

SOMMAIRE DU N° 114

SMF	
Mot du Président	3
MATHÉMATIQUES	
Un point de vue non-asymptotique pour la sélection de modèle, <i>P. Massart</i>	5
HISTOIRE	
J.L. Doob (27 février 1910 – 7 juin 2004), <i>M. Yor</i>	33
INFORMATIONS	
Bilan de la session 2007 du CNU section 26, <i>Le bureau de la section</i>	41
Quelques informations sur la transition MSTP-AERES, <i>P. Auscher</i>	47
CARNET	
Hommage à Paulette Libermann, <i>C.-M. Marle</i>	51
Paulette Libermann, <i>Y. Kosmann-Schwarzbach</i>	53
Ahmet Hayri Körezlioğlu, <i>A.S Ustunel</i>	55
COURRIER DES LECTEURS	
À propos d'un texte de P-A.Meyer	57
LIVRES	61

Éditorial

La SMF a depuis peu un nouveau Président en la personne de Stéphane Jaffard ; elle a aussi un nouveau Rédacteur en Chef qui sera en charge de sa Gazette. C'est un grand honneur pour moi. Je suis entré au Comité de Rédaction de la Gazette en 2002, et c'est alors Colette Anné qui dirigeait celui-ci. Je voudrais tout simplement la remercier. Bien sûr pour le travail qu'elle a accompli, bien sûr pour l'excellente ambiance dans laquelle nous avons travaillé. Mais surtout, je voudrais profiter de ces quelques mots pour louer une qualité qu'elle possède et sur laquelle elle veille jalousement (car elle est devenue si rare) : son côté rebelle. Rebelle et libre.

Bientôt la Gazette aura à sa tête un tandem que je formerai avec Frédéric Patras. Le prochain numéro sera l'occasion pour nous de décrire les transformations que nous souhaitons voir s'effectuer. Tous ensemble, tous ensemble...

Pour finir, je voudrais souhaiter un joyeux anniversaire à Jean-Pierre Bourguignon à qui nous devons tant... et plus. Il nous a beaucoup donné (et pas seulement comme Président de la SMF) et il sait que nous allons continuer à lui en demander encore.

— Zidine Djadli

SMF

Mot du Président

L'actualité la plus importante pour notre communauté est certainement l'adoption de la loi sur l'université cet été. Vous avez été nombreux à nous faire part de vos inquiétudes.

En juin dernier, le bureau de la SMF a adopté une prise de position commune avec les Sociétés Françaises de physique et de chimie pour attirer l'attention du gouvernement sur des dérives possibles consécutives à l'avant projet dont nous avons connaissance (vous le trouverez sur notre site web).

Le texte définitif n'a pas apaisé nos craintes; nous allons de nouveau nous adresser au gouvernement avant la publication des décrets d'application pour formuler plus précisément nos inquiétudes face aux conséquences possibles de cette loi, et proposer des pistes pour le futur.

Une initiative conjointe de la SMF et du CIMPA a conduit à la mise en ligne d'une base de données sur les mathématiques dans le monde, qui fonctionne sur le même principe que l'encyclopédie Wikipedia.

Nous vous invitons à l'enrichir, la faire connaître en partageant les informations de toutes natures dont vous disposez concernant les relations internationales dans le domaine des mathématiques <http://smf.emath.fr/International/>

La SMF, dans le cadre de son action internationale et dans sa volonté d'accroître le rayonnement des mathématiques dans le monde a décidé de mettre en place un système de parrainage d'institutions des pays pauvres.

Chaque laboratoire français pourra parrainer une ou plusieurs institutions appartenant à la liste des pays pauvres établie par l'OCDE en offrant l'adhésion institutionnelle à la SMF à l'organisme de son choix. La SMF de son côté, offrira un abonnement d'un an à l'une de ses revues à l'organisme parrainé et veillera à la cohérence des parrainages.

Chaque laboratoire recevra dans le courant de l'automne une documentation pour la mise en place pratique de l'opération. Je vous invite à diffuser autour de vous cette information et à susciter des parrainages dans votre laboratoire.

Les prix d'Alembert et Anatole Decerf sont des prix bisannuels, et ils seront remis lors de notre journée annuelle, le 21 juin prochain. Ils sont dotés chacun de 2000 euros. Je vous rappelle que le Prix d'Alembert, récompense une personne ou un groupe étant parvenu à intéresser le grand public aux développements des mathématiques et à les relier aux préoccupations de nos contemporains; le Prix Anatole Decerf récompense des travaux d'enseignement ou de vulgarisation de la pédagogie des mathématiques. Vous pouvez postuler, ou proposer des candidatures

dès maintenant. Elles sont à envoyer à la SMF. Je vous encourage à faire de la publicité autour de ce prix dont la mission est de stimuler les initiatives qui font connaître notre discipline au grand public ; pour plus d'information, cf. <http://smf.emath.fr/VieSociete/PrixAlembert/>

Pour terminer sur un ton plus léger, le dieu de la numérologie nous a adressé un sympathique clin d'œil en début d'été : en recensant le nombre de nos adhérents, nous nous sommes rendu compte le 20 07 2007 que la SMF comptait précisément... 2007 adhérents !

Le 2 octobre 2007

Stéphane Jaffard

MATHÉMATIQUES

Un point de vue non-asymptotique pour la sélection de modèle

Pascal Massart¹

1. Introduction

Si la formation initiale d'un mathématicien ne comporte pas nécessairement un cours de probabilités ou a fortiori de statistique mathématique, notre quotidien, lui, est riche d'expressions qui empruntent au vocabulaire de la statistique. Notre mémoire est encombrée de bribes de phrase lues ou entendues ici ou là : « *les statistiques du commerce extérieur sont mauvaises* », « *le chômage ce mois-ci a baissé en données corrigées des variations saisonnières* », « *le chouchou des sondages est apparu détendu à la sortie de son quartier général* », « *les dernières estimations le donnent gagnant au deuxième tour* », « *le contrôle positif à la testostérone a été confirmé après analyse de l'échantillon témoin* », etc...

Afin de familiariser le lecteur mathématicien non averti avec les concepts et le vocabulaire de base de la statistique, il peut donc être utile (voire même judicieux) de s'appuyer sur la connaissance empirique de la statistique qu'il possède comme tout citoyen avec l'intention d'intégrer progressivement cette connaissance dans un formalisme mathématique. C'est avec ce point de vue que nous chercherons dans un premier temps à introduire la sélection de modèle au travers d'un exemple que chacun d'entre nous a eu l'occasion de rencontrer.

1.1. L'exemple des histogrammes

Les histogrammes sont communément utilisés comme outil de statistique descriptive pour représenter graphiquement des données.

1.1.1. Un outil de statistique descriptive

Supposons donc que nous disposions d'un ensemble fini de nombres réels x_1, x_2, \dots, x_n , où chaque valeur x_i correspond à une donnée. Par exemple x_i peut représenter le revenu annuel d'un individu i . Généralement on possède une bonne idée a priori d'un intervalle $[a, b]$ dans lequel varient ces données et pour simplifier on peut supposer (quitte à effectuer une transformation affine une fois pour toute) que les valeurs considérées varient dans l'intervalle $[0, 1]$. Pour définir un histogramme on choisit une partition $m = \{I_0, \dots, I_D\}$ de $[0, 1]$

¹ Université Paris-Sud.

par $D + 1$ intervalles dont les extrémités sont données par une suite croissante $y_0 = 0 < y_1 < \dots < y_D < y_{D+1} = 1$. Autrement dit, on a, pour chaque $j < D$, $I_j = [y_j, y_{j+1}[$ et $I_D = [y_D, y_{D+1}]$. Pour chaque j , on calcule le nombre n_j de données tombant dans l'intervalle I_j , à savoir

$$n_j = \sum_{i=1}^n \mathbb{1}_{I_j}(x_i)$$

et l'histogramme des données correspondant à la partition m est tout simplement défini comme étant la fonction de $[0, 1]$ dans \mathbb{R}

$$(1) \quad x \mapsto \sum_{j=0}^D \frac{n_j}{n(|I_j|)} \mathbb{1}_{I_j}(x),$$

avec pour chaque j , $|I_j| = y_{j+1} - y_j$. Cette fonction est naturellement constante sur chacun des morceaux de la partition m et c'est le graphe ou plutôt l'épigraphe de cette fonction qui est usuellement utilisé comme outil de représentation graphique des données. Noter que cette fonction est positive ou nulle et d'intégrale égale à 1, c'est donc une *densité de probabilité* sur $[0, 1]$. Si les points y_j sont équirépartis, c'est-à-dire si les intervalles I_j sont tous de même longueur $(D + 1)^{-1}$, la partition est dite régulière et l'histogramme est dit régulier. Même pour cette représentation parfaitement élémentaire des données, on voit poindre quelques questions fondamentales. La première d'entre elles est sans doute : qu'est-ce qu'une « bonne » partition m ou autrement dit comment peut-on mesurer la qualité de représentation des données par un histogramme sur une partition donnée m ? La seconde qui n'est pas comme nous le verrons sans lien avec la première est : comment en pratique choisir une partition m ? Sans formaliser davantage ce problème pour le moment, on peut aisément intuitiver qu'une partition trop pauvre, c'est-à-dire avec un trop petit nombre d'intervalles comparé à n , risque fort de conduire à une représentation non informative des données, alors qu'à l'inverse une partition trop riche qui comporterait un faible nombre de données par intervalle, fournit une représentation très erratique et donc pour le moins difficilement interprétable. Bien entendu, ces considérations sont purement qualitatives et ne permettent pas d'avancer vers un critère de qualité pour un histogramme et c'est à présent vers ce nouvel objectif que nous désirons nous diriger.

1.1.2. L'aléatoire s'en mêle

Pour progresser dans l'analyse des histogrammes, il nous faut introduire un cadre mathématique permettant de modéliser le fait qu'usuellement, les données qu'on souhaite représenter par un histogramme possèdent une certaine variabilité intrinsèque. Une façon de tenir compte de cette variabilité est d'utiliser une modélisation stochastique. C'est ce cadre que nous emploierons à présent dans lequel chaque donnée x_i est interprétée comme la réalisation $x_i = X_i(\omega)$ d'une variable aléatoire X_i , définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et prenant ses valeurs dans $[0, 1]$. Dans le cadre le plus simple que nous adopterons ici, les variables aléatoires sont supposées indépendantes et de même loi de probabilité P , ce qui correspond à l'idée que les données observées correspondent aux répétitions d'un

même phénomène aléatoire. On dit qu'alors X_1, \dots, X_n constitue un n -échantillon de loi P , ce qui mathématiquement se résume par la formule suivante :

$$\mathbb{P} \{ \omega \in \Omega \mid X_1(\omega) \in A_1, \dots, X_n(\omega) \in A_n \} = \prod_{i=1}^n P(A_i)$$

pour toute famille d'ensembles boréliens A_1, \dots, A_n de $[0, 1]$.

Dans ce nouveau cadre probabiliste, l'histogramme défini par (1) devient la réalisation d'une fonction aléatoire que nous noterons \widehat{s}_m . Plus précisément, nous avons pour tout $\omega \in \Omega$ et tout $x \in [0, 1]$

$$\widehat{s}_m(x, \omega) = \sum_{j=0}^D \frac{N_j(\omega)}{n(|I_j|)} \mathbb{1}_{I_j}(x),$$

avec pour chaque j , $N_j(\omega) = \sum_{i=1}^n \mathbb{1}_{I_j}(X_i(\omega))$. Supposons à présent que la loi P admette une densité de probabilité s par rapport à la mesure de Lebesgue, l'interprétation probabiliste de l'histogramme permet d'envisager la qualité de celui-ci pour approcher s .

1.1.3. L'histogramme vu comme un estimateur

En pratique, même s'il est raisonnable d'admettre que les données observées sont issues de la répétition d'un même phénomène aléatoire et que la loi de probabilité P commune aux variables aléatoires X_i admet une densité s par rapport à la mesure de Lebesgue, il est par contre exclu de considérer que la loi de probabilité P et par voie de conséquence cette densité s est connue. C'est le but même de la statistique dite « inférentielle » que d'obtenir des renseignements sur P à partir de l'observation d'une réalisation du n -échantillon X_1, \dots, X_n .

Rappelons que pour chaque réalisation de ce n -échantillon \widehat{s}_m est une densité de probabilité sur $[0, 1]$. Ce nouveau point de vue consiste donc à considérer à présent l'histogramme \widehat{s}_m comme une approximation aléatoire de la densité inconnue s fondée sur X_1, \dots, X_n , c'est-à-dire dans le langage de la statistique un *estimateur* de s . La distorsion entre la densité estimée \widehat{s}_m et la « vraie » densité s peut-être mesurée par exemple par le carré de la *distance de Hellinger* $\|\sqrt{s} - \sqrt{\widehat{s}_m}\|^2$, où $\|\cdot\|$ désigne la norme dans $\mathbb{L}_2([0, 1])$. Cette façon de mesurer la distorsion n'est évidemment pas la seule possible mais elle a le mérite d'une part d'être simple et d'autre part d'être invariante par un changement de mesure dominante. Évidemment l'erreur $\|\sqrt{s} - \sqrt{\widehat{s}_m}\|^2$ tout comme \widehat{s}_m est aléatoire. Afin de résumer la qualité d'estimation par un nombre plutôt que par une variable aléatoire, on a coutume d'en prendre l'espérance pour considérer le *risque de Hellinger*

$$\mathbb{E}_s \left[\|\sqrt{s} - \sqrt{\widehat{s}_m}\|^2 \right] = \int \|\sqrt{s} - \sqrt{\widehat{s}_m}\|^2 d\mathbb{P}_s(\omega).$$

Afin de rappeler explicitement leur dépendance en la densité inconnue s , on a pris soin d'utiliser ci-dessus les symboles \mathbb{P}_s et \mathbb{E}_s pour noter respectivement la probabilité ou l'espérance d'un événement ou d'une variable aléatoire fonction des variables observées X_1, \dots, X_n lorsque celles-ci sont indépendantes et de même loi

de densité s . Nous disposons à présent avec le risque de Hellinger d'une mesure objective de la qualité d'un histogramme.

1.1.4. Le « modèle histogramme »

L'histogramme \widehat{s}_m est une densité de probabilité constante par morceaux sur la partition m . On peut donc se demander quel rôle particulier joue \widehat{s}_m parmi toutes les fonctions possédant cette propriété. Autrement dit, si nous introduisons le *modèle histogramme*

$$S_m = \left\{ \sum_{j=0}^D a_j \mathbb{1}_{I_j} \mid a_0, \dots, a_D \in \mathbb{R}_+ \text{ et } \sum_{j=0}^D a_j |I_j| = 1 \right\}$$

la question est : quelle propriété spécifique possède \widehat{s}_m comme élément de ce modèle S_m ? C'est le moment d'introduire une notion clef en statistique mathématique : *la vraisemblance*. t étant une densité de probabilité donnée, la vraisemblance en t est la densité de l'observation X_1, \dots, X_n sous l'hypothèse que $s = t$, évaluée au point observé X_1, \dots, X_n , à savoir

$$\prod_{i=1}^n t(X_i).$$

On vérifie aisément que \widehat{s}_m maximise cette vraisemblance lorsque t parcourt S_m , ce qu'on peut encore synthétiser de la manière suivante

$$\widehat{s}_m = \operatorname{argmax}_{t \in S_m} \sum_{i=1}^n \ln(t(X_i)).$$

1.1.5. Le problème du choix de modèle

Nous venons de voir qu'à chaque partition m on peut faire correspondre le modèle S_m des densités constantes par morceaux sur m et associer à ce modèle l'estimateur par histogramme qui se trouve être l'estimateur par maximum de vraisemblance sur ce modèle S_m . Formaliser le problème du choix d'un « bon » modèle S_m revient donc à formaliser celui de la sélection d'un « bon » estimateur par histogramme \widehat{s}_m . Une façon naïve de procéder consiste à raisonner comme suit. Partant d'une collection finie \mathcal{M} de partitions (par exemple la collection de toutes les partitions régulières à au plus n morceaux), le meilleur estimateur par histogramme est défini par la partition $m_0(s)$ minimisant

$$m \mapsto \mathbb{E}_s \left[\left\| \sqrt{s} - \sqrt{\widehat{s}_m} \right\|^2 \right]$$

sur \mathcal{M} . Si cette définition paraît abstraitement satisfaisante, elle est inutilisable pour sélectionner effectivement une bonne partition en pratique puisque le risque de Hellinger considéré ci-dessus dépend malheureusement de la densité inconnue s . Tout le problème consiste donc à sélectionner une partition \widehat{m} construite à partir des seules observations X_1, \dots, X_n (et ne dépendant surtout pas de s), de telle façon que la performance en terme de risque de Hellinger de l'estimateur par histogramme

sélectionné $\widehat{s}_{\hat{m}}$ soit comparable à celle de $\widehat{s}_{m_0(s)}$. L'idée que nous allons étudier en détail ici consiste à choisir \hat{m} minimisant sur \mathcal{M} le critère suivant

$$m \mapsto - \sum_{i=1}^n \ln \widehat{s}_m(X_i) + \text{pen}(m)$$

où $\text{pen} : \mathcal{M} \rightarrow \mathbb{R}_+$ est une fonction dite de *pénalisation* convenable. Bien entendu toute la difficulté réside dans la définition judicieuse de la fonction de pénalité qui sera largement discutée dans ce qui suit. L'idée de choisir un modèle via un critère de type log-vraisemblance pénalisée remonte au début des années 70 avec les travaux précurseurs de Mallows et d'Akaike (voir [1], [13] et [24]). Il est temps à présent de quitter le strict exemple des histogrammes afin d'élargir le cadre de notre réflexion.

1.2. Inférence statistique

Le problème de base de l'inférence statistique consiste à prendre une décision à propos d'une quantité s liée à la loi inconnue d'une variable aléatoire observée \mathbf{X} . La nature de \mathbf{X} peut être diverse : il se peut qu'on observe un vecteur aléatoire ou un processus stochastique ou encore une image bruitée. De même s peut être un vecteur ou une fonction ou encore une image. On peut par exemple chercher à construire une zone de confiance pour s , c'est-à-dire une région aléatoire qui contient s avec une probabilité donnée. Partant d'une procédure d'estimation \widehat{s} de s , (c'est-à-dire d'une fonction de l'observation \mathbf{X}) et d'une *fonction de perte* ℓ (typiquement ℓ est une distance ou le carré d'une distance comme le carré de la distance de Hellinger utilisé ci-dessus) permettant de préciser la qualité de \widehat{s} , l'approche naturelle pour réaliser une telle construction consiste à analyser la répartition de $\ell(s, \widehat{s})$. Il se trouve qu'en règle générale il est exclu d'évaluer de manière exacte la distribution de la procédure d'estimation. Il est alors essentiel de disposer d'outils pertinents d'approximation de cette répartition issus du Calcul des Probabilités.

1.2.1. La théorie asymptotique

Dans le cas où $\mathbf{X} = \mathbf{X}^{(n)}$ dépend d'un paramètre n (typiquement lorsque $\mathbf{X} = (X_1, \dots, X_n)$, où les variables X_1, \dots, X_n sont indépendantes et de même loi), la *théorie asymptotique* en statistique utilise les théorèmes limites (Théorème Central Limite, Principes de Grandes Déviations...) comme des outils d'approximation lorsque n tend vers l'infini. L'exemple historique le plus représentatif de cette approche est sans doute l'utilisation du Théorème Central Limite pour l'analyse du comportement lorsque n tend vers l'infini, de l'estimateur du maximum de vraisemblance sur un modèle paramétrique régulier. Si nous prenons à nouveau pour exemple le cadre de l'estimation de la densité dans lequel on observe X_1, \dots, X_n indépendantes et équidistribuées, dont la distribution commune admet une densité inconnue s par rapport à une mesure dominante μ . Un modèle S dans ce cas est simplement une partie de l'ensemble des densités de probabilité par rapport à μ . Un

modèle S étant donné, l'estimateur du maximum de vraisemblance \hat{s} (s'il existe!) est simplement défini comme minimiseur sur S du critère empirique

$$t \mapsto \sum_{i=1}^n -\ln t(X_i).$$

Lorsque le modèle S est paramétrique $S = \{s_\theta, \theta \in \Theta\}$, où Θ est un ouvert de \mathbb{R}^D , on écrit plutôt \hat{s} sous la forme $\hat{s} = s_{\hat{\theta}}$. Sous des conditions de différentiabilité convenables de s_θ par rapport au paramètre θ et à la condition que s appartienne effectivement au modèle S (et donc s'écrive $s = s_{\theta_0}$ pour un certain $\theta_0 \in \Theta$), le résultat classique de la théorie asymptotique évoqué plus haut concerne la normalité asymptotique de $\sqrt{n}(\hat{\theta} - \theta)$ lorsque n tend vers l'infini. On peut également garantir que la matrice de covariance apparaissant dans la loi gaussienne asymptotique est en un sens minimale, c'est ce qu'on appelle la propriété d'*efficacité asymptotique*. Plus récemment, avec les travaux séminaux de Dudley dans les années 70 sur les processus empiriques, la théorie des probabilités dans les espaces de Banach a profondément influencé le développement de la statistique asymptotique, conduisant à des avancées décisives dans le domaine de la théorie de l'efficacité asymptotique. Le lecteur intéressé trouvera dans les ouvrages de van der Vaart and Wellner [32] et van der Vaart [31] de nombreux résultats allant dans cette direction.

1.2.2. La sélection de modèle

Le problème majeur pour le statisticien est alors de définir un modèle S convenable. Il peut être délicat de deviner quel modèle paramétrique utiliser pour refléter un jeu de données réelles et il est clair qu'une erreur de modélisation, qui se traduit par un trop grand éloignement de s par rapport à S , peut conduire à une qualité d'estimation catastrophique. On peut être alors naïvement tenté de choisir un très grand modèle. Si on choisit S comme étant l'ensemble de toutes les densités ou comme un trop vaste sous-ensemble de celles-ci, il est bien connu que la procédure du maximum de vraisemblance devient inconsistante (voir [3]) ou sous-optimale (voir [6]). Déterminer par avance quel modèle utiliser pose donc des problèmes. Si on souhaite réaliser l'opération du choix d'un modèle convenable avec le plus d'objectivité possible, l'idée clef de la sélection de modèle consiste à s'appuyer sur les données elles-mêmes afin de construire un critère que le modèle choisi devra minimiser au sein d'une liste donnée, plutôt que de se fier au seul flair du modélisateur pour effectuer ce choix. Il s'agit donc ici de traiter un problème qui généralise celui du choix de la partition pour construire un histogramme. Plus précisément, si $(S_m)_{m \in \mathcal{M}}$ est une liste finie de modèles paramétriques réguliers où chacun des modèles S_m est défini par D_m paramètres et si $(\hat{s}_m)_{m \in \mathcal{M}}$ désigne la liste des estimateurs du maximum de vraisemblance correspondants, le critère de log-vraisemblance pénalisée d'Akaike (voir [1]) propose de sélectionner le modèle $S_{\hat{m}}$ tel que \hat{m} minimise le critère

$$m \mapsto -\sum_{i=1}^n \ln \hat{s}_m(X_i) + D_m$$

sur \mathcal{M} . La conception même de ce critère repose sur une heuristique qui s'appuie lourdement sur le comportement asymptotique de l'estimateur du maximum de

vraisemblance évoquée plus haut et notamment sur une de ses conséquences connue sous le nom de théorème de Wilks, qui assure que si $s \in S_m$ alors (sous des conditions de régularité convenables) la quantité

$$2 \left(- \sum_{i=1}^n \ln \widehat{s}_m(X_i) + \sum_{i=1}^n \ln s(X_i) \right)$$

converge en loi vers une loi du chi-deux à D_m degrés de liberté (c'est-à-dire la loi d'une somme de D_m carrés de variables aléatoires indépendantes et de même loi normale $\mathcal{N}(0, 1)$). D'autres critères proposés ultérieurement tels que le critère bayésien proposé par Schwartz, connu sous le nom de BIC (voir [28]) par exemple, possèdent exactement la même caractéristique : leur conception repose sur une approximation asymptotique qui sous-entend donc que la liste des modèles est fixée tandis que n tend vers l'infini.

