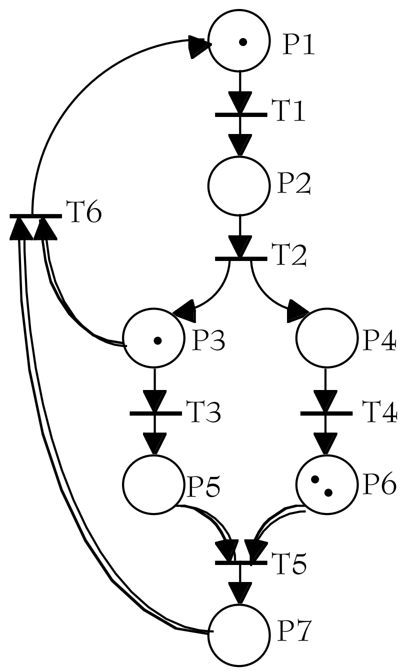
A-Processus déterministes :
Les Réseaux de Petri

Les Réseaux de Petri (RdP) permettent de modéliser des systèmes séquentiels. Ils ont été inventés par Carl Adam Petri, un mathématicien Allemand contemporain (d'où l'absence d'accent dans Petri). Il a défini un outil mathématique très général permettant de décrire les relations existant entre des conditions et des événements et de modéliser le comportement de systèmes dynamiques à événements discrets. Ces RdP datent de 1960-1962. C'est un outil très général, modélisant aussi bien les protocoles de communication informatiques que des systèmes de production. Il est à l'origine du Grafcet (ce dernier étant spécialisé dans la description de la commande de systèmes automatisés).

Chapitre 1 : Les réseaux de Petri autonomes

1.1 NOTIONS DE BASE

Un exemple de réseau de Petri



Nous retrouvons les éléments de base classiques : la place, la transition et l'arc reliant soit une place à une transition soit une transition à une place. On numérote les places P1, P2, … et les transitions T1, T2, …Chaque place Pi peut avoir un nombre mi de marques, mi □ℕ . On appelle M le vecteur des marquages dont les éléments sont les mi. Dans notre exemple, M=□(1,0,0,1,0,0,2,0).

<u>**Franchissement de transition :**</u>

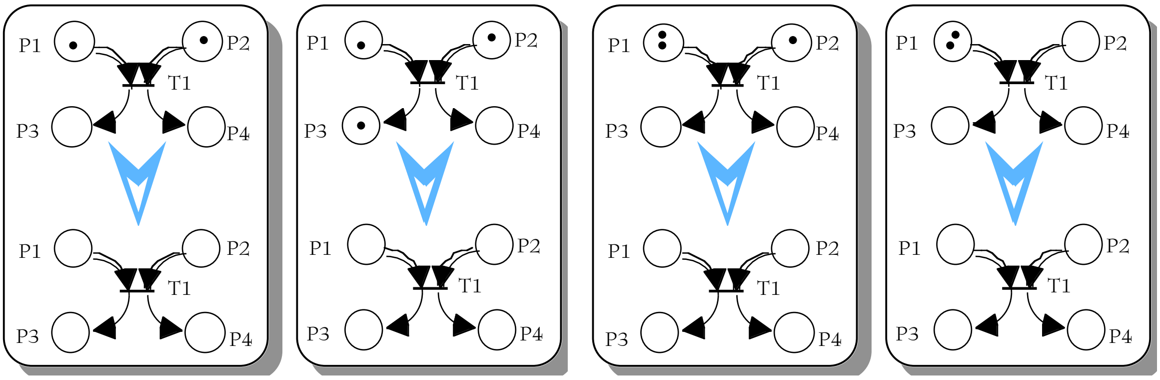
L'appellation autonome de ces réseaux de Petri vient du fait qu'on n'associe pas de condition,

d'événement, ni de temps aux transitions. Pour qu'une transition soit franchissable (ou validée), il faut et il suffit qu'il y ait au moins un jeton dans chaque place en amont de la transition.
Le franchissement (ou tir) d'une transition consiste à retirer une marque à chaque place en amont de la transition et à en rajouter une à chaque place en aval de la transition.

Remarques :

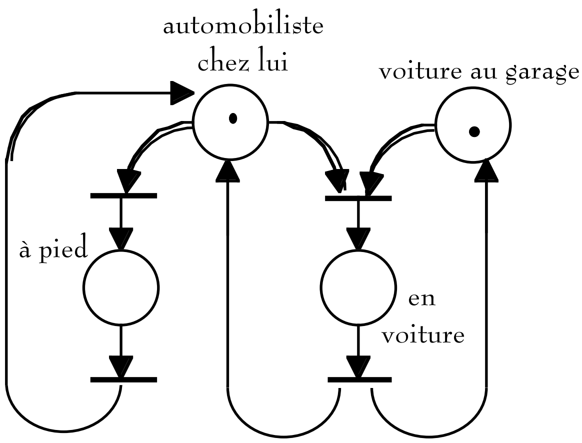
- Ces deux actions se font en même temps : le franchissement d'une transition n'est pas divisible.
- On peut remarquer qu'il n'y a pas conservation du nombre de jetons.
- Contrairement au Grafcet, lorsque deux transitions sont franchissables, on n'en franchit qu'une à la fois.
- lorsqu'une transition est validée, cela n'implique pas qu'elle soit immédiatement franchie. Ce n'est qu'une possibilité.

Pour les quatre situations ci-dessous, compléter le marquage du réseau après franchissement de T1 s'il est possible.



1.2 MODÉLISATION PAR RDP

Le RdP autonome récrit "**ce qui arrive**". Il permet la description du fonctionnement, de ce qui arrive, les transitions sont autonomes. Voici un exemple de description :

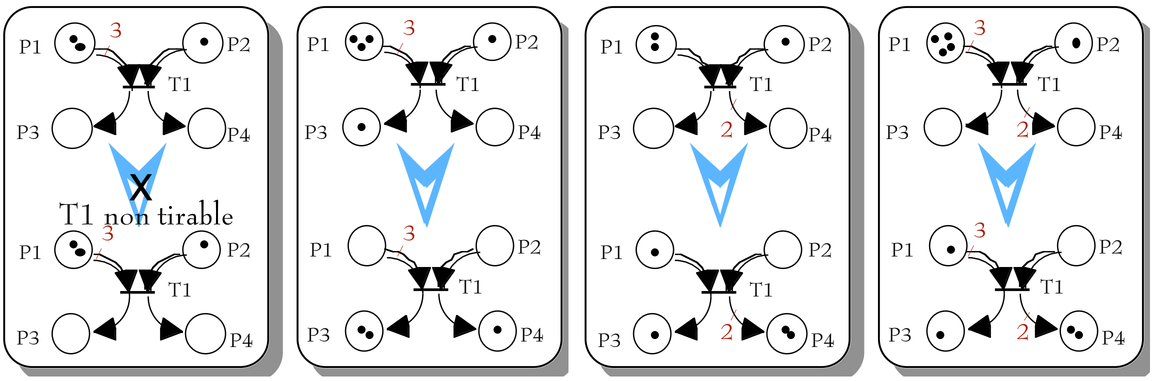


L'évolution du système est représenté par l'évolution du marquage.

1.3 RdP GÉNÉRALISÉS

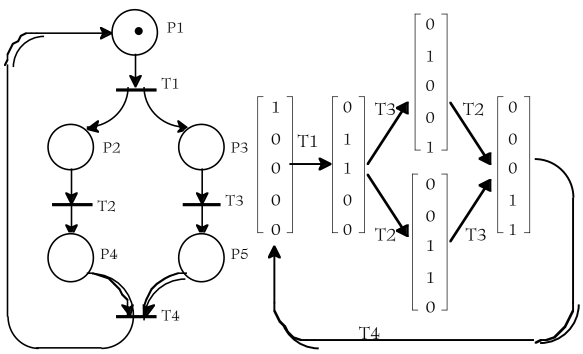
On peut affecter des poids aux arcs. Lorsque qu'il y a un poids n sur l'arc qui part d'une place Pi vers une transition Tj, cela signifie que la transition est franchissable s'il y a au moins n marques dans la place Pi. Lors du franchissement de Tj, on retirera n marques à la place Pi.
S'il y a un poids m sur l'arc qui part de la transition Tj vers la place Pk, lors du franchissement de Tj, on ajoutera m jetons dans la place Pk.

Exemples :



1.4 GRAPHE DES MARQUAGES ET ARBRES DE COUVERTURE

Le **graphe des marquages** est un graphe dont les noeuds correspondent à un marquage du RdP et les arcs, des franchissements de transitions.
Exemple :



Sur ce graphe des marquages, on peut trouver toutes les propriétés de ce RdP. On voit notamment que ce RdP est borné, sauf, vivant, réinitialisable, que T1T3T2T4 et T1T2T3T4 sont des séquences répétitives (donc le RdP est répétitif) et que $2.m_1+m_2+m_3+m_4+m_5= 2$ (donc le RdP est conservatif).

Un **arbre de couverture** est un graphe particulier dans lequel il n'y a pas de boucle ni de circuit.

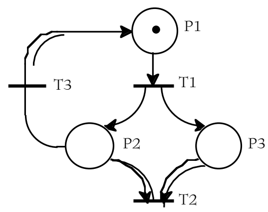
Algorithme de construction de l'arbre de couverture.

- **Pas 1** : A partir du marquage initial M_0, on indique toutes les transitions validées et les marquages successeurs correspondants. Si un de ces marquages est strictement supérieur à M_0, on met □ pour chacune des composantes supérieures aux composantes correspondantes de M_0.
- **Pas 2** : Pour chaque nouveau marquage M_i de l'arbre, on fait soit le pas 2.1 soit le pas 2.2
  - ◇ **Pas 2.1** : S'il existe sur le chemin de M_0 à M_i (ce dernier exclu) un marquage $M_j = M_i$, alors M_i

n'a pas de successeur.

◇ **Pas 2.2** : S'il n'existe pas de marquage $M_k = M_i$ sur le chemin de M_0 à M_i, alors on prolonge l'arbre en ajoutant tous les successeurs de M_i. Pour chaque successeur M_k de M_i : (a) une composante □ de Mi reste une composante □ de M_k ; (b) s'il existe un marquage M_j sur le chemin de M_0 à M_k tel que $M_k > M_j$, alors on met □ pour chacune des composantes supérieures aux composantes de M_j.

Exemple :

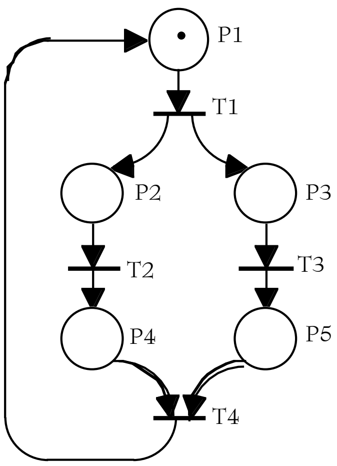


Faire l'arbre de couverture de ce RdP.

1.5 PROPRIÉTÉS

a) Notations et définitions

Le marquage d'un RdP est donné par un vecteur colonne dont la ième composante est le nombre de marques dans la place Pi. Pour faciliter l'écriture, on écrira le marquage sous sa forme transposée.
Exemple :



$$M_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

ou

$M_0 = (1,0,0,0,0)$

A partir de M_0, il y a une transition validée, T1. En franchissant T1, on arrive au marquage M_1, M_1□= (0,1,1,0,0). On note ceci : M_0 (T1 □ M_1.

On remarque que :

M_1 (T2 □ M_2 = (0,0,1,1,0)
M_1 (T3 □ M_3 = (0,1,0,0,1)
M_2 (T3 □ M_4 = (0,0,0,1,1)
M_4 (T4 □ M_0

On notera **<u>*M_0 l'ensemble des marquages accessibles à partir du marquage M_0</u>**. Pour notre exemple, *M_0 = {M_0, M_1, M_2, M_3, M_4}.

