

Modèles Markoviens Fluides et Discrets

Alain Jean-Marie

INRIA et LIRMM, University of Montpellier
161 Rue Ada, 34392 Montpellier Cedex 5, France
ajm@lirmm.fr

Réunion de l'AS 182
Paris, ENST
30 septembre 2004

Plan de l'exposé

Introduction _____ 2

- Modèles de trafic, discrets ou fluides.
- MMPP, MAP, BMAP
- MMRP

Un modèle d'accumulation Semi-Markovien

- Construction
- Distributions
- Temps d'atteinte

Applications _____ 28

- Un modèle d'allocation de capacité
- Un modèle de transmission de vidéo

Introduction

Les processus de Markov modulés sont des processus d'arrivée d'une certaine quantité :

$N(t) \in \mathbb{N}$ quantité de clients dans une file d'attente

$N(t) \in \mathbb{R}$ quantité d'information (bits, octets) arrivés à un élément de réseau

$N(t) \in \mathbb{R}$ quantité de «temps» écoulée (vidéo, temps-réel, biologie,...)

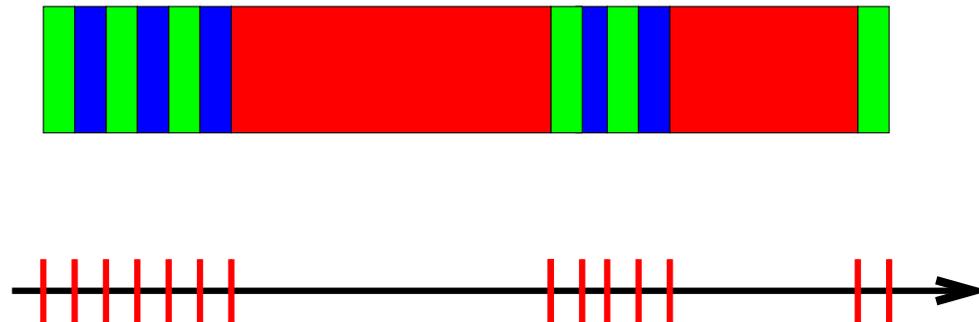
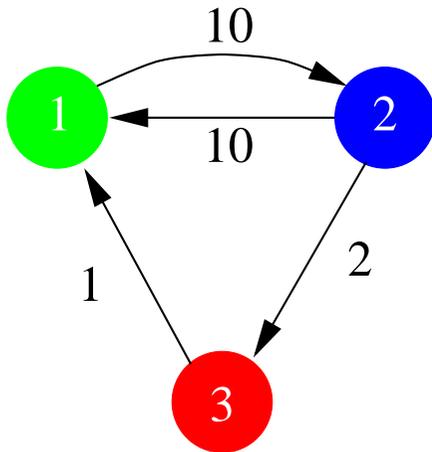
Ils constituent une famille de modèles pour lesquels on arrive à

- modéliser des processus complexes (rafales, dépendances dans le temps),
- maîtriser le nombre de sources,
- faire du calcul stochastique : distributions, files d'attente, asymptotiques...

MAP : Markov Arrival Process

Soit $\{X(t); t \in \mathbb{R}\}$ une chaîne de Markov en temps continu dans un espace fini.

$\{N(t); t \in \mathbb{R}\}$ compte le nombre de sauts de X dans $[0, t[$.



BMAP : Batch Markov Arrival Process

Aussi nommé « N-process » (N comme Neuts), processus « versatile » modulé par Markov.

$\{(X(t), N(t)); t \in \mathbb{R}\}$ est un chaîne de Markov en temps continu dans $\mathcal{E} \times \mathbb{N}$ dont le générateur a la structure :

$$Q = \begin{pmatrix} D_0 & D_1 & D_2 & \dots & \\ & D_0 & D_1 & D_2 & \\ & & D_0 & D_1 & \dots \\ & & & \dots & \dots \end{pmatrix}$$

On peut avoir

- des changements de phase $i \rightarrow j$ sans arrivée, avec taux $(D_0)_{i,j}$
- l'arrivées d'un groupe (batch) de taille k sans changement de la phase i , avec taux $(D_k)_{i,i}$
- changement de phase et arrivée d'un groupe de taille k , avec taux $(D_k)_{i,j}$.

MMPP : Markov Modulated Poisson Process

Soit $\{X(t); t \in \mathbb{R}\}$ une chaîne de Markov en temps continu sur un espace fini \mathcal{E} .

Soient $\lambda_i \geq 0$ des taux d'arrivée, pour chaque $i \in \mathcal{E}$.

On suppose que les arrivées ont lieu selon un processus de Poisson d'intensité $\lambda_{X(t)}$, c'est-à-dire d'intensité λ_i dans les intervalles où $X(t) = i$.

