Sur un théorème de limite conditionnelle

Adnan Aboulalaa

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Sur un théorème de limite conditionnelle. Application aux systèmes de capacités finies

Adnan Aboulalaâ

N° 2389

Novembre 1994

PROGRAMME 1

Architectures parallèles,

bases de données,

réseaux et systèmes distribués

apport de recherche

1994

ISSN 0249-6399

INRIA SOPHIA ANTIPOLIS

Sur un théorème de limite conditionnelle. Application aux systèmes de capacités finies

Adnan Aboulalaâ*

Programme 1 — Architectures parallèles, bases de données, réseaux et systèmes distribués

Projet MISTRAL

Rapport de recherche n° 2389 — Novembre 1994 — 28 pages

Résumé : Étant donné N copies indépendantes $X_t^1, \ldots, X_t^N$ d'un processus de sauts à valeurs dans un espace d'états fini, on étudie l'évolution d'un processus type conditionnellement à un événement du type $\{1/N \sum_{i=1}^N f(X_t^i) \in U \subset R^k, \forall t \in [0,T]\}$ lorsque N tend vers l'infini (en fait on considérera des événements qui sont proches de ce dernier). Il s'agit d'une extension au cas du temps mobile du 'principe de conditionnement de Gibbs'. Ce principe est appliqué à l'étude d'un système disposant de k types de ressources, en quantités finies, partagées par N clients.

Mots-clé : Loi limite conditionnelle, grandes déviations, information de Kullback, physique statistique, processus de sauts contraints, systèmes à capacités finies.

(Abstract: pto)

*. École Polytechnique et ENST. E-mail : aboulala@enst.fr

On a conditional limit theorem and application to loss systems

Abstract: Consider N independant copies $X_t^1, \ldots, X_t^N$ of a Markov jump processes with finite state space, we study the limiting behavior of a typical process, say X_t^1, given a condition like $\{1/N \sum_{i=1}^N f(X_t^i) \in U \subset R^k, \forall t \in [0,T]\}$ (in fact we will consider conditions which are close to the last one). This is a process version of a resulat which is called the Gibbs conditioning principle. This principle is applied to the analysis of a system which has several kinds of ressources shared by N traffic sources.

AMS 1991 Subject Classification: 60B12, 60F10, 60J75, 60K25

Key-words: Conditional limit law, large deviations, Kullback information, statistical physics, jump processes with constraints, loss systems.

1 Introduction

Le problème du conditionnement auquel on s'intéresse ici se pose de la manière suivante: on se donne une suite de variables i.i.d $(X^i)_{i\in\mathbb{N}}$ à valeurs dans un espace $\mathcal{S}$, le plus général possible. Chaque X^i peut être vue comme représentant l'état d'une particule (i) d'un gaz. On cherche alors la loi d'une de ces variables sous une contrainte du type

$$C(N,U) := \{\frac{1}{N}\sum_{i=1}^{N} f(X^i) \in U\}$$

où $f : \mathcal{S} \longrightarrow V$, ($V$ espace localement convexe en général). $f(X^i)$ peut s'interpréter comme l'énergie de la particule i et la condition correspond à une contrainte sur l'énergie globale du système. Autrement dit la question est de connaître la loi conditionnelle par rapport à $C(N,U)$ de X^1 (lorsque N tend vers l'infini). Ce problème a été traité par Vasicek (1980) [15] et Van Campenhout et Cover (1981) [14] dans des cas particuliers, puis par Csiszár (1984)[4] dans une grande généralité. Le résultat est que cette loi conditionnelle limite est obtenue en maximisant une certaine entropie (-l'information de Kullback d'une probabilité par rapport à la loi commune des X^i) sur un ensemble de mesures adéquat. Le grand principe de maximisation d'entropie en physique statistique, ou en d'autres domaines, y trouve ainsi un certain[1] fondement.
Cependant les problèmes en temps mobile n'ont apparement pas été considérés dans le passé. Le plus simple se pose de la manière suivante: les $X^i \equiv X^i_t, t \in [0,T]$ sont des processus markoviens de sauts à espace d'états fini Σ indépendants et identiquement distribués et l'on s'intéresse au comportement de $X^1_t, t \in [0,T]$ conditionnellement à un événement du type

$$C(N,U) = \{\frac{1}{N}\sum_{i=1}^{N} f(X^i_t) \in U \ \forall t \in [0,T]\}, f : \Sigma \longrightarrow \mathbb{R}^k, k \geq 1$$

lorsque N devient grand; (U est un sous ensemble convenable de $\mathbb{R}^k$ et $T > 0$ est fixé). Ce problème est motivé par l'étude d'un modèle de système de communication(cf. *infra*); il a également la même interprétation en terme de physique statistique que celle décrite ci-dessus. Un des objectifs de ce travail est d'apporter des éléments de réponse à cette question.
Le résultat concernant la loi limite conditionnelle en temps fixe est connu sous le nom de principe de conditionnement de Gibbs. Il a trouvé des applications dans des domaines variés (traitement d'images, conception VLSI...). On en propose une dans le domaine des systèmes de communication.

⋆⋆

1. Il s'agit du cas particulier de la distribution microcanonique, voir aussi Stroock-Zeitouni(91)[13].

Les modèles classiques des systèmes tels que les réseaux de communication, e.g. les modèles de Jackson ou de Kelly, permettent d'obtenir le régime stationnaire -probabilité invariante d'un processus de Markov représentant l'état du système- sous une forme explicite. Cependant le calcul effectif et direct de cette probabilité stationnaire n'est possible que pour les 'petits' systèmes(nombre faible de noeuds et de liaisons), à cause de la compléxité combinatoire des formules.
Ce problème de compléxité se retrouve en des termes pareils en physique statistique où l'on cherche également la distribution de probabilité des états d'équilibre qui peuvent donner lieu là aussi à des problèmes de combinatoire pour calculer la fonction de partition. Vu ces analogies et les succés remportés par la physique statistique, il est assez naturel d'examiner l'éventuelle utilité des méthodes de la physique statistique dans l'étude des grands systèmes de communication. Un des premiers travaux allant dans ce sens (et qui en a inspiré d'autres par la suite cf. e.g. [7],[11],[12]) est celui de Beneš [1]. Il décrit les états du système par un ensemble S dont les éléments x représentent les ressources occupées à un instant donné en chaque 'point' du réseau. Alors en désignant par $|x|$ la sommes de toutes les ressources occupées(ou le nombre total des appels en cours) lorsque le réseau est dans l'état x, et par q_x la probabilité d'un tel état, la distribution "d'équilibre" sur S est obtenue en maximisant l'entropie

$$H(q) = -\sum_{x\in S} q_x Log q_x$$

sous la contrainte que le tafic moyen dans le réseau est fixé:

$$\sum_{x\in S} |x| q_x = m$$

Moyenant ces postulats la distribution d'équilibre est donnée par

$$q_x = \frac{\lambda^{|x|}}{\Phi(\lambda)},\ \Phi(\xi) = \sum_{x\in S} \xi^{|x|},\ m = \lambda \frac{dLog\Phi(\lambda)}{d\lambda}$$

Et comme en physique statistique ou dans le modèle de Kelly le problème du calcul de la fonction de partition se retrouve dans ce cas cf. [1] pour plus de détails.

Le modèle que l'on propose ici est basé sur l'idée suivante : une manière d'évaluer les performances (ou les pertes) dans un système de capacité finie est de comparer le comportement 'libre' des utilisateurs i.e. lorsque le système a une capacité infinie et le comportement conditionnel de ceux ci par rapport à la contrainte de capacité finie. Malgré quelques analogies entre ce modèle et celui de Beneš (on s'intéresse notament au comportement macroscopique du système en faisant abstraction de sa topologie), ceux ci sont différents, outre du fait que les idées de base ne sont pas les mêmes, pour les raisons

suivantes:
• Le réseau n'est pas décrit par des variables représentant ses différents états, on s'intéresse plutôt au comportement des utilisateurs .
• On prend en compte effectivement la contrainte de limitation des ressources qui est réelle alors que V.Beneš prend comme contrainte un trafic moyen constant.
• la technique de maximisation d'entropie qui apparait comme un principe dans la théorie de Beneš, se retrouve dans le problème qu'on traitera comme un résultat (mais dans un contexte différent).
Enfin on peut dire, très schématiquement, que la démarche de V.Beneš correspond à l'ensemble canonique de la physique statistique alors que le modèle que l'on présente correspond à l'ensemble microcanonique (au sens de [13]).

⋆⋆

Le présent travail est organisé comme suit. Au §2. on présente le modèle adopté qui fait le lien entre la question du conditionnement et l'évaluation des performances des systèmes de capacités finies. Ceci est illustré par l'exemple des réseaux hiérarchiques. Le §3. est consacré aux "problèmes à temps fixe" où les v.a. X^i sont à valeurs dans un ensemble fini . On y rappelle le résultat sur la loi conditionnelle limite de X^1 et on indique d'autres résultats de convergence. Ceux ci sont ensuite appliqués à l'exemple des systèmes hiérarchiques. Au §4. on étudie les problèmes à temps mobile : les X^i sont cette fois des fonctions aléatoires et l'on s'intéresse à la loi conditionnelle limite de $X_t^1, t \in [0,T]$ sous une contrainte sur 'l'énergie moyenne' à chaque instant. Ce cas est bien entendu plus intéressant mais en contrepartie les résultats correspondant sont peu explicites. On traite d'abord le cas des chaînes de Markov §4.1 qui permet de deviner la forme de la loi conditionnelle limite éventuelle dans le cas temps continu; ce dernier cas est traité en modifiant légèrement la forme de l'événement $C(N,U)$, §4.2; on donnera les caractéristiques du processus X_t^1 sous cette loi limite.

