

THÈSES D'ORSAY

FABRICE GAMBOA

Méthode du maximum d'entropie sur la moyenne et applications

Thèses d'Orsay, 1989

[<http://www.numdam.org/item?id=BJHTUP11_1989__0248__P0_0>](http://www.numdam.org/item?id=BJHTUP11_1989__0248__P0_0)

L'accès aux archives de la série « Thèses d'Orsay » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.



NUMDAM

*Thèse numérisée par la bibliothèque mathématique Jacques Hadamard - 2016
et diffusée dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>*

UNIVERSITE PARIS SUD

CENTRE d'ORSAY

T H E S E

présentée pour obtenir

Le TITRE de DOCTEUR EN SCIENCES

SPECIALITE : MATHEMATIQUES

par

Fabrice GAMBOA

**Sujet : METHODE DU MAXIMUM D'ENTROPIE SUR LA MOYENNE
ET APPLICATIONS.**

Soutenue le 20 Décembre 1989 devant le Jury composé de :

MM. DACUNHA-CASTELLE	Didier	Président
AZENCOTT	Robert	
COMETS	Francis	
DEMOMENT	Guy	
GUYON	Xavier	
NAVAZA	Jorge	

Maximum entropy method on the mean and applications

Fabrice Gamboa

Abstract

An explicit solution for the problem of probability reconstruction when only the averages of random variables are known is given by the maximum entropy method. We use this method to reconstruct a function constrained to a convex set $\mathcal{C}$ (no linear constraint) using a finite number of its generalized moments (linear constraint). A sequence of entropy maximization problems is considered. The n^{th} problem consists in the reconstruction of a probability distribution on $\mathcal{C}_n$, the projection of $\mathcal{C}$ on $\mathbb{R}^n$, whose mean satisfies a constraint approximating the initial linear constraint (generalized moments). When n approaches infinity this gives a solution for the initial problem as the limit of the sequence of means of maximum entropy distributions on $\mathcal{C}_n$. We call this technique the maximum entropy method on the mean (M.E.M) because linear constraints are only on the mean of the distribution to be reconstructed. We mainly study the case where $\mathcal{C}$ is a band of continuous functions. We find a reconstruction family, each element of this family only depends of referenced measures used for the sequence of entropy problems. We show that the M.E.M method is equivalent to a concav criteria maximization. We then use the M.E.M method to construct a numerically computable criteria to solve generalized moments problem on a bounded band of continuous functions. In the last chapter we discuss statistical applications of the method.

Key words

Maximum of entropy
Generalized moments
Large deviations

Functionnal estimation
Moment problems

A.M.S Code: 62B10, 62A99.

Je voudrais tout d'abord remercier Didier Dacunha-Castelle de m'avoir proposé un sujet de thèse et d'avoir dirigé efficacement mon travail. Je voudrais lui exprimer ici reconnaissance et amitié.

Je remercie Jorge Navaza dont le travail est à l'origine de cette thèse, de m'avoir fait bénéficier de fructueuses discussions.

Je remercie Robert Azencott, Guy Demoment et Xavier Guyon d'avoir accepté d'être membres du jury, Francis Comets et Jorge Navaza d'en être les rapporteurs.

Une partie de ce travail a été fait en collaboration avec Elizabeth Gassiat, je la remercie pour sa constante bonne humeur, travailler avec elle est un plaisir.

La réalisation finale de cette thèse a été faite à la Escuela de Física y Matemática de l'Université Centrale de Caracas, je remercie son directeur José León de m'y avoir accueilli.

Enfin, je voudrais remercier Sylvie, Marc et Chichi pour leur aide dans la réalisation matérielle de cette thèse.

A Sylvie

CHAPITRE I

INTRODUCTION ET PLAN DETAILLE DE LA THESE

I.1 INTRODUCTION

La méthode du maximum d'entropie est une méthode pour reconstruire des mesures de probabilité, lorsque l'on ne dispose comme information que des valeurs moyennes d'un nombre fini de variables aléatoires [Wragg-Dow, 70] (i.e $E_p(\phi_j)=c_j$, $j=1,...,k$). Elle consiste à choisir parmi toutes les probabilités candidates, (i.e parmi toutes les probabilités P vérifiant, $E_p(\phi_j)=c_j$, $j=1,...,k$) celle qui maximise la μ - entropie:

$$S_{\mu}(P) = - \int_{\Omega} \text{Log} \frac{dP}{d\mu} dP ,$$

où μ est une mesure de référence donnée. L'origine de la méthode est le choix de modèles en mécanique statistique [Jayn, 57]. On trouvera dans [Jayn, 82] et [Shore-John, 80] des justifications sur le choix de la fonctionnelle d'entropie comme fonctionnelle à maximiser.

La méthode du maximum d'entropie a été largement utilisée dans des problèmes de théorie du signal, d'imagerie¹ ou de théorie de l'information. On trouvera par exemple dans la thèse de A. Mohammad Djafari [Moham, 87] le développement de la méthode pour des problèmes de tomographie, ainsi qu'une bibliographie très complète sur le sujet. En ce qui concerne les problèmes de déconvolution en image on pourra consulter les travaux de S.F. Gull et J. Skilling [Gull-Skil, 84], [Gull-Skil, 85]. A. Lippman [Lip, 86] quant à lui

¹La fonctionnelle d'entropie est en général utilisée ici comme opérateur de régularisation pour des problèmes inverses mal posés. Le plus souvent le signal (ou l'image) à reconstruire est positif, il est considéré après normalisation comme une probabilité. La méthode du maximum d'entropie consiste alors à choisir parmi toutes les reconstructions se trouvant dans un voisinage de l'observation (la forme de ce voisinage est déterminée par la nature du bruit de mesure, l'exemple le plus classique est celui où le bruit est gaussien, le voisinage est alors un ellipsoïde), celle qui maximise l'entropie.

donne une application particulièrement intéressante de la méthode, il l'utilise pour construire un système expert de reconnaissance de caractères.

G. Bricogne [Bric, 84] a, comme d'autres cristallographes, utilisé la méthode du maximum d'entropie pour résoudre le problème de reconstruction d'une densité électronique en cristallographie. Le problème cristallographique [Dac, 84] consiste à reconstruire une densité de probabilité sur une maille élémentaire (cube unité de $\mathbb{R}^3$) lorsque l'on ne dispose que d'un certain nombre de modules des coefficients de Fourier de la densité à reconstruire. Si maintenant on s'intéresse au problème cristallographique de la reconstruction de la densité *scattering length* pour la diffraction neutronique, la méthode du maximum d'entropie n'est plus applicable car la fonction à reconstruire n'est plus une probabilité (elle peut prendre des valeurs négatives). J. Navaza [Nav, 85], [Nav, 86] propose pour résoudre ce problème d'utiliser la méthode du maximum d'entropie d'une autre manière. L'idée est de considérer l'ensemble $\mathcal{C}$ des densités comprises entre a et b (les seuils a et b sont donnés par la physique du problème), de munir l'ensemble fonctionnel $\mathcal{C}$ d'une mesure de référence μ et de chercher la probabilité sur $\mathcal{C}$ du maximum de μ -entropie dont la moyenne (à valeur fonction) possède les modules des coefficients de Fourier observés. La reconstruction est alors la valeur moyenne de cette probabilité. Elle possède la propriété d'être comprise entre a et b (moyenne de fonctions comprises entre a et b), et par construction les modules de ses coefficients de Fourier sont ceux qui ont été observés. Bien évidemment le problème précédent ne peut pas être résolu de façon directe car il est difficile de définir une mesure de référence μ sur l'ensemble de fonctions $\mathcal{C}$ ayant une interprétation physique acceptable. Pour le résoudre on l'approche par une suite de problèmes de maximisation de l'entropie sur une suite d'espaces $(\mathcal{C}_n)_{n \geq 1}$ approchant $\mathcal{C}$, et de dimension finie.

Plus généralement donnons-nous un convexe $\mathcal{C}^2$ d'un espace de fonctions E, une famille finie $\phi = (\phi_j)_{j=1..k}$ de fonctions du dual de E, et considérons pour c vecteur de $\mathbb{R}^k$ donné le problème de reconstruction d'une fonction g appartenant à $\mathcal{C}$ dont les ϕ - moments valent c, c'est-à-dire trouver une fonction g de E qui vérifie la contrainte linéaire,

$$(L) \quad \int \phi(x) g(x) dx = c,$$

et la contrainte non linéaire,

²Replaçons nous dans le cadre de la note 1 (voir page 3). L'utilisation directe de la méthode du maximum d'entropie revient à utiliser un convexe $\mathcal{C}$ constitué de fonctions positives. Il est des cas (cristallographie, image, etc..) où l'on connaît plus précisément les enveloppes des signaux à reconstruire, un procédé de reconstruction prenant en compte cette information a priori peut donc être intéressant.

$$(NL) \quad g \in \mathcal{C}.$$

On peut alors essayer de reconstruire g en utilisant la technique dont nous avons parlé précédemment. Pour cela on se donne un quadruplé, $((T_n)_{n \geq 1}, (R_n)_{n \geq 1}, (\mu_n)_{n \geq 1}, (\phi_n)_{n \geq 1})$ où T_n (resp. R_n) est un opérateur linéaire défini sur E et à valeurs dans $\mathbb{R}^n$ (resp. défini sur $\mathbb{R}^n$ et à valeurs dans F contenant E), μ_n est une mesure de référence sur $\mathcal{C}_n$ image du convexe $\mathcal{C}$ par l'opérateur linéaire T_n et ϕ_n est un opérateur linéaire sur $\mathbb{R}^n$ à valeurs dans $\mathbb{R}^k$ approchant ϕ (dans un sens à préciser). On considère alors le problème suivant de reconstruction dans $\mathbb{R}^n$: Trouver un vecteur y_n du convexe $\mathcal{C}_n$ dont l'image par ϕ_n est c , c'est-à-dire trouver un vecteur y_n de $\mathbb{R}^n$ qui vérifie la contrainte linéaire,

$$(L_n) \quad \phi_n y_n = c,$$

et la contrainte non linéaire,

$$(NL_n) \quad y_n \in \mathcal{C}_n.$$

On cherche alors, P_n^{ME} probabilité de μ_n - entropie maximale dont l'espérance (à valeur dans $\mathbb{R}^n$) vérifie la contrainte linéaire (L_n) , en supposant que cette probabilité existe, son espérance y_n^{ME} est une solution du $n^{\text{ème}}$ problème de reconstruction:

$$y_n^{ME} \text{ définie par } y_n^{ME} = \int_{\mathcal{C}_n} \omega_n dP_n^{ME}(\omega_n),$$

où ω_n désigne l'élément générique de $\mathbb{R}^n$, vérifie

$$(L_n) \quad \phi_n y_n^{ME} = c,$$

et,

$$(NL_n) \quad y_n^{ME} \in \mathcal{C}_n.$$

Posons maintenant $g_n^{ME} = R_n y_n^{ME}$, sous des hypothèses raisonnables sur la suite $(\phi_n)_{n \geq 1}$ des approximations de ϕ , ainsi que sur les suites d'opérateurs $(T_n)_{n \geq 1}$ et $(R_n)_{n \geq 1}$, on peut espérer que la suite (g_n^{ME}) converge (dans un sens

à préciser) vers g_{∞}^{ME} solution du problème initial. Cette méthode de reconstruction consistant à utiliser une suite de problèmes de maximisation de l'entropie sur des espaces de dimension finie est appelée méthode du maximum d'entropie sur la moyenne (M.E.M) car la contrainte ne porte que sur les moyennes des lois à reconstruire. Elle est résumée dans le schéma 1.

Cette thèse est essentiellement consacrée à l'étude de la suite des problèmes M.E.M, dans le cas où le convexe fonctionnel $\mathcal{C}$ est la bande de fonctions continues:

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \},$$

où $C(U)$ désigne l'ensemble des fonctions continues définies sur U sous-ensemble de $\mathbb{R}^1$ et, a et b sont des fonctions données. Pour construire la suite de problèmes de dimension finie on utilise ici une suite $(x_i)_{i=1}$ de points de U dense dans U . Le $n^{\text{ème}}$ problème de maximisation de l'entropie sur la moyenne consiste alors à trouver une reconstruction définie seulement aux points $(x_i)_{i=1..n}$. Le convexe $\mathcal{C}_n$ est donc le pavé,

$$\mathcal{C}_n = \prod_{i=1}^n]a(x_i), b(x_i)[,$$

que l'on munit d'une mesure produit $\mu_n = \bigotimes_{i=1}^n F_{x_i}$, où $(F_x)_{x \in U}$ est une famille de mesures positives donnée. Le résultat principal est alors l'existence d'une reconstruction de la forme,

$$g_{\infty}^{ME}(x) = \psi'(\langle v_{\infty}^*(c), \phi(x) \rangle, x),$$

où $v_{\infty}^*(c)$ est un vecteur de $\mathbb{R}^k$ qui dépend de c , $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel sur $\mathbb{R}^k$ et $\psi'(\cdot, x)$ est la log-transformée de Laplace de la mesure F_x . Cette reconstruction est obtenue comme limite de la suite (g_n^{ME}) .

Nous montrons ensuite que la fonction g_{∞}^{ME} possède la propriété d'être le maximum d'une fonctionnelle concave (dépendant de la famille $(F_x)_{x \in U}$) sur l'ensemble des reconstructions possibles. La plupart des méthodes de reconstruction classiquement utilisées (moindres carrés, entropie, entropie de la prédiction, etc..) sont alors interprétables comme des problèmes M.E.M avec des familles de mesures $(F_x)_{x \in U}$ particulières. Ces résultats sont présentés au chapitre III.

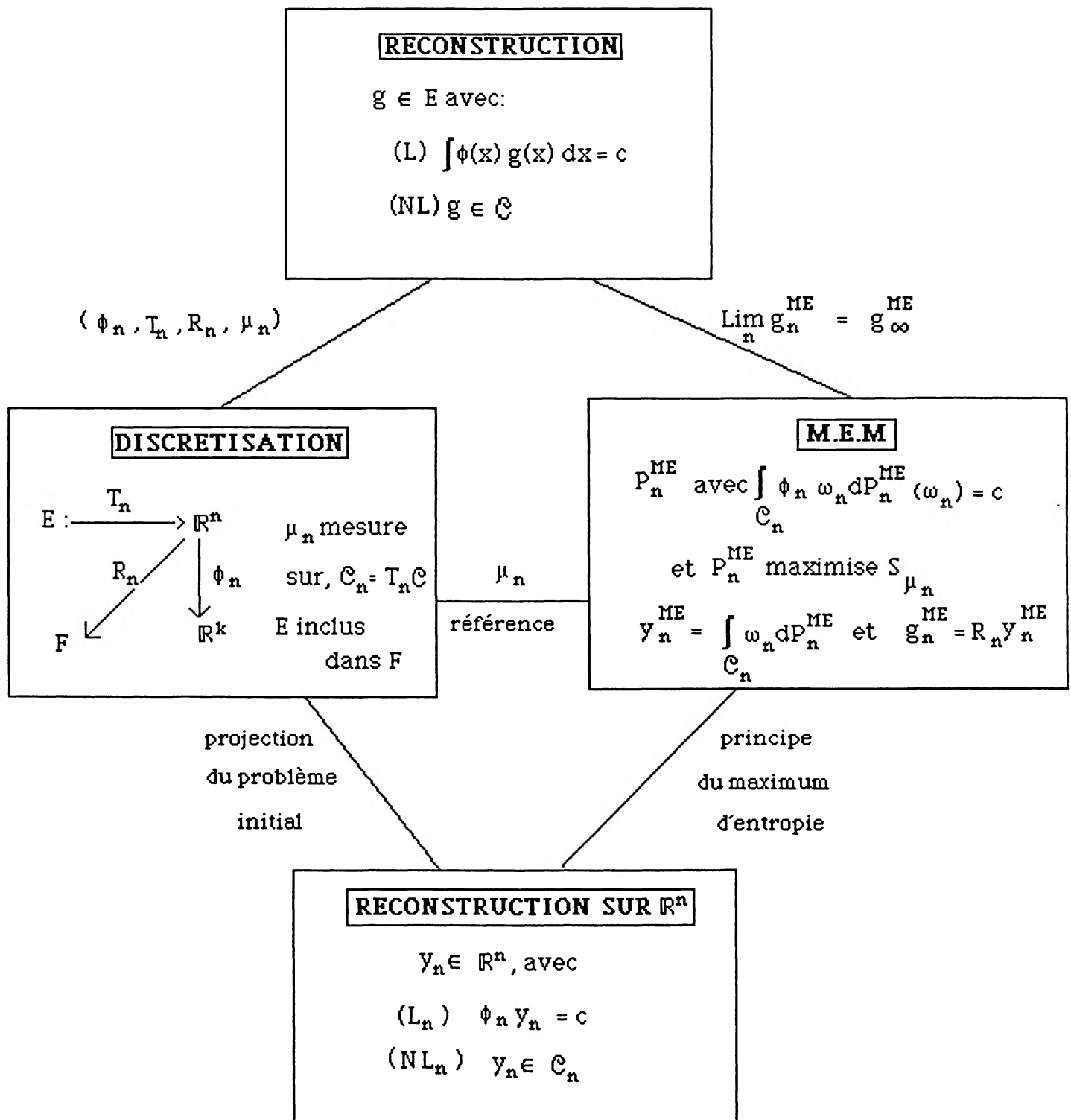


Schéma 1 la méthode M.E.M

Le problème des moments trigonométriques est un problème classique en analyse. Il consiste à trouver des conditions sous lesquelles une suite de réels $(c_j)_{j=0..k}$ est la suite des moments trigonométriques d'une mesure positive et à caractériser les extensions extrémales. Dans [Land,87] on trouvera une résolution du problème sous l'angle du principe du maximum d'entropie. Plus généralement, étant donnés une suite de fonctions continues $\phi(x) = (\phi_j(x))_{j=1..k}$, une suite de réels $(c_j)_{j=1..k}$ et un convexe de fonctions $\mathcal{C}$, le problème des moments généralisés consiste à donner des conditions sous lesquelles $(c_j)_{j=1..k}$ est la suite des ϕ -moments d'une distribution du convexe $cl(\mathcal{C})$, ($cl(\mathcal{C})$ désigne la fermeture de $\mathcal{C}$ dans un sens à préciser) et à caractériser les extensions extrémales. On trouvera dans le livre de M.G. Krein et A.A. Nudel'man [Krein-Nud, 77] une résolution du problème lorsque $\mathcal{C}$ est la bande de fonctions continues,

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \}.$$

Au chapitre IV nous étudions ce problème sous l'angle du M.E.M. Enfin le chapitre V est constitué d'applications statistiques de la méthode du maximum d'entropie sur la moyenne.

I.2 PLAN DETAILLE DE LA THESE

II La méthode du maximum d'entropie

Ce chapitre est une présentation de la méthode du maximum d'entropie pour reconstruire une probabilité. Nous présentons aussi d'autres constructions où la distribution du maximum d'entropie intervient. Dans [Gzyl, 88], H.K. Gzyl donne un panorama très complet et une bibliographie importante sur le sujet.

II.2 Le principe du maximum d'entropie

Après avoir donné la définition de la fonctionnelle d'entropie nous rappelons les résultats de J.E. Shore et R.J. Johnson [Shore-John, 80] qui montrent que le procédé de reconstruction d'une probabilité suivant le critère d'entropie maximale possède certaines propriétés d'invariance. Nous rappelons ensuite une construction ensembliste de l'entropie due à J. Chover [Chov, 61].

II.3 Minimisation de l'information de Kullback et maximisation de l'entropie

Nous étudions ici le problème variationnel de maximisation de la μ - entropie (ou de la minimisation de l'information de Kullback par rapport à μ , si μ est une probabilité) sous des contraintes sur les valeurs moyennes d'un nombre fini de variables aléatoires données.

II.3.1 Cas de contraintes linéaires

Nous rappelons les résultats de I. Csiszar [Csis, 75], [Csis, 84] dans le cas de contraintes d'égalité pour les valeurs moyennes. Le résultat principal est que sous certaines hypothèses la loi du maximum d'entropie est un élément du modèle exponentiel dont la statistique exhaustive est la variable (à valeurs vectorielles) définissant la contrainte. Nous donnons ensuite quelques exemples de reconstruction de probabilité par la méthode du maximum d'entropie.

II.3.2 Généralisation à des contraintes linéaires relaxées

Nous montrons dans le cas de contraintes linéaires relaxées, c'est-à-dire lorsque les valeurs moyennes sont contraintes à être dans un convexe compact K , et sous des hypothèses du même type que celles faites dans le cas de contraintes d'égalité, que la probabilité de μ - entropie maximale est encore un élément du modèle exponentiel précédemment défini. Ceci est illustré par d'autres exemples.

II.4 Autres constructions où la loi du maximum d'entropie intervient

II.4.1 Grandes déviations

Après avoir donné la définition de grandes déviations de R. Ellis [El, 85], nous rappelons deux exemples: grandes déviations d'une somme de variables indépendantes équidistribuées, grandes déviations d'une mesure empirique. Enfin nous donnons un troisième exemple (qui est une variante du premier) qui est directement en rapport avec la méthode M.E.M.

II.4.2 Maximum d'entropie et probabilité conditionnelles

Nous rappelons brièvement que la loi du maximum d'entropie est la limite d'une certaine suite de lois conditionnelles [Van-Cov, 81]. Ceci est illustré par un exemple pour des lois discrètes.

III La méthode du maximum d'entropie sur la moyenne

Ce chapitre est consacré à la méthode du maximum d'entropie sur la moyenne qui a été décrite dans l'introduction. Nous étudions tout d'abord dans

le premier paragraphe des exemples pour lesquels une utilisation directe de la méthode est possible, c'est-à-dire que l'on peut appliquer directement la méthode du maximum d'entropie sur $\mathcal{C}$. Dans le paragraphe suivant on s'intéresse à une formulation abstraite de la suite des problèmes M.E.M, ainsi qu'à la résolution du même problème d'entropie. Le paragraphe III.4 étudie la convergence de la suite des reconstructions (g_n^{ME}) dans trois exemples, l'exemple principal étant la bande de fonctions continues,

$$\mathcal{C} = \{g \in C(U), \forall x \in U \ a(x) < g(x) < b(x)\}.$$

Enfin le dernier paragraphe traite du lien entre la méthode M.E.M et la maximisation d'un critère concave.

III.2 Exemples de reconstruction

III.2.1 Moments généralisés

Si l'on considère le cas où le convexe $\mathcal{C}$ est l'ensemble des fonctions continues sur $[0,1]$, ou continues et strictement positives sur $[0,1]$ et que la contrainte linéaire est la contrainte de ϕ - moments,

$$\int_0^1 \phi(x) g(x) dx = c, \phi \text{ fonction sur } [0,1], \text{ à valeurs dans } \mathbb{R}^k,$$

on peut donner une formulation globale du problème. Le problème d'entropie consiste ici à trouver parmi une famille de processus celui dont l'information de Kullback par rapport à un processus à accroissements indépendants (processus de référence) est minimale et dont la dérivée de l'espérance vérifie la contrainte linéaire. La reconstruction est alors la dérivée de l'espérance du processus solution. En particulier lorsque le processus de référence est un processus croissant on obtient une reconstruction positive. Ce type de reconstruction est étudié lorsque le processus de référence est le processus de Poisson ou de Wiener, on retrouve alors respectivement la reconstruction d'entropie maximale et de norme L^2 minimale.

III.2.2 Application Interpolation par fonctions splines

Lorsque le convexe fonctionnel est l'ensemble des fonctions continues sur l'intervalle $[0,1]$ et que les contraintes linéaires sont des contraintes d'interpolation, on peut encore utiliser directement la méthode M.E.M. Le problème M.E.M consiste alors à trouver parmi une famille de processus celui dont l'information de Kullback par rapport à un processus de référence est minimale, et dont l'espérance (à valeur fonction) de l'image par une certaine transformation vérifie les contraintes d'interpolation. La reconstruction est

alors l'espérance du processus transformé sous la loi du maximum d'entropie. On étudie le cas où le processus de référence est le processus de Wiener et la transformation est une série d'intégrations. Dans ce cas la méthode M.E.M est la méthode d'interpolation par fonctions Splines qui est décrite dans [Schoen, 64].

III.3 La suite des problèmes M.E.M pour un convexe d'un espace de fonctions

III.3.1 Définition du problème

La suite des problèmes M.E.M est définie de façon abstraite pour un convexe $\mathcal{C}$ d'un espace de Banach. On détaille en particulier la construction pour les deux exemples qui seront traités par la suite: $\mathcal{C}$ la boule unité d'un espace de Hilbert séparable et $\mathcal{C}$ la bande d'un espace de fonctions continues.

III.3.2 Résolution du $n^{\text{ème}}$ problème d'entropie

On utilise ici les résultats du paragraphe II.3 pour résoudre le $n^{\text{ème}}$ problème d'entropie. On montre que, sous des hypothèses sur le support et sur la transformée de Laplace de la mesure de référence μ_n , il existe une solution au $n^{\text{ème}}$ problème d'entropie.

III.4 Convergence de la suite des problèmes M.E.M

III.4.1 Boule unité d'un espace de Hilbert séparable

On utilise une base Hilbertienne pour discrétiser l'espace. Le $n^{\text{ème}}$ problème d'entropie est alors un problème d'entropie sur la boule unité de $\mathbb{R}^n$. On montre que pour toute suite de mesures de référence $(\mu_n)_{n \geq 1}$, telle que μ_n soit invariante par le groupe des isométries sur $\mathbb{R}^n$, la suite des reconstructions (g_n^{ME}) converge vers la reconstruction de norme minimale.

III.4.2 Bande ouverte d'un espace de fonctions continues

Nous résolvons le problème décrit dans l'introduction dans divers cas de figure.

III.4.2.a Cas compact

Le résultat principal est le théorème 3.1 où il est montré la convergence de la suite des reconstructions dans le cas où les fonctions $a(\cdot)$ et $b(\cdot)$ sont des constantes de $\bar{\mathbb{R}}$ et que la famille de mesures de référence est constante (pour tout x , $F_x = F$). L'étude est ensuite complétée quand $a(\cdot)$ et $b(\cdot)$ sont des

fonctions continues à valeurs dans $\mathbb{R}$ et (ou) que la famille de mesures de référence n'est pas constante. Nous donnons aussi dans cette partie des exemples de reconstructions en utilisant des familles de mesures de référence particulières. A la fin du paragraphe nous revenons sur le troisième exemple de grandes déviations du chapitre II.

III.4.2.b Cas non compact

Lorsque le domaine de définition des fonctions n'est plus compact nous montrons qu'il existe une reconstruction ayant une forme identique à celle obtenue par la méthode M.E.M dans le cas d'un ensemble de définition compact.

III.4.3 Tube de variables aléatoires de L^∞

Cette application de la méthode M.E.M est une extension de l'exemple précédent. Il s'agit de reconstruire une variable aléatoire vectorielle définie sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, contrainte à être P-p.s à valeurs dans un convexe donné et à vérifier une contrainte linéaire. Une suite de problèmes M.E.M est construite à l'aide d'une suite de tribu "*convergeant*" vers $\mathcal{A}$. On obtient alors le même type de reconstruction que celles obtenues dans le cas de la bande de fonctions continues. Nous donnons ensuite un exemple d'application à un problème de commande.

III.5 Problème M.E.M et maximisation d'un critère concave

III.5.2 Maximisation de la fonctionnelle de Legendre

Dans le cadre de l'exemple du paragraphe III.4.2, nous montrons l'équivalence entre méthode du maximum d'entropie sur la moyenne et la maximisation du critère concave,

$$\Gamma(g) = \int_U \gamma(g(x), x) dx,$$

où $\gamma(\cdot, x)$ est l'opposée de la duale de Legendre de la Log-transformée de Laplace de F_x . Les méthodes classiques de reconstruction sont ensuite regardées sous l'angle du M.E.M.

III.5.3 ψ - Divergence et théorème de Pythagore

Le critère concave défini précédemment possède la propriété d'être associé à une divergence, ce qui lui confère des propriétés particulières. Ces propriétés sont étudiées ici.

IV Le problème des moments généralisés

Nous étudions sous l'angle du MEM le problème des moments généralisés.

IV.2 Cas d'une bande de $C(U)$

IV.2.1 Cas général

On énonce ici un corollaire découlant de la construction du MEM: le problème des moments généralisés possède une solution si et seulement si il en existe une de la forme, $\psi'(\langle v_{\infty}^*(c), \phi(x) \rangle, x)$, où $(\psi(\cdot, x))_{x \in U}$ est la famille des Log-transformées de Laplace d'une famille de mesures $(F_x)_{x \in U}$ convenablement bien choisie.

IV.2.2 Cas U compact bande bornée

Lorsque les bornes $a(\cdot)$ et $b(\cdot)$ sont des fonctions continues et que l'ensemble de définition des fonctions est compact, la famille de mesures à priori $F_x = \delta_{a(x)} + \delta_{b(x)}$, permet de construire un critère numériquement calculable pour décider si un point donné de $\mathbb{R}^k$ est une suite de ϕ - moments. Ce critère est une version "convexifiée" d'un critère donné par M.G. Krein et A.A. Nudel'man [Krein-Nud, 77]. Dans le cas où le convexe des extensions est non vide, le MEM permet aussi de caractériser ses points extrémaux.

IV.3 Problème des moments sur un polytope

Ce paragraphe constitue une extension de l'étude faite au paragraphe précédent pour un tube de variables aléatoires, quand le vecteur aléatoire à reconstruire est contraint à appartenir P-p.s à un polytope borné, c'est-à-dire à l'enveloppe convexe d'un nombre fini de points.

V Modèle exponentiel généralisé et problèmes statistiques

V.2 Modèle exponentiel généralisé

La méthode MEM donne une famille de modèles du type $\psi'(\langle v, \phi(\omega) \rangle + v_0)$, où ψ est la Log-transformée de Laplace d'une mesure positive F de support convenable. Par exemple lorsque F est la loi de Poisson

on retrouve le modèle exponentiel. Nous étudions brièvement des problèmes d'estimation pour cette famille de modèles.

V.3 M.E.M et statistique non paramétrique

V.3.1 Le problème sommatoire

Dans [Gas, 86] E. Gassiat étudie les qualités sommatoires (pour les séries de Fourier) de la méthode du maximum d'entropie, nous prolongeons son étude pour la méthode M.E.M.

V.3.2 Estimation de densité par M.E.M

On étudie l'estimation d'une densité de probabilité sur $[0, 2\pi]$ par la méthode des moments trigonométriques empiriques. Il est montré que sous des hypothèses de régularité sur la densité à estimer, l'estimation obtenue par la méthode M.E.M converge p.s en variation.

V.3.3 La méthode de Burg

Pour estimer la densité spectrale d'un processus stationnaire, la méthode de Burg consiste à prendre comme estimateur la densité spectrale du processus A.R qui possède les covariances estimées. Dans le cas d'un processus strictement stationnaire et fortement mélangeant, en utilisant les propriétés asymptotiques du périodogramme et la construction du M.E.M, nous montrons que cette méthode d'estimation peut aussi être interprétée comme une méthode Bayésienne pour un paramètre infini-dimensionnel.

CHAPITRE II

LA METHODE DU MAXIMUM D'ENTROPIE

II.1 INTRODUCTION

Ce chapitre est consacré à la présentation de la méthode du maximum d'entropie. Nous rappelons tout d'abord les résultats de J.E. Shore et R.J. Johnson [Shore-John, 80] montrant que pour reconstruire une probabilité à partir des valeurs des espérances d'un nombre fini de variables aléatoires, la méthode du maximum d'entropie est un procédé de reconstruction vérifiant certains axiomes d'invariance. On étudie ensuite dans le paragraphe II.3 le problème variationnel de maximisation de la μ - entropie sous des contraintes linéaires, puis linéaires relaxées. Enfin le paragraphe II.4 donne deux autres constructions où la loi du maximum d'entropie joue un rôle : Dans les grandes déviations d'une somme de variables aléatoires indépendantes renormalisée ou d'une mesure empirique, et comme limite d'une suite de lois conditionnelles.

II.2 LE PRINCIPE DU MAXIMUM D'ENTROPIE

Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré (μ est une mesure positive), on se propose d'établir un procédé de reconstruction permettant de reconstruire une probabilité P absolument continue par rapport à μ , lorsque l'on connaît uniquement les valeurs moyennes (sous la probabilité P) d'un nombre fini de variables aléatoires¹:

$$(M) \quad E_P[\phi] = \int_{\Omega} \phi(\omega) dP = c,$$

¹ E_P désigne l'espérance sous la probabilité P .

ϕ est une variable aléatoire sur Ω à valeurs dans $\mathbb{R}^k$ et, c est un élément de $\mathbb{R}^k$ (ϕ et c sont donnés). (M) sera appelée l'information. On exige de plus que le procédé de reconstruction vérifie les axiomes d'invariance suivants :

A1 : Unicité

La reconstruction, est (si elle existe) unique.

A2 : Indépendance par rapport aux systèmes de coordonnées

Si Ω est un espace normé alors la reconstruction est indépendante du choix de systèmes de coordonnées.

A3 : Conservation de l'indépendance

Si Ω est l'espace produit $\Omega = \prod_{i=1}^m \Omega_i$, si μ se factorise en $\mu = \bigotimes_{i=1}^m \mu_i$, où μ_i est une mesure sur Ω_i et si la variable aléatoire ϕ se décompose en,

$$\phi(\omega) = (\phi^{(1)}(\omega_1), \dots, \phi^{(m)}(\omega_m))$$

où pour tout i de $\{1, \dots, m\}$, ω_i appartient à Ω_i et $\phi^{(i)}$ est une variable aléatoire sur Ω_i à valeurs dans $\mathbb{R}^{k_i}$. Alors la reconstruction P^* pour l'information (M) se factorise en $P^* = \bigotimes_{i=1}^m P_i^*$ où P_i^* (probabilité sur Ω_i) est la reconstruction pour l'information

$$(M_i) \quad E_{P_i}(\phi^{(i)}) = c^{(i)},$$

où l'on a décomposé le vecteur c en, $c = (c^{(1)}, \dots, c^{(m)})$, $c^{(i)} \in \mathbb{R}^{k_i}$.

A4 : Conservation d'une partition

Si Ω se décompose en $\Omega = \bigcup_{i=1}^m \Omega_i$, et si (M) est de la forme :

$$\int_{\Omega_i} dP = c_i^{(0)}, \quad i = 1, \dots, m,$$

$$\int_{\Omega_i} \phi^{(i)} dP = c^{(i)}, \quad i = 1, \dots, m.$$

C'est-à-dire que ϕ et c se décomposent en

$$\phi = (1_{\Omega_1}, \dots, 1_{\Omega_m}, 1_{\Omega_1} \phi^{(1)}, \dots, 1_{\Omega_m} \phi^{(m)}),$$

$$c = (c^{(0)}, c^{(1)}, \dots, c^{(m)}),$$

où $\phi^{(i)}$ est pour tout i de $\{1, \dots, m\}$ une variable aléatoire sur Ω_i à valeurs dans $\mathbb{R}^{k_i}$, $c^{(i)}$ appartient à $\mathbb{R}^{k_i}$ et $c^{(0)}$ appartient à $\mathbb{R}^m$. Notons P^* la reconstruction sur Ω pour l'information (M) et P_i^* la reconstruction sur Ω_i pour l'information:

$$(M_i') \quad \int_{\Omega_i} \phi^{(i)} dP_i = c^{(i)} \quad c_i^{(0)}, \quad i = 1, \dots, m.$$

Alors la probabilité conditionnelle $P^*(\cdot | \Omega_i)$ est égale à P_i^* . En d'autres termes la reconstruction est invariante lorsque le problème se décompose en sous-parties si l'on traite le problème globalement ou conditionnellement par parties.

Avant de donner un procédé de reconstruction vérifiant les axiomes précédents, donnons la définition de l'entropie. On notera $\mathbb{P}(\Omega)$ (resp. $\mathbb{M}(\Omega)$) l'ensemble des probabilités (resp. des mesures positives) sur $(\Omega, \mathcal{A})$.

Information de Kullback et entropie

a- mesures bornées

Pour ν, ρ de $\mathbb{M}(\Omega)$, telles que

$$\nu(\Omega) = \int_{\Omega} d\nu < +\infty \quad \text{et} \quad \rho(\Omega) = \int_{\Omega} d\rho < +\infty, \quad (\nu \text{ et } \rho \text{ sont des mesures bornées}),$$

on pose

$$K(\nu, \rho) = \int_{\Omega} \text{Log} \frac{d\nu}{d\rho} d\nu + \rho(\Omega) - \nu(\Omega) \quad \text{si } \nu \ll \rho \text{ et } \text{Log} \frac{d\nu}{d\rho} \in L^1(\nu)$$

$$= +\infty \quad \text{sinon.}$$

$K(\nu, \rho)$ est l'information de Kullback de ν par rapport à ρ . $K(\nu, \rho)$ est une mesure de la dissemblance entre ν et ρ puisque l'on a

$$K(\nu, \rho) \geq 0 \quad \text{et} \quad K(\nu, \rho) = 0 \quad \text{équivaut à} \quad \nu = \rho.$$

b- propabilités

Soient P et Q de $\mathbb{P}(\Omega)$ on a:

$$K(P, Q) = E_P \left(\log \frac{dP}{dQ} \right) = \int_{\Omega} \log \frac{dP}{dQ} dP \quad \text{si } P \ll Q \text{ et } \log \frac{dP}{dQ} \in L^1(P) \\ = +\infty \quad \text{sinon.}$$

Dans [Bret, 79] J. Bretagnolle donne les principales propriétés de $K(P, Q)$. Soient P de $\mathbb{P}(\Omega)$ et ν de $\mathbb{M}(\Omega)$ on pose:

$$S_{\nu}(P) = - E_P \left(\log \frac{dP}{d\nu} \right) = - \int_{\Omega} \log \frac{dP}{d\nu} dP \quad \text{si } P \ll \nu \text{ et } \log \frac{dP}{d\nu} \in L^1(P) \\ = -\infty \quad \text{sinon.}$$

$S_{\nu}(P)$ est la ν - entropie de P, et l'on a pour ν mesure bornée de $\mathbb{M}(\Omega)$,

$$S_{\nu}(P) = - K(P, \nu) + \nu(\Omega) - 1.$$

On a alors le théorème suivant dû à J.E. Shore et R.N. Johnson [Shore-John, 80].