- C'est un cas particulier de processus BMAP, pour lequel
- les arrivées se font un par un : $D_k = 0, k \geq 2$;
 - les changements de phase ne provoquent pas d'arrivée

$$D_1 = \begin{pmatrix} \lambda_0 & & \\ & \lambda_1 & \\ & & \dots \end{pmatrix} .$$

Cas particulier, le processus IPP (Interrupted Poisson Process) : $\lambda_0 = 0, \lambda_1 = \lambda$.

Et le processus de Poisson non interrompu :

$$\lambda_i = \lambda, \quad \forall i \in \mathcal{E} .$$

MMRP : Markov Modulated Rate Process

Soit $\{X(t); t \in \mathbb{R}\}$ une chaîne de Markov en temps continu sur un espace fini \mathcal{E} .

Soient r_i (≥ 0 en général) des taux d'arrivée (accumulation, déplétion), pour chaque $i \in \mathcal{E}$.

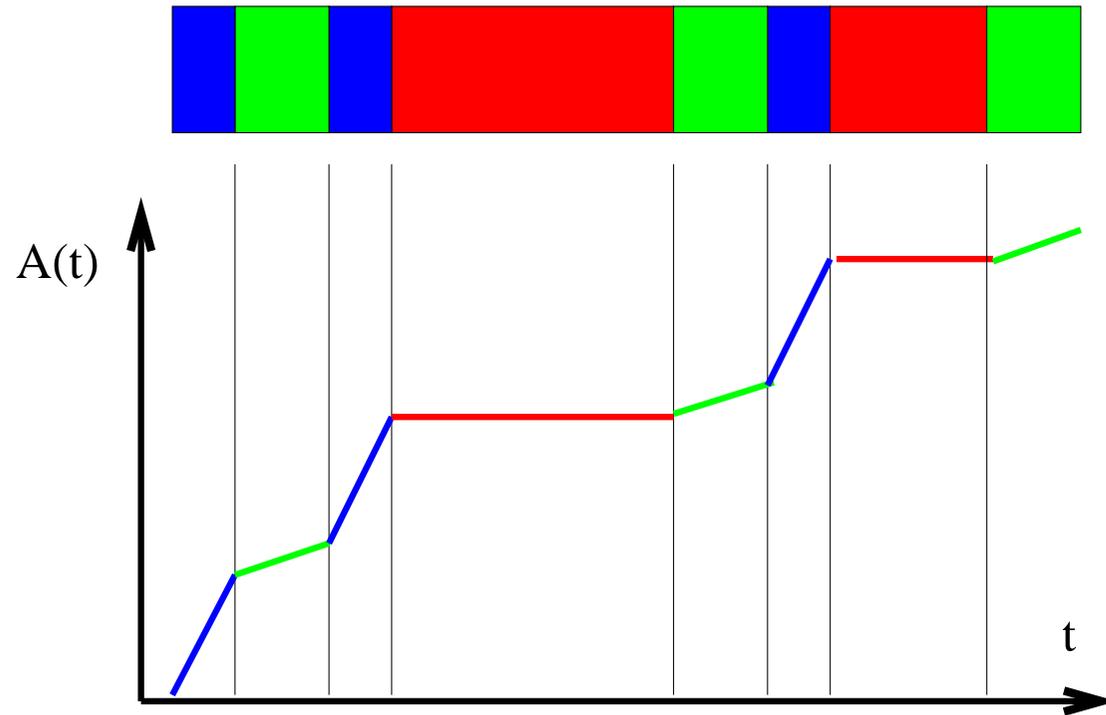
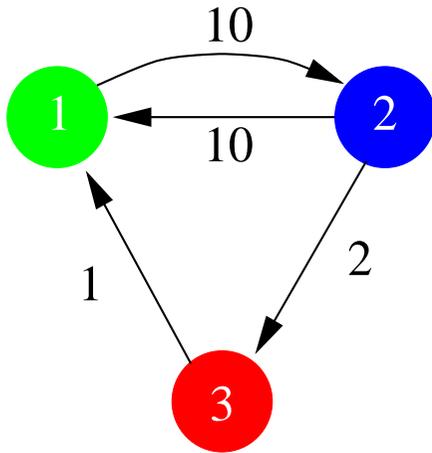
On suppose que les arrivées ont lieu selon un processus fluide de débit $r_{X(t)}$, c'est-à-dire d'intensité r_i dans les intervalles où $X(t) = i$.

Soit $N(t)$ la quantité de fluide arrivé à la date t :

$$\frac{dN}{dt}(t) = r_{X(t)} .$$

Note : anciennement nommé « Markov drift process ».