1.2.3. Le point de vue non asymptotique

Il se trouve que dans plusieurs situations d'intérêt motivées par les applications, il est utile de laisser croître la taille des modèles avec n . Nous verrons d'autres exemples un peu plus loin mais il est clair que pour ce qui concerne les histogrammes réguliers par exemple, il est légitime de permettre au nombre de morceaux de varier librement entre 1 et n . Il peut même être utile de laisser croître avec n le nombre de modèles d'une dimension donnée.

Exemple : la détection de ruptures

La détection de ruptures sur la moyenne d'un signal discret fournit un cas d'école de ce type. Soit s une fonction sur $[0, 1]$ représentant un signal inconnu. Si nous observons à chaque instant j/n un signal bruité X_j , de telle sorte que le vrai signal $s(j/n)$ à l'instant j/n apparaisse comme l'espérance de X_j la question de la détection de ruptures sur la moyenne se formalise par la recherche d'une partition optimale (en un sens à préciser) de $[0, 1]$ par des intervalles dont les extrémités appartiennent à $\{j/n, 0 \leq j \leq n\}$ sur laquelle s soit une fonction constante par morceaux. Les motivations proviennent de l'analyse de signaux sismiques pour lesquels les instants de rupture (c'est-à-dire les extrémités des intervalles de la partition) correspondent à des couches géologiques différentes. Dans ce cas, à chaque partition m correspond un modèle S_m de fonctions constantes sur chacun des intervalles de la partition m . Dans cet exemple, pour chaque entier $D \leq n$, le nombre de modèles de dimension D , c'est-à-dire en fait le nombre de partitions à D morceaux, vaut $\binom{n-1}{D-1}$ et croît donc polynomialement par rapport à n .

Dans de telles circonstances l'analyse asymptotique classique n'est plus pertinente et une autre approche devient nécessaire que nous appellerons *non asymptotique*. Par non asymptotique, nous ne voulons pas dire que nous cherchons des résultats valables lorsque n est systématiquement modéré. L'idée est plutôt que quelle que soit la valeur de n , (et peut-être même surtout lorsque n est grand), il est utile d'autoriser la liste aussi bien que la taille des modèles à dépendre de n afin de garantir que l'un d'entre eux soit proche de s . Lorsque la cible s est une fonction, ceci permet d'utiliser toutes les connaissances issues de la *théorie de l'approximation* afin de définir des modèles dont les propriétés d'approximation à

des échelles variables sont bien connues (nous pensons à des polynômes par morceaux à pas et degrés variables par exemple). Dans les vingt dernières années le phénomène de concentration de la mesure a fait l'objet d'une recherche intense et féconde tout particulièrement sous l'impulsion des remarquables travaux de Michel Talagrand qui ont abouti à la découverte de nouvelles inégalités très puissantes en probabilités (voir en particulier [29] et [30]). Le principal avantage des inégalités de concentration est qu'à l'inverse des théorèmes limites, elles fournissent des outils non asymptotiques. C'est donc en un sens sans surprise que nous verrons ces inégalités jouer un rôle crucial dans l'élaboration d'une théorie non asymptotique pour la sélection de modèles telle qu'elle a émergé durant ces dix dernières années (voir en particulier [7] et [5]). Notre point de vue sera ici d'expliquer les idées et motivations centrales de cette théorie en les explicitant sur des exemples que nous espérons parlants.

2. La sélection de modèle gaussienne

2.1. La régression linéaire gaussienne

Nous commencerons notre analyse avec le *modèle linéaire gaussien* qui est sans aucun doute l'un des modèles les plus simples et les plus utilisés en statistique. On observe dans ce cas des variables aléatoires X_1, \dots, X_n structurées par le modèle de régression linéaire suivant :

$$X_i = \sum_{j=1}^N \beta_j \varphi_j(i) + \sigma \xi_i \text{ pour } 1 \leq i \leq n,$$

où les variables aléatoires ξ_i sont indépendantes et de même loi normale $\mathcal{N}(0, 1)$ alors que les nombres $\varphi_j(i)$ sont eux connus et représentent des valeurs observées de variables explicatives φ_j . Ici, le terme variable est à considérer dans l'acception usuelle de variable « économique » ou « physique ». En pratique, X_i correspond à la valeur prise par une observation réalisée à la i^{e} expérience et le modèle ci-dessus signifie donc que cette valeur dépend linéairement des valeurs $\varphi_j(i)$ prises par les variables φ_j pour cette même expérience, plus un terme de perturbation aléatoire représenté par la variable aléatoire $\sigma \xi_i$. Les paramètres β_j sont bien entendu inconnus mais nous supposerons par contre, dans un premier temps, le paramètre σ connu. Bien qu'irréaliste en pratique cette hypothèse simplifie grandement l'analyse. Par ailleurs nous reviendrons sur le problème d'estimation de σ dans un second temps. Le cadre ci-dessus fournit bien un modèle paramétrique pour la densité du vecteur \mathbf{X} dans \mathbb{R}^n par rapport à la mesure de Lebesgue puisque les variables X_1, \dots, X_n sont indépendantes avec pour lois respectives la distribution normale de moyenne $s_i = \sum_{j=1}^N \beta_j \varphi_j(i)$ et de variance σ^2 . Pour reformuler ceci de manière équivalente, on constate que le vecteur aléatoire \mathbf{X} suit une loi gaussienne multidimensionnelle, de moyenne $s = (s_i)_{1 \leq i \leq n}$ et de matrice de covariance $\sigma^2 I_n$, où I_n désigne la matrice identité d'ordre n . Si (et c'est ce que nous supposerons dans la suite) les vecteurs φ_j sont linéairement indépendants, ils engendrent un espace de dimension N que nous noterons S_N et il devient équivalent d'estimer le vecteur de paramètres β dans \mathbb{R}^N ou le vecteur s dans S_N . C'est un problème paramétrique qui

peut se résoudre par la méthode du maximum de vraisemblance. La vraisemblance de \mathbf{X} et son logarithme valent alors respectivement

$$(2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - s_i)^2\right) \text{ et } -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - s_i)^2.$$

Introduisons à présent la norme euclidienne sur \mathbb{R}^n

$$\|x\|^2 = \frac{1}{n} \sum_{i=1}^n x_i^2, \text{ pour tout } x \in \mathbb{R}^n$$

et posons $\varepsilon = \sigma/\sqrt{n}$. Les raisons pour lesquelles nous avons renormalisé la norme euclidienne canonique et introduit le paramètre ε apparaîtront plus clairement par la suite. En tout cas, nous déduisons de ce qui précède que l'estimateur du maximum de vraisemblance \widehat{s}_N de s sur S_N est tout simplement la projection orthogonale du vecteur \mathbf{X} sur l'espace S_N . De plus l'invariance par rotation de la loi gaussienne multidimensionnelle $\mathcal{N}_n(0, I_n)$ garantit que la loi de $\varepsilon^{-2} \|\widehat{s}_N - s\|^2$ est identique à celle obtenue lorsque S_N est engendré par les N premiers vecteurs de la base canonique de \mathbb{R}^n . C'est donc une loi du chi-deux à N degrés de liberté. Par conséquent le *risque quadratique* de \widehat{s}_N se calcule explicitement par la formule

$$\mathbb{E}_s \left[\|\widehat{s}_N - s\|^2 \right] = N\varepsilon^2.$$

Il est intéressant de noter que le choix de \widehat{s}_N comme estimateur de s a encore un sens même si $s \notin S_N$. La formule de Pythagore permet de corriger l'expression du risque quadratique ci-dessus qui devient donc

$$(2) \quad \mathbb{E}_s \left[\|\widehat{s}_N - s\|^2 \right] = N\varepsilon^2 + \|s - s_N\|^2,$$

où s_N désigne la projection orthogonale de s sur S_N . Ce risque quadratique apparaît comme la somme de deux termes, l'un, appelé *terme de variance*, proportionnel au nombre de paramètres à estimer N et l'autre, appelé *terme de biais*, qui mesure la qualité d'approximation de la réalité que procure le modèle S_N . Ce second terme disparaît bien entendu lorsque le modèle est exact, c'est-à-dire contient s .

2.2. La sélection de variables

Dans le modèle linéaire gaussien exposé ci-dessus, le modèle S_N est supposé exact de telle sorte que le risque quadratique s'écrit $\mathbb{E}_s \left[\|\widehat{s}_N - s\|^2 \right] = N\varepsilon^2$. Cette approche qui repose sur le choix *a priori* d'un modèle peut conduire à des soucis de natures opposées. Afin de garantir une bonne qualité d'estimation on est tenté de prendre une valeur modérée pour N , c'est-à-dire de mettre une petite partie des variables explicatives dont on dispose dans le modèle. Si on omet des variables explicatives importantes non seulement s n'appartiendra pas à S_N mais surtout le terme de biais $\|s - s_N\|^2$ peut augmenter considérablement. A contrario, si pour contourner cette difficulté on utilise beaucoup de variables explicatives pour engendrer le modèle S_N , alors, même si le modèle est exact l'estimation sera de piètre qualité. Or il peut se faire que parmi les variables $\varphi_1, \dots, \varphi_N$, seules un petit nombre D d'entre elles soient réellement influentes. Cela signifie que si S_D désigne l'espace engendré par ces D variables (disons $\varphi_1, \dots, \varphi_D$) le terme de biais

$\|s - s_D\|^2$ va rester faible de sorte que $D\varepsilon^2 + \|s - s_D\|^2$ peut être sensiblement plus petit que $N\varepsilon^2$. On voit se dessiner ici une des premières idées importantes que nous souhaitons avancer : on peut tirer bénéfice de l'utilisation d'un modèle approché (ici S_D) plutôt que d'un modèle exact (ici S_N).

La problématique intéressante qui se dégage ici est donc celle de la *sélection de variables* qui, partant d'une famille (qui peut être vaste) de variables explicatives $\varphi_1, \dots, \varphi_N$, consiste à tenter de sélectionner les plus influentes d'entre elles. La seconde idée importante qui émerge ici est la suivante : la notion de risque d'estimation permet de bien formuler mathématiquement ce problème de sélection. En effet le « meilleur » sous-ensemble $\{\varphi_j, j \in m\}$ de variables est tout simplement celui qui minimise le risque quadratique de l'estimateur par projection orthogonale \widehat{s}_m sur l'espace S_m engendré par les $\varphi_j, j \in m$. Idéalement, on souhaiterait sélectionner m minimisant

$$\mathbb{E}_s \left[\|\widehat{s}_m - s\|^2 \right] = |m| \varepsilon^2 + d^2(s, S_m),$$

où $d^2(s, S_m) = \inf_{t \in S_m} \|s - t\|^2$. Bien entendu un tel sous-ensemble idéal $m(s)$ dépend de s qui est inconnu du statisticien et non pas de la seule observation \mathbf{X} . L'enjeu statistique est alors de construire une procédure \widehat{m} de sélection d'un sous-ensemble de $\{1, \dots, N\}$ qui ne dépende que de l'observation \mathbf{X} . Le critère de qualité que nous adopterons pour une telle procédure est celui qui découle naturellement de ce qui précède, c'est-à-dire un critère de performance en terme de risque pour l'estimateur par projection correspondant $\widehat{s}_{\widehat{m}}$. Plus précisément on souhaite que ce risque quadratique $\mathbb{E}_s \left[\|\widehat{s}_{\widehat{m}} - s\|^2 \right]$ soit aussi voisin que possible du risque quadratique de $\widehat{s}_{m(s)}$, soit

$$\inf_{m \subseteq \{1, \dots, N\}} |m| \varepsilon^2 + d^2(s, S_m).$$

2.3. Le modèle linéaire gaussien généralisé

Avant d'aller plus loin il est utile de fixer le cadre stochastique général dans lequel nous allons poser le problème de la sélection de modèle gaussienne. Considérons comme dans [8] le modèle linéaire gaussien généralisé défini de la manière suivante. Étant donné un espace de Hilbert séparable \mathbb{H} , on observe le processus \mathbf{X}^ε donné par

$$(3) \quad \mathbf{X}^\varepsilon(t) = \langle s, t \rangle + \varepsilon W(t) \text{ pour tout } t \in \mathbb{H},$$

où W désigne un *processus gaussien isonormal*, c'est-à-dire que W est une isométrie de \mathbb{H} sur un sous-espace gaussien de $\mathbb{L}_2(\Omega)$, s est un paramètre inconnu dans \mathbb{H} et ε un paramètre réel positif supposé connu.

2.3.1. Exemples

Voyons en détail quelles sont les possibilités de modélisation offertes par ce nouveau cadre en commençant par vérifier qu'il généralise bien le modèle linéaire gaussien fini-dimensionnel introduit plus haut.

Le modèle linéaire gaussien fini-dimensionnel

Dans ce cas on observe, comme indiqué plus haut,

$$(4) \quad X_i = s_i + \sigma \xi_i, \quad 1 \leq i \leq n,$$

où les variables aléatoires ξ_i sont indépendantes et de même loi normale $\mathcal{N}(0, 1)$. Si nous considérons le produit scalaire normalisé sur \mathbb{R}^n

$$\langle x, y \rangle = \frac{1}{n} \sum_{i=1}^n x_i y_i$$

associé à la norme $\|\cdot\|$ et si nous posons $W(t) = \sqrt{n} \langle \xi, t \rangle$, alors W est bien un processus gaussien isonormal et

$$\mathbf{X}^\varepsilon : t \mapsto \frac{1}{n} \sum_{i=1}^n X_i t_i$$

satisfait bien à (3) avec $\varepsilon = \sigma/\sqrt{n}$.

Le modèle de bruit blanc continu

Dans ce cas, on observe le processus $\{X^\varepsilon(x), x \in [0, 1]\}$ régi par l'équation différentielle stochastique suivante

$$(5) \quad dX^\varepsilon(x) = s(x) dx + \varepsilon dB(x) \quad \text{avec } X^\varepsilon(0) = 0,$$

où B désigne un mouvement brownien sur $[0, 1]$. Si nous définissons alors pour tout $t \in \mathbb{L}_2[0, 1]$, $W(t) = \int_0^1 t(x) dB(x)$, W est bien un processus gaussien isonormal sur $\mathbb{L}_2[0, 1]$ et $\mathbf{X}^\varepsilon(t) = \int_0^1 t(x) dX^\varepsilon(x)$ obéit bien à (3) dès lors que $\mathbb{L}_2[0, 1]$ est muni de son produit scalaire usuel $\langle s, t \rangle = \int_0^1 s(x) t(x) dx$. Typiquement s représente un signal et $dX^\varepsilon(x)$ représente le signal bruité reçu à l'instant x . Ce modèle s'étend aisément au cas multivarié si l'on considère un drap brownien multivarié B sur $[0, 1]^d$ et $\mathbb{H} = \mathbb{L}_2([0, 1]^d)$.

Le modèle de bruit blanc discret

Particularisons le modèle linéaire gaussien fini-dimensionnel au cas où $s_i = s(i/n)$, où s désigne une fonction définie sur $[0, 1]$. C'est-à-dire que

$$(6) \quad X_i = s(i/n) + \sigma \xi_i, \quad 1 \leq i \leq n,$$

où les variables aléatoires ξ_i sont indépendantes et de même loi normale $\mathcal{N}(0, 1)$. Si s est un signal, X_i représente le signal bruité à l'instant i/n . Ce modèle peut être vu comme une version discrétisée du modèle de bruit blanc continu. En effet, partant du bruit blanc continu, nous pouvons poser $\sigma = \varepsilon\sqrt{n}$ et

$$\xi_i = \sqrt{n}(B(i/n) - B((i-1)/n)), \quad \text{pour tout } i \in [1, n].$$

Le signal bruité reçu à l'instant i/n vaut alors

$$X_i = n(X^\varepsilon(i/n) - X^\varepsilon((i-1)/n)) = n \int_{(i-1)/n}^{i/n} s(x) dx + \sigma \xi_i.$$

Comme les propriétés du mouvement brownien garantissent que les variables ξ_i sont bien indépendantes et de même loi normale centrée réduite, nous revenons bien au modèle de signal discret avec $s_i = s^{(n)}(i/n)$, où $s^{(n)}(x) = n \int_{(i-1)/n}^{i/n} s(y) dy$

pour tout $x \in [(i-1)/n, i/n[$. Si le signal continu s est suffisamment régulier, la fonction en escalier $s^{(n)}$ représente une approximation convenable de s , ce qui établit un lien entre les modèles de bruit blanc discret et continu.

Le modèle de suite gaussienne

Bien que le processus donné par (3) ne soit assujéti à aucune base orthonormée particulière, sitôt qu'une telle base est donnée dans \mathbb{H} , on peut filtrer le processus sur cette base et obtenir ainsi une *suite gaussienne*. Supposons donc que \mathbb{H} soit de dimension infinie (le cas de la dimension finie ayant déjà été traité plus haut) et considérons une base orthonormée $\{\varphi_j, j \geq 1\}$ de \mathbb{H} . La suite des coefficients $\widehat{\beta}_j = \mathbf{X}^\varepsilon(\varphi_j)$ est alors structurée de la manière suivante

$$(7) \quad \widehat{\beta}_j = \beta_j + \varepsilon \xi_j, \quad j \in \mathbb{N}^* \text{ et } (\beta_j)_{j \geq 1} \in \ell_2.$$

Ici $(\beta_j)_{j \geq 1}$ représente la suite des coefficients de s dans la base $\{\varphi_j, j \geq 1\}$, c'est-à-dire que $\beta_j = \langle s, \varphi_j \rangle$ pour tout $j \in \mathbb{N}^*$. De plus les variables aléatoires ξ_j sont indépendantes et de même loi $\mathcal{N}(0, 1)$. On peut donc voir le modèle de suite gaussienne comme une extension naturelle du modèle linéaire gaussien fini-dimensionnel (4). L'intérêt du modèle de suite gaussienne provient de ce que si $\mathbb{H} = \mathbb{L}_2[0, 1]$ par exemple et si la base est bien choisie (base de Fourier ou base d'ondelettes) une condition de régularité sur la fonction s du type « s appartient à une boule d'un espace de Sobolev W » se traduit par une condition de sommabilité sur les coefficients de s du type « β appartient à un ellipsoïde ».

Revenons à présent sur le problème de la sélection de variables, à la lumière du nouveau cadre que nous venons d'introduire.

2.3.2. La sélection de variables en dimension infinie

Il est intéressant de noter que le problème de la sélection de variables continue à avoir du sens lorsque les variables sont « construites » par le statisticien afin de construire des modèles approchés de la cible qu'il cherche à estimer. Pour illustrer ce propos, plaçons-nous dans le modèle de bruit blanc gaussien continu. Afin de reconstruire le signal s à partir du signal bruité, une stratégie possible est de considérer une famille de fonctions linéairement indépendantes $\{\varphi_j, j \in \Lambda\}$, où Λ est soit un ensemble fini $\Lambda = \{1, \dots, N\}$ soit $\Lambda = \mathbb{N}^*$. Si l'on songe à la situation où $\{\varphi_j, j \in \Lambda\}$ représente une vaste famille finie d'éléments d'une base d'ondelettes par exemple, rechercher une représentation « parsimonieuse » de s revient à sélectionner un sous-ensemble fini m de Λ (de cardinal sensiblement plus faible que N si Λ est fini) afin de représenter s sur la famille de fonctions $\{\varphi_j, j \in m\}$. On vient de s'intéresser à la sélection de variables *complète*, c'est-à-dire celle pour laquelle on recherche un sous-ensemble de Λ parmi toutes les parties possibles de Λ . Un autre thème d'intérêt dans ce contexte est la sélection de variables dite *ordonnée*. En effet, si la famille $\{\varphi_j, j \in \Lambda\}$ considérée est cette fois la base trigonométrique (prise dans son ordre naturel), on peut être tenté, au vu des propriétés d'approximation bien connus de cette base, de se restreindre à la recherche de sous-ensembles ordonnés, c'est-à-dire du type $[1, D]$, $D \in \mathbb{N}^*$.

On voit émerger une autre idée d'importance. Dès lors que l'on s'intéresse à des modèles approchés dont la dimension peut varier sur une échelle très vaste, on ouvre la possibilité d'approcher un élément s d'un espace de dimension infinie,

c'est-à-dire de faire de l'estimation *non paramétrique*, alors même que tous les modèles approchés que nous manipulons sont eux paramétriques (mais évidemment de dimensions très diverses).

2.4. Sélection de modèle et oracles

Notre but est à présent d'indiquer un formalisme mathématique précis dans lequel poser le problème de la sélection de modèle gaussienne que nous formulons dans le cadre du modèle linéaire gaussien généralisé (3) introduit plus haut. Considérons donc une famille au plus dénombrable $\{S_m, m \in \mathcal{M}\}$, de modèles. Bien que ce ne soit pas strictement nécessaire nous supposons que chaque modèle est un sous-espace vectoriel de dimension finie D_m de \mathbb{H} . Nous verrons que cette restriction qui peut paraître à première vue très forte couvre d'une part un nombre considérable d'exemples et d'autre part simplifie grandement la présentation. Comme dans le cas où $\mathbb{H} = \mathbb{R}^n$, l'estimateur \hat{s}_m du maximum de vraisemblance de s sur le modèle S_m est tout simplement le minimiseur

$$(8) \quad \gamma^\varepsilon(t) = \|t\|^2 - 2\mathbf{X}^\varepsilon(t)$$

sur S_m . Il est aisé de le calculer explicitement. En effet, si $\{\varphi_j, 1 \leq j \leq D_m\}$ est une base orthonormée de S_m il s'exprime sous la forme

$$\hat{s}_m = \sum_{j=1}^{D_m} \mathbf{X}^\varepsilon(\varphi_j) \varphi_j.$$

Comme la projection orthogonale s_m de s sur S_m s'écrit

$$s_m = \sum_{j=1}^{D_m} \langle s, \varphi_j \rangle \varphi_j$$

on en déduit que

$$\varepsilon^{-2} \|\hat{s}_m - s_m\|^2 = \sum_{j=1}^{D_m} W^2(\varphi_j)$$

suit une loi du chi-deux à D_m degrés de liberté. Puisque $\|s - s_m\|^2 = d^2(s, S_m)$, la formule de Pythagore assure donc que le risque quadratique de \hat{s}_m s'écrit

$$\mathbb{E}_s \left[\|\hat{s}_m - s\|^2 \right] = D_m \varepsilon^2 + d^2(s, S_m).$$

Cette formule généralise celle obtenue pour le problème de la sélection de variables. Elle reflète parfaitement le paradigme du choix de modèle puis qu'on constate que sa minimisation implique de réaliser un bon équilibre entre le terme de variance $D_m \varepsilon^2$ et le terme de biais $d^2(s, S_m)$. Autrement dit, elle illustre bien l'idée intuitive qu'un bon modèle doit être un reflet convenable, sinon parfait, de la réalité, tout en restant d'une complexité raisonnable. Nous épouserons donc définitivement le point de vue du risque pour juger de la qualité d'un modèle. Chaque modèle S_m est ainsi représenté par le minimiseur \hat{s}_m du critère des moindres carrés γ^ε défini en (8) sur S_m et le « meilleur » modèle est celui qui minimise le risque quadratique $\mathbb{E}_s \left[\|\hat{s}_m - s\|^2 \right]$ lorsque m parcourt \mathcal{M} . Bien entendu le terme de biais dépendant de s , il en est de même d'un tel modèle que nous noterons donc $S_{m(s)}$. Selon

la terminologie introduite par Donoho et Johnstone (voir [15] par exemple), le minimiseur $\widehat{s}_{m(s)}$ du critère des moindres carrés correspondant est appelé *oracle*. Ce n'est évidemment pas un estimateur puisque $m(s)$ est inconnu du statisticien mais son risque quadratique

$$\mathbb{E}_s \left[\|\widehat{s}_{m(s)} - s\|^2 \right] = \inf_{m \in \mathcal{M}} \mathbb{E}_s \left[\|\widehat{s}_m - s\|^2 \right]$$

va servir de référence et de point de comparaison pour juger de la qualité des procédures de sélection qui seront définies à partir des seules données. Il est à noter que la notion de « meilleur » modèle définie ci-dessus diffère sensiblement de celle de modèle « exact ». En effet si s appartient à S_{m_0} il se peut parfaitement que le meilleur modèle soit de dimension inférieure à D_{m_0} et que \widehat{s}_{m_0} ne soit pas un oracle. Nous avons déjà, dans le contexte de la sélection de variables, expliqué les raisons pour lesquelles cette notion de meilleur modèle (au sens du risque) correspond bien à ce qui est recherché en pratique. Il se trouve donc que l'on puisse préférer un modèle approché à un modèle exact. Il est temps à présent d'en venir au coeur du problème, c'est-à-dire à la construction de procédures de sélection \widehat{m} , fondées uniquement sur l'observation, et telles que le risque de l'estimateur $\widehat{s}_{\widehat{m}}$ correspondant soit aussi proche que possible de celui d'un oracle.