Théorème 2.1 *Le procédé de reconstruction qui consiste à choisir parmi toutes les probabilités qui vérifient la contrainte (M), celle qui maximise la μ - entropie satisfait les axiomes A1, A2, A3, A4.*

Dans la suite, on appellera méthode du maximum d'entropie ce procédé de reconstruction. La méthode du maximum d'entropie est un procédé de reconstruction Bayésien. μ joue le rôle d'une mesure à priori, l'information (M) celui de l'observation et la loi du maximum d'entropie celui de la distribution à postériori. Il existe d'autres constructions de l'entropie, par exemple J. Chover [Chov, 61] obtient l'entropie à l'aide d'une construction ensembliste très générale. Décrivons brièvement cette construction.

On se donne une relation d'équivalence $\sim$ sur la tribu $\mathcal{A}$. Soit σ une partition finie de Ω en ensembles deux à deux disjoints, on pose:

$$\sigma = \{ E_{ij} \}, i = 1, \dots, m \text{ et } j = 1, \dots, n_i \quad \text{où} \quad E_{ij} \cap E_{i'j'} = \emptyset \text{ si } i' \neq i \text{ ou } j' \neq j$$

$$\text{et, } E_{ij} \sim E_{ij'} \quad \forall j, j'.$$

Pour P, probabilité sur $(\Omega, \mathcal{A})$, et i fixé, on construit l'entropie conditionnelle de la i-ème classe d'équivalence

$$S(P, \sigma, i) = - \sum_j \frac{P(E_{ij})}{P(\bar{E}_i)} \log \frac{P(E_{ij})}{P(\bar{E}_i)} \quad \text{si } P(\bar{E}_i) > 0$$

$$= 0 \quad \text{si } P(\bar{E}_i) = 0,$$

où l'on a posé: $\bar{E}_i = \bigcup_{j=1}^{n_i} E_{ij}$.

Posons ensuite, $S(P, \sigma) = \sum_{i=1}^m P(\bar{E}_i) (S(P, \sigma, i) - \log \bar{m}_i)$,

où $\bar{m}_i$ désigne le cardinal de l'ensemble $\{j, P(E_{ij}) > 0\}$.

Posons enfin, $S(P) = \inf_{\sigma} S(P, \sigma)$. On a alors le théorème suivant (cf [Chov, 61]):

Théorème 2.2 *Soit μ une probabilité sur $(\Omega, \mathcal{A})$. On suppose que μ n'est pas atomique et que la relation d'équivalence $\sim$ est définie par:*

Pour E et F de $\mathcal{A}$, $E \sim F$ équivaut à $\mu(E) = \mu(F)$.

Soit P une probabilité absolument continue par rapport à μ , alors

$$S(P) = - \int_{\Omega} \log \frac{dP}{d\mu} \frac{dP}{d\mu} d\mu = S_{\mu}(P).$$

II.3 MINIMISATION DE L'INFORMATION DE KULLBACK ET MAXIMISATION DE L'ENTROPIE.

Dans ce paragraphe, on s'intéresse à la résolution du problème de maximisation de l'entropie, ou de minimisation de l'information de Kullback sous une contrainte du type:

$$E_P(\phi) = \int_{\Omega} \phi dP \in K.$$

On étudiera tout d'abord le cas où K est un singleton $\{c\}$, puis on s'intéressera au cas où K est un ensemble convexe compact de $\mathbb{R}^k$.

II.3.1 Cas de contraintes linéaires

On se propose ici d'étudier le problème variationnel de maximisation de S_μ sur l'ensemble:

$$\mathbb{P}(c, \phi) = \{ P \in \mathbb{P}(\Omega), E_P(\phi) = \int_{\Omega} \phi \, dP = c \}$$

Définition 2.1 On notera P^{MEG} (probabilité du maximum d'entropie généralisé) si elle existe, la probabilité vérifiant: pour toute suite $(P_n)_{n \geq 1}$ de $\mathbb{P}(c, \phi)$ telle que

$$\lim_{n \rightarrow +\infty} S_\mu(P_n) = \sup_{P \in \mathbb{P}(c, \phi)} S_\mu(P),$$

on a $P^{\text{MEG}} = \lim_{n \rightarrow +\infty} P_n$, pour la distance en variation. Lorsque P^{MEG} appartient à $\mathbb{P}(c, \phi)$ on la notera P^{ME} .

On s'intéressera à donner des conditions suffisantes pour que P^{ME} existe.

Introduisons quelques notations:

Fonction de partition

On notera, lorsque cette intégrale existe

$$Z_{\mu, \phi}(t) = \int_{\Omega} \exp(\langle t, \phi(\omega) \rangle) \, d\mu(\omega), t \in \mathbb{R}^k,$$

où pour $a \in \mathbb{R}^k, b \in \mathbb{R}^k, \langle a, b \rangle = \sum_{i=1}^k a_i b_i$ est le produit scalaire sur $\mathbb{R}^k$.

Domaine de définition

$D(\mu, \phi)$ est le domaine de la fonction $Z_{\mu, \phi}$, c'est-à-dire:

$$D(\mu, \phi) = \{ t \in \mathbb{R}^k, Z_{\mu, \phi}(t) < +\infty \}.$$

Lagrangien

On pose

$$\begin{aligned} H_{\mu, \phi}(t, c) &= \text{Log}(Z_{\mu, \phi}(t)) - \langle t, c \rangle \text{ pour } t \in D(\mu, \phi), c \in \mathbb{R}^k, \\ &= +\infty \text{ pour } t \notin D(\mu, \phi), c \in \mathbb{R}^k. \end{aligned}$$

Arc d'Hellinger

Par ϕ -arc d'Hellinger de μ , on entendra la famille de probabilités

$$d\mu_{\phi, t}(\omega) = (Z_{\mu, \phi}(t))^{-1} \exp(\langle t, \phi(\omega) \rangle) d\mu(\omega), t \in D(\mu, \phi).$$

La famille de lois $\{\mu_{\phi, t}(\omega), t \in D(\mu, \phi)\}$ est aussi appelée modèle exponentiel, (voir chapitre V). On trouvera dans le livre de O. Barndorff-Nielsen [Barn, 78] une étude complète des modèles exponentiels.

On a alors,

Lemme 2.1 *On suppose que, $\sup_{P \in \mathcal{P}(c, \phi)} S_{\mu}(P) > -\infty$. Alors*

- a- *Si μ est une probabilité P^{MEG} existe*
- b- *Si $D(\mu, \phi)$ est non vide P^{MEG} existe.*

Démonstration

a- La démonstration est faite par I. Csiszar [Csis, 75] (théorème 2.1), elle utilise essentiellement l'inégalité entre l'information de Kullback et la distance en variation que nous rappelons ici²

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

où $\|P - Q\|$, (la distance en variation entre P et Q) est définie par

$$\|P - Q\| = \int_{\Omega} \left| \frac{dP}{dv} - \frac{dQ}{dv} \right| dv,$$

avec v appartenant à $\mathcal{M}(\Omega)$ et, P et Q absolument continues par rapport à v .

b- Il suffit de remarquer que pour P de $\mathcal{P}(c, \phi)$ et t de $D(\mu, \phi)$ fixé

$$K(P, \mu_{\phi, t}) = H_{\mu, \phi}(t, c) - S_{\mu}(P)$$

maximiser $S_{\mu}(\cdot)$ sur $\mathcal{P}(c, \phi)$ revient donc à minimiser $K(\cdot, \mu_{\phi, t})$ sur $\mathcal{P}(c, \phi)$.

Le théorème suivant dû à I. Csiszar [Csis, 84], résout complètement le problème quand μ est une probabilité (dans ce cas $S_{\mu}(P) = -K(P, \mu)$).

²Voir par exemple [Csis, 67].

Théorème 2.3 $D = \sup_{P \in \mathbb{P}(c, \phi)} S_{\mu}(P) = - \inf_{P \in \mathbb{P}(c, \phi)} K(P, \mu) > -\infty$

si et seulement si il existe P de $\mathbb{P}(c, \phi)$ absolument continue par rapport à μ .
Dans ce cas, on a

$$\begin{aligned} \frac{dP^{MEG}}{d\mu} &= \exp(-D + \langle t^*(c), \phi(\omega) \rangle - \langle t^*(c), c \rangle) \quad \text{si } \omega \in \{ \phi(\omega) \in M \} \\ &= 0 \quad \text{si } \omega \notin \{ \phi(\omega) \in M \}, \end{aligned}$$

où M est un sous-espace vectoriel de $\mathbb{R}^k$. La formule précédente est vraie avec $M = \mathbb{R}^k$, c'est-à-dire que P^{MEG} est un élément du ϕ -arc d'Hellinger de μ si et seulement si il existe P de $\mathbb{P}(c, \phi)$ équivalent à μ , dans ce cas on a:

$$\sup_{P \in \mathbb{P}(c, \phi)} S_{\mu}(P) = - \inf_{t \in \mathbb{R}^k} H_{\mu, \phi}(t, c) = S_{\mu}(P^{MEG})$$

Le cas intéressant en général est celui où P^{MEG} vérifie la contrainte linéaire ($P^{MEG} \in \mathbb{P}(c, \phi)$). Nous donnons dans le corollaire suivant des conditions suffisantes pour que cette propriété soit vérifiée.

Corollaire 2.1

a- Cas où μ est une probabilité.

Si $\mathbb{P}(c, \phi)$ contient au moins une probabilité P équivalente à μ et si $D(\mu, \phi)$ est ouvert, P^{ME} est l'élément du ϕ -arc d'Hellinger de μ qui vérifie:

$$E_{P^{ME}}(\phi) = c.$$

b- Cas où μ est une mesure positive

Même énoncé qu'en a- en remplaçant simplement $D(\mu, \phi)$ est ouvert par $D(\mu, \phi)$ est ouvert non vide.

Démonstration

Point a

Puisqu'il existe dans $\mathbb{P}(c, \phi)$ un P équivalent à μ , le théorème 2.3 donne:

$$\frac{dP^{MEG}}{d\mu} = \frac{d\mu}{d\mu} \phi.t^*(c),$$

où $t^*(c)$ minimise sur $\mathbb{R}^k$ la fonction $H_{\mu, \phi}(\cdot, c)$, le domaine de $H_{\mu, \phi}(\cdot, c)$ étant par hypothèse ouvert, on a ([Ciar, 82], théorème 7.4-4, p. 156) $\text{grad } H_{\mu, \phi}(t^*(c), c) = 0$. Mais (proposition 3.3.22 p. 73 de [Dac-Duf, 82]),

$$\text{grad } H_{\mu, \phi}(t^*(c), c) = \int_{\Omega} \phi(\omega) d\mu_{\phi, t^*(c)} - c.$$

Et donc P^{MEG} vérifie la contrainte linéaire : $E_{P^{\text{MEG}}}(\phi) = c$.

Point b

Puisque $D(\mu, \phi)$ est non vide, maximiser $S_{\mu}(\cdot)$ sur $\mathcal{P}(c, \phi)$ revient à minimiser $K(\cdot, \mu, \phi, c)$ (pour t de $D(\mu, \phi)$ fixé) sur $\mathcal{P}(c, \phi)$, nous avons déjà utilisé cet argument dans la démonstration du lemme 2.1. On conclut alors à l'aide du point a.

Remarque 2.1 Information de Kullback et image

Soient $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ deux espaces mesurables, T une application mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{A}')$ et P et Q de $\mathcal{P}(\Omega)$, telles que l'information de Kullback de P par rapport à Q soit finie. Notons P_T (resp. Q_T) l'image par T de la probabilité P (resp. Q), on a alors (voir par exemple [Bret, 79]) $K(P, Q) \geq K(P_T, Q_T)$, (l'entropie d'un système croît par transformation).

Soient maintenant P' et Q' des probabilités sur $(\Omega', \mathcal{A}')$, notons $P'_{T^{-1}}$ (resp. $Q'_{T^{-1}}$) l'image réciproque par T de P' (resp. Q'). On a alors comme l'a souligné I. Csiszar ([Csis, 84], lemme 3.3), par un simple changement de variable $K(P', Q') = K(P'_{T^{-1}}, Q'_{T^{-1}})$, (l'entropie d'un système reste constante par transformation réversible).

Conséquence

On peut toujours transformer un problème de minimisation de l'information de Kullback (ou de maximisation de l'entropie), sous une contrainte linéaire de dimension finie ($E_P(\phi) = c$), sur un espace mesurable $(\Omega, \mathcal{A})$ en un problème du même type sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Il suffit pour cela de travailler avec les lois images par ϕ .

Exemples

Nous illustrons ici les énoncés précédents par des exemples. Dans l'exemple 2.1, P^{MEG} existe mais P^{ME} n'existe pas. Les exemples 2.2 et 2.4 sont

³Pour A de $\mathcal{A}$, $P_{T^{-1}}(A) = P(T(A))$, $Q_{T^{-1}}(A) = Q(T(A))$.

des exemples dans lesquels P^{ME} existe et est un élément du ϕ - arc d'Hellinger de μ . Enfin dans l'exemple 2.3, P^{ME} existe mais ne fait pas forcément partie du ϕ - arc d'Hellinger de μ .

Exemple 2.1 Lois stables et contrainte de moyenne

On se propose ici de calculer la distribution de minimum d'information de Kullback par rapport à une loi stable sous une contrainte de moyenne (au sens de la définition 2.1).

Soient $\Omega = \mathbb{R}$ muni de la tribu de ses Boréliens $\mathcal{B}(\mathbb{R})$ et pour $x \in \mathbb{R}$ $\phi(x) = x$. Soit μ la densité d'une variable aléatoire de loi stable symétrique d'exposant α ($0 < \alpha < 2$), il est bien connu que,

$$D(\mu, \phi) = \{ t \in \mathbb{R} \mid Z_{\mu, \phi}(t) = \int_{\mathbb{R}} e^{tx} \mu(x) dx < \infty \} = \{ 0 \}.$$

D'autre part pour m réel donné, il existe une probabilité équivalente à μ d'espérance m . En utilisant le théorème 2.3, on voit que la loi du minimum d'information de Kullback par rapport à μ sur

$$\mathcal{P}(m, \phi) = \{ P \in \mathcal{P}(\mathbb{R}), \int_{\mathbb{R}} x dP(x) = m \},$$

(au sens de la définition 2.1) est μ .

Exemple 2.2 Loi Normale

Soient $\Omega = \mathbb{R}$ et pour $x \in \mathbb{R}$ $\phi(x) = (x, x^2)$. On munit $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ de la mesure de Lebesgue L . Pour $c = (m, \sigma^2 + m^2)$ fixé, on a

$$\mathcal{P}(c, \phi) = \{ P \in \mathcal{P}(\mathbb{R}), \int_{\mathbb{R}} x dP(x) = m, \int_{\mathbb{R}} x^2 dP(x) = m^2 + \sigma^2 \}.$$

On voit facilement que les hypothèses du corollaire 2.1 sont vérifiées. On a en effet que :

$$\begin{aligned} D(L, \phi) &= \{ (t_1, t_2) \in \mathbb{R}^2 \mid Z_{L, \phi}(t) = \int_{\mathbb{R}} \exp(t_1 x + t_2 x^2) dx < \infty \} \\ &= \{ (t_1, t_2) \in \mathbb{R}^2 \mid t_2 < 0 \} \end{aligned}$$

est un ouvert non vide.

La loi du maximum d'entropie par rapport à la mesure de Lebesgue sur $\mathbb{R}$, de moyenne et de variance fixées (m, σ^2) est donc la gaussienne $\mathcal{N}(m, \sigma^2)$.

Exemple 2.3 Loi discrète

Solent $x_1, \dots, x_{m+1}$ des points de $\mathbb{R}^m$ ne se trouvant pas dans un même hyperplan. On pose $\Omega = \{x_1, \dots, x_{m+1}\}$. $\mathcal{P}(\Omega)$ désignant la tribu complète, l'espace de probabilité $(\Omega, \mathcal{P}(\Omega))$ est muni de la mesure

$$\mu = \sum_{i=1}^{m+1} \delta_{x_i}.$$

On se donne sur Ω l'application $\phi(\omega) = \omega$ ($\omega \in \Omega$), et l'on définit A comme étant l'enveloppe convexe fermée dans $\mathbb{R}^m$ de Ω . Pour c de A , on s'intéresse au problème d'entropie pour la contrainte

$$\int_{\Omega} \omega \, dP(\omega) = c.$$

L'ensemble,

$$\mathbb{P}(c, \phi) = \{P \in \mathbb{P}(\Omega), \int_{\Omega} \omega \, dP(\omega) = c\},$$

est réduit au singleton,

$$\left\{ \sum_{i=1}^{m+1} a_i \delta_{x_i} \text{ où } c = \sum_{i=1}^{m+1} a_i x_i \text{ et } 0 \leq a_i \leq 1 \right\}.$$

La loi $\sum_{i=1}^{m+1} a_i \delta_{x_i}$ est donc la solution du problème d'entropie.

En particulier, si $c \in A \setminus \mathring{A}$ (au moins un des a_i est nul), cette loi n'a pas de densité de type exponentiel par rapport à μ . C'est un exemple de situation du premier point du théorème 2.3.

Exemple 2.4 Processus de Wiener

Solent $\Omega = C([0,1])$ l'ensemble des fonctions continues sur $[0,1]$ et $\mathcal{A}$ la tribu de ses Boréliens (pour $C([0,1])$ muni de la norme infinie). On munit $(\Omega, \mathcal{A})$ de la mesure de Wiener $\mu = W$. Soit $\phi \in \{L^2([0,1])\}^k$ (on suppose que ϕ est de rang k). On pose:

$$\phi^*(\omega) = \int_0^1 \phi(x) \, d\omega_x.$$

$\omega = (\omega_t)_{t \in [0,1]}$ désigne ici l'élément générique de $C([0,1])$. C'est-à-dire que lorsque $(C([0,1]), \mathcal{B}(C([0,1])))$ est considéré comme un espace de probabilité

muni de la probabilité μ , $(\omega_t)_{t \in [0,1]}$ est le mouvement Brownien standard, et donc, ϕ^* est l'intégrale stochastique classique de ϕ ([Dac-Duf, 83], paragraphe 1.2.3, p 20). La contrainte linéaire est alors $E_P\{\phi^*(\omega)\} = c$, (c fixé dans $\mathbb{R}^k$), c'est-à-dire

$$E_P\left\{\int_0^1 \phi(x) d\omega_x\right\} = c.$$

Le théorème de Girsanov ([Dac-Duf, 83], théorème 8.5.3.1, p 262) donne directement

$$Z_{W,\phi^*}(t) = E_W\{\exp(\langle t, \phi^*(\omega) \rangle)\} = \exp \frac{1}{2} \int_0^1 \langle t, \phi(x) \rangle^2 dx.$$

Remarquons que l'on aurait pu faire le calcul directement puisque sous W , ϕ^*

suit la loi $\mathcal{N}_k(0, \Gamma)$ avec $\Gamma = \left(\int_0^1 \phi_i(x) \phi_j(x) dx \right)_{i=1 \dots k, j=1 \dots k}$

Les hypothèses du corollaire 2.1 sont facilement vérifiables et l'on obtient

$$P^{ME}(\omega) = (Z_{W,\phi^*}(t^*(c)))^{-1} \exp[\langle t^*(c), \phi^*(\omega) \rangle] W(\omega),$$

où $t^*(c)$ est déterminé par l'équation

$$E_{P^{ME}}\left\{\int_0^1 \phi(x) d\omega_x\right\} = \int_{\Omega} \phi^*(\omega) dP^{ME}(\omega) = c.$$

Sous P^{ME} , $(\omega_t)_{t \in [0,1]}$ est solution de l'équation différentielle stochastique

$$d\omega_x = \langle t^*(c), \phi(x) \rangle dx + dB_x \text{ où } B \text{ est le Brownien standard.}$$

En particulier, on a:

$$E_{P^{ME}}[\omega_x] = \int_0^x \langle t^*(c), \phi(u) \rangle du \quad \text{et} \quad \frac{d}{dx} E_{P^{ME}}[\omega_x] = \langle t^*(c), \phi(x) \rangle$$

D'autre part (corollaire 2.1)

$$\begin{aligned} E_{P^{ME}}\left\{\int_0^1 \phi(x) d\omega_x\right\} &= \text{grad Log } Z_{W,\phi^*}(t^*(c)) \\ &= \int_0^1 \langle t^*(c), \phi(x) \rangle \phi(x) dx. \end{aligned}$$

Posant $g^{ME}(x) = \frac{a}{dx} E_P^{ME}[\omega_x] = \langle t^*(c), \phi(x) \rangle$,

on voit que g^{ME} est la solution de norme L^2 minimum du problème: trouver g de $L^2([0,1])$ qui vérifie la contrainte linéaire

$$\int_0^1 g(x) \phi(x) dx = c.$$

Nous reviendrons au chapitre III sur cet exemple.

II.3.2 Généralisation à des contraintes linéaires relaxées

On désire ici généraliser les résultats du corollaire 2.1 lorsque les contraintes sur la probabilité à reconstruire sont du type:

$$\int_{\Omega} \phi(\omega) dP(\omega) \in K$$

où K est un convexe compact de $\mathbb{R}^k$ donné. C'est-à-dire que l'on désire trouver des conditions suffisantes pour garantir l'existence de la probabilité du maximum de μ -entropie sur l'ensemble,

$$P(K, \phi) = \{ P \in \mathcal{P}(\Omega), E_P(\phi) = \int_{\Omega} \phi dP \in K \}.$$

Pour chaque c de K , on va utiliser le théorème 2.3 pour résoudre le problème de maximisation de S_{μ} sur $\mathcal{P}(c, \phi)$.

D'un point de vue pratique face à un problème de reconstruction de densité la mesure physique conduit à une région de confiance pour les valeurs des espérances. Généralement, quand on fait l'hypothèse que le bruit de mesure est blanc et gaussien, la région de confiance K est un ellipsoïde du type

$$K = \{ \alpha \in \mathbb{R}^k, \sum_{j=1}^k \frac{1}{\sigma_j^2} |\alpha_j - c_j|^2 \leq \chi_{\beta, k}^2 \},$$

où $\chi_{\beta, k}^2$ est le quantile d'ordre β d'une loi $\chi^2(k)$ et σ_j^2 est une variance. Enonçons maintenant le corollaire généralisant les résultats précédents. On pose,

$$\Sigma_{\mu}(K) = \{ c \in K, \exists P \in \mathcal{P}(c, \phi), P \sim \mu \}.$$

Corollaire 2.2 *On suppose que:*

a - $\Sigma_\mu(K)$ est non vide.

b- Si $c \in K$ et $c \notin \Sigma_\mu(K)$, alors pour toute P de $\mathcal{P}(c, \phi)$,
 P est étrangère à μ .

c- $D(\mu, \phi) = \mathbb{R}^k$.

Alors S_μ possède un unique maximum P^{ME} sur $\mathcal{P}(K, \phi)$. On a $P^{ME} = \mu_{\phi, t^*(K)}$,
 où $t^*(K)$ est l'unique minimum de la fonction

$$H_{\mu, \phi}(t, K) = \text{Log } Z_{\mu, \phi}(t) - \inf_{c \in K} \langle t, c \rangle.$$

Démonstration

Remarquons tout d'abord que, compte-tenu de la remarque 2.1, on peut toujours supposer que $(\Omega, \mathcal{A}) = (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Pour c de K , d'après le théorème 2.3, le corollaire 2.1 et les hypothèses il n'y a que deux possibilités:

1- $c \in \Sigma_\mu(K)$ et P^{ME} est $\mu_{\phi, t^*(c)}$, avec de plus $S_\mu(P^{ME}) = H_{\mu, \phi}(t^*(c), c)$

2- $c \notin \Sigma_\mu(K)$ et P^{MEG} n'existe pas, on a alors (compte-tenu de la remarque préliminaire et en utilisant le théorème II.5.1, p 42 de [E1, 85])

$$\sup_{P \in \mathcal{P}(c, \phi)} S_\mu(P) = - \inf_{t \in \mathbb{R}^k} H_{\mu, \phi}(t, c) = -\infty$$

On a:

$$\sup_{P \in \mathcal{P}(K, \phi)} S_\mu(P) \leq \sup_{c \in K} \sup_{P \in \mathcal{P}(c, \phi)} S_\mu(P).$$

En effet, soit $(P_n)_{n \geq 1}$ une suite de $\mathcal{P}(K, \phi)$ qui vérifie,

$$\lim_{n \rightarrow +\infty} S_\mu(P_n) = \sup_{P \in \mathcal{P}(K, \phi)} S_\mu(P), \quad \text{posons alors, } c^{(n)} = E_{P_n}(\phi).$$

$$\text{On a l'inégalité, } S_\mu(P_n) \leq \sup_{P \in \mathcal{P}(c^{(n)}, \phi)} S_\mu(P) \leq \sup_{c \in K} \sup_{P \in \mathcal{P}(c, \phi)} S_\mu(P).$$

Par passage à la limite on obtient l'inégalité annoncée. La fonction h de $\mathbb{R}^k$ dans $\mathbb{R} \cup \{-\infty\}$ définie par

$$h(c) = \inf_{t \in \mathbb{R}^k} H_{\mu, \phi}(t, c),$$

est concave, semi-continue supérieurement. Elle admet au moins un maximum c^* sur le domaine compact K .

On a,

$$\sup_{c \in K} \{ \inf_{t \in \mathbb{R}^k} H_{\mu, \phi}(t, c) \} = \sup_{c \in \Sigma_{\mu}(K)} H_{\mu, \phi}(t^*(c), c) = H_{\mu, \phi}(t^*(c^*), c^*)$$

$\mu_{\phi, t^*(c^*)}$ est donc une solution du problème de maximisation de S_{μ} sur $\mathbb{P}(K, \phi)$, comme de plus S_{μ} est une fonctionnelle strictement concave et que $\mathbb{P}(K, \phi)$ est un ensemble convexe, c'est l'unique solution du problème d'optimisation. Puisque K est compact et compte-tenu de l'hypothèse c, le corollaire 37.3.2 de [Rock, 70] (p. 393), de permutation de minimax donne,

$$H_{\mu, \phi}(t^*(c^*), c^*) = \inf_{t \in \mathbb{R}^k} \{ \sup_{c \in K} H_{\mu, \phi}(t, c) \}.$$

On obtient d'autre part facilement que $\sup_{c \in K} H_{\mu, \phi}(t, c) = H_{\mu, \phi}(t, K)$.

$H_{\mu, \phi}(\cdot, K)$ est strictement convexe, elle admet donc $t^*(c^*)$ comme unique minimum.

Remarque 2.2 Transformée de Cramer

Lorsque μ est une probabilité, la démonstration précédente n'utilise que le lien entre information de Kullback et transformée de Cramer (de la loi de ϕ sous la probabilité μ) et les propriétés de cette transformée.

Remarque 2.3 Un théorème de I. Csiszar

Dans le cadre très général où Ω est un espace vectoriel topologique localement convexe, I. Csiszar [Csis, 84] caractérise P^{MEG} lorsque μ est une probabilité et que la contrainte linéaire relaxée est du type $E(P) \in K$, où K est un convexe donné de Ω et $E(P)$ désigne la moyenne (ou résultante) de la probabilité P , c'est-à-dire que $E(P)$ est définie par

$$E(P) = \omega_0 \text{ si pour tout } t \text{ de } \Omega', E_P(\langle t, \omega \rangle_{\Omega', \Omega}) = \langle t, \omega_0 \rangle_{\Omega', \Omega}$$

(Ω' désigne le dual topologique de Ω et $\langle \cdot, \cdot \rangle_{\Omega', \Omega}$ est le produit de dualité).

Le résultat (théorème 3) est que, si l'intérieur de K a une intersection non vide avec l'enveloppe convexe du support de μ , alors P^{MEG} est de la forme

$$\frac{dP^{\text{MEG}}}{d\mu} = \frac{\exp(\langle t^*(K), \omega \rangle_{\Omega', \Omega})}{E_{\mu}\{\exp(\langle t^*(K), \omega \rangle_{\Omega', \Omega})\}},$$

où $t^*(K)$ est l'unique minimum de

$$H_{\mu, \omega}(t, K) = \text{Log } E_{\mu}\{\exp(\langle t, \omega \rangle_{\Omega', \Omega})\} - \inf_{\omega \in K} \langle t, \omega \rangle, \quad (t \in \Omega').$$

En considérant la loi de ϕ sous la probabilité μ , son théorème est donc beaucoup plus général que le corollaire 2.2, mais il ne donne pas des conditions suffisantes pour que P^{ME} existe. Remarquons enfin que dans le cadre précédent l'hypothèse du théorème de I. Csiszar est équivalente à l'hypothèse a du corollaire 2.2.

Exemples

Exemple 2.5 Loi normale

On reprend l'exemple 2.2 et l'on définit K par

$$K = \{ c \in \mathbb{R}^2, c_1^2 + c_2^2 \leq R^2 \}$$

En utilisant le corollaire précédent on trouve que la loi du maximum d'entropie est la loi normale centrée de variance R . Si maintenant on impose la contrainte: $-R \leq m \leq R$, et $0 \leq \sigma^2 \leq R$, l'ensemble

$$\mathbb{P}(K, \phi) = \{ P \in \mathbb{P}(\mathbb{R}), \int_{\mathbb{R}} x dP(x) = m, \int_{\mathbb{R}} x^2 dP(x) = m^2 + \sigma^2, (m, \sigma^2) \in K \}$$

avec, $K = \{ (m, \sigma^2) \in \mathbb{R}^2, -R \leq m \leq R, \text{ et } 0 \leq \sigma^2 \leq R \}$,

n'est plus convexe, et toute loi normale de moyenne m ($-R \leq m \leq R$) et de variance R est une solution du problème de maximisation de l'entropie. En effet, l'entropie d'une loi normale ne dépend que de sa variance (invariance par translation).

Exemple 2.6 Mesure discrète

Les simulations de E. Gassiat [Gas, 86] montrent l'intérêt de la méthode du maximum d'entropie lorsque l'on mesure une partie finie du spectre et que l'on désire reconstruire une loi absolument continue proche de " p/σ ". Une mesure de probabilité discrètes à n Diracs est complètement déterminée par la connaissance de ces n premiers coefficients de Fourier. L'estimation d'une telle mesure par une probabilité absolument continue ayant les mêmes coefficients de Fourier est impossible. On va donc relaxer les contraintes et appliquer le principe du maximum d'entropie:

$$\text{Soit } \nu = \sum_{l=1}^n a_l \delta_{x_l} \text{ avec } \sum_{l=1}^n a_l = 1, 0 < a_l \leq 1.$$

On mesure $c_j = \sum_{l=1}^n a_l \exp(ijx_l)$, $j = 1 \dots k$, $k \geq n$.

on veut alors résoudre le problème suivant : trouver une densité de probabilité p sur $[0, 2\pi]$ qui vérifie :

$$\sum_{j=1}^k \left| \int_0^{2\pi} p(x) \exp(ijx) dx - c_j \right|^2 \leq \varepsilon^2$$

et qui maximise l'entropie (par rapport à la mesure de Lebesgue)

$$S_L(p) = - \int_0^{2\pi} p(x) \log p(x) dx$$

Le convexe K est donc la boule de centre c et de rayon ε . En appliquant le corollaire précédent la solution est :

$$p^{ME}(x) = \{Z(t^*(K))\}^{-1} \exp \left\{ \sum_{j=1}^k (t_j^*(K) e^{ijx} + \bar{t}_j^*(K) e^{-ijx}) \right\}$$

où $t^*(K)$ minimise

$$H(t, K) = \log Z(t) + \varepsilon \|t\| - \sum_{j=1}^k (t_j c_j + \bar{t}_j \bar{c}_j),$$

avec
$$Z(t) = \int_0^{2\pi} \exp \left\{ \sum_{j=1}^k (t_j e^{ijx} + \bar{t}_j e^{-ijx}) \right\} dx.$$

En conclusion l'utilisation du principe du maximum d'entropie avec une contrainte linéaire relaxée, permet en pratique d'estimer des mesures de probabilité pour lesquelles la reconstruction avec une contrainte d'égalité n'est pas possible. Par exemple, lorsque l'on estime une mesure discrète par densité, ou en présence de données bruitées.

II.4 AUTRES CONSTRUCTIONS OU LA DISTRIBUTION DU MAXIMUM D'ENTROPIE INTERVIENT

II.4.1 Grandes déviations

Rappelons la définition de propriété de grandes déviations pour une suite de probabilités ([El, 85] p. 35 définition II.3.1)

Définition 2.2 Soient Ω un espace métrique complet et séparable et $\mathcal{A}$ la sigma algèbre de ses Boréliens. Soit $(Q_n)_{n \geq 1}$ une suite de mesures de probabilité sur $(\Omega, \mathcal{A})$. $(Q_n)_{n \geq 1}$ possède la propriété de grandes déviations si il existe une suite de réels positifs $(a_n)_{n \geq 1}$ qui tend vers l'infini en même temps que n et une fonction I de Ω dans $\mathbb{R}^+$ telles que:

- a I est semi-continue inférieurement.
- b L'ensemble $\{x \in \Omega, I(x) \leq b\}$ est compact pour tout b de $\mathbb{R}^+$.
- c $\lim_{n \rightarrow +\infty} \frac{1}{a_n} \log Q_n(G) \leq - \inf_{x \in G} I(x)$, pour tout ensemble ouvert G de Ω .
- d $\lim_{n \rightarrow +\infty} \frac{1}{a_n} \log Q_n(K) \geq - \inf_{x \in K} I(x)$, pour tout ensemble fermé K de Ω .

I est appelée fonction d'entropie de la suite $(Q_n)_{n \geq 1}$.

Exemples

Nous présentons ici trois exemples de suites de probabilités possédant une propriété de grandes déviations. Les deux premiers sont des exemples classiques: grandes déviations de la loi d'une moyenne de vecteurs aléatoires indépendants équidistribués, grandes déviations de la mesure empirique d'un n -échantillon. Le troisième exemple traite des grandes déviations d'une somme pondérée de variables aléatoires réelles indépendantes équidistribuées (c'est donc une variante du premier exemple), cet exemple est traité en détail car il sera utilisé dans les chapitres suivants (voir le corollaire 3.5 et le théorème 4.2)

Exemple 2.7 Niveau 1 [E1, 85] théorème II.4.1 (p. 37)

Soit $X_1, X_2, \dots$ une suite de vecteurs aléatoires indépendants identiquement distribués à valeurs dans $\mathbb{R}^k$, soit Q la distribution de X_1 . On suppose que la transformée de Laplace de la loi de X_1 est définie sur $\mathbb{R}^k$, c'est-à-dire que:

$$Z_{Q,X}(t) = E_Q[\exp \langle t, X_1 \rangle] = \int_{\mathbb{R}^k} \exp(\langle t, x \rangle) Q(dx),$$

est finie pour tout t de $\mathbb{R}^k$. On pose alors

$$I^{(1)}(c) = - \inf_{t \in \mathbb{R}^k} H_{Q,X}(t, c).$$

Soient S_n la n -ième somme partielle $\sum_{j=1}^n X_j$, et $Q_n^{(1)}$ la loi du vecteur $\frac{S_n}{n}$.

Alors la suite $(Q_n^{(1)})_{n \geq 1}$ possède une propriété de grandes déviations avec $a_n = n$ et $I^{(1)}$ comme fonction d'entropie. Et donc sous les hypothèses du corollaire 2.2 on a :

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \text{Log } Q_n^{(1)}\left(\frac{S_n}{n} \in K\right) \geq K(P^{ME}, Q),$$

où P^{ME} réalise l'infimum de $K(\cdot, Q)$ sur $\mathcal{P}(K, \phi)$.

Exemple 2.8 Niveau 2 théorème II.4.3 de [El, 85] (p. 40)⁴

Reprenant les notations de l'exemple précédent, soit L_n la mesure empirique

$$L_n = \frac{1}{n} \sum_{j=1}^n \delta_{X_j}.$$

Pour $P \in \mathcal{P}(\mathbb{R}^k)$ posons $I^{(2)}(P) = K(P, Q)$. Alors $Q_n^{(2)}$ la distribution de L_n sur $\mathcal{P}(\mathbb{R}^k)$ a une propriété de grandes déviations avec $a_n = n$ et $I^{(2)}$ comme fonction d'entropie.

