{p}_t(x,y) = p_t(y,x)\frac{\mathrm{P}(X(T-t) = y)}{\mathrm{P}(X(T) = x)}$$

A l'état stationnaire, l'équation devient : $\bar{A}(x,y) = A(y,x)\dfrac{\bar{\pi}(y)}{\pi(x)}$.

Définition : Processus réversible
X est réversible si $X = (X(t), 0 \leq t \leq T)$ et $\bar{X}^T = (X(T-t), 0 \leq t \leq T)$ ont même loi, i.e. si $A = \bar{A}$.

Théorème : $X = (X(t), t \geq 0)$ est réversible si et seulement si $\pi(x)A(x,y) = \pi(y)A(y,x)$.



Théorème : Un processus de naissance et de mort est réversible.

Démonstration : On a vu que $\pi(i) = \dfrac{\prod_{k=0}^{i-1} \lambda_k}{\prod_{k=1}^{i} \mu_k} \pi(0)$ et $\pi(i+1) = \dfrac{\lambda_i}{\mu_{i+1}} \pi(i)$.
On a $\pi(i)A(i,i+1) = \pi(i)\lambda_i$ d'une part, et $\pi(i+1)A(i+1,i) = \pi(i+1)\mu_{i+1}$ d'autre part. On retrouve bien l'égalité demandée. Les autres coefficients de A sont nuls.

Théorème de Kelly : Soit $X = (X(t), t \geq 0)$ un processus de Markov réversible sur E. On note A son générateur infinitésimal et π sa probabilité stationnaire. Alors on peut tronquer X sur $F \subset E$, on note $\tilde{X} = (\tilde{X}(t), t \geq 0)$ le processus tronqué :

- $\tilde{X}$ est réversible
- Son générateur infinitésimal $\tilde{A}$ est tel que $\tilde{A}(x,y) = A(x,y)$ pour $x, y \in F, x \neq y$, et $\tilde{A}(x,x) = -\sum_{y \in F, y \neq x} \tilde{A}(x,y)$ pour tout $x \in F$
- Sa probabilité stationnaire est $\tilde{\pi}(x) = c\pi(x)$ pour tout $x \in F$ où $c = \dfrac{1}{\sum_{x \in F} \tilde{\pi}(x)}$ est une constante de normalisation.

Démonstration : Si X est réversible, alors A vérifie les équations d'équilibres locales pour tout $x \in E$, et en particulier pour tout $x \in F$. C'est donc aussi le générateur infinitésimal du processus tronqué à F.

Exemple : La file M/M/S/S est la troncature de la file M/M/∞.

Théorème de Burke : Soit X le processus qui représente le nombre de clients de la file M/M/1. On écrit $\mathscr{A}(X) = (\mathscr{A}_1, \mathscr{A}_2, ..)$ les instants d'arrivées des clients et $\mathscr{D}(X) = (\mathscr{D}_1, \mathscr{D}_2, ..)$ les instants de départs, comme des fonctions de X.
X est réversible, donc $X = (X(t), 0 \leq t \leq T)$ et $\bar{X}^T = (X(T-t), 0 \leq t \leq T)$ sont égaux en loi, et $\mathscr{A}(\bar{X}^T)$ et $\mathscr{D}(X)$ sont égaux presque sûrement.
Le processus des départs d'une M/M/1 est donc un processus de Poisson d'intensité λ.

IV. Files d'attente

1. Systèmes à perte ou à attente

On distingue les systèmes à perte des systèmes à attente. Les systèmes à perte sont ceux où des clients peuvent être perdus quand le système (serveurs + salle d'attente) est plein. Les systèmes à attente ont, eux, une salle d'attente infinie.
En pratique, il n'existe pas de système à attente, mais on modélise par un système à attente tout système pour lequel le dimensionnement est tel que les pertes sont négligeables.
Pour un système à attente, un des principaux critères de performance est le temps d'attente.

Théorème : Formule de Little
On considère un système à attente où X est le processus qui représente le nombre de clients dans le système. On note λ le taux d'arrivées, et W le temps moyen de séjour dans la file (attente + service). Alors on a :

$$\mathrm{E}[X] = \lambda W.$$

Le temps moyen de séjour est donc égal au nombre moyen de clients divisé par le taux d'arrivées.

Exemple : M/M/1
Le temps moyen de séjour est $W = \dfrac{\mathrm{E}[X]}{\lambda} = \dfrac{\rho}{1-\rho}\lambda = \dfrac{1}{\mu - \lambda}$.

Lorsque le système est à perte, on cherchera plutôt à calculer la probabilité de perte, i.e. la probabilité qu'un client soit perdu. Dans un système à attente, on parlera de probabilité d'attente. Un client est perdu lorsqu'il arrive et que le système est plein. On cherche donc à calculer la probabilité que le système soit plein étant donnée l'arrivée d'un client :

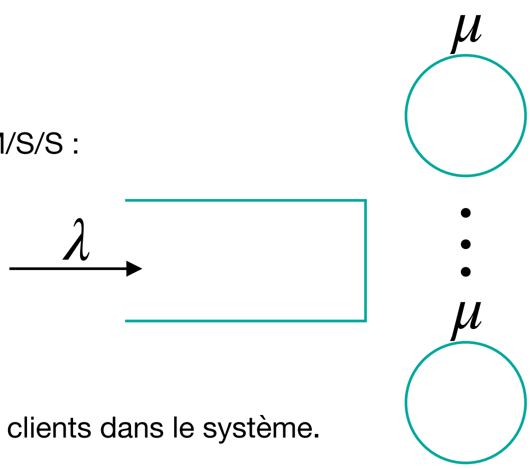
$$\begin{aligned}
\mathrm{P}_{\text{perte}} &= \mathrm{P}(\text{système plein} \mid \text{un client arrive}) \\
&= \frac{\mathrm{P}(\text{système plein, un client arrive})}{\mathrm{P}(\text{un client arrive})} \\
&= \frac{\sum_{x \text{ état bloquant}} \text{taux d'arrivées(état } x) \times \mathrm{P}(\text{état } x)}{\sum_x \text{taux d'arrivées(état } x) \times \mathrm{P}(\text{état } x)} \\
&= \frac{\text{taux de perte}}{\text{taux d'arrivées}}
\end{aligned}$$

Théorème : PASTA (Poisson Arrivals See Times Averages)
Quand le processus d'arrivées est un processus de Poisson, un point du processus (= un client de la file) voit la loi stationnaire.