2.5. Sélection de modèle par pénalisation

Décrivons tout d'abord formellement la méthode. Il s'agit d'une procédure de moindres carrés pénalisée. On se donne une fonction dite de *pénalité* $\text{pen} : \mathcal{M} \mapsto \mathbb{R}_+$ et on considère \widehat{m} minimisant

$$(9) \quad \gamma^\varepsilon(\widehat{s}_m) + \text{pen}(m)$$

lorsque m parcourt \mathcal{M} . Le modèle et l'estimateur sélectionnés sont alors respectivement définis par $S_{\widehat{m}}$ et $\widehat{s}_{\widehat{m}}$.

Cette méthode remonte au début des années 70 avec les critères dit du C_p de Mallows et d'Akaike (communément appelé AIC). Le problème essentiel est de comprendre quelle fonction de pénalité il convient de choisir. La proposition formulée par Mallows (voir [13] et [24]) dans le contexte du modèle linéaire gaussien fini-dimensionnel (qui, rappelons-le, correspond dans notre formalisme au cas où $\mathbb{H} = \mathbb{R}^n$) est de prendre comme fonction de pénalité $\text{pen}(m) = 2D_m\sigma^2/n$, ou encore $\text{pen}(m) = 2D_m\varepsilon^2$ puisque l'interprétation du modèle linéaire gaussien classique comme un modèle de type (3) passe par le changement de variable $\varepsilon = \sigma/\sqrt{n}$. Il se trouve que cette proposition est strictement identique à celle d'Akaike dans ce contexte de sélection de modèle linéaire au sein du modèle linéaire gaussien à variance σ^2 connue. Voyons sur quelle idée repose cette proposition et surtout en quoi elle est reliée aux notions de meilleur modèle et d'oracle.

2.6. L'heuristique de Mallows

L'idée de base est la suivante. Rappelons que notre souhait est que l'estimateur sélectionné $\widehat{s}_{\widehat{m}}$ imite l'oracle $\widehat{s}_{m(s)}$ et que $m(s)$ minimise le risque quadratique

$$\mathbb{E}_s \left[\|\widehat{s}_m - s\|^2 \right] = D_m\varepsilon^2 + \|s_m - s\|^2.$$

L'idée la plus naturelle serait d'estimer le risque quadratique de \widehat{s}_m puis de minimiser cette estimation. Sa mise en œuvre butte sur la difficulté du problème de l'estimation du terme de biais $\|s_m - s\|^2$. C'est ici qu'il convient d'amender légèrement (mais subtilement) l'idée initiale en remarquant que, d'après la formule de Pythagore, $\|s_m - s\|^2 = \|s\|^2 - \|s_m\|^2$, donc que $m(s)$ minimise également

$$(10) \quad -\|s_m\|^2 + D_m \varepsilon^2.$$

Contrairement au terme de biais, la quantité $\|s_m\|^2$ est facile à estimer. En effet, notons que

$$\|\widehat{s}_m\|^2 = \|s_m\|^2 + \|\widehat{s}_m - s_m\|^2 + 2 \langle s_m, \widehat{s}_m - s_m \rangle,$$

ce qui implique que

$$\mathbb{E}_s \left[\|\widehat{s}_m\|^2 \right] = \|s_m\|^2 + D_m \varepsilon^2.$$

$\|\widehat{s}_m\|^2 - D_m \varepsilon^2$ est donc un estimateur sans biais de $\|s_m\|^2$. Si nous substituons à $\|s_m\|^2$ son estimateur sans biais $\|\widehat{s}_m\|^2 - D_m \varepsilon^2$ dans (10) nous obtenons le critère du C_p de Mallows :

$$-\|\widehat{s}_m\|^2 + 2D_m \varepsilon^2.$$

2.7. Un théorème non asymptotique

L'heuristique de Mallows peut être justifiée (ou corrigée) en contrôlant l'écart entre $\|\widehat{s}_m\|^2$ et son espérance $\|s_m\|^2 + D_m \varepsilon^2$, uniformément en $m \in \mathcal{M}$. L'inégalité de concentration gaussienne est précisément un outil adapté à cet usage. Elle constitue la pierre angulaire de la preuve du théorème suivant (voir [8]) dans lequel nous obtenons simultanément une proposition de forme pour la pénalité et une borne non asymptotique pour le risque de l'estimateur sélectionné dont nous verrons, dans un second temps, qu'elle permet une comparaison effective avec le risque de l'oracle.

Théorème 1. Soit $(x_m)_{m \in \mathcal{M}}$ une famille de nombres positifs ou nuls tels que

$$\sum_{m \in \mathcal{M}} \exp(-x_m) = \Sigma < \infty.$$

Soit $K > 1$. Supposons que pour tout $m \in \mathcal{M}$

$$\text{pen}(m) \geq K \varepsilon^2 \left(\sqrt{D_m} + \sqrt{2x_m} \right)^2.$$

Si \widehat{m} minimise le critère pénalisé

$$-\|\widehat{s}_m\|^2 + \text{pen}(m),$$

alors l'inégalité suivante est valide

$$(11) \quad \mathbb{E}_s \|\widehat{s}_{\widehat{m}} - s\|^2 \leq C(K) \left\{ \inf_{m \in \mathcal{M}} \left(\|s_m - s\|^2 + \text{pen}(m) \right) + \Sigma \varepsilon^2 \right\},$$

où la constante $C(K)$ ne dépend que de K .

Il est important de comprendre dans quelle mesure le Théorème 1 permet une comparaison effective entre le risque de l'estimateur pénalisé \widehat{s}_m et celui de l'oracle $\inf_{m \in \mathcal{M}} \mathbb{E}_s \|\widehat{s}_m - s\|^2$. Pour ce faire, nous pouvons raisonner de la manière suivante. Rappelant à nouveau que le risque quadratique de \widehat{s}_m s'exprime sous la forme

$$\mathbb{E}_s \|\widehat{s}_m - s\|^2 = \|s_m - s\|^2 + D_m \varepsilon^2,$$

considérons la situation la plus simple dans laquelle pour un certain nombre L , le choix de $x_m = LD_m$ pour tout $m \in \mathcal{M}$ conduit d'après à $\sum_{m \in \mathcal{M}} \exp(-x_m) \leq 1$ (prendre 1 comme borne supérieure n'a rien de magique ici, 2 ferait tout autant l'affaire!). Si nous choisissons $\text{pen}(m) = KD_m (1 + \sqrt{2L})^2 \varepsilon^2$, nous voyons que le membre de droite de la borne de risque (11) est majoré (à un facteur près dépendant de K et de L) par $\inf_{m \in \mathcal{M}} \mathbb{E} \|\widehat{s}_m - s\|^2$. Dans ce cas nous obtenons bien une comparaison avec le risque idéal et l'estimateur sélectionné se comporte, à une constante près, comme un oracle.

Il est également intéressant de noter le lien qu'établit le Théorème 1 entre Statistique et Théorie de l'Approximation. Pour ce faire supposons que le nombre de modèles d'une dimension donnée soit fini et considérons une façon raisonnable de choisir les poids x_m comme fonction de la dimension de chacun des modèles, c'est-à-dire de la forme $x_m = x(D_m)$ avec

$$x(D) = \alpha D + \ln \# \{m \in \mathcal{M}; D_m = D\} \text{ et } \alpha > 0.$$

La pénalité peut alors être choisie de la manière suivante

$$\text{pen}(m) = \text{pen}(D_m) = K\varepsilon^2 \left(\sqrt{D_m} + \sqrt{2x(D_m)} \right)^2$$

et (11) devient

$$\mathbb{E}_s \|\widehat{s}_m - s\|^2 \leq C' \inf_{D \geq 1} \left\{ \inf_{m \in \mathcal{M}, D_m = D} \left(\|s_m - s\|^2 \right) + D\varepsilon^2 \left(1 + \sqrt{2x(D)} \right)^2 \right\},$$

où la constante positive C' ne dépend que de K et de α . À la lecture de cette inégalité on constate que les propriétés d'approximation de $\bigcup_{D_m = D} S_m$ sont absolument cruciales. On peut en particulier espérer un gain substantiel dans le terme de biais grâce à la redondance de modèles de dimension D pour un prix $x(D)$ relativement modeste puisque la dépendance de $x(D)$ en le nombre de modèles de dimension D est logarithmique. C'est typiquement l'effet constaté lorsqu'on utilise une base d'ondelettes pour débruiter un signal.

2.8. Exemples

De nombreux exemples d'applications du Théorème 1 sont développés dans [8]. Contentons-nous de reprendre ici les deux applications mentionnées plus haut : la sélection de variables et la détection de ruptures.

2.8.1. Sélection de variables

Soit $\{\varphi_j, j \in \Lambda\}$ une famille d'éléments linéairement indépendants de \mathbb{H} avec soit $\Lambda = \{1, \dots, N\}$, soit $\Lambda = \mathbb{N}^*$. Pour chaque sous-ensemble m de Λ , nous définissons le sous-espace S_m engendré par $\{\varphi_j, j \in m\}$ et nous considérons une collection \mathcal{M} de parties finies de Λ .

La sélection de variables ordonnées

Nous choisissons dans ce cas pour \mathcal{M} la collection de tous les sous-ensembles de Λ de la forme $\{1, \dots, D\}$. Puisque cette collection ne comporte qu'un seul modèle de dimension donnée D , on peut choisir comme poids $x_m = \alpha D_m$, ce qui conduit à

$$\Sigma = \sum_{m \in \mathcal{M}} e^{-x_m} \leq \sum_{D=1}^{\infty} e^{-\alpha D} = (e^\alpha - 1)^{-1}.$$

Comme α peut être choisi arbitrairement petit, le Théorème 1 autorise de prendre une pénalité de la forme $\text{pen}(m) = K' |m| \varepsilon^2$ avec $K' > 1$. Ce choix conduit en utilisant (11) à une inégalité de comparaison avec le risque de l'oracle de la forme

$$\mathbb{E}_s \|\widehat{S}_m - s\|^2 \leq C' \inf_{m \in \mathcal{M}} \mathbb{E}_s \|\widehat{S}_m - s\|^2,$$

où la constante C' ne dépend que de K' . Par conséquent l'estimateur sélectionné se comporte (à constante près) comme un oracle. De plus, il est possible de prouver que la contrainte $K' > 1$ est optimale au sens suivant. Si $K' < 1$, on peut démontrer que même si $s = 0$, le critère de choix de modèle explose, c'est-à-dire qu'avec une grande probabilité, le modèle sélectionné est systématiquement de grande dimension. Ce comportement a pour corollaire que le risque de l'estimateur sélectionné est d'ordre $N\varepsilon^2$, où N est arbitrairement grand si Λ est infini et $N = |\Lambda|$ sinon, ce qui prouve qu'en aucun cas l'estimateur sélectionné ne peut se comporter comme un oracle.

La sélection de variables complète

Nous considérons le cas où $\Lambda = \{1, \dots, N\}$. Dans le contexte de la sélection de variables complète, \mathcal{M} désigne la collection de tous les sous-ensembles de $\{1, \dots, N\}$. Si nous choisissons comme poids $x_m = |m| \log(N)$, alors

$$\Sigma = \sum_{m \in \mathcal{M}} \exp(-x_m) = \sum_{D \leq N} \binom{N}{D} \exp(-D \log(N)) \leq e$$

et nous pouvons prendre comme pénalité

$$\text{pen}(m) = K |m| \left(1 + \sqrt{2 \log(N)}\right)^2 \varepsilon^2$$

avec $K > 1$. Dans ces conditions (11) devient

$$(12) \quad \mathbb{E}_s \|\widehat{S}_m - s\|^2 \leq C'(K) \inf_{D \geq 1} \left\{ \inf_{m \in \mathcal{M}, D_m=D} \left(\|s_m - s\|^2 \right) + D \log(N) \varepsilon^2 \right\},$$

où $C'(K)$ ne dépend que de K . Nous constatons que le facteur supplémentaire $\log(N)$ est un prix relativement modeste à payer comparé au gain potentiel dans le terme de biais que procure la redondance de modèles de dimension identique. Il est intéressant de noter qu'aucune hypothèse d'orthogonalité entre les éléments $\{\varphi_j, j \leq N\}$ n'est nécessaire pour obtenir ce résultat. Si toutefois le système est

orthonormé, l'estimateur par pénalisation ci-dessus peut être explicitement calculé et l'on retrouve l'estimateur par seuillage introduit par Donoho et Johnstone dans le cadre du modèle de bruit blanc (voir [15]). En effet, lorsque pour un certain nombre $T > 0$

$$\text{pen}(m) = T^2 |m|,$$

et

$$\widehat{\beta}_j = \mathbf{X}^\varepsilon(\varphi_j), \text{ pour } 1 \leq j \leq N,$$

l'estimateur des moindres carrés sur le modèle S_m engendré par $\varphi_j, j \in m$ a pour expression

$$\widehat{s}_m = \sum_{j \in m} \widehat{\beta}_j \varphi_j.$$

Dans ces conditions, le critère pénalisé s'écrit

$$\text{crit}(m) = -\|\widehat{s}_m\|^2 + \text{pen}(m) = \sum_{j \in m} \left(-\widehat{\beta}_j^2 + T^2 \right).$$

Par conséquent l'ensemble \widehat{m} minimisant le critère $\text{crit}(m)$ lorsque m parcourt la collection de tous les sous-ensembles de Λ vaut exactement

$$\widehat{m} = \left\{ j \in \Lambda, -\widehat{\beta}_j^2 + T^2 \leq 0 \right\}.$$

En d'autres termes

$$\widehat{s}_{\widehat{m}} = \sum_{j=1}^N \widehat{\beta}_j \mathbb{1}_{|\widehat{\beta}_j| \geq T} \varphi_j$$

qui est bien un estimateur par seuillage, le seuil T fourni par notre Théorème étant finalement de la forme $T = \sqrt{K} \left(1 + \sqrt{2 \log(N)} \right) \varepsilon$. On peut à nouveau prouver que la contrainte $K > 1$ est fine.

Notons que les calculs précédents sur les poids peuvent être légèrement améliorés. Plus précisément il est possible de remplacer le facteur logarithmique $\log(N)$ ci-dessus par $\log(N/|m|)$. En effet rappelons la majoration classique suivante pour le coefficient binomial

$$(13) \quad \ln \binom{N}{D} \leq D \ln \left(\frac{eN}{D} \right).$$

Un choix de x_m de la forme $x_m = |m| L(|m|)$ implique que

$$\begin{aligned} \Sigma &= \sum_{D \leq N} \binom{N}{D} \exp[-DL(D)] \leq \sum_{D \leq N} \left(\frac{eN}{D} \right)^D \exp[-DL(D)] \\ &\leq \sum_{D \leq N} \exp \left[-D \left(L(D) - 1 - \ln \left(\frac{N}{D} \right) \right) \right]. \end{aligned}$$

Si nous fixons $L(D) = 1 + \theta + \ln(N/D)$ avec $\theta > 0$ nous obtenons que $\Sigma \leq \sum_{D=0}^{\infty} e^{-D\theta} = [1 - e^{-\theta}]^{-1}$. Avec $\theta = \ln 2$, le Théorème 1 nous autorise à prendre comme pénalité

$$\text{pen}(m) = K\varepsilon^2 |m| \left(1 + \sqrt{2(1 + \ln(2N/|m|))} \right)^2$$

avec $K > 1$ et nous en déduisons la borne de risque suivante pour l'estimateur pénalisé correspondant

$$(14) \quad \mathbb{E}_s \left[\|\widehat{s}_m - s\|^2 \right] \leq C'' \inf_{1 \leq D \leq N} \{ b_D^2(s) + D(1 + \ln(N/D)) \varepsilon^2 \},$$

où $b_D^2(s) = \inf_{m \in \mathcal{M}, |m|=D} (\|s_m - s\|^2)$. Cette inégalité améliore légèrement (12).

Par ailleurs, l'estimateur pénalisé reste facilement calculable lorsque le système $\{\varphi_j\}_{j \leq N}$ est orthonormé. En effet

$$\begin{aligned} & \inf_{m \in \mathcal{M}} \left\{ - \sum_{j \in m} \widehat{\beta}_j^2 + K\varepsilon^2 |m| \left(1 + \sqrt{2L(|m|)}\right)^2 \right\} \\ &= \inf_{D \leq N} \left\{ - \sup_{\{m \mid |m|=D\}} \sum_{j \in m} \widehat{\beta}_j^2 + K\varepsilon^2 D \left(1 + \sqrt{2L(D)}\right)^2 \right\} \\ &= \inf_{D \leq N} \left\{ - \sum_{j=1}^D \widehat{\beta}_{(j)}^2 + K\varepsilon^2 D \left(1 + \sqrt{2L(D)}\right)^2 \right\} \end{aligned}$$

où $\widehat{\beta}_{(1)}^2 \geq \dots \geq \widehat{\beta}_{(N)}^2$ désignent les carrés des coefficients estimés $\{\widehat{\beta}_j, j \leq N\}$, rangés par ordre décroissant. Nous constatons que la minimisation du critère pénalisé revient à sélectionner une valeur \widehat{D} de D minimisant

$$- \sum_{j=1}^D \widehat{\beta}_{(j)}^2 + K\varepsilon^2 D \left(1 + \sqrt{2L(D)}\right)^2$$

et finalement à exprimer l'estimateur pénalisé sous la forme

$$(15) \quad \widehat{s}_m = \sum_{j=1}^{\widehat{D}} \widehat{\beta}_{(j)} \varphi_{(j)}.$$

La performance de cet estimateur est en un sens optimale. On peut démontrer en effet que la borne de risque (14) est optimale au sens dit *du minimax* sur l'ensemble $\mathbb{S}_D = \bigcup_{|m|=D} S_m$, $D \leq N$. Autrement dit, il existe une constante absolue strictement positive κ telle que quel soit l'estimateur \widetilde{s} de s

$$\sup_{s \in \mathbb{S}_D} \mathbb{E}_s \|\widetilde{s} - s\|^2 \geq \kappa D (1 + \ln(N/D)) \varepsilon^2.$$

2.8.2. Détection de ruptures multiples

Considérons le problème de détection de ruptures sur la moyenne décrit ci-dessus dans le cadre du modèle de bruit blanc discret. Le signal bruité observé est donc de la forme

$$X_j = s(j/n) + \sigma \xi_j, \quad 1 \leq j \leq n,$$

où les erreurs ξ_j sont indépendantes et de même loi normale $\mathcal{N}(0, 1)$. Définissons l'espace vectoriel S_m des fonctions constantes par morceaux sur la partition m . Détecter les ruptures revient à sélectionner un modèle au sein de la famille

$\{S_m\}_{m \in \mathcal{M}}$, où \mathcal{M} désigne la collection de toutes les partitions possibles de $[0, 1]$ par des intervalles dont les extrémités se situent sur la grille $\{j/n, 0 \leq j \leq n\}$. Puisque le nombre de modèles de dimension D , c'est-à-dire le nombre de partitions à D morceaux est égal à $\binom{n-1}{D-1}$, cette collection de modèles possède des propriétés combinatoires analogues à celle de la collection de modèles correspondant à la sélection de variables complète au sein de $N = n - 1$ variables. Concernant le choix de la pénalité et les bornes de risque qui en résultent, les mêmes considérations que dans le cas de la sélection de variables complète étudié ci-dessus restent donc valides.