A partir du marquage M_0, on peut franchir successivement T1 puis T2. On se retrouve avec le marquage

M_2. T1T2 est une **<u>séquence de franchissement</u>**. On peut la noter S. On peut écrire alors :
M_0 (T1T2 □ M_2 ou bien M_0 (S □ M_2.

Une **<u>place</u>** Pi est **<u>bornée</u>** pour un marquage initial M_0 si pour tout marquage accessible à partir de M_0 le nombre de marques dans Pi est fini.

Un **<u>RdP est borné</u>** pour un marquage initial M_0 si toutes les places sont bornées pour M_0.

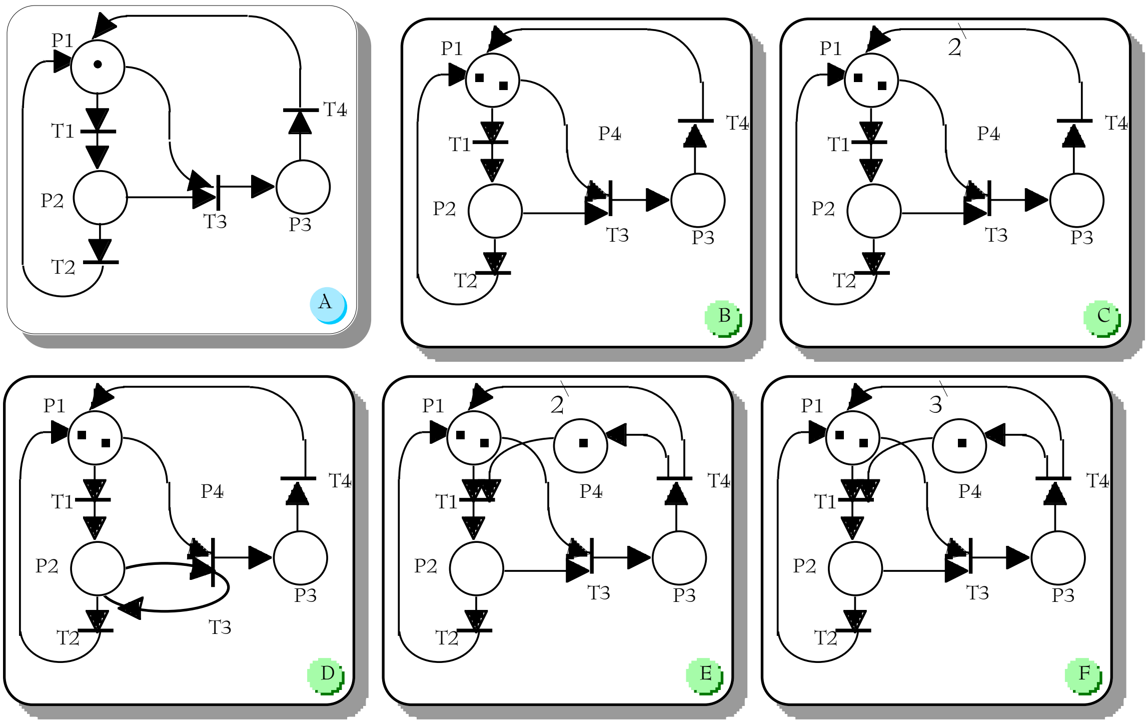
Un RdP est **<u>sauf</u>** pour un marquage initial M_0 s'il y a une marque au plus dans chaque place, pour tout marquage accessible à partir de M_0.

Une **<u>transition</u>** Tj est **<u>vivante</u>** pour un marquage initial M_0 si pour tout marquage accessible M_i□□□*M_0, il existe une séquence de franchissements S qui contienne la transition Tj, à partir de M_i. Autrement dit, quelle que soit l'évolution, il existera toujours une possibilité de franchir Tj.

Un **<u>RdP est vivant</u>** pour un marquage initial M_0 si toutes ses transitions sont vivantes.

Un blocage est un marquage tel qu'aucune transition n'est validée. Un RdP est dit sans blocage pour un marquage initial M_0 si aucun marquage accessible n'est un blocage.

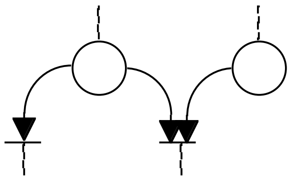
Pour les différents RdP ci-dessous, dire s'ils sont, à votre avis, vivant, sans blocage et bornés. Expliquer.



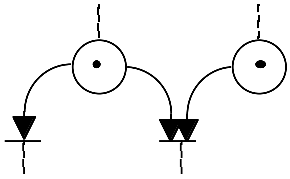
Un RdP a un **<u>état d'accueil</u>** M_a pour un marquage initial M_0 si pour tout marquage accessible M_i□□□*M_0, il existe Si tel que M_i (Si □ M_a.

Un RdP est **<u>réinitialisable</u>** pour un marquage initial M_0 si M_0 est un état d'accueil.

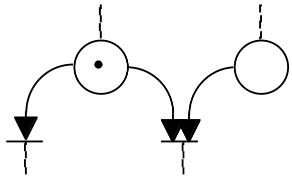
Définitions : On dit qu'il y a **conflit** quand une place a 2 transitions de sortie. On parle de **conflit structurel** car cela ne dépend pas du marquage. Dans certains cas, le franchissement de l'une des transitions peut empêcher le franchissement de l'autre.



Le conflit devient **conflit effectif** quand il y a effectivement conflit. Cela dépend du marquage :



conflit effectif

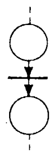
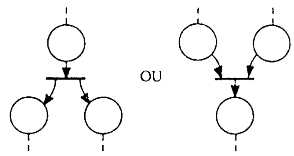
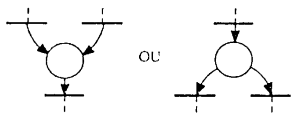
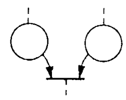
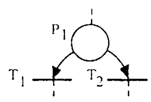
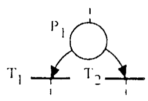
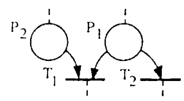
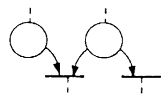
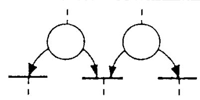
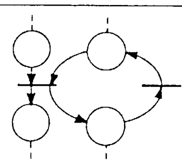
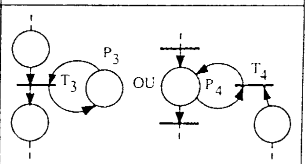


pas de conflit effectif

Si chaque transition ne peut être concernée que par un conflit au plus, , le RdP est **simple**.

Structures particulières :

- Un **graphe d'état** est un RdP telle que toute transition a une place d'entrée et une place de sortie.
- Un **graphe d'événement** est un RdP telle que toute place a exactement une transition d'entrée et une transition de sortie.
- Un RdP est **à choix libre** si pour tout conflit Pi, Tj, Tk, ni Tj ni Tk n'a d'autre place d'entrée. Dans ce cas, tous les conflits structurels sont effectifs.
- Un RdP est **pur** s'il n'existe aucune transition telle que une des places d'entrée soit également place de sortie.

Dans un réseau de Petri qui est ...	On peut trouver ...	On ne peut pas trouver ...
graphe d'états		
graphe d'événements		
sans conflit		
à choix libre		
simple		
pur		

1.6 INVARIANTS

A partir d'un marquage initial, le marquage d'un RdP peut évoluer par franchissement de transitions : et s'il n'y a pas de blocage, le nombre de franchissements de transitions est illimité. Néanmoins, on ne pourra pas atteindre n'importe quel marquage et on ne pourra pas franchir n'importe quelle séquence de transitions. Des invariants permettent de caractériser certaines propriétés des marquages accessibles et des transitions franchissables, quelle que soit l'évolution.

a) Composante conservative.

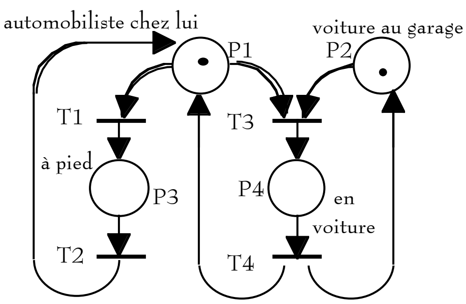
Soit un RdP à n places. On a un **invariant** linéaire de places s'il existe une pondération sur le marquage des places tel que :
$q_1\, m_1 + q_2\, m_2 + \ldots + q_n\, m_n = \text{cste}$,
pour tout marquage accessible $M \in {}^*M_0$, les q_i étant des nombres entiers naturels.

L'ensemble des places P' telles que leur pondération soit non nulle est une **composante conservative**. Le réseau est dit **conservatif** si et seulement si l'ensemble de toutes les places du RdP forme une composante conservative.

La propriété d'être composante conservative est indépendante du marquage initial. Par contre, la constante de l'invariant dépend du marquage initial.

En règle générale, une composante conservative a une signification physique. Elle peut signifier soit qu'un système est dans un seul état à la fois, soit la conservation du nombre d'entités.

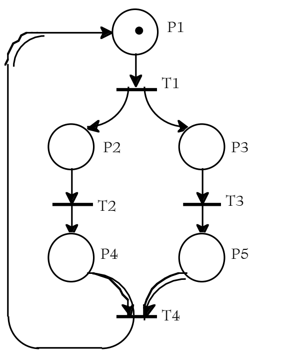
Dans l'exemple suivant :



les composantes conservatives sont :
- {P1, P3, P4} : $m_1 + m_3 + m_4 = 1$ (l'automobiliste est soit chez lui soit a pied soit en voiture)
- {P2, P4} : $m_2 + m_4 = 1$ (la voiture est soit au garage soit avec l'automobiliste)
- {P1, P2, P3, P4} : $m_1 + m_2 + m_3 + 2.m_4 = 2$.

b) Composantes répétitives

On revient à l'exemple :



Les séquences de franchissements qui sont possibles à partir du marquage M_0 sont les suivantes :
T_1, T_1T_2, $T_1T_2T_3$, $T_1T_2T_3T_4$, $T_1T_2T_3T_4T_1$, etc… La séquence $T_1T_2T_3T_4$ est particulière

car
M_0 ($T_1T_2T_3T_4$ $\rightarrow$ M_a. Cette séquence ramène à l'état initial. C'est une **séquence répétitive**. Une séquence répétitive qui contient toutes les transitions est une **séquence répétitive complète**.

Les séquences sont décrites en utilisant des expressions régulières. Les règles d'écriture de ces expressions sont :

T1+T2=T2+T1 T1 ou T2
T1T2 = (T1)(T2) T1 suivi de T2
T1T1 = $T1^2$ T1 2 fois de suite
T1T2T3 est une séquence de longueur 3
ε est une séquence de longueur 0
$T1^* = \varepsilon$+T1+T1T1+T1T1T1+…
T1(T2+T3) = T1T2 + T1T3
(T1+T2)T3 = T1T3 + T2T3

Toutes ces règles restent valables en remplaçant Ti par le symbole d'une expression régulière.

La séquence S1 est un **préfixe** de S2 s'il existe une séquence S3 telle que : S1S3 = S2

Une **séquence répétitive minimale** est une séquence répétitive telle qu'aucun de ses préfixes stricts (de longueur non nulle) n'est une séquence répétitive.

L'ensemble T' des transitions concernées par une séquence répétitive est une **composante répétitive**.

S'il existe une séquence répétitive telle que toutes les transitions du RdP soient concernées, le RdP est dit **répétitif**.

Dans l'exemple de l'automobiliste, il y a deux séquences répétitives minimales : T1T2 ou T3T4. Aucune de ces composantes ne contient toutes les transitions du RdP. Mais à partir de ces composantes, on peut construire d'autres composantes répétitives. L'ensemble de ces composantes peut s'écrire : (T1T2+T3T4)*. On peut construire des séquences répétitives complètes. Le RdP est répétitif.