Cela signifie que la probabilité qu'un client trouve la file dans l'état x (sans lui) est $\pi(x)$.
Ainsi, la probabilité de perte est égale à la probabilité de blocage :

$$\begin{aligned}
\mathrm{P}(\text{système plein} \mid \text{un client arrive}) &= \mathrm{P}(\text{système plein}) \\
&= \sum_{x \text{ état bloquant}} \mathrm{P}(\text{état } x)
\end{aligned}$$

2. Modèle d'Erlang-B



Le modèle d'Erlang-B correspond à une file M/M/S/S :

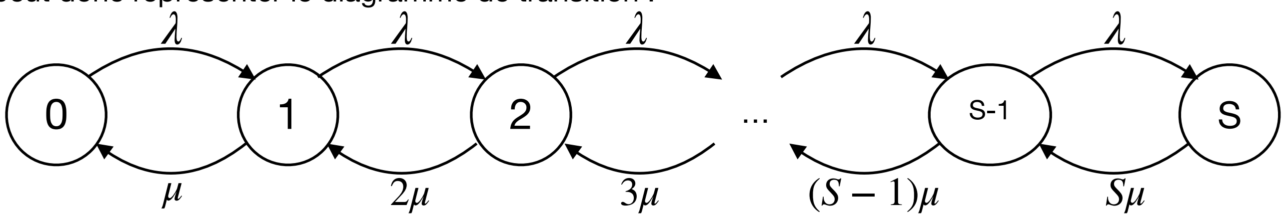
- Arrivées $\sim$ Poisson(λ)
- Temps de services $\sim$ exp(μ)
- S serveurs
- Pas de salle d'attente

On note $\rho = \dfrac{\lambda}{\mu}$ la charge en Erlang.

Soit X le processus qui représente le nombre de clients dans le système.
L'espace d'états est $E = \{0,1,2,\ldots,S\}$.
Les taux de transition sont :

- Arrivée : le temps avant la prochaine arrivée suit une loi exponentielle de paramètre λ, donc le taux de transition pour une arrivée est λ.
- Départ : si k serveurs sont occupés, alors le prochain départ est le premier des k clients en service à finir. Le temps avant le prochain départ est donc le minimum de k lois exponentielles de paramètre μ : $\min(\exp(\mu),\ldots\exp(\mu)) = \exp(k\mu)$. Le taux de transition pour un départ est donc de $k\mu$ lorsque k serveurs sont occupés.

On peut donc représenter le diagramme de transition :



X est un processus de naissance et de morts, on peut directement écrire son générateur infinitésimal :

$$A = \begin{pmatrix} -\lambda & \lambda & & & & & \\ \mu & -(\lambda+\mu) & \lambda & & & 0 & \\ & 2\mu & -(\lambda+2\mu) & \lambda & & & \\ & & 3\mu & -(\lambda+3\mu) & \lambda & & \\ & & & \ddots & \ddots & \ddots & \\ & 0 & & & (S-1)\mu & -(\lambda+(S-1)\mu) & \lambda \\ & & & & & S\mu & -S\mu \end{pmatrix}$$

X est un processus de Markov homogène, irréductible et récurrent sur un espace d'état fini, il admet donc une probabilité stationnaire :

$$\pi(i) = \frac{\lambda^i}{i!\mu^i}\pi(0) = \frac{\rho^i}{i!}\pi(0).$$

Avec la condition de normalisation, on a : $\displaystyle\sum_{j=0}^{S} \frac{\rho^j}{j!}\pi(0) = 1 \Leftrightarrow \pi(0) = \frac{1}{\sum_{j=0}^{S} \frac{\rho^j}{j!}}.$

Les arrivées suivent un processus de Poisson, donc d'après la propriété PASTA, la probabilité de perte est égale à la probabilité de blocage. Le seul état bloquant est celui où il y a S clients qui occupent les S serveurs.

On obtient ainsi la formule d'Erlang-B :

$$E(\rho, S) = \pi(S) = \frac{\frac{\rho^S}{S!}}{\sum_{j=0}^{S} \frac{\rho^j}{j!}}.$$

On note que la probabilité de perte ne dépende que de la charge ρ. De plus, on montre qu'elle est insensible à la distribution du processus d'arrivées.

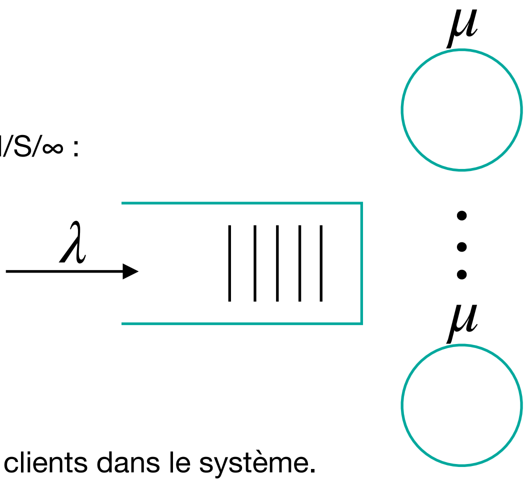
Utilisation pour le dimensionnement :

- On connait la charge ρ du système à dimensionner.
- On fixe un seuil de perte à ne pas dépasser.
- On augmente le nombre de serveurs S jusqu'à ce que $E(\rho, S)$ passe sous le seuil fixé.

Pour augmenter S, on peut se servir de la relation de récurrence :

- $E(\rho,0) = 1$ (perte sûre)
- $E(\rho, S) = \dfrac{E(\rho, S-1)}{\frac{S}{\rho} + E(\rho, S-1)}$.

3. Modèle d'Erlang-C



Le modèle d'Erlang-C correspond à une file M/M/S/∞ :

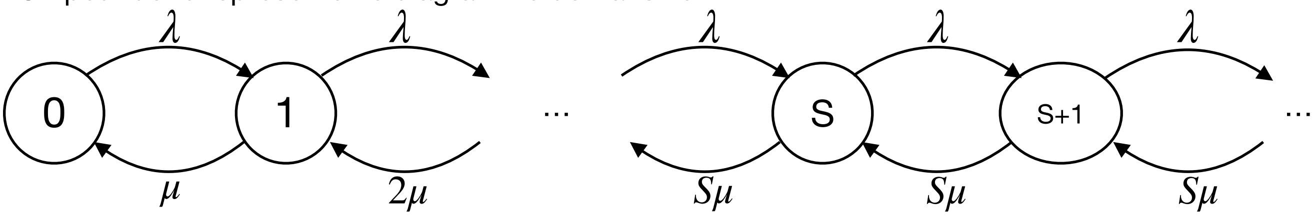
- Arrivées $\sim$ Poisson(λ)
- Temps de services $\sim$ exp(μ)
- S serveurs
- Salle d'attente infinie

On note $\rho = \dfrac{\lambda}{\mu}$ la charge en Erlang.