Exemple 2.9 Niveau 1 pondéré

Soit F une loi sur $\mathbb{R}$, on suppose que la transformée de Laplace de F $\exp\{\psi(\tau)\} = \int \exp(\tau y) dF(y)$, est définie pour tout τ réel. Soit ϕ une application continue sur $[0, 1]$ à valeurs dans $\mathbb{R}^k$. Soit $(\omega_i^j)_{i \in \mathbb{N}, j=1, \dots, l}$ un tableau triangulaire de variables aléatoires réelles tel que pour tout (i, j) ω_i^j a la loi F et ω_i^j est indépendant de ω_i^l , dès que l et j sont différents. On pose alors, pour n de $\mathbb{N}$

$$X_n = \frac{1}{n} \sum_{j=1}^n \phi\left(\frac{1}{n}\right) \omega_n^j = \phi_n \omega_n, \quad \phi_n = \frac{1}{n} \left(\phi\left(\frac{1}{n}\right) \right)_{j=1, \dots, n} \quad \text{et} \quad \omega_n = (\omega_n^1, \dots, \omega_n^n).$$

Calculons la transformée de Laplace de la loi du vecteur X_n . Pour t de $\mathbb{R}^k$, on a

⁴Voir aussi [Bret, 79].

$$\text{Log } E(\exp\{\langle t, X_n \rangle\}) = \text{Log } Z_{F^{\otimes n}, \phi_n \omega_n}(t) = \sum_{j=1}^n \psi\left(\frac{1}{n} \langle t, \phi\left(\frac{1}{n}\right) \rangle\right).$$

Posons $v = \frac{t}{n}$ et,

$$\Phi_{n, (\phi_n)}^{(F^{\otimes n})}(v) = \frac{1}{n} \text{Log } Z_{F^{\otimes n}, \phi_n \omega_n}(nv) = \frac{1}{n} \sum_{j=1}^n \psi\left(\langle v, \phi\left(\frac{1}{n}\right) \rangle\right).$$

Calculons la limite de cette suite de fonctions quand n tend vers l'infini

$$\Phi_{\infty, (\phi_n)}^{(F^{\otimes n})}(v) = \lim_{n \rightarrow +\infty} \Phi_{n, (\phi_n)}^{(F^{\otimes n})}(v) = \int_0^1 \psi(\langle v, \phi(x) \rangle) dx.$$

La fonction $\Phi_{n, (\phi_n)}^{(F^{\otimes n})}(\cdot)$ est, pour tout n , définie sur $\mathbb{R}^k$ (puisque l'on a supposé que la transformée de Laplace de F est définie sur $\mathbb{R}$), il en est de même pour la fonction $\Phi_{\infty, (\phi_n)}^{(F^{\otimes n})}(\cdot)$. Appliquant le théorème de Lebesgue, on déduit que cette dernière fonction est dérivable sur $\mathbb{R}^k$. Nous sommes donc sous les hypothèses du théorème II.6.1 de [El, 85] (p. 47) et la suite $(Q_n)_{n \geq 1}$ des lois des vecteurs X_n possède une propriété de grandes déviations avec $a_n = n$ et posant pour z de $\mathbb{R}^k$,

$$I(z) = \sup_{v \in \mathbb{R}^k} \left\{ \langle v, z \rangle - \int_0^1 \psi(\langle v, \phi(x) \rangle) dx \right\},$$

$I(\cdot)$ est la fonction d'entropie. En particulier on a une loi des grands nombres, la suite $(X_n)_{n \geq 1}$ converge en probabilité (cf [El, 85], théorème II.6.3, p 49) vers

$$\frac{\partial}{\partial v} \Phi_{\infty, (\phi_n)}^{(F^{\otimes n})}(0) = \int_0^1 \phi(x) \psi'(0) dx.$$

C'est-à-dire que, quand n grandit, la loi $F^{\otimes n}$ a "*tendance à se concentrer*" sur l'évènement

$$\{\omega_n \in \mathbb{R}^n, \phi_n \omega_n = \int_0^1 \phi(x) \psi'(0) dx\}.$$

Soit v^* de $\mathbb{R}^k$ fixé, on pose

$$c^* = \int_0^1 \phi(x) \psi(\langle v^*, \phi(x) \rangle) dx.$$

Considérons le tableau triangulaire précédent lorsque l'on fait le changement de probabilité: F est à remplacer par

$$\exp\{\langle v^*, \phi(\frac{1}{i}) \rangle \omega_i^j - \psi(\langle v^*, \phi(\frac{1}{i}) \rangle)\} F.$$

C'est-à-dire que le tableau triangulaire est tel que pour tout (i,j) , ω_i^j a la loi précédemment définie et que ω_i^j est indépendant de ω_i^l dès que j et l sont différents. On peut alors faire exactement le même raisonnement que précédemment, la nouvelle fonction d'entropie est

$$I(z) = \sup_{v \in \mathbb{R}^k} \left\{ \langle v, z \rangle - \int_0^1 \psi(\langle v + v^*, \phi(x) \rangle) dx \right\},$$

donc la suite $(X_n)_{n \geq 1}$ converge en probabilité vers

$$c^* = \int_0^1 \phi(x) \psi(\langle v^*, \phi(x) \rangle) dx.$$

Et donc la nouvelle probabilité a "*tendance à se concentrer*" sur l'événement

$$\left\{ \omega_n \in \mathbb{R}^n, \phi_n \omega_n = \int_0^1 \phi(x) \psi(\langle v^*, \phi(x) \rangle) dx = c^* \right\}.$$

Soit $[a,b]$ $(-\infty \leq a < b \leq +\infty)$, l'intérieur de l'enveloppe convexe du support de F , posons

$$\mathcal{C} = \{ e \in C([0,1]), \forall x \in [0,1], a < e(x) < b \} \text{ et } \mathcal{K} = \left\{ \int_0^1 \phi(x) e(x) dx, e \in \mathcal{C} \right\}.$$

Nous verrons au chapitre III (théorème 3.1) que pour tout c de $\mathcal{K}$, il existe un élément $v_\infty^*(c)$ de $\mathbb{R}^k$ qui vérifie

$$\int_0^1 \phi(x) \psi(\langle v_\infty^*(c), \phi(x) \rangle) dx = c.$$

Soit toujours c de $\mathcal{K}$, intéressons-nous à la suite des problèmes M.E.M sur $\mathbb{C}$ avec comme mesure de référence $F^{\otimes n}$ (voir sa définition dans l'introduction du chapitre I, ou au paragraphe III.4.2). Nous verrons au chapitre III que P_n^{ME} est de la forme

$$P_n^{\text{ME}}(\omega_n) = \prod_{i=1}^n \exp \{ \langle v_n^*(c), \phi(x_i) \rangle \omega_n^i - \psi(\langle v_n^*(c), \phi(x_i) \rangle) \} F^{\otimes n},$$

avec $\lim_{n \rightarrow +\infty} v_n^*(c) = v_\infty^*(c)$. Donc⁵ P_n^{ME} a "tendance à se concentrer" sur l'évènement

$$\{\omega_n \in \mathbb{R}^n, \phi_n \omega_n = \int_0^1 \phi(x) \psi'(\langle v_\infty^*(c), \phi(x) \rangle) dx = c\}.$$

II.4.2 Probabilités conditionnelles

La distribution du maximum d'entropie peut aussi être obtenue comme limite d'une suite de probabilités conditionnelles. Soit Q une loi absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Soit $X_1, \dots, X_n$ un échantillon de loi Q et ϕ une fonction mesurable de $\mathbb{R}$ dans $\mathbb{R}$, non Q -p.s constante. On a alors le théorème suivant dû à J.M. Van Campenhout et T.M. Cover [Van-Cov, 81]:

Théorème 2.4 *Soit c appartenant à l'intérieur du support de la loi de $\phi(X_1)$. On suppose que la transformée de Laplace de la loi de $\phi(X_1)$ est définie sur un ouvert. Alors*

$$\lim_{n \rightarrow +\infty} Q(X_1 \leq x / \frac{1}{n} \sum_{i=1}^n \phi(X_i) = c) = P^{\text{ME}}(-\infty, x) \quad (x \in \mathbb{R}),$$

où P^{ME} est la loi de Q -entropie maximale sous la contrainte $E_P(\phi) = c$.

Remarque 2.4 Extensions

Par souci de simplicité, nous avons énoncé le théorème précédent dans un cadre unidimensionnel et avec des hypothèses sous lesquelles P^{ME} existe. Dans [Csis, 84], I. Csiszar le généralise pour un système de contraintes multiples, et affaiblissant les hypothèses il remplace dans l'énoncé P^{ME} par

⁵Voir le début du paragraphe III.4 et le corollaire 3.5.

P^{ME} . C. Robert [Rob, 89] donne un théorème analogue pour la suite des lois conditionnelles d'une suite de mesures empiriques (le conditionnement étant du type la moyenne d'un vecteur aléatoire ϕ vaut c sous la loi empirique⁶).

Exemple 2.10 Loi discrète [Gzyl, 88]

Soit $X_1, \dots, X_n$ un échantillon de loi $Q(X_1 = x_i) = q_i$, $i=1, \dots, m$.

Alors,
$$\lim_{n \rightarrow +\infty} Q(X_1 = x_i / \frac{1}{n} \sum_{j=1}^n X_j = c) = P^{ME}(x_i)$$

où,
$$P^{ME}(x_i) = \frac{\exp(t^*(c)x_i) q_i}{\sum_{j=1}^m \exp(t^*(c)x_j) q_j}$$

et $t^*(c)$ minimise $H_{Q,X}(t, c) = \text{Log } Z_{Q,X}(t) - t c$.

⁶ ϕ et c sont donnés.

CHAPITRE III

LA METHODE DU MAXIMUM D'ENTROPIE SUR LA MOYENNE

III.1 INTRODUCTION

Pour reconstruire une densité g , à partir d'un nombre fini de moments généralisés du type

$$\int \phi_j(x) g(x) dx = c_j \quad j=1, \dots, k,$$

il a été souvent proposé la méthode du maximum d'entropie. Une des limitations de la méthode du maximum d'entropie est que si elle conduit toujours à une estimation positive de g , elle ne peut pas prendre en compte de manière utile des contraintes non linéaires aussi simples que:

$$a < g(x) < b \text{ ou } \int [g(x)]^2 dx < 1.$$

Ces contraintes non linéaires simples sont souvent définies par un ensemble convexe $\mathcal{C}$ d'un certain espace fonctionnel E . Il existe des cas où l'on peut encore utiliser la méthode du maximum d'entropie pour reconstruire la fonction g , en utilisant l'information à priori qu'elle appartient à $\mathcal{C}$. Pour cela, on se donne un espace fonctionnel mesurable $(\Omega, \mathcal{A})$ que l'on munit d'une mesure de référence μ et un opérateur linéaire R défini sur Ω et à valeurs dans $\mathcal{C}$. On cherche alors la probabilité P^{ME} , de μ - entropie maximale (portée par Ω) sous la contrainte

$$\int \phi(x) R\{E_P[\omega_x]\} dx = c.$$

La reconstruction est alors $g^{ME}(x) = R\{E_{P^{ME}}[\omega_x]\}$.

Nous avons déjà rencontré ce type de situation dans l'exemple 2.4. Nous donnons d'autres exemples au paragraphe III.2. D'autres exemples ont été traités dans [Dac-Gam, 89].

En général, il n'existe pas de mesure de référence μ "*raisonnable*", le plus souvent ce terme est attribué en référence à l'invariance de " $R\mu$ " pour certaines transformations sur $\mathcal{C}$ (l'expression " $R\mu$ " est entre guillemets car elle peut ne pas avoir de sens mathématique, par exemple dans l'exemple 2.4 " $R\mu$ " est "*la loi d'un bruit blanc sur $[0,1]$* "). Le choix d'une autre mesure à priori met en cause les fondements de la méthode du maximum d'entropie en introduisant une information à priori peu crédible. Il est donc plus intéressant de remplacer le problème de dimension infinie par une suite de problèmes de dimension finie pour lesquels les mesures de référence "*raisonnables*" existent. Décrivons alors la méthode qui sera développée au paragraphe III.3. On se donne deux suites d'opérateurs linéaires $(T_n)_{n \geq 1}$, $(R_n)_{n \geq 1}$ avec,

$$T_n : E \longrightarrow \mathbb{R}^n \text{ et } R_n : \mathbb{R}^n \longrightarrow F, \quad \text{avec } E \text{ inclus dans } F$$

(rappelons que E est un espace fonctionnel contenant $\mathcal{C}$ donné au départ, F est donné dans la construction via la suite $(R_n)_{n \geq 1}$). On pose alors

$$\mathcal{C}_n = T_n(\mathcal{C}).$$

La contrainte linéaire est discrétisée en $\phi_n T_n g(x) = c$, où ϕ_n est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^k$ approchant (dans un sens à préciser) ϕ . On munit $\mathcal{C}_n$ d'une mesure de référence μ_n , on calcule la probabilité P_n^{ME} (portée par $\mathcal{C}_n$) solution du problème de μ_n - entropie maximale sous la contrainte linéaire

$$E_{P_n}[\phi_n \omega_n] = \phi_n \int_{\mathcal{C}_n} \omega_n dP_n(\omega_n) = c,$$

où ω_n désigne l'élément générique de $\mathcal{C}_n$. La n -ième reconstruction est alors:

$$g_n^{\text{ME}} = R_n E_{P_n^{\text{ME}}}[\omega_n].$$

Bien évidemment le système (P_n^{ME}) n'est éventuellement cohérent que sur un ensemble du type $\mathbb{R}^I$ (I étant l'ensemble où varie x), et il ne définit pas une probabilité P^{ME} sur $\mathcal{C}$. Mais la suite des interpolations (g_n^{ME}) converge dans bon nombre de situations intéressantes vers une reconstruction g_∞^{ME} . C'est par exemple, le cas si $\mathcal{C}$ est une boule et E un espace de Hilbert, ou si $\mathcal{C}$ est une bande et E un espace de fonctions continues ou de variables aléatoires bornées

(voir le paragraphe III.4). cette procédure est en filigrane dans certaines techniques d'imagerie physique [Gull-Skil, 85], elle a été développée par J. Navaza [Nav, 85], [Nav, 86], dans le cadre du problème des phases en cristallographie.

Il est tout à fait remarquable que l'on puisse aussi obtenir directement g_{∞}^{ME} comme la solution d'un problème de maximisation d'une certaine fonctionnelle concave Γ sur l'intersection de $\mathcal{C}$ avec l'ensemble des fonctions vérifiant les contraintes linéaires. Cette méthode donne donc une vue plus unifiée des différents principes de reconstruction (moindres carrés, entropie usuelle, entropie de la prédiction, ...).

La fonctionnelle Γ possède des propriétés remarquables, en particulier elle est comme l'entropie liée à une divergence (pour l'entropie la divergence en question est l'information de Kullback). Ceci sera développé dans le paragraphe III.5.

III.2 EXEMPLES DE RECONSTRUCTION

III.2.1 Moments généralisés

Reprenons l'exemple 2.4. Soit μ la loi d'un processus à accroissements indépendants homogène défini sur $\mathbb{R}^{[0,1]}$ muni de sa filtration canonique. On s'intéresse ici au problème de reconstruction d'une fonction g définie sur $[0,1]$ dont on connaît les ϕ - moments,

$$\int_0^1 \phi_j(x) g(x) dx = c_j, \quad j=1, \dots, k,$$

où les ϕ_j sont des fonctions continues données. Nous écrirons les contraintes précédentes de façon plus condensée

$$\int_0^1 \phi(x) g(x) dx = c,$$

Formellement, si pour une loi P on a,

$$\int_0^1 \phi(x) [E_P(\omega_x)]' dx = c,$$

alors posant $g_P(x) = [E_P(\omega_x)]'$, g_P est une reconstruction possédant dans un sens à préciser les propriétés "des dérivées des trajectoires" sous la loi P .

Comme d'habitude, nous choisissons P maximisant la μ - entropie. En particulier, lorsque μ est la loi d'un P.A.I croissant, la reconstruction $g^{ME}(x) = [E_{P^{ME}}(\omega_x)]'$ est positive.

Exemples

Exemple 3.1 Processus de Wiener

Reprenons l'exemple 2.4, c'est un problème d'entropie du type précédent avec μ la mesure de Wiener. On avait trouvé que,

$$g^{ME}(x) = \frac{d}{dx} E_{P^{ME}}(\omega_x) = \langle t^*(c), \phi(x) \rangle$$

g^{ME} est la fonction de norme L^2 minimum qui vérifie les contraintes de ϕ - moments.

Exemple 3.2 Processus de Poisson

Intéressons-nous au problème précédent lorsque la mesure μ est N , la loi du processus de Poisson homogène d'intensité 1. La contrainte linéaire s'écrit :

$$E_P \left[\int_0^1 \phi(x) d\omega_x \right] = c.$$

Utilisant le corollaire 2.1, la probabilité maximisant la μ - entropie sous la contrainte linéaire précédente est :

$$\frac{dP^{ME}}{dN} = \exp \left[1 - \int_0^1 \exp \langle t^*(c), \phi(x) \rangle dx + \int_0^1 \langle t^*(c), \phi(x) \rangle d\omega_x \right],$$

où $t^*(c)$ minimise la fonction : $H_{\mu, \phi}(t, c) = \int_0^1 \exp \langle t, \phi(x) \rangle dx - \langle t, c \rangle$.

On en déduit que : $g^{ME}(x) = \exp \langle t^*(c), \phi(x) \rangle$. g^{ME} est donc une reconstruction positive, ce qui n'a rien d'étonnant puisque c'est la dérivée de l'espérance d'un processus croissant. C'est la reconstruction d'information de Kullback par rapport à la mesure de Lebesgue minimale, c'est-à-dire que :

$$g^{ME} \text{ minimise } K(g, L) = \int_0^1 g(x) \log g(x) dx - \int_0^1 g(x) dx + 1,$$

où L désigne la mesure de Lebesgue sur $[0, 1]$.

On voit sur cet exemple que lorsque l'on veut une reconstruction appartenant au convexe, $\mathcal{C} = \{ e \in C([0, 1]), \forall x \in [0, 1], e(x) > 0 \}$, on peut utiliser la formulation précédente avec un P.A.I croissant.

Remarque 3.1 P.A.I non homogène

Dans les deux exemples précédents, si l'on ne suppose plus les processus homogènes, g^{ME} est:

- la reconstruction de norme $L^2([0,1], \frac{dx}{f(x)})$ minimum dans l'exemple 3.1. f est la fonction de covariance ($E(\omega_s \omega_t) = f(\min(s,t))$), on suppose qu'elle est strictement croissante, dérivable sur $[0,1]$ et nulle en 0.
- La reconstruction de minimum d'information de Kullback par rapport à f dans l'exemple 3.2. f est l'intensité du processus de Poisson, f est une fonction intégrable et strictement positive sur $[0,1]$.

Remarque 3.2 "la loi d'un bruit blanc"

Heuristiquement, la construction précédente revient à prendre comme mesure de référence pour l'utilisation du principe du maximum d'entropie, "la loi d'un bruit blanc" c'est-à-dire que la dépendance entre les valeurs de la reconstruction g^{ME} en des points différents ne provient que des contraintes de ϕ -moments. En l'absence d'informations à priori supplémentaires, l'utilisation d'une autre mesure de référence peut mettre en cause le principe du maximum d'entropie (introduction d'informations arbitraires).

III.2.2 Application: Interpolation par fonctions splines

Supposons maintenant que nous nous intéressions à un problème d'interpolation du type: trouver g du convexe

$$\mathcal{C} = \{ e \in C^\alpha([0,1]), e(0) = e'(0) = \dots = e^{(\alpha)}(0) = 0 \}$$

qui vérifie $g(x_j) = c_j$, $j=1, \dots, k$, où les x_j et les c_j sont donnés ($c_j \in \mathbb{R}$, $x_j \in [0,1]$).

Soit X_α le processus de Wiener intégré α fois, c'est-à-dire le processus:

$$X_\alpha = \int_0^x \frac{(x-t)^{\alpha-1}}{\alpha-1} d\omega_t.$$

Pour résoudre le problème d'interpolation précédent, on cherche la probabilité P^{ME} de minimum d'information de Kullback par rapport à la mesure de Wiener W , sous les contraintes

$$E_P \left[\int_0^{x_j} \frac{(x_j - t)^\alpha}{\alpha!} d\omega_t \right] = c_j, \quad j=1, \dots, k.$$

Remarquons que compte-tenu de la remarque 2.1, cela revient à chercher la loi d'information de Kullback par rapport à la loi du processus $(X_x)_{x \in [0,1]}$ minimale sous les contraintes $E_P[X_{x_j}] = c_j, \quad j=1, \dots, k.$

On pose ensuite

$$g^{ME}(x) = E_{P^{ME}}[X_x], \quad \text{si } P^{ME} \text{ existe.}$$

g^{ME} est une solution du problème d'interpolation. Par application directe des exemples 3.1 et 2.4, on trouve:

$$g^{ME}(x) = \sum_{j=1}^k t_j^* \int_0^{\min(x, x_j)} \frac{(x-t)^\alpha (x_j-t)^\alpha}{(\alpha!)^2} dt,$$

où le vecteur t^* est calculé de manière à satisfaire les contraintes. g^{ME} est la reconstruction par fonctions Splines d'ordre α ([Schoenb, 64]), c'est-à-dire que g^{ME} est la fonction de $\mathcal{C}$ vérifiant les contraintes d'interpolation et minimisant la norme L^2 de la dérivée d'ordre $(\alpha + 1)$.

Remarques

Remarque 3.3 Construction statistique des fonctions splines

On trouve dans [Wahba, 83] une construction Bayésienne de la reconstruction par fonctions Splines dans le modèle de régression

$$y_x = g(x) + \varepsilon_x$$

g étant considéré comme un paramètre infini-dimensionnel, et ε_x un "bruit blanc" Gaussien. La loi à priori sur le paramètre g est la loi du processus de Wiener intégré α fois. On observe le processus $(y_x)_{x \in [0,1]}$ en k points

$$y_{x_1} = c_1, \dots, y_{x_k} = c_k$$

on estime alors g par $\hat{g}(x) = E(y_x / y_{x_1} = c_1, \dots, y_{x_k} = c_k)$ et on trouve que $\hat{g}$ minimise le critère

$$\|f^{(\alpha+1)}\|_L^2 + \lambda \sum_{j=1}^k |f(x_j) - c_j|^2$$

où λ est un rapport de variance. Il est tout à fait remarquable que l'on retrouve la même solution en utilisant le principe du maximum d'entropie

pour un système de contraintes d'interpolation relaxées (corollaire 2.2). ceci est en fait une conséquence des égalités:

$$E(\exp(X/Z)) = E(\exp(X)) E(X/Z), \quad E(X/Z) = \frac{E(XZ)}{\text{VAR}(Z)} Z,$$

valables si (X, Z) est un vecteur Gaussien centré.

Remarque 3.4 Conditions au bord

C'est pour simplifier les calculs que nous avons imposé les valeurs des dérivées en 0. Pour supprimer ces contraintes on peut utiliser un processus du type:

$$X_x = \int_0^x \frac{(x-t)^\alpha}{\alpha!} d\omega_t + \sum_{j=0}^{\alpha} Y_j x^j.$$

Les Y_j étant des variables aléatoires à préciser.

Remarque 3.5 ϕ - moments

La fonction g^{ME} que nous venons de construire peut aussi être construite directement par intégrations successives de la solution du problème de ϕ - moments de l'exemple 3.1, pour la famille de fonctions

$$\phi_j(x) = \frac{1}{|0, x_j|} (x), \quad j=1, \dots, k.$$

Nous allons maintenant étendre les constructions précédentes à des ensembles fonctionnels pour lesquels il n'existe pas de mesure de référence "raisonnable".

III.3 LA SUITE DES PROBLEMES M.E.M POUR UN CONVEXE D'UN ESPACE FONCTIONNEL.

III.3.1 Définition du problème.

Soient E un espace de Banach, ϕ un élément fixé de $(E')^k$ (E' désigne le dual topologique de E), $\mathcal{C}$ un convexe ouvert de E et c un élément de $\mathbb{R}^k$ donnés. On s'intéresse à la reconstruction d'un élément g de E vérifiant:

$$g \in \mathcal{C} \quad \text{et} \quad \langle \phi, g \rangle_{E', E} = c,$$

où $\langle \cdot, \cdot \rangle_{E', E}$ désigne la dualité E', E et $\langle \phi, \cdot \rangle_{E', E}$ est le vecteur de j -ième composante $\langle \phi_j, \cdot \rangle_{E', E}$.

Exemples

Ces exemples se poursuivent un peu plus bas et seront étudiés en détail au paragraphe III.4.

Exemple 3.3 Espace de Hilbert

E est un espace de Hilbert séparable, $\mathcal{C}$ sa boule unité et ϕ un élément de E^k .

Exemple 3.4 Bande d'un espace de fonctions continues

E est ici $C(U)$ ensemble des fonctions continues sur U compact de $\mathbb{R}^1$ donné, $\mathcal{C}$ est la bande:

$$\mathcal{C} = \{e \in C(U), \forall x \in U, a(x) < e(x) < b(x)\}$$

et ϕ est une fonction continue sur U à valeurs dans $\mathbb{R}^k$.

On voit sur ces exemples que le procédé de reconstruction utilisant directement le principe du maximum d'entropie (en reprenant les notations de l'introduction de ce chapitre R est ici l'identité), c'est-à-dire celui qui consiste à reconstruire g par

$$g^{ME} = \int_{\mathcal{C}} \omega \, dP^{ME}(\omega),$$

où P^{ME} maximise la μ - entropie (μ une mesure de référence), et vérifie la

$$\text{contrainte, } c = \langle \phi, \int_{\mathcal{C}} \omega \, dP^{ME}(\omega) \rangle_{E', E},$$

est difficile à mettre en oeuvre (il faut en premier lieu donner un sens aux expressions entre guillemets). D'autre part, il n'existe pas en général de mesure de référence μ sur $\mathcal{C}$ "raisonnable". C'est-à-dire invariante par certaines transformations. Par exemple, on peut exiger pour l'exemple 3.3 que μ soit invariante par toute isométrie de E dans E . Pour ces raisons, on construit une suite de problèmes de maximum d'entropie, sur des espaces de dimension finie approchant l'espace E . On utilise pour cela le schéma suivant: on se donne un quadruplé: $((T_n)_{n \geq 1}, (R_n)_{n \geq 1}, (\mu_n)_{n \geq 1}, (\phi_n)_{n \geq 1})$, avec pour tout n , T_n (resp. R_n) est un opérateur linéaire de E dans $\mathbb{R}^n$ (resp. de $\mathbb{R}^n$ dans F), et F contient E , $(\phi_n)_{n \geq 1}$ est une suite d'éléments de $\mathbb{R}^n \times \mathbb{R}^k$ approchant l'application

ϕ dans un sens que l'on précisera plus bas. Enfin, μ_n est une mesure positive sur $\mathcal{C}_n = T_n(\mathcal{C})$ qui possède les propriétés d'invariance. Le n -ième problème d'entropie consiste alors à trouver la probabilité P_n^{ME} du maximum de μ_n -entropie vérifiant les contraintes linéaires:

$$\phi_n \int_{\mathcal{C}_n} \omega_n dP_n(\omega_n) = c.$$

La n -ième reconstruction (en supposant que P_n^{ME} existe) est alors :

$$g_n^{ME} = R_n \int_{\mathcal{C}_n} \omega_n dP_n^{ME}(\omega_n).$$

On appelle le problème précédent le n -ième problème de maximisation de l'entropie sur la moyenne (M.E.M), car la contrainte linéaire ne porte que sur la moyenne de la loi.

Définition 3.1

a- On appelle suite de problème de maximisation de l'entropie sur la moyenne associée au quadruplé $((T_n)_{n \geq 1}, (R_n)_{n \geq 1}, (\mu_n)_{n \geq 1}, (\phi_n)_{n \geq 1})$ la suite des problèmes précédents.

b- On dit que la suite des problèmes de maximisation de l'entropie sur la moyenne converge s'il existe un élément g_∞^{ME} de $\mathcal{C}$ vérifiant:

$$\lim_{n \rightarrow +\infty} g_n^{ME} = g_\infty^{ME}, \text{ le sens de la limite étant à préciser et, } \langle \phi, g_\infty^{ME} \rangle_{E', E} = c.$$

Remarque 3.6 La suite (P_n^{ME})

La définition b n'implique pas la convergence de la suite (P_n^{ME}) vers une probabilité P_∞^{ME} sur $\mathcal{C}$.

Avant de nous intéresser à la résolution de n -ième problème d'entropie détaillons la construction du M.E.M sur les exemples précédents.

Exemples

Exemple 3.3 $\mathcal{C}$ la boule unité d'un espace de Hilbert séparable (suite)

Soit $(e_n)_{n \geq 1}$ une base Hilbertienne, la discrétisation naturelle de E est donnée par:

$$T_n e = (\langle e_1, e \rangle_{E', E}, \dots, \langle e_n, e \rangle_{E', E}), e \in E$$

$$\text{et, } R_n \omega_n = \sum_{i=1}^n \omega_n^i e_i, \omega_n = (\omega_n^1, \dots, \omega_n^n) \in \mathbb{R}^n.$$

En particulier, R_n est à valeurs dans E , ($F = E$). ϕ_n est définie par:

$$\forall \omega_n \in \mathbb{R}^n, \phi_n \omega_n = \sum_{i=1}^n \omega_n^i \langle e_i, \phi \rangle_{E',E}.$$

Rappelons que $\langle \cdot, \phi \rangle_{E',E}$ est le vecteur de j -ième composante $\langle \cdot, \phi_j \rangle_{E',E}$.
Le théorème de Parseval donne:

$$\forall e \in E, \lim_{n \rightarrow +\infty} \phi_n T_n e = \langle \phi, e \rangle_{E',E}.$$

Le convexe,

$$C_n = \{ \omega_n = (\omega_n^1, \dots, \omega_n^n) \in \mathbb{R}^n, \sum_{i=1}^n (\omega_n^i)^2 < 1 \},$$

image de C par l'application T_n , est muni de la probabilité uniforme.

Exemple 3.4 Bande d'un espace de fonctions continues (suite)

On se donne une suite $(x_i)_{i \geq 1}$ d'éléments de U qui vérifie la condition:

$$\forall e \in C(U), \frac{1}{n} \sum_{i=1}^n e(x_i) \xrightarrow[n \rightarrow \infty]{} \int_U e(x) \frac{dx}{|U|}$$

où l'on a posé,

$$|U| = \int_U dx.$$

L'existence d'une telle suite sera discutée en III.3. Remarquons que la condition précédente entraîne que la suite $(x_i)_{i \geq 1}$ est dense dans U . On définit ensuite sur $C(U)$ l'opérateur T_n par:

$$T_n e = (e(x_1), \dots, e(x_n)), e \in C(U)$$

et R_n par:

$$(R_n \omega_n)(x) = \frac{1}{\#J_n(x)} \sum_{j \in J_n(x)} \omega_n^j, \quad \omega_n = (\omega_n^1, \dots, \omega_n^n) \in \mathbb{R}^n, \quad x \in U,$$

où $J_n(x)$ est l'ensemble d'indices:

$$J_n(x) = \{ j \in \{1, \dots, n\}, \min_{i=1..n} \|x - x_i\| = \|x - x_j\| \},$$

et $\#J_n(x)$ désigne le cardinal de $J_n(x)$.

En particulier on a, $(R_n \omega_n)(x_j) = \omega_n^j$, $j=1, n$ et R_n est à valeurs dans $F=L^\infty(U, dx)$. La suite $(\phi_n)_{n \geq 1}$ des approximations de ϕ se définit naturellement par

$$\phi_n \omega_n = \frac{|U|}{n} \sum_{i=1}^n \phi(x_i) \omega_n^i.$$

On voit en particulier que l'hypothèse faite sur la suite $(x_j)_{j \geq 1}$ entraîne que

$$\forall e \in \mathcal{C}, \lim_{n \rightarrow +\infty} \phi_n T_n e = \int_U \phi(x) e(x) dx.$$

Enfin, on se donne une famille $(F_x)_{x \in U}$ de mesures positives sur $\mathbb{R}$, et l'on munit

$$\mathcal{C}_n = \prod_{i=1}^n |a(x_i), b(x_i)|$$

de la mesure de référence $\mu_n = \bigotimes_{i=1}^n F_{x_i}$.

Nous verrons que, sous des hypothèses sur la famille $(F_x)_{x \in U}$, la suite des problèmes MEM converge presque partout.

III.3.2 Résolution du n-ième problème d'entropie sur la moyenne

Dans ce paragraphe, on donne des conditions suffisantes pour garantir l'existence et l'unicité de la solution du n-ième problème d'entropie sur la moyenne. On s'intéresse aussi à la description de la loi du maximum d'entropie ainsi qu'à la n-ième reconstruction g_n^{ME} . Nous avons vu au chapitre II que pour garantir l'existence et l'unicité de la solution d'un problème de maximisation de l'entropie il suffit de vérifier deux conditions. Premièrement, l'existence d'une mesure équivalente à la mesure de référence et vérifiant la contrainte linéaire, deuxièmement, une hypothèse sur l'ensemble de définition de la transformée de Laplace de la loi de la variable aléatoire définissant la contrainte. Nous allons vérifier ces hypothèses pour le n-ième problème d'entropie sur la moyenne.

On fera toujours l'hypothèse que ϕ est de rang plein dans le sens suivant:

$$\alpha \in \mathbb{R}^k, \forall e \in E, \langle \alpha, \phi \rangle, e \rangle_{E,E} = 0 \Rightarrow \alpha = 0.$$

Remarquons que l'on peut toujours faire cette hypothèse, en effet, si ϕ n'est pas de rang plein, on en extrait un système de rang plein de dimension maximum et on exprime les autres composantes de ϕ sur ce système.

Définition 3.2 Soit c un élément de $\mathbb{R}^k$. On dit que c est ϕ -réalisable (resp. strictement réalisable) dans $\mathcal{C}$ s'il existe un élément e de $\mathcal{C}$ (resp. de $\text{cl}(\mathcal{C})$) qui vérifie $\langle \phi, e \rangle_{E^*E} = c$, où $\text{cl}(\mathcal{C})$ désigne dans un sens que l'on précisera la fermeture de $\mathcal{C}$.

Enonçons maintenant un lemme relatif à la réalisation de la contrainte discrétisée.

Lemme 3.1 On suppose que les hypothèses suivantes sont vérifiées:

$$(H1) \quad \forall e \in E, \lim_{n \rightarrow +\infty} \phi_n T_n e = \langle \phi, e \rangle_{E^*E}.$$

$$(H2) \quad c \text{ est } \phi \text{ strictement réalisable dans } \mathcal{C}.$$

Alors il existe un entier n_0 et une suite $(\beta_n)_{n \geq n_0}$ qui satisfont à:

$$\forall n, n \geq n_0, \beta_n \in \mathcal{C}_n \text{ et } \phi_n \beta_n = c.$$

Le lemme précédent dit que si c est ϕ strictement réalisable dans $\mathcal{C}$, et si ϕ_n est une approximation "raisonnable" de ϕ , alors, pour n assez grand, c est ϕ_n -réalisable dans $\mathcal{C}_n$.

Démonstration

Soit e de $\mathcal{C}$ qui vérifie $\langle \phi, e \rangle_{E^*E} = c$, et soit ε réel strictement positif tel que la boule $B(e, \varepsilon)$ de centre e , de rayon ε soit incluse dans $\mathcal{C}$. Notons B_n , l'image de $B(e, \varepsilon)$ par T_n . B_n est un ouvert convexe non vide inclus dans $\mathcal{C}_n$.

Raisonnons par l'absurde: Supposons qu'il existe une suite d'entiers $(n_j)_{j \geq 1}$ de limite $+\infty$ vérifiant

$$\forall j, \forall \beta_{n_j} \in \mathcal{C}_{n_j}, \quad \phi_{n_j} \beta_{n_j} \neq c.$$

L'ensemble $\{ \phi_{n_j} \beta_{n_j}, \beta_{n_j} \in B_{n_j} \}$ est un convexe de $\mathbb{R}^k$ qui ne contient pas le point c . Il existe donc un hyperplan $\alpha_{n_j} = (\alpha_{n_j}^1, \dots, \alpha_{n_j}^k)$ de norme 1 qui le sépare de c . On a donc par exemple,

$$\forall \delta \in B(e, \varepsilon), \langle \alpha_{n_j}, c \rangle \geq \langle \alpha_{n_j}, \phi_{n_j} T_{n_j} \delta \rangle.$$

La suite $(\alpha_{n_j})_{j \geq 1}$ est relativement compacte sur la sphère unité de $\mathbb{R}^k$. Soit α_∞ l'un de ses points d'adhérence, l'inégalité précédente et l'hypothèse (H1) entraînent donc que,

$$\forall \delta \in B(0, \epsilon), \langle \alpha_\infty, \langle \phi, \delta \rangle_{E,E} \rangle \leq 0,$$

donc, $\forall \delta \in E, \langle \alpha_\infty, \langle \phi, \delta \rangle_{E,E} \rangle = 0.$

Ce qui contredit l'hypothèse sur le rang de ϕ et achève la démonstration.

Exemple 3.5 Mesures discrètes sur le tore

Soit $E = C(\mathbb{T})$ où $\mathbb{T}$ est le tore $[0, 2\pi[$ et soit $\phi(x) = (1, \cos x, \sin x, \cos kx, \sin kx)$, posons $\mathcal{C} = \{e \in C(\mathbb{T}), \forall x \in \mathbb{T}, e(x) > 0\}$. Pour $c \notin \phi$ strictement réalisable dans $\mathcal{C}$ (c de type strictement positif) il est bien connu que si $n \geq n_0(k)$, il existe une mesure positive ponctuelle

$$dv = \sum_{j=1}^n a_j \delta_{x_j}, \quad \text{avec } a_j \geq 0, \quad j=1, \dots, n \text{ et } \int_0^{2\pi} \phi(x) dv(x) = c.$$

Le lemme précédent assure que pour $\mathcal{C} = \{e \in C(\mathbb{T}), \forall x \in \mathbb{T}, 0 < e(x) < 1\}$, si c est ϕ strictement réalisable dans $\mathcal{C}$, et si n est assez grand $n \geq n_0(k, c)$, il existe une mesure positive ponctuelle :

$$dv = \frac{2\pi}{n} \sum_{j=1}^n a_j \delta_{x_j}, \quad \text{avec } 0 < a_j < 1, \quad j=1, \dots, n \text{ et } \int_0^{2\pi} \phi(x) dv(x) = c.$$

Enonçons maintenant le résultat principal de ce paragraphe:

Corollaire 3.1 *Sous les hypothèses (H1), (H2),*

*(H3) L'enveloppe convexe fermée du support de μ_n est $\bar{\mathcal{C}}_n$
(fermeture pour la topologie usuelle sur $\mathbb{R}^n$).*

(H4) $D(\mu_n, \phi_n \omega_n) = \{t \in \mathbb{R}^k, Z_{\mu_n, \phi_n \omega_n}(t) = \int_{\mathcal{C}_n} \exp \langle t, \phi_n \omega_n \rangle d\mu_n(\omega_n) < \infty\}$

est un ouvert non vide de $\mathbb{R}^k$.

La solution du problème d'entropie sur la moyenne d'ordre n est, pour n assez grand:

$$p_n^{\text{ME}} = [Z_{\mu_n, \phi_n \omega_n}(t_n^*(c))]^{-1} \exp \langle t_n^*(c), \phi_n \omega_n \rangle \mu_n,$$

où $t_n^(c)$ est l'unique minimum de la fonction:*

$$H_{\mu_n, \phi_n \omega_n}(t, c) = \text{Log } Z_{\mu_n, \phi_n \omega_n}(t) - \langle t, c \rangle, \text{ définie sur } D(\mu_n, \phi_n \omega_n).$$

Démonstration

On utilise le corollaire 2.1. Il reste simplement à vérifier qu'il existe pour n assez grand une loi équivalente à μ_n vérifiant la contrainte linéaire. Pour cela utilisons le lemme 3.1, c étant ϕ strictement réalisable dans $\mathcal{C}$, si n est assez grand il existe β_n de $\mathcal{C}_n$ avec $\phi_n \beta_n = c$. Il faut donc voir que pour tout β_n de $\mathcal{C}_n$ il existe une loi P_n équivalente à μ_n qui vérifie:

$$\beta_n = \int_{\mathcal{C}_n} \omega_n dP_n(\omega_n).$$

Cette propriété est vérifiée grâce aux hypothèses (H3) et (H4), il existe en effet ξ_n de $\mathbb{R}^n$, tel que la probabilité (voir la remarque 3.8)

$$\mu_{\omega_n, \xi_n} = [Z_{\mu_n, \omega_n}(\xi_n)]^{-1} \exp(\xi_n, \omega_n) \mu_n,$$

vérifie la contrainte de moyenne précédente (voir le début de la démonstration du lemme 1.4 de [Az-Rug, 77]).