2 Le modèle

2.1 Modèle et problèmes

On considère un système qui possède k types de ressources et auquel peuvent accéder N clients. Ces ressources sont limitées en nombre de $L_1, ..., L_k$ unités respectivement.
Si les ressources étaient infinies un client évoluerait de manière aléatoire, identique (aux) et indépendante (des) autres utilisateurs, dans un espace d'état fini $\Sigma = \{e_0, ..., e_d\}$. On suppose aussi que l'on dispose d'une fonction $f : \Sigma \longrightarrow \mathbb{R}^k$ qui à chaque état associe les quantités utilisées de chaque ressources. Le comportement réel des utilisateurs est

donné par leur comportement *a priori* conditionné par le fait que les ressources utilisées à chaque instant ne peuvent pas dépasser, en quantité, les ressources disponibles. On se pose alors les problèmes suivants :

2.1.1 Problèmes à temps fixe

On suppose que les utilisateurs sont représentés par des variables i.i.d. $X^1, ..., X^N$ à valeurs dans Σ et définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On pose

$$\mathbb{P}(X^i = e_h) = \mu_h$$

La condition de limitation des ressources s'écrit

$$\sum_{i=1}^{N} f(X^i) \in U_0,\ U_0 = \prod_{j=1}^{k} [0, L_j] \subset \mathbb{R}^k$$

Le comportement réel des utilisateurs est donné par

$$\nu_h^N := \mathbb{P}(X^i = e_h | \sum_{i=1}^{N} f(X^i) \in U_0)$$

En particulier l'espérance de la quantité de ressources que demandent les utilisateurs est

$$E(\sum_{i=1}^{N} f(X^i)) = N \sum_{h=0}^{N} \mu_h f(e_h)$$

(on a même $1/N \sum_{i=1}^{N} f(X^i)) \longrightarrow \sum_{h=0}^{N} \mu_h f(e_h)\ \mathbb{P} - p.s.$) et l'espérance de la quantité de ressources réellement utilisée est

$$E(\sum_{i=1}^{N} f(X^i) | \sum_{i=1}^{N} f(X^i) \in U_0) = N \sum_{h=0}^{N} \nu_h^N f(e_h)$$

On établira un résultat de type convergence en probabilité(sous les contraintes)

$$\frac{1}{N} \sum_{i=1}^{N} f(X^i) \longrightarrow \sum_{h=0}^{d} \nu_h f(e_h)$$

(prop. 3.1.3.) où $\nu_h = \lim_N \nu_h^N$, prop. 3.1.1. Donc la détermination des probabilité conditionnelles ν_h donne des renseignements intéressants sur les performances du système. Ce cas peut s'appliquer en particulier aux problèmes des accés concurents à un instant donné.

Remarque 2.1.1 Dans toute la suite on posera $L_i = Nl_i, i = 1, \ldots d$; ν_h doit être vue comme une estimation asymptotique de ν_h^N qui dépendra des l_i, μ_h. Enfin l'événement par lequel on conditionne s'écrira désormais

$$C(N, U) = \{\frac{1}{N} \sum_{i=1}^{N} f(X^i) \in U\},\ U = \prod_{j=1}^{k} [0, lj] \subset \mathbb{R}^k$$

2.1.2 Problèmes à temps mobile

On fixe un intervalle de temps $[0, T]$. On suppose que l'évolution des utilisateurs est donnée par des processus $X_t^1, ..., X_t^N, t \in [0, T]$ indépendants et de même loi, dont l'espace des états est Σ, définis sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Ici la condition de limitation des ressources s'écrit

$$C(N, U) = \{\frac{1}{N}\sum_{i=1}^{N} f(X_t^i) \in U, \ \forall t \in [0, T]\} \ U = \Pi_{j=1}^{k}[0, l_j] \subset \mathbb{R}^k$$

et le comportement réel des utilisateurs est donné par la probabilité P^N sur Ω

$$P^N(.) = \mathbb{P}(X^1 \in .|C(N, U))$$

Le but étant d'obtenir des informations sur le comportement de X_t^1 sous la probabilité P^N lorsque N tend vers l'infini. En ce qui concerne les ressources utilisées, le problème est le même qu'en temps fixe.

2.2 Exemple

• Temps mobile : On se donne un réseau hiérarchisé sur d niveaux . Par exemple le réseau telephonique où l'on trouve les abonnés (en nombre N très grand) au niveau 0, les centres locaux au niveau 1, les centres à autonomie d'acheminement (CAA) au niveau 2, les centres de transit secondaires (CTS) au niveau 3, et enfin les centres de transit principaux (CTP).
Ici on fait abstraction de la topologie du réseau et on se donne seulement le nombre de ressources par niveau : un appel local nécessite seulement un canal disponible dans le commutateur de niveau 1 , on suppose qu'il y en a $l_1 N$ au total; un appel international necessite l'utilisation d'un canal au niveau 1, un canal du niveau 2 etc... Bref on suppose qu'à une communication correspond un niveau i et qu'elle utilise une unité de ressource de chaque niveau $j \leq i$. Soit $Nl_1, ..., Nl_d$ les quantités globales de ressources de chaque niveau dont dispose le réseau.
Un abonné i disposant de toutes les ressources qu'il faut est représenté par un processus de sauts X_t^i à espaces d'états $\Sigma = \{e_0, ..., e_d\}$; ces processus sont indépendants et de même loi. Il occupe l'état de repos e_0 (où il ne communique pas) pendant une durée τ_0 puis passe à l'état $e_h, h \geq 1$ (l'abonné est en communication et utilise simultanément une unité de ressources des niveaux $1, \ldots, h$) où il reste une durée τ_h puis revient à e_0; ces durées sont des v.a. exponentielles ... On prend alors comme générateur de X_t^i :

$$A(e_0, e_h) = \alpha_h, \ h \geq 1$$

$$A(e_h, e_0) = \beta_h, \ h \geq 1$$

$$A(e_h, e_k) = 0 \text{ si } h \neq k \text{ dans les autres cas}$$

La contrainte de limitation des ressources à chaque niveau s'écrit :

$$\left\{ \begin{array}{c} \sum_1^N 1_{X_t^i = e_d} \leq N l_d \\ \sum_1^N 1_{X_t^i \in \{e_{d-1}, e_d\}} \leq N l_{d-1} \\ \dots \\ \sum_1^N 1_{X_t^i \in \{e_1, \dots, e_d\}} \leq N l_1 \end{array} \right. \tag{1}$$

pour tout $t \in [0, \infty[$- pour des raisons techniques on prendra plutôt $\forall t \in [0,T], T < \infty$ cf.§4. L'événement $C(N,U)$ traduisant le fait que ces contraintes sont satisfaites s'écrit

$$C(N,U) = \{\frac{1}{N} \sum_{i=1}^N f(X^i) \in U \forall t \in [0,T]\},\ U = \prod_0^d [0, l_i], \quad f : \Sigma \longrightarrow R^d$$

$$f(e_0) = (0, \dots, 0), f(e_1) = (1, 0, \dots, 0), f(e_2) = (1, 1, 0, \dots, 0), \dots, f(e_d) = (1, \dots, 1)$$

Notons que l'on a déja

$$P(X_t^i = e_0 | C(N,U)) \geq P(X_t^i = e_0)$$

(vérification facile). Comme il se doit la probabilité pour qu'un abonné soit dans l'état de repos conditionnellement aux contraintes est plus grande que celle correspondant à des ressources infinies.

Remarque 2.2.1 Une modélisation standard de la situation est la suivante. On considère d organes de services $S_1, \dots, S_d$ (qui représentent les ressources des d niveaux). Le i-ème organe a une capacité de C_i serveurs. À l'entrée de S_1 il y a des d type de flux d'arrivées poissoniennes d'intensités $\lambda_1, \dots, \lambda_d$. Lorsqu'un client de type k arrive à S_1, soit il trouve des serveurs libres simultanément dans $S_1, \dots, S_k$ et alors il occupe un serveur de chaque niveau $i, i \leq k$ pendant une durée v.a. exponentielle de paramètre μ_k puis les libère ensuite, soit il trouve qu'un organe de service $S_i, i \leq k$ est saturé et dans se cas il est éjecté du système.
L'état du système peut alors être représenté par un $(Y_t^1, \dots, Y_t^d) \in \mathbb{N}^d$ où Y_t^i est le nombre de client dans S_i à l'instant t. C'est un processus de sauts dont le générateur est donné par :

$$\left\{ \begin{array}{c} A((x_1, \dots, x_k, \dots, x_d), (x_1 + 1, \dots, x_k + 1, x_{k+1}, \dots)) = \lambda_k \ si \ x_1 < C_1, \dots, x_k < C_k, \\ A((x_1, \dots, x_k, \dots, x_d), (x_1 - 1, \dots, x_k - 1, x_{k+1}, \dots) = \mu_k x_k \ si \ 0 < x_1 \leq C_1, \dots, 0 < x_k \leq C_k; \\ A(x,y) = 0 \text{ si } x \neq y \text{ dans les autres cas.} \end{array} \right. \tag{2}$$

Le problème serait alors de chercher la mesure invariante de ce processus.
Il faut noter que ce modèle est équivalent à celui qu'on a adopté sur le point suivant : les N processus X_t^i représentant les clients engendrent d flux d'arrivées poissoniennes avec des durées de services exponentielles. En effet si $T_k^{i,h}, k \in \mathbb{N}$ désignent les instants de sauts à l'état e_h de X_t^i, on obtient pour chaque i et h un processus de Poisson de paramètre $-A(e_h, e_h)$(puisque, pour un processus de sauts, les durées de séjour dans les états visités sont des v.a. exp. indépendantes, conditionnellement à la position de ces états). La superposition de ces instants pour les N processus donne le flux d'arrivée global de type h d'intensité $\lambda_h = -NA(e_h, e_h)$ (les X^i sont indépendants). Par ailleurs on peut considérer que le flux des demandes qui ne seront pas rejetées correspond à l'évolution des X_t^i conditionnellement à la limitation de la capacité du système. C'est précisement à cette évolution que l'on s'intéressera.