Soit X le processus qui représente le nombre de clients dans le système.
L'espace d'états est $E = \mathbb{N}$.
Les taux de transition sont :

- Arrivée : le temps avant la prochaine arrivée suit une loi exponentielle de paramètre λ, donc le taux de transition pour une arrivée est λ.
- Départ : si k serveurs sont occupés, alors le prochain départ est le premier des k clients en service à finir. Le temps avant le prochain départ est donc le minimum de k lois exponentielles de paramètre μ : $\min(\exp(\mu), \ldots \exp(\mu)) = \exp(k\mu)$. Le taux de transition pour un départ est donc de $k\mu$ lorsque k serveurs sont occupés. Au plus S serveurs sont occupés, les clients au delà du S-ième sont en attente.

On peut donc représenter le diagramme de transition :



X est un processus de naissance et de morts, on peut directement écrire son générateur infinitésimal :

$$A = \begin{pmatrix} -\lambda & \lambda & & & & & & 0 \\ \mu & -(\lambda+\mu) & \lambda & & & & & \\ & 2\mu & -(\lambda+2\mu) & \lambda & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & S\mu & -(\lambda+S\mu) & \lambda & & \\ & 0 & & & S\mu & -(\lambda+S\mu) & \lambda & \\ & & & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Si $\rho < S$, X est homogène, récurrent et irréductible, on peut donc écrire sa probabilité stationnaire :

$$\pi(i) = \begin{cases} \frac{\rho^i}{i!}\pi(0) & \text{pour } 0 \le i \le S \\ \frac{\rho^i}{S!S^{i-S}}\pi(0) & \text{pour } i \ge S+1 \end{cases} \quad \text{avec } \pi(0) = \frac{1}{\sum_{j=0}^{S}\frac{\rho^j}{j!} + \frac{\rho^S}{S!}\frac{\rho/S}{1-\rho/S}}.$$

D'après la propriété PASTA, la probabilité d'attente est égale à la probabilité que le système soit dans un état où les serveurs sont pleins.
C'est la formule d'Erlang-C :

$$C(S,\rho) = \sum_{i=S}^{+\infty}\pi(i) = \frac{\rho^S}{S!}\frac{1}{1-\frac{\rho}{S}}\pi(0).$$

On peut exprimer la probabilité d'attente du modèle d'Erlang-C en fonction de la probabilité de perte du modèle d'Erlang-B :

$$C(S,\rho) = \frac{SE(\rho,S)}{S-\rho+\rho E(\rho,S)} \ge E(\rho,S).$$

La probabilité d'attendre lorsqu'il y a une salle d'attente est supérieure à la probabilité de perte quand il n'y en a pas pour le même système par ailleurs.

4. Modèle d'Engset

Le modèle d'Engset diffère des modèles d'Erlang car il suppose que le nombre de clients est fini. Les arrivées ne sont donc plus Poissonniennes, on considère qu'il n'y a pas une infinité de clients potentiels, mais seulement un ensemble de M clients susceptibles de rejoindre le système, et lorsqu'un des M clients est déjà dans le système, il n'est alors plus susceptible d'y entrer tant que son service n'est pas fini. En revanche, on considère toujours que les inter-arrivées suivent une loi exponentielle. On a alors le modèle suivant :

- Temps entre 2 arrivées pour un client $\sim \exp(\lambda)$
- Temps de services $\sim \exp(\mu)$
- M clients possibles en tout
- S serveurs (avec $S \le M$)
- Pas de salle d'attente

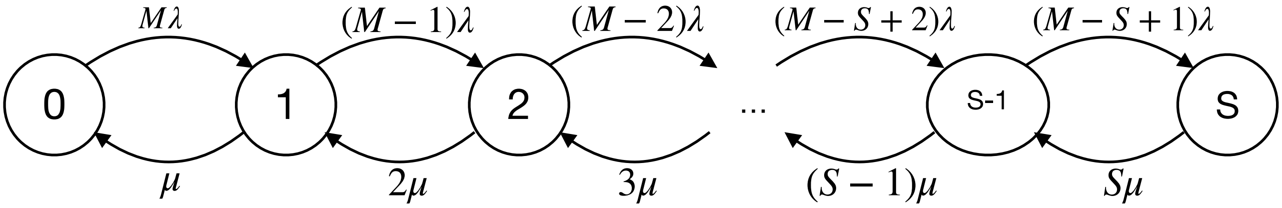
On note $\rho = \dfrac{\lambda}{\mu}$ la charge en Erlang.

L'espace d'états est $E = \{0,1,\ldots,S\}$.

Les taux de transition sont :

- Arrivée : si k clients sont déjà en service dans le système, alors seulement $M-k$ clients sont susceptibles d'arriver. La prochaine arrivée est le plus petit des $M-k$ temps inter-arrivées des $M-k$ clients hors du système, qui suivent chacun une loi exponentielle de paramètre λ. Ainsi le temps avant la prochaine arrivée est le minimum de $M-k$ lois exponentielles de paramètre λ, et suit une loi exponentielle de paramètre $(M-k)\lambda$.
- Départ : si k serveurs sont occupés, alors le prochain départ est le premier des k clients en service à finir. Le temps avant le prochain départ est donc le minimum de k lois exponentielles de paramètre μ : $\min(\exp(\mu),\ldots\exp(\mu)) = \exp(k\mu)$. Le taux de transition pour un départ est donc de $k\mu$ lorsque k serveurs sont occupés.

On peut donc représenter le diagramme de transition :



X est un processus de naissance et de morts, on peut directement écrire son générateur infinitésimal :

$$A = \begin{pmatrix} -M\lambda & M\lambda & & & & 0 \\ \mu & -((M-1)\lambda+\mu) & (M-1)\lambda & & & \\ & 2\mu & -((M-2)\lambda+2\mu) & (M-2)\lambda & & \\ & & \ddots & \ddots & \ddots & \\ & 0 & & (S-1)\mu & -((M-S+1)\lambda+(S-1)\mu) & (M-S+1)\lambda \\ & & & & S\mu & -S\mu \end{pmatrix}$$

X est un processus de Markov homogène, irréductible et récurrent sur un espace d'état fini, il admet donc une probabilité stationnaire :

$$\pi(i) = \frac{M(M-1)\ldots(M-i+1)\lambda^i}{i!\mu^i}\pi(0) = \frac{M!}{(M-i)!i!}\rho^i\pi(0) = \binom{M}{i}\rho^i\pi(0).$$

Avec la condition de normalisation, on a $\pi(0) = \dfrac{1}{\sum_{j=0}^{S}\binom{M}{j}\rho^j}$.

Ainsi, $\pi(i) = \dfrac{\binom{M}{i}\rho^i}{\sum_{j=0}^{S}\binom{M}{j}\rho^j}$.