REMARQUE : Après avoir présenté les propriétés que peuvent posséder certains réseaux, il importe maintenant de savoir comment on va pouvoir trouver si tel réseau présente telle ou telle propriétés. Il existe principalement 3 classes de méthodes pour rechercher les propriétés d'un RdP : le graphe des marquages ou arbre de couverture (voir 1.6), l'algèbre linéaire (1.8) ou des méthodes de réduction.

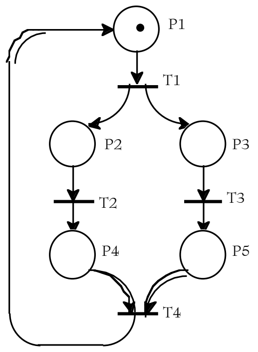
1.7 ALGÈBRE LINÉAIRE

Un **RdP non marqué est un quadruplet Q=<P, T, Pré, Post>** tel que :

- $P = \{P_1, P_2, \dots, P_n\}$ est un ensemble fini et non vide de places ;
- $T = \{T_1, T_2, \dots, T_m\}$ est un ensemble fini et non vide de transitions ;
- Pré : PxT $\rightarrow$ $\{0, 1\}$ est l'application d'incidence avant ;
- Post : PxT $\rightarrow$ $\{0, 1\}$ est l'application d'incidence arrière.

Pré (Pi, Tj) est le poids de l'arc Pi $\rightarrow$ Tj. Ce poids est 1 si cet arc existe et 0 sinon. (Dans le cas d'un RdP généralisé, la valeur de Pré (Pi, Tj) est le poids porté par l'arc, s'il existe, 0 sinon.). De la même façon, Post (Pi, Tj) est le poids de l'arc Tj $\rightarrow$ Pi.

Exemple :



Dans cet exemple,
$Pré(P_1, T_1) = 1$;
$Pré(P_1, T_2) = 0$;
$Post(P_1, T_1) = 0$;
$Post(P_2, T_1) = 1$;

Quelques notations :
$°Tj = \{Pi \in P \mid Pré(Pi, Tj) > 0\}$ = ensemble des places d'entrée de Tj
$Tj° = \{Pi \in P \mid Post(Pi, Tj) > 0\}$ = ensemble des places de sortie de Tj
$°Pi = \{Tj \in T \mid Post(Pi, Tj) > 0\}$ = ensemble des transitions d'entrée de Pi
$Pi° = \{Tj \in T \mid Pré(Pi, Tj) > 0\}$ = ensemble des transitions de sortie de Pi

Un **RdP marqué est un doublet $R = \langle Q, M_0 \rangle$** dans lequel Q est un RdP non marqué et M_0 est un marquage initial.
On peut exprimer la condition de validation d'une transition de la façon suivante : la transition Tj est validée pour un marquage M si et seulement si $M(Pi) \geq Pré(Pi, Tj)$ pour tout $Pi \in °Tj$.

On appelle **matrice d'incidence avant** la matrice :
$W^- = [w^-_{ij}]$, où $w^-_{ij} = Pré(Pi, Tj)$.
On appelle **matrice d'incidence arrière** la matrice :
$W^+ = [w^+_{ij}]$, où $w^+_{ij} = Post(Pi, Tj)$.

Pour l'exemple précédent, on a :

$$W^- = \begin{array}{c} \begin{array}{cccc} T_1 & T_2 & T_3 & T_4 \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right] \end{array} \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{array} \quad \text{et} \quad W^+ = \begin{array}{c} \begin{array}{cccc} T_1 & T_2 & T_3 & T_4 \end{array} \\ \left[\begin{array}{cccc} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{array}$$

On appelle **matrice d'incidence** la matrice $W = W^+ - W^- = [w_{ij}]$. Dans notre exemple :

$$W = \begin{array}{c} \begin{array}{cccc} T_1 & T_2 & T_3 & T_4 \end{array} \\ \left[\begin{array}{cccc} -1 & 0 & 0 & +1 \\ +1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & 0 & -1 \\ 0 & 0 & +1 & -1 \end{array}\right] \end{array} \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{array}$$

Une colonne de cette matrice correspond à la modification du marquage apporté par le franchissement de la transition correspondante. (Ex : en franchissant T_1 si elle est franchissable, on retire une marque dans P_1 et on en ajoute une dans P_2 et une dans P_3).

Équation fondamentale : Si on passe d'un marquage Mi à un marquage Mk par une séquence de franchissement S, cad Mi(S Mk, on a l'équation :
Mk = Mi + W.S
où S est la séquence de franchissement S mise sous forme vectorielle, l'élément j de ce vecteur étant le nombre de fois qu'apparaît la transition Tj dans la séquence.

Exemple : calcul du marquage résultant du tir de T_2 à partir du marquage : 1 marque dans P_2, une marque dans P_3. (Pour le RdP précédent).

$$\underbrace{\begin{bmatrix}0\\1\\1\\0\\0\end{bmatrix}}_{Mi} + \underbrace{\begin{bmatrix}-1 & 0 & 0 & +1\\ +1 & -1 & 0 & 0\\ +1 & 0 & -1 & 0\\ 0 & +1 & 0 & -1\\ 0 & 0 & +1 & -1\end{bmatrix}}_{W} . \underbrace{\begin{bmatrix}0\\1\\0\\0\end{bmatrix}}_{\underline{S}} = \begin{bmatrix}0\\1\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\-1\\0\\+1\\0\end{bmatrix} = \underbrace{\begin{bmatrix}0\\0\\1\\1\\0\end{bmatrix}}_{Mk}$$

Pour le RdP B du paragraphe 1.5, calculer le marquage après la séquence de franchissement : T1T3T4T1T3T4T1T2.

Composante conservative : Soit F un vecteur de pondération des places F = $(q_1, q_2, \dots, q_n)$ tel que chaque q_i est un nombre entier positif ou nul. Soit P(F) l'ensemble des places dont le poids n'est pas nul. P(F) est une composante conservative si et seulement si $\underline{F}^T.\underline{W} = \underline{0}$.
Le vecteur F est appelé un **P-semi-flot**.

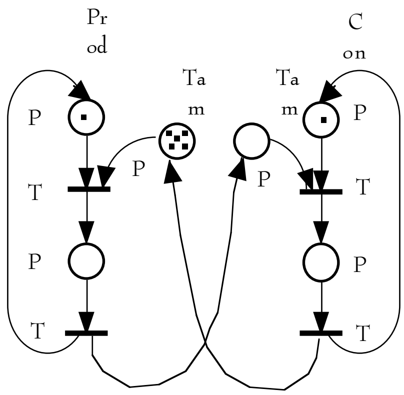
Dans notre exemple, pour F= (1,1,0,1,0), $F^T.W = 0$ donc P(F) = $\{P_1, P_2, P_4\}$ est une composante conservative. L'invariant linéaire de place est : $m_1 + m_2 + m_4 = 1$. Il existe un autre P-semi-flot pour cet exemple : G = (1, 0, 1, 0, 1).

Propriété : Si F_1 et F_2 sont deux P-semi-flots d'un même RdP, alors (a,b) ℕ, $a.F_1 + b.F_2$ est un P-semi-flot.

Dans notre exemple, F+G = (2, 1, 1, 1, 1) est un P-semi-flot. Comme toutes les places sont concernées par ce P-semi-flot, ce RdP est conservatif.

Propriété : Soient P(F) une composante conservative et F = $(q_1, q_2, \dots, q_n)$ le vecteur pondération correspondant. Toutes les places de P(F) sont bornées et l'on a : $M(Pi) \leq (F^T.M_0)/q_i$.

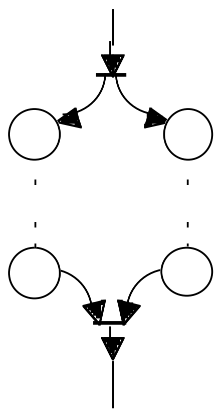
Composantes répétitives : Soit S une séquence de franchissement et T(S) l'ensemble des transitions qui apparaissent dans cette séquence. T(S) est une composante répétitive si et seulement si $\underline{W.S} = \underline{0}$. Le vecteur $\underline{S}$ est appelé un **T-semi-flot**. T(S) est une composante répétitive croissante si $W.\underline{S} > 0$. Remarque : tout T-semi-flot ne correspond pas forcément une composante répétitive car il ne correspond pas nécessairement à une séquence de franchissement.



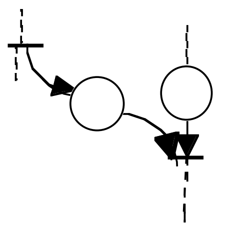
Pour ce RdP, donnez la matrice W, un P-semi-flot et un T-semi-flot, s'il y en a.

1.8 MODÉLISATION PAR RDP

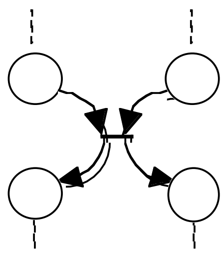
Les RdP permettent de représenter graphiquement certaines relations, de visualiser certaines notions. En voici quelques exemples.



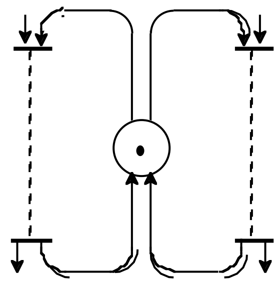
Parallélisme



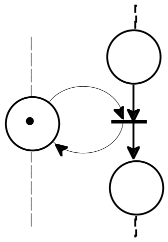
Synchronisation (sémaphore)



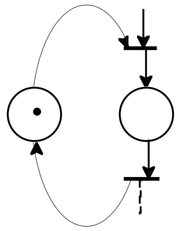
Synchronisation (rendez-vous)



Partage de ressource



Lecture



Capacité limitée

<u>Gestion des cabines et des paniers d'une piscine.</u>

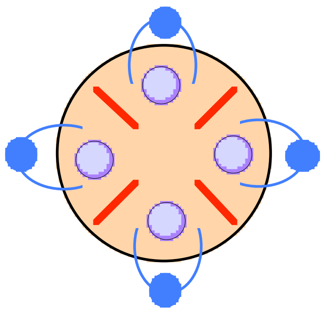
A l'entrée, un client qui a trouvé une cabine libre y entre et se change en posant ses vêtements dans la cabine. Il demande ensuite un panier qu'il remplit pour libérer la cabine. Après la baignade le client rentre dans une cabine avec son panier, le vide et le libère. Ensuite, il se rhabille et libère la cabine.

Modéliser le protocole avec un RdP en prenant comme hypothèse qu'il y a 5 paniers et 3 cabines. Le nombre de clients à la baignade est-il borné? Le RdP est-il borné ? Montrer qu'il y a blocage.

<u>Les philosophes.</u>

Comme l'indique la figure ci-dessous, quatre philosophes (phi1 à phi4) sont autour d'une table, disposant de 4 baguettes (b1 à b4) disposées entre eux. Un philosophe peut avoir essentiellement deux états : il pense ou il mange. Pour manger il a besoin des deux baguettes qui sont de chaque côté de lui. A l'état initial, tous les philosophes pensent et les baguettes sont posées sur la table.

Décrire par un RdP le protocole suivant : lorsqu'un philosophe veut manger, il prend la baguette à sa droite, puis celle à sa gauche et se met à manger. Quand il a fini, il repose la baguette de droite puis celle de gauche. Montrer qu'il y a blocage.