2.9. Estimation adaptative et Théorie de l'Approximation

Le principal avantage de la borne de risque fournie par le Théorème 1 est qu'elle vaut pour toute valeur de s . Son principal désavantage est qu'elle ne compare le risque de l'estimateur sélectionné qu'avec celui des estimateurs appartenant à la collection $\{\hat{s}_m\}_{m \in \mathcal{M}}$ dont il est issu mais pas avec celui d'une procédure d'estimation quelconque. Exprimé en des termes familiers on peut donc craindre d'avoir sélectionné un estimateur « borgne au royaume des aveugles ». Bien entendu si nous souhaitons effectuer une comparaison de risque avec un estimateur quelconque, il nous faudra abandonner l'idée de la réaliser ponctuellement (c'est-à-dire pour toute valeur de s) car il est clair qu'un estimateur constamment égal à s_0 est parfait si $s = s_0$ puisque de risque nul même s'il est par ailleurs stupide. Il convient donc pour donner du sens à une telle comparaison de s'intéresser aux performances des estimateurs en plusieurs points simultanément. Une approche classique est de considérer le risque maximal sur certains sous-ensembles. C'est le point de vue dit *minimax*. Le risque minimax sur un sous-ensemble \mathcal{T} de \mathbb{H} est ainsi défini par

$$R_M(\mathcal{T}, \varepsilon) = \inf_{\hat{s}} \sup_{s \in \mathcal{T}} \mathbb{E}_s \left[\|\hat{s} - s\|^2 \right],$$

où l'infimum porte sur l'ensemble de tous les estimateurs possibles de s . La performance d'un estimateur donné \hat{s} , peut alors être mesurée par le rapport

$$\frac{\sup_{s \in \mathcal{T}} \mathbb{E}_s \left[\|\hat{s} - s\|^2 \right]}{R_M(\mathcal{T}, \varepsilon)}.$$

Si ce rapport est borné indépendamment des valeurs de ε , \hat{s} sera dit *approximativement minimax* sur \mathcal{T} . Pour illustrer notre propos, revenons au cas du modèle de bruit blanc continu en dimension 1, pour lequel $\mathbb{H} = \mathbb{L}_2[0, 1]$. Un exemple typique de choix pour l'ensemble \mathcal{T} est une boule d'un certain espace de Banach de fonctions régulières comme par exemple, une boule de rayon R d'un espace de Sobolev de régularité α . Si nous notons $W^\alpha(R)$ une telle boule, un sérieux désavantage de l'approche minimax est qu'un estimateur approximativement minimax peut très bien dépendre de α et de R , quantités inconnues en pratique. Il est donc préférable d'exiger d'une procédure d'estimation donnée, qu'elle soit approximativement minimax sur toute une famille d'ensembles \mathcal{T} simultanément (dans notre exemple toutes les boules $W^\alpha(R)$ lorsque α et R varient). C'est précisément le point de vue de l'adaptation au sens du minimax qui a fait l'objet de très nombreux travaux en statistique depuis le début des années 90. Mentionnons les nombreuses contributions de Donoho, Johnstone, Kerkycharian et Picard qui ont étudié les propriétés

d'adaptation des estimateurs par seuillage de coefficients d'ondelettes sur des familles de boules d'espaces de Besov (voir [14] pour un panorama). D'une manière ou d'une autre, toutes les constructions d'estimateurs adaptatifs reposent sur une procédure de sélection (ou d'agrégation) d'une famille d'estimateurs préliminaires qui peuvent bien entendu différer de la sélection par pénalisation comme c'est le cas par exemple de la méthode de Lepskii (voir [21] et [22]). Néanmoins, le principe est toujours le même. Dès lors qu'une procédure de sélection d'estimateurs se comporte comme un oracle, il suffit de vérifier grâce à des arguments de Théorie de l'Approximation que l'oracle lui-même est approximativement minimax sur la famille d'ensembles $\{\mathcal{T}_\theta\}_{\theta \in \Theta}$ d'intérêt pour conclure à l'adaptativité de l'estimateur sélectionné. Pour en revenir à la sélection de modèle proprement dite, il convient que la famille de modèles $\{S_m\}_{m \in \mathcal{M}}$ possède de bonnes qualités d'approximation vis-à-vis de la famille d'ensembles $\{\mathcal{T}_\theta\}_{\theta \in \Theta}$. Les performances de l'estimateur pénalisé \tilde{s} , sont ensuite évaluées pour chaque $\theta \in \Theta$ par

$$\sup_{s \in \mathcal{T}_\theta} \inf_{m \in \mathcal{M}} \left(\|s_m - s\|^2 + \text{pen}(m) \right).$$

Illustrons à présent ce principe.

2.9.1. Exemple : adaptation sur des ellipsoïdes

En guise d'illustration, supposons à nouveau que \mathbb{H} soit de dimension infinie et considérons une base orthonormée $\{\varphi_j, j \geq 1\}$ de \mathbb{H} . Considérons pour chaque suite $(\theta_j)_{j \geq 1}$ décroissant vers 0, l'ellipsoïde de \mathbb{H} défini par

$$\mathcal{E}_2(\theta) = \left\{ s \in \mathbb{H}, \sum_{j \geq 1} \left(\frac{\langle s, \varphi_j \rangle}{\theta_j} \right)^2 \leq 1 \right\}.$$

Étudions les propriétés d'adaptation à la famille des ellipsoïdes ci-dessus de l'estimateur associé à la procédure de sélection de variable ordonnée décrite au paragraphe 2.8.1. Rappelons qu'en pareil cas

$$\mathcal{M} = \{[1, D], D \geq 1\}.$$

De plus si $m = [1, D]$, S_m désigne le sous-espace engendré par $\{\varphi_j, 1 \leq j \leq D\}$ et la pénalité en m s'écrit $\text{pen}(m) = K'D\varepsilon^2$ avec $K' > 1$. Ce choix conduit en utilisant (11) à une inégalité de comparaison avec le risque de l'oracle de la forme

$$\mathbb{E}_s[\|\hat{s}_m - s\|^2] \leq C' \inf_{m \in \mathcal{M}} \mathbb{E}[\|\hat{s}_m - s\|^2] = C' \inf_{D \geq 1} \left(\sum_{j > D} \langle s, \varphi_j \rangle^2 + D\varepsilon^2 \right).$$

À présent si $s \in \mathcal{E}_2(\theta)$, le terme de biais dans l'inégalité ci-dessus se contrôle aisément

$$\sum_{j > D} \langle s, \varphi_j \rangle^2 = \sum_{j > D} \frac{\langle s, \varphi_j \rangle^2}{\theta_j^2} \theta_j^2 \leq \theta_{D+1}^2,$$

d'où

$$\sup_{s \in \mathcal{E}_2(\theta)} \mathbb{E}_s[\|\hat{s}_m - s\|^2] \leq C' \inf_{D \geq 1} (\theta_{D+1}^2 + D\varepsilon^2).$$

Or, pourvu que $\theta_1 \geq \varepsilon$, il est possible de prouver (voir par exemple [26]) que le risque minimax sur l'ellipsoïde $\mathcal{E}_2(\theta)$ est effectivement minoré, à une constante absolue multiplicative près par $\inf_{D \geq 1} (\theta_{D+1}^2 + D\varepsilon^2)$, ce qui démontre que \widehat{s}_m est approximativement minimax sur chacun des ellipsoïdes $\mathcal{E}_2(\theta)$, et ce quelle que soit la suite $\theta = (\theta_j)_{j \geq 1}$ décroissant vers 0, telle que $\theta_1 \geq \varepsilon$. Si $\mathbb{H} = \mathbb{L}_2[0, 1]$ et si $\{\varphi_j, j \geq 1\}$ désigne la base de Fourier, la propriété d'adaptation sur les ellipsoïdes ci-dessus implique que l'estimateur pénalisé est adaptatif sur la collection de toutes les boules de Sobolev $W^\alpha(R)$, avec $\alpha > 0$ et $R \geq \varepsilon$.

Cet exemple est représentatif de nombreux autres du même type (voir [26]) qui tendent en définitive à analyser les performances de diverses stratégies de sélection de variables lorsque celles-ci sont élaborées avec pour objectif d'approcher au mieux l'objet à estimer. On peut pour ce faire sélectionner des variables au sein d'une même base mais aussi profiter de la souplesse offerte par un résultat comme le Théorème 1 pour utiliser des variables provenant de différentes bases. Autrement dit rien n'oblige a priori à travailler avec une seule et même base. Afin de résoudre un problème de traitement du signal ou d'image donné, on peut donc tenter de choisir la meilleure représentation possible parmi plusieurs disponibles. On peut le faire globalement ou même à chaque niveau de résolution si on travaille avec des représentations multi-échelles. On trouvera dans les travaux de Stéphane Mallat et de ses collaborateurs plusieurs méthodes et résultats allant dans cette direction (voir en particulier [23] et [20]).

2.10. Conclusions

Les points saillants suivants émergent de l'étude de la sélection de modèle gaussienne.

- Le C_p de Mallows peut sous-pénaliser et il doit être corrigé lorsque le nombre de modèles de même dimension est trop élevé.
- On peut utiliser la sélection de modèle comme outil d'estimation non paramétrique et choisir des listes de modèles inspirées par la Théorie de l'Approximation afin de produire des estimateurs adaptatifs.
- La condition $K > 1$ apparaissant dans l'énoncé du Théorème 1 est fine.
- Quelle pénalité doit être au bout du compte recommandée ? On peut tenter d'optimiser la borne de risque fournie par le Théorème 1. C'est le travail effectué dans [9], dont la conclusion est que $K = 2$ est en général un bon choix.
- En pratique, le niveau de bruit est inconnu mais on peut finalement retenir de la théorie la formule suivante : pénalité "*optimale*" = $2 \times$ pénalité "*minimale*". Le point clef pour tirer profit de cette remarque est que la pénalité minimale peut être devinée à partir des données grâce au phénomène d'explosion : tant que la pénalité n'est pas assez lourde, le critère pénalisé choisit des modèles de très grande dimension. Une fois la pénalité minimale estimée, il reste à la multiplier par 2 pour obtenir la pénalité (présumée optimale) désirée. Cette stratégie fournit donc une pénalité dépendant des données qui ne nécessite pas la connaissance a priori du niveau de bruit ε (voir [17] pour les détails d'implémentation de cette méthode).

3. Extension au cas non gaussien

Revenons au problème plus général d'estimation d'une quantité $s \in \mathcal{S}$ liée à la loi de probabilité inconnue d'une observation X . Outre le cadre du modèle linéaire gaussien généralisé décrit précédemment, les cadres typiques auxquels on peut penser sont les suivants :

– estimation de la densité où $X = (X_1, \dots, X_n)$ les X_i , $1 \leq i \leq n$ étant des variables aléatoires indépendantes et de même loi admettant la densité s par rapport à une mesure dominante μ .

– modèle de régression avec apprentissage, où les variables aléatoires $X_i = (\xi_i, Y_i)$ sont des copies indépendantes d'un couple (ξ, Y) . La réponse Y à la variable explicative ξ est supposée de carré intégrable. La fonction de régression s est définie par $s(x) = \mathbb{E}_s [Y \mid \xi = x]$.

La méthode du maximum de vraisemblance possède une généralisation naturelle appelée estimation par minimum de contraste.

3.1. Sélection par minimum de contraste pénalisé

Considérons un critère $\gamma(\mathbf{X}, \cdot)$ tel que

$$t \mapsto \mathbb{E}_s [\gamma(\mathbf{X}, t)]$$

atteigne un minimum au point s sur \mathcal{S} . Un tel critère est appelé *contraste*. On peut associer à ce contraste la fonction de perte naturelle ℓ définie par

$$(16) \quad \ell(s, t) = \mathbb{E}_s [\gamma(\mathbf{X}, t)] - \mathbb{E}_s [\gamma(\mathbf{X}, s)] \geq 0.$$

Les exemples les plus connus de contraste sont l'opposé de la log-vraisemblance d'une part et le contraste des moindres carrés d'autre part. Voici comment les définir dans les cadres de l'estimation de la densité et de la régression.

– Densité

On observe $X = (X_1, \dots, X_n)$, où X_1, \dots, X_n sont des variables aléatoires indépendantes de même loi admettant pour densité s par rapport à une mesure dominante μ . Le choix de

$$\gamma(\mathbf{X}, t) = -\frac{1}{n} \sum_{i=1}^n \log(t(X_i))$$

conduit à la fonction de perte

$$\ell(s, t) = K(s, t).$$

$K(s, t)$ désigne l'information de Kullback-Leibler de la probabilité $s\mu$ relativement à $t\mu$, définie par

$$K(s, t) = \int s \log\left(\frac{s}{t}\right) d\mu$$

si $s\mu$ est absolument continue par rapport à $t\mu$ et $K(s, t) = +\infty$ sinon. Supposant cette fois que $s \in \mathbb{L}_2(\mu)$, il est possible de définir le critère des moindres carrés par

$$\gamma(\mathbf{X}, t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n t(X_i),$$

où $\|\cdot\|$ note la norme dans $\mathbb{L}_2(\mu)$. Le fonction de perte correspondante vaut dans ce cas

$$\ell(s, t) = \|s - t\|^2,$$

pour tout $t \in \mathbb{L}_2(\mu)$.

– Régression

On observe $\mathbf{X} = (\xi_1, Y_1), \dots, (\xi_n, Y_n)$ indépendantes et de même loi. Soit μ la loi commune aux variables ξ . Le contraste des moindres carrés se définit alors par

$$\gamma(\mathbf{X}, t) = \frac{1}{n} \sum_{i=1}^n (Y_i - t(\xi_i))^2$$

et la fonction de perte qui lui est associée vaut tout simplement

$$\ell(s, t) = \|s - t\|^2,$$

pour tout $t \in \mathbb{L}_2(\mu)$.

Si nous considérons donc un contraste $\gamma(\mathbf{X}, \cdot)$ et un modèle $S \subseteq \mathcal{S}$, un *estimateur de minimum de contraste* de s sur S est un minimiseur de $t \mapsto \gamma(\mathbf{X}, t)$ sur S . Étant donné une collection au plus dénombrable de modèles $(S_m)_{m \in \mathcal{M}}$, on peut associer à chaque modèle S_m un estimateur du minimum de contraste \hat{s}_m sur S_m . Le problème de sélection de modèle devient alors un problème de sélection d'estimateurs. La notion d'oracle introduite plus haut pour la sélection de modèle gaussienne s'étend aisément. Plus précisément, si nous considérons $m(s)$ minimisant $m \rightarrow \mathbb{E}_s[\ell(s, \hat{s}_m)]$ sur \mathcal{M} , $\hat{s}_{m(s)}$ est appelé oracle. $\hat{s}_{m(s)}$ représente donc une sélection idéale puisque son risque relativement à la fonction de perte naturelle ℓ est minimum.

La méthode de sélection par minimum de contraste pénalisé consiste à se donner une fonction de pénalité $\text{pen} : \mathcal{M} \rightarrow \mathbb{R}_+$ et à sélectionner \hat{m} minimisant le critère

$$\gamma(\mathbf{X}, \hat{s}_m) + \text{pen}(m)$$

sur \mathcal{M} . L'estimateur puis le modèle sélectionnés sont respectivement définis par $\hat{s}_{\hat{m}}$ et $S_{\hat{m}}$.

3.2. Le rôle des inégalités de concentration

Indiquons à présent quelques éléments sur la façon dont peuvent se démontrer les résultats non asymptotiques sur les critères de choix de modèle par minimisation d'un critère pénalisé dont le Théorème 1 peut être considéré comme un prototype. Introduisons donc le processus centré

$$\bar{\gamma}(\mathbf{X}, t) = \gamma(\mathbf{X}, t) - \mathbb{E}_s[\gamma(\mathbf{X}, t)], \quad t \in \mathcal{S}.$$

Par définition un estimateur pénalisé $\hat{s}_{\hat{m}}$ satisfait pour tout $m \in \mathcal{M}$ et tout point $s_m \in S_m$

$$\gamma(\mathbf{X}, \hat{s}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma(\mathbf{X}, \hat{s}_m) + \text{pen}(m) \leq \gamma(\mathbf{X}, s_m) + \text{pen}(m),$$

ou de façon équivalente en substituant $\bar{\gamma}(\mathbf{X}, \cdot) + \mathbb{E}_s[\gamma(\mathbf{X}, \cdot)]$ à $\gamma(\mathbf{X}, \cdot)$,

$$\bar{\gamma}(\mathbf{X}, \hat{s}_{\hat{m}}) + \text{pen}(\hat{m}) + \mathbb{E}_s[\gamma(\mathbf{X}, \hat{s}_{\hat{m}})] \leq \bar{\gamma}(\mathbf{X}, s_m) + \text{pen}(m) + \mathbb{E}_s[\gamma(\mathbf{X}, s_m)].$$

Soustrayant $\mathbb{E}_s [\gamma(\mathbf{X}, s)]$ à chaque membre de l'inégalité ci-dessus, nous obtenons l'importante relation suivante

$$\begin{aligned} \ell(s, \widehat{s}_m) &\leq \ell(s, s_m) + \text{pen}(m) \\ &\quad + \overline{\gamma}(\mathbf{X}, s_m) - \overline{\gamma}(\mathbf{X}, \widehat{s}_m) - \text{pen}(\widehat{m}) \end{aligned}$$

On voit donc que la fonction de pénalité doit simultanément être choisie

- suffisamment grande pour contrer les fluctuations de $\overline{\gamma}(\mathbf{X}, s_m) - \overline{\gamma}(\mathbf{X}, \widehat{s}_m)$,
- mais pas trop non plus car idéalement on souhaiterait que $\ell(s, s_m) + \text{pen}(m) \leq \mathbb{E}_s [\ell(s, \widehat{s}_m)]$.

Par conséquent le secret d'une calibration convenable de la pénalité réside dans notre capacité à évaluer finement les fluctuations de $\overline{\gamma}(\mathbf{X}, s_m) - \overline{\gamma}(\mathbf{X}, \widehat{s}_m)$. C'est précisément ce que procure l'utilisation des inégalités de concentration combinée avec un procédé de localisation. Il s'agit en effet d'obtenir des estimées qui sont sensibles au fait que les fluctuations de $\overline{\gamma}(\mathbf{X}, t) - \overline{\gamma}(\mathbf{X}, u)$ sont d'autant plus faibles que t est proche de u . Il convient donc d'obtenir pour chaque $m' \in \mathcal{M}$, un bon contrôle de

$$\sup_{t \in S_{m'}} \frac{\overline{\gamma}(\mathbf{X}, s_{m'}) - \overline{\gamma}(\mathbf{X}, t)}{\omega(s_{m'}, t)},$$

pour une pondération convenable $\omega(s_{m'}, t)$ permettant de facturer la proximité entre $s_{m'}$ et t .

L'inégalité de concentration gaussienne (voir [12]) et l'inégalité de Talagrand pour les processus empiriques (voir [30]) constituent les prototypes des inégalités utiles pour réaliser le contrôle ci-dessus. Rappelons un énoncé de l'inégalité de Talagrand. Étant donné X_1, \dots, X_n indépendantes et de même loi, et une classe \mathcal{F} (au plus dénombrable) de fonctions centrées en espérance et uniformément bornées par 1, on définit $Z = \sup_{f \in \mathcal{F}} \sum_{i=1}^n f(X_i)$ et $v = \mathbb{E} [\sup_{f \in \mathcal{F}} \sum_{i=1}^n f^2(X_i)]$. Alors, pour tout x positif, excepté sur un événement de probabilité moindre que $K \exp(-x)$ l'inégalité suivante est valide

$$Z \leq \mathbb{E}[Z] + \sqrt{2v\kappa x} + cx$$

où K , κ et c sont des constantes universelles. Si on suit l'approche de Ledoux fondée sur des inégalités de type Sobolev logarithmiques (voir [18] et [19]), on peut expliciter la valeur des constantes et prendre $K = 1$, $\kappa = 4$ et $c = 2$ (voir [25]). Si on modifie la définition du facteur de variance v en posant $v = 2\mathbb{E}[Z] + n \sup_{f \in \mathcal{F}} \mathbb{E}[f^2(X_1)]$, il est même possible d'obtenir les valeurs optimales $\kappa = 1$ et $c = 1/3$ (voir [10]) pour les constantes. Comme indiqué pour la première fois dans [7] dans le contexte de l'estimation de densité par moindres carrés pénalisés, l'inégalité de Talagrand pour les processus empiriques permet de prouver des théorèmes de sélection de modèle, analogues au Théorème 1 pour la sélection de modèle gaussienne. Parmi les travaux s'appuyant sur cette même idée, citons [11] pour la log-vraisemblance pénalisée sur des log-splines, [2] pour des critères de type Mallows dans le contexte de la régression avec des erreurs non gaussiennes (voir également [4] si les erreurs sont faiblement dépendantes) et enfin [27] pour l'estimation de l'intensité d'un processus de Poisson inhomogène par sélection de modèle.

4. Références

- [1] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In P.N. Petrov and F. Csaki, editors, *Proceedings 2nd International Symposium on Information Theory*, pages 267–281. Akademia Kiado, Budapest, 1973.
- [2] BARAUD, Y. Model selection for regression on a fixed design. *Probability Theory and Related Fields* **117**, n°4 467-493 (2000).
- [3] BAHADUR, R.R. Examples of inconsistency of maximum likelihood estimates. *Sankhya Ser.A* **20**, 207-210 (1958).
- [4] BARAUD, Y., COMTE, F. and VIENNET, G. Model selection for (auto-)regression with dependent data. *ESAIM : Probability and Statistics* **5** 33–49 (2001) <http://www.emath.fr/ps/>.
- [5] BARRON, A.R., BIRGÉ, L., MASSART, P. Risk bounds for model selection via penalization. *Probab. Th. Rel. Fields.* **113**, 301-415 (1999).
- [6] BIRGÉ, L. and MASSART, P. Rates of convergence for minimum contrast estimators. *Probab. Th. Relat. Fields* **97**, 113-150 (1993).
- [7] BIRGÉ, L. and MASSART, P. From model selection to adaptive estimation. In *Festschrift for Lucien Lecam : Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds.), 55-87 (1997) Springer-Verlag, New-York.
- [8] BIRGÉ, L. and MASSART, P. Gaussian model selection. *Journal of the European Mathematical Society*, n°3, 203-268 (2001).
- [9] BIRGÉ, L., MASSART, P. Minimal penalties for Gaussian model selection. *Probab. Th. Relat. Fields* **138**, 33-73 (2007).
- [10] BOUSQUET, O. A Bennett concentration inequality and its application to suprema of empirical processes. *C.R. Math. Acad. Sci. Paris* **334** n°6, 495-500 (2002).
- [11] CASTELLAN, G. Density estimation via exponential model selection. *IEEE Trans. Inform. Theory* **49** n°8, 2052-2060 (2003).
- [12] CIREL'SON, B.S., IBRAGIMOV, I.A. and SUDAKOV, V.N. Norm of Gaussian sample function. In *Proceedings of the 3rd Japan-U.S.S.R. Symposium on Probability Theory*, Lecture Notes in Mathematics **550** 20-41 (1976) Springer-Verlag, Berlin.
- [13] DANIEL, C. and WOOD, F.S. *Fitting Equations to Data*. Wiley, New York (1971).
- [14] DONOHO, D.L. and JOHNSTONE, I.M., KERKYACHARIAN, G. and PICARD, D. Wavelet shrinkage :Asymptopia? *J. R. Statist. Soc. B* **57**, 301-369 (1995).
- [15] DONOHO, D.L. and JOHNSTONE, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425-455 (1994).
- [16] DUDLEY, R.M. *Uniform Central Limit Theorems*. Cambridge Studies in advanced mathematics **63**, Cambridge University Press (1999).
- [17] LEBARBIER, E. Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing*, **85** n°4, 717-736 (2005).
- [18] LEDOUX, M. On Talagrand deviation inequalities for product measures. *ESAIM : Probability and Statistics* **1**, 63-87 (1996) <http://www.emath.fr/ps/>.
- [19] LEDOUX, M. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs **89**, American Mathematical Society.
- [20] LE PENNEC, E. et MALLAT, S. Sparse Geometric Image Representation with Bandelets. *IEEE Trans. on Image Processing*, **14**, n°4, 423-438, (2005).
- [21] LEPSKII, O.V. On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **36**, 454-466 (1990).
- [22] LEPSKII, O.V. Asymptotically minimax adaptive estimation I : Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36**, 682-697 (1991).
- [23] MALLAT, S. *A Wavelet Tour of Signal Processing*. Academic Press, 1999
- [24] MALLOWS, C.L. Some comments on C_p . *Technometrics* **15**, 661-675 (1973).
- [25] MASSART, P. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. of Probability*. **28**, n°2, 863-884 (2000).
- [26] MASSART, P. *Concentration inequalities and model selection*. In *Lectures on Probability Theory and Statistics, École d'Eté de Probabilités de St-Flour XXXIII-2003* (J. Picard, ed.). Lecture notes in Mathematics n° 1896 (2007) Springer, Berlin.

- [27] REYNAUD-BOURET, P. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Relat. Fields* **126**, n°1, 103-153 (2003).
- [28] SCHWARTZ, G. Estimating the dimension of a model. *Ann. of Statistics* **6**, 461-464 (1978).
- [29] TALAGRAND, M. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.* **81** 73-205 (1995).
- [30] TALAGRAND, M. New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563 (1996).
- [31] VAN DER VAART, A. *Asymptotic statistics*. Cambridge University Press (1998).
- [32] VAN DER VAART, A. and WELLNER J. *Weak Convergence and Empirical Processes*. Springer, New York (1996).