Les arrivées ne suivent pas un processus de Poisson, la propriété PASTA ne s'applique pas. La probabilité de perte est alors :

$$\begin{aligned} \mathrm{P}_{\text{perte}} &= \frac{\text{taux de pertes}}{\text{taux d'arrivées}} \\ &= \frac{\sum_{x \text{ état bloquant}} \text{taux d'arrivées(état } x) \times \mathrm{P}(\text{état } x)}{\sum_x \text{taux d'arrivées(état } x) \times \mathrm{P}(\text{état } x)} \\ &= \frac{(M-S)\lambda\pi(S)}{\sum_{j=0}^{S}(M-j)\lambda\pi(j)} \end{aligned}$$

$$\mathrm{P}_{\text{perte}} = \frac{(M-S)\lambda \binom{M}{S} \rho^S \pi(0)}{\lambda \sum_{j=0}^{S} (M-j)\binom{M}{j} \rho^j \pi(0)} = \frac{\binom{M-1}{S} \rho^S}{\sum_{j=0}^{S} \binom{M-1}{j} \rho^j}.$$

On obtient ainsi la formule d'Engset :

$$\text{Eng}(\rho, S, M) = \frac{\binom{M-1}{S} \rho^S}{\sum_{j=0}^{S} \binom{M-1}{j} \rho^j}.$$

On remarque que la probabilité de perte est égale à la probabilité stationnaire d'être dans l'état avec S serveurs occupés, i.e. la probabilité de blocage pour $M-1$ clients.

De plus, on a la relation de récurrence suivante :

- $\text{Eng}(\rho, 0, M) = 1$
- $\text{Eng}(\rho, S, M) = \dfrac{\text{Eng}(\rho, S-1, M)}{\frac{S}{(M-S)\rho} + \text{Eng}(\rho, S-1, M)}.$

5. Modèle d'Erlang multi-classe

Le modèle d'Erlang multi-classe est une variante du modèle d'Erlang-B dans lequel les clients sont répartis en classes.

- Il y a N classes
- Arrivées des clients de la classe i pour $1 \leq i \leq N \sim$ Poisson(λ_i)
- Temps de services de la classe i pour $1 \leq i \leq N \sim \exp(\mu_i)$
- Il y a S serveurs

On note $\rho_i = \dfrac{\lambda_i}{\mu_i}$ la charge de la classe i, et $S_i \leq S$ le nombre de serveurs utilisés par les clients de la classe i pour $1 \leq i \leq N$.

On remarque que la totalité des arrivées est une superposition de processus de Poisson, c'est donc un processus de Poisson d'intensité $\displaystyle\sum_{i=1}^{N} \lambda_i$.

L'espace d'états est $E = \{(n_1, n_2, \ldots, n_N) \mid \displaystyle\sum_{i=1}^{N} n_i S_i \leq S\}$.

Si le nombre de serveurs S est infini, alors on a N processus indépendants de type Erlang-B, qui admettent donc la probabilité stationnaire :

$$\pi(n_1, n_2, \ldots, n_N) = \frac{\rho_1^{n_1}}{n_1!} \frac{\rho_2^{n_2}}{n_2!} \cdots \frac{\rho_N^{n_N}}{n_N!} \pi(0).$$

Par troncation, on obtient la probabilité stationnaire lorsque $S < +\infty$:

$$\pi(n_1, n_2, \ldots, n_N) = \frac{\rho_1^{n_1}}{n_1!}\frac{\rho_2^{n_2}}{n_2!}\ldots\frac{\rho_N^{n_N}}{n_N!}\pi(0) \text{ pour } (n_1, \ldots, n_N) \mid \sum_{i=1}^{N} n_i S_i \leq S$$

avec $\pi(0) = \dfrac{1}{\sum_{(n_1,\ldots,n_N)\in E} \frac{\rho_1^{n_1}}{n_1!}\ldots\frac{\rho_N^{n_N}}{n_N!}}$.

La probabilité de blocage des clients de la classe i est alors :

$$\mathrm{P}_{\text{blocage}} = \frac{\sum_{\text{états bloquants pour } i} \pi(n_1, \ldots, n_N)}{\sum_{(n_1,\ldots,n_N)\in E} \pi(n_1, \ldots, n_N)}.$$

On peut remarquer que cette probabilité de blocage n'est pas monotone pour les classes de trafic avec les charges les plus faibles.

V. Dimensionnement des réseaux mobiles

1. Cellule 2G isolée

Un opérateur cellulaire choisit le nombre de porteuses attribuées à chacune de ses cellules 2G, chaque porteuse est ensuite découpée en 8 slots temporels. Chaque communication occupe ensuite 1 slot = un huitième d'une porteuse donnée. Dimensionner une cellule est le fait de déterminer le nombre de canaux de communications qu'une cellule doit avoir pour satisfaire les demandes des utilisateurs, en 2G il s'agit donc de déterminer le nombre de porteuses à attribuer à chaque cellule.

On suppose que les arrivées d'appels suivent un processus de Poisson d'intensité λ, et que les durées de communication suivent une loi exponentielle de paramètre μ. On retrouve donc un modèle d'Erlang-B M/M/S/S. Et on cherche le nombre de canaux de communications, i.e. serveurs S nécessaires pour absorber la charge $\rho = \dfrac{\lambda}{\mu}$ avec un seuil de perte acceptable fixé. Typiquement, le seuil en téléphonie fixe est à 0,1%, et à 2% en téléphonie mobile.

Le problème est résolu par la formule d'Erlang-B :

$$E(\rho, S) = \frac{\frac{\rho^S}{S!}}{\sum_{j=0}^{S} \frac{\rho^j}{j!}}.$$

2. Réseau cellulaire

On considère maintenant une cellule faisant partie d'un réseau de cellules toutes identiques. On considère toujours des arrivées Poissoniennes d'intensité λ, et des durées de communication exponentielles de paramètre μ.

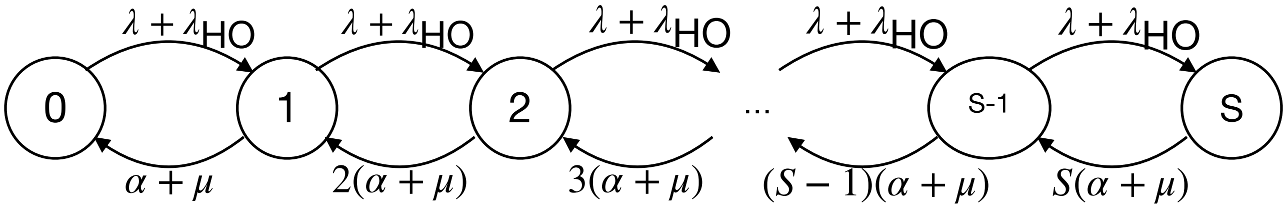
On doit considérer en plus les mouvements d'utilisateurs entre cellules : un appel passé dans une cellule peut être poursuivi dans une autre après le déplacement de l'utilisateur. Ce phénomène s'appelle le handover. On a donc des arrivées d'appel en cours de communication venant de cellules voisines : on suppose que ces arrivées suivent un processus de Poisson d'intensité λ_{HO}.