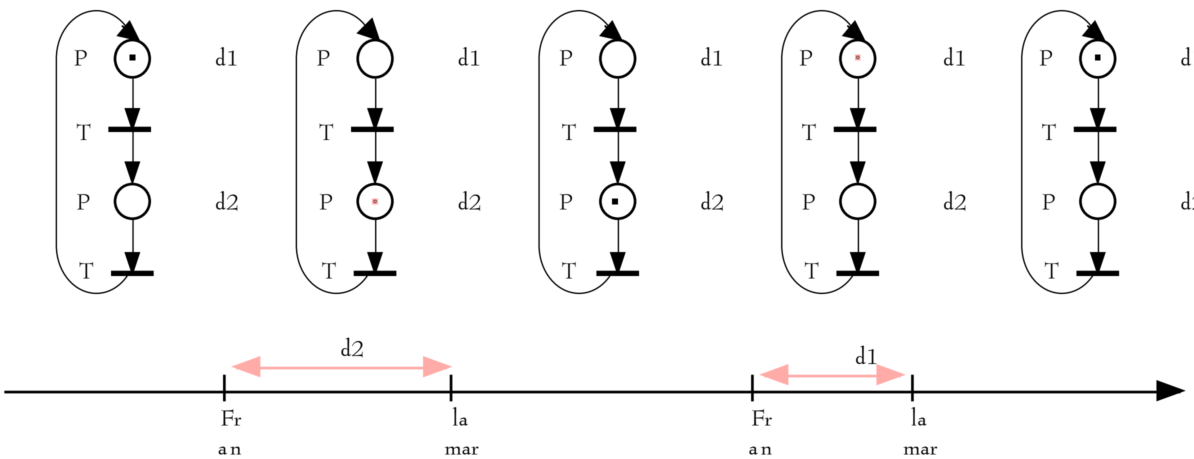
Chapitre 2 : Les réseaux de Petri temporisés

Un RdP temporisé est permet de décrire un système dont le fonctionnement dépend du temps. Les RdP temporisés sont utiles pour l'évaluation des performances d'un système. Soit les temporisations sont associées aux places (RdP P-temporisé) soit aux transitions (RdP T-temporisé).

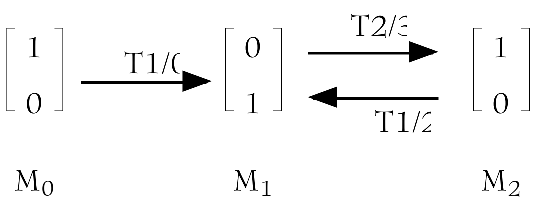
2.1 RÉSEAUX DE PETRI P-TEMPORISÉS

On associe une temporisation (valeur rationnelle positive) à chaque place. On notera di la temporisation de la place Pi.

Principe de fonctionnement : lorsqu'une marque arrive dans une place temporisée, on dit qu'elle est **indisponible** pendant un temps di. Quand le temps est écoulé, la marque devient disponible. Exemple :



On parle de **fonctionnement à vitesse maximale** quand on franchit toute transition dès qu'elle devient franchissable. Dans ce cas de fonctionnement, on peut compléter le graphe des marquages en associant à chaque arc du graphe reliant un marquage Mi à un marquage Mj par la transition Tk le temps séparant l'obtention du marquage Mi du tir de la transition Tk. Pour l'exemple précédent, cela donne :



On remarque qu'après le passage de Mo à M1, on a un fonctionnement périodique.

Propriété : Un RdP borné fonctionnant à vitesse maximale a toujours un comportement périodique au bout d'un temps fini.

Un RdP P-temporisé fonctionne en vitesse propre si toute marque ne reste dans une place que pendant sa durée d'indisponibilité.

La **fréquence de franchissement** Fj d'une transition Tj est le nombre moyen de franchissements d'une transition par unité de temps, lorsque le régime stationnaire est établi.

Calcul des fréquences de franchissement :

Dans notre exemple, le nombre moyen de marques dans la place P1 est : d1.F2 car les marques entrent à une fréquence F2 et y restent un temps d1. De même, dans P2, il y a en moyenne d2.F1 marques. Comme le RdP est conservatif, on a :

$$d1.F2 + d2.F1 = Mo(P1) + Mo(P2).$$

On observe F1 = F2. d1=2 et d2=3. On en déduit F1 = F2 = 1/5.

On aurait pu écrire aussi qu'en régime stationnaire, le nombre de marques qui entrent dans une place est égal au nombre de marques qui en sortent. Ce dernier nombre est égal, pour la place 1 de notre exemple d1.F1. Ce raisonnement conduit à une autre équation :

$$d1.F1 + d2.F2 = Mo(P1) + Mo(P2).$$

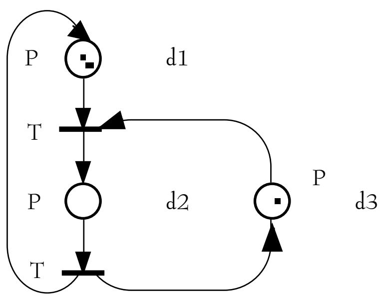
Ce qui donne le même résultat.

De façon générale, on a les relations suivantes :

Une relation liant les temporisations, les fréquences et le marquage initial qui est associée à chaque invariant de marquage. Cette relation est une inéquation car les fréquences réelles peuvent être inférieures à celles qui correspondraient à un fonctionnement à vitesse propre. Cette équation peut s'écrire : $X^T.D.W^+.F \leq X^T.Mo$ où X est un P-semi-flot, D, une matrice diagonale telle que Dii = di, la temporisation associée à la place Pi, F, le vecteur des fréquences de franchissement et Mo, le marquage initial.

Une relation entre les fréquences de franchissement des transitions correspondant à chaque invariant de franchissement.

De ces relations on déduit les fréquences de franchissement correspondant au fonctionnement à vitesse maximale (quand le problème est soluble). Ceci permet d'évaluer certaines performances de systèmes, le franchissement d'une transition pouvant correspondre à l'accomplissement d'une tâche, et le marquage moyen d'une place à un nombre moyen de clients en attente d'un service.



Pour cet exemple, donner les fréquences maximales de franchissement des deux transitions.

2.2 RÉSEAUX DE PETRI T-TEMPORISÉS

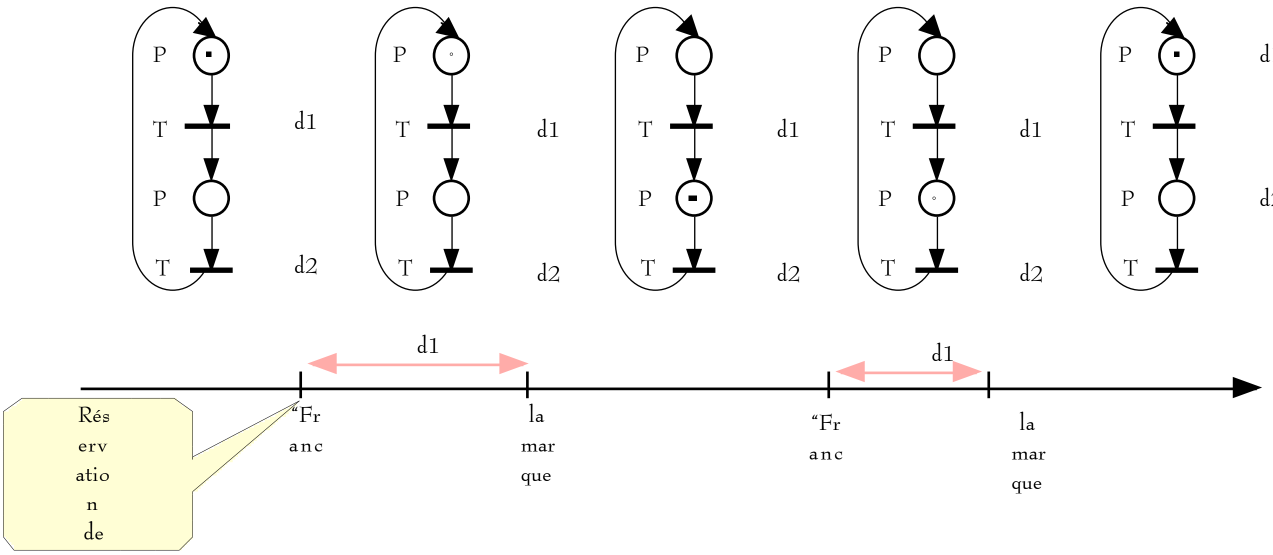
On associe cette fois la temporisation aux transitions.

<u>Principe de fonctionnement</u>

Une marque peut avoir deux états : -

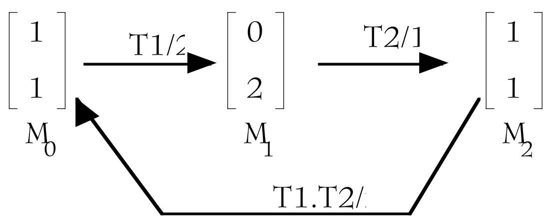
- disponible (ou non réservée)
- ou bien réservée pour le franchissement d'une transition.

Exemple de fonctionnement :



Remarque : on peut toujours passer d'un RdP T-temporisé à un RdP P-temporisé.

On peut, de la même façon que pour les RdP P-temporisés, le fonctionnement à vitesse maximale et le fonctionnement à vitesse propre. Dans le cas de l'exemple précédent, avec 1 marque dans la place P1 et 1 marque dans P2 au départ, on obtient le graphe des marquages suivant :



De même que pour les RdP P-temporisés, le calcul des fréquences maximales de franchissement se fait en égalant le nombre moyen de marques dans une place Pi au produit de la fréquence de sortie Tj par dj (ou la somme des Tj.dj s'il y a plusieurs transitions de sortie de la place Pi).

Processus stochastiques

Généralités :

Un processus stochastique est une famille de variables aléatoires X(t) à valeur réelles où t est un paramètre réel. L'espace des paramètres ou espace temps T correspondant prendra essentiellement l'une des deux formes suivantes :

- $T = \{0, 1, 2, \ldots\}$; on parlera de processus stochastique à temps discret et on écrira Xn au lieu de X(t). Ce sera notamment le cas des chaînes de Markov à temps discret.
- $T = [0, \infty)$; on dira alors que $\{X(t)\ ;\ t \geq 0\}$ est un processus stochastique à temps continu. Ce sera le cas de tous les autres processus étudiés dans ce cours.

On appellera espace des états l'ensemble S des valeurs prises par toutes les variables aléatoires d'un processus stochastique. Nous nous limiterons aux processus stochastiques à états discrets ; S sera identifié à un sous ensemble $\{1, 2, \ldots\}$.

La forme prise par un processus stochastique lors d'une expérimentation du phénomène aléatoire en question est appelé réalisation ou trajectoire de ce processus.

Voici quelques exemples de phénomènes physiques susceptibles d'être modélisé par des processus stochastiques.

- Le nombre d' appels arrivant dans un central téléphonique pendant un intervalle de temps [0, t] ;
- Le nombre de défaillances se produisant par jour dans un système technique
- La fortune du joueur après avoir joué n parties.
- Le nombre de clients dans une file d'attente à un instant donné.

Partie A : Les chaînes de Markov

Une chaîne de Markov (CM) est un cas particulier de processus stochastique. Il en existe deux types : les CM à temps discret et les CM à temps continu. Dans ce dernier cas, on les appelle aussi processus de Markov.

1. Chaînes de Markov à temps discret

1.1 Introduction-Définition

Une grenouille (Zora) est sur un étang comprenant 4 nénuphars (numérotés de 1 à 4). La grenouille saute de nénupar en nénuphar. Le nénuphar destination du saut ne dépend que du nénuphar de départ.

La grenouille n'a pas de mémoire. Si Xn représente le numéro du nénuphar où est la grenouille après n sauts, on peut définir : $p_{ij}(n) = p(X_{n+1} = j \mid X_n = i)$. C'est la probabilité conditionnelle que la grenouille fasse un saut du nénuphar i au nénuphar j à la date n. Ceci signifie implicitement que ces probabilités conditionnelles ne dépendent pas du comportement antérieur à l'instant n de la grenouille :

$$p_{ij}(n) = p(X_{n+1} = j \mid X_n = i) = \begin{array}{l} p(X_{n+1} = j \mid X_n = i, X_{n-1} = 1) \\ p(X_{n+1} = j \mid X_n = i, X_{n-1} = 2) \\ p(X_{n+1} = j \mid X_n = i, X_{n-1} = 3) \\ p(X_{n+1} = j \mid X_n = i, X_{n-1} = 4) \end{array}$$

La classe des processus verifiant cette propriété est caractérisée par le fait que l'état présent du processus, cad son ***état*** à l'instant n, résume toute l'information nécessaire pour connaître son évolution future. En d'autres termes, la prévision de cette dernière ne peut être améliorée par une connaissance supplémentaire du passé du processus, c'est à dire par la connaissance de ses états aux instants ≤ n-1.