Remarques

Remarque 3.7 Les hypothèses du corollaire 3.1

Les hypothèses (H1), (H2), (H3), (H4) ne sont pas restrictives. (H1), (H2), (H3) sont des hypothèses naturelles sur la construction, en particulier l'hypothèse (H3) sur le support de la mesure μ_n exprime la connaissance à priori. (H4) est toujours vérifiée si $\mathcal{C}_n$ est compact.

Remarque 3.8 Une famille exponentielle sur $\mathcal{C}_n$

Considérons la famille exponentielle:

$$\mu_{\omega_n, \beta} = [Z_{\mu_n, \omega_n}(\beta)]^{-1} \exp(\beta, \omega_n) \mu_n,$$

où β est un élément du convexe :

$$D_{\mu_n, \omega_n} = \left\{ \beta \in \mathbb{R}^n, Z_{\mu_n, \omega_n}(\beta) = \int_{\mathcal{C}_n} \exp(\beta, \omega_n) d\mu_n(\omega_n) < \infty \right\}$$

et $(,)$ désigne ici le produit scalaire sur $\mathbb{R}^n$. La solution du n -ième problème d'entropie sur la moyenne est :

$$P_n^{ME} = \mu_{\omega_n, \phi_n^T t_n^*(c)},$$

où ϕ_n^T désigne la matrice transposée de ϕ_n . En particulier, utilisant les propriétés des familles exponentielles on a:

$$g_n^{ME} = R_n \int_{C_n} \omega_n dP_n^{ME}(\omega_n) = R_n \frac{\partial}{\partial \beta} \text{Log } Z_{\mu_n, \omega_n}(\phi_n^T t_n^*(c)).$$

En fait, on peut montrer que $\mu_{\omega_n, \beta}$ est la probabilité de μ_n - entropie maximale vérifiant la contrainte de moyenne :

$$\int_{C_n} \omega_n dP_n(\omega_n) = \frac{\partial}{\partial \beta} \text{Log } Z_{\mu_n, \omega_n}(\beta).$$

Cette loi joue un rôle particulier dans la construction précédente, nous utiliserons ses propriétés dans les applications suivantes.

III.4 CONVERGENCE DE LA SUITE DES PROBLEMES M.E.M

La convergence de la suite g_n^{ME} vers un élément g_∞^{ME} est très liée à des propriétés de grandes déviations d'une certaine suite de probabilités (nous en avons déjà parlé dans l'exemple 2.9, voir aussi le corollaire 3.5). Pour simplifier, supposons que la suite $(\mu_n)_{n \geq 1}$ est une suite de mesures de probabilité. Posons alors $X_n = \phi_n \omega_n$, le convexe C_n étant muni de la loi μ_n . S'il existe une suite de réels positifs $(a_n)_{n \geq 1}$ de limite $+\infty$ telle que la fonction:

$$\Phi_{\infty, (\phi_n)}^{(\mu_n)}(v) = \lim_{n \rightarrow +\infty} \frac{1}{a_n} \text{Log} \int_{C_n} \exp \{ a_n \langle v, X_n \rangle \} d\mu_n(\omega_n),$$

soit définie et différentiable sur $\mathbb{R}^k$ et que pour chaque n la fonction :

$$\Phi_{n, (\phi_n)}^{(\mu_n)}(v) = \frac{1}{a_n} \text{Log} \int_{C_n} \exp \{ a_n \langle v, X_n \rangle \} d\mu_n(\omega_n),$$

soit définie sur $\mathbb{R}^k$, alors la suite des lois des vecteurs X_n possède une propriété de grandes déviations (théorème II.6.1 de [El, 85], p 47) avec comme fonction d'entropie :

$$I(z) = \sup_{v \in \mathbb{R}^k} \{ \langle v, z \rangle - \Phi_{\infty, (\phi_n)}^{(\mu_n)}(v) \}, z \in \mathbb{R}^k.$$

En particulier, la suite des variables X_n converge en probabilité vers, (voir [El, 85] théorème II.6.3, P 49)

$$\frac{\partial}{\partial v} \Phi_{\infty, (\phi_n)}^{(\mu_n)}(0).$$

Considérons maintenant la suite de variables (X_n) si $\mathcal{C}_n$ est muni de la loi du maximum d'entropie $P_n^{ME} = \mu_{\omega_n, \phi_n^T t_n^*(c)}$. Si l'on fait l'hypothèse que la suite $(v_n^*(c))$ définie par $v_n^*(c) = t_n^*(c) a_n^{-1}$ converge vers $v_\infty^*(c)$, on obtient la convergence en probabilité de la suite (X_n) vers c . En d'autres termes, quand n grandit P_n^{ME} a "tendance" à se concentrer sur les ω_n pour lesquels la contrainte linéaire est vérifiée, c'est-à-dire sur l'ensemble $\{\omega_n \in \mathbb{R}^n, \phi_n \omega_n = c\}$. Reprenant les notations du paragraphe précédent, on a vu que:

$$g_n^{ME} = R_n \frac{\partial}{\partial \beta} \text{Log } Z_{\mu_n, \omega_n}(\phi_n^T t_n^*(c)),$$

et l'on a,

$$\frac{\partial}{\partial v} \Phi_{n, (\phi_n)}^{(\mu_n)} \left(\frac{t_n^*(c)}{a_n} \right) = \phi_n \frac{\partial}{\partial \beta} \text{Log } Z_{\mu_n, \omega_n}(\phi_n^T t_n^*(c)) = c.$$

On peut donc espérer que sous des hypothèses raisonnables sur le quadruplé $((T_n)_{n \geq 1}, (R_n)_{n \geq 1}, (\mu_n)_{n \geq 1}, (\phi_n)_{n \geq 1})$, nous aurons:

$$\frac{\partial}{\partial v} \Phi_{\infty, (\phi_\infty)}^{(\mu_\infty)}(v_\infty^*(c)) = \langle \phi, g_\infty^{ME} \rangle_{E^*, E} = c.$$

Nous étudierons la convergence de la suite des problèmes de maximisation de l'entropie sur la moyenne dans le cas des exemples du paragraphe précédent. Le cas de la bande de fonctions continues est ensuite étudié lorsque l'ensemble de définition de la variable est non compact (U non compact). Enfin, une extension est donnée lorsque $\mathcal{C}$ est un tube d'un espace de variable aléatoire du type $\{L^\infty(\Omega, \mathcal{A}, P)\}^I$ où $(\Omega, \mathcal{A}, P)$ est un espace probabilisé. La convergence de la suite des problèmes MEM a été étudiée dans un cadre plus général dans [Gam-Gas, 89]. Dans la suite de ce chapitre, nous supposons toujours que c est ϕ strictement réalisable dans $\mathcal{C}$.

III.4.1 Boule unité d'un espace de Hilbert séparable

Nous reprendrons ici les mêmes notations que dans le paragraphe précédent. Calculons la fonction Z_{μ_n, ω_n} ,

$$\text{pour } \beta \in \mathbb{R}^n, Z_{\mu_n, \omega_n}(\beta) = \int_{\mathcal{C}_n} \exp(\beta, \omega_n) d\mu_n(\omega_n)$$

$$= \int_{-1}^1 \exp(\|\beta\| u) (1 - u^2)^{(n-1)/2} \gamma_n du,$$

avec $\gamma_n = \frac{v_{n-1}}{v_n}$, où v_n est le volume de la boule unité de $\mathbb{R}^n$

et $\|\cdot\|$ désigne la norme Euclidienne dans $\mathbb{R}^n$.

Il est facile de voir ici que les hypothèses (H1), (H3), (H4) sont vérifiées, la n-ième reconstruction par maximum d'entropie est donc:

$$g_n^{ME} = R_n \frac{\partial}{\partial \beta} \text{Log } Z_{\mu_n, \omega_n}(\phi_n^T t_n^*(c)),$$

où $t_n^*(c)$ minimise:

$$H_{\mu_n, \phi_n \omega_n}(t, c) = \text{Log} \int_{-1}^1 \exp(\|\phi_n^T t\| u) (1 - u^2)^{(n-1)/2} \gamma_n du - \langle t, c \rangle.$$

On a $\frac{\partial}{\partial \beta} \text{Log } Z_{\mu_n, \omega_n}(\beta) = \xi(\beta) \beta$, ($\beta \neq 0$), où l'on a posé:

$$\xi(\beta) = (\|\beta\|)^{-1} (Z_{\mu_n, \omega_n}(\beta))^{-1} \int_{-1}^1 u \exp(\|\beta\| u) (1 - u^2)^{(n-1)/2} \gamma_n du$$

donc
$$g_n^{ME} = \xi(\phi_n^T t_n^*(c)) \sum_{j=1}^n \langle t_n^*(c), \langle e_j, \phi \rangle_{E,E} \rangle e_j.$$

Renormalisons la suite $t_n^*(c)$ en posant $v_n^*(c) = t_n^*(c) \xi(\phi_n^T t_n^*(c))$ et utilisons la contrainte linéaire pour calculer $v_n^*(c)$. On a par construction:

$$\sum_{j=1}^n \langle v_n^*(c), \langle e_j, \phi \rangle_{E,E} \rangle \langle e_j, \phi \rangle_{E,E} = c.$$

C'est-à-dire que $v_n^*(c) = M_n^{-1}(\phi) c$, où $M_n(\phi)$ est la matrice

$$M_n(\phi) = \left(\sum_{l=1}^n \langle e_l, \phi_i \rangle_{E,E} \langle e_l, \phi_j \rangle_{E,E} \right)_{1 \leq i, j \leq n}.$$

On en déduit alors que la suite des problèmes de maximisation de l'entropie sur la moyenne converge dans E , et l'on a

$$g_{\infty}^{ME} = \langle M_{\infty}^{-1}(\phi) c, \phi \rangle$$

où l'on a posé : $M_{\infty}(\phi) = (\langle \phi_i, \phi_j \rangle_{E,E})_{1 \leq i,j \leq k}$. g_{∞}^{ME} est donc l'élément de E qui minimise la norme $\| \cdot \|_E$ et qui satisfait la contrainte linéaire. En particulier, puisque l'on a supposé que c était ϕ strictement réalisable dans $\mathcal{C}$, g_{∞}^{ME} appartient à $\mathcal{C}$.

Remarques

Remarque 3.9 Renormalisation

Soit β une suite de $\mathbb{R}^2$, notons $\beta^{(n)}$ la projection de β sur $\mathbb{R}^n$, c'est-à-dire

$$\beta = (\beta_j)_{j \in \mathbb{N}} \text{ avec } \sum_{j=1}^{\infty} \beta_j^2 < \infty, \text{ et } \beta^{(n)} = (\beta_j)_{j=1, \dots, n}.$$

$$\begin{aligned} \text{On a, } \lim_{n \rightarrow +\infty} \sqrt{n} \xi(\sqrt{n} \beta^{(n)}) &= \lim_{n \rightarrow +\infty} \frac{\int_{-\sqrt{n}}^{\sqrt{n}} v \exp(-\|\beta^{(n)}\| v) \left(1 - \frac{v^2}{n}\right)^{(n-1)/2} dv}{\int_{-\sqrt{n}}^{\sqrt{n}} \exp(-\|\beta^{(n)}\| v) \left(1 - \frac{v^2}{n}\right)^{(n-1)/2} dv} \\ &= \|\beta\|_{12}. \end{aligned}$$

Ce qui prouve que la renormalisation utilisée pour calculer la suite $v_n^*(c)$ est équivalente à $\frac{\sqrt{n}}{\| \langle M_{\infty}^{-1}(\phi) c, \phi \rangle \|_E}$. La suite des lois des variables aléatoires:

$$X_n = \phi_n \omega_n = \sum_{j=1}^n \langle \phi, e_j \rangle_{E,E} \omega_n^j$$

possède donc lorsque la variable aléatoire ω_n a loi μ_n , une propriété de grandes déviations avec $a_n = \sqrt{n}$.

Remarque 3.10 Mesure Gaussienne

Si l'on munit l'espace de Hilbert E d'une mesure Gaussienne μ ([Gih-Skor, 74], p. 348), le problème d'entropie directement résolu dans E conduit à la même reconstruction. Remarquons que cela ne remet pas en cause la méthode M.E.M (sous sa forme suite de problème d'entropie), car le support

de μ étant π tout entier, on ne pouvait pas savoir a priori que la reconstruction obtenue à l'aide de μ appartiendrait à $\mathcal{C}$.

Remarque 3.11 Les mesures "*raisonnables*"

En l'absence d'informations supplémentaires les seules suites de mesures de références "*raisonnables*" (invariantes par le groupe des isométries sur $\mathbb{R}^n$), absolument continues ont une densité du type $f_n(\|\omega_n\|)$ avec f_n une fonction strictement positive définie sur $[0,1]$. Elles conduisent toutes à la même reconstruction,

$$g_\infty^{ME} = \langle M_\infty^{-1}(\phi) c, \phi \rangle.$$

III.4.2 Bande ouverte d'un espace de fonctions continues

Nous étudions dans ce paragraphe la convergence de la suite des problèmes MEM lorsque $E = C(U)$, ensemble des fonctions continues sur U , et que le convexe $\mathcal{C}$ est défini par:

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \}.$$

Lorsque $a(x) = 0$ et $b(x) = +\infty$, la reconstruction "*classique*" par maximum d'entropie (celle qui consiste à choisir parmi toutes les reconstructions possibles celle d'entropie maximale) est très utilisée en pratique pour des problèmes de traitement du signal [Max, 85]. La limitation de la méthode est de ne proposer que des reconstructions positives, il est des cas où la physique du problème donne des informations supplémentaires sur la fonction à reconstruire. D'autre part les axiomes (chapitre II. paragraphe II.2) vérifiés par la méthode du maximum d'entropie, s'ils sont tous justifiés pour la reconstruction d'une probabilité, il n'en est pas de même pour la reconstruction d'une fonction. L'exemple que nous étudions ici est inspiré du travail de J. Navaza [Nav, 85], [Nav, 86] qui porte sur la reconstruction de densités électroniques dont on connaît l'intervalle de variation, et dont on mesure les modules des coefficients de Fourier. Nous étudierons tout d'abord le cas où l'ensemble de définition U de la fonction à reconstruire est compact. Ayant obtenu une reconstruction g_∞^{ME} (limite d'une suite de problèmes MEM), nous nous intéresserons à son extension sur un ensemble de définition U non compact en considérant une suite de compact U_n tendant vers U .

III.4.2.a Cas compact

Nous traiterons tout d'abord le cas le plus simple d'une bande fixe:

$$\forall x \in U, a(x) = a, b(x) = b, -\infty \leq a < b \leq +\infty,$$

avec une suite de mesures de référence du type $\mu_n = F^{\otimes n}$, où F est une mesure positive de support inclus dans $\overline{[a,b]}$, puis du type $\bigotimes_{i=1}^n F_{x_i}$, le cas d'une bande variable sera étudié ensuite.

III.4.2.a.1 Bande fixe

a- Existence de la suite $(x_i)_{i \geq 1}$

On s'intéresse ici à la construction d'une suite d'éléments de U vérifiant:

$$\forall e \in C(U), \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n e(x_i) = \int_U e(x) \frac{dx}{|U|}.$$

On a alors le lemme suivant :

Lemme 3.2 Soient U un compact de $\mathbb{R}^1$ et ν une mesure sigma finie de support U . Alors il existe une suite $(x_j)_{j \geq 1}$ d'éléments de U vérifiant:

$$\forall e \in C(U), \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n e(x_i) = \int_U e(x) \frac{d\nu(x)}{|U|}.$$

Démonstration

Soit l'espace de probabilité $(\Omega, \mathcal{A}, P) = \left\{ U^{\mathbb{N}^*}, \mathcal{B}(U)^{\otimes \mathbb{N}^*}, \left(\frac{d\nu}{\nu(U)} \right)^{\otimes \mathbb{N}^*} \right\}$

pour $\omega \in \Omega$, $\omega = (\omega_j)_{j \geq 1}$, on pose $X_j(\omega) = \omega_j$. La mesure empirique

$$F_n(\omega) = \frac{1}{n} \sum_{j=1}^n \delta_{X_j(\omega)} \text{ converge étroitement vers } \frac{d\nu}{\nu(U)},$$

pour presque tout ω (théorème de Glivenko Cantelli, th 4.4.25 ([Dac-Duf, 82], p. 104). Un élément $(x_j)_{j \geq 1}$ quelconque de l'ensemble de convergence étroite de F_n répond au problème.

b- Etude du problème d'entropie de taille n

Introduisons de nouvelles notations. On pose

$$D(F) = \left\{ \tau \in \mathbb{R}, \int_a^b \exp(\tau y) dF(y) < \infty \right\},$$

$$\text{et pour } \tau \in D(F), \quad \psi(\tau) = \text{Log} \int_a^b \exp(\tau y) dF(y).$$

Lemme 3.3 *On fait les hypothèses suivantes:*

(K1) *L'enveloppe convexe fermée du support de F est $\overline{[a,b]}$.*

(K2) *$D(F)$ est un ouvert non vide de $\mathbb{R}$ et il existe v de $\mathbb{R}^k$, tel que pour tout x de U , $\langle v, \phi(x) \rangle \in D(F)$.*

Alors pour n assez grand le problème de maximisation de l'entropie sur la moyenne possède l'unique solution:

$$\frac{dP_n^{\text{ME}}}{d\mu_n}(\omega_n) = \prod_{i=1}^n \exp \{ \langle v_n^*(c), \phi(x_i) \rangle \omega_n^i - \psi(\langle v_n^*(c), \phi(x_i) \rangle) \}$$

$$\text{et } g_n^{\text{ME}}(x) = \frac{1}{\#J_n(x)} \sum_{j \in J_n(x)} \psi(\langle v_n^*(c), \phi(x_j) \rangle),$$

où $v_n^*(c)$ est le minimum de la fonction:

$$H_{\mu_n, \phi_n, \omega_n}^{\text{ren}}(v, c) = \frac{|U|}{n} \sum_{i=1}^n \psi(\langle v, \phi(x_i) \rangle) - \langle v, c \rangle.$$

Démonstration

Utilisons le corollaire 3.1. Par construction de la suite ϕ_n , elle vérifie l'hypothèse (H1). L'hypothèse (K1) (resp. (K2)) entraîne (H3) (resp. (H4)). Donc, l'unique solution du n -ième problème MEM est pour n assez grand:

$$P_n^{\text{ME}} = [Z_{\mu_n, \phi_n, \omega_n}(t_n^*(c))]^{-1} \exp \langle t_n^*(c), \phi_n \omega_n \rangle \mu_n,$$

où $t_n^*(c)$ est l'unique minimum de

$$H_{\mu_n, \phi_n, \omega_n}(t, c) = \text{Log } Z_{\mu_n, \phi_n, \omega_n}(t) - \langle t, c \rangle,$$

mais ici, on a,
$$\text{Log } Z_{\mu_n, \phi_n, \omega_n}(t) = \sum_{i=1}^n \psi\left(\frac{|U|}{n} \langle t, \phi(x_i) \rangle\right).$$

Renormalisons le paramètre t et la fonction $H_{\mu_n, \phi_n, \omega_n}$ en posant $v = \frac{|U|}{n} t$.

D'où,
$$H_{\mu_n, \phi_n, \omega_n}^{\text{ren}}(v, c) = \frac{|U|}{n} H_{\mu_n, \phi_n, \omega_n}(t, c) = \frac{|U|}{n} \sum_{i=1}^n \psi(\langle v, \phi(x_i) \rangle) - \langle v, c \rangle.$$

c- Convergence de la suite des problèmes M.E.M

On a le résultat principal suivant:

Théorème 3.1 *Sous les hypothèses (K1), (K2) et,*

$$(K3) \ D(F, \phi) = \left\{ v \in \mathbb{R}^k, \int_U \psi(\langle v, \phi(x) \rangle) dx < \infty \right\} \text{ est non vide et}$$

$$\text{coïncide avec } D'(F, \phi) = \left\{ v \in \mathbb{R}^k, \forall x \in U, \langle v, \phi(x) \rangle \in D(F) \right\}.$$

La suite des problèmes M.E.M converge simplement. On a :

$$g_{\infty}^{\text{ME}}(x) = \psi'(\langle v_{\infty}^*(c), \phi(x) \rangle)$$

et $v_{\infty}^*(c)$ minimise,
$$H_{F, \phi}(v, c) = \int_U \psi(\langle v, \phi(x) \rangle) dx - \langle v, c \rangle.$$

Démonstration

Remarquons tout d'abord que la fonction $H_{F, \phi}(\cdot, c)$ est semi-continue inférieurement (lemme de Fatou), et que $D(F, \phi)$ est un ouvert non vide. Posons

$$\mathcal{K} = \left\{ \int_U \phi(x) e(x) dx, e \in \mathcal{C} \right\}.$$

$\mathcal{K}$ est convexe ouvert car ϕ est de rang plein (théorème de l'image ouverte).

Pour $y \in]a, b[$, on pose :

$$\gamma(y) = -y \psi^{-1}(y) + \psi(\psi^{-1}(y)) = \inf_{\tau \in \mathbb{R}} \{\psi(\tau) - \tau y\}.$$

Lorsque F est une probabilité, $-y$ est sa transformée de Cramer. Soient c un élément de $\mathcal{K}$ et g de $\mathcal{C}$ tels que:

$$\int_U \phi(x) g(x) dx = c.$$

Posons :

$$c^{(n)} = \frac{|U|}{n} \sum_{j=1}^n \phi(x_j) g(x_j).$$

On a alors l'égalité, (voir le théorème 2.3):

$$H_{\mu_n, \phi_n \omega_n}^{\text{ren}}(v_n^*(c^{(n)}), c^{(n)}) = \frac{|U|}{n} S_{\mu_n}(P_n^{\text{ME}}),$$

où P_n^{ME} est la loi du maximum de μ_n - entropie lorsque la contrainte linéaire

est :

$$\phi_n \int_{\mathcal{C}_n} \omega_n dP_n(\omega_n) = c^{(n)}.$$

Soit μ_{ω_n, β_n} la loi sur $\mathcal{C}_n$ définie par (remarque 3.8):

$$\begin{aligned} \mu_{\omega_n, \beta_n} &= [Z_{\mu_n, \omega_n}(\beta_n)]^{-1} \exp(\beta_n, \omega_n) \mu_n, \\ &= \prod_{i=1}^n \exp\{\psi^{-1}\{g(x_i)\} \omega_n^i - \psi(\psi^{-1}\{g(x_i)\})\} \mu_n, \end{aligned}$$

où $\beta_n = (\psi^{-1}\{g(x_1)\}, \dots, \psi^{-1}\{g(x_n)\})$. μ_{ω_n, β_n} vérifie

$$\frac{|U|}{n} \sum_{j=1}^n \int_{\mathcal{C}_n} \omega_n^j \phi(x_j) d\mu_{\omega_n, \beta_n}(\omega_n) = c^{(n)}$$

et,

$$\frac{|U|}{n} S_{\mu_n}(\mu_{\omega_n, \beta_n}) = \frac{|U|}{n} \sum_{j=1}^n \gamma(g(x_j)).$$

Puisque par définition P_n^{ME} est la loi du maximum d'entropie on a l'inégalité

$$H_{\mu_n, \phi_n \omega_n}^{\text{ren}}(v, c^{(n)}) \geq H_{\mu_n, \phi_n \omega_n}^{\text{ren}}(v_n^*(c^{(n)}), c^{(n)}) \geq \frac{|U|}{n} \sum_{j=1}^n \gamma(g(x_j)).$$

La fonction $\gamma(g(\cdot))$ est continue sur U (car, $\forall x \in U, a < g(x) < b$). Par passage à la limite, on obtient donc

$$H_{F,\phi}(v, c) \geq \int_U \psi(g(x)) \, dx.$$

Soit v' un point de $D(F, \phi)$ posons

$$c' = \int_U \phi(x) \psi'(\langle v', \phi(x) \rangle) \, dx$$

$\psi'(\langle v, \phi(x) \rangle)$ est uniformément majorée par une constante si v appartient à un voisinage de v' d'adhérence compacte et incluse dans $D(F, \phi)$. En effet, soit B un tel voisinage, on a pour $v \in B$,

$$\begin{aligned} |\psi'(\langle v, \phi(x) \rangle)| &\leq \sup_{x \in U} |\psi'(\langle v, \phi(x) \rangle)| \\ &= |\psi'(\sup_{x \in U} \langle v, \phi(x) \rangle)| \leq |\psi'(\sup_{v \in B} \{ \sup_{x \in U} \langle v, \phi(x) \rangle \})| < \infty. \end{aligned}$$

Sinon, par compacité, il existe v de $\bar{B}$ et x de U avec $|\psi'(\langle v, \phi(x) \rangle)| = +\infty$, et donc $\langle v, \phi(x) \rangle \notin D(F)$ ce qui contredit l'hypothèse (K3). On peut donc permuter intégration et dérivation en appliquant le théorème de Lebesgue, et donc v' minimise la fonction $H_{F,\phi}(\cdot, c')$. Compte-tenu de l'inégalité démontrée plus haut (inégalité encadrée), en appliquant le lemme 3.4 (voir plus bas) et en utilisant le même argument, on en déduit que $H_{F,\phi}(\cdot, c)$ est minimum en $v_\infty^*(c)$ avec

$$\int_U \phi(x) \psi'(\langle v_\infty^*(c), \phi(x) \rangle) \, dx = c.$$

Pour terminer la démonstration, montrons que g_n^{ME} converge pour tout x de U vers g_∞^{ME} . Soit x de U , on a

$$\begin{aligned} |g_n^{ME}(x) - g_\infty^{ME}(x)| &\leq \sup_{j \in J_n(x)} |\psi'(\langle v_n^*(c), \phi(x_j) \rangle) - \psi'(\langle v_\infty^*(c), \phi(x) \rangle)| \\ &\leq \sup_{j \in J_n(x)} |\psi'(\langle v_n^*(c), \phi(x_j) \rangle) - \psi'(\langle v_\infty^*(c), \phi(x_j) \rangle)| \\ &\quad + \sup_{j \in J_n(x)} |\psi'(\langle v_\infty^*(c), \phi(x_j) \rangle) - \psi'(\langle v_\infty^*(c), \phi(x) \rangle)|. \end{aligned}$$

Puisque la suite $(x_j)_{j \geq 1}$ est dense dans U et que les fonctions ψ' et ϕ sont continues, le second terme tend vers 0. Le premier terme, quant à lui, est majoré par:

$$\sup_{x \in U} | \psi'(\langle v_n^*(c), \phi(x) \rangle) - \psi'(\langle v_\infty^*(c), \phi(x) \rangle) |.$$

D'autre part, la suite de fonctions convexes,

$$\{ H_{\mu_n, \phi_n, \omega_n}^{\text{ren}}(\cdot, c) \}_{n \geq 1}$$

converge uniformément sur tout compact inclus dans $D(F, \phi)$ vers $H_{F, \phi}(\cdot, c)$ (théorème VI.3.3 de [El, 85], p 214), ce qui entraîne que la suite de points $\{ v_n^*(c) \}_{n \geq n_0}$ converge vers $v_\infty^*(c)$. Et donc la suite de fonctions $\{ \langle v_n^*(c), \phi(\cdot) \rangle \}_{n \geq 1}$ converge uniformément vers $\langle v_\infty^*(c), \phi(\cdot) \rangle$. Finalement puisque la fonction ψ' est uniformément continue sur tout compact inclus dans $D(F)$ on obtient

$$\lim_{n \rightarrow +\infty} \left\{ \sup_{x \in U} | \psi'(\langle v_n^*(c), \phi(x) \rangle) - \psi'(\langle v_\infty^*(c), \phi(x) \rangle) | \right\} = 0.$$

Pour la démonstration précédente, nous avons eu besoin du lemme suivant:

Lemme 3.4 *Soit $K(v, c)$ une fonction strictement convexe, concave définie sur $V \times W$, V (resp. W) est un convexe ouvert de $\mathbb{R}^k$ (resp. de $\mathbb{R}^m$). On suppose que:*

- a- Pour c de W fixé, $K(\cdot, c)$ est semi-continue inférieurement (s.c.i) sur $\mathbb{R}^k$, si l'on convient que $\forall v \notin V, K(v, c) = +\infty$.*
- b- Il existe c' de W tel que $K(\cdot, c')$ soit minimum en un point v' de V .*
- c- Il existe une fonction L à valeurs finies sur W telle que:*

$$\forall v \in V, K(v, c) \geq L(c).$$

Alors, pour tout c de W , il existe $v^(c)$ de V qui vérifie,*

$$K(v^*(c), c) = \inf_{v \in V} K(v, c).$$

Démonstration

Montrons tout d'abord qu'une fonction f strictement convexe, s.c.i sur $\mathbb{R}^k$, définie sur un ouvert non vide V , admet un minimum si et seulement si:

$$\lim_{\|v\| \rightarrow +\infty} f(v) = +\infty,$$

avec la convention que f vaut $+\infty$ quand elle n'est pas définie. Il est facile de voir que si $\lim_{\|v\| \rightarrow +\infty} f(v) = +\infty$, alors f possède un minimum (corollaire III.20, [Brez, 83], p. 46).

Reciproquement, supposons que f admette un minimum en v^* , par translation on peut toujours supposer que $v^* = 0$. Raisonnons par l'absurde et supposons qu'il existe une suite $(v_n)_{n \geq 1}$ de $\mathbb{R}^k$ vérifiant:

$$\begin{aligned} \lim_{n \rightarrow +\infty} \|v_n\| &= +\infty, \\ \lim_{n \rightarrow +\infty} \frac{v_n}{\|v_n\|} &= \lim_{n \rightarrow +\infty} \alpha_n = \alpha, \text{ avec } \|\alpha\| = 1 \end{aligned}$$

$$\text{et} \quad \lim_{n \rightarrow +\infty} f(v_n) = \xi.$$

Soit t un réel positif tel que $\frac{t}{\|v_n\|} < 1$, on a

$$f(t \alpha_n) = f\left(\frac{t}{\|v_n\|} v_n\right) \leq \frac{t}{\|v_n\|} f(v_n) + \left(1 - \frac{t}{\|v_n\|}\right) f(0).$$

En prenant la limite de cette inégalité on obtient, $\lim_{n \rightarrow +\infty} f(t \alpha_n) \leq f(0)$ et la semi-continuité de f implique, $\lim_{n \rightarrow +\infty} f(t \alpha_n) \geq f(t \alpha)$. Utilisant ces deux inéquations on en déduit que $f(t \alpha) \leq f(0)$, ce qui contredit que 0 est l'unique minimum de f .

Démontrons maintenant le lemme. Soit c de W , il existe c^{**} de W et un réel λ de $]0,1[$ tels que : $c = \lambda c' + (1-\lambda) c^{**}$, (c' a été défini dans l'énoncé du lemme 3.4). La concavité de la fonction $K(v, \cdot)$ donne,

$$K(v, c) \geq \lambda K(v, c') + (1-\lambda) K(v, c^{**}) \geq \lambda K(v, c') + (1-\lambda) L(c^{**}),$$

on conclut à l'aide la démonstration préliminaire.

Exemples

Exemple 3.6 Loi normale

Soient F la loi normale centrée réduite et $\mathcal{C} = C(U)$, on a,

$$\psi(\tau) = \frac{\tau^2}{2} \quad (\tau \in \mathbb{R}) \quad \text{et donc,} \quad g_{\infty}^{ME}(x) = \psi'(\langle v_{\infty}^*(c), \phi(x) \rangle) = \langle v_{\infty}^*(c), \phi(x) \rangle.$$

On retrouve donc encore la reconstruction de norme L^2 minimum.

Exemple 3.7 Loi de Poisson

Soient F la loi de poisson d'intensité λ et $\mathcal{C} = \{e \in C(U), \forall x \in U, e(x) > 0\}$.

On a, $\psi(\tau) = \lambda (e^\tau - 1) \quad (\tau \in \mathbb{R})$, d'où

$$g_\infty^{\text{ME}}(x) = \psi'(\langle v_\infty^*(c), \phi(x) \rangle) = \lambda \exp \langle v_\infty^*(c), \phi(x) \rangle.$$

En particulier, lorsque $\lambda = 1$ ou que $\phi_1 = 1$, on retrouve la reconstruction de dx - entropie maximale.

Remarque 3.12 Recollement du problème M.E.M

Dans les deux exemples précédents, la suite des problèmes M.E.M n'est en fait, rien d'autre que la discrétisation des problèmes de maximisation de l'entropie du paragraphe III.2.1.

Exemple 3.8 Mesure de Bernouilli

Soient $F = \delta_0 + \delta_1$ et $\mathcal{C} = \{e \in C(U), \forall x \in U, 0 < e(x) < 1\}$, remarquons que F est (à une constante multiplicative près) la mesure positive de plus petit support vérifiant l'hypothèse (K1). On a,

$$\psi(\tau) = \text{Log}(e^\tau + 1) \quad (\tau \in \mathbb{R}), \text{ et}$$

$$g_\infty^{\text{ME}}(x) = \psi'(\langle v_\infty^*(c), \phi(x) \rangle) = \frac{1}{1 + \exp - \langle v_\infty^*(c), \phi(x) \rangle}.$$

Exemple 3.9 Loi uniforme

En l'absence d'informations supplémentaires, il est raisonnable d'utiliser sur le convexe,

$$\mathcal{C} = \{e \in C(U), \forall x \in U, 0 < e(x) < 1\}$$

la loi uniforme comme mesure de référence ([Nav, 85]). On a alors,

$$\psi(\tau) = \text{Log} \frac{e^\tau - 1}{\tau} \quad (\tau \in \mathbb{R}),$$

$$g_\infty^{\text{ME}}(x) = \frac{\exp \langle v_\infty^*(c), \phi(x) \rangle}{\exp \langle v_\infty^*(c), \phi(x) \rangle - 1} - \frac{1}{\langle v_\infty^*(c), \phi(x) \rangle}$$

Remarque 3.13 Les hypothèses du théorème 3.1

Les hypothèses (K1) et (K2) sont assez naturelles et ne sont pas restrictives. L'hypothèse (K3) est toujours vérifiée si les bornes a et b sont finies (car dans ce cas $D(F) = \mathbb{R}$). Donnons un exemple où l'hypothèse (K3) n'est pas vérifiée. F est la mesure de Lebesgue sur $\mathbb{R}^+$, $\mathcal{C}$ est défini par:

$$\mathcal{C} = \{e \in C(U), \forall x \in U, e(x) > 0\}$$

et $\phi(x)$ est de la forme,

$$\phi(x) = (1, \cos(m_1 \cdot x), \sin(m_1 \cdot x), \dots, \cos(m_k \cdot x), \sin(m_k \cdot x)),$$

où $\cdot$ désigne ici le produit scalaire sur $\mathbb{R}^1$ et m_j ($j=1..k$), est un multi-indice de $\mathbb{N}^1$. Pour simplifier les notations dans la suite lorsque nous utiliserons ce système de fonctions nous parlerons de système d'exponentielles complexes et nous le noterons de façon plus condensé $\phi(x) = (\exp(i m_j \cdot x))_{j=0..k}$, ($m_0=(0,..0)$). Le multiplicateur de Lagrange v sera donc un élément de $\mathbb{R} \times \mathbb{C}^k$, c'est-à-dire que

$$\langle v, \phi(x) \rangle = \sum_{j=1}^k (v_j e^{im_j \cdot x} + \bar{v}_j e^{-im_j \cdot x}) + v_0.$$

La transformée de Laplace F est ici $\exp\{\psi(\tau)\} = -\tau^{-1}$. Elle est définie pour τ strictement négatif. On a donc:

$$D(F, \phi) = \{v \in \mathbb{R} \times \mathbb{C}^k, \forall x \in U, \langle v, \phi(x) \rangle \leq 0\}$$

$$\neq \{v \in \mathbb{R} \times \mathbb{C}^k, \forall x \in U, \langle v, \phi(x) \rangle < 0\} = D^-(F, \phi).$$

Le théorème 3.1 n'est donc pas applicable, pourtant il est bien connu pour $U = [0, 2\pi]$ ([Land, 87]), qu'il existe pour le problème des moments trigonométriques une reconstruction du type, $-\{\langle v_\infty^*(c), \phi(x) \rangle\}^{-1}$, c'est celle qui maximise l'entropie de Burg. Nous reviendrons sur cet exemple un peu plus loin (exemple 3.10).

Le corollaire suivant donne la convergence des problèmes M.E.M, sans supposer (K3) mais en supposant une certaine régularité de l'application:

$$D(F, \phi) : \longrightarrow \mathbb{R}^k$$

$$v \longrightarrow \int_U \phi(x) \psi'(\langle v, \phi(x) \rangle) dx$$

Corollaire 3.2 *On fait les hypothèses (K1), (K2),*

(K'3) $D^-(F, \phi) = \{v \in \mathbb{R}^k, \forall x \in U, \langle v, \phi(x) \rangle \in D(F)\}$ est non vide.

(K4) Pour tout v' de $\overline{D^-(F, \phi)}/D^-(F, \phi)$ on a,

$$\lim_{v \in D^-(F, \phi)} \left\| \int_U \phi(x) \psi'(\langle v, \phi(x) \rangle) dx \right\| = +\infty.$$

Alors, on a les mêmes conclusions qu'au théorème 3.1.

Démonstration

Soit c de $\mathcal{K}$, d'après la démonstration du théorème 3.1, pour tout c' dans un voisinage de c inclus dans $\mathcal{K}$ on a,

$$\inf_{v \in D'(F, \phi)} H_{F, \phi}(v, c) > -\infty.$$

Le théorème VI.5.7 de [El, 85], p 224 (propriétés de la dualité de Legendre) permet d'affirmer, compte-tenu de l'hypothèse (K4), que l'intérieur de l'ensemble:

$$\{c \in \mathbb{R}^k, \inf_{v \in D'(F, \phi)} H_{F, \phi}(v, c) > -\infty\}$$

est inclus dans l'ensemble:

$$\left\{ \int_U \phi(x) \psi'(\langle v, \phi(x) \rangle) dx, v \in D'(F, \phi) \right\}, \text{ d'où le résultat.}$$

Exemple 3.10 Entropie de Burg

Reprenons la remarque 3.13, dans cet exemple on a, $\mathcal{C} = \{e \in C(U), \forall x \in U, e(x) > 0\}$, U est un compact de $\mathbb{R}$ ou $\mathbb{R}^2$, $\phi(\cdot)$ est un système d'exponentielles complexes et F est la mesure de Lebesgue sur $\mathbb{R}^+$. Si v est tel que le polynôme trigonométrique $\langle v, \phi(x) \rangle$ soit négatif et possède des racines, on a,

$$\int_U - \frac{dx}{\langle v, \phi(x) \rangle} = \infty.$$

Et donc le lemme de Fatou donne l'hypothèse (K4). Ceci généralise donc à U compact de $\mathbb{R}^2$ quelconque la reconstruction par maximum d'entropie de Burg [Burg, 75], [Far, 84]. Remarquons que pour U inclus dans $\mathbb{R}^1$ avec $1 \geq 3$, il existe des polynômes trigonométriques négatifs, s'annulant et d'inverse intégrable. Le corollaire précédent n'est donc pas applicable. L'existence d'une extension factorisable du type, $-\{\langle v_\infty^*(c), \phi(x) \rangle\}^{-1}$ est étudiée pour le cas $\mathbb{R}^1$ ($1 \in \mathbb{N}^*$) dans [Seg, 87].