Dans une cellule donnée, la totalité des arrivées est donc une superposition de 2 processus de Poisson, et suit donc un processus de Poisson d'intensité $(\lambda + \lambda_{\text{HO}})$.

Symétriquement, il y a aussi des départs en handover en cours de communication vers des cellules voisines. On considère donc le temps de séjour d'un utilisateur dans la cellule, on suppose que ce temps de séjour suit une loi exponentielle de paramètre α. Le temps de service dans une cellule donnée est donc le minimum entre le temps de communication et le temps de séjour dans la cellule : c'est une loi exponentielle de paramètre $(\alpha + \mu)$.

La charge d'une cellule est donc $\rho = \dfrac{\lambda + \lambda_{\text{HO}}}{\alpha + \mu}$.

On retrouve un modèle d'Erlang-B avec des paramètres différents :



Les paramètres λ, α, et μ dépendent des utilisateurs, des tailles des cellules… Ils dépendent du système. En revanche, le paramètre λ_{HO} peut être exprimé à partir des trois autres à l'équilibre.

A l'équilibre, rentrent dans la cellule :

- Des nouveaux utilisateurs au taux λ
- Des utilisateurs en handover au taux λ_{HO}

Et sortent de la cellule :

- Des utilisateurs qui ont terminé leur appel au taux = taux d'entrée x probabilité de raccrocher = $(\lambda + \lambda_{\text{HO}})\dfrac{\mu}{\alpha + \mu}$
- Des utilisateurs qui partent en handover au taux $(\lambda + \lambda_{\text{HO}})\dfrac{\alpha}{\alpha + \mu}$

A l'équilibre, étant donné que toutes les cellules sont identiques, ce qui entre en handover est égal à ce qui sort en handover :

$$\lambda_{\text{HO}} = (\lambda + \lambda_{\text{HO}})\frac{\alpha}{\alpha+\mu} \Leftrightarrow \lambda_{\text{HO}}(1 - \frac{\alpha}{\alpha+\mu}) = \lambda\frac{\alpha}{\alpha+\mu}$$
$$\Leftrightarrow \lambda_{\text{HO}}(\alpha + \mu - \alpha) = \alpha\lambda$$
$$\Leftrightarrow \lambda_{\text{HO}} = \frac{\alpha}{\mu}\lambda$$

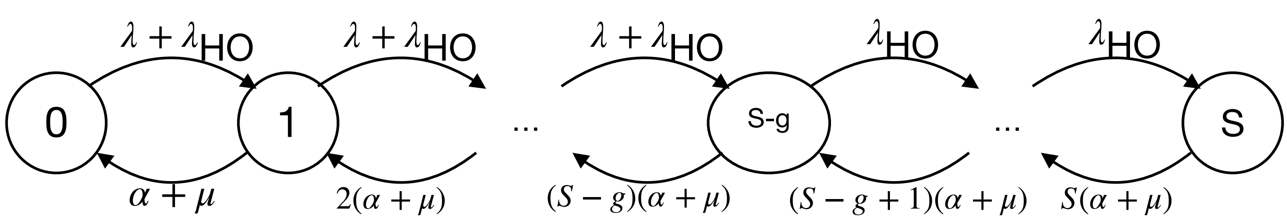
On a donc la charge :

$$\rho = \frac{\lambda + \lambda_{\text{HO}}}{\alpha + \mu} = (1 + \frac{\alpha}{\mu})\frac{\lambda}{\alpha+\mu} = \frac{\lambda}{\mu}.$$

La charge est inchangée. La formule d'Erlang-B est insensible au processus d'arrivées et ne dépend que de la charge. Donc le dimensionnement en cellule isolée reste valide.

Cependant, on a considéré que les handovers et les nouveaux appels étaient traités de la même manière. Or, généralement, les handovers ont la priorité sur les nouveaux appels car il est plus désagréable de se faire couper une communication que de ne pas pouvoir passer un appel.

On considère donc un système où il y a S canaux de communications, dont g canaux de garde sont réservés exclusivement pour les arrivées en cours de communication par handovers. Ainsi tant qu'il y a strictement moins de $S - g$ canaux utilisés dans la cellule, tous les appels sont acceptés : nouveaux et handovers, les arrivées suivent un processus de Poisson d'intensité $(\lambda + \lambda_{\text{HO}})$.
En revanche, lorsqu'il y a moins de g canaux libres, soit entre $S - g$ et S canaux occupés, alors seuls les handovers sont acceptés, et les arrivées suivent un processus de Poisson d'intensité λ_{HO}. On a le graphe de transition suivant :



X est un processus de Markov homogène, récurrent et irréductible sur un espace d'état fini. C'est un processus de naissances et de morts on peut donc directement écrire son générateur infinitésimal et sa probabilité stationnaire.

$$A = \begin{pmatrix}
-(\lambda+\lambda_{HO}) & \lambda+\lambda_{HO} & & & & & & 0 \\
\alpha+\mu & -(\lambda+\lambda_{HO}+\alpha+\mu) & \lambda+\lambda_{HO} & & & & & \\
 & \ddots & \ddots & \ddots & & & & \\
 & & (S-g)(\alpha+\mu) & -(\lambda_{HO}+(S-g)(\alpha+\mu)) & \lambda_{HO} & & & \\
 & & & \ddots & \ddots & \ddots & & \\
 & 0 & & & (S-1)(\alpha+\mu) & -(\lambda_{HO}+(S-1)(\alpha+\mu)) & \lambda_{HO} \\
 & & & & & S(\alpha+\mu) & -S(\alpha+\mu)
\end{pmatrix}$$

et $\pi(i) = \begin{cases} \dfrac{\rho^i}{i!}\pi(0) & \text{pour } 0 \leq i \leq S-g \\ \dfrac{\rho^{S-g}\rho_{HO}^{i-S+g}}{i!}\pi(0) & \text{pour } S-g \leq i \leq S \end{cases}$ avec $\rho = \dfrac{\lambda+\lambda_{HO}}{\alpha+\mu}$ et $\rho_{HO} = \dfrac{\lambda_{HO}}{\alpha+\mu}$.

On obtient $\pi(0)$ par normalisation.