Cette propriété (sans mémoire) est connue sous le nom de propriété de Markov. Elle s'exprime par :

$$p(X_{n+1} = j \mid X_n = i) = p(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0=i_0) \text{ pour } n\geq 0 \text{ et } j, i, i_{n-1}, \ldots, i_0 \in S$$

Une chaîne de Markov à temps discret est un processus stochastique qui vérifie cette propriété.

Nous ne nous intéresserons qu'au chaînes de Markov homogènes, c'est à dire pour lesquels les probabilités de transition sont indépendantes du temps ($p_{ij}(n) = p_{ij}$).

1.2 Matrice et graphe des transitions

La matrice de transition d'une CM à temps discret est la matrice P composée des p_{ij}, probabilités de transition. Cette matrice est carrée, de dimension le nombre d'états possibles. Tous les termes sont positifs ou nuls et inférieurs ou égaux à 1 (puisqu'elle ne contient que des probabilités). La somme des termes de chaque ligne est 1 (puisqu'il y a toujours 1 état de destination).

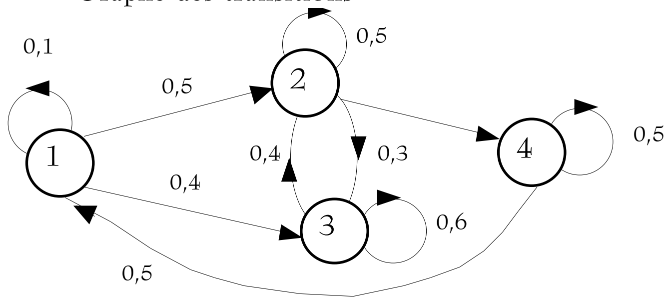
Le graphe des transitions est formé de points représentant les états du processus et d'arcs correspondant aux transitions possibles, cad pour lesquelles les probabilités p_{ij} sont non nulles.

Exemple de Zora :

Matrice de trannsition :

$$P = \begin{bmatrix} 0,1 & 0,5 & 0,4 & 0 \\ 0 & 0,5 & 0,3 & 0,2 \\ 0 & 0,4 & 0,6 & 0 \\ 0,5 & 0 & 0 & 0,5 \end{bmatrix}$$

Graphe des transitions



TD A.1: fiabilité de deux éléments en parallèle

Soit un dispositif technique comprenant deux éléments montés en parallèle et fonctionnant indépendamment l'un de l'autre. Chaque élément a une fiabilité égale à p au cours d'une journée (cad qu'il a une probabilité de 1-p de tomber en panne). Il n'y a pas de possibilité de réparation. Si Xn est le nombre d'éléments en panne au début de la n-ième journée, décrire la chaîne de Markov correspondante, sa matrice et son graphe de transition.

Modifier la chaîne précédente en stipulant qu'une machine en panne sera réparée au cours de la journée suivante.

La suite du cours nous donnera la possibilité de répondre aux questions du type :
- Quelle est la distribution du nombre d'éléments en panne après 1, 2, 3, …, n jours?
- Quelle est la durée de bon fonctionnement de ce dispositif?

1.3 Loi de probabilité de Xn

Soit $\pi_k(n) = P(Xn=k)$, pour $n \geq 0$ et $k \in S$. C'est la probabilité que le processus soit dans l'état k à l'instant n. On peut définir ainsi un vecteur ligne $\pi(n) = [\pi_1(n), \pi_2(n), \ldots]$ dont la somme des termes vaut 1. On a la relation suivante :

$$\pi(n+1) = \pi(n).P \quad \text{où P est la matrice de transition.}$$

On peut en déduire :

$$\pi(n) = \pi(0).P^n$$

Dans l'exercice précédent (deux éléments en parallèle, sans réparation), calculer la probabilité π en fonction de n.
AN : p=0,9.

1.4 Comportement asymptotique

Nous avons pu calculer $\pi(n)$, les probabilités d'état en fonction de n. Ces dernières dépendent de $\pi(0)$. En fait, nous avons étudié le régime transitoire du processus.

On constate souvent que la distribution $\pi(n)$ converge vers une distribution limite quand $n \to \infty$. Quand c'est la cas, cette dernière défini le régime permanent du processus. Ce régime permanent n'est pas influencé par le choix de la distributuion initiale. On admet que le régime permanent est atteint au bout d'un nombre fini de transitions.

Existe-t-il un régime permanant pour la chaîne de Makov définie par :

1.

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

2.

$$P = \begin{bmatrix} 0,2 & 0,8 \\ 0 & 1 \end{bmatrix}$$

3.

$$P = \begin{bmatrix} 0,2 & 0,8 \\ 0,1 & 0,9 \end{bmatrix}$$

<u>Théorème</u> : Si la matrice de transition P est telle qu'au moins une de ses puissances n'a que des termes strictement positifs, alors, quelle que soit la distribution initiale □(0), quand n tend vers l'infini :

$$□(n) \rightarrow □$$

$$P^n \rightarrow P^*$$

p est un vecteur de probabilité strictement positif et P*, une matrice dont toutes les lignes sont identiques au vecteur limite □.

<u>Théorème</u> : Si la valeur propre 1 de P est simple et si toute autre valeur propre de P est de module strictement inférieur à 1, alors les conclusions du théorème précédent sont les mêmes.

Une distribution de probabilité discrète □ est appelée <u>stationnaire</u> par rapport à une matrice stochastique□P si : □.P = □.

Pour trouver les composantes du vecteur de distribution stationnaire, il existe 2 méthodes :

- résoudre le système :

$$□.P = □$$
$$\sum_{k □ S} □_k = 1$$

- en résolvant les <u>équations de balance</u> : on interprète les probabilité $□_k$ comme des masses associées aux états k et le produits $□_k p_{kj}$ comme des flux de masse entre les deux états k et j. La répartition des masses $□_k$ est stationnaire si, lors d'une transition, le flux d'entrée et égal au flux de sortie pour chacun des états

<u>Théorème</u> : Si □ est la distribution limite d'une chaîne de Markov, alors □ est l'unique distribution stationnaire de cette chaîne.

TD A.2 : Trouver les distributions limite des chaînes suivantes

1. Montrer que le chaîne de Markov définie par P converge et calculer la distribution limite

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

2.Soit la matrice :

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

a) Dessiner le graphe des transitions correspondant.
b) Calculer les valeurs propres de P. La chaîne est-elle convergente ?
c)Calculer la distribution stationnaire de ce processus.

1.5 Chaînes de Markov absorbantes

Un état k d'une chaîne est dit absorbant si le processus ne peut plus quitter cet état une fois qu'il y est entré. En d'autres termes, $p_{kk} = 1$. Une chaîne de Markov est dite absorbante s'il existe au moins un état absorbant et s'il on peut passer de n'importe quel état à un état absorbant.

Lorsqu'on a affaire à une chaîne de Markov absorbante, on est généralement intéressé par les deux questions uivantes :

- Combien de temps faudra-t-il pour que le processus soit absorbé, étant donné son état initial ? On appellera n_i le temps moyen jusqu'à l'absorption en partant de i.
- S'il existe plusieurs états absorbants, quelle est la probabilité pour un processus d'être absorbé par un état donné. On appellera b_{ij} la probabilité que le processus soit absorbé dans j si son état initial est i.

Théorème : Les quantités n_i sont solution du système d'équations :

$$n_i = 1 + \sum_{k \in S'} p_{ik} n_k$$

où i est un état non absorbant et S' l'ensemble de tous les états non absorbants

Théorème : Soit j un état absorbant et S' l'ensemble de tous les état non absorbants. Alors les probabilités b_{ij} ($i \in S'$) sont solution du système d'équations :

$$b_{ij} = p_{ij} + \sum_{k \in S'} p_{ik} b_{kj}$$

TD A.3

1 : calcul de durée de vie d'une pièce

Une certaine pièce d'équipement électrique peut se trouver dans 3 états :
1: bonne ; 2 : condition marginale ; 3 : défaillante.
A la fin de chaque jour de service, l'état de la pièce est enregistré. La matrice de transition obtenue est :

$$P = \begin{bmatrix} 0{,}7 & 0{,}2 & 0{,}1 \\ 0{,}5 & 0{,}3 & 0{,}2 \\ 0 & 0 & 1 \end{bmatrix}$$

Calculer la durée de vie moyenne d'une pièce se trouvant initialement en bonne condition.

2. promenade du scarabée

Soit un tétrahèdre régulier dont les sommets sont numérotés de 1 à 4. Un scarabée qui se trouve initialement sur le sommet 1 se déplace pendant chaque unité de temps le long d'une arête. Arrivé en un sommet, ilchoisit avec la même probabilité l'une des trois arêtes pour continuer sa promenade.

- quel est le temps moyen pour atteindre pour la première fois le sommet 4 ?
- Quelle est la probabilité d'atteindre 4 s'il y a un mangeur de scarabée en 2

2. Chaîne de Markov à temps continu

2.1 La loi exponentielle

La loi exponentielle joue un rôle fondamental dans la suite du cours. Considérons la durée T de bon fonctionnement d'un dispositif technique et admettons que T soit une loi exponentielle de taux λ. La fonction densité de probabilité de cette loi est :

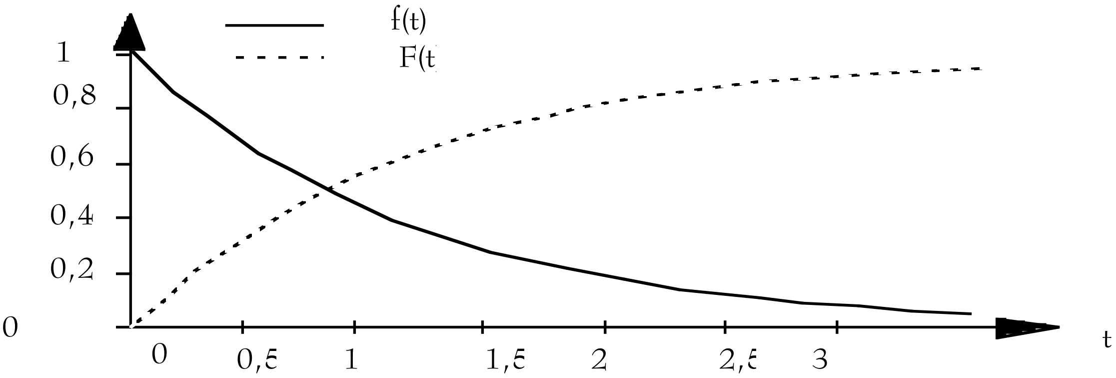
$$f(t) = \lambda e^{-\lambda t} \qquad (t \geq 0)$$

Sa fonction de répartition est :

$$F(t) = P(T>t) = e^{-\lambda t}$$

Si un temps aléatoire suit une loi exponentielle de taux λ, le temps moyen sera : $1/\lambda$.

La figure ci-dessous représente f(t) et F(t) pour $\lambda = 1$.



Propriétés :

- La loi exponentielle est sans mémoire. Cela signifie que la probabilité de bon fonctionnement pendant un intervalle (u, u+t] dépend uniquement de la longueur t de cet intervalle et non pas de sa position relative à l'axe temporel :
$$P(T > t+u \mid T \geq u) = P(T > t) = e^{-\lambda t}$$
La loi exponentielle est la seule loi qui possède cette propriété.

- La probabilité que ce dispositif tombe en panne dans l'intervalle (t, t+dt] est λ.dt.
- Soit $T_1, T_2, T_3, \ldots, T_n$ des variables aléatoires indépendantes distribuées selon des lois exponentielles de paramètres $\lambda_1, \lambda_2, \ldots, \lambda_n$. Alors $T = \min(T_1, T_2, T_3, \ldots, T_n)$ soit une loi

exponentielle de paramètre $\lambda_1+\lambda_2+\ldots+\lambda_n$.