Généralisons maintenant le théorème 3.1 dans le cas où la mesure de référence μ_n vaut $\bigotimes_{i=1}^n F_{x_i}$. Pour cela, étendons tout d'abord les notations du paragraphe précédent. On pose,

$$D(F_x) = \left\{ \tau \in \mathbb{R}, \int_a^b \exp(\tau y) dF_x(y) < \infty \right\},$$

et pour $x \in U, \tau \in D(F_x)$,

$$\psi(\tau, x) = \text{Log} \int_a^b \exp(\tau y) dF_x(y), \quad \psi'(\tau, x) = \frac{\partial}{\partial \tau} \psi(\tau, x).$$

Enfin pour y appartenant à l'intérieur de l'enveloppe convexe du support de F_x on pose: $\psi'^{-1}(y, x) = \tau$, si $\psi'(\tau, x) = y$.

Corollaire 3.3 *Sous les hypothèses*

(K1) *Pour tout x de U , l'enveloppe convexe fermée du support de F_x est $\overline{[a, b]}$.*

(K'2) $\forall x \in U, D(F_x) = \mathbb{R}$.

(K''3) ψ, ψ' sont continues sur $\mathbb{R} \times U$ et ψ'^{-1} est continue sur $[a, b] \times U$.

On a les mêmes conclusions qu'au théorème 3.1 en remplaçant partout $\psi'(\cdot)$ par $\psi'(\cdot, x)$.

Démonstration

La démonstration est identique à celle du théorème 3.1 en remplaçant $\gamma(y)$ par $\gamma(y, x) = -y \psi'^{-1}(y, x) + \psi\{\psi'^{-1}(y, x), x\}$.

Exemples

Donnons-nous une fonction f continue sur U (on suppose que f est strictement positive sur U) et reprenons alors les exemples du début du paragraphe.

Exemple 3.11 Loi Normale (suite de l'exemple 3.6)

$\mathcal{C} = C(U)$ et F_x la normale centrée de variance $f(x)$. On a, pour τ réel,

$$\psi(\tau, x) = \frac{f(x)}{2} \tau^2 \quad \text{et donc,} \quad g_{\infty}^{ME}(x) = \psi'(\langle v_{\infty}^*(c), \phi(x) \rangle, x) = \langle v_{\infty}^*(c), \phi(x) \rangle f(x).$$

g_{∞}^{ME} est la reconstruction de norme minimale dans l'espace $L^2(U, \{f(x)\}^{-1}dx)$.
Changer la mesure de référence en changeant la variance de la loi normale, revient donc à changer directement la mesure sur U .

Exemple 3.12 Loi de Poisson (suite de l'exemple 3.7)

$\mathcal{C} = \{e \in C(U), \forall x \in U, e(x) > 0\}$ et F_x la loi de Poisson d'intensité $f(x)$.

Pour τ de $\mathbb{R}$, $\psi(\tau, x) = f(x)(e^{\tau} - 1)$, d'où $g_{\infty}^{ME}(x) = f(x) \exp \langle v_{\infty}^*(c), \phi(x) \rangle$.

Donc g_{∞}^{ME} minimise l'information de Kullback par rapport à $f(x)$ et l'on peut faire la même remarque que dans l'exemple 3.11.

III.4.2.3.2 Bande variable

Intéressons-nous maintenant au problème de reconstruction précédent lorsque les bornes $a(x)$ et $b(x)$ sont des fonctions continues sur U . On a donc,

$$\mathcal{C} = \{e \in C(U), \forall x \in U, a(x) < e(x) < b(x)\}.$$

La mesure de référence sur le convexe de dimension finie

$$\mathcal{C}_n = \prod_{i=1}^n [a(x_i), b(x_i)],$$

est toujours une mesure produit du type $\mu_n = \bigotimes_{i=1}^n F_{x_i}$. Reprenant alors les notations du paragraphe précédent on a

Corollaire 3.4 *Sous les hypothèses (K'3),*

(K'1) *Pour tout x de U , l'enveloppe convexe fermée du support de F_x est $[a(x), b(x)]$.*

(K'2) *Pour tout x de U , $\int_{[a(x), b(x)]} dF_x < +\infty$.*

On a les mêmes conclusions qu'au théorème 3.1 en remplaçant partout $\psi(\cdot)$ par $\psi'(\cdot, x)$.

Démonstration

L'hypothèse (K''2) assure que pour tout x de U la transformée de Laplace de la mesure F_x est définie sur $\mathbb{R}$. On procède ensuite de la même manière que pour la démonstration du corollaire 3.3.

Exemple 3.13 Mesure de Bernouilli (suite de l'exemple 3.8)

Choisissons comme famille de mesures de référence,

$$F_x = \delta_{a(x)} + \delta_{b(x)}, \quad (x \in U).$$

F_x est (à une constante multiplicative près) la mesure de support minimal vérifiant l'hypothèse (K'1). On a, pour τ de $\mathbb{R}$ et x de U ,

$$\psi(\tau, x) = \text{Log} (e^{a(x)\tau} + e^{b(x)\tau}), \quad \text{donc}$$

$$g_{\infty}^{\text{ME}}(x) = \frac{a(x) \exp\{a(x) \langle v_{\infty}^*(c), \phi(x) \rangle\} + b(x) \exp\{b(x) \langle v_{\infty}^*(c), \phi(x) \rangle\}}{\exp\{a(x) \langle v_{\infty}^*(c), \phi(x) \rangle\} + \exp\{b(x) \langle v_{\infty}^*(c), \phi(x) \rangle\}}.$$

Dans le cadre du problème M.E.M précédent (bande variable), intéressons nous brièvement aux propriétés de grandes déviations de la suite des lois des variables aléatoires $X_n = \phi_n \omega_n$, (voir l'exemple 2.9 et le début du paragraphe III.4), où ω_n est l'élément générique de

$$C_n = \prod_{i=1}^n [a(x_i), b(x_i)],$$

lorsque celui-ci est muni de la probabilité P_n^{ME} .

Nous allons voir que dans le cas unidimensionnel, la convergence en probabilité de (X_n) est un cas particulier de la convergence en loi d'une suite de processus construite à partir de la suite (ω_n) , on a:

corollaire 3.5

a- Sous les hypothèses du corollaire 3.4, la suite des lois des variables aléatoires X_n possède, quand la probabilité sur $\mathcal{C}_n$ est P_n^{ME} , une propriété de grandes déviations avec $a_n = n$ et,

$$I(z) = \sup_{v \in \mathbb{R}^k} \{ \langle v, z \rangle - \int_U \{ \psi(\langle v + v_\infty^*(c), \phi(x) \rangle, x) - \psi(0, x) \} dx \},$$

comme fonction d'entropie. En particulier, $(X_n)_{n \geq 1}$ converge en probabilité vers c .

b- Lorsque $U = [\alpha, \beta]$, on pose $X_n(x(j)) = \frac{|U|}{n} \sum_{i=1}^j \phi(x(i)) \omega_n^i$, $j=1, \dots, n$,

puis, pour x de $[x(1), x(n)]$, $x(j-1) < x < x(j)$, $x = \lambda x(j-1) + (1-\lambda) x(j)$, $\lambda \in]0, 1[$

$$X_n(x) = \lambda X_n(x(j-1)) + (1-\lambda) X_n(x(j)),$$

enfin, $X_n(x) = X_n(x(n))$ si $x > x(n)$ et $X_n(x) = X_n(x(1))$ si $x < x(1)$.

Où $(x(j))_{j=1, \dots, n}$ désigne la suite $(x_j)_{j=1, \dots, n}$ réordonnée. Alors la suite de processus $(X_n(\cdot))$ converge étroitement quand la probabilité sur $\mathcal{C}_n$ est P_n^{ME} vers la fonction

$$\int_{\alpha}^x \phi(t) \psi'(\langle v_\infty^*(c), \phi(t) \rangle, t) dt.$$

DémonstrationPoint a

D'après le corollaire 3.4, la suite $(v_n^*(c))$ converge vers $v_\infty^*(c)$, avec de plus

$$\int_U \phi(x) \psi'(\langle v_\infty^*(c), \phi(x) \rangle, x) dx = c,$$

le reste de la démonstration a déjà été donné dans l'exemple 2.9, en remplaçant partout $\frac{1}{n}$ par $\frac{|U|}{n}$, $\frac{j}{n}$ par x_j et $\psi'(\cdot)$ par $\psi'(\cdot, x)$.

Point b

C'est une application directe du théorème 7.3.22 (p. 217) et bilan (p.216) de [Dac-Duf, 83].

Exemple 3.14 Mesure de Bernouilli (suite des exemples 3.8 et 3.13)

Dans le cas le plus simple où la famille de mesures de référence est donnée par $F_x = \delta_{a(x)} + \delta_{b(x)}$, ($x \in U$), on trouve que

$$I(z) = \sup_{v \in \mathbb{R}^k} \{ \langle v, z \rangle + |U| \log 2 - \int_U \log [\exp\{a(x) \langle v + v_\infty^*(c), \phi(x) \rangle\} + \exp\{b(x) \langle v + v_\infty^*(c), \phi(x) \rangle\}] dx \}.$$

$I(z)$ possède la propriété suivante (voir chapitre IV)

$$0 \leq I(z) \leq |U| \log 2 \iff z \in \bar{\mathcal{K}},$$

$$I(z) = +\infty \iff z \notin \bar{\mathcal{K}}.$$

Où $\mathcal{K}$ est rappelons le défini par $\mathcal{K} = \{ \int_U \phi(x) e(x) dx, e \in \mathcal{C} \}$. La propriété

précédente permet donc de caractériser le convexe $\mathcal{K}$. Remarquons enfin que la propriété précédente est tout simplement l'analogie de la propriété de la fonctionnelle de grandes déviations d'une loi binomiale renormalisée: soit $X_1, \dots, X_n$ un n -échantillon de loi de Bernouilli de paramètre p , posons

$$S_n = \sum_{i=1}^n X_i.$$

La fonctionnelle de grandes déviations de la loi de $\frac{S_n}{n}$ est (voir l'exemple 2.7)

$$I(z) = \sup_{v \in \mathbb{R}} (vz - \log(1 - p + pv)) = z \log \frac{z}{p} + (1 - z) \log \frac{1 - z}{1 - p}, \quad (z \in [0, 1])$$

$$= +\infty, \quad (z \notin [0, 1]).$$

III.4.2.b Extension au cas non compact

Nous allons étendre la construction précédente lorsque:

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \},$$

(a et b sont des fonctions continues sur U) et que U n'est pas compact. Pour cela, on va considérer la suite de compacts $U_n = U \cap \{ x \in \mathbb{R}^1, \|x\| \leq n \}$. Sur U_n , on résout la suite des problèmes M.E.M, puis on fera ensuite tendre n vers l'infini. Énonçons tout d'abord un lemme qui nous servira par la suite.

Lemme 3.5 sous les hypothèses (K1), (K3),

(K2) Pour tout x de U , F_x est une probabilité

(K4) $D(F, \phi) = \{v \in \mathbb{R}^k, \int_U \psi(\langle v, \phi(x) \rangle) dx < \infty\}$ est un ouvert non vide.

(K5) $a(x) \sup_{j=1, \dots, k} |\phi_j(x)|$ et $b(x) \sup_{j=1, \dots, k} |\phi_j(x)|$ sont intégrables sur U .

Pour tout c de $\mathcal{K}$, il existe une fonction g appartenant à $\mathcal{C}$ qui vérifie

$$\int_U \phi(x) g(x) dx = c \quad \text{et} \quad \int_U |\gamma \{g(x), x\}| dx < \infty.$$

Démonstration

Montrons tout d'abord que $\mathcal{K}$ est un ouvert de $\mathbb{R}^k$. Soit $C(U_n, U)$ l'ensemble des fonctions continues et bornées sur U qui sont nulles en dehors de U_n . On munit $C(U_n, U)$ de la norme

$$\|e\|_{C(U_n, U)} = \sup_{x \in U} |e(x)|, \quad (e \in C(U_n, U)).$$

$C(U_n, U)$ ainsi normé est un espace de Banach. Si n est assez grand l'application

$$C(U_n, U): \longrightarrow \mathbb{R}^k$$

$$e \longrightarrow \int_U \phi(x) e(x) dx,$$

est de rang plein. Soit g de $\mathcal{C}$, posons

$$\mathcal{C}_g^n = \{e \in C(U_n, U), \forall x \in U_n, a(x) < g(x) + e(x) < b(x)\}$$

$$\text{et} \quad \mathcal{K}_g^n = \left\{ \int_U \phi(x) (e(x) + g(x)) dx, e \in \mathcal{C}_g^n \right\}.$$

$\mathcal{K}_g^n$ est, pour n assez grand, un ouvert (théorème de l'image ouverte appliqué à $C(U_n, U)$, ϕ étant de rang plein), on conclut en remarquant que $\mathcal{K} = \bigcup_{g \in \mathcal{C}} \mathcal{K}_g^n$.

Posons maintenant pour x de U ,

$$h(x) = \int_{[a(x), b(x)]} y dF_x(y) = \psi(o, x).$$

$h(x)$ est l'espérance de la loi F_x portée par l'intervalle $[a(x), b(x)]$. h est continue par hypothèse et h appartient à $\mathcal{C}$. Un simple calcul donne d'autre part que,

pour tout x de U $\gamma\{h(x), x\}$ est nul. Soit $(g_n)_{n \geq 1}$ une suite de fonctions définie par:

$$g_n(x) = g(x) \quad \text{si } x \in U_n$$

$$g_n(x) = h(x) \quad \text{si } x \in U/U_{n+1}$$

$$g_n(x) = k_n(x) \quad \text{si } x \in U_{n+1}/U_n.$$

où la fonction k_n est telle que g_n soit un élément de $\mathcal{C}$. On a alors

$$\lim_{n \rightarrow +\infty} \int_U \phi(x) g_n(x) dx = \int_U \phi(x) g(x) dx.$$

Posons $\mathcal{K}_n = \left\{ \int_U \phi(x) g_n(x) dx, g \in \mathcal{C} \right\}$.

$\mathcal{K}_n$ est convexe, soit c appartenant à $\mathcal{K}$, supposons que pour tout n , c n'appartienne pas à $\mathcal{K}_n$. Il existe donc un hyperplan $\alpha_n = (\alpha_n^1, \dots, \alpha_n^k)$ de norme 1 qui sépare $\mathcal{K}_n$ de c on a alors par exemple

$$\langle \alpha_n, c \rangle \geq \langle \alpha_n, c' \rangle, \quad \forall c' \in \mathcal{K}_n.$$

Soit α_∞ un point d'adhérence de la suite $(\alpha_n)_{n \geq 1}$, on a donc

$$\langle \alpha_\infty, c \rangle \geq \langle \alpha_\infty, c' \rangle, \quad \forall c' \in \mathcal{K}.$$

ce qui contredit le fait que $\mathcal{K}$ est ouvert. En conclusion, il existe g de $\mathcal{C}$ et un entier m tels que

$$\int_U \phi(x) g_m(x) dx = c \quad \text{et} \quad \gamma\{g_m(x), x\} \text{ soit } dx\text{-intégrable sur } U.$$

Enonçons maintenant un corollaire généralisant la reconstruction par M.E.M lorsque U n'est plus compact.

Corollaire 3.6 *Sous les hypothèses du lemme 3.5, il existe $v_\infty^*(c)$ de $\mathbb{R}^k$ tel que :*

$$g_\infty^{\text{ME}}(x) = \psi'(\langle v_\infty^*(c), \phi(x) \rangle, x)$$

$$\text{vérifie} \quad \int_U \phi(x) g_\infty^{\text{ME}}(x) dx = c \quad \text{et} \quad g_\infty^{\text{ME}} \in \mathcal{C}.$$

$v_\infty^*(c)$ est l'unique minimum de la fonction $H_{F,\phi}(\cdot, c)$ définie sur $D(F, \phi)$,

$$H_{F,\phi}(v, c) = \int_U \psi(\langle v, \phi(x) \rangle, x) dx - \langle v, c \rangle.$$

Démonstration

Soit g de $\mathcal{C}$ qui vérifie $\int_U \phi(x) g(x) dx = c$. Posons

$$c^{(n)} = \int_{U_n} \phi(x) g(x) dx.$$

D'après le corollaire 3.4, il existe une fonction

$$g_{\infty}^{ME}(x) = \psi'(\langle v_{\infty}^*(c^{(n)}), \phi(x) \rangle, x),$$

qui vérifie $\int_{U_n} \phi(x) g_{\infty}^{ME}(x) dx = c^{(n)}$.

$v_{\infty}^*(c^{(n)})$ est le minimum de la fonction:

$$H_{F,\phi}^{(n)}(v, c^{(n)}) = \int_{U_n} \psi(\langle v, \phi(x) \rangle, x) dx - \langle v, c^{(n)} \rangle.$$

La suite de fonctions $\{ H_{F,\phi}^{(n)}(\cdot, c^{(n)}) \}_{n \geq 1}$ converge en tout point de $D(F, \phi)$ quand n tend vers ∞ vers la fonction

$$H_{F,\phi}(v, c) = \int_U \psi(\langle v, \phi(x) \rangle, x) dx - \langle v, c \rangle.$$

Comme par hypothèse, $a(x) \parallel \phi(x) \parallel$ et $b(x) \parallel \phi(x) \parallel$ sont intégrables sur U le théorème de Lebesgue donne l'existence du gradient de $H_{F,\phi}(\cdot, c)$ sur $D(F, \phi)$ et

$$\frac{\partial}{\partial v} H_{F,\phi}(v, c) = \int_U \phi(x) \psi'(\langle v, \phi(x) \rangle, x) dx - c.$$

Reste à voir que $H_{F,\phi}(v, c)$ possède un minimum sur $D(F, \phi)$, pour cela on procède de la même façon que dans la démonstration du théorème 3.1, compte-tenu du lemme 3.5.

Exemple 3.15 Mesure de Bernoulli (suite de l'exemple 3.13)

Soient U la demi droite $U = \mathbb{R}^+$, $b(x) = e^{-x}$, $a(x) = 0$, et $\phi(x) = 1$, $x \in \mathbb{R}^+$. Donnons-nous comme famille de mesures de référence

$$F_x = \frac{1}{2} (\delta_0 + \delta_{e^{-x}}).$$

On trouve alors

$$g_{\infty}^{\text{ME}}(x) = \frac{e^{-x}}{1 + \exp \{v_{\infty}^*(c) e^{-x}\}}.$$

Remarque 3.14 La mesure sur U

Dans tout le paragraphe III.4.2, on a muni l'ensemble U de la mesure de Lebesgue. On peut en fait la remplacer partout par ν mesure positive sigma finie de support U. On peut en effet faire la même construction, compte-tenu du lemme 3.2.

III.4.3 Tube de variables aléatoires de L^{∞}

On traite ici une extension des exemples précédents (bandes de fonctions continues) lorsque la fonction à reconstruire est à valeurs vectorielles. Plus exactement, on se propose d'utiliser la méthode du M.E.M pour reconstruire une variable aléatoire vectorielle à valeurs dans $\mathbb{R}^r$ et définie sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. La contrainte linéaire est ici du type

$$E_P \{ \phi g \} = c$$

et la contrainte non linéaire est

$$g \in A \text{ P-p.s.}$$

E_P désigne l'espérance sous la probabilité P , A est un convexe ouvert non vide de $\mathbb{R}^r$ donné et $\phi = (\phi_{ij})_{i=1 \dots k, j=1 \dots r}$ est une matrice aléatoire $k \times r$ de $\{ L^{\infty}(\Omega, \mathcal{A}, P) \}^{k \times r}$ donnée. Lorsque $r = 1$, $\Omega = U$ compact de $\mathbb{R}^1$, P la probabilité uniforme sur U et que ϕ est supposée continue, on retrouve l'exemple traité en III.4.2.a.

III.4.3.a Construction d'une suite de problèmes M.E.M

Puisque la contrainte non linéaire est du type $g \in A$ P-p.s., il est naturel de chercher une reconstruction dans $E = \{ L^{\infty}(\Omega, \mathcal{A}, P) \}^r$. Le convexe $\mathcal{C}$ est alors donné par:

$$\mathcal{C} = \{ e \in \{ L^{\infty}(\Omega, \mathcal{A}, P) \}^r, e \in \mathcal{C} \text{ P-p.s, } \mathcal{C} \text{ ensemble fermé inclus dans } A \}.$$

Précisons le procédé de discrétisation utilisé. On suppose que l'espace de probabilité $(\Omega, \mathcal{A}, P)$ est tel qu'il existe une famille d'événements $(A_n^m)_{n \geq 1, m=1, \dots, b_n}$ qui vérifie les propriétés :

$$(P1) \quad A_n^m \cap A_n^{m'} = \emptyset, n \geq 1, m \neq m', m=1, \dots, b_n, m'=1, \dots, b_n.$$

(P2) Notant $\mathcal{A}_n$ la sigma-algèbre engendrée par les événements $A_n^m, m=1, \dots, b_n$ et $E_p(\cdot / \mathcal{A}_n)$ l'espérance conditionnelle sur la tribu $\mathcal{A}_n$, on a :

$$\forall e \in L^\infty(\Omega, \mathcal{A}, P), \lim_{n \rightarrow +\infty} E_p(e / \mathcal{A}_n) = e, P\text{-p.s.}$$

$$(P3) \quad P(A_n^m) = a_n^{-1}, m=1, \dots, b_n.$$

Remarque 3.15 La définition de $\mathcal{C}$

La condition: e appartient à $\mathcal{C}$, P-p.s, avec $\mathcal{C}$ ensemble fermé inclus dans A garantit qu'il existe P-p.s une séparation entre la frontière de A et l'ensemble des valeurs de la variable e (i.e l'ensemble $\{e(\omega), \omega \in \Omega\}$). Remarquons qu'au paragraphe III.4.2 cette propriété était vérifiée grâce à la continuité des fonctions.

Exemple 3.16 $(\Omega, \mathcal{A}, P) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), m(x) dx)$

$$\text{Posons alors } M(x) = \int_{-\infty}^x m(u) du \text{ et } M^{-1}(y) = \inf \{x, M(x) \geq y\},$$

la famille d'événements (A_n^m) définie par: $A_n^m = \{x, M^{-1}(\frac{m-1}{n}) < x \leq M^{-1}(\frac{m}{n})\}$, $m=1, n$ vérifie les propriétés (P1), (P2), (P3).

On définit alors sur $\{L^\infty(\Omega, \mathcal{A}, P)\}^r$ l'opérateur de discrétisation T_n (à valeurs dans $\mathbb{R}^{r \times b_n}$) par:

$$T_n e = (E_p(e / 1_{A_n^m}))_{m=1, \dots, b_n}, e \in \{L^\infty(\Omega, \mathcal{A}, P)\}^r.$$

Puis, l'opérateur R_n défini sur $\mathbb{R}^{r \times b_n}$ à valeurs dans $\{L^\infty(\Omega, \mathcal{A}, P)\}^r$ (E et F coïncident ici) par:

$$R_n \omega_{b_n} = \sum_{m=1}^{b_n} 1_{A_n^m} \omega_{b_n}^m,$$

où $\omega_{b_n} = (\omega_{b_n}^1, \dots, \omega_{b_n}^{b_n})$ est l'élément générique de $\mathbb{R}^{r \times b_n}$ et $\omega_{b_n}^m \in \mathbb{R}^r, m=1, \dots, b_n$.

Enfin l'opérateur linéaire ϕ_n est défini par:

$$\phi_n \omega_{b_n} = \sum_{m=1}^{b_n} \mathbb{E}_P \{ \phi 1_{A_n^m} \} \omega_{b_n^m}, \quad \omega_{b_n} \in \mathbb{R}^{r \times b_n}.$$

On a donc en particulier pour e de $\{L^\infty(\Omega, \mathcal{A}, P)\}^r$ l'égalité

$$\phi_n T_n e = \mathbb{E}_P \{ \mathbb{E}_P(\phi / \mathcal{A}_n) \mathbb{E}_P(e / \mathcal{A}_n) \}.$$

Donc, compte-tenu de la propriété (P2) sur la famille (A_n^m) , on a bien

$$\forall e \in \{L^\infty(\Omega, \mathcal{A}, P)\}^r, \lim_{n \rightarrow +\infty} \phi_n T_n e = \mathbb{E}_P(\phi e).$$

Nous supposons comme d'habitude que ϕ est de rang plein, c'est-à-dire

$$\alpha \in \mathbb{R}^k, \forall e \in \{L^\infty(\Omega, \mathcal{A}, P)\}^r, \alpha^T \mathbb{E}_P(\phi e) = 0 \Rightarrow \alpha_1 = \dots = \alpha_k = 0.$$

Remarquons que cela équivaut à

$$\alpha \in \mathbb{R}^k, \alpha^T \phi = 0 \text{ P-p.s.} \Rightarrow \alpha_1 = \dots = \alpha_k = 0.$$

On trouve encore ici que le convexe $\mathcal{C}_n$ (égal par définition à $T_n \mathcal{C}$) est $\mathcal{C}_n = A^n$. On le munit comme précédemment d'une mesure produit $\mu_n = F^{\otimes n}$ avec F une mesure positive sur A . Etendons de façon naturelle les notations utilisées en III.4.b en posant

$$D(F) = \left\{ \tau \in \mathbb{R}^r, \int_A \exp(\tau \cdot y) dF(y) < \infty \right\},$$

où $(\cdot)$ désigne ici le produit scalaire usuel sur $\mathbb{R}^r$.

Pour τ de $D(F)$, on pose, $\psi(\tau) = \text{Log} \int_A \exp(\tau \cdot y) dF(y)$ et $\psi'(\tau) = \frac{\partial}{\partial \tau} \psi(\tau)$.

On a alors l'analogie du lemme 3.3

Lemme 3.6 On fait les hypothèses suivantes:

(L1) L'enveloppe convexe fermée du support de F est $\bar{A}$.

(L2) $D(F)$ est un ouvert non vide de $\mathbb{R}^r$ et il existe v de $\mathbb{R}^k$, tel que ,
 $v^T \phi \in D(F)$, P-p.s.

Alors pour n assez grand, le problème de maximisation de l'entropie sur la moyenne possède l'unique solution :

$$\frac{dP_n^{ME}}{d\mu_n}(\omega_{b_n}) = \prod_{m=1}^{b_n} \exp \{ v_n^{*T}(c) E_P(\phi/A_n^m) \omega_{b_n}^m - \psi[v_n^{*T}(c) E_P(\phi/A_n^m)] \}$$

et $g_n^{ME} = \sum_{m=1}^{b_n} \psi' \{ v_n^{*T}(c) E_P(\phi/A_n^m) \} 1_{A_n^m}$. Où $v_n^*(c)$ est le minimum de la fonction:

$$H_{\mu_n, \phi, \omega_n}^{ren}(v, c) = a_n^{-1} \sum_{m=1}^{b_n} \psi \{ v^T E_P(\phi/A_n^m) \} - \langle v, c \rangle.$$

On démontre ce lemme de la même façon que l'on a démontré le lemme 3.3

III.4.3.b Convergence de la suite des problèmes M.E.M

On a le résultat suivant:

Corollaire 3.7 Sous les hypothèses (L1), (L2),

(L3) L'ensemble $D(F, \phi) = \{ v \in \mathbb{R}^k, E_P \{ | \psi(v^T \phi) | \} < \infty \}$ est non vide et coïncide avec

$$D'(F, \phi) = \{ v \in \mathbb{R}^k, \exists C \text{ fermé inclus dans } D(F) \text{ avec } v^T \phi \in C, \text{ P-p.s.} \}.$$

La suite des problèmes MEM converge dans $\{ L^\infty(\Omega, \mathcal{A}, P) \}^r$.

On a $g_\infty^{ME} = \psi'(v_\infty^{*T}(c) \phi)$ et $v_\infty^*(c)$ minimise la fonction

$$H_{F, \phi}(v, c) = E_P \{ \psi(v^T \phi) \} - \langle v, c \rangle.$$

Démonstration

Soit v un élément de $D'(F, \phi)$, par définition il existe un ensemble fermé C inclus dans $D(F)$ tel que $v^T \phi \in C$, P-p.s. Remarquons que puisque ϕ est bornée P-p.s, C peut être choisi compact. Soit C' l'enveloppe convexe fermée de C , $D(F)$ étant convexe ouvert il contient C' .

On a

$$\| \psi \{ v^T E_P(\phi / \mathcal{A}_n) \} \|_{\infty} \leq \sup_{y \in C} | \psi(y) |$$

et donc,
$$\lim_{n \rightarrow +\infty} E_P [\psi \{ v^T E_P(\phi / \mathcal{A}_n) \}] = E_P [\psi \{ v^T \phi \}].$$

On pose ensuite pour y de A et τ de $D(F)$:

$$\psi^{-1}(y) = \tau \text{ si } \psi(\tau) = y \text{ et } \gamma(y) = - (y \cdot \psi^{-1}(y)) + \psi(\psi^{-1}(y)).$$

On conclut alors comme dans la démonstration du théorème 3.1 en utilisant que pour toute variable g de $\{ L^{\infty}(\Omega, \mathcal{A}, P) \}^r$ qui vérifie $E_P(\phi g) = c$ et g élément de A , P -p.s, on a, $H_{F,\phi}(v, c) \geq E_P(\gamma(g))$, $v \in D(F, \phi)$.

Exemple Commande d'un système

Soit le système dynamique

$$\begin{aligned} X'(\sigma) &= M X(\sigma) + N u(\sigma), \sigma \geq 0 \\ X(0) &= X_0 \neq 0 \end{aligned}$$

où $X(\cdot)$ est un vecteur de $\mathbb{R}^k$, $u(\cdot)$ un vecteur de $\mathbb{R}^r$ et N une matrice de rang plein. Pour T réel positif donné et A ouvert de $\mathbb{R}^r$ fixé, on s'intéresse à la construction d'une commande (en supposant qu'il en existe une) $u(\sigma)$, pour σ appartenant à $[0, T]$, à valeurs dans A et ramenant le système en 0 à la date T . Réécrivons ce problème comme un problème de reconstruction. On a

$$X(t) = \exp(tM) X_0 + \exp(tM) \int_0^t \exp(-\sigma M) N u(\sigma) d\sigma$$

et l'on veut :

$$\int_0^T \exp(-\sigma M) N u(\sigma) \frac{d\sigma}{T} = \frac{X_0}{T}.$$

La commande devant aussi satisfaire la contrainte: $\forall \sigma \in [0, T], u(\sigma) \in A$. Utilisons les résultats précédents pour construire une telle commande.

Exemple 3.17 $A =]0, 1]^r$

Choisissons comme mesure de référence $F = (\delta_0 + \delta_1)^{\otimes r}$. Notant, $[v]_j$ la j -ième composante du vecteur v , la commande donnée par la construction M.E.M définie précédemment est

$$[u]_j(\sigma) = \frac{\exp \{ [N^T \exp(-M^T \sigma) v_\infty^*]_j \}}{1 + \exp \{ [N^T \exp(-M^T \sigma) v_\infty^*]_j \}},$$

où v_∞^* est défini par l'équation $\int_0^T \exp(-\sigma M) N u(\sigma) d\sigma = X_0$.

Exemple 3.18 Boule unité de $\mathbb{R}^r$.

On prend comme mesure de référence F , la probabilité uniforme. Alors, on a (voir le paragraphe III.4.1) pour $\tau \in \mathbb{R}^r$,

$$\psi(\tau) = \text{Log} \int_{-1}^1 \exp(\|\tau\| \xi) (1 - \xi^2)^{(n-1)/2} \gamma_n d\xi.$$

Pour σ de l'intervalle $[0, T]$, la commande, $u(\sigma) = \psi\{N^T \exp(-M^T \sigma) v_\infty^*\}$, où v_∞^* est défini par

$$\int_0^T \exp(-\sigma M) N u(\sigma) d\sigma = X_0.$$

est donc une solution du problème de commande.

Etant donnés X_0 , A et T , on peut aussi s'intéresser à l'existence d'une commande à valeurs dans A ramenant le système en 0 à la date T . Nous verrons (chapitre IV) que l'on peut répondre complètement à la question lorsque A est un pavé borné de $\mathbb{R}^r$.

III.5 PROBLEME M.E.M ET MAXIMISATION D'UN CRITERE CONCAVE

III.5.1 Préambule

Si μ_n est une mesure positive dont l'enveloppe convexe du support est $\text{cl}(\mathcal{C}_n)$ ($\mathcal{C}_n$ ouvert de $\mathbb{R}^n$), intéressons-nous à la valeur de la μ_n - entropie pour un élément du ω_n - arc d'Hellinger de μ_n .

$$\mu_{\omega_n, \beta} = [Z_{\mu_n, \omega_n}(\beta)]^{-1} \exp(\beta, \omega_n) \mu_n, (\beta \in D(\mu_n, \omega_n)).$$

La μ_n - entropie de $\mu_{\omega_n, \beta}$ est

$$S_{\mu_n}(\mu_{\omega_n, \beta}) = \text{Log } Z_{\mu_n, \omega_n}(\beta) - (\beta, \int_{C_n} \omega_n d\mu_{\omega_n, \beta}).$$

Considérons maintenant un point y de C_n , si l'on suppose que le domaine de définition de la transformée de Laplace de μ_n , $D(\mu_n, \omega_n)$ est un ouvert non vide de $\mathbb{R}^n$, il existe (corollaire 2.1) $\beta^*(y)$ avec

$$\int_{C_n} \omega_n d\mu_{\omega_n, \beta^*(y)} = y.$$

$\mu_{\omega_n, \beta^*(y)}$ est la probabilité du maximum de μ_n - entropie vérifiant la contrainte linéaire précédente et l'on a :

$$\begin{aligned} S_{\mu_n}(\mu_{\omega_n, \beta^*(y)}) &= \text{Log } Z_{\mu_n, \omega_n}(\beta^*(y)) - (\beta^*(y), \int_{C_n} \omega_n d\mu_{\omega_n, \beta^*(y)}) \\ &= \Gamma_n(y) = \inf_{\beta \in \mathbb{R}^n} \{ \text{Log } Z_{\mu_n, \omega_n}(\beta) - (\beta, y) \} \end{aligned}$$

$\Gamma_n(y)$ est donc une fonction concave de y (c'est l'opposée de la transformée de Legendre de $\text{Log } Z_{\mu_n, \omega_n}(\cdot)$). Soit maintenant le problème de μ_n - entropie maximale lorsque la contrainte linéaire est du type:

$$\phi_n \int_{C_n} \omega_n dP_n(\omega_n) = c,$$

où ϕ_n est une matrice $k \times n$ et c un vecteur de $\mathbb{R}^k$ (tous les deux sont des données). En supposant que c soit tel que ce problème d'entropie possède l'unique solution P^{ME} (corollaire 2.1), on a

$$S_{\mu_n}(P^{\text{ME}}) = \inf_{v \in \mathbb{R}^k} \{ \text{Log } Z_{\mu_n, \omega_n}(\phi_n^T t) - \langle t, c \rangle \}.$$

Soit y de C_n réalisant c par l'application ϕ_n ($c = \phi_n y$), on a alors

$$\begin{aligned} S_{\mu_n}(P^{\text{ME}}) &= \inf_{v \in \mathbb{R}^k} \{ \text{Log } Z_{\mu_n, \omega_n}(\phi_n^T t) - \langle t, c \rangle \} \\ &\geq \inf_{\beta \in \mathbb{R}^n} \{ \text{Log } Z_{\mu_n, \omega_n}(\beta) - (\beta, y) \}. \end{aligned}$$

Posant $y^{\text{ME}} = \int_{C_n} \omega_n dP^{\text{ME}}(\omega_n)$, on en déduit que :

$$\Gamma_n(y^{ME}) \geq \Gamma_n(y), \quad \forall y \in \mathcal{C}_n, \quad \phi_n y = c.$$

En d'autres termes, la moyenne de la probabilité du maximum d'entropie maximise la fonctionnelle $\Gamma_n(\cdot)$. Et donc, si l'on ne s'intéresse qu'à la moyenne y^{ME} , c'est la solution du problème de maximisation de Γ_n sur l'ensemble:

$$\mathbb{L}_n(c) = \{y \in \mathcal{C}_n, \phi_n y = c\}.$$

Nous allons voir que cette propriété est conservée lorsque l'on s'intéresse à g_∞^{ME} .

III.5.2 Maximisation de la fonctionnelle de Legendre

III.5.2.a Bande de $C(U)$

Comme cadre général à la minimisation de normes, à la maximisation de l'entropie usuelle ou celle de Burg, on peut proposer la situation suivante : soit $(\exp \{ \psi(t, x) \})_{x \in U}$ la famille des transformées de Laplace de la famille de mesures $(F_x)_{x \in U}$. On suppose que la famille $(F_x)_{x \in U}$ vérifie les hypothèses du théorème 3.1 ou de l'un des corollaires du paragraphe III.4. Le problème M.E.M a alors pour solution:

$$g_\infty^{ME}(x) = \psi'(\langle v_\infty^*(c), \phi(x) \rangle, x) \text{ avec } \int_U \psi'(\langle v_\infty^*(c), \phi(x) \rangle, x) \phi(x) dx = c.$$

Les fonctions de type négatif jouent donc un rôle très particulier. Considérons la fonctionnelle de Legendre

$$\Gamma(e) = \int_U \gamma(e(x), x) dx, \text{ où}$$

$$\forall x \in U, \forall y \in [a(x), b(x)], \gamma(y, x) = -y \psi'^{-1}(y, x) + \psi \{ \psi'^{-1}(y, x), x \}.$$

$\Gamma(\cdot)$ est définie pour e de $\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \}$.