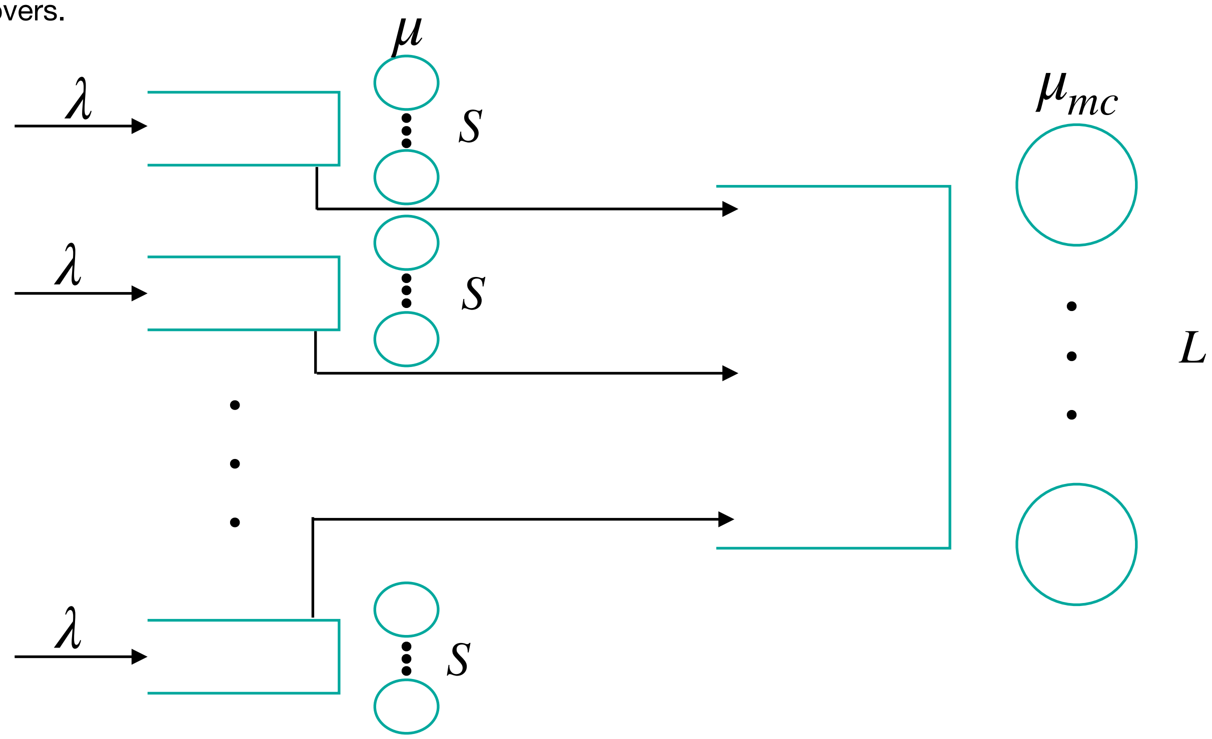
Les arrivées suivent un processus de Poisson, on peut donc appliquer la propriété PASTA. La probabilité de perte des appels en handovers est $\pi(S)$, tandis que la probabilité de perte des nouveaux appels est $\displaystyle\sum_{i=S-g}^{S} \pi(i)$.

3. Réseau hiérarchique

On considère maintenant un réseau cellulaire où toutes les cellules ne sont pas identiques. Il y a deux types de cellules :

- des cellules "normales" fonctionnant selon le modèle Erlang-B M/M/S/S
- des macro-cellules "parapluie" dont la portée recouvre plusieurs cellules normales et qui doivent absorber le débordement de ces dernières.

On cherche à dimensionner une cellule parapluie, ou macro-cellule, qui absorbe le débordement de n cellules identiques ayant chacune S canaux de communications. On note L le nombre de serveurs de la macro-cellule à déterminer. Pour simplifier les choses, on ne considère plus les handovers.



On note respectivement μ et μ_{mc} les paramètres des lois exponentielles régissant les temps de services dans les petites cellules, et dans la macro-cellule.

On commence par étudier le processus correspondant au débordement d'une seule petite cellule :

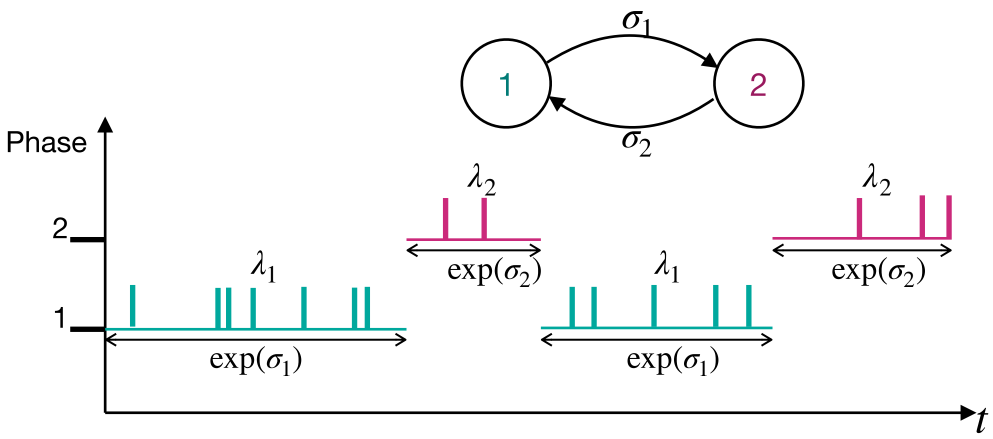
- Pendant x% du temps, la cellule sert la totalité de ses utilisateurs, elle ne déborde pas. Les arrivées dans la macro-cellule de débordement sont nulles. On peut les voir comme un processus de Poisson d'intensité 0.
- Pendant le reste du temps, (100-x)%, la cellule est pleine et tous les utilisateurs qui arrivent débordent. Les arrivées dans la macro-cellule de débordement suivent alors un processus de Poisson d'intensité λ.

Ainsi le débordement d'une cellule n'est pas un processus de Poisson, car l'intensité change. C'est un processus de Poisson à phases.

Définition : MMPP (Markov Modulated Poisson Process)
Un processus MMPP est un processus à phases tel que :

- Dans chaque phase, c'est un processus de Poisson dont l'intensité dépend de la phase.
- Le temps de séjour dans chaque phase suit une loi exponentielle dont le paramètre dépend de la phase.

Exemple : MMPP à 2 phases
On considère un MMPP à 2 phases dont les intensités dans chaque phase sont respectivement λ_1 et λ_2, et dont les temps de séjour moyens dans chaque phase sont respectivement $1/\sigma_1$ et $1/\sigma_2$.



Soit X un processus MMPP à m phases, X est caractérisé par :

- m intensités $\lambda_1, \lambda_2, \ldots, \lambda_m$ que l'on peut écrire sous forme matricielle : $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$
- Le générateur infinitésimal Q_J du processus de changements de phases.
  On note J le processus de changements de phases de générateur infinitésimal Q_J et de probabilité stationnaire ν_j.