• Temps fixe : C'est le même problème que précédement sauf que les X^i sont simplement des v.a. à valeurs dans un ensemble fini; la condition $C(N, U)$ est dans ce cas

$$\left\{ \begin{array}{c} \sum_1^N 1_{X^i = e_d} \leq N l_d \\ \sum_1^N 1_{X^i \in \{e_{d-1}, e_d\}} \leq N l_{d-1} \\ \dots \\ \sum_1^N 1_{X^i \in \{e_1, \dots, e_d\}} \leq N l_1 \end{array} \right. \tag{3}$$

Ce problème sera traité au §3.2.

3 Problèmes à temps fixe

3.1 Probabilités conditionnelles et convergence

Le cadre est celui du 2.1.1: $\Sigma = \{e_0, ..., e_d\}, M_1(\Sigma) =$ l'ensemble des mesures de probabilité sur Σ. Pour une mesure $\nu \in M_1(\Sigma)$ soit sa divergence (ou son information de Kullback) par rapport à μ

$$K(\nu|\mu) = \sum_{i=0}^{d} \nu_i Log \frac{\nu_i}{\mu_i}$$

et soit le sous-ensemble de $M_1(\Sigma)$

$$\begin{array}{rcc} \Gamma & = & \{\nu \in M_1(\Sigma) : \int f d\nu \in U\} \\ & \equiv & \{(\nu_0, ..., \nu_d) \in \mathbb{R}_+^{d+1}, \sum_0^d \nu_i = 1, \ \ \sum_0^d \nu_i f(e_i) \in U\} \end{array} \tag{4}$$

La fonction $K(.|\mu)$ est strictement convexe sur Γ. Si U est un convexe fermé (ce qui est le cas au 2.1.1 et ce sera toujours le cas dans la suite) Γ l'est aussi (c'est même un compact) et $K(.|\mu)$ y admet un minimum unique que l'on notera ν^*. Dans ces conditions:

Proposition 3.1.1 *La suite des lois conditionnelles de X^1 par rapport à $C(N,U)$ soit $\nu^N \in M_1(\Sigma)$ convergent en variation*[2] *vers ν^* lorsque N tend vers l'infini.*

Preuve. Il s'agit d'une application immédiate du 'principe de conditionnement de Gibbs'. cf. Dembo-Zeitouni[5] Th.3.3.3 ou Csiszár [4] □

Remarque 3.1.2 On a par la loi des grands nombres

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \longrightarrow \bar{f} := \sum_{h=0}^{d} \mu_h f(e_h) \ \ p.s.$$

Donc la probabilité $\mathbb{P}(C(N,U))$ tend vers 1 ou 0 selon que $\bar{f}$ appartient ou non à U. Dans le premier cas on a directement $\nu_h = \mu_h, h = 0,\dots d$ sans utiliser le principe de Gibbs. Ce cas n'est pas très intéressant; il correspond à la situation où les ressources disponibles sont $\geq$ aux ressources requises par les utilisateurs en moyenne.

Introduisons maintenant quelques notations. On prend $\Omega = \Sigma^{\mathbb{N}}$, $X^i : \Omega \longrightarrow \Sigma$ les applications coordonnées, $\mathcal{B}^n = \sigma(X^1,...,X^n)$; on se donne $\mu \in M_1(\Sigma)$ et on pose $\mathbb{P} = \mu^{\otimes \mathbb{N}} \in M_1(\Omega)$
Pour tout $k \geq 1$ on considère $P_n^{(k)} \in M_1(\Sigma^N, \mathcal{B}^k)$ telle que

$$P_n^{(k)}(.) = \mathbb{P}(.|C(n,U))_{|\mathcal{B}^k}$$

Les éléments de $M_1(\Sigma^{\mathbb{N}}, \mathcal{B}^k)$ seront identifiés aux éléments de $M_1(\Sigma^k, \mathcal{B}^k)$. D'après la proposition précédente, on dispose de $P^{(1)} := \lim_{n\longrightarrow\infty} P_n^{(1)}$. Alors

Proposition 3.1.3 *1/. Pour tout k fixé, $X^1,...,X^k$ sont conditionnellement asymptotiquement indépendantes i.e. $P_n^{(k)}(.)$ converge en information (cf. §4.2) vers $P^{(1)\otimes k}$*
2/. Pour toute application $\phi : \Sigma \longrightarrow \mathbb{R}^m, m \geq 0$ et tout fermé $A \subset \mathbb{R}^m$ tel que $E^{P^{(1)}}(\phi(X^1)) \notin A$ on a

$$P_n^{(n)}\Big(\frac{1}{n}\sum_{i=1}^{n}\phi(X^i) \in A\Big) \longrightarrow 0$$

lorsque n tend vers l'infini, du moins si $\inf_{x\in A} \Lambda_\mu^*(x) > 0$, Λ_μ^* étant la transformée de Cramér associée à μ.

2. Ici cela correspond tout simplement à la distance $d(\nu,\nu') = \sup_{I\subset\Sigma} |\nu(I) - \nu'(I)| = \frac{1}{2}\sum_{e_i\in\Sigma} |\nu(e_i) - \nu'(e_i)|$

Preuve.
D'après un théorème de Csiszár [4] Th.4 $X^1, ..., X^n$ sont "asymptotiquement quasi-indépendantes sous la condition $C(n,U)$" (cf. la référence pour plus de précisions). Cela entraîne ([4] p. 772) 1/ d'une part et d'autre part que s'il existe un $\alpha > 0$, $\beta > 0$

$$\forall B_n \in \mathcal{B}^n P^{(1)\otimes n}(B_n) \leq \beta exp(-\alpha n), n = 1, 2, ...$$

alors

$$\lim_{n\infty} P_n^{(n)}(B_n) = 0$$

Or si $A \subset R^m$ est un fermé tel que $E^{P^{(1)}}(\phi(X^1)) \notin A$. Pour $B_n = \{\frac{1}{n}\sum_{i=1}^n \phi(X^i \in A\}$, on a

$$P^{(1)\otimes n}(B_n) \leq 2exp(-\alpha n), n = 1, 2, ...$$

pour $\alpha = \inf_{x\in A} \Lambda_\mu^*(x)$ (cf. le théorème de Cramér-Chernov). 2/ s'en déduit. □

Proposition 3.1.4 *La suite* $\mathbb{P}_n(.) = \mathbb{P}(.|C(n,U))$ *converge étroitement vers* $P^{(1)\otimes\mathbb{N}}$.

Preuve.
On sait que les probabilités marginales de $\mathbb{P}_n$ convergent étroitement vers celles de $P^{(1)\otimes\mathbb{N}}$ (d'après le 1/ de la proposition précédente, la convergence en information entraînant la convergence en variation). D'autre part, en munissant $\Sigma^{\mathbb{N}}$ de la metrique induite par la metrique usuelle de $\mathbb{R}^{\mathbb{N}}$ (après avoir identifier Σ à $\{0, 1, ..., d\} \subset \mathbb{R}$ par exemple), c'est un compact cf. Billingsley [2] p.219. Il en résulte que la suite $(P_n)_{n\geq 1}$ est tendue; on conclut en appliquant le théorème de Prohorov (cf.[2]) □

3.2 Application : l'exemple des systèmes hiérarchiques

On se place dans le cadre du paragraphe 2.2 alinéa "temps fixe". Pour calculer la probabilité conditionnelle limite, on est donc amené à résoudre le problème d'optimisation suivant

$$\begin{cases} \inf_\nu K(\nu|\mu) = \sum_0^d \nu_h Log\frac{\nu_h}{\mu_h} \\ g(\nu) \in A \end{cases} \tag{5}$$

où

$$g(\nu_0, ..., \nu_d) = (\sum_{h=0}^d \nu_h, \sum_{h=1}^d \nu_h, \ldots, \nu_d)$$

$$A = \{1\} \times \{(x_1, ..., x_d) \in R_+^d : x_i \leq l_i, i = 1, ...d\}$$

On peut réécrire ce problème sous la forme standard d'un problème d'optimisation convexe et on dispose de $\lambda = (\lambda_0, ..., \lambda_d), \lambda_i \geq 0 \; si \;\; i \geq 1$ (multiplicateurs de Lagrange) tels que:

$$\nu_0^* = \mu_0 e^{-\lambda_0 - 1}$$

$$\nu_i^* = \mu_i e^{-\sum_0^i \lambda_i - 1}$$

Avec

$$\lambda_i = 0 \quad ou \;\; bien \quad \sum_{j=1}^{i} \nu_j^* = 0 \;\; i = 1, ..., d.$$

En introduisant Z à la place de λ_0 :

$$Z = e^{-\lambda_0 - 1}$$

("fonction de partition"), la solution vérifie:

$$\begin{cases} \nu_0^* = \dfrac{\mu_0}{Z} \\ \nu_i^* = \dfrac{\mu_i}{Z} e^{-\sum_{j=1}^{i} \lambda_j} \\ \lambda_i = 0 \;\; \text{ou bien} \;\; \sum_{j=i}^{d} \nu_j^* = l_i \;\; i = 1, ..., d. \\ Z = \mu_0 + \sum_{i=1}^{d} \mu_i e^{-\sum_{j=1}^{i} \lambda_j} \end{cases} \tag{6}$$

• Intuitivement les probabilités μ_i peuvent être interprétées comme étant les proportions des gens qui seraient en communication à un niveau i si les ressources étaient infinies; les probabilités ν_i^* quant à elles peuvent être vues comme les proportions de gens qui sont en communication à un niveau i compte tenu des contraintes de limitation des ressources. On peut alors voir le rapport $\nu_i^* / \mu_i \leq 1$ comme étant la proportion de gens ayant reussi à communiquer au niveau i et $1 - \nu_i^* / \mu_i$ la proportion de gens n'ayant pas réussi à communiquer à ce niveau.
En effet, si l'on considère la fonction $\phi_h : \Sigma \longrightarrow R \;\; e_i \longrightarrow \delta_{hi}$ alors $\frac{1}{n} \sum_{i=1}^{n} \phi_h(X^i)$ qui représente la proportion des ressources de type h utilisée tend, conditionnellement à la contrainte de limitation des ressources vers ν_h^* au sens de la propostion 3.1.3. Ainsi la quantité ν_h^* peut être interprétée comme la proportion des ressources de type h *effectivement* utilisée.
• Les ressources disponibles ne sont pas intégralement utilisées, bien qu'elles soient inférieures en quantité aux ressources requises. En effet elles seront utilisées dans leur intégralité si $E^{\nu^*} f(X^1) := \sum_{h=1}^{d} \nu_h^* f(e_h) = (l_1, \ldots, l_d)$ ce qui n'est pas vrai en général. Par contre on a $E^{\nu^*} f(X^1) \in \partial U, \; U = \prod_1^d [0, l_h]$ car $K(\nu|\mu)$ est strictement convexe et son minimum absolu qui est atteint en μ qui vérifie $E^{\mu}(f(X^1) \notin U$ comme on le suppose.