On a alors le théorème suivant:

Théorème 3.2 g_{∞}^{ME} est l'unique solution de $g = \arg \max_{e \in \mathbb{L}(c)} \Gamma(e)$, où

$$\mathbb{L}(c) = \{ e \in \mathcal{C}, \int_U e(x) \phi(x) dx = c \}.$$

Démonstration

On a vu au paragraphe III.4 (dans la démonstration du théorème 3.1) que

$$\forall e \in \mathbb{L}(c), \Gamma(e) \leq H_{F,\phi}(v_{\infty}^*(c), c).$$

D'autre part,

$$\begin{aligned} \Gamma(g_{\infty}^{ME}) &= \int_U \gamma \{ \psi'(\langle v_{\infty}^*(c), \phi(x) \rangle, x), x \} dx \\ &= \int_U \psi(\langle v_{\infty}^*(c), \phi(x) \rangle, x) dx - \langle v_{\infty}^*(c), c \rangle \\ &= H_{F,\phi}(v_{\infty}^*(c), c). \end{aligned}$$

Donc, g_{∞}^{ME} est une solution du problème d'optimisation. On conclut en remarquant que $\gamma(\cdot, x)$ est une fonction strictement concave, ce qui assure l'unicité de la solution. Remarquons que l'on aurait pu conclure directement par un calcul standard de variations (par les équations d'Euler).

Exemples

Exemple 3.19 Loi normale et moindres carrés¹

Soit F_x la loi normale de variance $f(x)$ (f est positive et continue sur U).

On a

$$\psi(\tau, x) = \frac{f(x)}{2} \tau^2 \quad (\tau \in \mathbb{R}, x \in U)$$

D'où :

$$\Gamma(e) = \frac{1}{2} \int_U \frac{e^2(x)}{f(x)} dx, \quad e \in \mathcal{C}(U).$$

On retrouve donc la méthode des moindres carrés.

¹Suite des exemples 3.6 et 3.11.

Exemple 3.20 Loi de Poisson et entropie²

Soit F_x la loi de Poisson de paramètre $f(x)$ (f est positive et continue sur U). On a

$$\psi(\tau, x) = f(x) (e^\tau - 1), (\tau \in \mathbb{R}, x \in U).$$

Donc pour e de $C(U)$ qui vérifie, $\forall x \in U, e(x) > 0$, on a

$$\Gamma(e) = - \int_U e(x) \operatorname{Log} \frac{e(x)}{f(x)} dx + \int_U \{ e(x) - f(x) \} dx = -K(e, f),$$

En particulier, lorsque $\phi_1 = 1$ et $c_1 = 1$, on retrouve la f -entropie.

Exemple 3.21 Mesure de Lebesgue et entropie de Burg³

Soit F_x la mesure de Lebesgue sur $\mathbb{R}^+$. On a

$$\psi(\tau, x) = -\operatorname{Log}(-\tau), (\tau \in \mathbb{R}^{*-}, x \in U).$$

Le critère associé est l'entropie de Burg :

$$\Gamma(e) = \int_U \operatorname{Log} e(x) dx + \int_U dx.$$

Il est défini pour e de $C(U)$ tel que, $\forall x \in U, e(x) > 0$. Dans le cas où $\phi_1 = 1$, on peut aussi interpréter l'entropie de Burg comme un problème usuel avec la mesure à priori G associée à la mesure aléatoire définie par le semi groupe $\gamma(t, 1)$ de lois indéfiniment divisibles (voir le paragraphe III.2.1). La construction par M.E.M de l'entropie de la prédiction permet d'autre part de donner une nouvelle interprétation de la méthode de Burg (estimation d'une densité spectrale), ce point de vue sera développé au paragraphe V.3.

Exemple 3.22 Mesure de Bernoulli et entropie de Fermi-Dirac⁴

Soit $F_x = \frac{1}{2} \{ \delta_{a(x)} + \delta_{b(x)} \}$, on a

$$\psi(\tau, x) = \operatorname{Log} \frac{1}{2} \{ e^{a(x)\tau} + e^{b(x)\tau} \}, (\tau \in \mathbb{R}, x \in U).$$

Le critère est donc l'entropie de Fermi-Dirac [Gull-Skil, 85]

²Suite des exemples 3.7 et 3.12.

³Suite de l'exemple 3.10.

⁴Suite des exemples 3.8 et 3.13

$$\Gamma(e) = \int_U - \frac{e(x) - a(x)}{b(x) - a(x)} \operatorname{Log} \left\{ \frac{e(x) - a(x)}{b(x) - a(x)} \right\} dx \\ + \int_U - \frac{b(x) - e(x)}{b(x) - a(x)} \operatorname{Log} \left\{ \frac{b(x) - e(x)}{b(x) - a(x)} \right\} dx + \int_U \operatorname{Log} 2 \, dx.$$

Il est défini sur la bande, $\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \}$. L'entropie de Fermi-Dirac possède la propriété d'être une fonctionnelle bornée sur le convexe $\mathcal{C}$. Cette propriété nous permettra au chapitre IV de résoudre une famille de problèmes de moments généralisés.

III.5.2.b Tube de $\{ L^\infty(\Omega, \mathcal{A}, P) \}^r$

Reprenons les notations du paragraphe III.4.3 et posons

$$\mathbb{L}(c) = \{ e \in \mathcal{C}, \operatorname{Ep} \{ \phi e \} = c \}$$

et pour e du convexe $\mathcal{C}$, $\Gamma(e) = \operatorname{Ep} \{ \gamma(e) \}$,

où pour y de A , $\gamma(y) = - (y \cdot \psi'^{-1}(y)) + \psi \{ \psi'^{-1}(y) \}$.

On a alors l'analogie du théorème 3.2:

Corollaire 3.8 *Sous les hypothèses (L1), (L2), (L3)*

$g_{\infty}^{\operatorname{ME}}$ est l'unique solution de $g = \arg \max_{e \in \mathbb{L}(c)} \Gamma(e)$.

Exemple 3.23 Mesure de Lebesgue

Prenons $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, P la loi de Cauchy. Soit F la mesure de Lebesgue sur $\mathbb{R}^+$. On trouve alors pour e variable aléatoire sur $(\Omega, \mathcal{A})$, P -p.s. positive

$$\Gamma(e) = \int_{-\infty}^{+\infty} \frac{\operatorname{Log} e(x)}{\pi (1 + x^2)} \, dx.$$

Dans [Chov, 61] on trouve une autre construction de la fonctionnelle Γ (voir le paragraphe II.2)

III.5.3 ψ - divergence et théorème de Pythagore

III.5.3.a Information de Kullback et théorème de Pythagore

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et ξ un sous-ensemble convexe de $\mathbb{P}(\Omega)$ ensemble des probabilités sur $(\Omega, \mathcal{A})$. On suppose qu'il existe une probabilité P^{ME} de ξ réalisant l'infimum de $K(\cdot, P)$ sur ξ . P^{ME} possède alors la caractérisation suivante ([Csis, 75] théorème 2.2):

$$Q \in \xi, Q = P^{ME} \iff \forall R \in \xi, K(R, P) \geq K(R, Q) + K(Q, P).$$

Dans le cas particulier où ξ est une intersection finie d'hyperplan :

$$\xi = \{ Q \in \mathbb{P}(\Omega), E_Q[\phi] = c \} = \mathbb{P}(c, \phi),$$

et sous les hypothèses du corollaire 2.1, on a :

$$\frac{dP^{ME}}{dP} = \frac{\exp \langle t^*(c), \phi \rangle}{Z_{P, \phi}(t^*(c))}.$$

Donnons-nous une probabilité R de ξ , d'information de Kullback par rapport à P finie. On a alors :

$$\begin{aligned} K(R, P) &= \int_{\Omega} \text{Log} \frac{dR}{dP} dR = \int_{\Omega} \text{Log} \frac{dR}{dP^{ME}} dR + \int_{\Omega} \text{Log} \frac{dP^{ME}}{dP} dR \\ &= K(R, P^{ME}) + K(P^{ME}, P). \end{aligned}$$

C'est l'analogie du théorème de Pythagore, P^{ME} pouvant être interprétée comme une "projection" de P sur le sous-espace ξ . Nous allons voir que la construction du M.E.M induit une propriété analogue pour une dissemblance entre élément du convexe fonctionnel $\mathcal{C}$. Cette dissemblance est fonction de la suite des mesures de référence $(\mu_n)_{n \geq 1}$. Nous limiterons notre étude au cas où $\mathcal{C}$ est la bande de fonctions continues:

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a < e(x) < b \}, (-\infty \leq a < b \leq +\infty)$$

et U est un ensemble compact. Mais des énoncés analogues peuvent être donnés pour les autres exemples.

III.5.3.b Cas de la ψ - divergence

Le concept de divergence vise à englober différentes "distances" statistiques classiques comme l'information de Kullback, la distance en variation, du χ^2 , etc... Les fonctions de type négatif liées aux entropies en moyenne donnent une classe de divergence:

Définition 3.3 Soient f et g deux éléments du convexe

$$\mathcal{C} = \{e \in C(U), \forall x \in U, a < e(x) < b\}, (-\infty \leq a < b \leq +\infty)$$

et ψ la Log transformée de Laplace d'une mesure positive F vérifiant les hypothèses (K1), (K2) du paragraphe III.4.2.a.1. On appelle ψ - divergence de f sur g la quantité (notée dans la suite $D_\psi(f, g)$):

$$\int_U f(x) [\psi^{-1}\{f(x)\} - \psi^{-1}\{g(x)\}] - [\psi(\psi^{-1}\{f(x)\}) - \psi(\psi^{-1}\{g(x)\})] dx$$

Ce type de divergence est étudié d'un point de vue statistique dans [Ag- Bouc,80]. Cela définit bien une divergence au sens classique du terme puisque l'on a:

Propriété 3.1

$$a-D_\psi(f, g) \geq 0, \forall f, g \in \mathcal{C}.$$

$$b-D_\psi(f, g) = 0 \iff f = g.$$

$$c- \text{Si } F \text{ est une probabilité et que, } \forall x \in U, g(x) = \psi(0), \\ \text{alors, } D_\psi(f, g) = -\Gamma(f).$$

Démonstration

Points a et b

Pour t, t' de $[a, b]$, on pose,

$$\eta(t, t') = t [\psi^{-1}(t) - \psi^{-1}(t')] - [\psi(\psi^{-1}(t)) - \psi(\psi^{-1}(t'))].$$

On a alors, $\frac{\partial \eta}{\partial t}(t, t') = \psi^{-1}(t) - \psi^{-1}(t')$, cette dérivée change une seule fois de signe en $t = t'$.

Point c

Si $g(x) = \psi'(0)$, $x \in U$, un simple calcul donne : $D_\psi(f, g) = -\Gamma(f)$.

Exemples 3.24 Quelques divergences classiques

- $]a, b[=]-\infty, +\infty[$, la normale centrée réduite donne bien sûr $\|f - g\|_L^2$.
- $]a, b[=]0, +\infty[$, la loi de Poisson redonne l'information de Kullback.
- Soit F l'exponentielle de convolution de la loi exponentielle de paramètre 1. On a pour τ strictement inférieur à 1, $\psi(\tau) = \frac{1}{1-\tau} - 1$. D'où

$$D_\psi(f, g) = \int_U \frac{[f(x)^{1/2} - g(x)^{1/2}]^2}{g(x)^{1/2}} dx,$$

cette divergence est définie pour f et g des fonctions strictement positives sur U . C'est la distance d'Hellinger pondérée.

- Pour F la loi exponentielle de paramètre 1, on trouve la 0 - divergence

$$D_\psi(f, g) = \int_U \left\{ -\text{Log} \left[\frac{f(x)}{g(x)} \right] + \frac{f(x)}{g(x)} - 1 \right\} dx,$$

définie pour f et g des fonctions strictement positives sur U .

- $]a, b[=]0, 1[$ et $F = \delta_0 + \delta_1$, on trouve pour f et g des fonctions du convexe

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, 0 < e(x) < 1 \},$$

$$D_\psi(f, g) = K(f, g) + K(1-f, 1-g)$$

Enonçons maintenant l'analogue du théorème de Pythagore pour la ψ - divergence:

Théorème 3.3 On suppose que les hypothèses (K1), (K2) du paragraphe III.4.2.a.1 sont vérifiées, et qu'il existe $v_{\infty}^*(c)$ de $\mathbb{R}^k$, tel que :

$$g_{\infty}^{ME}(x) = \psi'(\langle v_{\infty}^*(c), \phi(x) \rangle),$$

appartienne à $\mathbb{L}(c) = \{e \in \mathcal{C}, \int_U e(x) \phi(x) dx = c\}$.

Alors g_{∞}^{ME} vérifie :

- a- $\forall g \in \mathbb{L}(c), \Gamma(g) = \Gamma(g_{\infty}^{ME}) - D_{\psi}(g, g_{\infty}^{ME})$. En particulier, si F est une probabilité, on a $\forall g \in \mathbb{L}(c) D_{\psi}(g, \psi'(0)) = D_{\psi}(g_{\infty}^{ME}, \psi'(0)) + D_{\psi}(g, g_{\infty}^{ME})$.
- b- Posons $\mathcal{E}(\psi, \phi) = \{ \psi'(\langle v, \phi(x) \rangle), v \in D(F, \phi) \}$. Alors pour g fixé de $\mathbb{L}(c)$, g_{∞}^{ME} réalise l'infimum de $D_{\psi}(g, \cdot)$ sur $\mathcal{E}(\psi, \phi)$.

Démonstration

Point a

Pour g de $\mathbb{L}(c)$ on a

$$\begin{aligned} D_{\psi}(g, g_{\infty}^{ME}) &= \int_U g(x) \left[\psi'^{-1}[g(x)] - \langle v_{\infty}^*(c), \phi(x) \rangle \right] dx \\ &\quad - \int_U \left[\psi(\psi'^{-1}[g(x)]) - \psi(\langle v_{\infty}^*(c), \phi(x) \rangle) \right] dx \\ &= -\Gamma(g) + \Gamma(g_{\infty}^{ME}). \end{aligned}$$

Point b

Pour $e = \psi(\langle v, \phi(x) \rangle)$, v de $D(F, \phi)$ et g de $\mathbb{L}(c)$ fixée, on a

$$D_{\psi}(g, e) = -\Gamma(g) + H_{F, \phi}(v, c),$$

donc $D_{\psi}(g, \psi(\langle v, \phi(x) \rangle))$ est minimum pour $v = v_{\infty}^*(c)$.

CHAPITRE IV

LE PROBLEME DES MOMENTS GENERALISES

IV.1 INTRODUCTION

Etant donnés un convexe $\mathcal{C}$ d'un espace de fonctions E , $(\phi_j)_{j \in I}$ une famille finie de fonctions du dual de E , et $(c_j)_{j \in I}$ une famille de réels, le problème des moments généralisés consiste à étudier l'existence et à caractériser les distributions G vérifiant:

$$\forall j \in I, \int_U \phi_j(x) dG(x) = c_j, \quad \text{et} \quad G \in \text{cl}(\mathcal{C}),$$

où $\text{cl}(\mathcal{C})$ désigne la fermeture de $\mathcal{C}$ pour une topologie à préciser. Un exemple classique est le problème des moments trigonométriques sur le tore $\mathbb{T}$ ([Gren-Szeg, 58], chap 4, p 56). Dans cet exemple E est l'ensemble des fonctions continues sur le tore, $\mathcal{C}$ est l'ensemble des fonctions de E strictement positives sur $\mathbb{T}$, sa fermeture est le cône des mesures positives $\mathcal{M}(\mathbb{T})$ (c'est la fermeture faible de $\mathcal{C}$), enfin $(\phi_j)_{j \in I}$ est le système d'exponentielles complexes $(\exp(ijx))_{j \in \{0, 1, \dots, k\}}$. L'existence d'une solution est alors liée à la positivité de la matrice de Toeplitz construite à partir des coefficients $(c_j)_{j=0 \dots k}$. Une étude de ce problème est faite en utilisant l'entropie de Burg (cf exemples 3.10, 3.21) par H.J. Landau [Land, 87]. Plus généralement, le livre de M.G. Krein et A.A. Nudel'man [Krein-Nud, 77] traite de tous les problèmes de moments classiques. Nous étudions sous l'angle du M.E.M (paragraphe IV.2), le cas où E est $C(U)$, l'espace des fonctions continues sur U , $\mathcal{C}$ est la bande

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \}$$

et $\phi = (\phi_j)_{j=1..k}$ est une fonction continue sur U . Nous appellerons ce problème, problème des ϕ - moments sur $\mathcal{C}$. Pour chaque famille de mesures positives $(F_x)_{x \in U}$ vérifiant les hypothèses permettant d'assurer l'existence d'une solution au problème M.E.M, si pour $c = (c_j)_{j=1..k}$ il existe une solution au problème des ϕ - moments sur $\mathcal{C}$, appartenant à $\mathcal{C}$, alors d'après le chapitre III, il en existe une de la forme $\psi'(\langle v_\infty^*(c), \phi(x) \rangle, x)$, où $\psi'(\cdot, x)$ est la Log-transformée de Laplace de la mesure F_x . Lorsque U est compact et que $a(x)$ et $b(x)$ sont finies, nous résolvons complètement le problème des ϕ - moments sur $\mathcal{C}$ en utilisant le M.E.M (paragraphe IV.2.2). Plus précisément, nous utilisons la famille de mesures de référence $(F_x)_{x \in U}$, avec $F_x = \delta_{a(x)} + \delta_{b(x)}$ (cf. exemples 3.8, 3.13, 3.14 et 3.22) pour construire un critère numériquement calculable permettant de décider si une suite de réels $(c_j)_{j=1..k}$ donnée est une suite de ϕ - moments pour un élément de $cl(\mathcal{C})$ (le sens de $cl(\mathcal{C})$ est précisé plus bas).

Pour $c \notin \mathcal{C}$ strictement réalisable dans $\mathcal{C}$, nous caractérisons ensuite les points extrémaux du convexe des réalisations de c dans $\mathcal{C}$, c'est-à-dire de l'ensemble

$$\bar{\mathcal{L}}(c) = \{ e \in cl(\mathcal{C}), \int_U e(x) \phi(x) dx = c \},$$

($cl(\mathcal{C})$ est la fermeture de $\mathcal{C}$ dans $L^\infty(U, dx)$ (voir IV.2.2), toutes les distributions de $cl(\mathcal{C})$ sont donc ici des fonctions).

Nous étendons au paragraphe IV.3 ces résultats au cas multidimensionnel où E est l'espace de variables aléatoires $\{ L^\infty(\Omega, \mathcal{A}, P) \}^r$ et $\mathcal{C}$ le convexe

$$\mathcal{C} = \{ e \in \{ L^\infty(\Omega, \mathcal{A}, P) \}^r, e \in C, P\text{-p.s.}, C \text{ ensemble fermé inclus dans } A \}^1,$$

où A est un polytope ouvert, borné, c'est-à-dire l'intérieur de l'enveloppe convexe d'un nombre fini de points ne se trouvant pas tous dans le même hyperplan.

¹ Voir la remarque 3.15 pour la définition de $\mathcal{C}$.

IV.2 Cas d'une bande de $C(U)$

Nous étudierons dans ce paragraphe l'existence d'une fonction g de

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \}, \quad \text{telle que} \quad \int_U g(x) \phi(x) dx = c,$$

où c est un élément de $\mathbb{R}^k$ fixé. Par commodité d'écriture, les résultats seront énoncés pour U compact et $a(\cdot)$ et $b(\cdot)$ des fonctions constantes. Pour chacun d'eux, nous dirons s'il est possible de l'étendre à U , a et b quelconques.

IV.2.1 Cas général

Un des problèmes de moments bien connu est celui où U est le tore $\mathbb{T}$, l'intervalle $]a, b[$ est $\mathbb{R}^{+*}$ et ϕ est le système d'exponentielles complexes $(\exp(ijx))_{j \in \{0, 1, \dots, k\}}$. Dans ce cas c est ϕ strictement réalisable dans $\mathcal{C}$ si et seulement si c est de type strictement positif, c'est-à-dire si la matrice de Toeplitz $T_k(c) = (c_{i-j})_{1 \leq i, j \leq k}$, ($c_{-j} = \bar{c}_j$) est définie positive. Dans ce cas, la solution classiquement obtenue est de la forme $|P_k(\exp(ix))|^{-2}$, où P_k est le polynôme de degré k

$$P_k(z) = \sum_{j=0}^k a_j z^j, \quad \text{avec,} \quad (1, 0, \dots, 0) T_k^{-1}(c) = a_0 (a_0, a_1, \dots, a_k), \quad [\text{Far}, 84].$$

Cette fonction est la densité spectrale du processus gaussien d'entropie maximale dont la covariance est une extension de c , elle peut aussi être obtenue par MEM si l'on prend comme mesure de référence F la mesure de Lebesgue sur $\mathbb{R}^+$ (voir les exemples 3.10, 3.21 et le paragraphe V.3.3). Dans le cas général, la construction du MEM donne une famille de solutions du type, $\psi(\langle v_\infty^*(c), \phi(x) \rangle)$ où ψ varie dans un certain sous-ensemble de fonctions de type négatif, plus précisément, utilisant les résultats (ainsi que les notations et les hypothèses) du chapitre III, on a:

Théorème 4.1 Soient ϕ une application continue définie sur U à valeurs dans $\mathbb{R}^k$ et c de $\mathbb{R}^k$ donnés. Soit (P) le problème : trouver g fonction continue sur U , vérifiant :

$$\forall x \in U, a < g(x) < b \text{ et } \int_U g(x) \phi(x) dx = c, \text{ (} a \text{ et } b \text{ éléments de } \bar{\mathbb{R}} \text{ donnés)}$$

Alors les conditions suivantes sont équivalentes :

1- (P) a une solution.

2- Il existe ψ , telle que $\exp \psi$ soit la transformée de Laplace d'une mesure positive vérifiant les hypothèses (K1), (K2), (K3) (ou (K1), (K2), (K'3), (K4)) et telle que (P) ait pour solution $\psi'(\langle v_\infty^*(c), \phi(x) \rangle)$, où $v_\infty^*(c)$ est défini par

$$\int_U \psi'(\langle v_\infty^*(c), \phi(x) \rangle) \phi(x) dx = c.$$

3- Pour toute ψ telle que $\exp \psi$ soit la transformée de Laplace d'une mesure positive vérifiant les hypothèses (K1), (K2), (K3) (ou (K1), (K2), (K'3), (K4)) (P) a pour solution $\psi'(\langle v_\infty^*(c), \phi(x) \rangle)$, où $v_\infty^*(c)$ est défini par

$$\int_U \psi'(\langle v_\infty^*(c), \phi(x) \rangle) \phi(x) dx = c.$$

Extensions

Dans la cas précédent, (bande fixe), on peut donner un énoncé analogue lorsqu'on utilise une famille de mesure $(F_x)_{x \in U}$ en remplaçant les hypothèses (K1), (K2), (K3) par celles du corollaire 3.3. Cet énoncé est encore vrai dans le cas général (U non compact et/ou $a(x)$ et $b(x)$ des fonctions continues) en remplaçant les hypothèses (K1), (K2), (K3) par les hypothèses correspondantes (corollaires 3.4 et 3.6). Dans le cas non compact, le deuxième point est encore vrai car on peut toujours construire une solution au problème des ϕ - moments sur $\mathbb{C}$ à partir de la famille de lois $(F_x)_{x \in U}$ avec $F_x = 2^{-1} (\delta_{a(x)} + \delta_{b(x)})$. En effet, on a dans ce cas

$$\forall \tau \in \mathbb{R}, \forall x \in U, \psi(\tau, x) = \text{Log} (e^{a(x)\tau} + e^{b(x)\tau}) - \text{Log } 2.$$

Utilisant alors l'égalité

$$\forall \alpha \in \mathbb{R}, \forall \beta \in \mathbb{R}, \forall \lambda \in [0,1], \text{Log} (\lambda e^\alpha + (1-\lambda) e^\beta) = \alpha + \text{Log} (\lambda + (1-\lambda) e^{\beta-\alpha}),$$

on en déduit que pour τ de $\mathbb{R}$ et x de U , $|\psi(\tau, x)| \leq \text{Max}(|a(x)\tau|, |b(x)\tau|)$.

Cette inégalité et l'hypothèse d'intégrabilité de

$$a(x) = \sup_{j=1, \dots, k} |\phi_j(x)| \quad \text{et de} \quad b(x) = \sup_{j=1, \dots, k} |\phi_j(x)|,$$

entraînent que $\psi'(\langle v, \phi(x) \rangle, x)$ est pour tout v de $\mathbb{R}^k$ intégrable sur U . On est donc sous les hypothèses du corollaire 3.6 et l'on peut construire une solution au problème des ϕ -moments sur $\mathcal{C}$ à partir de la famille $(F_x)_x \in U$.

IV.2.2 Cas U compact, a, b finis

Nous nous intéressons ici à la construction d'un critère calculable (l'analogue de la condition $T_k(c)$ est positive dans le cas des moments trigonométriques) pour décider si une suite donnée de réels $c = (c_j)_{j=1, \dots, k}$ est ϕ -réalisable dans

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a(x) < e(x) < b(x) \},$$

c'est-à-dire décider s'il existe une fonction e de $cl(\mathcal{C})$ avec

$$\int_U e(x) \phi(x) dx = c.$$

$cl(\mathcal{C})$ est ici la fermeture de $\mathcal{C}$ pour la norme $L^\infty(U, dx)$. Il est naturel de fermer $\mathcal{C}$ de cette façon. En effet, on a²

$$\overline{\mathcal{K}} = \overline{\left\{ \int_U e(x) \phi(x) dx, e \in \mathcal{C} \right\}} = \left\{ \int_U e(x) \phi(x) dx, e \in cl(\mathcal{C}) \right\}.$$

M.G. Krein et A.A. Nudel'man [Krein-Nudel, 77] (théorème 1.2, p.240) montrent par des arguments de convexité le critère suivant:

1- c est ϕ -réalisable dans $\mathcal{C}$ si et seulement si

² L'inclusion de $\overline{\mathcal{K}}$ dans $\left\{ \int_U e(x) \phi(x) dx, e \in cl(\mathcal{C}) \right\}$ est évidente, l'inclusion dans l'autre sens se démontre par exemple par des arguments de séparation identiques à ceux utilisés dans la démonstration du lemme 3.1.

$$\forall v \in \mathbb{R}^k, \langle v, c \rangle \leq b \int_U \langle v, \phi(x) \rangle^+ dx - a \int_U \langle v, \phi(x) \rangle^- dx,$$

où $\langle v, \phi(x) \rangle^+$ (resp. $\langle v, \phi(x) \rangle^-$) désigne la partie positive (resp. négative) de $\langle v, \phi(x) \rangle$.

2- c est ϕ -réalisable non ϕ strictement réalisable dans $\mathcal{C}$ si et seulement si il existe $v_0 \in \mathbb{R}^k$ tel que,

$$\langle v_0, c \rangle = b \int_U \langle v_0, \phi(x) \rangle^+ dx - a \int_U \langle v_0, \phi(x) \rangle^- dx.$$

Dans ce cas et uniquement dans ce cas il existe un unique g de $\mathcal{C}$ qui réalise c . De plus, g est de la forme $g(x) = b 1_{\langle \alpha, \phi(x) \rangle > 0} + a 1_{\langle \alpha, \phi(x) \rangle \leq 0}$.

La construction du maximum d'entropie sur la moyenne donne un autre critère:

Théorème 4.2 Pour c et v éléments de $\mathbb{R}^k$, on pose

$$h(v) = \int_U \text{Log} (e^{a \langle v, \phi(x) \rangle} + e^{b \langle v, \phi(x) \rangle}) dx,$$

$$H(v, c) = h(v) - \langle v, c \rangle \text{ et,}$$

$$h^*(c) = - \inf_{v \in \mathbb{R}^k} H(v, c).$$

On a alors:

1- c est ϕ -réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) \leq 0$.

2- Si c est ϕ strictement réalisable dans $\mathcal{C}$ alors $h^*(c) < 0$.

3- c n'est pas ϕ -réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) = \infty$.

4- On suppose que ϕ est de rang plein presque partout. C'est-à-dire que pour tout α de $\mathbb{R}^k$, la fonction $\langle \alpha, \phi(x) \rangle$ ne s'annule que sur un ensemble négligeable. Alors c est ϕ -réalisable non ϕ strictement réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) = 0$. Dans ce cas le problème des ϕ -moments sur $\mathcal{C}$ possède une unique solution g . g est de la forme

$$g(x) = b 1_{\langle \alpha, \phi(x) \rangle > 0} + a 1_{\langle \alpha, \phi(x) \rangle \leq 0}.$$

Démonstration

Rappelons que nous notons $\mathcal{K}$ l'ensemble $\mathcal{K} = \{ \int_U \phi(x) e(x) dx, e \in \mathcal{C} \}$.

Point 2

Considérons le problème M.E.M pour la mesure de référence $F = \delta_a + \delta_b$, (voir les exemples 3.8, 3.13). Si c est ϕ strictement réalisable dans $\mathcal{C}$, il existe $v_\infty^*(c)$ de $\mathbb{R}^k$, tel que la fonction, $g_\infty^{\text{ME}}(x) = \psi'(\langle v_\infty^*(c), \phi(x) \rangle)$, avec $\psi(\tau) = \text{Log}(e^{a\tau} + e^{b\tau})$, ($\tau \in \mathbb{R}$), vérifie

$$\int_U g_\infty^{\text{ME}}(x) \phi(x) dx = c, \quad \text{et de plus}$$

$$\Gamma(g_\infty^{\text{ME}}) = \int_U \gamma(g_\infty^{\text{ME}}(x)) dx = H(v_\infty^*(c), c) = -h^*(c),$$

où pour y de $]a, b[$, $\gamma(\cdot)$ est la fonction

$$\gamma(y) = -\frac{y-a}{b-a} \text{Log}(y-a) - \frac{b-y}{b-a} \text{Log}(b-y) + \text{Log}(b-a).$$

En particulier, $\gamma(y)$ est strictement positif pour tout y de $]a, b[$. On en déduit que $h^*(c) < 0$.

Points 1 et 4

Soit c ϕ -réalisable non ϕ strictement réalisable dans $\mathcal{C}$, comme h^* est convexe, s.c.i (comme dual de Legendre d'une fonction continue [El, 85], th VI.5.3, p. 221) on a,

$$-\text{Log } 2 |U| \leq \lim_{\substack{c' \rightarrow c \\ c' \in \mathcal{K}}} h^*(c') = h^*(c) \leq 0.$$

Soit $(c_i)_{i \geq 1}$, une suite d'éléments de $\mathcal{K}$ convergeant vers c et soit $(v_i)_{i \geq 1}$ la suite de $\mathbb{R}^k$ définie par $v_i = v_\infty^*(c_i)$, c'est-à-dire que l'on a pour tout i

$$\int_U \psi'(\langle v_i, \phi(x) \rangle) \phi(x) dx = c_i.$$

On a alors, $\lim_{i \rightarrow +\infty} \|v_i\| = +\infty$, car sinon c serait ϕ strictement réalisable. Quitte à extraire une sous-suite, on peut toujours supposer que

$$\lim_{i \rightarrow +\infty} \frac{v_i}{\|v_i\|} = \alpha.$$

On a,

$$-H(v_i, c) = \langle v_i, c \rangle - \int_U \text{Log} (e^{a \langle v_i, \phi(x) \rangle} + e^{b \langle v_i, \phi(x) \rangle}) dx$$

$$\text{et, } \int_U \text{Log} (e^{a \langle v_i, \phi(x) \rangle} + e^{b \langle v_i, \phi(x) \rangle}) dx$$

$$= b \int_U \langle v_i, \phi(x) \rangle^+ dx - a \int_U \langle v_i, \phi(x) \rangle^- dx$$

$$+ \int_U \text{Log} (1 + \exp \{ (b - a) \langle v_i, \phi(x) \rangle \}) 1_{\langle v_i, \phi(x) \rangle \leq 0} dx$$

$$+ \int_U \text{Log} (1 + \exp \{ (a - b) \langle v_i, \phi(x) \rangle \}) 1_{\langle v_i, \phi(x) \rangle > 0} dx$$

En supposant que ϕ est de rang plein presque partout et en utilisant le théorème de Lebesgue, les deux dernières intégrales tendent vers 0 quand n tend vers ∞ . On a,

$$h^*(c) = \lim_{i \rightarrow +\infty} \langle v_i, c_i - b \int_U \phi(x) 1_{\langle v_i, \phi(x) \rangle > 0} dx - a \int_U \phi(x) 1_{\langle v_i, \phi(x) \rangle \leq 0} dx \rangle$$

et puisque $|h^*(c)| < +\infty$ on a donc forcément

$$\langle \alpha, c - b \int_U \phi(x) 1_{\langle \alpha, \phi(x) \rangle > 0} dx - a \int_U \phi(x) 1_{\langle \alpha, \phi(x) \rangle \leq 0} dx \rangle = 0.$$

Prenant maintenant la suite $v_i' = i \alpha$ et en utilisant la décomposition précédente, on obtient finalement $\lim_{i \rightarrow +\infty} H(v_i', c) = 0$ et donc $h^*(c) = 0$.

Montrons maintenant la dernière assertion. On a,

$$c_i = \int_U \phi(x) \frac{a + b \exp \{ (b - a) \langle v_i, \phi(x) \rangle \}}{1 + \exp \{ (b - a) \langle v_i, \phi(x) \rangle \}} 1_{\langle \alpha, \phi(x) \rangle \leq 0} dx$$

$$+ \int_U \phi(x) \frac{b + a \exp \{ (a - b) \langle v_i, \phi(x) \rangle \}}{1 + \exp \{ (a - b) \langle v_i, \phi(x) \rangle \}} 1_{\langle \alpha, \phi(x) \rangle > 0} dx.$$

Prenant la limite de cette égalité, et utilisant le théorème de Lebesgue, on

trouve que
$$c = b \int_U \phi(x) 1_{(\langle \alpha, \phi(x) \rangle > 0)} dx + a \int_U \phi(x) 1_{(\langle \alpha, \phi(x) \rangle \leq 0)} dx .$$

La fonction, $g_{\infty}^{ME}(x) = b 1_{(\langle \alpha, \phi(x) \rangle > 0)} + a 1_{(\langle \alpha, \phi(x) \rangle \leq 0)}$ réalise donc c .

Point 3.

Le théorème VI.5.7 de [El, 85] (p. 224) permet de dire que l'intérieur du domaine de la fonction h^* est inclus dans l'ensemble,

$$\left\{ \int_U \psi'(\langle v, \phi(x) \rangle) \phi(x) dx, v \in \mathbb{R}^k \right\}$$

qui est égal à $\mathcal{K}$ d'après la construction M.E.M. En combinant ce résultat aux résultats précédents, on en déduit que le domaine de la fonction h^* coïncide avec l'adhérence du convexe $\mathcal{K}$. Cela entraîne que $h^*(c) = +\infty$ si c n'est pas ϕ -réalisable dans $\mathcal{C}$.

Remarques

Remarque 4.1 Grandes déviations

Nous reprenons ici les notations du paragraphe III.4 et II.4.1 La fonction h^* du théorème précédent est la fonction d'entropie (à une constante additive près), pour les grandes déviations de la suite des lois des vecteurs aléatoires (de $\mathbb{R}^k$) $X_n = \phi_n \omega_n$, quand les composantes du vecteur aléatoire ω_n (de $\mathbb{R}^n$) suivent des lois de Bernouilli indépendantes, (voir le corollaire 3.5 et l'exemple 3.14). Le fait que $h^*(c)$ vaille l'infini pour c non ϕ strictement réalisable dans $\mathcal{C}$, exprime simplement que la probabilité de l'évènement $\{X_n = c\}$ est nulle pour n assez grand. Ceci peut être considéré comme une réciproque du lemme 3.1.

Remarque 4.2 Extensions

Le théorème précédent est encore vrai si l'on remplace a et b par des fonctions continues $a(x)$ et $b(x)$, mais il ne peut pas être étendu au cas où U n'est pas compact, car il faut renormaliser la mesure de référence pour qu'une reconstruction par M.E.M existe, et donc l'inégalité

$$\forall y \in]a, b[, \quad 0 \leq \gamma(y) < \text{Log } 2,$$

est à remplacer par l'inégalité

$$\forall y \in]a(x), b(x)[, \forall x \in U, \quad -\text{Log } 2 \leq \gamma(y, x) < 0,$$

on peut donc seulement dire que la fonction $h^*(c)$ vaut l'infini si c n'est pas ϕ -réalisable dans $\mathcal{C}$.

Remarque 4.3 Le critère de M.G. Krein et A.A. Nudel'man³

Puisqu'on a toujours pour τ de $\mathbb{R}$, $\text{Log}(1 + e^\tau) > 0$, l'inégalité du critère donné par M.G. Krein et A.A. Nudel'man [Krein-Nudel, 77] (théorème 1.2, p. 240) est plus fine que celle du théorème précédent. Mais le théorème donne un critère numériquement calculable car la fonction $H(\cdot, c)$ est strictement convexe. Ce n'est pas le cas de la fonction

$$G(v, c) = b \int_U \langle v, \phi(x) \rangle^+ dx - a \int_U \langle v, \phi(x) \rangle^- dx - \langle v, c \rangle,$$

qui est convexe mais n'est pas strictement convexe.

Remarque 4.4 ϕ - moments relaxés

Si l'on s'intéresse au problème des ϕ - moments relaxés : trouver g tel que

$$g \in \mathcal{C} \quad \text{et} \quad \int_U g(x) \phi(x) dx \in K,$$

où K est un convexe compact de $\mathbb{R}^k$ donné, le théorème précédent reste vrai en remplaçant $H(v, c)$ par

$$H(v, K) = \int_U \text{Log} (e^{a \langle v, \phi(x) \rangle} + e^{b \langle v, \phi(x) \rangle}) dx - \inf_{c \in K} \langle v, c \rangle.$$

Pour démontrer cela on procède comme pour la démonstration du corollaire 2.2.

Remarque 4.5 Décidabilité numérique du problème des ϕ moments

D'un point de vue numérique, l'assertion 3 du théorème précédent (c n'est pas ϕ -réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) = \infty$) est intéressante car il suffit d'exhiber un vecteur v de $\mathbb{R}^k$ qui vérifie $H(v, c) < 0$, pour conclure que c n'est pas ϕ -réalisable dans $\mathcal{C}$. On peut par exemple envisager un algorithme de Monte-Carlo pour tester cette condition.