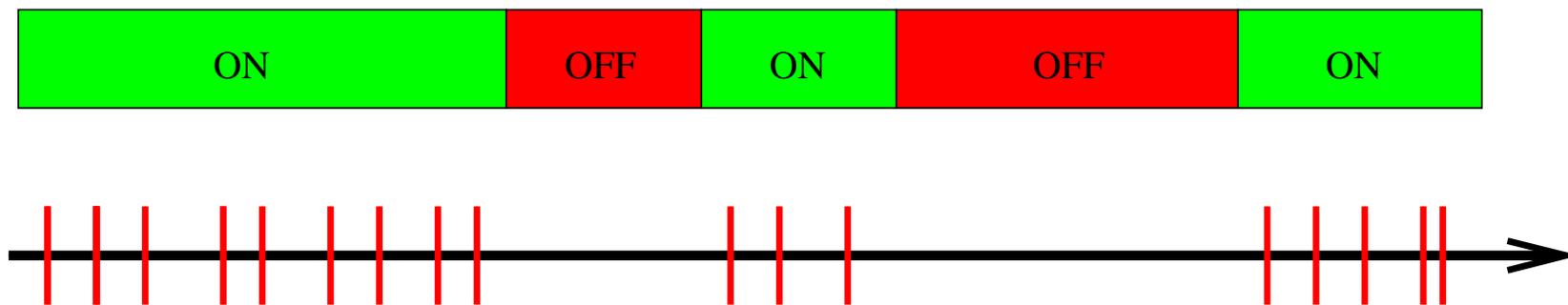
Exemple. \mathcal{E} a trois états, $r_3 = 0$, $0 < r_1 < r_2$:



Sources On/Off

Processus On/Off :

- alternance de périodes On et Off, de durées IID (deux distributions)
- pendant les périodes On, un processus fluides (débit constant) ou discrets (Poisson ou périodique).



→ bon modèle de télécommunication voix/vidéo numériques, ainsi que sources TCP, etc.

→ processus de renouvellement alterné.

MMPP, BMAP, etc ont tous des durées exponentielles. On/Off peut avoir des durées non-exponentielles (en particulier, queues lourdes!).

Propriétés et analyse

Les propriétés communes les plus utiles sont :

- Les processus sont « Markov renouvellement » (semi-Markoviens).
 - Ils se décrivent par des matrices. Par exemple, pour MMPP/MMRP
 - Le générateur infinitésimal de la chaîne de Markov sous-jacente, \mathbf{Q} ,
 - La matrice des taux d'arrivée (rate matrix) $\mathbf{\Lambda}$.
 - Leur superposition mène à un **calcul matriciel** simple.
- une généralisation qui a ces propriétés

Semi-Markov Accumulation Process

Un processus unifiant MMPP et MMRP. Voir aussi les sources « Markov Regenerative »
[Çinlar, Kulkarni \(1996\)](#).

Ingrédients. Soient :

- $\mathcal{E} = \{1, 2, \dots, K\}$ un espace d'états fini ;
- $\{S_i(n)\}_{n \in \mathbb{N}}$ une suite de v.a.i.i.d., $i \in \mathcal{E}$; séjours
- $\{Z(n)\}_{n \in \mathbb{N}}$ une CM en temps discret homogène à états dans \mathcal{E} ; phases
- $\{x_i(t)\}_{t \in \mathbb{R}}$ un processus stationnaire à incréments indépendents, $i \in \mathcal{E}$;
arrivées « continues »
- $\{\Delta_{ij}(n)\}_{n \in \mathbb{N}}$ une suite de v.a.i.i.d., $(i, j) \in \mathcal{E} \times \mathcal{E}$. arrivées par sauts

On définit d'abord un processus semi-Markovien $\{X(t)\}_{t \in \mathbb{R}}$, conjointement à une suite de dates de sauts $\{T_n\}_{n \in \mathbb{N}}$:

$$\begin{aligned} T_{n+1} &= T_n + S_{Z(n)}(n) \\ X(t) &= Z(n+1) & T_n \leq t < T_{n+1} \\ X(t) &= Z(0) & 0 \leq t < T_0, \end{aligned}$$

$T_0 \geq 0$ v.a. donnée.

Le processus $\{N(t)\}_{t \in \mathbb{R}}$ est construit comme :

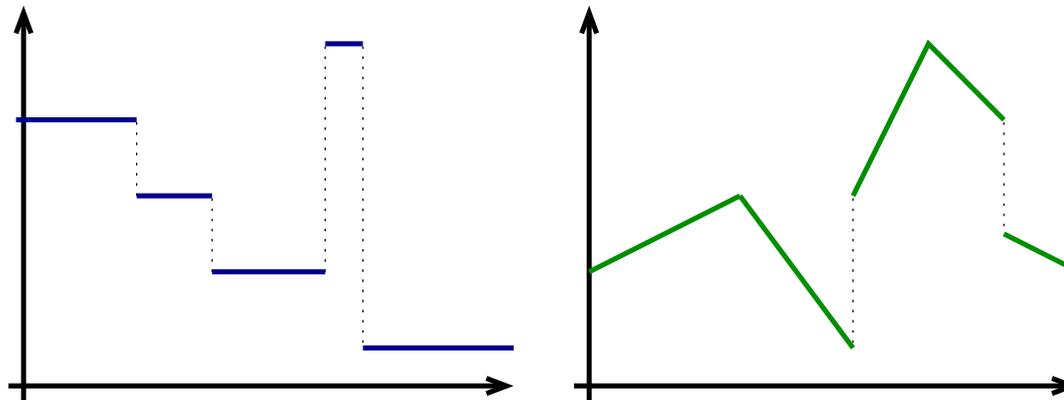
$$\begin{aligned} N(t) &= N(T_n^+) + (x_{Z(n+1)}(t) - x_{Z(n+1)}(T_n)) & T_n < t < T_{n+1} \\ N(t) &= x_{Z(0)}(t) - x_{Z(0)}(T_n) & 0 \leq t < T_0 \\ N(T_n^+) &= N(T_n^-) + \Delta_{Z(n), Z(n+1)}(n). \end{aligned}$$