4 Problèmes à temps mobile

4.1 Cas des chaînes de Markov (temps discret)

Ce paragraphe peut être considéré comme une introduction heuristique au cas des processus de sauts à temps continu.
Soit $T \in \mathbb{N}$. Ici le temps t décrit les entiers $\leq T$. Soit $X_t^i, i \in \mathbb{N}$ une suite de chaînes de Markov indépendantes de même loi (même matrice de transition Q et probabilié initiale π) et dont l'espace des états est l'ensemble fini $\Sigma = \{e_0, ..., e_d\}$.. Ces chaines sont définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$
Soit enfin une fonction:

$$f : \Sigma \longrightarrow \mathbb{R}^k$$

et U un convexe de $\mathbb{R}^k$.
On considère l'événement

$$C(N, U) = \{\frac{1}{N}\sum_{i=1}^{N} f(X_t^i) \in U, \ \forall t = 0, ..., T\}$$

(qui représente la contrainte) et on s'intéresse à la limite de

$$P^N(.) = \mathbb{P}(X \in .|C(N, U))$$

lorsque N tend vers l'infini.
Dans ce cas on peut toujours appliquer le principe de Gibbs pour les mesures à support fini. Posons en effet:

$$X^i = (X_0^i, ..., X_T^i)$$

les X^i sont alors i.i.d à valeurs dans Σ^{T+1} avec

$$\mathbb{P}(X^i = (x_0, ..., x_T)) = \pi(x_0)Q(x_0, x_1)....Q(x_{T-1}, x_T)$$

et l'on peut écrire

$$C(N, U) = \{\frac{1}{N}\sum_{i=1}^{N} F(X^i) \in U^{T+1}\}, \ U^{T+1} \subset \mathbb{R}^{k(T+1)}$$

avec

$$F : \Sigma^k \longrightarrow \mathbb{R}^{k(T+1)} \ F(x_0, ..., x_T) = (f(x_0), ..., f(x_T)), \ x_i \in S_0$$

Notre but est d'obtenir une expression explicite de la limite de $P^N(.)$ qu'on notera $P_T^*(.)$.
La loi commune des X^i sera notée $\mu \in M_1(\Sigma^{T+1})$. La loi limite conditionnelle cherchée, soit $\mu^* \in M_1(\Sigma^{T+1})$, est donnée par

$$\mu^* = \arg\min_{\nu \in \Gamma} K(\nu|\mu)$$

$$\Gamma = \{\nu \in M_1(\Sigma^{T+1}) : \int F d\nu \in U^T \subset \mathbb{R}^{k(T+1)}\},$$

$$K(\nu|\mu) = \sum_{x \in \Sigma^{T+1}} \nu(x) Log \frac{\nu(x)}{\mu(x)}$$

Plus explicitement on a à résoudre le problème d'optimisation suivant:

$$\left\{ \begin{array}{l} \inf_{\nu \in M_1(\Sigma^{T+1})} \sum_{x_0,\ldots,x_T} \nu(x_0,\ldots,x_T) Log \frac{\nu(x_0,\ldots,x_T)}{\mu(x_0,\ldots,x_T)} \\ \sum_{x_0,\ldots,x_T} f(x_0)\nu(x_0,\ldots,x_T) \in U \subset \mathbb{R}^k \\ \sum_{x_0,\ldots,x_T} f(x_1)\nu(x_0,\ldots,x_T) \in U \subset \mathbb{R}^k \\ \ldots \\ \sum_{x_0,\ldots,x_T} f(x_T)\nu(x_0,\ldots,x_T) \in U \subset \mathbb{R}^k \end{array} \right. \qquad (7)$$

Ce problème d'optimisation admet une solution unique, la fonction à optimiser étant strictement convexe et le domaine est convexe. On dispose alors de multiplicateurs de Lagrange $\lambda_0, \ldots, \lambda_T \in \mathbb{R}^k$ tels que

$$\frac{\mu^*(x_0,\ldots,x_T)}{\mu(x_0,\ldots,x_T)} = \frac{e^{\sum_{t=0}^{T} \lambda_t . f(x_t)}}{Z_T}$$

(Z_T est un facteur de normalisation qui correspond à la contrainte $\sum \nu(x_0,\ldots,x_T) = 1$). En écrivant $\mu(x_0,\ldots,x_T)$ sous forme de produit des probabilités de transitions, il est facile de voir que μ^* se factorise en un produit de probabilités de transitions $Q^*_t(x_t, x_{t+1})$ qui dépendent donc du temps, de sorte que sous P^*_T, X^1_t est une chaîne de Markov non homogène.

Remarque 4.1.1 On ne peut pas prendre T infini; l'événement par lequel on conditionnerait pourrait être de probabilité nulle. Une condition suffisante pour que les événements de type $C(N,U)$ soient non négligeables est que $f(X^1_t)$ reste dans U pour $t = 0,\ldots T$ avec une probabilité non nulle; ce qui est vrai dans le cas typique où $U = \prod_{h=1}^{k}[0, l_h]$ et $f(e_0) = 0$

4.2 Cas des processus de sauts (temps continu)

On se donne cette fois un intervalle $[0,T] \subset \mathbb{R}$ et une suite de processus markoviens de sauts $X^i_t, t \in [0,T], i = 1,\ldots,$ dont l'espace des états $\Sigma = \{e_0,\ldots,e_d\}$ est fini, définis sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On suppose que ces processus sont indépendants et ont la même loi (générateur A et loi initiale π). Comme au paragraphe précédent, on se donne un convexe $U \subset \mathbb{R}^k$ et une application $f : \Sigma \longrightarrow \mathbb{R}^k$, et on considère l'événement

$$C_0(N,U) = \{\frac{1}{N}\sum_{i=1}^{N} f(X_t^i) \in U, \ \forall t \in [0,T]\}$$

On s'intéresse alors à la limite de

$$P^N(.) = \mathbb{P}(X^1 \in .|C_0(N,U))$$

lorsque N tend vers l'infini. Dans la suite on notera P_X la loi commune des X^i sous $\mathbb{P}$.

Remarque 4.2.1 Ici encore on ne peut pas prendre T infini. En revanche une condition suffisante pour que $C(N,U)$ soit non négligeable est $P(f(X_t^1) \in U \ \forall t \in [0,T]) \neq 0$. En particulier s'il existe un état e_i tel que $f(e_i) \in U$ cette condition est realisée puisque X_t^1 reste $= e_i$ pour tout $t \in [0,T]$ avec une probabilité non nulle.On se placera dans ce cas dans toute la suite.

D'après l'expression de la probabilité limite pour les chaînes de Markov, on peut s'attendre à ce que cette limite en temps continue soit donnée par une formule du type

$$P^*(X_.^1 \in dx_.) = \frac{e^{\int_0^T \lambda_t . f(x_t)dt}}{Z} P_X(X_.^1 \in dx_.)$$

Rappelons d'abord une définition. Soit P_n une suite de probabiliés sur un espace $(\mathcal{S},\mathcal{B})$. On dit que P_n converge en information (ou au sens de Csiszár) vers une probabilié $Q \in M_1(\mathcal{S})$ si $K(P_n|Q) \longrightarrow 0$. Avec

$$K(P|Q) = \int_{\mathcal{S}} Log\frac{dP}{dQ}dP \ \text{ si } P \ll Q, \text{ et } \infty \text{ sinon.}$$

Cette convergence entraîne la convergence en variation de P_n vers Q, et on a même (cf. [4]):

$$|P-Q| \leq \sqrt{2K(P,Q)} \ \ \forall P,Q \in M_1(\mathcal{S})$$

On a alors le résultat général suivant sur les lois conditionnelles limites (cf.[4] Th.4.):

Théorème 4.2.2 (Csiszár) *Soit Ψ une application mesurable d'un ensemble (mesurable) $(S,\mathcal{S})$ vers un espace vectoriel topologique localement convexe (e.v.t.l.c.) V. Soient $X^1, X^2, ...$ une suite i.i.d de v.a. à valeur dans S de loi commune P_X et Q la probabilité*

image de P_X par Ψ; Q est supposée "convexe-tendue"[3].
Alors pour tout $C \subset V$, convexe dont l'intérieur a une intersection non vide avec le support de Q la suite des probabilités conditionnelles

$$P_X^N(.) = P(X^1 \in .|\frac{1}{N}\sum_{i=1}^{N}\Psi(X^i) \in C)$$

converge en information vers la mesure de probabilité P^* définie par

$$\frac{dP^*}{dP_X}(.) = \frac{1}{Z}exp\mathcal{V}^*(\Psi(.))$$

avec

$$\mathcal{V}^* = \arg\max_{\mathcal{V}\in V'}[\inf_{v\in C}\mathcal{V}(v) - LogE^Q(exp(\mathcal{V}(.)))]$$

(ce maximum étant atteint)