Remarque 4.6 Le rôle de $\delta_a + \delta_b$

Soit F_2 une mesure positive et bornée sur $[a, b]$ ($-\infty < a < b < +\infty$), notons $F_1 = p \delta_a + \delta_b$. ($p > 0$ donné). Soient τ de $\mathbb{R}$ et λ de $]0, 1]$, on a

³Voir aussi la remarque 4.8.

$$\begin{aligned} & \text{Log} \int \exp(\tau y) (\lambda dF_1(y) + (1 - \lambda) dF_2(y)) \\ &= \text{Log} \left(\lambda \int \exp(\tau y) dF_1(y) \right) + \text{Log} \left(1 + \frac{1 - \lambda}{\lambda} \frac{\int \exp(\tau y) dF_2(y)}{\int \exp(\tau y) dF_1(y)} \right). \end{aligned}$$

Et donc

$$\text{Log} \left(\lambda \int \exp(\tau y) dF_1(y) \right) \leq \text{Log} \int \exp(\tau y) (\lambda dF_1(y) + (1 - \lambda) dF_2(y)).$$

Soient F une mesure positive et bornée sur $[a, b]$ et y appartenant à l'intérieur de l'enveloppe convexe du support de F , on pose

$$\gamma_F(y) = - \sup_{\tau \in \mathbb{R}} (\tau y - \text{Log} \int \exp(\tau t) dF(t)).$$

D'après les inégalités précédentes on a

$$\forall y \in]a, b[, \gamma_{\lambda F_1}(y) \leq \gamma_{\lambda F_1 + (1 - \lambda) F_2}(y), (\lambda \in]0, 1]).$$

Donc, puisque $\gamma_{\lambda F_1}(\cdot)$ est finie sur $[a, b]$, il en est de même de $\gamma_{\lambda F_1 + (1 - \lambda) F_2}(\cdot)$.

Réciproquement, soit F une mesure bornée d'enveloppe convexe du support $[a, b]$, supposons que $|\gamma_F(a)|$ et $|\gamma_F(b)|$ soient finis. On a pour y de $]a, b[$

$$\gamma_F(y) = \psi(\tau) - \tau \psi'(\tau), \text{ avec } \psi(\tau) = \text{Log} \int \exp(\tau t) dF(t) \text{ et } \psi'(\tau) = y.$$

Donc, on a par exemple $\gamma_F(a) = \lim_{\tau \rightarrow -\infty} (\psi(\tau) - \tau \psi'(\tau))$, et

$$\exp\{\gamma_F(y)\} = \int \exp\{\tau t - \tau \psi'(\tau)\} dF(t).$$

Puisque $\lim_{\tau \rightarrow -\infty} \exp\{\gamma_F(y)\}$ est non nulle F charge le point a . En conclusion une mesure F , positive bornée sur $[a, b]$, vérifie $|\gamma_F(a)| < +\infty$ et $|\gamma_F(b)| < +\infty$, si et seulement si elle s'écrit

$$F = \lambda F_1 + (1 - \lambda) F_2, \quad (\lambda \in]0, 1], F_2 \in \mathcal{M}([a, b]), F_2([a, b]) < +\infty).$$

On peut donc énoncer le théorème précédent en utilisant n'importe quelle mesure du type $F = \lambda F_1 + (1 - \lambda) F_2$, $\lambda \in]0,1[$, avec $F_1 = \delta_a + \delta_b$, ($p=1$, pour que $\gamma_F(a) = \gamma_F(b)$) et $F_2(a) = F_2(b) = 0$.

Quelle est la signification de cette propriété pour le M.E.M ?

Soit F une mesure bornée d'enveloppe convexe fermée du support $[a,b]$ et soit c de $\mathbb{R}^k$ ϕ -réalisable, non ϕ strictement réalisable dans $\mathcal{C}$. Alors il existe pour tout $n > n_0$, une loi P_n absolument continue par rapport à $F^{\otimes n}$, qui vérifie $\phi_n E_{P_n}(\omega_n) = c$, si et seulement si F s'écrit

$$F = \lambda F_1 + (1 - \lambda) F_2, \quad (\lambda \in]0,1[, F_2 \in \mathcal{M}([a,b]), F_2([a,b]) < +\infty).$$

Lorsque le convexe

$$\bar{\mathcal{L}}(c) = \{ e \in \text{cl}(\mathcal{C}), \int_U e(x) \phi(x) dx = c \},$$

n'est pas réduit à un unique point (c'est-à-dire d'après le théorème précédent lorsque c est ϕ strictement réalisable dans $\mathcal{C}$) on peut s'intéresser à la caractérisation de ses points extrémaux. Rappelons qu'un élément e d'un convexe ξ est extrémal s'il vérifie la propriété

$$(e = (1 - \lambda) e_0 + \lambda e_1 \text{ avec } \lambda \in]0,1[, e_0, e_1 \in \xi) \Rightarrow (e_0 = e_1 = e)$$

Lorsque le convexe ξ est compact il coïncide avec l'enveloppe convexe fermée de ses points extrémaux (théorème de Krein-Milman [Brez, 83], th I.12 p. 13). Les points extrémaux suffisent alors pour caractériser tous les éléments du convexe ξ . Dans l'exemple du problème des moments trigonométriques, les éléments extrémaux du convexe des solutions sont des combinaisons de masse.

Considérons maintenant la fonction $\tilde{h}^*$ sur l'ensemble $\mathbb{R}^k \times \mathcal{B}(U)$ (où $\mathcal{B}(U)$ est l'ensemble des boréliens de U) à valeurs dans $\mathbb{R}$ définie pour c de $\mathbb{R}^k$ et A de $\mathcal{B}(U)$ par,

$$\begin{aligned} \tilde{h}^*(c,A) = \sup_{v \in \mathbb{R}^k} \{ \langle v, c \rangle - \int_A \text{Log}(e^{a \langle v, \phi(x) \rangle} + e^{b \langle v, \phi(x) \rangle}) dx \\ - a \int_{U/A} \langle v, \phi(x) \rangle dx \}. \end{aligned}$$

Cette fonction permet de caractériser les points extrémaux de $\bar{\mathbb{L}}(c)$ lorsqu'il existe ξ de $\mathbb{R}^k$ tel que $\langle \xi, \phi(\cdot) \rangle$ soit une fonction strictement positive presque partout sur U , et que ϕ est de rang plein presque partout:

Théorème 4.3

1- Les points extrémaux du convexe $\bar{\mathbb{L}}(c)$ sont du type $b 1_A + a 1_{U/A}$

$$\text{avec,} \quad c = b \int_A \phi(x) \, dx + a \int_{U/A} \phi(x) \, dx .$$

2- Si $b 1_A + a 1_{U/A}$ est un point extrémal de $\bar{\mathbb{L}}(c)$, alors $\tilde{h}^*(c, A) = 0$ et il existe un élément α de $\mathbb{R}^k$ avec $\langle \alpha, \phi(x) \rangle > 0$ p.p sur A .

3- Si $\tilde{h}^*(c, A) = 0$ alors il existe un élément α de $\mathbb{R}^k$ tel que la fonction

$$b 1_A 1_{\langle \alpha, \phi(x) \rangle > 0} + a 1_{U/A} 1_{\langle \alpha, \phi(x) \rangle \leq 0} ,$$

soit un point extrémal de $\bar{\mathbb{L}}(c)$.

Démonstration

Point 1

Puisque pour tout A mesurable inclus dans U , la fonction $b 1_A + a 1_{U/A}$, est uniquement décomposable (décomposition en combinaison convexe) en $\lambda (b 1_A + a 1_{U/A}) + (1 - \lambda) (b 1_A + a 1_{U/A})$, avec λ appartenant à $[0, 1]$, si elle vérifie

$$c = b \int_A \phi(x) \, dx + a \int_{U/A} \phi(x) \, dx, \text{ elle est extrémale.}$$

Réciproquement, soit g un point extrémal de $\bar{\mathbb{L}}(c)$, puisqu'il ne peut pas appartenir à l'intérieur (dans $L^\infty(U, dx)$) de $cl(\bar{\mathbb{C}})$, g s'écrit $g = g_1 1_{U/A} 1_{U/B} + b 1_B + a 1_A$, où A et B sont des ensembles mesurables inclus dans U avec A ou B de mesure (de Lebesgue) non nulle. On peut en plus supposer que A et B sont de mesure (de Lebesgue) maximale. Posons alors

$$c' = \int_{U'} g_1(x) \phi(x) \, dx \quad \text{et} \quad U' = U/(A \cup B).$$

On conclut en remarquant que, compte-tenu de l'hypothèse de maximalité de mesure sur A et B , g_1 ne peut pas être un point extrémal du convexe

$\bar{\mathbb{L}}'(c') = \{ e \in \mathcal{cl}(\mathbb{C}), \int_{U'} e(x) \phi(x) dx = c' \}$ et donc $g_1 = 0$, c'est-à-dire que les ensembles U/A et U/B sont tous les deux de mesure de Lebesgue nulle.

Point 2

Si $b 1_A + a 1_{U/A}$ appartient à $\bar{\mathbb{L}}(c)^4$, on a

$$c = b \int_A \phi(x) dx + a \int_{U/A} \phi(x) dx.$$

$$\begin{aligned} \text{Posons pour } v \text{ de } \mathbb{R}^k, \tilde{H}(v, c, A) &= \int_A \text{Log}(e^{a \langle v, \phi(x) \rangle} + e^{b \langle v, \phi(x) \rangle}) dx \\ &+ a \int_{U/A} \langle v, \phi(x) \rangle dx - \langle v, c \rangle \\ &= \int_A \text{Log}(1 + e^{(a-b) \langle v, \phi(x) \rangle}) dx. \end{aligned}$$

Supposons que $\tilde{H}(\cdot, c, A)$ atteigne son minimum en un point v^* alors,

$$\text{grad } \tilde{H}(v^*, c, A) = 0 \Rightarrow \int_A (a - b) \phi(x) \frac{1}{1 + e^{(b-a) \langle v^*, \phi(x) \rangle}} dx = 0.$$

Ce qui est impossible puisque $\langle \xi, \phi(\cdot) \rangle^5$ est une fonction strictement positive sur U . Le minimum n'est donc pas atteint et il existe une suite $(v_i)_{i \geq 1}$ de $\mathbb{R}^k$ avec,

$$\lim_{i \rightarrow +\infty} \|v_i\| = +\infty,$$

$$\lim_{i \rightarrow +\infty} \frac{v_i}{\|v_i\|} = \alpha,$$

et

$$\lim_{i \rightarrow +\infty} \tilde{H}(v_i, c, A) = -\tilde{h}^*(c, A).$$

Posons $\alpha_i = \frac{v_i}{\|v_i\|}$, on a alors

⁴On suppose que $A \neq \emptyset$, le cas $A = \emptyset$ est trivial.

⁵Voir la ligne qui précède l'énoncé du théorème 4.3.

$$\begin{aligned} \tilde{H}(v_i, c, A) = & \int_A \text{Log}(1 + e^{(a-b)\langle v_i, \phi(x) \rangle}) 1_{\langle \alpha_i, \phi(x) \rangle > 0} dx \\ & + \int_A \text{Log}(1 + e^{(a-b)\langle v_i, \phi(x) \rangle}) 1_{\langle \alpha_i, \phi(x) \rangle \leq 0} dx, \end{aligned}$$

puisque l'on a l'inéquation, $-\tilde{h}^*(c, A) \leq \int_A \text{Log } 2 \, dx = \tilde{H}(0, c, A)$,

l'ensemble: $A \cap \{\langle \alpha, \phi(x) \rangle \leq 0\}$ est négligeable, et donc $\lim_{i \rightarrow +\infty} \tilde{H}(v_i, c, A) = 0$, c'est-à-dire $\tilde{h}^*(c, A) = 0$.

Point 3

On utilise le point 4- du théorème 4.2 avec $\bar{A}$ comme ensemble de définition des fonctions et

$$c = \int_{U/A} \phi(x) \, dx \text{ comme } \phi\text{-moments}.$$

Remarque 4.7 Extension

Le théorème précédent reste vrai si l'on remplace a et b par des fonctions $a(\cdot)$ et $b(\cdot)$ continues sur U .

IV.3 Problème des ϕ - moments sur un polytope

Soit A un polytope borné convexe, ouvert non vide de $\mathbb{R}^r$, ce qui signifie que A est l'intérieur de l'enveloppe convexe d'un nombre fini de points qui ne se trouvent pas tous dans un même hyperplan. Pour c de $\mathbb{R}^k$ et $(\Omega, \mathcal{A}, P)$ un espace probabilisé satisfaisant les hypothèses du paragraphe III.4.3, on se propose d'étudier le problème des ϕ - moments sur $\bar{C}$ suivant:

trouver g de $\text{cl}(\bar{C}) = \{e \in \{L^\infty(\Omega, \mathcal{A}, P)\}^r, e \in \bar{A}, P\text{-p.s.}\}$,

$$\text{avec } E_P \{ \phi g \} = c, \quad \text{où } \phi = (\phi_{ij})_{i=1..k, j=1..r}.$$

On a alors l'analogie du théorème 4.2:

Corollaire 4.1 Notons $y_1, \dots, y_m$ les sommets de A ,

$$H(v, c) = \mathbb{E}_p \left\{ \log \sum_{j=1}^m \exp(v^T \phi y_j) \right\} - \langle v, c \rangle \quad \text{et}$$

$$h^*(c) = - \inf_{v \in \mathbb{R}^k} H(v, c).$$

Alors :

1- c est ϕ -réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) \leq 0$.

2- Si c est ϕ strictement réalisable dans $\mathcal{C}$ alors $h^*(c) < 0$.

3- c n'est pas ϕ -réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) = \infty$.

4- On suppose que ϕ vérifie,

$$\forall \alpha \in \mathbb{R}^k, \alpha \neq 0, P(\alpha^T \phi(y_i - y_j) \neq 0) = 1, \forall i, j, i \neq j.$$

Alors c est ϕ -réalisable non ϕ strictement réalisable dans $\mathcal{C}$ si et seulement si $h^*(c) = 0$. De plus, il existe un unique g de $\mathcal{C}$ réalisant c . g est de la forme

$$g = \sum_{j=1}^m y_j \mathbb{1}_{\{\alpha^T \phi y_j > \alpha^T \phi y_i, i \neq j\}}.$$

Démonstration

Nous reprenons ici les notations de III.4.3.

Point 2

On considère le problème MEM pour la mesure de référence

$$F(y) = \sum_{j=1}^m \delta_{y_j}(y). \text{ On a alors}$$

$$\text{pour } \tau \text{ de } \mathbb{R}^F, \psi(\tau) = \log \sum_{j=1}^m \exp\{\langle \tau, y_j \rangle\}$$

$$\text{et pour } y \text{ de } A, \gamma(y) = -\langle y, \psi^{-1}(y) \rangle + \psi(\psi^{-1}(y)).$$

Faisons le changement de variable $\tau = \psi^{-1}(y)$, on obtient

$$\begin{aligned} \gamma(y) &= -\langle \tau, \psi(\tau) \rangle + \psi(\tau) = -\langle \tau, \mathbb{E}_\tau \{ Y \} \rangle + \sup_{i=1..m} \langle \tau, y_i \rangle \\ &\quad + \log \sum_{j=1}^m \exp\{\langle \tau, y_j \rangle - \sup_{i=1..m} \langle \tau, y_i \rangle\}, \end{aligned}$$

où P_τ désigne la probabilité sur A définie par,

$$dP_\tau = \frac{\sum_{j=1}^m \exp(\tau \cdot y_j) \delta_{y_j}(Y)}{\sum_{j=1}^m \exp(\tau \cdot y_j)},$$

et $Y = (Y_1, \dots, Y_1)$ désigne le vecteur coordonnées (voir l'exemple 2.3).

Remarquant que $(\tau \cdot \text{Ep}_\tau\{Y\}) - \sup_{i=1, \dots, m} (\tau \cdot y_i) < 0$, car $\text{Ep}_\tau\{Y\}$ appartient à l'intérieur de l'enveloppe convexe des points $\{y_1, \dots, y_m\}$, on en déduit que pour tout y de A , $\gamma(y) > 0$. On procède ensuite de la même façon que dans la démonstration du point 2 du théorème 4.2.

Point 1.3 et 4

Identique à la démonstration du théorème 4.2.

Remarque 4.8 Le critère de M.G. Krein et A.A. Nudel'man

Sans supposer que A est un polytope M.G. Krein et A.A. Nudel'man [Krein-Nud, 77] (théorème 7.1, p. 279) donnent le critère suivant: c est ϕ réalisable dans $\mathcal{C}$ si et seulement si

$$\forall v \in \mathbb{R}^k, \langle v, c \rangle \leq \text{Ep} \left\{ \sup_{y \in A} v^T \phi y \right\}.$$

On voit donc que, compte-tenu de la remarque 4.3, le corollaire précédent est comme le théorème 4.2 une version "convexifiée" de ce critère. Dans le cadre plus général où A est un compact (mais pas nécessairement un polytope), on peut obtenir le critère de M.G. Krein et A.A. Nudel'man à partir du M.E.M. On considère une approximation polygonale de $\mathcal{F}(A)$ (la frontière de A), c'est-à-dire que l'on se donne une suite $(y_i)_{i \geq 1}$ de points de $\mathcal{F}(A)$, dense dans $\mathcal{F}(A)$. A l'étape m on considère le polytope A_m engendré par les points $y_1, \dots, y_m$. on construit le critère donné par le corollaire 4.1: c est ϕ réalisable dans $\mathcal{C}_m$ si et seulement si

$$\forall v \in \mathbb{R}^k, \langle v, c \rangle \leq \text{Ep} \left\{ \log \sum_{j=1}^m \exp(v^T \phi y_j) \right\},$$

où l'on a posé, $cl(\mathcal{C}_m) = \{ e \in \{ L^\infty(\Omega, \mathcal{A}, P) \}^r, e \in \bar{A}_m, P\text{-p.s.} \}$. Multipliant l'inégalité précédente par m^{-1} et faisant le changement de variable $u = m^{-1} v$, le critère se réécrit: c est ϕ réalisable dans $\mathcal{C}_m$ si et seulement si

$$\forall u \in \mathbb{R}^k, \langle u, c \rangle \leq E_P \left\{ \frac{1}{m} \text{Log} \sum_{j=1}^m \exp(m u^T \phi y_j) \right\}.$$

Quand m tend vers l'infini on obtient (à l'aide du théorème de Lebesgue) le critère de M.G. Krein et A.A. Nudel'man. Cela revient à considérer la méthode M.E.M avec comme mesure de référence *la mesure de Lebesgue sur $\mathcal{F}r(A)$ de masse infinie*. Sauf dans le cas particulier où A est un polytope, il n'existe pas de mesure F positive bornée sur A telle que la fonction

$$\gamma_F(y) = - \sup_{t \in \mathbb{R}^r} \left(\langle t, y \rangle - \text{Log} \int \exp(t.t) dF(t) \right)$$

soit finie sur toute la frontière de A . Démontrons cette propriété, soit y de $\mathcal{F}r(A)$ posons $\mathcal{P}(y,t) = \{ P \in \mathcal{P}(A), \int t dP(t) = y \}$. Puisque A est convexe il n'y a que deux possibilités:

- a- Il existe y_1 et y_2 de $\mathcal{F}r(A)$, $y \neq y_1$, $y \neq y_2$, tels que y appartienne au segment $[y_1, y_2]$, dans ce cas $\mathcal{P}(y,t)$ contient une infinité d'éléments. $\mathcal{P}(y,t)$ contient toutes les lois sur le segment $[y_1, y_2]$ qui sont d'espérance y .
- b- Pour tout y_1 et y_2 de $\mathcal{F}r(A)$, $y \neq y_1$, $y \neq y_2$, y n'appartient pas au segment $[y_1, y_2]$. Dans ce cas $\mathcal{P}(y,t) = \{ \delta_y \}$.

D'autre part d'après le théorème II.5.1 de [El, 85] (p.42), on a

$$\gamma_F(y) = \sup_{P \in \mathcal{P}(y,t)} S_F(P).$$

On conclut en remarquant que lorsque A n'est pas un polytope, il existe une infinité non dénombrable de points y pour lesquels $\mathcal{P}(y,t) = \{ \delta_y \}$ (cas b). C'est

pour cette raison qu'il faut considérer une mesure de "*masse infinie*" pour construire un critère analogue à celui du corollaire 4.1.

Exemple 4.1 Contrôlabilité d'un système par une commande bornée

Reprenons l'exemple 3.17 du système dynamique,

$$\begin{aligned} X'(\sigma) &= M X(\sigma) + N u(\sigma), \sigma \geq 0 \\ X(0) &= X_0 \neq 0 \end{aligned}$$

On désire ramener le système à 0 à la date T par une commande à valeurs dans $[0,1]^r$. Il existe une telle commande si et seulement si $h^*(X_0) < 0$, où

$$h^*(c) = - \inf_{v \in \mathbb{R}^k} H(v, c),$$

$$H(v, c) = \int_0^T \sum_{j=1}^r \text{Log} \{ 1 + \exp [N^T \exp (- M^T \sigma) v]_j \} d\sigma - \langle v, c \rangle$$

et $[\alpha]_j$ désigne la j-ième composante du vecteur α .

CHAPITRE V

MODELE EXPONENTIEL GENERALISE ET PROBLEMES STATISTIQUES

V.1 INTRODUCTION

La méthode du maximum d'entropie est souvent utilisée en mécanique statistique pour choisir des modèles. En situation classique on construit un modèle à partir d'une mesure de référence μ d'une observable ϕ , variable aléatoire dans $\mathbb{R}^k$ donnant les états "*macroscopiques*" par $E_P(\phi) = c$ ($c \in \mathbb{R}^k$) si P est la loi du modèle défini sur l'espace de probabilité des états "*microscopiques*" (tout comme ϕ). Le choix de P se fait en maximisant la μ -entropie $S_\mu(P)$ sous la contrainte $E_P(\phi) = c$. Si le problème a une solution lorsque c décrit l'ensemble des valeurs possibles, on obtient le modèle défini par les densités (voir le paragraphe II.3)

$$\frac{dP^{ME}}{d\mu} = \{Z_{\mu, \phi}[t^*(c)]\}^{-1} \exp(\langle t^*(c), \phi \rangle) d\mu$$

$t^*(c)$ est le paramètre canonique en dualité avec c .

Supposons maintenant nous donner à priori la contrainte supplémentaire du type $\frac{dP}{d\mu} \in \mathcal{C}$, $\mathcal{C}$ ensemble donné de densité. Lorsque $\mathcal{C}$ est une bande nous avons vu que la méthode du maximum d'entropie sur la moyenne donne une famille de modèles du type $\psi'(\langle v^*(c), \phi \rangle)$ ($\phi_1 = c_1 = 1$), où ψ est la Log-transformée de Laplace d'une mesure positive de support convenable. Cette famille de modèles que nous appellerons modèles exponentiels généralisés sera étudiée au paragraphe V.2.

La méthode du maximum d'entropie sur la moyenne permet de reconstruire une densité lorsque l'on ne connaît que les valeurs de ses

ϕ - moments, dans le paragraphe V.3.1 nous nous intéressons aux qualités de la reconstruction quand les valeurs des ϕ - moments sont connues de façon exacte. Nous donnons dans le cas d'une bande du type,

$$\mathcal{C} = \{ e \in C(U), \forall x \in U, a < e(x) < b \}, a \text{ et } b \text{ finis, } U \text{ compact,}$$

des inégalités sommatoires, c'est-à-dire des inégalités quantifiant comment se rapproche la reconstruction de la "vraie" densité quand la dimension de ϕ augmente. Une étude similaire est ensuite faite dans le cas de la reconstruction maximisant l'entropie de Burg. Le paragraphe V.3.2 est consacré à la même étude dans un cadre statistique où les ϕ - moments sont estimés. l'exemple étudié est celui de l'estimation d'une densité de probabilité sur $[0, 2\pi]$ dont on estime les coefficients de Fourier à l'aide d'un n -échantillon. Enfin dans le paragraphe V.3.3 nous donnons, en utilisant la construction M.E.M, une interprétation Bayésienne de la méthode de Burg pour estimer la densité spectrale d'un processus.

V.2 MODELE EXPONENTIEL GENERALISE

V.2.1 Définition

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé possédant les propriétés de discrétisation décrites au paragraphe III.4.3, soit ϕ une variable aléatoire de $\{L^\infty(\Omega, \mathcal{A}, P)\}^k$ de rang k , on supposera qu'aucune des composantes de ϕ n'est P - p.s égale à 1. Soient a et b avec $0 \leq a < b \leq +\infty$. Enfin, soit F une mesure positive sur $[a, b]$ qui vérifie les hypothèses du paragraphe III.4.3:

(L1) L'enveloppe convexe fermée du support de F est $\overline{[a, b]}$.

(L2) $D(F) = \{ \tau \in \mathbb{R}, \psi(\tau) = \text{Log} \int_{[a, b]} \exp(\tau y) dF(y) < \infty \}$ est un ouvert non vide de $\mathbb{R}$.

(L3) L'ensemble $D(F, \phi) = \{ v \in \mathbb{R}^k, E_P \{ |\psi(\langle v, \phi \rangle)| \} < \infty \}$ est non vide et coïncide avec

$$D'(F, \phi) = \{ v \in \mathbb{R}^k, \exists C \text{ fermé inclus dans } D(F) \text{ avec } \langle v, \phi \rangle \in C, P\text{-p.s.} \}.$$

On a alors, la définition suivante:

Définition 5.1 On appelle *F-modèle exponentiel généralisé* pour l'observable ϕ la famille de lois de densité par rapport à P

$$\{\psi(v_0(v) + \langle v, \phi \rangle), v \in \Theta\}, \text{ avec}$$

$$\Theta = \{v \in \mathbb{R}^k, \exists v_0(v) \in \mathbb{R}, \int_{\Omega} \psi(v_0(v) + \langle v, \phi \rangle) dP = 1\}.$$

Si F est la loi de Poisson ($a=0, b=+\infty$) on retrouve bien sur le modèle exponentiel classique dont une étude complète est faite dans [Barn, 78]. Le terme exponentiel généralisé est sans doute abusif. Du modèle exponentiel on a gardé le principe de construction par observable et maximum d'une fonctionnelle liée à l'entropie (paragraphe III.5), et la dualité de paramétrage que nous allons détailler ci-dessous. Le reste des structures (liées à l'indépendance) est perdu.

Lemme 5.1 On suppose que $a < 1 < b$. Alors

$a\text{-}\Theta$ contient $D(F, \phi)$.

$$b\text{-Posons } \tilde{\mathcal{K}} = \{E_P[\phi \psi(v_0(v) + \langle v, \phi \rangle)], v \in \overset{\circ}{\Theta}\}.$$

$\tilde{\mathcal{K}}$ est ouvert et l'application: $\overset{\circ}{\Theta} \longrightarrow \tilde{\mathcal{K}}$

$$v \longrightarrow E_P[\phi \psi(v_0(v) + \langle v, \phi \rangle)]$$

est biunivoque et analytique

Démonstration

Point a

Posons $D(F) =]-\infty, \beta[$, $-\infty < \beta \leq +\infty$, on a alors,

$$\lim_{\tau \rightarrow -\infty} \psi(\tau) = a \text{ et } \lim_{\tau \rightarrow \beta} \psi(\tau) = b.$$

Puisque l'application

$$\mathbb{R} \longrightarrow \mathbb{R}$$

$$v_0 \longrightarrow E_P[\psi(v_0 + \langle v, \phi \rangle)],$$

est continue, le théorème des valeurs intermédiaires donne le résultat.

Point b

$\tilde{\mathcal{K}}$ peut être considéré comme l'intersection de,

$\mathcal{K} = \{ E_p \left[\begin{pmatrix} 1 \\ \phi \end{pmatrix} e \right], e \in \mathbb{C} \}$ avec l'hyperplan de $\mathbb{R}^{k+1}$ $c_0 = 1$. On a vu au chapitre III que $\mathcal{K}$ est ouvert dans $\mathbb{R}^{k+1}$, donc $\tilde{\mathcal{K}}$ est ouvert dans $\mathbb{R}^k$. La biunivocité résulte directement de la construction du MEM (chapitre III), quant à l'analyticité du théorème des fonctions implicites.

V.2.2 Problèmes d'estimation

Soit X une variable aléatoire réelle sur $(\Omega, \mathcal{A})$. Dans ce paragraphe ϕ désignera une fonction mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. On dira que X a la loi $\psi(v_0(v) + \langle v, \phi \rangle)$, (resp. $A^{-1}(v) \psi'(v_0(v) + \langle v, \phi \rangle)$), si $(\Omega, \mathcal{A})$ est muni de la probabilité $\psi(v_0(v) + \langle v, \phi(X) \rangle)P$, (resp. $A^{-1}(v) \psi'(v_0(v) + \langle v, \phi(X) \rangle)P$).

Soit $X_1, \dots, X_n$ un échantillon de loi $\psi(v_0(v^*) + \langle v^*, \phi \rangle)$ où v^* est un élément fixé de $\overset{0}{\Theta}$, la structure statistique la plus intéressante est l'existence

de l'estimateur sans biais $\hat{c}_n = \frac{1}{n} \sum_{i=1}^n \phi(X_i)$ pour estimer

$$c = E_p[\phi \psi(v_0(v^*) + \langle v^*, \phi \rangle)].$$

Dans le cas où b est fini on a

Théorème 5.1

a- L'estimateur des moments $\hat{v}_n$ défini par $\hat{c}_n = E_p[\phi \psi(v_0(\hat{v}_n) + \langle \hat{v}_n, \phi \rangle)]$, est consistant asymptotiquement normal (à la vitesse $\sqrt{n}$), de variance

$$V(v^*) = A^2(v^*) M^{-1}(v^*) N(v^*) M^{-1}(v^*), \text{ avec}$$

$$A(v) = E_p[\psi'(v_0(v) + \langle v, \phi \rangle)],$$

M(v) désigne la matrice de variance covariance de $\phi(X)$ quand X a la loi

$$\frac{1}{A(v)} \psi'(v_0(v) + \langle v, \phi \rangle) \text{ et } N(v) \text{ cette matrice quand } X \text{ a la loi du modèle}$$

b- $(v_0(\hat{v}_n), \hat{v}_n)$ est aussi l'estimateur du minimum du contraste

$$U_n(v_0, v) = E_p[\psi(v_0 + \langle v, \phi \rangle)] - v_0 - \langle v, \hat{c}_n \rangle.$$

¹voir le théorème 5.1

Démonstration

point b

Cela découle directement de la construction M.E.M.

point a

Puisqu'on a supposé que b est fini ψ, ψ', ψ'' sont uniformément bornées par des constantes, le résultat s'obtient donc par application directe du théorème 3.3.11 de [Dac-Duf, 83] (p. 97).

Remarques

Remarque 5.1 Estimateur du maximum de vraisemblance²

Sauf pour les modèles exponentiels l'estimateur du maximum de vraisemblance (e.m.v) ne coïncide pas avec l'estimateur des moments, l'e.m.v est solution de l'équation implicite

$$\sum_{i=1}^n [A^{-1}(v) B(v) + \phi(X_i)] \frac{\psi''(v_0(v) + \langle v, \phi(X_i) \rangle)}{\psi'(v_0(v) + \langle v, \phi(X_i) \rangle)} = 0,$$

avec $B(v) = E_p[\phi \psi''(v_0(v) + \langle v, \phi \rangle)]$.

Remarque 5.2 Extension

Sans supposer que b est fini on peut encore montrer que le théorème précédent est vrai en utilisant le même type d'arguments que ceux utilisés dans la démonstration du théorème 3.1.

V.3 M.E.M ET STATISTIQUE NON PARAMETRIQUE

V.3.1 Le problème sommatoire

Nous nous plaçons ici dans le cadre de la reconstruction d'une fonction par M.E.M dans le cas d'une bande

$$\mathcal{C} = \{e \in C(U), \forall x \in U, a \leq e(x) \leq b\}, a \text{ et } b \text{ finis, } U \text{ compact.}$$

Nous nous intéressons à quantifier la qualité de la reconstruction. Lorsqu'on se place dans le cas où U est l'intervalle $[0, 2\pi]$ et que $\phi(x) = (\exp(ijx))_{j \in \{0, \dots, k\}}$ les qualités de la méthode sommatoire par maximum d'entropie, c'est-à-dire

²Dans [Ag-Bouc, 80], on trouvera d'autres considérations sur L'e.m.v dans un M.E.G.

les qualités de la reconstruction d'une probabilité par maximum d'entropie quand on connaît ses coefficients de Fourier ont été signalées dans [Gas, 86] (voir aussi les simulations en annexe). Dans ce cas, la méthode du maximum d'entropie est très supérieure aux méthodes usuelles pour le "moyen" et si la fonction à reconstruire est de forte concentration. Le théorème suivant localise pour une fonction g de $\mathcal{C}$ la reconstruction par M.E.M qui possède les mêmes ϕ -moments que g .

Théorème 5.2 *Soit F une mesure positive sur $[a,b]$ (a et b des réels finis). Si l'hypothèse (K1) est vérifiée et si F est de masse totale finie alors on a l'inégalité sommatoire:*

$$\forall v \in \mathbb{R}^k, \|g - g_{\infty}^{ME}\|_{L^1(U)} \leq m \{ 2|U| [m \|\langle v, \phi \rangle - \psi^{-1}(g)\|_{L^1(U)} + \|g\|_{L^q(U)} \|\langle v, \phi \rangle - \psi^{-1}(g)\|_{L^p(U)}] \}^{1/2},$$

avec $\frac{1}{p} + \frac{1}{q} = 1$, $m = \max(|a|, |b|)$ et g une fonction continue sur U qui vérifie

$$\forall x \in U, a < g(x) < b \quad \text{et} \quad \int_U g(x) \phi(x) dx = c.$$

Démonstration

Soit c un point de $\mathcal{K}$, intéressons nous au $n^{\text{ème}}$ problème d'entropie sur la moyenne avec les coefficients approchés

$$c^{(n)} = \frac{|U|}{n} \sum_{j=1}^n \phi(x_j) g(x_j)$$

où g est un élément fixé de $\mathcal{C}$ qui vérifie

$$\int_U g(x) \phi(x) dx = c.$$

Soit $\mu_{\omega_n^i, \beta_n}^i$ la $i^{\text{ème}}$ marginale de μ_{ω_n, β_n} (μ_{ω_n, β_n} a été définie dans la démonstration du théorème 3.1), on a,

$$\mu_{\omega_n^i, \beta_n}^i(\omega_n^i) = \exp \{ \psi^{-1}\{g(x_i)\} \omega_n^i - \psi(\psi^{-1}\{g(x_i)\}) \} F,$$

où ω_n^i désigne ici l'élément générique de $[a,b]$.

Soit $P_{n,i}^{ME}$ la $i^{\text{ème}}$ marginale de P_n^{ME} . On a alors, en utilisant l'inégalité classique entre information de Kullback et distance en variation³,

$$\begin{aligned} |U| m \sqrt{2 K(\mu_{\omega_n, \beta_n}^i, P_{n,i}^{ME})} &\geq |U| m \int_{|a,b|} \left| \frac{dP_{n,i}^{ME}}{dF} - \frac{d\mu_{\omega_n, \beta_n}^i}{dF} \right| dF \\ &\geq |U| \left| \psi'(\langle v_n^*(c^{(n)}), \phi(x_i) \rangle) - g(x_i) \right|, i=1, \dots, n. \end{aligned}$$

Par addition on obtient alors,

$$\begin{aligned} \frac{|U| m}{n} \sum_{i=1}^n \sqrt{2 K(\mu_{\omega_n, \beta_n}^i, P_{n,i}^{ME})} \\ \geq \frac{|U|}{n} \sum_{i=1}^n \left| \psi'(\langle v_n^*(c^{(n)}), \phi(x_i) \rangle) - g(x_i) \right| \end{aligned}$$

l'information de Kullback est additive donc (utilisant l'inégalité de Schwartz),

$$\frac{|U| m}{n} \sum_{i=1}^n \sqrt{2 K(\mu_{\omega_n, \beta_n}^i, P_{n,i}^{ME})} \leq m \sqrt{|U|} \sqrt{\frac{2|U|}{n} K(\mu_{\omega_n, \beta_n}, P_n^{ME})}.$$

D'autre part,

$$\lim_{n \rightarrow +\infty} \frac{|U|}{n} \sum_{i=1}^n \left| \psi'(\langle v_n^*(c^{(n)}), \phi(x_i) \rangle) - g(x_i) \right| = \|g - g_\infty^{ME}\|_{L^1(U)}$$

et⁴ $\lim_{n \rightarrow +\infty} \frac{|U|}{n} K(\mu_{\omega_n, \beta_n}, P_n^{ME}) = D_\psi(g, g_\infty^{ME})$, d'où l'inégalité

$$\|g - g_\infty^{ME}\|_{L^1(U)} \leq m \sqrt{|U|} D_\psi(g, g_\infty^{ME}),$$

et d'après le théorème 3.3,

$$\forall v \in \mathbb{R}^k, m \sqrt{|U|} D_\psi(g, g_\infty^{ME}) \leq m \sqrt{|U|} D_\psi(g, \psi'(\langle v, \phi(x) \rangle)).$$

Reste donc à majorer $D_\psi(g, \psi'(\langle v, \phi(x) \rangle))$ pour obtenir le résultat. On a

$$D_\psi(g, \psi'(\langle v, \phi(x) \rangle)) = \int_U g(x) \left| \psi'^{-1}[g(x)] - \langle v, \phi(x) \rangle \right| dx$$

³Voir la démonstration du lemme 2.1.

⁴La Fonctionnelle D_ψ a été définie au paragraphe III.5.3.2.

$$- \int_U [\psi\{\psi^{-1}[g(x)]\} - \psi\langle v, \phi(x) \rangle] dx.$$

La première intégrale est majorée par

$$\|g\|_{L^q(U)} \|\langle v, \phi \rangle - \psi^{-1}(g)\|_{L^p(U)}, \text{ avec } \frac{1}{p} + \frac{1}{q} = 1,$$

quant à la seconde, en utilisant le théorème des accroissements finis, elle est majorée par $m \|\langle v, \phi \rangle - \psi^{-1}(g)\|_{L^1(U)}$.