2.2 Définitions

Définition : Une chaîne de Markov homogène à temps continu et à nombre fini d'états est un processus stochastique $\{X(t), t \geq 0\}$ tel que :

$$P[X(t+dt) = j \mid X(t) = i] = p_{ij}(dt).$$

Où $p_{ij}(dt)$ est la probabilité d'être dans l'état j après un intervalle de temps dt sachant que l'état courant est i. On peut définir par q_{ij}, le taux de transition de l'état i à l'état j (j≠i).

$$q_{ij} = \lim_{dt \to 0}\left(\frac{p_{ij}(dt)}{dt}\right)$$

Le temps de passage de la chaîne de l'état i à l'état j est un temps aléatoire qui suit une loi exponentielle de taux q_{ij}.

Le temps pendant lequel la chaîne de Markov reste dans l'état i est le plus petit des temps de passage de cet état i à un autre état. D'après la dernière propriété des lois exponentielle, le temps de séjour de la chaîne dans cet état i suit une loi exponentielle de taux (qu'on appellera taux de sortie de l'état i) :

$$q_{ii} = -\sum_{j \neq i} q_{ij}$$

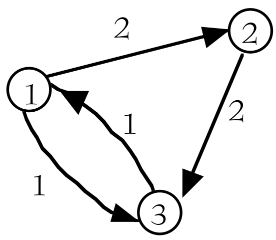
Remarque : On peut considérer que le fonctionnement d'une chaîne de Markov à temps continu a le fonctionnement suivant : dans un état, le processus reste un temps aléatoire qui suit une loi exponentielle de taux $-q_{ii}$. A la sortie de cet état, le processus choisi son état destinantion suivant le probabilités : $-q_{ij}/q_{ii}$.
La matrice Q ainsi construite est appelée matrice de transition ou génératrice de la chaîne de Markov.

On peut associer à chaque chaîne de Markov un graphe orienté construit de la façon suivante : chaque état i du processus est représenté par un noeud. Il y a un arc entre un noeud i et un noeud j ≠i si et seulement si l'élément i, j de la matrice Q est non nul. On donne alors un poids à l'arc i, j : le taux de transition de l'état i à l'état j. C'est l'élément i, j de la matrice Q.

Exemple : Soit la chaîne de Markov (qui pourrait représenter le comportement de Zora) définie par son générateur :

$$Q = \begin{pmatrix} -3 & 2 & 1 \\ 0 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix}$$

Sa représentation graphique est en figure ci dessous.



2.3 Loi de probabilité de x(t)

Si x(t) est le vecteur dont les éléments sont les probabilités de présence du processus dans ses états, on peut écrire :

$$\dot{x}(t)=x(t).Q$$

x(0) est la distribution initiale de probabilité de présence dans les états.

La résolution de l'équation différentielle ci-dessus donne :

$$x(t) = x(0).\ e^{Q.t}$$

Dans la pratique, on ne s'intéressera qu'à deux types de problèmes : le comportement asymptotique de ces chaînes (voir paragraphe suivant) soit les temps d'absorption des chaînes absorbantes (voir paragraphe d'après).

2.4 Comportement asymptotique

On constate souvent souvent que la distribution x(t) converge vers une <u>distribution limite</u> quand $t \to \infty$. Quand c'est la cas, cette dernière défini le <u>régime permanent</u> du processus. Ce régime permanent n'est pas influencé par le choix de la distributuion initiale.

On différentie deux catégories d'états : les états <u>transitoires</u> dont la probabilité de présence du processus dans cet état tend vers 0 quand t tend vers l'infini. Les autres états sont appelés <u>ergodiques</u>.

Une distribution de probabilité □ est appelée <u>stationnaire</u> par rapport à une génératrice Q si :

□.Q = 0.

Pour trouver les composantes du vecteur de distribution stationnaire, il existe 2 méthodes :

- résoudre le système :

  □.Q = 0

  $\sum_{k □ S}$ □$_k$ = 1

- en résolvant les <u>équations de balance</u> : on interprète les probabilité x_k comme des masses associées aux états k et le produits $x_k q_{kj}$ comme des flux de masse entre les deux états k et j. La répartition des masses x_k est stationnaire si, lors d'une transition, le flux d'entrée et égal au flux de sortie pour chacun des états

Si □ est la distribution limite d'une chaîne de Markov, alors □ est l'unique distribution stationnaire de cette chaîne.

2.5 Chaînes de Markov absorbantes

Un état k d'une chaîne est dit absorbant si le processus ne peut plus quitter cet état une fois qu'il y est entré. En d'autres termes, $q_{kk} = 0$. Une chaîne de Markov est dite absorbante s'il existe au moins un état absorbant et s'il on peut passer de n'importe quel état à un état absorbant. Nous ne nous intéresserons qu'aux chaînes de Markov n'ayant qu'un seul état absorbant. (si tel n'est pas le cas, il suffit de 'rassembler' tous les états absorbants en un seul). Le générateur Q peut donc se mettre sous la forme :

$$Q = \left[\begin{array}{c|c} S & v \\ \hline 0\ . & 0 \end{array}\right]$$

où v est un vecteur colonne vérifiant :

$$v = - S.e$$

pour assurer la nullité des sommes par ligne des éléments de Q, e étant un vecteur colonne de dimension appropriée dont tous les éléments valent 1. La distribution initiale est concentrée sur les états transitoires sans perte de généralité :

$$x(0) = (\square , 0)$$

où $\square$ est un vecteur ligne contenant les probabilités initiales de présence dans les états transitoires.

Ce qui nous intéresse dans ces chaînes absorbantes c'est le temps que va mettre le processus pour arriver dans l'état absorbant.

<u>Propriété</u> : La fonction de répartition de cette loi de probabilité est la fonction probabilité de présence du processus dans l'état absorbant en fonction du temps.

En résolvant les équations différentielles d'évolution de la chaîne, on trouve :

$$f(t) = \square.e^{St}.v$$

Il peut être plus commode de calculer la Transformée de Laplace de cette densité de probabilité :

$$F(p) = TL\,(f(t)) = \square.(pI-S)^{-1}.v$$

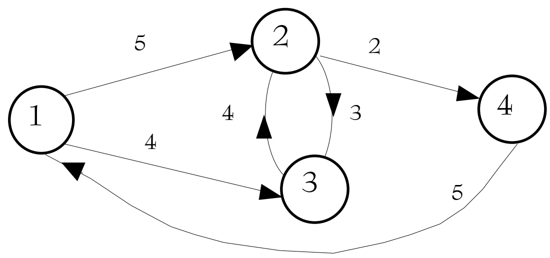
où I est la matrice identité de dimension adéquate.

Le temps moyen d'absorption se calcule alors :

$$T = \int_{t=0}^{\infty} t.f(t)\ dt$$

TD A.4 Promenade de ZORA

Nous nous intéressons à présent au comportement de la grenouille Zora au cours du temps (et non plus aux sauts seulement). Nous observons le comportement modélisé par la chaîne de Markov à temps continu suivante :



1. Trouver la génératrice correspondante. Calculer les probabilités p_{12}, p_{41}, p_{23}, p_{34}.

2. Le taux de sortie de l'état 2 est 5. Cela signifie que le temps de séjour de Zora sur le nénuphar 2 suit une loi exponentielle de taux 5. Quel est le temps de séjour moyen de Zora sur ce nénuphar si l'unité de temps est l'heure.

3. Calculer les probabilités stationnaires de présence de Zora sur chacun des nénuphars.

4. Il y a un héron au dessus du nénuphar 4. Calculer la densité de probabilité du temps que mettra Zora pour rencontrer le héron sachan qu'au départ, Zora est sur le nénuphar 1.

Partie B : Les Réseaux de files d'attente

1. Introduction

Les Réseaux de Files d'Attente (RFA) permettent de modéliser et d'analyser les systèmes de type clients/serveurs.

Un peu d'histoire :

Ils ont été introduits pour l'étude des premiers systèmes téléphoniques, pour connaitre le taux d'occupation des lignes et des standards. Plus tard, ils ont servi en recherche opérationnelle et en fiabilité. Plus tard encore, en informatique, ils ont permis de prévoir le taux d'utilisation d'un serveur ou d'un réseau. En ce qui concerne les systèmes de production, il est clair que l'on peut aisément considérer les pièces comme des clients et les machines comme des serveurs. Les applications sont alors nombreuses : taux d'occupation machine, temps de séjour d'une pièce dans le système de production, nombre de pièces dans les stocks intermédiaires. Voir figure 1.

2. Présentation d'une station

Une station est un système où les clients arrivent pour recevoir un service. Si le serveur ou les serveurs sont occupés, ils attendent leur tour.
Schema d'une station :

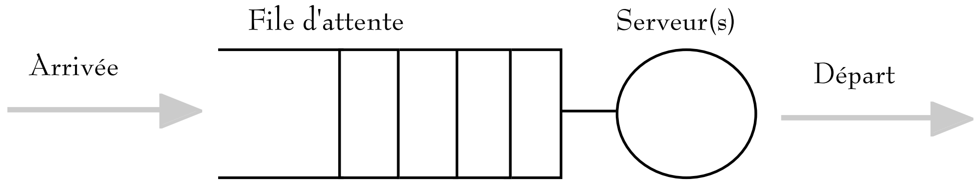


Figure 2 : représentation d'une station

Définitions :

La file d'attente : Elle peut être finie ou infinie. Dans ce dernier cas, elle peut modéliser un stock à capacité illimitée. Quand un client essaye d'entrer dans une station dont la file d'attente limitée est pleine, on considère que le client est perdu.

Le ou les serveurs : On parle d'un **monoserveur** lorsqu'il y a un serveur qui traite les clients les uns après les autres. C'est le cas des guichets par exemple. On parle de **multiserveur** quand plusieurs clients sont servis en même temps, parcequ'il y a plusieurs serveurs. C'est le cas d'un rayon fromage dans les supermarchés par exemple. La notion de **serveur infini** est plus abstraite. Il correspond au cas où tous les clients qui arrivent peuvent être servis immédiatement. Cela suppose que la file d'attente derrière un tel serveur n'est pas utile. C'est le cas des rayons boîtes de conserve dans les supermarchés par exemple. Certains convoyeurs peuvent également être considérés comme tels.

Discipline de service : Ordre dans lequel les clients dans la file seront retirés pour être servis. La discipline par défaut est PAPS (Premier Arrivé, Premier Servi) ou FIFO (First In, First Out). Si une autre discipline est utilisée (comme DAPS ou LIFO ou bien aléatoirement ou autre), il faut la préciser.

Processus d'arrivée : décrit le temps entre deux arrivées successives de clients. Ce temps peut être déterministe (il faut donner sa valeur) ou bien aléatoire (il faut alors préciser la loi : loi de Poisson, d'Erlang ou autre et les paramètres qui permettent de la définir).

Processus de service : décrit le temps que met un serveur pour traîter un client. Mêmes remarques que ci-dessus.

Notation de Kendall

Pour préciser en 2 lettres et 2 nombres les principaux paramètres de la file d'attente : **A/B/C/K.**

A et **B** : Processus d'arrivée et Processus de service

- M : loi de Poisson (nous en verrons une définition au chapitre suivant) ou loi exponentielle
- D : Déterministe
- G : générale

C : le nombre de serveurs

K : La capacité de la file (omise si infinie)

Exemples :

- Une caisse de supermarché dont l'arrivée de clients suit une loi de poisson et telle que le temps de service (temps de traitement d'un client) suive une loi exponentielle peut être considérée comme une station M/M/1
- Un salon de Coiffure dont l'arrivée des clients suit une loi de poisson et telle que le temps de service (temps de coiffage d'un client) suive une loi gérérale (à préciser) qui comprend 2 coiffeurs et 4 places assises pour attendre peut être considéré comme une station M/G/2/4
- Un convoyeur à vitesse constante et qui n'a qu'une entrée et une sortie et dont l'arrivée des clients suit une loi de poisson peut être considéré comme une station M/D/∞
- Un petit train touristique dont l'arrivée des touristes suit une loi de poisson et telle que le temps de service (temps mis par le train pour faire le tour de la ville) est constant et qui a 100 places assises ne peut pas être considéré comme une station M/D/100.