Bien que ce résultat soit très général, les conditions imposées à C et à Q constituent des obstructions importantes à son utilisation dans le cas qui nous intéresse. On se place dans le cadre de ce théorème en prenant

$$S = \{\omega : [0,T] \longrightarrow \Sigma \text{ constante par morceaux}\}$$

$$\Psi : S \longrightarrow V \quad \Psi(\omega)(t) = f(\omega(t)), \; C = \{\phi \in V : \phi(t) \in U \; \forall t \in [0,T]\}$$

où V est un e.v.t.l.c $\subset (\mathbb{R}^k)^{[0,T]}$ à préciser. Les choix de V que l'on considère ici tournent autour des espaces $L^p_{([0,T],R^k)}$. Ils ont l'avantage d'avoir des espaces duaux bien connus. Cependant, si on adopte $V = L^p_{([0,T],R^k)}$ avec $p < \infty$, l'intérieur des ensembles C du type ci-dessus serait vide. D'autre part si on prenait $V = L^\infty_{([0,T],R^k)}$, il est facile de voir que la loi image Q du processus X_t^1 n'est pas séparable (i.e. son support n'est pas séparable pour la topologie considérée) et donc non tendue[4](cf. Billingsley[2] Appendice III). La solution retenue pour l'instant consiste à travailler sur un espace L^p, en modifiant légèrement la forme des contraintes $C_0(N,U)$ par lesquelles on conditionne : on parlera de contraintes de type L^p; étant entendu que notre problème initial est celui des contraintes de type L^∞.
Pour simplifier la présentation, on se placera désormais dans le cas $k = 1, T = 1$ avec U de la forme $[0,E]$; la fonction f est généralement positive (à valeurs dans $\mathbb{R}_+^k$ pour $k \geq 1$): elle donne les quantités de ressources utilisées pour chaque état e_i ou l'énergie en physique statistique. L'hypothèse de l'existence d'un $e_i \in \Sigma$ tel que $f(e_i) \in U$ se traduit ici par $f(e_i) \leq E$. La traduction des résultats lorsque $k > 1$ et U est un convexe de $\mathbb{R}^k$ ne pose pas de difficulté.

3. i.e. il existe une suite d'ensembles $K_1 \subset K_2, \ldots$ convexes et compacts tels que $\lim_{n\longrightarrow\infty} Q(K_n) = 1$. C'est toujours vrai si V est un espace de Banach *séparable*.

4. si Q était tendue on aurait supp$Q \subset \cup K_n$ où K_n sont des compacts. Dans ce cas les K_n seraient séparables car ce sont des espaces compacts metrisables et donc supp$Q$ le serait aussi.

4.2.1 Cas des contraintes de type L^p, $p < \infty$.

Posons :

$$C(N,p) = \{\frac{1}{N}\sum_{i=1}^{N} f \circ X^i \in B_p(E)\}$$

où $B_p(E)$ est la boule fermée de rayon E de $L^p_{([0,1],\mathbb{R})}$. Notons que $B_p(E) \subset B_{p'}(E)$ si $p' \leq p$ (inégalité de Hölder), et donc $C(N,p) \subset C(N,p')$ si $p' \leq p$. Remarquons aussi que $C_0(N,U) = C(N,\infty)$ et il est facile de voir que $\bigcap_p C(p,N) = C(N,\infty)$. D'autre part on a

$$\forall \epsilon > 0 \ \exists p \in \mathbb{N} \ \forall \phi \in B_p(E) \ \exists I_\epsilon \subset [0,1] \ : dt(I_\epsilon) < \epsilon, \ \phi(t) \leq E \ \forall t \in [0,1] - I_\epsilon$$

(vérification facile). Ces considérations montrent que le fait de conditionner par des événements $C(p,N)$ avec p suffisament grand donnerait une idée sur la loi limite conditionnelle qu'on cherchait au départ (si elle existe). Pour $p \in \mathbb{N}$ posons alors

$$P_p^N(.) = \mathbb{P}(X^1 \in .|C(N,p))$$

Proposition 4.2.3 *1. La suite des probabilités P_p^N converge en information vers une probabilité P_p^* donnée par :*

$$\frac{dP_p^*}{dP_X}(\omega) = \frac{e^{\int_0^1 f(\omega(t))\lambda_p(t)dt}}{Z_p}, \ Z_p = E^{P_X}(e^{\int_0^1 f(X_t^1)\lambda_p(t)dt}) \tag{8}$$

où $\lambda_p \in L^q([0,1])$, $(1/p + 1/q = 1)$ est l'unique solution du problème d'optimisation suivant :

$$\lambda_p = \arg \max_{\lambda \in L^q([0,1])} [\inf_{\phi \in B_p(E)} \int_0^1 \phi(t)\lambda(t)dt - Log E^{P_X}(exp(\int_0^1 f(X_t^1)\lambda(t)dt))] \tag{9}$$

2.Le processus X_t^1, $t \in [0,T]$ est markovien (non homogène) sous la probabilité P_p^*.

Preuve.
1.La première partie est une application du théorème de Csiszár avec ici $V = L^p_{([0,1])}$qui est un espace de Banach séparable, $C = B_p(E)$ dont l'intérieur a une intersection non vide avec le support de la loi image de X^1 par Ψ si on suppose qu'il existe un $e_i \in \Sigma$: $f(e_i) \leq E$ par exemple. Les conditions du théorème sont vérifiées dans ce cas.
2.Rappelons que si P et Q sont deux probabilités sur un même espace $(\Omega, \mathcal{A})$ telles que $dQ/dP = M$ alors pour toute sous tribu $\mathcal{B}$ de $\mathcal{A}$ et toute v.a. Φ on a

$$E^Q(\Phi|\mathcal{B}) = \frac{E^P(\Phi M|\mathcal{B})}{E^P(M|\mathcal{B})} \tag{10}$$

Pour $t \in [0,T]$ désignons par $\mathcal{F}_{\leq t} = \sigma(X_s^1, s \leq t)$,$\mathcal{F}_{\geq t} = \sigma(X_s^1, s \geq t)$ et posons

$$M_T(x.) = \frac{e^{\int_0^T f(x_s)\lambda_p(s)ds}}{Z_p},\; M_{\leq t}(x.) = e^{\int_0^t f(x_s)\lambda_p(s)ds},\; M_{\geq t}(x.) = e^{\int_t^T f(x_s)\lambda_p(s)ds}$$

Soit $Y_{\leq t}, Y_{\geq t}$ deux variables aléatoires mesurables par rapport à $\mathcal{F}_{\leq t}$ et $\mathcal{F}_{\geq t}$ respectivement. Alors en désignant par $E(.)$ et $E^*(.)$ les espérances par rapport aux probabilités P_X et P_X^* respectivement, on a grâce à la formule rappelée ci-dessus (remarquer qu'on a simplifier par Z_p dans la deuxième égalité)

$$\begin{aligned} E^*(Y_{\leq t}Y_{\geq t}|X_t^1) &= \frac{E(Y_{\leq t}Y_{\geq t}M_T|X_t^1)}{E(M_T|X_t)} && (11) \\ &= \frac{E(Y_{\leq t}Y_{\geq t}M_{\leq t}M_{\geq t}|X_t^1)}{E(M_{\leq t}M_{\geq t}|X_t^1)} && (12) \end{aligned}$$

et puisque X_t^1 est markovien sous P_X et compte tenu de ce que $M_{\leq t}$ et $M_{\geq t}$ sont mesurables resp. par rapport $\mathcal{F}_{\leq t}$ et $\mathcal{F}_{\geq t}$, cette derniére quantité est

$$\begin{aligned} &= \frac{E(Y_{\leq t}M_{\leq t}|X_t^1)E(M_{\geq t}Y_{\geq t}|X_t^1)}{E(M_{\leq t}|X_t^1)E(M_{\geq t}^1|X_t^1)} && (13) \\ &= \frac{E(Y_{\leq t}M_{\leq t}|X_t^1)E(M_{\geq t}|X_t^1)}{E(M_{\leq t}|X_t^1)E(M_{\geq t}|X_t^1)} \frac{E(Y_{\geq t}M_{\geq t}|X_t^1)E(M_{\leq t}|X_t^1)}{E(M_{\geq t}|X_t^1)E(M_{\leq t}|X_t^1)} && (14) \\ &= \frac{E(Y_{\leq t}M_T|X_t^1)}{E(M_T|X_t^1)} \frac{E(Y_{\geq t}M_T|X_t^1)}{E(M_T|X_t^1)} && (15) \end{aligned}$$

où dans le dernier passage on a utilisé encore une fois la propriété markovienne de X_t^1 sous P_X et divisé les numérateurs et dénominateurs des deux fractions par Z_p.
On applique une deuxième fois 10 et l'on obtient

$$E^*(Y_{\leq t}Y_{\geq t}|X_t^1) = E^*(Y_{\leq t}|X_t^1)E^*(Y_{\geq t}|X_t^1)\square$$

Dans la suite de ce paragraphe on fixe p et on pose $V_p = L^p_{([0,T])}$, $P^* := P_p^*$, $\lambda := \lambda_p$

Proposition 4.2.4 *Si f est à valeurs dans $\mathbb{R}_+$ (ce qui est le cas aussi bien pour l'application qui nous intéresse qu'en physique statistique) alors*

$$\lambda(t) \leq 0 \;\; dt - p.p. \; sur[0,T]$$

La signification physique de cette proposition est claire : la probabilité d'une trajectoire est d'autant plus faible que les états occupés dans cette trajectoire ont des énergies élevées.