Exemple : MMPP à 2 phases
On a donc $m = 2$. Les paramètres du processus sont :

- $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$
- $Q_J = \begin{pmatrix} -\sigma_1 & \sigma_1 \\ \sigma_2 & -\sigma_2 \end{pmatrix}$ et $\nu_J Q_J = 0 \Leftrightarrow \nu_J(1) = \dfrac{\sigma_2}{\sigma_1 + \sigma_2}$ et $\nu_J(2) = \dfrac{\sigma_1}{\sigma_1 + \sigma_2}$.

On cherche donc à modéliser le débordement d'une file M/M/S/S par un MMPP. Il faut que le temps de séjour dans chaque phase suive une loi exponentielle, donc on ne peut pas se contenter de 2 phases, car le temps passé à ne pas déborder ne suit pas une loi exponentielle. Cependant, la cellule ne déborde pas lorsqu'il y 0, 1, 2, ou S-1 serveurs occupés, donc quand le processus représentant la file M/M/S/S est dans l'état 0, 1, 2, ou S-1. Or le processus représentant la file M/M/S/S est markovien et reste donc un temps exponentiel dans chacun de ces états. Ainsi, il faut que chaque état du processus de la file corresponde à une phase du processus de débordement. Et les transitions entre phases correspondent aux transitions entre états de file M/M/S/S dont on connait déjà le générateur infinitésimal.
Dans les S premières phases, de 0 à S-1 serveurs occupés, les arrivées sont nulles et suivent un processus de Poisson d'intensité nulle. Et dans la dernière phase, à S serveurs occupés, les arrivées suivent un processus de Poisson d'intensité λ.

Finalement le processus de débordement d'une cellule à S serveurs est un MMPP dont les paramètres sont :

- $m = S + 1$ phases
- $\Lambda = \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \lambda \end{pmatrix}$
- $Q_J = \begin{pmatrix} -\lambda & \lambda & & & & \\ \mu & -(\lambda+\mu) & \lambda & & & \\ & 2\mu & -(\lambda+2\mu) & \lambda & & \\ & & \ddots & \ddots & \ddots & \\ & & & (S-1)\mu & -(\lambda+(S-1)\mu) & \lambda \\ & & & & S\mu & -S\mu \end{pmatrix}$

  avec $\nu_J(i) = \dfrac{\frac{\rho^i}{i!}}{\sum_{j=0}^{S} \frac{\rho^j}{j!}}$

On s'intéresse maintenant au débordement de n cellules à S serveurs, ce qui correspond à la superposition de n processus MMPP.

Théorème : Soient $X_1, \ldots, X_n$ n processus MMPP indépendants. Pour tout $1 \leq i \leq n$, le processus X_i a m_i phases et est caractérisé par les matrices Λ_i et Q_i.
Le processus de superposition des X_i est un processus MMPP avec :

- $m = \prod_{i=1}^{n} m_i$
- $\Lambda = \Lambda_1 \oplus \ldots \oplus \Lambda_n$
- $Q = Q_1 \oplus \ldots \oplus Q_n$ avec $\nu = \nu_1 \otimes \ldots \otimes \nu_n$.

Les symboles $\otimes$ et $\oplus$ sont les opérations de Kronecker :

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}$$

$A \oplus B = A \otimes I_B + I_A \otimes B$ où I_A est la matrice identité de la même taille que A.

Le débordement de n cellules à S serveurs est donc un MMPP avec :

- $m = (S+1)^n$ phases
- $\Lambda_{mc} = \Lambda \oplus \ldots \oplus \Lambda$
- $Q_{mc} = Q_J \oplus \ldots \oplus Q_J$ avec $\nu_{mc} = \nu_J \otimes \ldots \otimes \nu_J$

Avec Λ, Q_J, ν_J tels que définis précédemment pour le débordement d'une unique cellule.

Le dimensionnement de la macro-cellule de débordement revient dont à l'étude de la file MMPP/M/L/L :

- Arrivées $\sim$ MMPP(Q_{mc}, Λ_{mc})
- Temps de service $\sim \exp(\mu_{mc})$
- L serveurs
- Pas de salle d'attente

On appelle X_{mc} le processus d'occupation de la macro-cellule. Le taux de transition entre 2 états de X_{mc} dépend de la phase dans laquelle est le processus MMPP qui régit les arrivées. Ainsi, X_{mc} n'est pas un processus de Markov.

On note J_{mc} le processus de changement de phases du processus MMPP d'arrivées dont le générateur infinitésimal est Q_{mc}.

Le processus couple (X_{mc}, J_{mc}) est alors un processus de Markov :

- X_{mc} a $L+1$ états (de 0 à L serveurs occupés)
- J_{mc} a $m = (S+1)^n$ états

Donc (X_{mc}, J_{mc}) a $(L+1)(S+1)^n$ états. On écrit son générateur infinitésimal A_{mc} par blocs :

- On a $L+1$ blocs de taille $(S+1)^n$.
- Chacun des $L+1$ blocs correspond à un état de X_{mc}.
- Pour que X_{mc} passe de l'état i à $i+1$, le taux de transition dépend de l'état de J_{mc} : c'est la matrice Λ_{mc}.
- Pour que X_{mc} passe de l'état i à $i-1$, le taux de transition ne dépend que du nombre de serveurs occupé et vaut $i\mu_{mc}$. Par bloc, on retrouve donc une matrice $i\mu I_m$ où I_m est la matrice identité de taille m.
- Sur la diagonale, dans chaque bloc on retrouve le générateur de J_{mc} pour les transitions entre les phases du MMPP pour chaque état possible de X_{mc}.
- On retrouve aussi sur la diagonale les coefficients nécessaires pour que la somme sur les lignes soit nulle.

$$A_{mc} = \begin{pmatrix} Q_{mc} - \Lambda_{mc} & \Lambda_{mc} & & & & 0 \\ \mu_{mc} I_m & Q_{mc} - \Lambda_{mc} - \mu I_m & \Lambda_{mc} & & & \\ & 2\mu_{mc} I_m & Q_{mc} - \Lambda_{mc} - 2\mu I_m & \Lambda_{mc} & & \\ & & \ddots & \ddots & \ddots & \\ & & & (L-1)\mu_{mc} I_m & Q_{mc} - \Lambda_{mc} - (L-1)\mu I_m & \Lambda_{mc} \\ 0 & & & & L\mu_{mc} I_m & Q_{mc} - L\mu I_m \end{pmatrix}$$

Le processus couple (X_{mc}, J_{mc}) est un processus de Markov homogène, irréductible et récurrent (espace d'états fini). Il admet donc une probabilité stationnaire π qui vérifie $\pi A_{mc} = 0$.

Pour la résolution, on écrit π par bloc $\pi = (x_0, x_1, \ldots, x_L)$ où x_i, pour $0 \leq i \leq L$, est la probabilité que i serveurs soient occupés quelle que soit la phase de J_{mc}.