  Expliquez pourquoi.

3. Etude d'une station

On peut vouloir étudier une station en isolation. Bien sûr, lorque les processus d'arrivée et de service sont déterministes, le comportement de la station est également déterministe et une étude est inutile. Par contre, lorsque ces processus sont aléatoires, selon s'il y a un ou plusieurs serveurs, si la file est à capacité limitée ou non, il devient difficile d'intuiter le nombre moyen de clients en attente par exemple.

Performances souhaitées

L'étude d'une station doit permettre de connaître son comportement et plus particulièrement, certaines grandeurs qui semblent importantes. Certaines d'entre elles sont indépendantes des processus d'arrivée/de service, elles dépendent seulement de leur moyenne :

- débit des clients : **X**
- taux d'utilisation du serveur

D'autres, dépendent des processus d'arrivée/de service :

- nombre moyen de clients dans la station : **Q** (clients en attente + clients en train d'être servis)
- temps de réponse du client (entre l'entrée et la sortie de la station) : **W** (temps d'attente + temps de service)

Une loi relie certaines de ces grandeurs, la loi de Little :

$$\mathbf{Q = W.X}$$

Pour connaître ces grandeurs, il faut et il suffit de calculer :

P(n) = probabilité d'avoir n clients dans la station, n □ □.

Etudier une station c'est trouver P(n), pour tout n. Dans le cadre de méthodes analytiques, on trouvera P(n) par calcul. On peut aussi le trouver par simulation. Une fois qu'on a P(n), pour tout n, on peut connaître :

$$Q = \sum_{n=0}^{\infty} n\, P(n)$$

car Q est la moyenne du nombre de clients dans la station. Le débit des clients vaut, s'il n'y a qu'<u>1</u> serveur :

$$X = \square.(1 - P(0))$$

où □.est le débit du serveur lorsque celui-ci est occupé. Si la file d'attente est illimitée et que l'on appelle □ le débit d'entrée des clients et si □ < □ alors X = □ en régimle permanent. On peut alors connaître le temps de passage myen du client dans la station :

$$W = Q / X$$

Etude de la M/M/1

Nous allons particulièrement étudier une station dont les processus d'arrivée des clients et de service sont repectivement un processus de poisson et une loi exponentielle. Ces processus dits markoviens sont particulièrement pratiques dans le cadre d'un étude analytique car le comportement de la station peut être décrit par un processus Markovien "de naissance et de mort".

<u>**Définition**</u> : On dit qu'un <u>processus de comptage</u> (d'événements arrivant aléatoirement) $\{N(t)\}$ est un <u>**processus de Poisson**</u> s'il satisfait aux trois conditions suivantes :

- Le processus N(t) est homogène dans le temps. Ceci veut dire que la probabilité d'avoir k événements dans un intervalle de longueur donnée t ne dépend que de t et non pas de la position de l'intervalle par rapport à l'axe temporel.
- Le processus N(t) est à accroissements indépendants ce qui signifie que pour tout système d'intervalles disjoints, les nombres d'événements s'y produisant sont des variables aléatoires indépendantes.
- La probabilité que deux événements ou plus se produisent dans un petit intervalle dt est négligeable par rapport à la probabilité qu'il n'y ait qu'un seul événements. En termes plus précis : $p_k(t) = o(dt)$ si $k \geq 2$, $p_k(t) =$ □$.dt + o(dt)$ pour k=1 et $p_k(t) = 1-$□$.dt + o(t)$ pour k=0.

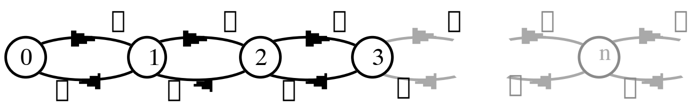
Le coefficient □ est appelé densité ou intensité du processus poissonien.

<u>**Théorème**</u> : Le coefficient □ est égal au nombre moyen d'événements par unité de temps.

<u>**Théorème**</u> : Un processus de comptage est un processus de Poisson de paramètre □ si les intervalles de temps entre deux événements consécutifs sont des varaibles aléatoires indépendantes obéissant à la même loi exponentielle de paramètre □.

Il y a donc un lien étroit entre processus de Poisson et loi exponentielle. C'est pourquoi ils sont représentés, dans la notation de Kendall par la même lettre : M comme Markovien. On parlera de processus de poisson pour une loi d'arrivée de clients parceque c'est un processus de comptage ou de renouvellement et on parlera de temps suivant une loi exponentielle pour les processus de service.

Lorsque les processus d'arrivée et de service sont markoviens, on peut parler d'état de la station. Ces états sont : 0 client dans la station, 1 clients, 2 clietns, etc... On passe de l'état i à l'état i+1 si un client arrive c'est à dire en suivant une loi exponentielle de taux □. On passe de l'état i à l'état i-1 si un client est servi c'est à dire en suivant une loi exponentielle de taux □ (□ est le taux de service du serveur). Pour qu'un tel système ne diverge pas (nombre de clients fini), il est indispensable que □ ≤ □.Voici une représentation shématique de ce processus :



<u>Figure 4 : processus de naissance et de mort</u>

Il nous faut maintenant chercher P(n), la probabilité d'avoir n clients dans la station. C'est aussi la probabilité d'être dans l'état n du processus markovien ci-dessus. On peut écrire les équations d'évolution du processus :

$$\frac{\partial P(0)}{\partial t} = \square\, P(1) - \square P(0)$$
$$\frac{\partial P(1)}{\partial t} = \square P(0) + \square P(2) - (\square + \square)\, P(1)$$

puis dire qu'en régime stationnaire, la partie gauche des équations est nulle. Ce qui nous donne, pour tout n :

$$P(n) = \left(\frac{\lambda}{\mu}\right)^n . P(0)$$

Pour calculer P(0), il suffit de rappeler que les P(i) sont les probabilités d'être dans un état et que :

$$\sum_{i=0}^{\infty} p(i) = 1$$

D'où, dans le cas de la M/M/1 : $P(0) = 1 - \rho$ avec : $\rho = \lambda/\mu$ $(\rho < 1)$

Nous pouvons alors connaître les performances d'une telle station :

Débit : $X = \mu\,(1 - P(0)) = \lambda$

Nombre moyen de clients dans la station :

$$Q = \sum_{n=0}^{\infty} n.P(n) = \frac{\rho}{1-\rho}$$

Nombre moyen de clients en attente (nombre de clients dans la station - clients en train d'être servis) :

$$Q - \rho$$

Temps de réponse moyen :

$$W = \frac{Q}{X} = \frac{1}{\mu-\lambda}$$

Temps d'attente (temps de réponse - temps de service) :

$$W - \frac{1}{\mu}$$

Ces résultats ne sont valables que pour une M/M/1.

Analyse d'une file unique M/M/1

Etude de l'attente dans un organisme public:

Un organisme public est ouvert, chaque jour ouvrable, de 9h à 17h sans interruption. Il accueille en moyenne 64 usagers par jour. Un guichet unique sert à traiter le dossier de chaque usager, ceci dans un temps moyen de deux minutes et demie. Les usagers font la queue dans l'ordre de leur arrivée. Même si la queue est importante, on ne refuse aucun usager.

Une étude statistique a permis de conclure que la durée aléatoire des services suit une loi exponentielle et que les arrivées des usagers forment un processus de Poisson. On suppose que le régime permanent est rapidement atteint.

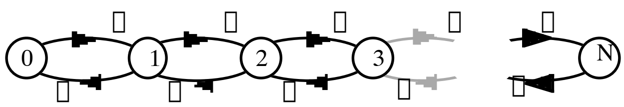
1. Donner la notation de Kendall de cette file d'attente, le temps moyen passé à attendre et le temps moyen passé dans l'organisme par chaque usager.
2. Quelle est, en moyenne et par heure, la durée pendant laquelle l'employé du guichet ne s'occupe pas des usagers ?

3. Quelle est la probabilité d'observer une file d'attente de plus de k usagers derrière celui en cours de service ?

Extensions de la M/M/1

la M/M/1/N

Cette fois, on limite la capacité de la station (serveur + attente) à N clients. Le processus de naissance et de mort équivalent se transforme en :



On a toujours :

$$P(n) = \left(\frac{\lambda}{\mu}\right)^n . P(0)$$

Cette fois, pour calculer P(0), on utilise :

$$\sum_{i=0}^{N} p(i) = 1$$

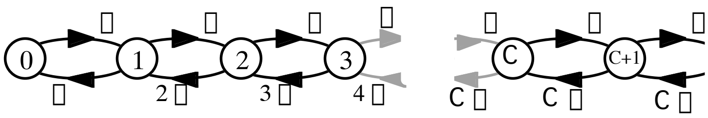
D'où :

$$P(0) = \frac{1 - \rho}{1 - \rho^{N+1}}$$

On peut alors calculer les performances souhaitées : Q, W, X.

la M/M/C

La capacité de la station n'est plus limitée mais il y a C serveurs en parallèle. Chacun de ces serveurs a un taux μ (chaque serveur met en moyenne un temps de 1/μ pour servir un client).
La nouvelle condition de stabilité est $\lambda < C.\mu$
Cette station se comporte comme le processus de naissance et de mort suivant :



Etude des performances de deux machines en parallèle:
On dispose de deux machines en parallèle qui ont un stock d'entrée commun et on envisage de les remplacer par une machine deux fois plus rapide. Les pièces arrivent suivant un processus de Poisson de taux λ, et le temps de service des machines est distribué exponentiellement (taux μ pour la machine unique et μ/2 pour les deux machines en parallèle).
Comparer les performances de la machine unique à celles des deux machines en parallèle.

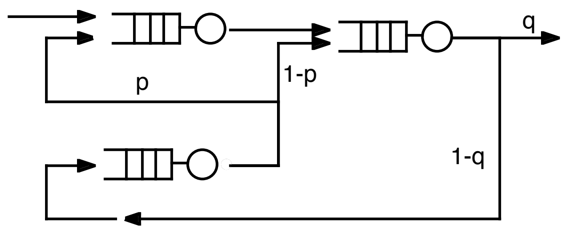
4. Modéliser par un Réseau de files d'attente (RFA)

Un RFA est un réseau de stations interconnectées. Il peut décrire un processus comprenant plusieurs serveurs à travers desquels circulent des clients. Par exemple, dans un supermarché, certains clients passent de la station "fromage" à la station "fruits et légumes" pour finir aux "caisses". On distingue plusieurs types de RFA :

RFA Ouvert et RFA Fermé

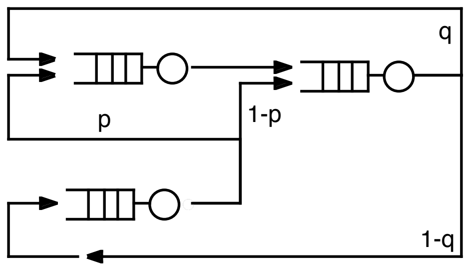
RFA ouvert : Les clients arrivent de l'extérieur, circulent dans le réseau à travers différentes stations, puis quittent le réseau. Pour décrire un tel réseau, il faut préciser :
- la description de chaque station (hormis le processus d'arrivé qui cette fois est fixé par le réseau lui-même),
- le processus d'arrivée des clients dans le réseau,
- le cheminement des clients dans le réseau (routage).

Exemple :



RFA fermé : Il n'y a ni arrivée ni départ de clients. Leur nombre est donc constant.Pour décrire un tel réseau, il faut préciser :
- la description de chaque station (hormis le processus d'arrivé qui est fixé par le réseau),
- le nombre total de clients dans le réseau,
- le cheminement des clients dans le réseau (routage).