Preuve.
Pour $\beta \in V_p^{'}$ notons

$$H(\beta) = \inf_{\phi \in C} \int_0^T \phi(t)\beta(t)dt - LogE^{P_X} exp \int_0^T f(X^1(t)\beta(t)dt$$

où $C = B_p(E)$. Par définition on a

$$\lambda = \arg\max_{\beta \in V_p^{'}} H(\beta)$$

On va montrer que $\forall \beta \in V_p^{'}$ il existe $\beta_1 \in V_p^{'}$ négative telle que $H(\beta_1) \geq H(\beta)$. Soit donc $\beta \in V_p^{'}$ et soit $supp_\beta^-$ (resp. $supp_\beta^+$) le plus grand sous ensemble de $[0,T]$ sur lequel β est négative (resp. positive) i.e. $\forall \phi \in V_p \; supp(\phi) \subset supp_\beta^- \Longrightarrow \beta(\phi) \leq 0$ (resp. $\forall \phi \in V_p \; supp(\phi) \subset supp_\beta^+ \Longrightarrow \beta(\phi) \geq 0)$. Soit $\beta_1 \in V_p^{'}$ définie par

$$\int_0^T \phi(t)\beta_1(t)dt = \int_0^T 1_{supp_\beta^-}\phi(t)\beta(t)dt - \int_0^T 1_{supp_\beta^+}\phi(t)\beta(t)dt.$$

Il est clair que β_1 est négative. Alors

$$\inf_{\phi \in C} \int_0^T \phi(t)\beta(t)dt = \inf_{\phi \in C} \int_0^T \phi(t)\beta_1(t)dt \tag{16}$$

En effet, $\forall \phi \in C$ il existe $\phi_1 \in C$ définie par $\phi_1(t) = \phi(t)$ si $t \in supp_\beta^-$ et $\phi_1(t) = -\phi(t)$ si $t \in supp_\beta^+$ telle que

$$\int_0^T \phi(t)\beta(t)dt = \int_0^T \phi_1(t)\beta_1(t)dt.$$

D'autre part compte tenu de ce que f est positive on a

$$LogE^{P_X} exp \int_0^T f(X_t^1)\beta(t)dt \geq LogE^{P_X} exp \int_0^T f(X_t^1\beta_1(t)dt \tag{17}$$

16 et 17 entraînent bien que $H(\beta_1) \geq H(\beta)$. $\square$

•Semi-groupe et générateurs de X_t^1 sous P^*

Les démonstrations des deux propositions qui suivent sont données en Appendice.

Proposition 4.2.5 *Les générateur A_t^* du processus X_t^1 sous P_X^* sont donnés par*

$$A_t^*(x,y) = A(x,y)\frac{E(e^{\int_t^T f(X_s^1)\lambda(s)ds}|X_t^1 = y)}{E(e^{\int_t^T f(X_s^1)\lambda(s)ds}|X_t^1 = x)} \text{ pour } x \neq y$$

• Semi-groupe de X_t^1 sous P^*: Il suffit de calculer les probabilités $P^*(X_t^1 = x)$. On a

$$\begin{aligned} P^*(X_t = x) &= \frac{E(1_{X_t=x} e^{\int_0^T \lambda(s).f(X_s)ds})}{E(e^{\int_0^T \lambda(s).f(X_s)ds})} \\ &= \frac{E(e^{\int_0^T \lambda(s).f(X_s)dt} | X_t = x) P_X(X_s = x)}{E(e^{\int_0^T \lambda(s).f(X_s)ds})} \\ &= \frac{E(e^{\int_0^t \lambda(s).f(X_s)ds} | X_t = x) E(e^{\int_t^T \lambda(s).f(X_s)ds} | X_t = x)}{E(e^{\int_0^T \lambda(s).f(X_s)ds})} \times P_X(X_t = x) \end{aligned} \tag{18}$$

La proposition suivante fournit un moyen pour le calcul des termes intervenant dans l'expression du semi-groupe ou des générateurs.

Proposition 4.2.6 *En tout point $t \in [0, T]$ où λ est continue, les fonctions $h_x(t) = E^{P_X}(e^{\int_t^T f(X_s^1)\lambda(s)ds} | X_t^1 = x)$ et $k_x(t) := E^{P_X}(e^{\int_0^t f(X_s^1)\lambda(s)ds} | X_t^1 = x)$ satisfont les équations suivantes :*

$$\frac{dh_x(t)}{dt} = -f(x).\lambda(t) h_x(t) - \sum_y A(x, y) h_y(t)$$

$$\frac{dk_x(t)}{dt} = (-\frac{P'(t, x)}{P(t, x)} - \lambda(t).f(x)) k_x(t) + \frac{1}{P(t, x)} \sum_y A(y, x) P(t, y) k_y(t)$$

avec $P(t, x) = P_X(X_t^1 = x)$.

Enfin la (moyenne de la) quantité des ressources réellement utilisées à un instant $t \in [0, T]$ est donnée par

$$E^*(f(X_t^1)) = \sum_{h=0}^{d} f(e_h) P^*(X_t^1 = e_h)$$

4.2.2 Quelques remarques sur le cas des contraintes L^∞

Ici on suppose que f est à valeurs positives. On prend $E = 1$ (il suffit de poser $f := f/E$) et on suppose qu'il existe $e_i \in \Sigma$ tel que $f(e_i) < E = 1$. On notera aussi $V = L^\infty_{([0,1],\mathbb{R})}$ et B_1 la boule fermé unité de V et $V_p = L^p_{([0,1],\mathbb{R})}$.

Remarque 4.2.7 Pour tout p fixé P_p^n converge en variation vers $P_p^* \in M_1(S)$ lorsque n tend vers l'infini. D'autre part puisque $\bigcap_p C(n, p) = C(n, U)$, on a la convergence en variation de P_p^n vers P^n lorsque $p \longrightarrow \infty$ pour tout n fixé. Compte tenu de ce que $M_1(S)$ est complet pour la distance variationnelle (cf. Neveu[9]), si on arrive à montrer que l'une des convergences est uniforme, on conclura à l'existence de la limite de P^n qu'on cherchait au départ (théorème de la double limite). La question reste ouverte.

Si la limite de $P^n(.) = \mathbb{P}(.|C(n,U))$ existe (ce qui est parfaitement plausible), d'après la proposition 4.2.3 on peut s'attendre à ce qu'elle soit donnée par une formule du type 8 :

$$\frac{dP^*}{dP_X}(\omega) = \frac{e^{\int_0^1 f(\omega(t))d\lambda(t)}}{Z}$$

où $\lambda \in V^{'}$ (mesure de Radon) est donnée par

$$\lambda = \arg\max_{\alpha\in V^{'}} H(\alpha) = [\inf_{\phi\in B_1}\int_0^1 \phi(t)d\alpha(t) - LogEe^{\int_0^1 f(X_t)d\alpha(t)}]$$

Les éléments de $V^{'}$ s'expriment sous la forme d'une intégrale (de Radon, cf. Dunford-Schwartz[6] ou Yosida[16]): $\alpha(\phi) = \int_0^1 \phi(t)d\alpha(t),\ \ \phi \in V$; elles s'identifient aux mesures sur $[0,1]$ (additives seulement!) et absolument continues par rapport à la mesure de Lebesgue; on notera aussi $\alpha(I) := \int_0^1 1_I d\alpha(t)$ pour un borélien $I \subset [0,1]$. On a alors le résultat suivant:

Remarque 4.2.8 *La fonctionnelle*

$$J(\alpha) = -H(\alpha) = -\inf_{\phi\in B_1}\int_0^1 \phi(t)d\alpha(t) + LogEe^{\int_0^1 f(X_t)d\alpha(t)}$$

admet un minimum unique sur $V^{'}$.

Preuve.

J est définie sur $V^{'}$ qui sera muni de la topologie faible usuelle.

• Existence:

(i) J est semi-continue inférieurement(s.c.i.) : en effet on a d'une part $\alpha \longrightarrow -\int_0^1 \phi(t)d\alpha(t)$ est continue pour toute $\phi \in V$, ce qui entraîne que $\alpha \longrightarrow \sup_{\phi\in V} -\int_0^1 \phi(t)d\alpha(t)$ est s.c.i. (l'enveloppe supérieure d'une famille de fonctions continues est s.c.i.). D'autre part il faut noter que $\inf_{\alpha\in V^{'}} J(\alpha) = \inf_{\alpha\in K} J(\alpha)$ où K est l'ensemble des formes linéaires continues sur V qui sont négatives i.e. $\alpha(\phi) \leq 0$ si $\phi \geq 0$ (cf. la proposition 4.3.2 dont la vérification s'applique également au cas présent)et alors il est facile de voir que $\alpha \longrightarrow LogEe^{\int_0^1 f(X_t)d\alpha(t)}$ est continue sur K (en utilisant le théorème de convergence dominée, ici f est positive). D'où le résultat.

(ii) $J(\alpha) \longrightarrow +\infty$ si $\|\alpha\| \longrightarrow \infty$ avec $\|\alpha\| = \sup_{\|\phi\|_\infty=1}\int_0^1 \phi(t)d\alpha(t)$. En effet

$$\begin{aligned} J(\alpha) &= \sup_{\phi\in B_1}(-\int_0^1 \phi(t)d\alpha(t)) + LogEe^{\int_0^1 f(X_t)d\alpha(t)} \\ &= \|\alpha\|(1 + \frac{LogEe^{\int_0^1 f(X_t)d\alpha(t)}}{\|\alpha\|}) \end{aligned}$$

Par hypothèse, on a $f(e_0) < E = 1$. Soit $\epsilon = \mathbb{P}(X_t^1 = e_0, \forall t \in [0,1])$; on sait que $\epsilon > 0$ et on a

$$LogEe^{\int_0^1 f(X_t)d\alpha(t)} \geq Log\epsilon e^{\int_0^1 f(e_0)d\alpha(t)}$$

et donc

$$\begin{aligned} J(\alpha) &\geq \|\alpha\|(1+\frac{Log\epsilon + f(e_0)\int_0^1 d\alpha(t)}{\|\alpha\|}) \\ &\geq \|\alpha\|(1+\frac{Log\epsilon}{\|\alpha\|} - f(e_0)) \end{aligned}$$