HISTOIRE

Joseph Leo Doob

(27 février 1910 – 7 juin 2004)

Marc Yor¹

Biographie de J.L. Doob : quelques points clé

L'accession du domaine des probabilités au rang de discipline mathématique à part entière ne s'est pas faite sans difficultés – écrire cela est un bel euphémisme! – ni avec rapidité.

Après Kolmogorov², qui pose les fondements axiomatiques de la théorie des probabilités, J.L. Doob est, incontestablement, l'un des mathématiciens – à l'origine, spécialiste des fonctions analytiques – qui a le plus œuvré pour la mathématisation rigoureuse et la création de ce qui allait devenir, en particulier à l'aide de ses propres travaux, la théorie des probabilités.

Né en 1910 à Cincinatti (Ohio, E.U), décédé en 2004 à Urbana (Illinois, E.U.), Doob a effectué toute sa carrière de professeur de mathématiques à l'université d'Urbana-Champaign (Illinois) entre 1935 et 1978. Très attaché à la nature des environs de Champaign, il y demeure jusqu'à sa mort en juin 2004.

Doob est l'auteur, entre 1957 et 1963, d'articles très importants sur les mouvements browniens conditionnés, et donne une preuve probabiliste du théorème de Fatou et de nombreuses extensions de ce théorème pour les limites à la frontière de (rapports de) fonctions harmoniques. Le second paragraphe de cette Notice sera plus particulièrement consacré à ces travaux.

Doob est également l'auteur de trois livres, publiés respectivement en 1953, 1984, et 1993, qui ont eu des retentissements sensiblement différents parmi les probabilistes, et plus généralement, dans la communauté mathématique.

– Le premier livre : *Stochastic Processes* (1953), qualifié par Paul-André Meyer (2000) de « Bible des Probabilités nouvelles », expose en particulier la théorie des processus stochastiques en temps continu, pour laquelle de nombreuses propriétés de mesurabilité et/ou régularité se posent de façon cruciale. Doob montre comment

¹ Université Paris VI, Institut Universitaire de France.

² Tambov, 25 avril 1903 - Moscou, 20 octobre 1987.

la théorie de la mesure, qu'il est nécessaire de développer de façon adéquate, permet de résoudre un bon nombre de ces problèmes.

On trouve également dans ce livre ce qui fera son succès principal, à savoir l'exposé de la théorie des martingales en temps discret et en temps continu, ainsi – ce que de jeunes probabilistes peuvent avoir du mal à apprécier aujourd'hui! – qu'une étude détaillée des équations différentielles stochastiques d'Itô; soulignons en effet que ce livre paraît en 1953 et que la « reconnaissance » générale du calcul d'Itô n'aura lieu véritablement qu'à partir de 1969, avec le merveilleux petit livre de Henry P. McKean : *Stochastic Integrals* (New York : Academic Press).

– Dans un second livre monumental (de plus de 800 pages) : *Classical Potential Theory and its Probabilistic Counterpart* (1984), Doob a assemblé de façon magistrale et exploité toute sa connaissance très profonde de deux domaines : la théorie newtonienne du potentiel et la théorie des probabilités. Dans une première partie, il expose avec beaucoup de soin et de pédagogie la théorie newtonienne du potentiel, aussi bien sous sa forme classique qu'avec certains de ses apports personnels. Puis dans une seconde partie, après avoir détaillé la théorie des probabilités proprement dite, il montre, en se référant de façon très précise aux points correspondants de la première partie, comment une argumentation probabiliste permet de retrouver ces résultats, ou bien d'en donner une version probabiliste. Il explique ainsi que la contrepartie probabiliste d'une fonction harmonique est une martingale³. Plus généralement, la décomposition de Riesz d'une fonction surharmonique se traduit, en probabilités, par la décomposition de Doob-Meyer d'une surmartingale en la différence d'une martingale et d'un processus croissant (prévisible). Ce résultat a eu une importance fondamentale pour le développement ultérieur de la théorie des processus stochastiques.

On peut toutefois regretter que ce traité, par son ampleur, semble plutôt avoir effrayé les lecteurs auxquels il était destiné – potentialistes et probabilistes, en premier lieu – et qui lui ont souvent préféré des ouvrages de dimensions plus réduites, qui venaient – à quelques années près – d'être écrits sur ces sujets⁴; je pense par exemple aux livres de Murali Rao, de Richard Durrett ou de Sydney C. Port et Charles J. Stone⁵. L'existence même de ces livres, leurs titres, et leurs contenus, montrent combien les apports des travaux de J.L. Doob dans ces domaines ont été fondamentaux.

– Le troisième livre : *Measure Theory* (1993) présente – ce sont les termes utilisés par Doob lui-même dans l'introduction de ce livre – la façon dont tout analyste en formation devrait aborder la théorie de la mesure, y incluant en particulier

³ Voir le petit glossaire ci-dessous.

⁴ Une autre explication, un peu plus dynamique, à ce défaut de lecture..., est peut-être que le début des années 80 est marqué par l'intérêt des probabilistes pour la géométrie différentielle stochastique d'une part, et le calcul des variations stochastiques (calcul de Malliavin) d'autre part, ainsi que le montrent très clairement les *Proceedings de la Conférence de Durham* (Angleterre) en juillet 1980, organisée par D. Williams (voir *Stochastic integrals : Proceedings of the LMS Durham Symposium, July 7-17, 1980*, sous la dir. de David Williams [Berlin : Springer, Lecture Notes in Mathematics 851, 1981]).

⁵ M. Rao, *Brownian Motion and Classical Potential Theory* (Math. Inst., Aarhus Univ., 1977); R. Durrett, *Brownian Motion and Martingales in Analysis* (Belmont, Calif., 1984); S. C. Port et C. J. Stone, *Brownian Motion and Classical Potential Theory* (New York : Academic Press, Probability and Mathematical Statistics, 1978).

les concepts probabilistes d'indépendance, d'espérance conditionnelle et de martingale, ainsi que les théorèmes correspondants (de convergence de martingales, par exemple) et leurs principales applications en analyse même.

En résumé, les traités de Doob sont rigoureux, sans excès démesuré de formalisme, et vont droit au but, l'auteur indiquant lui-même fortement pourquoi telle ou telle notion est importante : une manière d'écrire qui n'est pas si répandue chez les mathématiciens...

Un survol de l'œuvre de J.L. Doob

Si l'on peut attribuer – sans trop de risque d'erreur – la découverte de la notion de martingale à Jean Ville (1910 - 1988) exposée dans sa thèse : *Étude critique de la notion de collectif*, Paris (1939), ce sont les travaux de J. Doob qui développent la théorie des martingales, en établissant les théorèmes de convergence, et de nombreuses utilisations importantes des martingales. Je ne mentionnerai en particulier que deux résultats fondamentaux :

– le premier est le *théorème d'arrêt* qui exprime que la propriété de constance en espérance d'une martingale, espérance prise en tout temps t déterministe, s'étend lorsque l'on remplace t par n'importe quel temps d'arrêt borné T . C'est le théorème d'arrêt qui permet d'explicitier la transformée de Laplace (ou la fonction caractéristique) – et par suite la loi – de nombreuses fonctionnelles du mouvement brownien, et plus généralement de processus de Markov. En conséquence, la stratégie : « trouver la martingale » a été développée systématiquement depuis les travaux de Doob, pour l'obtention de tels résultats.

– un second résultat fondamental concerne la majoration en norme L^p ($p > 1$) du supremum d'une sous-martingale positive $(X_u, u \leq t)$, par un multiple de la norme L^p de X_t . De telles inégalités sont précieuses pour estimer la « grandeur » d'une sous-martingale ; combinées avec les inégalités de Burkholder-Davis-Gundy comparant le supremum d'une martingale et la racine carrée de sa variation quadratique, elles permettent de nombreuses estimations : par exemple, elles jouent un rôle clé pour montrer la convergence de la méthode d'itération de Picard pour la résolution d'équations différentielles stochastiques à coefficients lipschitziens.

Si l'on pose à un probabiliste contemporain la question : *quels sont les apports principaux de J.L. Doob en théorie des probabilités ?*, il est très vraisemblable que la réponse fournie consiste en les deux résultats ci-dessus, auxquels seront ajoutées la découverte et l'utilisation de la notion de h -transformée – souvent appelée tout simplement transformée de Doob – d'un processus de Markov. Cette transformation peut être lue au niveau du semi-groupe de ce processus, que l'on multiplie « à l'intérieur » et que l'on divise, « à l'extérieur », par h , fonction harmonique, ou plus généralement excessive pour le processus. Cette transformation modifie souvent très profondément la nature du processus d'origine : ainsi, avec $h(x) = x$, la transformée de Doob du mouvement brownien réel X , transformée que l'on notera ici \tilde{X} , a pour effet de conditionner le mouvement brownien X à ne pas revenir en 0 ; mieux encore, ce conditionnement fait s'échapper ce mouvement brownien transformé \tilde{X} vers l'infini. Ce nouveau processus \tilde{X} n'est autre que la norme euclidienne du mouvement brownien 3-dimensionnel, et l'étude du couple (X, \tilde{X}) est à la base de nombreux résultats sur le mouvement brownien réel. Ce

cas particulier de h -relation a suscité d'innombrables extensions, et continue de donner lieu à beaucoup de recherches, en particulier pour des processus multi-dimensionnels.

Les thèmes fortement entrelacés des limites angulaires, du théorème de Fatou et de la frontière de Martin représentent une part essentielle de l'apport de Doob aux « théorèmes limites à la frontière ». Cela va me permettre d'illustrer précisément, sur cet exemple fondamental, comment Doob a rédigé son second traité (1984) :

– à la page 641, Doob écrit : « dans la section 1.XII.19, il a été montré que, si h est une fonction strictement positive, harmonique dans un domaine Greenien D de \mathbb{R}^N et si v est une fonction positive surharmonique sur D , alors v/h a une limite fine minimale en presque tout point (relativement à M_h , mesure représentante de h , comme intégrale d'un noyau de Martin de D). Nous allons voir une formulation probabiliste équivalente de ce théorème 2.X.8 qui affirme que v/h admet une limite le long de certaines trajectoires browniennes conditionnées⁶. »

– le théorème 2.X.8 figure à la page 689, après le développement par Doob de la théorie des h -processus, depuis le début du Chapitre 2.X, p.668, jusqu'à la page 688.

– le cas particulier fondamental, où D est la boule de rayon δ dans \mathbb{R}^N , est présenté dans les pages 691-693.

La discussion approfondie de ce thème dans le traité de 1984 reprend les articles fondamentaux de Doob, publiés en particulier au *Bulletin de la SMF* (1957) et aux *Annales de l'Institut Fourier* (1959), dans lesquels il donne respectivement une preuve probabiliste et une preuve non-probabiliste du théorème de Fatou généralisé⁷.

Quelques prolongements de l'œuvre de J.L. Doob

Pour résumer l'apport essentiel de J.L. Doob en théorie des probabilités, Michel Emery a introduit l'image du *triangle stochastique* de Doob ayant pour sommets :

- S_1 : un espace de probabilité filtré ;
- S_2 : l'ensemble des temps d'arrêt sur cet espace ;
- S_3 : l'espace des martingales de cet espace.

L'un des objets des travaux de Paul-André Meyer, qui l'un des premiers, a prolongé l'œuvre de Doob, a été de classier les temps d'arrêt d'un espace de probabilité filtré. Pour ce faire, il lui a été nécessaire d'introduire 3 familles fondamentales de processus associés à un tel espace :

- les *processus prévisibles*, qui sont « strictement » dans le passé de la filtration ;
- les *processus optionnels*, qui représentent passé et présent,
- et enfin, les *processus progressivement mesurables*, tout juste adaptés à la filtration ambiante.

Les travaux de classification de Meyer ont consisté – entre autres – à décrire parmi les temps d'arrêt, ceux qui sont « prévisibles », ou au contraire ceux qui ne le sont pas du tout. En particulier, tous les temps d'arrêt d'un espace de probabilité filtré sont prévisibles si et seulement si toutes les martingales de cet espace sont

⁶ Traduction : Doob se réfère au h -processus.

⁷ On trouvera les références complètes à certains articles de Doob dans la bibliographie ci-dessous.

continues⁸, ce qui donne ainsi une illustration de l'idée du « triangle stochastique de Doob ».

Avec le développement du calcul stochastique d'Itô, qui permet d'intégrer par rapport à une martingale tout processus prévisible borné, il a été naturel de chercher à unifier ce calcul avec celui de l'intégrale de Lebesgue-Stieltjes, et pour cela on s'est intéressé à l'espace des *semi-martingales*, sommes d'une martingale et d'un processus à variation bornée, adapté à la filtration ambiante. Un théorème célèbre de Bichteler-Dellacherie caractérise les semi-martingales comme étant les « bons intégrateurs » des processus prévisibles bornés, résultat très satisfaisant quant à la nature de l'intégrale stochastique.

Un autre essai d'extension du triangle stochastique consiste à se demander quels sont les temps aléatoires ρ – qui ne sont plus nécessairement des temps d'arrêt – tels que, lorsque l'on arrête une martingale en ρ , on obtienne encore tout au moins une semi-martingale, bien sûr dans une (nouvelle) filtration adéquate, pour laquelle ρ serait maintenant un temps d'arrêt.

Cette question a été posée simultanément – et indépendamment, me semble-t-il – par D. Williams et P.A. Meyer en 1976-1977 ; elle a été résolue en partie par M. Barlow et T. Jeulin, entre autres, qui ont montré que si ρ est la fin d'un ensemble prévisible dans la filtration d'origine, la question admet une réponse positive.

Ceci a été le point de départ de la théorie des grossissements de filtrations, dont l'objet est de déterminer pour quelles sur-filtrations d'une filtration donnée, toutes les martingales d'origine restent des semi-martingales.

On peut dire, à l'heure actuelle, que deux types de grossissement ont été développés :

- les *grossissements initiaux* où toute l'information nouvelle est apportée à l'origine du temps,
- les *grossissements progressifs* où l'information nouvelle est apportée au fur et à mesure de l'écoulement du temps⁹.

Je viens d'évoquer, dans les points précédents, comment les sommets S_2 et S_3 du triangle stochastique de Doob ont fait l'objet de développements importants ; il en a été de même, en fait même à une période antérieure, pour le sommet S_1 , où l'on a cherché à comprendre ce que devenait une martingale relative à un espace de probabilité filtré, lorsque l'on change, de façon absolument continue, la probabilité de référence. Là encore, les martingales d'origine sont transformées en semi-martingales pour le nouvel espace de probabilité filtré ; ceci a été décrit au moyen des multiples versions du théorème de Girsanov, établi sous des conditions de plus en plus générales, à l'origine dans un cadre markovien, puis finalement tout simplement dans le cadre général de S_1 , par Van Schuppen-Wong (1974).

Remarquons qu'en mathématiques financières, le problème inverse suivant a une grande importance et a été résolu pour une grande classe de sommets S_1 : étant donnée une *semi-martingale* X relative à un espace de probabilité filtré, où P est la probabilité de référence, trouver toutes les probabilités Q , localement équivalentes à P , telles que sous Q , X soit une martingale.

⁸ C'est le cas pour l'espace de probabilité filtré du mouvement brownien.

⁹ Le lecteur intéressé trouvera des développements de ces points dans R. Mansuy et M. Yor *Random Times and Enlargements of Filtrations in a Brownian Setting* (Berlin : Springer, Lecture Notes in Mathematics 1873, 2006).

Ainsi, chacun des sommets du triangle stochastique de Doob a fait l'objet de nombreux développements, qui ont souvent nécessité beaucoup d'efforts, et dont on peut dire à bien des égards, qu'ils sont loin d'être achevés, montrant ainsi la profondeur et l'originalité de l'œuvre de J.L. Doob.

D. Revuz m'a fait remarquer, à juste titre, qu'il faut également mentionner l'importance de G. Hunt qui en associant à tout processus de Markov une théorie du potentiel a poursuivi le programme « Probabilités-Potentiel » de Doob.

Petit Glossaire du « Triangle Stochastique »

Espace de probabilité : un triplet constitué de :

- un espace de référence, souvent noté Ω ;
- une tribu \mathcal{F} sur Ω ; c'est-à-dire une famille de sous-ensembles de Ω , possédant certaines propriétés de stabilité ;
- une probabilité P , c'est-à-dire : une application " σ -additive" associant à tout ensemble A de \mathcal{F} un nombre $P(A)$, à valeurs dans $[0,1]$.

Filtration : une famille croissante $(\mathcal{F}_t)_{t \geq 0}$ de sous tribus de \mathcal{F} ; \mathcal{F}_t « mathématise » le passé jusqu'en l'instant t .

Temps d'arrêt T : une application de (Ω, \mathcal{F}) à valeurs dans $[0, \infty]$ telle que, pour tout t , $(T \leq t)$ appartient à \mathcal{F}_t . L'exemple typique de temps d'arrêt est le premier instant où un phénomène « observable pour la filtration (\mathcal{F}_t) » a lieu. A l'opposé, un « dernier instant » n'est typiquement pas un temps d'arrêt.

Martingale (relative à un espace de probabilité filtré) : une famille $(X_t)_{t \geq 0}$ de variables intégrables, telles que X_t soit \mathcal{F}_t -mesurable, et $E[X_t | \mathcal{F}_s] = X_s$ ($s \leq t$). Typiquement : le gain jusqu'en l'instant t dans un « jeu équitable ».

Références

Quelques articles écrits par J.L. Doob

- Semi-martingales and subharmonic functions. *Trans. Amer. Math. Soc.* 77 (1954), pp. 86-121.
- Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. Fr.* 85 (1957), p. 431-458.
- A non-probabilistic proof of the relative Fatou theorem. *Ann. Inst. Fourier Grenoble* 9 (1959), pp. 293-300.
- A relativized Fatou theorem. *Proc. Nat. Acad. Sci. USA* 45 (1959), pp. 215-222.
- What is a martingale? *Amer. Math. Monthly* (1971), pp. 451-463.

Autres références biographiques concernant J.L. Doob

- N.H. Bingham. Doob : A Half Century On. *J. App. Proba.* 42 (2005), pp. 257-266.
- P.A. Meyer. Measure Theory, by J.L. Doob. *Bull. Amer. Math. Soc.* 31 (1994), pp. 233-235.
- P.A. Meyer. Les processus stochastiques de 1950 à nos jours. In : *Développements des Mathématiques, 1950-2000*. ed : J. Pier, Birkhäuser (2000).
- J.L. Snell. A conversation with Joe Doob. *Stat. Sci.* 12(4), (1997), pp. 301-311.

INFORMATIONS

Bilan de la session 2007 du CNU, section 26

Le bureau de la section¹

Les membres du CNU sont en fin de mandat, et le Conseil sera renouvelé cet automne. Un très bref bilan du travail du CNU sur les quatre dernières années pourrait être le suivant.

Le travail de qualification aux fonctions de Maître de Conférences et aux fonctions de Professeur est lourd (plus de 500 dossiers examinés par an), mais utile. Le processus de qualification par le CNU délivre un label de qualité garantissant la compétence des enseignants-chercheurs en mathématiques dans les universités françaises.

Le choix annuel des promus à la hors-classe des MdC, à la première classe des professeurs et à la classe exceptionnelle des professeurs est difficile, car le nombre de promotions ouvertes à ces grades est largement inférieur au nombre de collègues qui devraient naturellement en bénéficier, au vu de la qualité de leur travail scientifique, de leur investissement pédagogique et des services rendus à la communauté dans l'administration de la recherche ou de leurs établissements. Ce déséquilibre entre l'offre et les besoins est particulièrement criant pour les promotions à la première classe des professeurs.

En sus de la comparaison des dossiers, nous avons aussi veillé au respect de quelques équilibres, en particulier en ce qui concerne la représentation des « sous-disciplines », et la distribution des âges des promus. Les choix que nous avons faits ont été adoptés à l'unanimité par notre Conseil.

La distribution des semestres de Congé pour Recherche ou Conversion Thématique est aussi délicate, d'autant plus que le CNU ne dispose que d'un contingent extrêmement réduit pour une demande forte en nombre et en qualité.

Nous remercions l'Institut Henri Poincaré qui a accueilli dans ses locaux historiques et agréables sept de nos huit sessions, et l'université de Tours qui nous a invités en mai 2007.

Voici le bilan de la quatrième année du présent Conseil.

¹ Emmanuel Lesigne, François Golse, Bernard Gleyse et Olivier Raimond.

Qualifications : bilan 2007

Qualifications aux fonctions de Maître de Conférences

Le nombre de candidats inscrits était de 510. Le nombre de dossiers non parvenus aux rapporteurs est de 125. Sur les 385 dossiers examinés, 252 candidats ont été qualifiés (soit 65%). Environ les 4/5 des refus de qualification sont justifiés par une inadéquation de la candidature au domaine disciplinaire recouvert par la section.

Comme les années passées, deux critères importants ont été utilisés dans l'évaluation des dossiers, en particulier pour les candidats dont le parcours ne s'inscrivait pas de façon canonique dans les thématiques de la section.

(1) L'aptitude à enseigner les mathématiques.

(2) L'activité scientifique. Dans les domaines d'application des mathématiques, cette activité ne doit pas se limiter à une description de modèles classiques et une utilisation de méthodes et algorithmes éprouvés. L'évaluation prend en compte l'apport méthodologique, la mise en place de modèles originaux, le développement de nouveaux algorithmes, la validation par des applications réalistes.

Recommandations aux candidats (et aux directeurs de thèse)

Le dossier de candidature doit faire apparaître clairement :

– La capacité à enseigner les mathématiques dans un cursus de Licence de Mathématiques.

– Un travail de recherche en mathématiques appliquées. L'utilisation d'un outil mathématique standard dans un travail de recherche relevant d'une autre discipline ne semble pas suffisante à elle seule pour la qualification en Section 26.

– Une activité liée à la recherche en mathématiques appliquées dans la période précédant la demande de qualification.

Le dossier de candidature doit être présenté avec soin et clarté. Nous demandons que les rapports préalables à la soutenance de thèse de doctorat soient joints au dossier (quand ils existent et sont publics, ce qui est le cas des doctorats français). Le dossier doit contenir un CV détaillé, les références complètes des travaux du candidat, et au minimum quelques-uns de ceux-ci.

La présence d'une publication dans une revue à comité de lecture n'est pas exigée pour les thèses récentes. Mais elle représente un élément d'appréciation décisif pour les thèses plus anciennes. La publication d'un article en seul auteur, ou sans son directeur de thèse, peut être un élément positif d'appréciation.

En ce qui concerne les candidats dont la formation et/ou la mention du doctorat ne relèvent pas des mathématiques (informatique, biologie, physique, mécanique, traitement du signal, économie,...), il est impératif qu'une large part du dossier de qualification soit consacrée à la mise en évidence :

– de la part des mathématiques dans leur formation initiale ;

– de leur contribution scientifique dans le domaine des mathématiques appliquées.

Pour les candidats titulaires d'un doctorat récent, il est naturel d'attendre qu'un ou plusieurs membres du jury de thèse, et si possible un des rapporteurs, relèvent de la section du CNU dans laquelle le candidat demande la qualification. (Cette condition n'est bien sûr pas absolue).