avec $N(0) = 0$.

Ce processus généralise bien comme processus

- les MMPP avec des incréments Poisson $x(t) = N_\lambda(0, t)$,
- les MMRP avec des incréments fluides à débit constant $x(t) = rt$,
- les BMAP avec les incréments $\Delta_{ij}(n)$.

On a aussi les combinaisons des deux (dont les « Piecewise Deterministic Process »).



Le processus de Wiener est à incréments indépendants : utilité possible pour des sources « bruitées », ou pour un grand nombre de sources ?

Il est connu que pour ces processus (*e.g.* Doob (1952)) :

$$\mathbb{E}(e^{-\nu(x(t)-x(s))}) = e^{-(t-s)\phi(\nu)},$$

avec par exemple

$$\begin{aligned}\phi(\nu) &= r\nu && \text{pour un taux d'accumulation constant } r \\ \phi(\nu) &= r(1 - e^{-\nu}) && \text{pour un processus de Poisson de débit } r \\ \phi(\nu) &= r\nu + \frac{1}{2}\sigma^2\nu^2 && \text{pour un processus de diffusion de dérive } r \text{ et variance } \sigma^2.\end{aligned}$$

Distributions

Sont intéressants à calculer :

- $N^{i;r}(T)$: quantité accumulée dans $[0, T)$ sachant que $X(0) = i$ et $T(0) = r$.
- $N^{i;S_i}(T)$: idem sachant que $T_0 = 0$,
- $N^{i;\tilde{S}_i}(T)$, où \tilde{S}_i is le temps de récurrence avant de S_i .

Mais aussi :

- $\tau^{i;r}(x)$: temps d'atteinte de la quantité x , sachant que $X(0) = i$ et $T(0) = r$.
- autocorrélations, densité spectrale,
- corrélations croisées entre plusieurs processus définis sur la même chaîne...

Comme cas particuliers de $N(T)$, on aura :

- le nombre de sauts dans $[0, T)$,
- la durée de séjour dans un état pendant $[0, T)$.

Distribution de $N(T)$

On peut calculer la transformée de Laplace de la quantité arrivée en T unités de temps :

$$\begin{aligned} K_{i,j}(\mu, \nu) &= \int_0^\infty e^{-\mu T} \mathbb{E}(e^{-\nu N^{i;S_i}(T)} \mathbf{1}_{\{X(T)=j\}}) dT , \\ &= \int_0^\infty e^{-\mu T} \int_0^\infty e^{-\nu q} d\mathbb{P}_i\{N(T) \leq q, X(T) = j\} dT . \end{aligned}$$

Définissons les matrices, fonctions de (μ, ν) :

$$\begin{aligned} \mathbf{K} &= ((K_{i,j}(\mu, \nu)))_{(i,j) \in \mathcal{E}\mathcal{E}} & \mathbf{L} &= \text{diag} \left(\frac{1}{\mu + \phi_i(\nu)} \right)_{i \in \mathcal{E}} \\ \mathbf{S} &= \text{diag} (S_i^*(\mu + \phi_i(\nu)))_{i \in \mathcal{E}} & \mathbf{\Delta} &= ((p_{ij} \Delta_{ij}^*(\nu)))_{(i,j) \in \mathcal{E} \times \mathcal{E}} . \end{aligned}$$

Alors,

$$\begin{aligned} \mathbf{K} &= \mathbf{L} (\mathbf{I} - \mathbf{S}) + \mathbf{S} \Delta \mathbf{K} \\ \mathbf{K} &= (\mathbf{I} - \mathbf{S} \Delta)^{-1} \mathbf{L} (\mathbf{I} - \mathbf{S}) . \end{aligned}$$

Preuve : arguments standard, en conditionnant sur la date du premier saut. Voir *e.g.* [Cox & Miller \(1965\)](#) pour $K = 2$ (processus de renouvellement alterné).