Comme $f(e_0) < 1$ on en déduit que $J(\alpha)$ tend vers $+\infty$ lorsque $\|\alpha\| \longrightarrow \infty$.
(iii) Soit α_n une suite minimisante de J i.e. $J(\alpha_n) \longrightarrow \inf_{\alpha\in V'} J(\alpha)$. D'après (ii) α_n est bornée; on dispose alors d'une valeur d'adhérence λ pour cette suite (théorème d'Alaoglu-Bourbaki). Comme J est s.c.i. (pour la topologie faible), on a $\forall\epsilon > 0$ il existe un voisinage A de λ tel que

$$J(\alpha) \geq J(\lambda) - \epsilon \ \ \forall\alpha \in A.$$

λ étant une valeur d'adhérence de (α_n), A contient une infinité d'éléments de la suite et on en déduit que

$$\limsup_n J(\alpha_n) \geq J(\lambda) - \epsilon,$$

et comme $\epsilon > 0$ est arbitraire

$$\limsup_n J(\alpha_n) \geq J(\lambda),$$

mais $\limsup_n J(\alpha_n) = \lim_n J(\alpha_n) = \inf_{\beta\in V'} J(\beta)$, ce qui entraîne que $J(\lambda) \leq \inf_{\beta\in V'} J(\beta)$ et par conséquent $J(\lambda) = \inf_{\beta\in V'} J(\beta)$. Le minimum est donc atteint[5] en λ.
• Unicité : J est strictement convexe: on a d'abord la convexité de $\alpha \longrightarrow sup_{\phi\in V} - \int_0^1 \phi(t)d\alpha(t)$ (envellope supérieure de fonctions convexes) puis $\alpha \longrightarrow LogEe^{\int_0^1 f(X_t)d\alpha(t)}$ est strictement convexe; en effet Soient $\lambda \in]0,1[$ et $\alpha_1, \alpha_2 \in V'$. Par l'inégalité de Hölder on vérifie aisément que

$$E[e^{\lambda\int_0^1 f(X_t)d\alpha_1(t)} e^{(1-\lambda)\int_0^1 f(X_t)d\alpha_2(t)}] \leq Ee^{\lambda\int_0^1 f(X_t)d\alpha_1(t)} Ee^{(1-\lambda)\int_0^1 f(X_t)d\alpha_2(t)} \tag{19}$$

en composant par Log, on obtient la convexité de $\alpha \longrightarrow LogEe^{\int_0^1 f(X_t)d\alpha(t)}$. La stricte convexité tient à ce que l'égalité dans 19 n'est possible que si $e^{\int_0^1 f(X_t)d\alpha_1(t)} = \delta e^{\int_0^1 f(X_t)d\alpha_2(t)}$ p.s., soit $\int_0^1 f(X_t)d\alpha_1(t) = \int_0^1 f(X_t)d\alpha_2(t) + Log\delta \ \ P_X - p.s.$; ce qui entraîne d'abord que $\delta = 1$ puis que $\alpha_1 = \alpha_2$. En effet si pour $\beta \in V'$ on a $\int_0^1 f(X_t)d\beta(t) = 0 \ P_X - p.s.$, il est facile de vérifier que pour tout intervalle $I \subset [0,1]$ on a $\beta(I) = 0$ (on suppose évidement qu'il existe un $e_i \in \Sigma$ tel que $f(e_i) \neq 0$), puis en utilisant le fait que β est absolument continue par rapport à la mesure de Lebesgue, $\beta(B) = 0$ pour tout borélien $B \subset [0,1]$; et donc $\beta \equiv 0$. Ceci achève la démonstration. □

5. Ici on a un problème d'optimisation sur un espace dual (V'); on n'a pas besoin de l'hypothèse de réflexivité de l'espace (non vérifiée dans ce cas) qui est généralement requise par les théorèmes d'existence des minima. D'autre part on a évité le passage par l'extraction d'une sous suite de α_n qui n'est pas nécessairement licite dans V'.

Remarque 4.2.9 *La suite des $\lambda_p \in V_p^{'} \subset V^{'}$ solutions de 9 est bornée dans $V^{'} = (L^{\infty}_{([0,1])})^{'}$ (qui est muni de la norme usuelle: $\|\lambda\| = \sup_{\|\phi\|_\infty \leq 1} \lambda(\phi)$).*
Preuve.
Posons $V_p = L^p_{([0,1])}$ et

$$J_p(\lambda) = \|\lambda\|_q + LogEe^{\int_0^1 f(X_t)\lambda(t)dt}, \ \lambda \in V_p, \ 1/p + 1/q = 1$$

Par définition $\lambda_p = \inf_{\lambda \in V_p^{'}} J_p(\lambda)$. Supposons que λ_p n'est pas bornée pour la norme$\|.\|$ de $V^{'}$; il existe alors une suite p_n telle que $\|\lambda_{p_n}\| \longrightarrow \infty$, ce qui entraîne que $\|\lambda_{p_n}\|_{q_n} \longrightarrow \infty$ (puisque $\|\lambda\|_q \geq \|\lambda\| \ \forall q \geq 1, \ \lambda \in V_p$). Et alors il est facile de voir que $J_{p_n}(\lambda_{p_n}) \longrightarrow \infty$. En effet, comme au (ii) de la remarque précédente on a

$$\begin{aligned} J_{p_n}(\lambda_{p_n}) &\geq \|\lambda_{p_n}\|_{q_n}(1 + \frac{Log\epsilon + f(e_0)\int_0^1 \lambda_{p_n}(t)dt}{\|\lambda_{p_n}\|_{q_n}}) \\ &\geq \|\lambda_{p_n}\|_{q_n}(1 + \frac{Log\epsilon}{\|\lambda_{p_n}\|_{q_n}} - f(e_0)) \end{aligned}$$

qui tend vers l'infini lorsque $\|\lambda_{p_n}\|_{q_n}$ tend vers l'infini ($\epsilon = \mathbb{P}(X_t^1 = e_0, \forall t \in [0,1]) > 0$).
Mais

$$J_{p_n}(\lambda_{p_n}) = \inf_{\lambda \in V_{p_n}^{'}} J_{p_n}(\lambda) \leq J_{p_n}(0) = 0$$

On arrive ainsi à une contradiction. □

La suite λ_p admet donc une valeur d'adhérence λ_∞ (pour la topologie faible); malheueusement V' (et V) n'a pas les "bonnes" propriétés permettant d'affirmer l'existence d'une sous-suite de λ_p qui converge. La question est alors de savoir si cette valeur d'adhérence est unique, si oui a-t-on $\lambda_\infty = \lambda$ (le point minimum de la remarque précédente) et enfin si cette forme linéaire correspond à la limite P^* cherchée.

5 Appendice

• Preuve de la proposition 4.2.5.
On va faire la preuve pour $t = 0$. Il s'agit de calculer pour $x, y \in \Sigma$

$$\lim_{t \longrightarrow 0} \frac{P_t^*(x,y)}{t} = \lim_{t \longrightarrow 0} \frac{E_x^*(1_{X_t^1 = y})}{t} = \lim_{t \longrightarrow 0} \frac{E_x(1_{X_t^1 = y} e^{\int_0^T f(X_s^1)\lambda(s)ds})}{E_x(e^{\int_0^T f(X_s^1)\lambda(s)ds})}$$

$(E_x(.) := E(.|X_0^1 = x), E_x^*(.) := E^*(.|X_0^1 = x))$. Or on a

$$E_x(1_{X_t^1 = y} e^{\int_0^T f(X_s^1)\lambda(s)ds}) = E_x(1_{X_t^1 = y} e^{\int_0^T f(X_s^1)\lambda(s)ds} | X_t = y) P_x(X_t^1 = y)$$

et en utilisant la propriété de Markov

$$E_x(e^{\int_0^T f(X_s^1)\lambda(s)ds} | X_t^1 = y) = E_x(e^{\int_0^t f(X_s^1)\lambda(s)ds} | X_t^1 = y) E_x(e^{\int_t^T f(X_s^1)\lambda(s)ds} | X_t^1 = y)$$

Par ailleurs on a $E_x(e^{\int_0^t f(X_s^1)\lambda(s)ds} | X_t^1 = y) \longrightarrow 1$ quand t tend vers 0 (en effet $E_x(|e^{\int_0^t f(X_s^1)\lambda(s)ds} - 1||X_t^1 = y) = O(t)$.D'autre part grâce à l'homogénéité de X^1 sous P_X on a

$$E_x(e^{\int_t^T f(X_s^1)\lambda(s)ds} | X_t^1 = y) = E_y(e^{\int_t^T f(X_{s-t}^1)\lambda(s)ds})$$

Il suffit de montrer maintenant que cette quantité tend vers $E_y(e^{\int_0^T f(X_s^1)\lambda(s)ds})$ lorsque t tend vers 0. Ceci découle du théorème de convergence dominée appliqué à $\int_t^T f(X_{s-t})\lambda(s)ds = \int_0^T 1_{s \geq t} f(X_{s-t})\lambda(s)ds$ en remarquant que pour P_X-presque toute trajectoire X_s^1, $f(X_{s-t})^1 \longrightarrow f(X_s^1)$ lorsque t tend vers 0 sauf pour un nombre fini de points. Bien entendu le théorème de convergence dominée est appliqué ensuite pour le passage à la limite sous E_y. Ceci achève la démonstration. □

••Preuve de la proposition 4.2.6.
On va calculer la dérivée à droite de $h_x(t)$; la dérivée à gauche se calcule de manière analogue. Dans la suite les symboles $d/d\epsilon$ désigneront des dérivées à droite ($\epsilon > 0$) et X_t^1 sera noté X_t.
1/ Écrivons

$$\begin{aligned} h_x(t) &= \sum_y E(1_{X_{t+\epsilon} = y} e^{\int_t^T f(X_s).\lambda(s)ds} | X_t = x) \\ &= \sum_y E(e^{\int_t^T f(X_s).\lambda(s)ds} | X_t = x, X_{t+\epsilon} = y) P_X(X_{t+\epsilon} = y | X_t = x) \end{aligned}$$