Remarque 5.3 Taille du convexe des extensions

Reprenons l'exemple de la sommation trigonométrique ($U = [0, 2\pi]$, $\phi(x) = (\exp(ijx))_{j=0, \dots, k}$). Le théorème précédent garantit que lorsque le nombre de coefficients de Fourier utilisés augmente, la reconstruction g_∞^{ME} converge dans $L^1([0, 2\pi], dx)$ vers la fonction à reconstruire pourvu que celle-ci appartienne au convexe $\mathcal{C}$. Cependant, elle ne montre pas la qualité de la sommation lorsque la fonction g est concentrée, bien au contraire dans ce cas les constantes de l'inégalité précédente (qui dépendent de g) sont d'autant plus grandes que la fonction est concentrée. Il semble en fait, dans ce cas, que le point de vue précédent ne soit pas le bon, et qu'il soit plus intéressant de considérer la taille (dans un sens à préciser) du convexe de toutes les extensions (le nombre de coefficients k étant fixé)

$$\mathbb{L}(c) = \{e \in \mathcal{C}(\mathcal{C}), \int_U e(x) \phi(x) dx = c\}.$$

A. Seghier [Seg, 88] propose une mesure de ce convexe basée sur l'entropie lorsque $a=0$ et $b = +\infty$.

Intéressons-nous maintenant au cas de la reconstruction maximisant l'entropie de Burg (voir les exemples 3.10 et 3.21, et le chapitre IV) c'est-à-dire que la mesure de référence F est la mesure de Lebesgue sur $\mathbb{R}^+$ et que le convexe $\mathcal{C}$ est l'ensemble des fonctions continues et strictement positives sur U . A. Seghier [Seg, 87] étudie le comportement de $D_\psi(g, g_\infty^{ME})$ dans le cas du problème des moments trigonométriques. Il montre l'inégalité suivante

$$D_\psi(g, g_\infty^{ME}) \leq$$

$$\left\| \frac{1}{g} \right\|_{L^\infty}^{-2} \inf_{(v_0, v) \in \mathbb{R} \times \mathbb{C}^k} \left\| \frac{1}{g} - v_0 - \sum_{j=1}^k (v_j \exp(ijx) + \bar{v}_j \exp(-ijx)) \right\|_{L^\infty}^2,$$

où g est une fonction continue et strictement positive sur $[0, 2\pi]$ qui vérifie

$$\int_0^{2\pi} \exp(ijx) g(x) dx = c_j, j = 0, \dots, k \quad \text{et}$$

g_{∞}^{ME} est la solution du problème M.E.M pour les coefficients $(c_j)_{j=0, \dots, k}$.

Nous complétons ici cette inégalité sommatoire.

Corollaire 5.1

a- Si ϕ est une fonction continue de U dans $\mathbb{R}^k$ vérifiant, lorsque la mesure de référence F est la mesure de Lebesgue sur $\mathbb{R}^+$, l'hypothèse (K3) du théorème 3.1 ou les hypothèses (K'3), (K4) du corollaire 3.2, alors on a

$$\left\| \frac{g - g_{\infty}^{\text{ME}}}{g + g_{\infty}^{\text{ME}}} \right\|_{L^1(U)} \leq \frac{e}{\sqrt{2}} \sqrt{|U| D_{\Psi}(g, g_{\infty}^{\text{ME}})}$$

où g est un élément de $\mathcal{C}$ qui vérifie $\int_U g(x) \phi(x) dx = c$ et

$$g_{\infty}^{\text{ME}}(x) = - \frac{1}{\langle \nabla_{\infty}^*(c), \phi(x) \rangle}.$$

b- Lorsque U est l'intervalle $[0, 2\pi]$ et que ϕ est le système d'exponentielles $\phi(x) = (\exp(ijx))_{j=0, \dots, k}$ on a l'inégalité sommatoire

$$\forall (\nabla_0, \nabla) \in \mathbb{R} \times \mathbb{C}^k,$$

$$\left\| \frac{g - g_{\infty}^{\text{ME}}}{g + g_{\infty}^{\text{ME}}} \right\|_{L^1(U)} \leq e\sqrt{\pi} \left\| \frac{1}{g} \right\|_{L^{\infty}} \left\| \frac{1}{g} - \nabla_0 - \sum_{j=1}^k (\nabla_j \exp(ijx) + \bar{\nabla}_j \exp(-ijx)) \right\|_{L^{\infty}}.$$

Démonstration

Remarquons tout d'abord que l'on a l'inégalité,

$$\int_0^{+\infty} \left| \theta \exp(-\theta x) - \theta' \exp(-\theta' x) \right| dx \geq \frac{2 \left| \frac{1}{\theta} - \frac{1}{\theta'} \right|}{e \operatorname{Max} \left(\frac{1}{\theta}, \frac{1}{\theta'} \right)},$$

pour θ et θ' éléments de $\mathbb{R}^{+*}$ et $\theta \neq \theta'$. En effet, supposons par exemple que $\theta > \theta'$, on a

$$\begin{aligned}
& \int_0^{+\infty} | \theta \exp(-\theta x) - \theta' \exp(-\theta' x) | dx \\
&= 2 \int_0^{\frac{\theta}{\theta - \theta'}} | \theta \exp(-\theta x) - \theta' \exp(-\theta' x) | dx \\
&= 2 \exp\left(\frac{t}{1-t} \operatorname{Log} t\right) (1-t), \text{ où l'on a posé } t = \frac{\theta'}{\theta}.
\end{aligned}$$

On a d'autre part l'inégalité, $\forall t, 0 < t < 1, \frac{t}{1-t} \operatorname{Log} t \geq -1$, donc

$$\int_0^{+\infty} | \theta \exp(-\theta x) - \theta' \exp(-\theta' x) | dx \geq \frac{2}{e} (1-t) = \frac{2}{e} \frac{\theta'}{\theta} \left(\frac{1}{\theta'} - \frac{1}{\theta} \right).$$

En permutant les rôles de θ et θ' , on obtient l'inégalité annoncée.

Reprenons les notations de la démonstration du théorème 5.2, $\mu_{\omega_n, \beta_n}^i$ est ici la loi exponentielle de paramètre $\frac{1}{g(x_i)}$, quant à $P_{n,i}^{ME}$ c'est la loi exponentielle de paramètre $\langle v_n^*(c^{(n)}), \phi(x_i) \rangle$. En utilisant l'inégalité préliminaire, on en déduit que

$$\sqrt{|U|} \sqrt{\frac{2|U|}{n} K(\mu_{\omega_n, \beta_n}, P_n^{ME})} \geq \frac{2}{e} \frac{|U|}{n} \sum_{i=1}^n \frac{\left| g(x_i) + \frac{1}{\langle v_n^*(c^{(n)}), \phi(x_i) \rangle} \right|}{g(x_i) - \frac{1}{\langle v_n^*(c^{(n)}), \phi(x_i) \rangle}}.$$

Par passage à la limite (de manière identique à la démonstration du théorème 5.2) on obtient l'inégalité annoncée.

V.3.2 Estimation de densité par M.E.M

Supposons vouloir estimer une densité de probabilité sur $[0, 2\pi]$ à partir de la méthode (dite parfois de projection) utilisant les moments trigonométriques empiriques. Soit donc

$$g(x) = \frac{1}{2\pi} \sum_{j \in \mathbb{Z}} c_j \exp(-ijx) \text{ une densité de probabilité sur } [0, 2\pi].$$

Soit $X_1, \dots, X_N$ un N -échantillon de loi g et

$$\hat{c}_j = \frac{1}{N} \sum_{l=1}^N \exp(ijX_l) \text{ les coefficients de Fourier empiriques.}$$

L'estimateur non paramétrique de g est

$$\hat{g}_N(x) = \frac{1}{2\pi} \sum_{|j| \leq J(N)} \hat{c}_j \exp(-ijx),$$

où $J(N)$ est une fonction croissante de N tendant vers l'infini en même temps que N , et à déterminer. Cette estimation correspond à une sommation de type linéaire pour les séries de Fourier. Si maintenant on sait à priori que la densité à estimer appartient au convexe

$$\mathcal{C} = \{e \in C(U), \forall x \in [0, 2\pi], a < e(x) < b\},$$

pour chaque mesure F sur $\overline{[a, b]}$ vérifiant les hypothèses garantissant l'existence d'une solution au problème MEM, il existe un estimateur non paramétrique de g du type

$$\hat{g}_N^{ME}(x) = \psi' \left\{ \sum_{|j| \leq J(N)} v_j^*(\hat{c}) \exp(ijx) \right\}, \text{ où } v_{-j}^*(\hat{c}) = \bar{v}_j^*(\hat{c})$$

et $v^*(\hat{c})$ est l'unique vecteur de $\mathbb{R} \times \mathbb{C}^{2J(N)}$ qui vérifie,

$$\int_0^{2\pi} \exp(ijx) \psi' \left\{ \sum_{|j| \leq J(N)} v_j^*(\hat{c}) \exp(ijx) \right\} dx = \hat{c}_j, j = 0, \dots, J(N).$$

En annexe nous donnons quelques simulations illustrant ce type d'estimation quand F est la loi de Poisson ($b = +\infty$, dx-entropie) et $F = \delta_a + \delta_b$ (entropie de Fermi-Dirac, voir l'exemple 3.22).

Le corollaire suivant montre pour g suffisamment régulière la convergence de la suite $(\hat{g}_N^{ME})$ lorsque b est fini.

Corollaire 5.2 *On suppose que $g \in \mathcal{C} \cap C^4([0, 2\pi])$, $g(0) = g(2\pi)$ et que la mesure de référence F vérifie les hypothèses (K1), (K2) du théorème 3.1. Alors on peut choisir la suite $(J(N))$ de telle façon que la suite $(\hat{g}_N^{ME})$ converge presque sûrement en variation vers g .*

Démonstration

Prenons⁵ $J(N) = N^{1/8}$, alors les suites $(\hat{g}_N)$, $(\hat{g}_N^{\text{ME}})$, $(\hat{g}_N')$ convergent uniformément presque sûrement vers g, g', g'' . D'autre part on a,

$$\|g - \hat{g}_N^{\text{ME}}\|_{L^1([0,2\pi])} \leq \|\hat{g}_N - \hat{g}_N^{\text{ME}}\|_{L^1([0,2\pi])} + \|g - \hat{g}_N\|_{L^1([0,2\pi])}.$$

Reste donc à voir que $\|\hat{g}_N - \hat{g}_N^{\text{ME}}\|_{L^1([0,2\pi])}$ tend vers 0 quand N grandit.

Puisqu'on a convergence uniforme presque sûre de la suite $(\hat{g}_N)$, presque sûrement, pour N assez grand ($N > N_0(\omega)$), $\hat{g}_N$ appartient à $\mathcal{C}$, on a donc d'après le théorème 5.2

$$\|\hat{g}_N - \hat{g}_N^{\text{ME}}\|_{L^1} \leq b \sqrt{2 \left(\|\hat{g}_N\|_{L^\infty} + b \right) \left\| \sum_{|j| > J(N)} [\psi'^{-1}(\hat{g}_N)]_j \exp(-ijx) \right\|_{L^1}}$$

où $[\psi'(\hat{g}_N)]_j$ désigne le $j^{\text{ème}}$ coefficient de Fourier de $\psi'(\hat{g}_N)$.

D'autre part, on a l'inégalité (obtenue à l'aide d'intégrations par parties),

$$\left\| \sum_{|j| > J(N)} [\psi'^{-1}(\hat{g}_N)]_j \exp(-ijx) \right\|_{L^1} \leq \sum_{|j| > J(N)} \frac{1}{j^2} |[\psi'^{-1}(\hat{g}_N)]''|_j.$$

Soit f appartenant à $\mathcal{C} \cap C^2([0,2\pi])$ on a,

$$[\psi'^{-1}(f)]'' = \frac{f'' - \psi''[\psi'^{-1}(f)] - \frac{f'^2}{\psi''[\psi'^{-1}(f)]}}{(\psi''[\psi'^{-1}(f)])^2}.$$

Posant,

$$\alpha(f) = \min \left\{ \psi''[\psi'^{-1}(\inf_{x \in [0,2\pi]} f(x))] , \psi''[\psi'^{-1}(\sup_{x \in [0,2\pi]} f(x))] \right\},$$

on a alors,

⁵Ce résultat nous a été communiqué par A. Mokkadem, sa démonstration utilise des inégalités de moments.

$$\| (\psi^{-1}(f))'' \|_{L^\infty} \leq \frac{\| f'' \|_{L^\infty}}{\alpha(f)} + \frac{(b-a)^3 \| f' \|_{L^\infty}^2}{\alpha^3(f)}.$$

On en déduit l'inégalité,

$$\| (\psi^{-1}(\hat{g}_N))'' \|_j \leq 2\pi \left[\frac{\| \hat{g}_N \|_{L^\infty}}{\alpha(\hat{g}_N)} + \frac{(b-a)^3 \| \hat{g}_N \|_{L^\infty}^2}{\alpha^3(\hat{g}_N)} \right],$$

et

$$\lim_{N \rightarrow +\infty} \left\| \sum_{|j| > J(N)} [\psi^{-1}(\hat{g}_N)]_j \exp(-ijx) \right\|_{L^1} = 0.$$

V.3.3 La méthode de Burg

Nous donnons ici une nouvelle construction de la méthode de Burg pour estimer la densité spectrale d'un processus. Soit $(X_i)_{i \in \mathbb{Z}}$ un processus réel, centré, strictement stationnaire et fortement mélangeant de densité spectrale continue et paire g , on a alors par définition

$$\int_0^{2\pi} g(x) e^{ijx} dx = E\{X_i X_{i+j}\} = r_j.$$

Supposons que le processus précédent soit observé depuis l'instant 1 jusqu'à l'instant n et que l'on désire estimer sa densité spectrale g supposée inconnue. Pour cela on construit le périodogramme

$$I_n(x) = \frac{1}{2\pi n} \left| \sum_{j=1}^n X_j e^{ijx} \right|^2.$$

I_n n'est pas un bon estimateur ponctuel de g puisque sous des hypothèses sur la sommabilité des cumulants d'ordre 2 et 4 du processus $(X_i)_{i \in \mathbb{Z}}$ ([Ros, 85], théorème 3, p.131), $I_n(x)$ tend pour x de $]0, \pi[$, en loi quand n tend vers l'infini vers un bruit blanc de loi $\frac{g(x)}{2} \chi^2(2)$. Par contre le coefficient de Fourier d'ordre j de I_n

$$\hat{r}_j^n = \int_0^{2\pi} I_n(x) e^{ijx} dx = \frac{1}{n} \sum_{l=1}^n X_l X_{l+j},$$

converge vers r_j à la vitesse $\sqrt{n}$. Enfin remarquons que l'on a aussi

$$\lim_{n \rightarrow +\infty} E(I_n(x)) = g(x).$$

Pour estimer g nous avons donc deux types d'information:

a- Information à priori

$$I_n(x) \xrightarrow{\mathcal{L}} \frac{g(x)}{2} \chi^2(2) \text{ (bruit blanc)}$$

$$\begin{matrix} n \longrightarrow \infty \\ \text{et } \lim_{n \rightarrow +\infty} E\{I_n(x)\} = g(x). \end{matrix}$$

b- Observation

$\hat{r}_k^n$ pour $k=0, \dots, n-1$, a les mêmes propriétés d'exhaustivité que $X_1, \dots, X_n$.

Compte-tenu de l'information à priori, il est naturel d'utiliser la méthode du maximum d'entropie sur la moyenne sur

$$\mathcal{C} = \{e \in C(\mathbb{T}), \forall x \in \mathbb{T}, e(x) > 0\},$$

avec comme mesure de référence la loi $\gamma(\alpha, \lambda)$ (loi gamma de paramètres α , λ fixés). En effet, la probabilité minimisant l'information de Kullback par rapport à une loi gamma à espérance fixée est encore une loi gamma (par application directe du corollaire 2.1). Les contraintes linéaires du problème M.E.M sont alors construites à partir du point b- (observation). Finalement l'estimateur obtenu est

$$\hat{g}_n(x) = \frac{-1}{\sum_{j=1}^{n-1} (v_j^* e^{ijx} + \bar{v}_j^* e^{-ijx}) + v_0^*},$$

avec
$$\int_0^{2\pi} \frac{e^{ilx}}{\sum_{j=1}^{n-1} (v_j^* e^{ijx} + \bar{v}_j^* e^{-ijx}) + v_0^*} dx = \hat{r}_l^n, \text{ pour } l = 0 \dots n-1.$$

C'est justement l'estimateur de Burg [Burg, 75], [Far, 84] qui a l'avantage de se calculer rapidement puisque comme nous l'avons vu au paragraphe IV.2 les coefficients v_j^* peuvent être obtenus par la résolution d'un système linéaire.

REFERENCES BIBLIOGRAPHIQUES

- [Ag-Bouc, 80]. **N.L. Aggarwal, B. Bouchon**: Minimum of h-divergence and maximum likelihood estimation, *Jour Inf Sci*, 3, pp 296-307. (1980).
- [Az-Rug, 77]. **R. Azencott, G. Ruget**: Mélanges d'équations différentielles et grands écarts à la loi des grands nombres, *Z. Wahr*, 38, pp 1-54. (1977).
- [Barn, 78]. **O. Barndorff-Nielsen**: Information and exponential families in statistical theory, Wiley. (1978).
- [Bret, 79]. **J. Bretagnolle**: Formule de Chernoff pour les lois empiriques de variables à valeurs dans des espaces généraux, *Grandes déviations et applications statistiques*, Ast 68, pp 33-52, Soc Math Fran. (1979)
- [Brez, 83]. **H. Brezis**: Analyse fonctionnelle, théorie et applications, Masson. (1983)
- [Bric, 84]. **G. Bricogne**: Maximum entropy and the foundations of direct methods, *Acta Cryst*, A40, pp 410-445. (1984).
- [Burg, 75]. **J.P. Burg**: Maximum entropy spectral analysis, Ph D thesis, Dept of Geophysics, Stanford University. (1975).
- [Chov, 61]. **J. Chover**: On normalized entropy and the extensions of a positive definite function, *J Math Mec*, 10, pp 927-945. (1961).
- [Ciar, 82]. **P.G. Ciarlet**: Introduction à l'analyse numérique matricielle et à l'optimisation, Masson. (1982).
- [Cot-Fort-Mal, 83]. **M. Cottrell, J.C. Fort, G. Malgouyres**: Large deviations and rare events in the study of stochastic algorithms, *I.E.E. Trans on Aut Cont*, Vol 28, pp 907-920. (1983).
- [Csis, 67]. **I. Csiszar**: Information type measures of probability distributions and indirect observations, *Studia Sci, Math Hungar*, 2, pp 299-318. (1967).
- [Csis, 75]. **I. Csiszar**: I-divergence geometry of distributions, *Ann Prob*, 3, pp 146-158. (1975).
- [Csis, 84]. **I. Csiszar**: Sanov property, generalized I-projection, and a conditional limit theorem, *Ann Prob*, 12, pp 768-793. (1984)
- [Dac-Duf, 82]. **D. Dacunha-Castelle, M. Duflo**: Probabilités et statistiques, Tome 1, Masson. (1982).
- [Dac-Duf, 83]. **D. Dacunha-Castelle, M. Duflo**: Probabilités et statistiques, Tome 2, Masson. (1983).

- [Dac, 84]. **D. Dacunha-Castelle**: Reconstruction des phases en cristallographie par maximum d'entropie, Seminaire Bourbaki, 628. (1984).
- [Dac-Gam, 89]. **D. Dacunha-Castelle, F. Gamboa**: Maximisation de l'entropie sous contraintes non linéaires, Prépublication, Orsay. (1989).
- [El, 85]. **R.S. Ellis**: Entropy large deviations and statistical mechanics, Springer Verlag. (1985).
- [Far, 84]. **D.R. Farrier**: Extension of maximum entropy spectral analysis, I.E.E Proceedings, Vol 131.(1984).
- [Gam-Gas, 89]. **F. Gamboa, E. Gassiat**: Maximum d'entropie et problème des moments: cas multidimensionnel, Prépublication, Orsay. (1989).
- [Gas, 86]. **E. Gassiat**: Problème sommatoire par maximum d'entropie, Notes C.R.A.S, t.303, Serie I, pp 675-680. (1986).
- [Gih-Skor, 74]. **I.I. Gihman, A.V. Skorohod**: The theory of stochastic processes, Springer Verlag. (1974).
- [Gren-Szeg, 58]. **U. Grenader, G. Szegő**: Toeplitz forms and their applications, University of California Press. (1958).
- [Gull-Skil, 84]. **S.F. Gull, J. Skilling**: Maximum entropy method in image processing, I.E.E Proceedings, Vol 131, pp 646-659. (1984).
- [Gull-Skil, 85]. **S.F. Gull, J. Skilling**: The entropy of an image, Maximum entropy and Bayesian methods in inverse problems, D. Reidel Publishing company, pp 287-301. (1985)
- [Gzyl, 88]. **H.K. Gzyl**: El metodo de maxima entropia, Prepublication, Universidad de los Andes, Merida. (1988). (A paraitre en anglais chez Springer Verlag (1990))
- [Jayn, 57]. **E.T. Jaynes**: Information theory and statistical mechanics, Phys Rev, Vol 106, pp 620-630. (1957).
- [Jayn, 82]. **E.T. Jaynes**: On the rational of maximum entropy methods, Proc of the I.E.E.E 70, pp 939-952. (1982).
- [Krein-Nud, 77]. **M.G. Krein, A.A. Nudel'man**: The Markov moment problem and extremal problems, Am Math Soc, Vol 50. (1977).
- [Land, 87]. **H.J. Landau**: Maximum entropy and the moment problem, Bull Am Math Soc, Vol 16, pp 47-77. (1987).
- [Lip, 86]. **A. Lippman**: A maximum entropy for expert systems, Brown University thesis. (1986).
- [Max, 85]. Maximum entropy and Bayesian methods in inverse problems, D. Reidel Publishing company. (1985)

[Moham, 87]. **A. Mohammad-Djafari**: Synthèse de Fourier multidimensionnelle à maximum d'entropie: Application à la reconstruction tomographique d'images, Thèse d'état, Paris 11.(1987).

[Nav, 85]. **J. Navaza**: On the maximum entropy estimate of electron density function, Act Cryst, A41, pp 232-241. (1985).

[Nav, 86]. **J. Navaza**: The use of non local constraints in maximum entropy electron density reconstruction, Act Cryst, A42, pp 212-222. (1986).

[Rob, 89]. **C. Robert**: An entropy concentration theorem, Prépublication. Grenoble (1989).

[Rock, 70]. **R.T. Rockafellar**: Convex analysis, Princeton University Press. (1970).

[Ros, 85]. **M. Rosenblatt**: Stationary sequences and random fields, Birckuser Boston. (1985).

[Schoenb, 64]. **I.J. Schoenberg**: Spline functions and the problem of graduation, Proc Nat Acad Sci U.S.A, 52, pp 947-950. (1964)

[Seg, 87]. **A. Seghier**: Reconstruction de la densité spectrale par maximum d'entropie cas d-dimensionnel, Notes C.R.A.S, t.305, Serie I, pp 517-520. (1987).

[Seg, 88]. **A. Seghier**: Extension de fonctions de type positif et entropie associée. Cas multidimensionnel, Prépublication, Orsay. (1988).

[Shore-John, 80]. **J.E. Shore, R.W. Johnson**: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, I.E.E.E Trans Inf Theory, Vol 26, pp 26-36. (1980).

[Van-Cov, 81]. **J.M. Van Campenhout, T.M. Cover**: Maximum entropy and conditional probability, I.E.E.E Trans Inf Theory, Vol 27, pp 483-489. (1981).

[Wahba, 83]. **G. Wahba**: Bayesian confidence intervals for the cross-validated smoothing spline, J.R Stat Soc B, 45, pp 133-150. (1983).

[Wragg-Dow, 70]. **A. Wragg, D.C. Dowson**: Fitting continuous probability density functions over $[0, +\infty[$ using information theory ideas, I.E.E.E Trans Inf Theory, Vol 16, pp 226-230. (1970).

[Zh-Os-Har, 87]. **X. Zhuang, E. Ostevold, R.M. Haralick**: A differential equation approach to maximum entropy image reconstruction, I.E.E.E Trans A.S.S.P, Vol 35, pp 208-225. (1987).

ANNEXE

SIMULATIONS: RECONSTRUCTION ET ESTIMATION D'UNE DENSITE

Les simulations portent sur la reconstruction et l'estimation par la méthode du maximum d'entropie sur la moyenne de la densité sur $[0,1]$:

$$g_A(x) = \frac{0,3 \exp\{-A (x - 0,2)^2\}}{C_1(A)} + \frac{0,7 \exp\{-A (x - 0,6)^2\}}{C_2(A)}$$

avec

$$C_1(A) = \int_0^1 \exp\{-A (x - 0,2)^2\} dx$$

$$C_2(A) = \int_0^1 \exp\{-A (x - 0,6)^2\} dx,$$

pour $A=50$ et $A=70$, ce qui correspond à deux concentrations différentes. On utilise pour cela les coefficients de Fourier

$$c_j = \int_0^1 \exp(i2\pi jx) g_A(x) dx, \quad j = 0, \dots, k,$$

dans le cas de la reconstruction, ou leur estimation

$$\hat{c}_j = \frac{1}{N} \sum_{l=1}^N \exp(i2\pi jX_l), \quad j = 0, \dots, J(N),$$

où $X_1, \dots, X_N$ est un N -échantillon de loi g_A dans le cas de l'estimation.

La reconstruction (resp. l'estimation) est de la forme

$$g_{\infty}^{ME}(x) = \psi \left\{ v_0^*(c) + \sum_{j=1}^k (v_j^*(c) \exp(i2\pi jx) + \bar{v}_j^*(c) \exp(-i2\pi jx)) \right\}$$

$$(\text{resp. } \hat{g}_{\infty}^{\text{ME}}(x) = \psi' \left\{ v_0^*(\hat{c}) + \sum_{j=1}^k (v_j^*(\hat{c}) \exp(i2\pi jx) + \bar{v}_j^*(\hat{c}) \exp(-i2\pi jx)) \right\}).$$

Il s'agit de calculer le paramètre $v^*(c)$ (resp. $v^*(\hat{c})$) ce qui est fait en minimisant la fonction $H_{F,\phi}(\cdot, c)$ (resp. $H_{F,\phi}(\cdot, \hat{c})$). Nous avons utilisé le programme de J. Navaza qui réalise la minimisation à l'aide de la méthode de Newton [Ciar, 82] (paragraphe 7.5, p 158), c'est-à-dire que l'on construit une suite de points $(v_n)_{n \geq 1}$ par l'équation de récurrence

$$v_{n+1} = v_n - \alpha_n M^{-1}(v_n) \text{ grad } H_{F,\phi}(v_n, c),$$

où $M(v)$ est le Hessien de la fonction $H_{F,\phi}(\cdot, c)$ au point v et $(\alpha_n)_{n \geq 1}$ est une suite de réels positifs.

a- Reconstruction

Les résultats sont rassemblés dans le tableau 1. Nous avons utilisé trois mesures de référence différentes:

- F est la loi normale centrée réduite (absence de contrainte non linéaire) dans ce cas on retrouve la sommation de Dirichlet et $v_{\infty}^*(c) = \bar{c}$.
- F est la loi de Poisson d'intensité 1, l'information à priori est donc la positivité, c'est la méthode du maximum d'entropie "classique".
- $F = \delta_0 + \delta_b$ avec $b = 1,1 \|g_A\|_{L^{\infty}}$, la contrainte non linéaire est une contrainte de bande.

La suite $(\alpha_n)_{n \geq 1}$ est prise constante. Pour chacune des méthodes on calcule pour diverses valeurs de k les distances L^1, L^2, L^{∞} de g_A à sa reconstruction.

b- Estimation

Les résultats sont regroupés dans les tableaux 2 ($N=100$) et 3 ($N=1000$). Pour le calcul du point $v^*(\hat{c})$ on a utilisé l'initialisation $v_0 = v_{\infty}^*(c)$, où $v_{\infty}^*(c)$ a été préalablement calculé dans la partie reconstruction. La suite $(\alpha_n)_{n \geq 1}$ est prise constante. Pour chacune des méthodes d'estimation on estime l'espérance des distances L^1, L^2, L^{∞} de g_A à son estimation, à l'aide de 100 reconstructions.

TABLEAU 1

k	A	Dirichlet			Entropie			M.E.M		
2	50	0,18	0,20	0,33	0,10	0,13	0,35	0,089	0,10	0,29
2	70	0,31	0,35	0,62	0,11	0,16	0,43	0,11	0,14	0,39
3	50	0,034	0,04	0,15	0,071	0,082	0,15	0,067	0,084	0,21
3	70	0,072	0,086	0,17	0,086	0,10	0,21	0,071	0,094	0,27
4	50	0,011	0,018	0,10	0,046	0,055	0,12	0,055	0,066	0,16
4	70	0,036	0,041	0,074	0,042	0,057	0,15	0,058	0,079	0,23
5	50	0,008	0,016	0,11	0,026	0,033	0,086	0,033	0,04	0,084
5	70	0,007	0,01	0,069	0,024	0,031	0,075	0,027	0,041	0,09
6	50	0,007	0,014	0,11	0,027	0,032	0,084	0,032	0,039	0,086
6	70	0,004	0,008	0,06	0,024	0,029	0,068	0,026	0,039	0,07
7	50	0,007	0,014	0,10	0,019	0,024	0,087	0,02	0,026	0,092
7	70	0,003	0,007	0,058	0,012	0,016	0,047	0,014	0,019	0,052
8	50	0,006	0,013	0,10	0,018	0,023	0,089	0,015	0,020	0,086
8	70	0,003	0,007	0,057	0,012	0,016	0,047	0,014	0,019	0,052
<i>norme</i>		// //1	// //2	// //∞	// //1	// //2	// //∞	// //1	// //2	// //∞

TABLEAU 2. N=100

JON	A	Dirichlet			Entropie			M.E.M		
2	50	0,22	0,27	0,54	0,18	0,24	0,57	0,17	0,22	0,50
2	70	0,33	0,39	0,77	0,18	0,26	0,67	0,17	0,25	0,61
3	50	0,18	0,23	0,45	0,19	0,26	0,61	0,19	0,25	0,58
3	70	0,18	0,23	0,50	0,19	0,27	0,68	0,18	0,25	0,64
4	50	0,2	0,26	0,56	0,21	0,3	0,75	0,22	0,3	0,71
4	70	0,2	0,26	0,57	0,20	0,3	0,80	0,21	0,3	0,77
5	50	0,23	0,30	0,7	0,24	0,33	0,90	0,24	0,33	0,83
5	70	0,22	0,29	0,67	0,23	0,35	0,96	indéterminées¹		
6	50	0,26	0,34	0,83	0,26	0,37	1,0	indéterminées		
6	70	0,24	0,33	0,82	0,25	0,37	1,1	indéterminées		
<i>espérance</i>		// //1	// //2	// //∞	// //1	// //2	// //∞	// //1	// //2	// //∞

¹C'est-à-dire qu'il existe des réalisations telles que $\hat{c}$ ne soit pas ϕ -réalisable dans $\mathcal{C}$.

TABLEAU 3. N=1000

J(N)	A	Dirichlet			Entropie			M.E.M		
2	50	0,18	0,21	0,38	0,11	0,14	0,36	0,098	0,13	0,32
3	50	0,063	0,08	0,19	0,091	0,11	0,25	0,087	0,11	0,29
4	50	0,064	0,083	0,18	0,077	0,1	0,25	0,082	0,11	0,27
5	50	0,072	0,3	0,22	0,078	0,1	0,27	0,08	0,10	0,26
6	50	0,079	0,1	0,25	0,083	0,11	0,30	0,08	0,11	0,29
espérance		// //1	// //2	// //∞	// //1	// //2	// //∞	// //1	// //2	// //∞

Ces simulations bien, qu'incomplètes indiquent l'intérêt pour k (resp. $J(N)$) petit de la méthode de reconstruction (resp. d'estimation) par M.E.M, utilisant un à priori. Lorsque la concentration de la densité à reconstruire (resp. à estimer) est grande ou que le nombre de coefficients de Fourier est grand la minimisation de la fonction $H_{F,\phi}(\cdot, c)$ (resp. $H_{F,\phi}(\cdot, \hat{c})$) est difficile, la suite $(\alpha_n)_{n \geq 1}$ doit alors être choisie avec beaucoup d'attention; dans ce cas il est difficile de faire un grand nombre de simulations pour estimer les espérances des différents risques. Nous renvoyons en ce qui concerne les reconstructions à [Gas,86], [Nav, 85] et [Nav, 86] où des simulations plus complètes sont données.

Evaluer la qualité d'une reconstruction se fait généralement en utilisant une distance choisie arbitrairement. Toutefois, rien ne dit que ce critère soit cohérent avec un critère "de type visuel". Il n'est pas évident ici que les normes L^p soient des distances "pertinentes". Ce problème a déjà été soulevé (dans un cadre maximum d'entropie "classique") par A. Mohammad-Djafari [Moham, 87], qui donne un exemple en tomographie où la norme L^2 s'oppose à un critère visuel. Il pourrait être intéressant de considérer la ψ -divergence (voir le paragraphe III.5.3), associée à la mesure de référence $F = \delta_0 + \delta_b$, puisque dans un sens $h^*(c)$ (voir le théorème 4.2) "mesure la taille" du convexe²

$$\bar{\mathcal{L}}(c) = \left\{ e \in \text{cl}(\mathcal{C}), \int_0^1 e(x) \phi(x) dx = c \right\},$$

où $\text{cl}(\mathcal{C}) = \{ e \in L^\infty([0,1], dx), 0 \leq e(x) \leq b, p.p. \}$.

²Voir aussi la remarque 5.3.

On pourrait pour cela utiliser un algorithme inspiré de l'article de M. Cottrell et al [Cot-Fort-Mal, 83], qui consisterait à commencer par faire une estimation grossière $\tilde{v}_\infty^*(c)$ de $v_\infty^*(c)$ (par exemple à l'aide d'une itération d'un algorithme de gradient), puis d'estimer à l'aide du changement de probabilité $\mu_{\omega_n, \phi_n^T \tilde{v}_\infty^*(c)}$ ³ (la probabilité initiale est $\mu_n = 2^{-n} (\delta_0 + \delta_1)^{\otimes n}$), la probabilité d'un événement du type $A_n = \{ \phi_n \omega_n \in B(c, \varepsilon) \}$ (i.e $\mu_n(A_n)$), où $B(c, \varepsilon)$ est une petite boule centrée en c (ε réel positif petit). On utiliserait alors l'approximation $\mu_n(A_n) \approx \exp\{ n(h^*(c) + \text{Log } 2) \}$. Un algorithme de ce type s'il est efficace (suffisamment rapide et précis) pourrait être intéressant, car il permettrait de résoudre le problème des ϕ -moments sur $\mathbb{C}$ et pourrait par exemple être utilisé dans une boucle dichotomique pour la résolution de problèmes de programmation linéaire de grande dimension.

D'un point de vue algorithmique pour le même type de reconstruction, on trouvera dans [Moham, 87] l'étude d'un algorithme de gradient conjugué dans l'espace primal, une solution algorithmique originale pour des problèmes de déconvolution d'image par la méthode du maximum d'entropie est d'autre part donnée dans [Zh-Os-Har, 87], enfin dans [Lip, 86] on trouvera pour le cas d'un lattice un algorithme stochastique pour calculer le minimum de $H_{F, \phi}(\cdot, c)$.

³Les notations employées sont celles du chapitre III. n est un entier assez grand fixé.

TABLE DES MATIERES

Abstract.....	1
 I INTRODUCTION ET PLAN DETAILLE DE LA THESE.....	 3
I.1 Introduction.....	3
I.2 Plan détaillé de la thèse.....	8
 II LA METHODE DU MAXIMUM D'ENTROPIE.....	 15
II.1 Introduction.....	15
II.2 Le principe du maximum d'entropie.....	15
II.3 Minimisation de l'information de Kullback et maximisation de l'entropie.....	19
II.3.1 Cas de contraintes linéaires.....	20
II.3.2 Généralisation à des contraintes linéaires relaxées.....	27
II.4 Autres constructions où la distribution du maximum d'entropie intervient.....	31
II.4.1 Grandes déviations.....	31
II.4.2 Maximum d'entropie et probabilités conditionnelles.....	36
 III LA METHODE DU MAXIMUM D'ENTROPIE SUR LA MOYENNE.....	 38
III.1 Introduction.....	38
III.2 Exemples de reconstruction.....	40
III.2.1 Moments généralisés.....	40
III.2.2 Application: interpolation par fonctions splines.....	42

III.3 La suite des problèmes M.E.M pour un convexe d'un espace fonctionnel.....	44
III.3.1 Définition du problème.....	44
III.3.2 Résolution du n-ième problème d'entropie.....	48
III.4 Convergence de la suite de problèmes M.E.M.....	52
III.4.1 Boule unité d'un espace de Hilbert séparable.....	53
III.4.2 Bande ouverte d'un espace de fonctions continues.....	56
III.4.2.a Cas compact.....	57
III.4.2.a.1 Bande fixe	
III.4.2.a.2 Bande variable	
III.4.2.b Extension au cas non compact.....	71
III.4.3 Tube de variables aléatoires de L^∞	75
III.5 Problème M.E.M et maximisation d'un critère concave.....	80
III.5.1 Préambule.....	80
III.5.2 Maximisation de la fonctionnelle de Legendre.....	82
III.5.2.a Bande de $C(U)$	82
III.5.2.b Tube de L^∞	85
III.5.3 ψ - divergence et théorème de Pythagore.....	86
III.5.3.a Information de Kullback et théorème de Pythagore.....	86
III.5.3.b Cas de la ψ - divergence.....	87
IV LE PROBLEME DES MOMENTS GENERALISES.....	90
IV.1 Introduction.....	90
IV.2 Cas d'une bande de $C(U)$.....	92
IV.2.1 Cas général.....	92
IV.2.2 Cas U compact, bande bornée.....	94
IV.3 Problème des moments sur un polytope.....	104

V MODELE EXPONENTIEL GENERALISE ET PROBLEMES STATISTIQUES.....	109
V.1 Introduction.....	109
V.2 Modèle exponentiel généralisé.....	110
V.2.1 Définition.....	110
V.2.2 Problèmes d'estimation	112
V.3 M.E.M et statistique non paramétrique.....	113
V.3.1 Le problème sommatoire.....	113
V.3.2 Estimation de densité par M.E.M.....	118
V.3.3 La méthode de Burg.....	121
 REFERENCES BIBLIOGRAPHIQUES.....	 123
 ANNEXE.....	 126