Une fois \mathbf{K} obtenue, les autres conditions initiales ($T_0 \neq 0$) peuvent être déduites. Par exemple, pour le cas stationnaire :

$$\begin{aligned} \tilde{K}_{i,j}(\mu, \nu) &= \int_0^\infty e^{-\mu T} \mathbb{E} \left[e^{-\nu N^{i; \tilde{S}_i}(T)} \mathbf{1}_{\{X(T)=j\}} \right] dT \\ \tilde{\mathbf{K}} &= \mathbf{L} (\mathbf{I} - \tilde{\mathbf{S}}) + \tilde{\mathbf{S}} \Delta \mathbf{K} . \end{aligned}$$

Temps d'atteinte de N

On se restreint au cas fluide avec $r_i \geq 0$, $\Delta_{ij} \equiv 0$. Soit

$$\tau^{i;r}(x) = \inf\{t \mid N(t) \geq x\}$$

le temps d'atteinte du niveau x , sachant que $X(0) = i$ et $T(0) = r$.

On peut calculer :

$$\mathcal{I}_{i,j}(\mu, \nu) = \int_0^\infty e^{-\mu x} \mathbb{E} \left[e^{-\nu \tau^{i;S_i}(x)} \mathbf{1}_{\{X(\tau^{i;S_i}(x))=j\}} \right] dx .$$

Avec les matrices

$$\mathcal{I} = ((\mathcal{I}_{i,j}(\mu, \nu)))_{(i,j) \in \mathcal{E} \times \mathcal{E}} \quad \bar{\mathbf{L}} = \text{diag} \left(\frac{r_i}{r_i \mu + \nu} \right)_{i \in \mathcal{E}},$$

on a :

$$\begin{aligned} \mathcal{I} &= \bar{\mathbf{L}} (\mathbf{I} - \mathbf{S}) + \mathbf{SP}\mathcal{I} \\ \mathcal{I} &= (\mathbf{I} - \mathbf{SP})^{-1} \bar{\mathbf{L}} (\mathbf{I} - \mathbf{S}) . \end{aligned}$$

Cas connu 1 : chaînes de semi-Markov

Soit $R(n) = N(T(n))$. Le processus $(R(n), Z(n))$ is un « Markov Additive Process » avec récompense distribuée comme $x_i(S_i) - x_i(0)$.

Soit $\bar{\mathbf{S}} = \text{diag}(S_i^*(\phi_i(\nu)))$ et

$$M_{i,j}(\nu) = \mathbb{E}(e^{-\nu R(n)} \mathbf{1}_{\{Z(n)=j\}} \mid X(0) = i, T(0) = 0) ,$$

Alors :

$$\mathbf{M} = (\bar{\mathbf{S}}\mathbf{P})^n .$$

Cas connu 2 : nombre de sauts

Soit $J(T)$ le nombre de sauts d'une chaîne de Markov en temps continu dans l'intervalle $[0, T)$. On a, pour :

$$J_{i,j}^*(z, t) = \int_0^\infty e^{-st} \mathbb{E} \left(z^{J(t)} \mathbf{1}_{\{X(t)=j\}} \mid X(0) = i \right) ds ,$$

$$\begin{aligned} (\mathbf{\Lambda} + s\mathbf{I}) \mathbf{N}^* &= z(\mathbf{\Lambda} + \mathbf{Q}) \mathbf{N}^* + \mathbf{I} , \\ \mathbf{N}^*(z, s) &= (s\mathbf{I} - z\mathbf{Q} - (z-1)\mathbf{\Lambda})^{-1} . \end{aligned}$$

Par exemple, pour le nombre moyen :

$$\mathbb{E}\mathbf{N}(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \sum_{p=0}^{n-1} \mathbf{Q}^p (\mathbf{Q} + \mathbf{\Lambda}) \mathbf{Q}^{n-p-1} .$$

Applications

Modèles de transmission vidéo : [Biersack, Jean-Marie et Nain \(2002\)](#).

La vidéo numérique est constituée comme un fluide, émis par une source (« serveur »), acheminé par le réseau, consommé par le « client ».

Plusieurs questions ont été abordées :

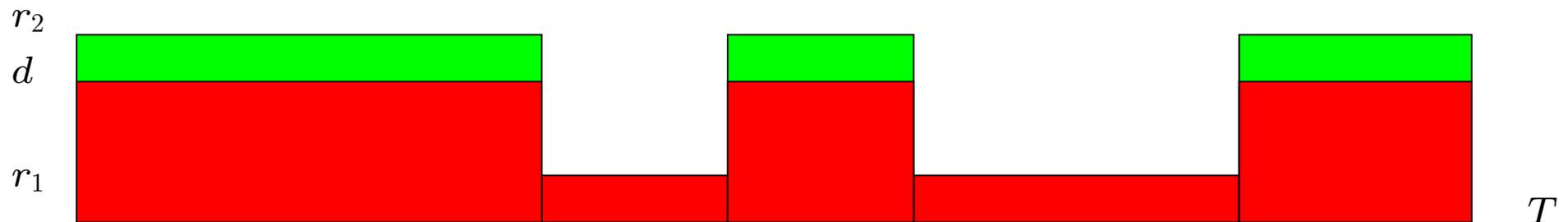
- combien de temps pour télécharger une vidéo ?
- combien de ressources pour obtenir un temps donné ?
- quelle probabilité que le client prenne de l'avance ?
- combien de ressources pour une probabilité donnée ?