Pour $y \neq x$, d'après la démonstration de la proposition précédente on a $E(e^{\int_t^T f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = y) \longrightarrow E(e^{\int_0^T f(X_s).\lambda(s)ds}|X_t = y)$ et par conséquent il est facile de voir que

$$\frac{d[E(e^{\int_t^T f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = y)P_X(X_{t+\epsilon} = y|X_t = x)]}{d\epsilon} =$$

$$E(e^{\int_0^T f(X_s).\lambda(s)ds}|X_t = y)A(x,y)$$

De même pour $y = x$ on a

$$\frac{d[E(e^{\int_t^T f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = x)P_X(X_{t+\epsilon} = x|X_t = x)]}{d\epsilon} =$$

$$E(e^{\int_0^T f(X_s).\lambda(s)ds}|X_t = x)A(x,x) + \frac{d[E(e^{\int_t^T f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = x)]}{d\epsilon}$$

Pour calculer le dernier terme, on écrit encore une fois par la propriété de Markov

$$\begin{aligned} E(e^{\int_t^T f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = x) \quad = \quad & E(e^{\int_t^{t+\epsilon} f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = x)\times \\ & E(e^{\int_{t+\epsilon}^T f(X_s).\lambda(s)ds}|X_{t+\epsilon} = x) \end{aligned} \tag{20}$$

le deuxième terme du second menbre tend vers $h_x(t)$ lorsque ϵ tend vers 0 et sa dérivée par rapport à ϵ n'est autre que dh_x/dt. Regardons maintenant la dérivée de l'autre terme à savoir

$$E(e^{\int_t^{t+\epsilon} f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = x) = \frac{E(1_{X_{t+\epsilon}=x} e^{\int_t^{t+\epsilon} f(X_s)\lambda(s)ds})}{P(X_{t+\epsilon} = x|X_t = x)}$$

Il faut noter qu'il tend vers 1 quand ϵ tend vers 0. Calculons maintenant la dérivée du numérateur. On a

$$\begin{aligned} E(1_{X_{t+\epsilon}=x} e^{\int_t^{t+\epsilon} f(X_s)\lambda(s)ds}|X_t = x) \quad &= \quad E(1_{X_{t+\epsilon}=x}(1 + \textstyle\int_t^{t+\epsilon} f(X_s)\lambda(s)ds + o(\epsilon))|X_t = x) \\ &= \quad E(1_{X_{t+\epsilon}=x}|X_t = x) \\ &+ \quad E(1_{X_{t+\epsilon}=x} e^{\int_t^{t+\epsilon} f(X_s)\lambda(s)ds}|X_t = x) + o(\epsilon) \end{aligned} \tag{21}$$

On en déduit que la dérivée cherchée est

$$\frac{d[E(1_{X_{t+\epsilon}=x} e^{\int_t^{t+\epsilon} f(X_s)\lambda(s)ds}|X_t = x)]}{d\epsilon} = A(x,x) + f(x)\lambda(t)$$

(on a utilisé le fait que $E(1_{X_{t+\epsilon}=x} - 1_{\{X_s=x, t\leq s\leq t+\epsilon\}}|X_t = x)$ tend vers 0 lorsque ϵ tend vers 0.) Par conséquent

$$\begin{aligned} \frac{d}{d\epsilon}\left[\frac{E(1_{X_{t+\epsilon}=x} e^{\int_t^{t+\epsilon} f(X_s)\lambda(s)ds})}{P_X(X^1_{t+\epsilon} = x|X_t = x)}\right] \quad &= \quad -A(x,x) + A(x,x) + f(x)\lambda(t) \qquad (22) \\ &= \quad f(x)\lambda(t) \qquad (23) \end{aligned}$$

Finalement

$$\frac{d[E(e^{\int_t^T f(X_s).\lambda(s)ds}|X_t = x, X_{t+\epsilon} = x)P_X(X_{t+\epsilon} = y|X_t = x)]}{d\epsilon} =$$

$$h_x(t)(A(x,x) + f(x)\lambda(t)) + \frac{dh_x(t)}{dt}$$

et

$$\sum_y A(x,y)h_y(t) + h_x(t)f(x)\lambda(t) + \frac{dh_x(t)}{dt} = 0 \quad \square$$

2/Venons en maintenant à $k_x(t)$. La preuve est analogue à la précédente :

$$k_x(t) = \sum_y E(1_{X_{t-\epsilon}=y} e^{\int_0^t f(X_s).\lambda(s)ds}|X_t = x) \tag{24}$$

$$= \sum_y E(1_{X_t=x}1_{X_{t-\epsilon}=y} e^{\int_0^t f(X_s).\lambda(s)ds}) \times \frac{1}{P(t,x)} \tag{25}$$

et l'on regardera la dérivée de $g(\epsilon, t, x, y) = E(1_{X_t=x}1_{X_{t-\epsilon}=y} e^{\int_0^t f(X_s).\lambda(s)ds})$ par rapport à ϵ. Par la propriété de Markov on a

$$g(\epsilon, t, x, y) = E(e^{\int_0^{t-\epsilon} f(X_s).\lambda(s)ds}|X_{t-\epsilon} = y)E(1_{X_t=x} e^{\int_{t-\epsilon}^t f(X_s).\lambda(s)ds}|X_{t-\epsilon} = y)$$

La dérivée (par rapport à ϵ) du premier terme du second membre n'est autre que $-dk_y(t)/dt$; Pour l'autre terme on a

$$\begin{aligned} E(1_{X_t=x} e^{\int_{t-\epsilon}^t f(X_s).\lambda(s)ds}|X_{t-\epsilon} = y) &= E(1_{X_t=x}(1 + \int_{t-\epsilon}^t f(X_s).\lambda(s)ds + o(\epsilon))|X_{t-\epsilon} = y) \\ &= E(X_t = x|X_{t-\epsilon} = y) \\ &+ E(1_{X_t=x} \int_{t-\epsilon}^t f(X_s).\lambda(s)ds|X_{t-\epsilon} = y) + o(\epsilon) \end{aligned}$$

ce qui montre que ce terme tend vers 0 lorsque $\epsilon \longrightarrow 0$ si $y \neq x$ et vers 1 si $y \neq x$ et que sa dérivée vaut

$$A(y,x) - f(x).\lambda(t)\delta_{xy}$$

En somme on a

$$\frac{dE(1_{X_t=x} e^{\int_{t-\epsilon}^t f(X_s).\lambda(s)ds}|X_{t-\epsilon} = y)}{d\epsilon} = -\frac{dk_x(t)}{dt}\delta_{xy} + k_x(t)(A(y,x) - f(x).\lambda(t)\delta_{xy})$$

et

$$\begin{aligned} \frac{dg(\epsilon, t, x, y)}{d\epsilon} &= P(t,x)(-\frac{dk_x(t)}{dt}\delta_{xy} + k_x(t)(A(y,x) - f(x).\lambda(t)\delta_{xy})) \\ &+ -P'(t,x)k_x(t) \end{aligned}$$

L'équation régissant $k_x(t)$ se déduit du fait que

$$\sum_y \frac{dg(\epsilon, t, x, y)}{d\epsilon} = 0 \quad \Box\Box$$

Remarque. Lorsque $\lambda \equiv 0$ l'équation à laquelle satisfait $k_x(t)$ donne

$$\sum_y A(y, x)P(t, y) = P^{'}(t, x)$$

On retrouve ainsi l'équation de Kolmogorov ('forward').

$\star$

$\star\star$

Remerciements
Ce travail a été effectué au sein du projet Mistral; il m'est agréable de remercier Monsieur F.Baccelli pour la qualité de son accueil, ses encouragements et pour l'intérêt qu'il a porté à ce sujet. D'autre part, je suis redevable à S.A.Zuyev d'avoir posé un problème de performances de réseaux hiérarchiques qui a été à l'origine de la méthode considérée ci-dessus; je le remercie pour son aide et les longues discussions que nous avons eu sur ce problème.

Références

[1] Beneš V.E. (1963) A Thermodynamic theory of traffic in connecting networks. *Bell. Syst. Tech. J. 567-607*

[2] Billingsley P.(1968) *Convergence of Probability Measures* . Wiley

[3] Csiszár I. (1975) I-divergence geometry of probability distributions and minimization problems. *Ann. of Prob. 3, 146-158*

[4] Csiszár I. (1984) Sanov property, generalized I-projection and a conditional limit theorem. *Ann. of Prob. 12, 768-793*

[5] Dembo A., Zeitouni O.(1993) *Large Deviation Techniques and Applications.* Jones and Bart. Pub.

[6] Dunford N., Schwartz J.T.(1988) *Linear operators.* Vol 1. Wiley

[7] Kouvatsos D.D., Xenios N.P. (1989) Maximum entropy analysis of general networks with blocking. In H.G.Perros, T.Altiock (eds) *Workshop on queuing networks with blocking.* North Holland.

[8] Lanford O.E. (1973) Entropy and equilibrium states in classical statistical mechanics and mathematical problems. *Lect. N. In Phy. 20, 1-113.* Springer-Verlag.

[9] Neveu J.(1970) *Bases mathématiques du calcul des probabilités.* 2ème éd. Masson.

[10] Pinsker M.S. (1964) *Information and information stability of random variables and processes.* Holden-Day, San Francisco.

[11] Pinsky E., Yemini Y.(1984) A statistical mechanics of some interconnection networks. In E.Gelenbe(ed) *Performace'84.* North-Holland.

[12] Pinsky E., Yemini Y.(1986) The canonical approximation in performance analysis. In T.Hasegawa *et al.* (eds) *Computer Networking and Performance Evaluation* . North-Holland.

[13] Stroock D.W., Zeitouni O.(1991) Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In Durrett and Kesten eds. *Festchrift in honour of F.Spitzer.* Birkhäuser.

[14] Van Campenhout J.M., Cover T.M. (1981) Maximum entropy and conditional probability. *IEEE Trans. Inf. Th. IT 27 483-489*

[15] Vasicek O.A. (1980) A conditional law of large numbers. *Ann. of Prob. 8 142-147*

[16] Yosida K. (1980) *Functional Analysis*, 6th ed. Springer-Verlag.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur

INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

ISSN 0249-6399