Enfin, signalons l'existence de guides édités par les sociétés savantes (livret du candidat SMF-SMAI, voir www.emath.fr) qui donnent des conseils très utiles aux candidats sur les postes universitaires.

Qualifications aux fonctions de Professeur

Le nombre de candidats inscrits était de 139. Le nombre de dossiers non parvenus aux rapporteurs est de 14. Sur les 125 dossiers examinés, 96 candidats ont été qualifiés (soit 77%). Un refus de qualification sur deux est justifié par une inadéquation de la candidature au domaine disciplinaire recouvert par la section.

Les points essentiels examinés dans un dossier de candidature à la qualification aux fonctions de Professeur sont les suivants :

- La capacité à enseigner les mathématiques dans un cursus de Master de Mathématiques.
- Un travail de recherche significatif en mathématiques appliquées, avec une activité avérée dans la période récente.
- La démonstration d'une réelle autonomie scientifique.
- L'aptitude à l'encadrement et à la direction de recherches.

Sur la base de ces critères, la majorité des dossiers examinés ne posait aucun problème.

Promotions

Nous donnons dans cette section un bilan du travail du CNU sur les promotions en 2007, auquel nous avons ajouté un bilan des promotions locales l'année précédente.

Les dossiers de candidature à une promotion doivent contenir un descriptif de l'ensemble de la carrière (et non des trois dernières années, comme c'est demandé par l'administration). À côté du CV et de la liste complète des travaux (classés par type de publication), le dossier doit comporter des informations précises sur les activités pédagogiques, administratives, et les services rendus à la communauté universitaire.

Chaque dossier de candidature est examiné par deux rapporteurs du CNU, désignés par le bureau, après consultation du bureau élargi.

Promotions à la hors-classe des MdC

Nombre de promotions offertes : 12

Nombre de collègues promouvables : 254

Nombre de candidats : 99

Liste des promus :

BENHENNI Karim (Grenoble II), BONANSEA ép. THIEULLEN Michèle (Paris VI), FAURE ép. LAUTON, Michèle (IUT Sceaux), GANDOLFO Daniel (Toulon), GRUN REHOMME Michel (Paris II), GUERRIER Viviane (IUFM Lyon), HABBAL Abderrahmane (Nice), HADIJI Rejeb (Paris XII), MEYER ép. BEL-LAMY Nadine (IUT Évry), POSTEL ép. DUTRONC Marie (Paris VI), SANDRI Dominique (Lyon I), TURLOT Jean-Christophe (IUT Pays de l'Adour).

Nous encourageons nos collègues promouvables à éviter une autocensure excessive.

Pour les promotions à la hors-classe, le CNU examine l'ensemble d'une carrière de MdC. À côté du travail de recherche et de l'activité d'enseignant, un investissement particulier dans le domaine pédagogique ou au service de la communauté scientifique est apprécié. Un objectif de ces promotions étant d'offrir une fin de carrière valorisée à des collègues méritants, le CNU est vigilant à une juste répartition des âges des collègues promus.

L'âge moyen des promus est 51 ans. Les âges s'étendent de 45 à 60.

Promotions à la première classe des PR

Nombre de promotions offertes : 12

Nombre de collègues promouvables : 258

Nombre de candidats : 139

Liste des promus :

BECKER Roland (Pau), BONNEL Henri (Nouméa), BRIANE Marc (INSA Rennes), DEL MORAL Pierre (Nice), FANG Shizan (Dijon), FRANCOIS Olivier (INP Grenoble), FRANCO Christian (Lille III), GAREL Bernard (INP Toulouse), GARNIER Josselin (Paris VII), GUES Olivier (Aix-Marseille I), PERRIER Valérie (INP Grenoble), SAINT RAYMOND Laure (ParisVI).

Pour l'examen des promotions à la première classe des professeurs, le CNU dégage de chaque dossier de candidature les éléments suivants :

- domaine scientifique, âge et ancienneté comme professeur,
- faits marquants de la carrière, distinctions scientifiques,
- responsabilités diverses (direction d'équipe, de projet ou d'établissement, responsabilités pédagogiques, activités éditoriales, appartenance à différentes commissions,...),
- activité scientifique (nombre et qualité des publications, communications),
- valorisation de la recherche, collaborations extra-mathématiques,
- encadrement doctoral (thèses encadrées et devenir des docteurs).

Les candidats sont invités à mettre clairement ces éléments en avant dans leurs dossiers.

Le CNU veille à une répartition équilibrée des sous-disciplines (analyse des EDP et analyse numérique, calcul scientifique, didactique, optimisation, probabilités, statistiques) qui n'exclut pas les dossiers transversaux ou atypiques.

L'âge moyen des promus est 44 ans. Les âges s'étendent de 32 à 57.

Promotions au 1^{er} échelon de la classe exceptionnelle des PR

Nombre de promotions offertes : 6

Nombre de collègues promouvables : 227

Nombre de candidats : 91

Liste des promus :

AZAIS Jean-Marc (Toulouse III), BOUCHITTE Guy (Toulon), JAFFARD Stéphane (Paris XII), KABANOV Youri (Besançon), KUTOYANTS Iouri (Le Mans), LE ROUX Alain (Bordeaux I).

Le CNU attend des candidats à une promotion au premier échelon de la classe exceptionnelle qu'ils aient fait preuve de compétences exceptionnelles dans les différentes missions d'un professeur des universités, que ce soit par l'excellence de leurs travaux de recherche, ou en jouant un rôle majeur dans la communauté

scientifique en termes d'encadrement, de diffusion et de structuration de la recherche.

L'âge moyen des promus est 55 ans. Les âges s'étendent de 45 à 61.

Promotions au 2^e échelon de la classe exceptionnelle des PR

Nombre de promotions offertes : 3

Nombre de collègues promouvables : 24

Nombre de candidats : 11

Liste des promus :

BARLES Guy (Tours), LEDOUX Michel (Toulouse 3), MATVEEV Vladimir (Dijon).

Parmi les candidats dont le dossier démontre une activité soutenue dans les différentes missions dévolues aux professeurs d'université, le critère essentiel pour le changement d'échelon est l'ancienneté dans la classe exceptionnelle, ainsi que l'âge.

L'âge moyen des promus est 53 ans. Les âges s'étendent de 48 à 63.

Promotions locales 2006

Les sections du CNU ne distribuent que la moitié (49,5%) des promotions ouvertes aux enseignants-chercheurs. (Ces promotions sont distribuées entre sections du CNU proportionnellement au nombre de promouvables.) Les autres promotions sont attribuées par les établissements d'enseignement supérieur.

Le bilan des promotions locales 2007 n'est pas encore disponible, mais voici le bilan des promotions locales en 2006 dans notre section.

Hors-Classe des Maîtres de Conférences

12 promotions avaient été attribuées par le CNU. 11 promotions ont été obtenues localement. Voici la liste des promus.

BORDIER Gérard (Paris X), CHABOUR Rachid (Metz), GHERMANI Belaid (Paris XII), HUARD Alain (INSA Toulouse), LASSNER François (Versailles-St Quentin), MAURIES ép. IGERSEIM Jacqueline (Strasbourg II), NACHAOUI Abdeljalil (Nantes), RAOUX Thierry (Reims), RICHARD ép. JUNG Françoise (Grenoble), ROUBAUD ép. GERMAIN Marie-Christine (Aix-Marseille I), VARENNE Jean-Pierre (Nice).

Première classe des professeurs

15 promotions avaient été attribuées par le CNU. 13 promotions ont été obtenues localement. Voici la liste des promus.

ALOUGES François (Paris XI), DANA Rose (Paris IX), FRANCHI Jacques (Strasbourg I), DESBAT Laurent (Grenoble I), GIACOMIN Giambattista (Paris VII), GUILLOPE Colette (Paris XII), JAOUA Mohamed (Nice), MISCHLER Stéphane (Paris IX), LABBAS Rabah (Le Havre), LELANDAIS ép. FLANDRIN Evelyne (Paris V), RESPONDEK Witold (INSA Rouen), SOULIER Philippe (Paris X), SUQUET Charles (Lille I).

Classe exceptionnelle des professeurs

Le CNU avait attribué 7 promotions au premier échelon de la classe exceptionnelle. 4 promotions ont été obtenues localement. Voici la liste des promus.

FABRIE Pierre (Bordeaux I), GUERRE ép. DELABRIERE Sylvie, JOUINI Elyes (Paris IX), PONTIER ép. SUEUR Monique (Toulouse III).

Le CNU avait attribué 4 promotions au second échelon de la classe exceptionnelle. Il y a eu 5 promotions locales :

BARAS Pierre (Chambéry), BERTOIN Jean (Paris VI), COMETS Francis (Paris VII), MASSART Pascal (Paris XI), PRUM Bernard (Évry).

**Congés pour recherche ou conversion thématique,
pour l'année 2007-2008**

Le nombre de semestres de CRCT que le CNU pouvait attribuer cette année est 8. Ce nombre est ridiculement faible par rapport au nombre de semestres demandés (plus de 100 dans notre section cette année), et à la qualité des projets annoncés.

Le CNU a proposé d'accorder un semestre de CRCT à :

BODART Olivier (MdC Clermont II), BONANSEA ép. THIEULLEN (MdC Paris VI), BRETON Jean-Christophe (MdC La Rochelle), BOUSSOUIRA ép. ALA-BAU Fatiha (Prof. Metz), FRANCOIS Olivier (Prof. INP Grenoble), GOATIN Paola (MdC Toulon), MATHIEU Pierre (Prof. Aix-Marseille I), MAURY Bertrand (Prof. Paris XI).

Suite à sa nomination à l'IUF, Pierre MATHIEU a renoncé au bénéfice de son CRCT. Le semestre ainsi libéré a été proposé à Alexandre CABOT, MdC à Toulon.

Motion sur les modalités d'attribution de PEDR

La motion suivante a été adoptée par le CNU 26, et transmise au ministère en février 2007.

« La 26^e section du Conseil National des Universités, réunie le 6 février 2007, demande la clarification de la procédure et des critères d'attribution des Primes d'Encadrement Doctoral et de Recherche. L'action minimale consisterait en la publication d'un compte-rendu public après chaque réunion de la commission d'attribution. Ce compte-rendu détaillerait les critères utilisés, présenterait un bilan chiffré, et donnerait la liste des membres du jury.

De plus les règles de désignation des membres de ce jury doivent être publiques.

De nombreux collègues méritant sans conteste la PEDR ont essuyé des refus ces dernières années ; l'incompréhension de la communauté est grande. L'obscurité et la pénurie alimentent un important mécontentement. »

Quelques informations sur la transition MSTP-AERES

Pascal Auscher

Ce texte sans caractère officiel a pour but d'éclairer notre communauté mathématique des modifications que l'arrivée de l'Agence d'Évaluation de la Recherche et de l'Enseignement Supérieur (AERES) entraîne dans le paysage de l'évaluation et de l'administration de la recherche.

L'AERES a été installée (i.e. munie d'un président et d'un CA) dans un décret de mars 2007 (faisant suite à la création de l'AERES en novembre 2006 prévue par la loi de programme sur la recherche d'avril 2006). Ce même décret a mis fin à l'existence de la Mission Scientifique Technique et Pédagogique (MSTP) structure interne au Ministère Délégué à l'Enseignement Supérieur et à la Recherche. Elle comprenait 10 départements scientifiques dont le Département Scientifique 1, Mathématiques et leurs Interactions, qui était composé alors de Pascal Auscher (Directeur Scientifique, succédant à Aline Bonami), Michel Kern (coordinateur), Pierre Arnoux, Marc Hoffmann, Abdellah Mokrane et Michel Pierre (chargés de mission).

Ses missions étaient principalement d'évaluer, pour le compte du ministère, les unités de recherche, les formations (Master et Écoles Doctorales), de donner un avis sur divers appels d'offres et sur les projets CPER, d'évaluer les projets internationaux (type Programme Hubert Curien (PHC), ex PAI, gérés par EGIDE, programmes ACCES...), de proposer un classement des allocations de recherche non contractualisées et enfin d'organiser l'expertise de la Prime d'Encadrement Doctoral et de Recherche (PEDR).

La MSTP a été chargée d'assurer une période de transition. C'est pour cela que la DS1 a continué d'exister jusqu'à fin juillet 2007. Elle a terminé l'expertise des formations Master Vague A, des unités de la vague B et a assuré l'expertise de la campagne PEDR 2007 pour la Direction Générale de l'Enseignement Supérieur (DGES). Pour l'expertise de la campagne PEDR 2008, rien n'est arrêté par la DGES.

L'AERES est maintenant opérationnelle. C'est une structure plus légère que la MSTP en nombre de scientifiques et toutes les activités de la MSTP n'ont pas été reprises.

Il faut d'abord savoir que l'AERES est une autorité administrative indépendante : elle ne dépend d'aucun ministère. Elle a donc vocation à étendre son action à tous les établissements de recherche et d'enseignement supérieur quel que soit le ministère de rattachement.

Elle est divisée en trois sections, chacune chargée d'une mission d'évaluation :

- section 1 chargée de l'évaluation des universités et des organismes (CNRS, INRA...),
- section 2 chargée de l'évaluation des unités de recherche et des structures fédératives,

– section 3 chargée de l'évaluation des formations : Master et Écoles Doctorales pour l'instant.

Les mathématiques y sont représentées par Pascal Auscher¹ et Michel Pierre², qui interviendront dans les sections 2 et 3.

La section 1 a repris les missions qu'effectuait le Comité National d'Évaluation (CNE) qui a aussi cessé d'exister.

La section 2 est chargée d'évaluer les laboratoires de recherche qui perçoivent de l'argent public, en particulier ceux qui relèvent de la contractualisation avec le Ministère de la Recherche mais aussi avec d'autres ministères (défense...). Le principe des comités de visite que l'on connaissait pour les UMR est étendu à tous les laboratoires quels qu'ils soient en mathématiques, UMR, EA (équipe d'accueil), JE (jeune équipe), ERT (équipe de recherche en technologie) et les fédérations de recherche.

La section 2 organise ces comités et choisit les experts. En ce qui concerne les UMR en mathématiques, il y aura une concertation très étroite avec le nouveau DSA au CNRS, Jean-Marc Gambaudo, pour assurer une continuité des évaluations organisées jusqu'à l'année dernière par le CNRS. Pour toutes les unités, les tutelles (universitaires ou autres) seront consultées et pourront proposer des experts ; le Comité National du CNRS continuera à être représenté (pour les unités CNRS) ; le CNU (25 ou 26) aura un représentant.

Pour cette vague (vague C), les comités de visite des laboratoires auront lieu entre les 15 octobre 2007 et la mi-mars 2008. Certains comités seront groupés (sites proches...) pour une meilleure vue d'ensemble. À la fin de l'évaluation, l'AERES rendra publique ses conclusions. L'agence ne prendra pas de décisions (qui reviendront entièrement aux tutelles, notamment en matière politique et financière). L'AERES pourra a posteriori donner un avis sur ces décisions lors de l'évaluation suivante. Certains PPF pourraient aussi être évalués par la section 2 pourvu qu'ils aient un aspect fédérateur (par exemple, les PPF-bibliothèques de mathématiques) : ce n'est pas encore arrêté.

La section 3 organise les évaluations des écoles doctorales en mettant en place les comités de visite sur site (pour les établissements de la vague B, c'est en cours). Elle organise aussi l'évaluation des Masters en conservant l'expertise sur dossier que pratiquait la MSTP, en utilisant les dossiers transmis par les établissements au Ministère de la Recherche. Les conclusions seront aussi rendues publiques.

En ce qui concerne les missions assurées par la MSTP non reprises par l'AERES, la situation est pour le moment la suivante : la Direction Générale de la Recherche et de l'Innovation (DGRI) du Ministère de la Recherche traite les projets internationaux des ministères de la Recherche et des Affaires Étrangères. J'attire l'attention des collègues que ce sont d'autres collègues de la discipline qui lisent les dossiers : il n'est donc pas inutile d'y parler de mathématiques.

¹ Délégué scientifique.

² Délégué scientifique adjoint

La DGRI transmet aussi un avis sur les demandes de subventions aux colloques au cabinet de Madame la Ministre³. Cette année, la DGRI a traité la campagne d'allocations de recherche non contractualisées. La responsable du secteur Math-STIC à la DGRI/A3 est Brigitte Rozoy. En attendant le remplacement de Mireille Martin-Deschamps, qui vient de terminer ses fonctions, les dossiers concernant les mathématiques sont traités par Michel Kern⁴ et Alain Lichnewsky⁵.

Il va de soi que la loi sur les universités du 10 août dernier va amener d'autres modifications.

³ Ces demandes, pour les mathématiques, sont à adresser à la ministre et la porte d'entrée est la DGRI/A3 qui couvre le secteur mathématiques : Sophie Cluet, Directrice du Département DGRI/A3, 1, rue Descartes, 75231 Paris Cedex 05.

⁴ INRIA.

⁵ Université Paris-Sud.

Paulette Libermann à Dinard

CARNET

Hommage à Paulette Libermann¹

(14 novembre 1919 – 10 juillet 2007)

Charles-Michel Marle²

Paulette Libermann est décédée le 10 juillet 2007, à Montrouge près de Paris, dans une maison de retraite où elle avait été transportée à la suite d'une intervention chirurgicale. C'est une grande figure de la géométrie différentielle, une actrice et un témoin des spectaculaires développements de cette discipline au cours de la seconde moitié du vingtième siècle qui disparaît.

Elle était née le 14 novembre 1919 à Paris, dans une famille juive d'origine russe et ukrainienne. Admise au concours d'entrée à l'École normale supérieure de Sèvres en 1938, elle y fut l'élève d'Élie Cartan, André Lichnerowicz et Jacqueline Ferrand. En application des lois racistes édictées par le gouvernement de Vichy, il lui fut interdit de se présenter au concours de l'agrégation de mathématiques de 1941. Madame Cotton, directrice de l'École, réussit à obtenir pour les trois élèves juives une bourse de quatrième année à l'école, avant d'être brutalement mise à la retraite en 1941. Grâce à elle, Paulette Libermann put s'initier à la recherche mathématique sous la direction d'Élie Cartan qui lui proposa un sujet, du niveau de ce que fut plus tard la « thèse de troisième cycle ». Elle garda toute sa vie une grande admiration pour lui et resta toujours une amie fidèle de la famille Cartan.

En juin 1942, quand le port de l'étoile jaune fut imposé, elle se réfugia avec ses parents et ses deux sœurs à Lyon. Miraculeusement, sa famille proche et elle-même échappèrent aux rafles du sinistre Klaus Barbie. Réintégrée à l'École de Sèvres à l'automne 1944, un mois après la libération de Lyon, elle put enfin se présenter à l'agrégation. Jeune agrégée, elle fut nommée professeur, d'abord à Douai, puis dès la rentrée de 1945 au lycée de jeunes filles de Strasbourg. Dans cette ville, elle fit la connaissance de mathématiciens de premier plan, notamment Charles Ehresmann, Georges Reeb et René Thom. Tout en assurant son enseignement au lycée, elle prépara une thèse sous la direction d'Ehresmann, qui occupait la seconde place, juste après Élie Cartan, dans son panthéon personnel.

Un bref passage au CNRS lui permit de mettre la dernière main à la rédaction de sa thèse, qu'elle soutint en 1953. Elle fut ainsi la première sévrienne titulaire d'un doctorat de mathématiques. Nommée professeur à l'université de Rennes, elle y enseigna, encadra un jeune chercheur et continua ses travaux en géométrie

¹ Je remercie Madame Corinne Mounier-Veil, nièce de Mademoiselle Libermann, ainsi que mes collègues et amies Michèle Audin et Yvette Kosmann-Schwarzbach, pour leur aide lors de la préparation de ce texte.

² Université Pierre et Marie Curie, Institut de Mathématiques.

différentielle. Elle fut élue professeur à la Faculté des Sciences de l'université de Paris en 1966. Lors de l'éclatement de celle-ci, en accord avec ses convictions politiques, elle choisit d'être rattachée à l'université Paris VII (aujourd'hui université Denis Diderot). Restée très attachée à l'École de Sèvres, elle y enseigna à deux reprises et fit passer le concours d'entrée de 1965 à 1967.

C'est en 1967, peu après sa nomination à Paris, que je fis sa connaissance, lorsqu'elle me proposa un sujet d'études pour ma « deuxième thèse », car à l'époque, on devait, outre le sujet principal, traiter un second sujet pour pouvoir soutenir une thèse.

À Paris, Paulette Libermann anima un séminaire, d'abord avec Ehresmann jusqu'à la mort de celui-ci en 1979, puis soit seule, soit avec l'aide de collègues plus jeunes (Yvette Kosmann-Schwarzbach, Jean-Pierre François, Pierre Cartier) jusqu'en 1990. Elle invitait dans ce séminaire non seulement des personnalités déjà célèbres, mais aussi de jeunes mathématiciens alors encore peu connus : Claude Albert, Jean-Pierre Bourguignon, Dan Burghilea, Claudette Buttin (hélas disparue en 1972), Marc Chaperon, Pierre Dazord, Jean-Paul Dufour, Yvette Kosmann, Antonio Kumpera, Daniel Lehmann, Josiane Lehmann-Lejeune, Pierre Molino, Francisco Javier Turiel, René Ouzilou, Jean-Paul Penot, Jean Pradines, Alan Weinstein, et d'autres. Elle était remarquablement bien informée des travaux des jeunes géomètres, français et étrangers. Les séances de son séminaire étaient généralement suivies de conversations amicales entre elle-même, les participants et le conférencier, dans un des cafés proches de l'université. Tous les sujets étaient abordés : Paulette Libermann était toujours parfaitement au courant de l'actualité politique, en France et dans le reste du monde, et se sentait très concernée par de nombreux problèmes de société, tout particulièrement, mais pas exclusivement, l'enseignement et la recherche.

Elle fut invitée pour des séjours de longue durée dans plusieurs universités et centres de recherche, à St Hugh's College à Oxford pour travailler avec Whitehead, à l'université de Californie à Berkeley, au Mathematical Sciences Research Institute de Berkeley, à l'Instituto Nacional de Matemática Pura e Aplicada de Rio de Janeiro en 1961. Jusqu'en 2004, elle a régulièrement présenté, lors des rencontres internationales auxquelles elle était invitée, des conférences sur ses travaux, toujours très appréciées. Son dernier article, sur les concepts de géométrie différentielle chez Ehresmann, a paru en 2007. Jusqu'en décembre 2006, elle a fidèlement assisté au séminaire de Géométrie et Mécanique, fondé par notre regretté collègue Pham Mau Quan, aujourd'hui appelé séminaire de Géométrie hamiltonienne ; elle y a encore participé en mars de cette année, après une première hospitalisation. Elle intervenait souvent en posant des questions toujours judicieuses aux conférenciers. Ses connaissances étaient très étendues et son jugement très sûr.