Modèle de transport

Le processus du débit **disponible** alterne entre deux valeurs $r_1 < r_2$.

Le serveur choisit d'envoyer au débit d avec $r_1 \leq d \leq r_2$.

Soit $C_T(d)$ la quantité transmise à la date T avec un débit d , $r_1 < d < r_2$.



On se ramène à un processus On/Off avec débit crête = 1 :

$$C_T(d) = r_1 T + (d - r_1) N(T) .$$

Calcul des transformées de Laplace

$$\int_0^\infty e^{-\theta x} \int_0^\infty e^{-sT} \mathbb{P}\{N(T) > x\} dT dx$$
$$= \frac{1}{m_1 + m_2} \left\{ \frac{m_2}{s(s + \theta)} + \frac{\theta(1 - G_1^*(s))(1 - G_2^*(s + \theta))}{s^2(s + \theta)^2(1 - G_1^*(s)G_2^*(s + \theta))} \right\} .$$

m_1, m_2 : moyennes des périodes On et Off.

Peut s'obtenir également avec le calcul matriciel des MMRP, qui se généralise au cas de plusieurs états du canal.

→ calcul de **moments**,

$$\text{Var}(C_T) = T^2 \frac{m_1^2 v_2 + m_2^2 v_1}{(m_1 + m_2)^3} + O(T) .$$

→ calcul de **débits équivalents** : pour $q \in [0, 1)$ donné, trouver γ tel que

$$\mathbb{P}\{C(d) \geq \gamma T\} \geq q .$$

Distribution du temps de séjour pour un SMAP à deux états

Cas Markovien. Soit $N_1(T)$, la quantité de temps passé en 1 dans l'intervalle $[0, T)$.
Les matrices caractéristiques sont :

$$\mathbf{Q} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix} \quad \mathbf{L}^{-1} = \begin{pmatrix} \mu + \nu & 0 \\ 0 & \mu \end{pmatrix},$$

$$\mathbf{\Delta} = \mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \frac{\alpha}{\alpha + \mu + \nu} & 0 \\ 0 & \frac{\beta}{\beta + \mu} \end{pmatrix}.$$

Inversion des transformées de Laplace

En utilisant l'équation qui donne \mathbf{K} , on a :

$$\mathbf{K} = \frac{1}{(\mu + \beta)(\mu + \nu + \alpha) - \alpha\beta} \begin{pmatrix} \mu + \beta & \alpha \\ \beta & \mu + \nu + \alpha \end{pmatrix} .$$

Inverser par rapport à ν est direct :

$$\int_0^\infty e^{-\mu T} \frac{d}{dq} \mathbb{P}(N^{1;S_1}(T) \leq q, X(T) = 1) dT = e^{-\alpha q} e^{-\mu q} \exp\left(\frac{\alpha\beta q}{\beta + \mu}\right) .$$

À inverser en μ .

Avec les tables de transformées de Laplace ([Spiegel \(1965\)](#)), on obtient directement les densités :

$$q_{ij}(x) = \frac{d}{dx} \mathbb{P}(N(T) \leq x, X(T) = j | X(0) = i)$$

$$q_{11}(x) = e^{-\alpha T} \delta_T(x) + e^{-\alpha x} e^{-\beta(T-x)} I_1(f(x)) \sqrt{\frac{\alpha\beta x}{T-x}} \mathbf{1}_{\{T \geq x\}}$$

$$q_{12}(x) = \alpha e^{-\alpha x} e^{-\beta(T-x)} I_0(f(x)) \quad q_{21}(x) = \beta e^{-\alpha x} e^{-\beta(T-x)} I_0(f(x))$$