Exemple :



RFA mono ou multi classe

Il peut circuler dans un réseaux plusieurs classes de clients. Les clients d'une même classe ont tous un comportement identique (temps de service et cheminement identique). Pour décrire un réseau multiclasse, il faut décrire
- chaque station pour chaque classe de clients,

- le cheminement pour chaque classe de clients,
- le processus d'arrivée pour chaque classe de clients (pour les réseaux ouverts) ou,
- le nombre total de clients pour chaque classe de clients (pour les réseaux fermés).

Remarques :

Il existe même des réseaux multiclasses mixtes : ouvert pour certaines classes et fermées pour les autres.

Un client peut changer de classe. (On définit alors une chaîne : l'ensemble des classes auxquelles peut appartenir un même client.)

Exemples de modélisation parun réseau de files d'attente

1. Modélisation d'une station de ski:

On considère une petite station de ski qui contient 3 pistes et 3 téléskis permettant l'accès à ces 3 pistes. Avant d'aller skier, les skieurs doivent obligatoirement passer à l'une des deux caisses pour acheter un forfait qui les autorisera à accéder aux pistes. Ils prennent alors le téléski 1 (en faisant éventuellement la queue dans l'ordre d'arrivée) pour se rendre en haut de la piste 1. Arrivés là haut, ils peuvent descendre la piste 1 et reprendre le même téléski (on suppose que le temps moyen passé sur une piste par un skieur est indépendant du nombre de skieurs sur la piste). Ils ont également la possibilité de prendre l'un ou l'autre des téléskis 2 et 3 qui les emmèneront plus haut et leur donneront accès aux pistes 2 et 3, respectivement. Arrivés en bas des pistes 2 ou 3, ils peuvent reprendre l'un quelconque des téléskis 2 et 3, ou alors redescendre la piste 1. A la fin de la journée, ils quittent la station après avoir rejoint le bas de la piste 1.
Modéliser cette station de ski par un réseau de files d'attente.

2. Modélisation d'un système de production:

On considère un atelier flexible dans lequel circulent trois types de pièces p_l. p_2, et p_3. Il comporte les quatre postes de travail suivants:

- un poste de chargement, C
- deux centres d'usinage, CUl et CU2 et
- un poste d'inspection, I.

Les pièces brutes qui arrivent à l'atelier sont fixées sur des palettes et se déplacent ensuite dans l'atelier en respectant la gamme suivante:

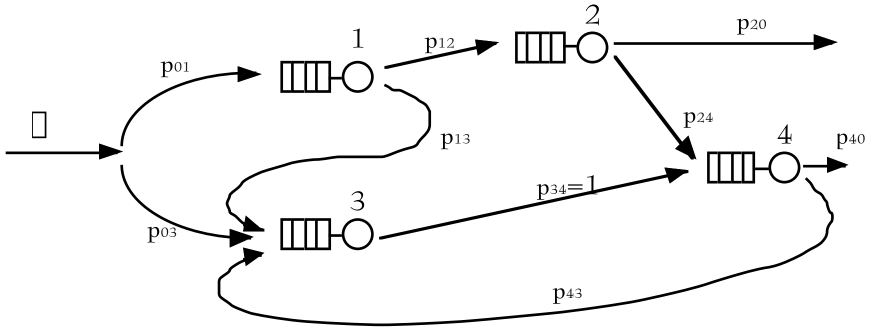
- p_1 : Chargement ⇨ CU1 ⇨ Inspection
- p_2: Chargement ⇨ CU1 ⇨ CU2 ⇨ Inspection
- p_3: Chargement ⇨ CU2 ⇨ Inspection (1/5)

Modéliser cet atelier par un réseau de files d'attente dans les deux cas suivants:

1. Les palettes sont en nombre suffisant.
2. On a un seul type de palettes, en nombre limité N. On a toujours des pièces brutes à l'entrée de l'atelier.

5. Analyse d'un réseau

On considère un RFA ouvert contenant M stations interconnectées de manière quelconque.



Les clients arrivent de l'extérieur du réseau avec un taux λ (pas d'hypothèse sur la distribution d'arrivée pour le moment). Ils se dirigent vers la file i avec une probabilité p_{0i}

$$\text{où} \quad \sum_{i=1}^{M} p_{0i} = 1 \text{ et } p_{0i} \geq 0.$$

Lorsqu'un client a fini son service dans la file i, il va à la station j avec la probabilité p_{ij}.

$$\text{On a : } p_{i1} + p_{i2} + \ldots + p_{iM} \leq 1.$$

Le client quitte le réseau avec la proba :

$$p_{i0} = 1 - \sum_{j=1}^{M} p_{ij}.$$

On définit le taux de visite d'un client à une station i, v_i comme le nombre moyen de passages à cette station i au cours du séjour du client dans le réseau. Les v_i sont solution (unique) de :

$$v_i = p_{0i} + \sum_{j=1}^{M} v_j p_{ji}; \qquad i = 1, \ldots, M;$$

On a aussi besoin de connaître μ_i, le taux de service de la station i.
Supposons le réseau en régime permanent. Soit X_i le nombre moyen de clients arrivant (et quittant) la station i par unité de temps. (un régime stationnaire n'est possible que si $X_i < \mu_i$). En considérant tous les clients qui arrivent à la station i, on peut écrire un ensemble d'équations linéaires reliant les quantités inconnues X_i :

$$X_i = \lambda . p_{0i} + \sum_{j=1}^{M} X_j p_{ji}; \qquad i = 1, \ldots, M;$$

Remarque : on a $X_i = \lambda . v_i$.
Peu de restrictions ont été faites jusqu'à présent mais si l'on veut aller plus loin dans l'étude du réseau et trouver les probas stationnaires et les autres performances, il faut faire un certain nombre d'hypothèses

Hypothèses de la forme produit.

Les réseaux vérifiant ces hypothèses sont appelés réseaux de Jackson.

1. Le processus d'arrivée des clients extérieurs est un processus de poisson.
2. Le cheminement des clients dans le réseau est Markovien c'est à dire qu'il ne dépend que de la dernière station visitée et pas des autres.
3. Chaque station contient ci serveurs de temps de service moyen $t_i = 1/\mu_i$.
   - ◇ les files sont illimitées
   - ◇ elles sont gérées en PAPS
   - ◇ le temps de service est distribué exponentiellement.

Dans ces conditions, la probabilité stationnaire p(n) est donnée par :

$$p(n) = p(n_1, n_2, \cdots, n_M) = \prod_{i=1}^{M} p_i(n_i)$$

où $p(n_i)$ est la proba stationnaire d'avoir ni clients dans une station $M/M/c_i$ en isolation, avec une charge $\rho i = X_i/\mu_i$.

Remarques : Le nombre de clients dans une station donnée ne dépend pas du nombre de clients dans les stations voisines.

La station i se comporte comme si elle était alimentée par un processus de Poisson de taux Xi.

Les hypothèses ne sont que des conditions suffisantes.

Performances

On étudie chaque file en isolation => longueur moyenne Qi, temps de réponse moyen Wi. Puis, pour le réseau, on a tout simplement :

$$Q = \sum_{i=1}^{M} Q_i \quad \text{et} \quad W = \frac{Q}{\lambda} \quad \text{ou} \quad W = \sum_{i=1}^{M} v_i . W_i$$

Etude du temps de réponse dans un atelier :

On considère un atelier contenant deux centres d'usinage (CU1 et CU2), un poste d'inspection (IN) et un poste de réparation (REPAR). Les pièces arrivent dans l'atelier avec un taux moyen λ et vont, soit au poste CU1 avec une probabilité p_{01}. soit au poste CU2 avec une probabilité p_{02}. (Au poste CU1, le temps moyen de service est T1, et au poste CU2, il est de T2). Les pièces quittant CU1 et CU2 sont mises sur un convoyeur (CV) qui les emmène au poste d'inspection au bout d'un temps T3. Après avoir passé un temps moyen de T4 à l'inspection, une proportion p_{40} des pièces quittent l'atelier et les autres vont au poste de réparation où elles passent un temps moyen T5 avant de retourner à l'inspection.

1. Donner le modèle réseau de files d'attente de l'atelier.
2. En supposant qu'il existe un régime stationnaire, donner le débit de chacune des stations du réseau en fonction de λ.
   On prendra $p_{01} = p_{02} = 0.5$ et $p_{40} = 0.9$.
3. Quelles sont les conditions sur λ pour qu'il existe un régime stationnaire ?
4. Donner les hypothèses pour avoir un réseau de files d'attente à forme produit. Sous ces hypothèses, calculer le temps moyen passé par une pièce dans l'atelier, W, en fonction de λ.
   On prendra T1 = T2 = 2, T3 = 10, T4 = 1 et T5 = 5.

Différence Réseaux de Petri/Réseaux de Files d'attente

RFA :

- existe depuis longtemps. De nombreux logiciels de simulation sont partis d'une base de RFA : QNAP 2, Simulog, Arena.
- intérêt : étude de systèmes pas trop complexes mais avec des lois de services ou des processus d'arrivée complexes (même en analyse).
- modèle naturel pour les systèmes manufacturiers

RDP :

- assez récent peu de logiciels mais en développement : Design CPN, Petri Maker
- modélisation rigoureuse, avec une base mathématique et graphique.
- Facile à simuler et à analyser
- intérêt : modélise des systèmes complexes mais ne tient pas compte des lois particulières (en analyse).

Bibliographie

Processus stochastiques, Alan Ruegg, Méthodes mathématiques pour l'ingénieur, Presses polytechniques romandes 1989, 150p.

Très pédagogique. On retrouve les principaux résultats.

Introduction aux files d'attente, E. Gelembe, G. Pujolle, Collection techniques et sciences des télécommunications, Eyrolles 1982.

Contient les résultats de base.

Réseaux de files d'attente - modélisation et traitement numérique, E. Gelembe, G. Pujolle, J. Labetoulle, R. Marie, M. Metivier, W. Stewart, Editions hommes et techniques 1980.

Résultats plus pointus

Queueing Systems, L. Kleinrock, Wiley, New York, Vol 1 : 1975, Vol 2 : 1976.

Très complet.

Du grafcet aux réseaux de Petri 2ème édition, R. David, H. Alla, Traité des nouvelles technologies, série Automatique, Hermès, 1992, 499p.

Contient les résultats de base en RdP.

Résultats et perspectives en automatique, chapitre 4 : "Evaluation de performances des systèmes de production", pp171-234, R. David coordonateur, Traité des nouvelles technologies, série automatique, 1988.

Fait le point sur les différentes méthodes d'évaluation de performances de systèmes séquentiels

Les réseaux de Petri pour la conception et la gestion des systèmes de production, J.M. Proth, X. Xie,,Masson 1994.

Utilisation des RdP pour les systèmes de production.

Etapes d'une analyse

Système de production

Modélisation :

- Simplification plus ou moins grande du système physique (ex : des stocks finis peuvent se modéliser en stock infinis)
- Hypothèses concernant les données (ex : lois de service)

Modèle (RFA, RdP, ...)

Très fidèle, non exploitable

Non fidèle, très exploitable

Analyse

Simulation
reproduction pas à pas
de l'évolution du modèle

Méthodes analytiques
équations / résolution
théorie des processus Markoviens

Simulation :

Simplifications physiques réduites (on peut pratiquement tout prendre en compte).
Permet de connaître le transitoire et le permanent.
Approche la réalité de près mais pas exact.

Difficile de revenir en arrière pour savoir comment changer les paramètres pour améliorer les performances.
Analyse des résultats compliquée d'où une nécessaire simplification des données.
Dure longtemps.

Méthodes analytiques :

Modèle exploitable facilement.
Peu de simplification des données.
Rapide.
Exact ou approximatif.

Simplifications physiques indispensables et importantes.
Ne permet la connaissance que du régime permanent.

Performances du modèle

Analyse des résultats

Figure 1 : analyse d'un système de production