On obtient les $L+1$ équations suivantes :

$$\pi A_{mc} \Leftrightarrow \begin{cases} x_0(Q_{mc}-\Lambda_{mc})+\mu_{mc}x_1=0 \\ x_0\Lambda_{mc}+x_1(Q_{mc}-\Lambda_{mc}-\mu_{mc}I_m)+2\mu_{mc}x_2=0 \\ \vdots \\ x_{L-2}\Lambda_{mc}+x_{L-1}(Q_{mc}-\Lambda_{mc}-(L-1)\mu_{mc}I_m)+L\mu_{mc}x_L=0 \\ x_{L-1}\Lambda_{mc}+x_L(Q_{mc}-L\mu_{mc}I_m)=0 \end{cases}$$

$$\Leftrightarrow \begin{cases} x_1=-\frac{1}{\mu_{mc}}x_0(Q_{mc}-\Lambda_{mc}) \\ x_i=-\frac{1}{i\mu_{mc}}(x_{i-2}\Lambda+x_{i-1}(Q_{mc}-\Lambda_{mc}-(i-1)\mu_{mc}I_m)) \text{ pour } 2\leq i\leq L \end{cases}$$

On pose alors :

- $R_0=I_m$
- $R_1=-\dfrac{1}{\mu_{mc}}(Q_{mc}-\Lambda_{mc})$
- $R_i=-\dfrac{1}{i\mu_{mc}}(R_{i-2}\Lambda_{mc}+R_{i-1}(Q_{mc}-\Lambda_{mc}-(i-1)\mu_{mc}I_m))$ pour $2\leq i\leq L$.

Et on a alors pour tout $0\leq i\leq L$, $x_i=x_0R_i$.

Il ne reste qu'à obtenir x_0.
Les x_i sont par constructions des vecteurs de taille m. Pour $1\leq j\leq m$, $x_i(j)$ est la probabilité que le processus couple (X_{mc},J_{mc}) soit dans l'état (i,j), i.e. qu'il y ait i serveurs occupés et que le processus d'arrivées soit dans la j-ème phase. Et $\sum_{i=0}^{L}x_i(j)$ est alors la probabilité que le processus d'arrivées soit dans la j-ème phase, quelque soit le nombre de serveurs de la macro-cellule occupés. La probabilité stationnaire de J_{mc} est ν_{mc}, on a alors :

$$\sum_{i=0}^{L}x_i(j)=\nu_{mc}(j) \text{ pour tout } 0\leq j\leq L$$

Si on écrit ν_{mc} comme un vecteur, on a donc : $\sum_{i=0}^{L}x_i=x_0\sum_{i=0}^{L}R_i=\nu_{mc}$.
Ce qui permet d'obtenir :

$$x_0=\nu_{mc}(\sum_{i=0}^{L}R_i)^{-1}.$$

On a maintenant la probabilité stationnaire du processus (X_{mc},J_{mc}). On cherche maintenant à calculer la probabilité de perte afin de dimensionner la macro-cellule.
Les arrivées ne suivent pas un processus de Poisson, on ne peut donc pas appliquer la propriété PASTA.
On reprend la définition initiale de la probabilité de perte :

$$\mathrm{P}_{\text{perte}}=\mathrm{P}(\text{système plein}\,|\,\text{un client arrive}).$$

On a $(X_{mc}(t), t \geq 0)$ le processus représentant le nombre de clients, i.e. serveurs occupés, dans la macro-cellule. On note $(A(t), t \geq 0)$ le processus MMPP de paramètres (Q_{mc}, Λ_{mc}) des arrivées dans la macro-cellule. On note $\Delta A(t)$ la variation de A à l'instant t, ainsi $\Delta A(t) = 1$ signifie qu'il y a eu une arrivée à l'instant t. Ainsi,

$$\mathrm{P}_{\text{perte}} = \mathrm{P}(X_{mc}(t) = L \,|\, \Delta A(t) = 1) = \frac{\mathrm{P}(X_{mc}(t) = L, \Delta A(t) = 1)}{\mathrm{P}(\Delta A(t) = 1)}$$

On a d'une part :

$$\begin{aligned}
\mathrm{P}(X_{mc}(t) = L, \Delta A(t) = 1) &= \sum_{i=1}^{m} \mathrm{P}(X_{mc}(t) = L, \Delta A(t) = 1, J_{mc}(t) = i) \\
&= \sum_{i=1}^{m} \mathrm{P}(\Delta A(t) = 1 \,|\, X_{mc}(t) = L, J_{mc} = i)\mathrm{P}(X_{mc}(t) = L, J_{mc}(t) = i) \\
&= \sum_{i=1}^{m} \mathrm{P}(\Delta A(t) = 1 \,|\, J_{mc}(t) = i)\mathrm{P}(X_{mc}(t) = L, J_{mc}(t) = i)
\end{aligned}$$

$\mathrm{P}(\Delta A(t) = 1 \,|\, J_{mc}(t) = i)$ est le taux d'arrivées en phase i, c'est donc $\lambda_i = \Lambda_{mc}(i, i)$.
Le processus (X_{mc}, J_{mc}) est markovien homogène, irréductible et récurrent, il est donc ergodique :

$$\mathrm{P}(X_{mc}(t) = L, J_{mc}(t) = i) \xrightarrow{t \to +\infty} \pi(L, i) = x_L(i).$$

Et, $\mathrm{P}(X_{mc}(t) = L, \Delta A(t) = 1) \xrightarrow{t \to +\infty} \sum_{i=1}^{m} \lambda_i x_L(i)$

D'autre part :

$$\begin{aligned}
\mathrm{P}(\Delta A(t) = 1) &= \sum_{i=1}^{m} \mathrm{P}(\Delta A(t) = 1, J_{mc}(t) = i) \\
&= \sum_{i=1}^{m} \mathrm{P}(\Delta A(t) = 1 \,|\, J_{mc}(t) = i)\mathrm{P}(J_{mc}(t) = i)
\end{aligned}$$

On retrouve le taux d'arrivées en phase i, λ_i.
De plus, le processus $(J_{mc}(t), t \geq 0)$ est markovien, homogène, irréductible et récurrent, il est donc ergodique également :

$$\mathrm{P}(J_{mc}(t) = i) \xrightarrow{t \to +\infty} \nu_{mc}(i).$$

On a finalement :

$$\mathrm{P}_{\text{perte}} = \frac{\sum_{i=1}^{m} \lambda_i x_L(i)}{\sum_{j=1}^{m} \lambda_j \nu_{mc}(j)} = \sum_{i=1}^{m} \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j \nu_{mc}(j)} x_L(i).$$

La probabilité de perte est donc la somme sur les phases des probabilités de perte dans chaque phase pondérées par la proportion d'arrivées dans chaque phase.