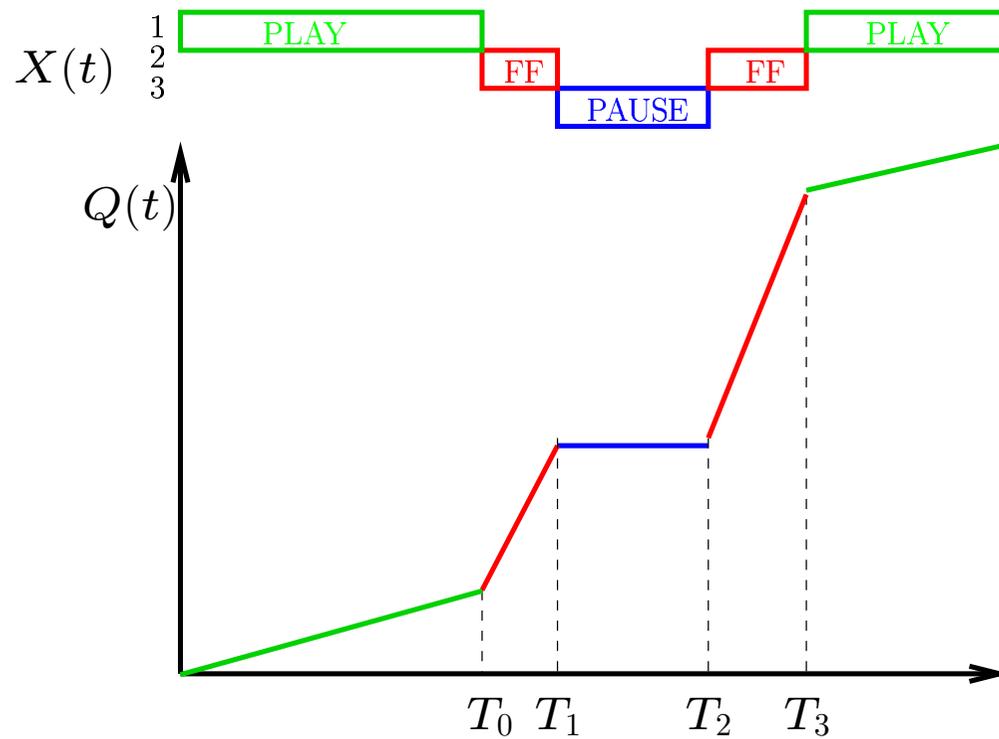
$$q_{22}(x) = e^{-\beta T} \delta_o(x) + e^{-\alpha x} e^{-\beta(T-x)} I_1(f(x)) \sqrt{\frac{\alpha\beta(T-x)}{x}} \mathbf{1}_{\{T \geq x\}},$$

et $f(x) = 2\sqrt{\alpha\beta x(T-x)}$.

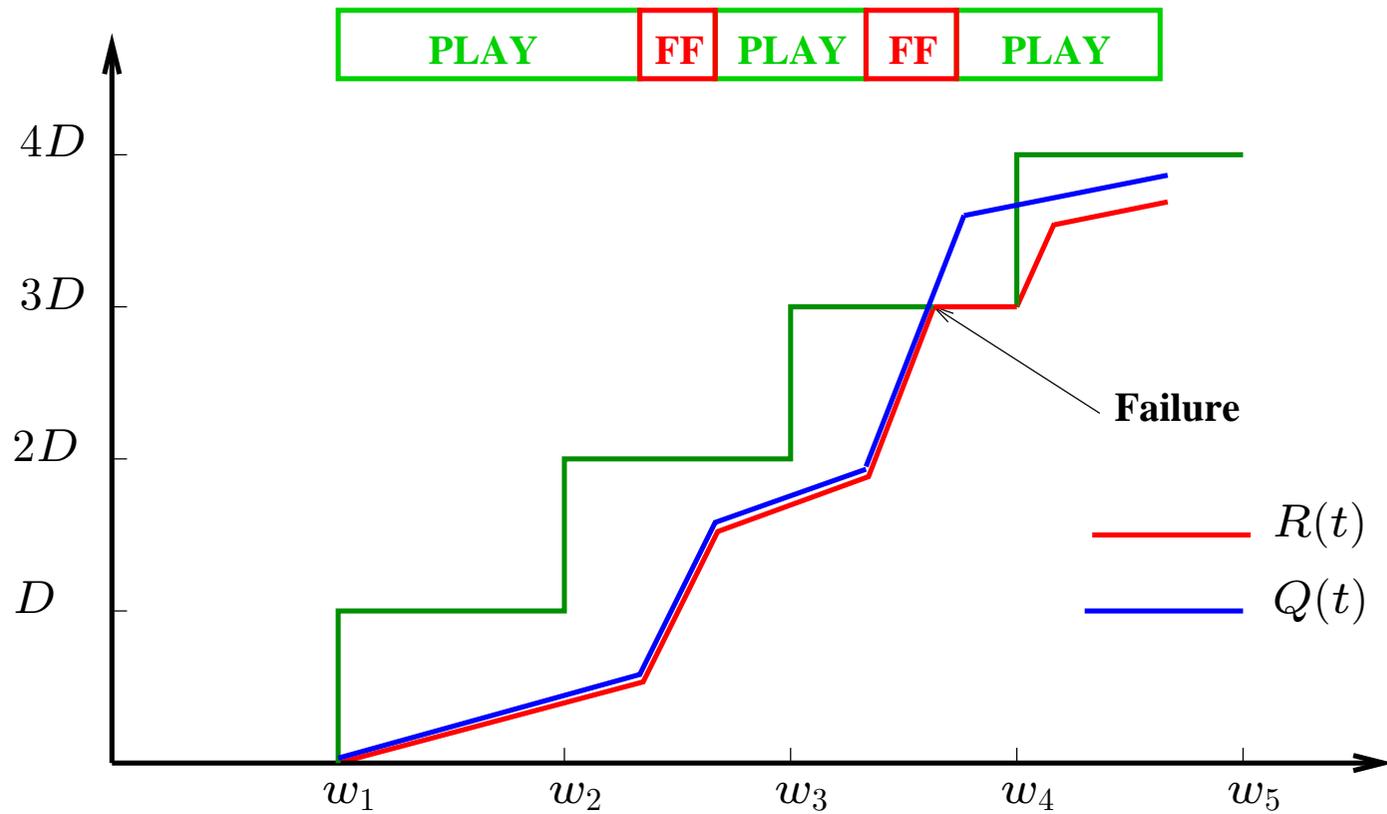
Les fonctions de distribution $\mathbb{P}(N(T) \leq x, X(T) = j | X(0) = i)$ s'ensuivent par intégration numérique.