Les travaux de Paulette Libermann se situent dans la continuité de ceux de Cartan et d'Ehresmann. Nombre de ses résultats ont été utilisés par d'autres mathématiciens. Ainsi par exemple, dans sa thèse, intitulée « Sur le problème d'équivalence des structures infinitésimales régulières », elle a étudié les variétés symplectiques munies de deux feuilletages lagrangiens transverses et montré l'existence, sur les feuilles de ces feuilletages, d'une connexion plate canonique. Bien plus tard, ce résultat a figuré dans les travaux d'Alan Weinstein. Elle a aussi très tôt remarqué l'importance des feuilletages d'une variété symplectique qu'elle a appelés

« symplectiquement complets », tels que le crochet de Poisson de deux fonctions, localement définies, constantes sur chaque feuille, soit lui aussi constant sur chaque feuille, et prouvé que cette propriété équivaut à l'existence d'une structure de Poisson sur l'espace des feuilles, telle que la projection canonique soit une application de Poisson. Parallèlement à Jean Pradines, elle a étudié les concepts fondamentaux relatifs aux groupoïdes et algébroides de Lie. Elle a contribué à bien d'autres sujets : systèmes de Pfaff, géométrie des espaces de jets, prolongements de fibrés vectoriels et de fibrés principaux, variétés symplectiques et variétés de contact. Outre une importante partie du livre que nous avons publié ensemble, elle a écrit plus de 70 articles de recherche, notes, articles d'encyclopédies, communications à des congrès.

Paulette Libermann est l'auteur d'une œuvre mathématique importante. Elle connaissait remarquablement les travaux de la plupart des géomètres, ses maîtres et ses contemporains. Autant que la mathématicienne, c'est l'amie, chaleureuse, attentive et toujours prête à échanger des idées, qui nous manquera.

Hommage à Paulette Libermann (1919-2007)

Yvette Kosmann-Schwarzbach

Paulette Libermann était une mathématicienne dans la tradition d'Élie Cartan et de Charles Ehresmann. Elle fut une élève brillante au lycée Lamartine ; elle faisait parfois les devoirs pour ses camarades ce qui entraînait que ces bonnes françaises l'acceptaient, elle dont les parents parlaient yiddish à la maison quand ils ne voulaient pas que les enfants – Paulette avait deux soeurs – comprennent. Reçue à l'École de Sèvres (École Normale Supérieure de Jeunes Filles) en 1938, elle n'avait pu passer l'agrégation qu'après la guerre, en 1945. Sa famille, réfugiée à Lyon, habitait près de la Place Bellecour un immeuble qui jouxtait celui occupé par la Gestapo. Elle me racontait comment ses parents et elle n'avaient survécu que grâce à un hasard providentiel et elle gardait une sévère rancune contre la directrice de l'École de Sèvres qui n'avait pas su protéger ses élèves de la déportation.

À l'École, elle fut l'élève d'Élie Cartan qui, disait-elle, corrigeait longuement à l'encre rouge les copies des jeunes sévriennes, elle s'y initia à la recherche sous sa direction, et elle garda toujours ses liens avec la famille Cartan. Elle resta aussi l'amie de Jacqueline Ferrand qui devint « caïmane » (agrégée préparatrice) à l'École en 1939 où elle entreprit d'y moderniser l'enseignement des mathématiques. Partie enseigner au lycée à Strasbourg, elle y commença une longue et fructueuse carrière de recherche. C'est là que, tout en enseignant, elle écrivit sa thèse sous la direction d'Ehresmann et devint ainsi la première sévrienne à soutenir une thèse de mathématiques et, plus tard, une des premières femmes professeur d'université en France.

Les sujets de ses recherches sont nombreux ; quelques termes évoquent les principaux d'entre eux : pseudo-groupes, jets, tenseurs de structure, connexions de Cartan, structures presque hermitiennes, problème d'équivalence, systèmes de Pfaff,

feuilletages, géométrie symplectique et de contact et, plus récemment, des sujets plus à la mode, tels que applications moment et algébroides de Lie.

Nommée professeur à l'université de Rennes, elle y continua ses recherches et publications en géométrie différentielle. C'est aussi l'époque où un accident de voiture où elle était passagère causa la « jambe raide » qui l'handicapa toujours.

C'est en 1966 qu'elle est nommée professeur à la Faculté des sciences de Paris, et c'est alors que nous avons fait connaissance. Je revois la scène : au thé de l'Institut Henri Poincaré, qui avait lieu une fois par semaine et réunissait les enseignants et les chercheurs en mathématiques – institution qui a disparu depuis –, dans la salle 116 du premier étage, une dame, de petite taille, lie conversation avec moi. J'avais commencé deux ans plus tôt une thèse d'état sous la direction d'André Lichnerowicz. Nous sommes toujours restées en relation depuis. Je crois que ce que Mademoiselle Libermann voyait en moi, c'était elle-même, vingt ans plus jeune : une mathématicienne, célibataire (je ne me suis mariée que dix ans plus tard), « d'origine juive » comme écrivent les journalistes, c'est-à-dire juive mais peu ou pas pratiquante. Pourtant, plus tard, j'ai su que Paulette jeûnait à Kippour et ne mangeait que des matsot à Pessah, et nous savons aussi que pendant de nombreuses années, elle avait plaisir à fêter Rosh Hashanah chez nous et à partager notre seder. Cette année, à Pessah déjà, elle a jugé qu'elle n'était pas en état de venir. C'était après une petite intervention à la jambe et une chute où elle s'était cassé l'épaule. Son épaule s'est remise et elle a pu assister à quelques séminaires. Elle faisait toujours des projets de voyages pour les divers congrès qui l'invitaient. Nous savons, hélas, la suite ; après son opération en avril, sa santé a décliné et, de Clinique de Turin en Villa Beausoleil, tout en gardant sa conversation toujours en alerte sur les événements du monde mathématique, elle a perdu ses forces et s'est éteinte au terme d'une longue carrière d'exception.

Sa vie se partageait entre sa famille, dont elle me parlait beaucoup – elle avait même une arrière petite-nièce – et le monde mathématique.

Son œuvre mathématique tient sa place dans la géométrie différentielle de la deuxième moitié du vingtième siècle ; les notions qu'elle a introduites ont toujours fait l'objet de développements ultérieurs et parfois ... ont été redécouvertes ! À Paris, elle a animé un séminaire, d'abord avec Ehresmann, puis seule et, plus tard, nous l'avons dirigé ensemble, puis avec Jean-Pierre Francoise, et une année avec Pierre Cartier. Le livre de géométrie symplectique qu'elle a écrit avec Charles-Michel Marle est un ouvrage considérable, qui fait autorité et reste une excellente référence, toujours citée vingt ans après sa parution. Mademoiselle Libermann était de tous les congrès de géométrie symplectique et de Poisson ; sa silhouette menue, sa présence aimable et bavarde nous manquent et nous manqueront irrémédiablement.

Ahmet Hayri Körezlioğlu

(1930 – 2007)

A. Suleyman Ustunel

Ahmet Hayri Körezlioğlu, appelé « Korez » par les intimes est décédé le 26 juin 2007 à Ankara. Korez a été un de ceux qui ont vraiment contribué à un élargissement des horizons des établissements où ils ont travaillé, à la fois en enseignement et en recherche. Comme c'était un ancien de l'ENST (Promo 1955) et comme il avait travaillé à la fois comme chercheur à METU (Middle East Technical University), à l'université de Paris, à Istituto di Fisica Teorica (Naples), au CNET et dans des entreprises comme Télécommunications Radioélectriques et Téléphoniques, il connaissait à fond les problèmes de la recherche et de l'enseignement français, qu'ils soient de nature administrative ou structurelle.

Plutôt que de donner les activités de Korez en ordre chronologique, nous allons les exposer suivant leurs importances. Tout d'abord Korez a été l'instigateur de l'enseignement moderne des probabilités à METU (Ankara) et à l'ENST : en effet, à l'époque où il était jeune docteur (ès Mathématiques), les probabilités françaises commençaient tout juste à prendre leur essor après la longue période de maturation de l'analyse fonctionnelle qui avait commencé avec E. Borel, M. Fréchet, J. Hadamard, suivis par l'École Bourbaki (J. Dieudonné, L. Schwartz, H. Cartan, etc). Aux États-Unis, Feller venait d'écrire ses deux volumes légendaires qui ont été suivis de quelques années par le superbe livre de J. Neveu et les travaux de P.-A. Meyer. À cette époque, Korez a organisé plusieurs groupes de travail pour étudier les ouvrages de Neveu et de Meyer et il a réussi à attirer de jeunes chercheurs dans les domaines des probabilités et des processus stochastiques. On peut dire que « les Bases Mathématiques du Calcul de Probabilités » de Neveu et « Probabilités et Potentiel » de Meyer ont été ses livres de chevet ainsi que ceux de ces jeunes chercheurs. Grâce à ces activités, les enseignements des probabilités et de l'analyse à METU et à l'ENST ont pris un caractère beaucoup plus moderne et il a eu plusieurs élèves qui ont fait de la recherche pour porter plus loin le flambeau allumé par lui (U. Capar, J. Szpirglas, G. Mazziotto, A.S. Ustunel, M. Kocatepe, G. Lefort, C. Martias, G. Loubaton, A. Saida, N. Privault, D. Kofman, L. Decreusefond, etc).

Ces activités ont duré de 1970 jusqu'à 1980, l'année où il a pris en main le groupe de probabilités au CNET. Il faut se souvenir qu'à l'époque, le sens du mot recherche n'était pas vraiment connu en France : le parti socialiste avait oublié de préparer un rapport pour l'état de la recherche et de l'enseignement supérieur français ; un rapport a été demandé ultérieurement à Laurent Schwartz qui a fait un superbe compte rendu de la situation (et celui-ci est toujours d'actualité). Korez a commencé par organiser des séminaires hebdomadaires au CNET sur les files d'attente et sur le filtrage des processus aléatoires (son sujet favori) où nous avons vu passer les chercheurs les plus importants dans ces domaines : M. Zakai, E. Wong, V. Beneš, G. Kallianpur et beaucoup d'autres encore. En ces années-là, il a aussi organisé une conférence sur les processus à deux paramètres à l'ENST, en collaboration avec Mazziotto et Szpirglas où toutes les sommités mondiales se sont retrouvées. Les actes de cette conférence qui ont été publiés chez Springer

(*Lecture Notes in Mathematics*, vol. 863) sont toujours une des références les plus consultées du domaine.

En 1985 il a été appelé à l'ENST pour la création du département réseaux qui a été unifié plus tard avec le département informatique sous l'appellation infres. Nous le voyons, avec son enthousiasme débordant, organiser des séminaires, des conférences, des groupes de travail sur les réseaux, sur les files d'attente avec Bacelli et Brémaud, sur l'analyse stochastique, etc. À cette époque, Paul Malliavin venait d'inventer « Le Calcul de Malliavin » et évidemment Korez a organisé un groupe de travail en 1985 pour étudier celui-ci. On peut dire que toute une génération de chercheurs français (A. Millet, M. Pontier, S. Tindel, N. Privault, T. Choukri, etc) ont trouvé leur voie grâce à ces séminaires. En particulier, il a organisé la première école d'été de Silivri sur l'Analyse Stochastique en 1986, entièrement consacrée au calcul de Malliavin où il avait invité tous les noms importants comme H. H. Kuo, P. Krée, D. Ocone, D. Nualart, M. Sanz, etc. Toute une génération de 68 a appris le calcul de Malliavin grâce à cette école et les actes qui ont été publiés chez Springer (*Lecture Notes in Math.*, vol. 1316) sont encore une des meilleures sources pour apprendre le sujet.

Après avoir fait un long séjour à Wuhan où il a pris en main la thèse de N. Privault, Korez a rendu visite plusieurs fois à G. Kallianpur à l'université de North Carolina at Chapel Hill où il a travaillé sur le problème du filtrage en deux paramètres ; il l'a caractérisé comme un problème à un paramètre mais de dimension infinie, de sorte qu'il pouvait utiliser les techniques d'analyse fonctionnelle. Il a organisé à nouveau l'École de Silivri en 1988, 1990 et en 1992 où nous avons pu assister à la démonstration du théorème de Atiyah-Singer (version probabiliste, par S. Watanabe), à la théorie des formes de Dirichlet (par F. Hirsch et par M. Röckner), aux processus de branchement (par D. Dawson), aux équations stochastiques aux dérivées partielles (par J. B. Walsh, E. Pardoux et P. Kotelenetz), aux mathématiques financières (D. Duffie), aux phénomènes de transition de phases dans les réseaux (V. Anantharam), etc. Pendant ces séminaires, Korez a fondé des collaborations fructueuses avec R. Mazumdar sur les réseaux, avec W. Runggaldier sur les méthodes de discrétisation dans le filtrage non linéaire sur lequel il a produit de nombreux papiers.

Dans les années 1994, il a fondé le projet Medcampus, grâce auquel il a créé des liens étroits avec des mathématiciens de Méditerranée qui a contribué en particulier au développements des mathématiques appliquées dans les pays du Maghreb.

Evidemment, pour un tel homme la vie de retraité n'a aucun sens ; on le voit en 1998 à la tête de l'Institut de Mathématiques Appliquées à METU où il a élevé de jeunes étudiants dans les domaines des mathématiques financières et de l'assurance. En particulier, la conférence qu'il a organisée à Antalya en 2006 sur les mathématiques financières a trouvé du retentissement partout dans le monde, malheureusement il est parti sans avoir vu les actes de celle-ci.

COURRIER DES LECTEURS

À propos d'un texte de P.-A.Meyer

En janvier 1998, le probabiliste Paul-André Meyer avait rédigé un texte où il jetait un regard sur sa carrière passée avec humour et profondeur. Ce texte fut publié en 2006 en complément du Séminaire de Probabilités XXXIX dédié à la mémoire de Paul-André Meyer, et contient un passage sur lequel je désirerais revenir brièvement. Il s'agit d'une critique de ce qu'on appelle les Social Sciences.

Commençons par citer ce qu'écrit P.- A. Meyer.

« Je devrais être un adhérent des Social Sciences à la mode, tant j'ai été conscient toute ma vie du caractère collectif de la science, du mouvement souterrain d'informations qui se trouvent à point nommé là où il faut (...) du caractère injuste de beaucoup d'attributions, du caractère fragile de la connaissance (...). Il me faut expliquer pourquoi je n'y adhère pas. Je sais parfaitement ce qu'on peut dire de la Science. Les gens lui demandent beaucoup trop, mais moi non. Non seulement j'ai depuis longtemps renoncé à tout savoir, mais je me suis persuadé que si je savais tout, je ne saurais pas grand chose. Cependant, mon enthousiasme pour la science n'a pas diminué, et je suis heureux d'avoir participé à la vie scientifique, même de façon relativement marginale. L'histoire des sciences telle qu'elle est racontée dans les Social Sciences est dépourvue de toute grandeur humaine.

Elle ne s'exprime qu'en termes de pressions sociales, réseaux d'influences, copinages et hostilités. Je me rappelle avoir voyagé entre Paris et Strasbourg dans le même compartiment que trois profs. Pendant tout le trajet ils n'ont parlé qu'avancement : proviseurs, syndicats, commissions, concours. C'est l'univers des Social Sciences. Moi qui ai été un étudiant-2CV des années 50, j'y reconnais les étudiants-Austin des années 70, nés dans un monde pauvre en postes. Les Social Sciences ne sont que l'expression des classes moyennes frustrées. »

Avant de faire quelques commentaires sur ce texte, il est sans doute nécessaire d'expliquer en deux mots ce que sont ces Social Sciences décriées par Meyer. Il y a d'ailleurs une ambiguïté dans cette expression, et je suppose vu ce qui en est dit que Meyer a en tête le mouvement de pensée connu sous le nom Social Studies of Knowledge. Née dans les années 1960 dans les pays Anglo-Saxons, cette école d'histoire des sciences désirait utiliser une méthodologie qui s'était progressivement imposée en histoire générale dans la première moitié du 20^e siècle. Promue par de grands mouvements de pensée (tels l'« école des Annales » de Bloch et Fèvre), cette méthodologie avait remis en cause l'univers souvent aseptisé de l'histoire positiviste (les Grands Hommes, les grandes dates...)

qui prévalait depuis des lustres. Il s'agissait de montrer au contraire qu'une société était une sorte d'organisme dont les mouvements et les choix s'expliquaient en termes de contacts et d'échanges, et la tâche de l'historien était de mettre ces derniers à nu pour en montrer la teneur et le fonctionnement. Ceci devint aussi du coup un objet majeur de l'histoire des sciences : faire apparaître les réseaux, les voies d'échanges, parfois (rarement) les combines, mais souvent les stratégies. Au lieu d'être un domaine éthéré peuplé de savants déconnectés de la vie réelle, le milieu scientifique se trouvait en prise directe avec le gouvernement de la cité, détenteur d'un pouvoir dont il ne pouvait prétendre se désengager. Il y a quelques années, la *Gazette* a publié un texte à ce sujet¹, et j'y renvoie le lecteur intéressé.

L'argument donné par Meyer sur les classes moyennes frustrées est sans doute partiellement exact, dans la mesure où les années creuses en postes avaient vermoulu l'ensemble du système universitaire, et installé certaines formes de pratiques revendicatives ronronnantes. Néanmoins, je doute qu'on puisse décemment tout expliquer par ce fait. Les années 1960 et 1970 furent aussi celles d'une « première mondialisation » qui prit la relève de la période des guerres mondiales et des guerres coloniales, une ouverture au monde à suffisamment grande échelle pour que la conscience des réseaux devienne plus forte. Or, les mathématiciens ont pendant longtemps prétendu vivre dans un monde à part, en se décrivant parfois dans des termes proches de ceux qu'un romantisme décadent employait avec délectation au sujet des poètes inspirés par leur muse (relire les lignes enflammées de Lautréamont sur

les mathématiques en donne une assez bonne idée!). Simone Weil fustigeait cette fâcheuse tendance : « *Il y a un mathématicien qui compare volontiers la mathématique à une sculpture dans une pierre particulièrement dure. (...) Mais si l'on a la vocation d'être sculpteur, il vaut mieux être sculpteur que mathématicien.* » Et au fond, les coups de boutoir, même pas toujours très maîtrisés, des Social Sciences ne sont-ils pas aussi une manière de dire aux scientifiques qu'ils sont des éléments parmi d'autres des sociétés humaines et non des OVNI, et de les inviter à prendre la parole ? J'ai déjà eu l'occasion de publier il y a trois ans dans la *Gazette*² deux textes sur les mathématiques financières, ou plutôt sur la façon peu prudente dont un certain nombre de collègues se sont précipités sur elles en les clamant inoffensives. Or je persiste et signe : il y a une idéologie véhiculée quand on parle à longueur de papiers de rentabiliser un portefeuille ou d'estimer la valeur d'une action. Que cette idéologie soit bénéfique ou maléfique est un autre débat. Prétendre l'évacuer sous couvert de la « neutralité de la science » est une escroquerie intellectuelle. Une jolie illustration me semble donnée par le slogan dont l'université de Rutgers a orné son affiche sur les formations doctorales en mathématiques : « Use the power of mathematics to change the world ». Sous une naïveté brute, il a le mérite de faire apparaître sans détour que les mathématiques sont un outil de pouvoir.

Paul-André Meyer regrettait l'absence de grandeur humaine dans les tableaux offerts par les Social Sciences. En fait, cette expression semble pour le moins très ambiguë. Je ne pense pas (surtout avec ce qu'il dit à côté) qu'il signalait par là uniquement son regret

¹ Octobre 2004, n° 102.

² Janvier et octobre 2003, n° 95 et 98.

qu'une plus grande place ne soit accordée aux « héros ». La réserve qu'il exprimait venait probablement plutôt du fait que certains tenants purs et durs des Social Sciences ne semblaient laisser aucune initiative, aucune marge d'action, aucun libre arbitre aux hommes de science. Ceux-ci seraient entièrement soumis aux glissements induits par la tectonique sociale, et rien ne se passerait au niveau de l'individu. Une telle position est naturellement intenable : il y a de « grands hommes » qui font de « grandes choses », des destins singuliers, des talents fulgurants. Mais la grandeur humaine, plutôt que de se résumer à la somme de ces derniers ne se trouve-t-elle pas tout autant dans de petites conquêtes au quotidien qui montrent la force créative que l'homme tire de se vivre en société? En fait, il semble impossible de séparer les deux aspects, qui ne sont donc en rien contradictoires mais bien plutôt complémentaires car l'histoire des hommes est un entremêlement permanent d'événements singuliers s'inscrivant dans de grands mouvements. Pour ce qui est d'une discipline comme les mathématiques, leur bonne santé me paraît bien mieux révélée par une multiplicité de représentants « moyens » que par l'existence de telle ou telle personnalité starifiée par le système, quoi qu'en disent certaines présentations quelque peu hollywoodiennes qui confondent « grandeur humaine » et « médiatisation ».

Il est certain que comme tout nouveau courant, les Social Studies n'ont

pas été à l'abri de certains abus, et c'est sans doute ceux-là qui ont heurté P-A. Meyer. Dans les années 1970, des personnes furent légitimement troublées de voir des gens qui prétendaient mettre en place une histoire des sciences « non positiviste » utiliser à cette fin des méthodes étrangement dogmatiques, à coup d'affirmations péremptoires et d'anathèmes répétés. La transposition inconditionnelle des techniques de l'analyse sociale ne fut pas étrangère à certains de ces débordements, mélangeant allègrement toutes sortes de considérations idéologiques au détriment d'une légitime prudence. Ainsi en fut-il pour certains auteurs dont l'obsession fut de ne lire l'histoire récente des sciences qu'à travers l'emprise du complexe militaro-industriel sur la société. Notons cependant que cette obsession fut aussi celle de mathématiciens parfaitement « purs » dont il est difficile de croire que Meyer les aurait inclus dans les Social Scientists. Pour ce qui est de l'analyse sociologique, il faut là aussi tâcher d'avoir un recul d'historien. Cette façon de concevoir l'histoire des sciences fut surtout une réponse et un contre-pied à une vision internaliste qui dominait et domine encore souvent, spécialement parmi les mathématiciens qui se vivent comme des êtres « à part ». C'est pour éviter cette tendance que les études des Social Studies peuvent être intéressantes. Le point de vue de Meyer doit ainsi être nuancé.

Laurent Mazliak
Université Paris VI

LIVRES

Non-linear elliptic equations in conformal geometry

SUN-YUNG ALICE CHANG

European Mathematical Society, 2004. 100 p.

ISBN : 978-3-03719-006-7. 24€

Une transformation conforme est un difféomorphisme qui préserve les angles (en tout point la différentielle étant par conséquent la composée d'une rotation et d'une homothétie). Dans son sens original, la géométrie conforme étudie les propriétés géométriques qui sont préservées par les transformations conformes. Ce sujet de recherche a connu de nombreux développements au cours des 30 dernières années. Le livre d'Alice Chang forme une version révisée d'un cours qui a été donné pour les étudiants de niveau Master 2 de l'ETH de Zürich. Il constitue en quelque sorte une introduction à la géométrie conforme. L'approche développée dans ce livre est essentiellement analytique : c'est l'un des attraits et l'une des richesses de cet ouvrage, montrer comment aborder des problèmes de géométrie conforme en utilisant des techniques d'analyse. Nous allons donner quelques exemples afin de mieux faire comprendre de quoi il s'agit.

Deux métriques g et g' sur une variété compacte et lisse (M, g) de dimension $n \geq 2$ sont dites conformes si $g' = e^\varphi g$ où φ est une fonction lisse. La classe conforme de g est donc l'ensemble de toutes les métriques pouvant s'écrire de cette façon. Il est maintenant classique que l'opérateur de Laplace-Beltrami Δ_g est un opérateur géométrique naturel au sens où celui-ci véhicule de nombreuses propriétés géométriques (on prendra garde au fait qu'ici l'opérateur de Laplace-Beltrami est l'opposé de celui considéré par les analystes, de telle façon que toutes ses valeurs propres sont positives ou nulles). Si on prend le cas d'une surface l'opérateur de Laplace-Beltrami est conformément covariant au sens où pour g et g' deux métriques conformes ($g' = e^\varphi g$) on a $\Delta_{g'} = e^{-\varphi} \Delta_g$. De plus on a une relation très simple qui relie la courbure de Gauss K_g pour la métrique g à la courbure de Gauss $K_{g'}$ pour la métrique g' :

$$\Delta_g \varphi + K_g = K_{g'} e^\varphi.$$

En intégrant cette relation sur la surface on s'aperçoit immédiatement que la quantité $\int K_g dv_g$ est constante sur une classe conforme donnée. En fait, la formule de Gauss-Bonnet nous dit que c'est même un invariant topologique.