Modèle de lecture

Considérons la lecture d'une vidéo comme un processus fluide.



Étant donnée une transmission par blocs à débit constant, quelle est la probabilité d'un « ratage » ?



Questions

Aspects stochastiques Lois limites (temps/espace), asymptotiques

Autocorrélations, mémoire

Files d'attente ? Par exemple pour le *workload* de la *MMPP/GI/1*,

$$\mathbf{W}^*(s) = s(1 - \rho) \mathbf{g} [s\mathbf{I} + \mathbf{Q} - (1 - H^*(s))\mathbf{\Lambda}]^{-1},$$

où

$$\mathbf{g} = \mathbf{g}\mathbf{G} \quad \text{et} \quad \mathbf{G} = \int_0^\infty \exp((\mathbf{Q} - \mathbf{\Lambda} + \mathbf{\Lambda}\mathbf{G})x) dH(x).$$

Aspects analytiques Calcul systématique de **moments** et d'**asymptotiques** : à partir de

$$\mathbf{K}(\mu, \nu) = (\mathbf{I} - \mathbf{S}\mathbf{\Delta})^{-1} \mathbf{L} (\mathbf{I} - \mathbf{S})$$

savoir dériver en $\nu = 0$ et $\mu = 0$.

Bornes (Chernoff...), grandes déviations, etc.

Aspects numériques Calcul systématique de distributions par inversion de transformées de Laplace doubles.

Bibliographie

Modèles fluides

D. Anick, D. Mitra, and M.M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *Bell Sys. Tech. J.*, 61 :1871–1894, October 1982.

D. Mitra. Stochastic theory of a fluid models of producers and consumers coupled by a buffer. *Adv. Appl. Prob.*, 20 :646–676, 1988.

T.E. Stern and A.I. Elwalid. Analysis of separable Markov-modulated rate models for information-handling systems. *Adv. Appl. Prob.*, 23 :105–139, 1991.

A.I. Elwalid, D. Mitra, and T.E. Stern. Statistical multiplexing of Markov modulated sources :theory and computational algorithms. In A. Jensen and V.B. Iversen, editors, *Proc. 13th International Teletraffic Congress*, pages 495–500, Copenhagen, 1991. Elsevier Science.

A.I. Elwalid, D. Mitra, and T.E. Stern. A theory of statistical multiplexing of Markov modulated sources : Spectral expansions and algorithms. In W.J. Stewart, editor, *Numerical solution of Markov Chains*, 1991.

A.I. Elwalid and D. Mitra. Statistical multiplexing with loss priorities in rate-based congestion control of high speed networks. *IEEE Trans. Comm.*, 42(11) :2989–3002, November 1994.

A.I. Elwalid and D. Mitra. Markovian arrival and service communication systems : Spectral expansions, separability and Kronecker-product forms. In W.J. Stewart, editor, *Computations in the Markov Chains*, pages 507–546. Kluwer, 1995.

MMPP, MAP, BMAP...

M.F. Neuts. The fundamental period of a queue with Markov-modulated arrivals. In *Probability, Statistics and Mathematics : papers in honour of Samuel Karlin*. Academic Press, NY, 1989.

W. Fischer and K. Meier-Hellstern. The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, 18 :149–171, 1992.

D.M. Lucantoni, G.L. Choudhury, and W. Whitt. The transient *BMAP/G/1* queue. *Commun.*

Statist.-Stochastic Models, 10(1) :145–182, 1994.

A. Jean-Marie, Z. Liu, P. Nain and D. Towsley, “Computational Aspects of the Workload Distribution in the MMPP/GI/1 Queue”. *JSAC*, 1999.

Asymptotiques, bornes et bande passante équivalente

W. Whitt. Tail probabilities with statistical multiplexing and effective bandwidth. *Telecommun. Syst.*, 3 :71–107.

D. Artiges and P. Nain. Upper and lower bounds for the multiplexing of multiclass Markovian on/off sources. *Performance Evaluation*, **27&28**, pp. 673–698, 1996.

V.G. Kulkarni. Effective bandwidth for Markov regenerative sources. *Queueing Systems*, **24**, pp. 137–153, 1996.

Z. Liu, P. Nain, and D. Towsley. Exponential bounds with applications to call admission. *JACM*, 44 (2) :366–394, 1997.