En dimension plus grande que 2, le même genre de relation d'invariance conforme est vraie mais pour un opérateur (d'ordre 2) légèrement différent du laplacien, opérateur appelé laplacien conforme $L_g = \frac{4(n-1)}{n-2} \Delta_g + \text{Scal}_g$. On a, si $\tilde{g} = e^{2w} g$, pour tout $\varphi \in C^\infty(M)$

$$L_{\tilde{g}}(\varphi) = e^{-\frac{n+2}{2}w} L_g \left(e^{\frac{n-2}{2}w} \varphi \right).$$

D'une façon plus générale, on dira qu'un opérateur A_g dépendant de la métrique est conformément covariant de bi-degré (a, b) si, sous le changement conforme de métrique $\tilde{g} = e^{2w}g$, les opérateurs $A_{\tilde{g}}$ et A_g sont reliés par la relation

$$A_{\tilde{g}}(\varphi) = e^{-bw}A_g(e^{aw}\varphi), \quad \forall \varphi \in C^\infty(M).$$

Un opérateur d'ordre 4 particulièrement intéressant fut découvert en 1983 par S. Paneitz. Cet opérateur défini sur les variétés de dimension 4 est donné par

$$P_g\varphi = \Delta_g^2\varphi - \operatorname{div}_g \left(\frac{2}{3} \operatorname{Scal}_g g - 2 \operatorname{Ric}_g \right) d\varphi.$$

Cet opérateur est conformément covariant de bidegré $(0, 4)$, i.e. si $\tilde{g} = e^{2w}g$

$$P_{\tilde{g}}(\varphi) = e^{-4w}P_g(\varphi), \quad \forall \varphi \in C^\infty(M).$$

L'opérateur de Paneitz sur les variétés de dimension 4 présente beaucoup de similitudes avec l'opérateur laplacien sur les surfaces. Sur une variété de dimension 4 on a $P_g w + 2Q_g = 2Q_{\tilde{g}}e^{4w}$, où Q_g est l'invariant de courbure défini par (on appelle souvent Q_g la Q -courbure) $Q_g = \frac{1}{12}(-\Delta \operatorname{Scal}_g + \operatorname{Scal}_g^2 - 3|\operatorname{Ric}_g|^2)$. L'analogie entre K et Q est encore plus flagrante si on considère la formule de Gauss-Bonnet. Sur une surface on a

$$2\pi\chi(M) = \int_M K_g dv_g,$$

où $\chi(M)$ est la caractéristique d'Euler-Poincaré de M . En dimension 4, on a

$$(1) \quad 4\pi^2\chi(M) = \int_M \left(\frac{1}{8} |\operatorname{Weyl}_g|_g^2 + Q_g \right) dv_g.$$

Il faut noter que, puisque $|\operatorname{Weyl}_g|_g^2 dv_g$ est invariant par changement conforme, c'est le terme $Q_g dv_g$ qui mesure le changement conforme. Notons aussi que compte tenu de la formule de Gauss-Bonnet et de l'invariance conforme de $|\operatorname{Weyl}_g|_g^2 dv_g$, la quantité $\int_M Q_g dv_g$ (que l'on note en général k_P) est un invariant conforme.

Le livre d'Alice Chang couvre ces sujets d'une manière introductive. La première partie de l'ouvrage porte sur l'étude de problèmes de géométrie conforme sur les surfaces, en particulier le problème de l'uniformisation. Bien que la première preuve qui en fut donnée relève de l'analyse complexe (preuve due à Poincaré, Koebe et Weil, et récemment une preuve simplifiée a été donnée par Demailly), la preuve qu'Alice Chang donne est celle due à Osgood, Philipps et Sarnak, preuve utilisant la fonction ζ associée aux valeurs propres de l'opérateur de Laplace-Beltrami et la recherche de métriques extrémales pour la fonctionnelle associée. La seconde partie de son ouvrage porte sur la généralisation à des opérateurs d'ordre supérieur ou égal à 4 (opérateur de Paneitz sur les variétés de dimension 4 par exemple). Là encore, en s'appuyant sur des travaux de Branson et Orsted elle introduit la fonction ζ associée à ces opérateurs et aborde les problèmes de recherche de métriques extrémales (par exemple la recherche d'une métrique extrémale ayant une Q -courbure constante). L'accent est mis, dans cette partie, sur le rôle central que joue l'invariant k_P dans ces problèmes. L'ouvrage aborde aussi, bien que de manière plus marginale, le cas des opérateurs différentiels de degré plus élevé et covariants conformes de bidegré (a, b) . Enfin la troisième partie porte sur l'étude des fonctions symétriques

liées à certains tenseurs. Par exemple si on regarde le tenseur de Schouten $A_g = Ric_g - \frac{1}{6}R_g g$ (où Ric_g est le tenseur de Ricci de g et R_g la courbure scalaire) sur une variété de dimension 4, on peut voir celui-ci comme un endomorphisme symétrique sur l'espace tangent de la variété et considérer ses 4 valeurs propres λ_i ; dans ce cas on définit la deuxième fonction symétrique par

$$\sigma_2(A_g) = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_1\lambda_4 + \lambda_2\lambda_3 + \lambda_2\lambda_4 + \lambda_3\lambda_4.$$

σ_2 possède des propriétés remarquables et l'ouvrage d'Alice Chang nous en propose quelques exemples (essentiellement en dimension 4, avec par exemple une preuve complète de l'existence d'une métrique à courbure de Ricci strictement positive dans une classe conforme donnée sous l'hypothèse que la courbure scalaire de la variété est strictement positive et que k_P est strictement positif).

L'ouvrage d'Alice Chang est écrit avec une très grande clarté et constitue une introduction idéale pour tous ceux qui voudraient s'initier à l'analyse non linéaire sur les variétés. L'ouvrage n'a comme seul défaut d'être trop court... Mais il semble que ce soit la règle dans cette collection.

Zidine Djadli,
Université Joseph Fourier, Grenoble

Random Fragmentation and Coagulation Processes

JEAN BERTOIN

University Press, Cambridge (UK), 2006. 288 p.

ISBN : 9780521867283. 40£

Les processus de fragmentation et de coalescence stochastiques décrivent des systèmes d'objets qui se disloquent ou coagulent de façon aléatoire et répétée au cours du temps. Le présent ouvrage est le premier à se concentrer sur l'étude de tels processus, où l'un de ces phénomènes a lieu (mais non les deux à la fois) d'une façon markovienne. Les hypothèses de base faites sur les modèles considérés sont que :

- Les différents objets ne sont représentés que par leur taille (masse, diamètre,...). Ceci exclut toute structure spatiale ou géométrique, et correspond à l'hypothèse de « champ moyen » en physique.
- Pour le cas de la fragmentation, les processus vérifient une propriété de branchement : les différents objets présents à l'instant t se disloquent de façon indépendante les uns des autres après t .
- Pour le cas de la coalescence, la façon dont un groupe d'objets coagule ne dépend que des caractéristiques de ce groupe (nombre et tailles des objets) et non du reste du système.

L'auteur donne un panorama des principaux modèles vérifiant ces hypothèses et dépendant d'un petit nombre de paramètres, en fonction desquels certains phénomènes typiques sont mis en évidence. Ainsi, Bertoin part de l'étude des chaînes de fragmentation (chapitre 1), où chaque objet reste dans le même état pendant un temps strictement positif avant de se fragmenter. Un outil crucial est que l'on dispose dans ce cas d'une notion discrète de généalogie du processus. Afin de procéder à la description de modèles plus complexes, où cette notion de généalogie disparaît, il est nécessaire de développer différentes notions de

partitions aléatoires, dont la notion de partition échangeable des entiers est la plus cruciale. Ceci est l'objet du chapitre 2. Sont alors introduites les fragmentations échangeables (chapitre 3), où les dislocations peuvent se produire sitôt après la formation des objets. Le chapitre 4 porte sur l'étude des processus de coalescence échangeables, dont la construction présente de fortes analogies avec celle des fragmentations échangeables. Enfin, le cinquième et dernier chapitre porte sur les coalescents stochastiques, où des paires d'objets fusionnent à des taux dépendant des masses des objets, et sur les relations existant entre ces processus et la célèbre équation de coagulation de Smoluchowski.

La littérature portant sur ces modèles et leurs ramifications forme un volume très important. Par souci de clarté et de pédagogie, l'auteur a choisi de présenter pour chacun de ces modèles les théorèmes fondamentaux et quelques applications, afin de permettre au lecteur de se faire facilement une idée des nombreuses méthodes impliquées dans leur étude. À la fin de chaque chapitre se trouve une abondante section de notes bibliographiques qui sont un précieux guide pour le lecteur désireux d'en apprendre plus sur un point particulier. L'ouvrage ne contient pas d'introduction à la théorie des probabilités, et s'adresse donc à un lecteur ayant déjà de bonnes notions dans ce domaine, disons de niveau M2. Le texte est cependant « auto-contenu », et rappelle au moment voulu les définitions et propriétés des chaînes de Markov de saut pur, des subordinateurs, des lois de Poisson-Dirichlet et de la théorie de Kingman des partitions échangeables, qui sont nécessaires à l'exposé.

Nous donnons un rapide aperçu linéaire des résultats. Dans le chapitre 1, les processus de fragmentation considérés sont tels que chaque fragment a un temps de vie strictement positif, au terme duquel il se brise en plusieurs morceaux, non-nécessairement en nombre fini. Ces processus peuvent se décrire par le biais de chaînes markoviennes branchantes, où les objets du système sont décrits par un élément de

$$\mathcal{S}^\downarrow = \{s = (s_1, s_2, \dots) : s_1 \geq s_2 \geq \dots \geq 0 \text{ et } \lim s_i = 0\}$$

donnant la suite de leurs tailles. À chaque paire (α, ν) où $\alpha \in \mathbb{R}$ et ν est une mesure finie sur \mathcal{S}^\downarrow portée par les suites s telles que $s_1 \leq 1$, Bertoin associe un processus de fragmentation *auto-similaire* d'indice α et de *mesure de dislocation* ν , dont la dynamique est la suivante. Chaque fragment de taille x laisse place avec un taux égal à $\nu(\mathcal{S}^\downarrow)x^\alpha$ à une famille de fragments de tailles $xs_j, j \geq 1$, où s est une suite tirée aléatoirement selon la mesure $\nu/\nu(\mathcal{S}^\downarrow)$. On note $(X(t), t \geq 0)$, où $X(t) = (X_1(t), X_2(t), \dots) \in \mathcal{S}^\downarrow$, le processus ainsi construit. Une définition différente, inspirée des cascades multiplicatives de Mandelbrot, en est également donnée, utilisant un système de marques aléatoires sur l'arbre généalogique constitué par l'ensemble des mots d'entiers. Bertoin base son étude des chaînes de fragmentations sur des méthodes issues de l'étude des marches aléatoires branchantes et des cascades multiplicatives. Ainsi, sous l'hypothèse *malthusienne* d'existence d'un $p^* \geq 0$ et de $p > 1$ tels que

$$\int_{\mathcal{S}^\downarrow} \left(1 - \sum_{i \geq 1} s_i^{p^*}\right) \nu(ds) = 0 \quad \text{et} \quad \int_{\mathcal{S}^\downarrow} \left(\sum_{i \geq 1} s_i^{p^*}\right)^p \nu(ds) < \infty,$$

l'auteur introduit une *martingale intrinsèque* et une notion de lignée généalogique marquée au hasard. Ces deux principaux outils permettent de déterminer le comportement des chaînes de fragmentation en fonction des paramètres α et ν . Lorsque $\alpha < 0$, les petits fragments tendent à être détruits plus vite, ce qui entraîne un emballement du système et une surprenante propriété de *formation de poussière* : tous les fragments sont de taille nulle au bout d'un temps fini. Quand $\alpha = 0$, les tailles des fragments décroissent à vitesse exponentielle, ce qui est mis en évidence par le comportement en temps grand de la mesure

$$\rho_t = \sum_{i \geq 1} X_i(t)^{\rho^*} \delta_{t^{-1} \log X_i(t)}.$$

Cette dernière tend à se concentrer en une masse de Dirac, tandis que dans l'échelle $t^{-1/2}$, les logarithmes des tailles des fragments se répartissent de façon gaussienne autour de la valeur typique, un phénomène déjà observé par Kolmogorov dans des cas simples. Dans le cas $\alpha > 0$, les tailles des fragments décroissent comme $t^{-\alpha}$, et non plus à vitesse exponentielle. Le chapitre se conclut par l'étude de martingales additives et les résultats de grandes déviations pour ρ_t qui s'en déduisent, permettant l'étude de fragments atypiquement grands ou petits.

Le chapitre 2 se penche sur les relations qu'entretiennent trois notions différentes de partitions (de masse, d'intervalle ou des entiers). L'élément le plus important est la théorie de Kingman, montrant qu'une partition aléatoire des nombres entiers, invariante par l'action des permutations, est déterminée par la loi des fréquences asymptotiques (masses) de ses blocs. Cette dernière est une loi sur l'ensemble \mathcal{P}_m des partitions d'une masse unité, c'est-à-dire le sous-ensemble de S^1 formé des éléments de somme au plus 1. Le chapitre décrit également une famille importante de partitions de masse aléatoires, obtenues via une partition de l'intervalle $(0, 1)$ induite par un subordonateur, ce qui permet en particulier de définir les lois de Poisson-Dirichlet à deux paramètres. Ce chapitre est une excellente introduction à ces lois classiques et aux manières de calculer avec elles, et contient également quelques résultats moins connus de continuité en loi pour les réordonnements des partitions de masse biaisés par la taille (Proposition 2.3) et pour les partitions échangeables en fonction de leurs fréquences asymptotiques (Proposition 2.9).

Dans le chapitre 3, Bertoin introduit les fragmentations échangeables, qui sont en quelque sorte l'analogie des chaînes de fragmentation du chapitre 1 mais dans le cas où la mesure ν peut être infinie. L'idée maîtresse est une méthode de « discrétisation de l'espace », consistant à jeter à l'origine des temps des marques aléatoires indépendantes U_1, U_2, \dots sur l'objet se fragmentant, puis à définir un processus échangeable à valeurs dans les partitions des entiers, dont les blocs à l'instant t sont constitués des indices des marques se retrouvant dans un même sous-objet de l'objet initial. Les masses des sous-objets se retrouvent en prenant les fréquences asymptotiques des blocs de la partition au temps t . Ceci impose néanmoins une restriction importante, qui est que les fragmentations ne peuvent être que conservatives ou dissipatives, c'est-à-dire que la masse totale des objets du système ne peut que décroître au cours du temps (on parlera à présent de masse des objets plutôt que de leur taille). Les fragmentations échangeables sont donc à valeurs dans les partitions de \mathbb{N} , et sont invariantes par l'action des permutations. Les fragmentations dites *homogènes* $(\Pi(t), t \geq 0)$ sont telles que sachant

$\Pi(t) = \pi$, l'évolution de Π au-delà du temps t s'obtient en intersectant (fragmentant) chaque bloc de π avec une copie indépendante de Π . La propriété cruciale est que l'étude de ces processus se ramène à celle des restrictions $\Pi|_{[n]}$ de Π aux n premiers entiers. L'étude des taux de transition de ces chaînes restreintes permet de classifier les lois des fragmentations échangeables homogènes en fonction d'un couple (c, ν) , où

- $c \geq 0$ est un paramètre d'érosion déterminant le taux auquel des singletons aléatoires se détachent de chaque fragment : ceci correspond à un phénomène de « fonte » continue des fragments, qui n'existait pas dans les cas étudiés au chapitre 1

- $\nu(ds)$ est une certaine mesure σ -finie sur \mathcal{P}_m gouvernant, comme au chapitre 1, le taux auquel un objet de masse x se disloque soudainement en fragments de masses xs_1, xs_2, \dots

Bertoin donne une construction poissonnienne des fragmentations homogènes, qui rappelle fortement la décomposition de Lévy-Itô des processus de Lévy, et montre la propriété de Markov des fragmentations de masses qui leur sont associées. L'évolution du fragment marqué est également étudiée, et se décrit comme un sous-ordinator multiplicatif dont les caractéristiques dépendent de façon simple de ν et c . En utilisant une notion de *ligne d'arrêt*, on peut enfin construire une famille de fragmentations échangeables auto-similaires d'indice α en partant de la construction des fragmentations homogènes, qui sont celles pour lesquelles $\alpha = 0$, et en effectuant des changements de temps cohérents dans l'évolution des différents fragments. Le chapitre s'achève par une loi des grands nombres pour la mesure empirique ρ_t généralisant celle prouvée au chapitre 1.

Le chapitre 4 étudie les coalescents échangeables, dont la caractérisation présente de nombreuses similarités avec les fragmentations échangeables. Bertoin commence par introduire le coalescent de Kingman, intervenant dans la limite du modèle de population de Wright-Fisher. Pour ce coalescent, toutes les collisions font intervenir deux fragments, pris uniformément au hasard parmi toutes les paires de fragments. C'est à nouveau un argument de restriction des partitions aux premiers entiers qui permet de faire partir ce processus d'un nombre infini de blocs. Plus généralement, un coalescent échangeable $(\Pi(t), t \geq 0)$ est un processus échangeable tel que sachant $\Pi(t) = \pi$, l'évolution de Π après le temps t s'obtient en coagulant les blocs de π à l'aide d'une copie indépendante de Π . Ici, la coagulation $\text{Coag}(\pi, \pi')$ d'une partition π par π' est la partition obtenue en numérotant les blocs de π en π_1, π_2, \dots et en prenant les réunions des π_i dont les indices sont dans le même bloc de π' . Comme pour les fragmentations échangeables, la loi d'un coalescent échangeable se caractérise par un couple (c, ν) , où

- $c \geq 0$ est le taux auquel chaque paire de blocs fusionne, et correspond à une partie « coalescent de Kingman », et

- ν est une mesure σ -finie sur \mathcal{P}_m , qui détermine les coagulations soudaines et simultanées : avec un taux $\nu(ds)$, on prend une partition échangeable π' avec fréquences asymptotiques \mathbf{s} , et on fait sauter le processus de l'état π à l'état $\text{Coag}(\pi, \pi')$.

Il est ensuite donné une représentation des coalescents dits « simples », pour lesquels ν est portée par les éléments \mathbf{s} tels que $s_2 = 0$, à l'aide de *flots de ponts*. Ici,

un pont est un processus croissant de $[0, 1]$ dans lui-même, valant 0 en 0 et 1 en 1, à accroissements échangeables. Par un procédé d'approximation, partant d'un coalescent où la mesure ν est finie, il est montré comment on peut construire un flot aléatoire $(B_{t,t'}, -\infty < t \leq t' < \infty)$ de tels ponts, tels que le processus des sauts de $(B_{0,t}, t \geq 0)$ vu comme processus à valeurs dans \mathcal{P}_m est le processus des masses du coalescent simple associé à ν . L'un des intérêts des coalescents est qu'ils peuvent être interprétés comme des modèles de population étudiés lorsque l'on retourne le temps. On peut ainsi interpréter le flot de pont dual, $(B_{-t',-t}, -\infty < t \leq t' < \infty)$ comme un modèle de population, décrit comme processus à valeurs mesures qui généralise les processus de Fleming-Viot, et dont les propriétés de *fixation* (situation où la population à l'instant présent descend d'un unique ancêtre) sont étudiées. Enfin, le chapitre se termine par l'étude du coalescent de Bolthausen-Sznitman, qui apparaît entre autres dans l'étude de verres de spin (modèle GREM), et dont le flot de ponts associé s'exprime en termes de ponts de subordinateurs stables, faisant intervenir des propriétés notables de « composition » et de dualité coalescence-fragmentation des lois de Poisson-Dirichlet.

Enfin, le chapitre 5 se penche sur l'étude de processus de coalescence d'objets massifs, où chaque paire d'objets de masses x, y fusionne en un unique objet à un taux $\kappa(x, y)$, pour une fonction κ symétrique et positive, ce qui peut être vu comme généralisation du coalescent de Kingman ($\kappa(x, y) = 1$). Cependant, les méthodes employées sont très différentes, du fait de l'absence de structure échangeable. *A priori*, un tel processus n'est correctement défini que si le système d'objets est fini. Le résultat fondamental de cette partie est que si le noyau κ vérifie des conditions de type « Lipschitz-local » : pour tout $a > 0$, il existe $c_a > 0$ tel que

$$|\kappa(x, y) - \kappa(x', y')| \leq c_a(|x - x'| + |y - y'|), \quad x, y, x', y' \in [0, a],$$

alors il est possible de construire un tel processus issu de n'importe quel état initial, avec masses $s_1 \geq s_2 \geq \dots$ de limite nulle et de somme finie, et ce processus s'obtient par approximation par des systèmes finis. Le chapitre se poursuit par l'obtention de limites hydrodynamiques lorsque le noyau κ est sous-multiplicatif, $\kappa(x, y) \leq xy$. Précisément, on considère le processus $(X^{[k]}(t), t \geq 0)$ issu de k objets de masse 1 et donnant à l'instant t les masses des différents objets contenus dans le système, rangés par ordre décroissant. Si

$$n_t^{[k]}(x) = \frac{1}{k} \text{Card} \{j \geq 1 : X_j^{[k]}(t/k) = x\}, \quad x \geq 1$$

est la concentration des objets de masse x à l'instant t/k , alors pour $0 \leq t \leq 1$, $n_t^{[k]}(x)$ converge dans L^2 lorsque $k \rightarrow \infty$ vers une fonction déterministe $(n_t(x), x \in \mathbb{N}, t \geq 0)$ qui est solution de l'équation de Smoluchowski :

$$\frac{dn_t(x)}{dt} = -n_t(x) \sum_{y \in \mathbb{N}} n_t(y) \kappa(x, y) + \frac{1}{2} \sum_{y=1}^{x-1} n_t(y) n_t(x-y) \kappa(y, x-y).$$

La preuve repose sur une construction remarquable du coalescent multiplicatif $\kappa(x, y) = xy$ en termes du graphe d'Erdős-Rényi, et un argument de couplage du coalescent sous-multiplicatif avec le coalescent multiplicatif. Le livre s'achève avec l'étude du coalescent additif $\kappa(x, y) = x + y$, que l'on peut construire, partant de n objets de masse 1, à l'aide d'un processus de forêts coalescentes à n sommets.

Ceci permet de construire le *coalescent additif standard*, issu au temps $-\infty$ d'objets de masses infinitésimales. La boucle est bouclée par le fait que le processus renversé en temps du coalescent additif standard est le processus des masses d'une fragmentation auto-similaire échangeable (!) dont les caractéristiques (α, c, ν) sont données explicitement.

Le lecteur aura sans doute pu se rendre compte de la grande richesse des méthodes intervenant dans l'étude des processus de fragmentation et de coalescence. Un ouvrage donnant une introduction à ces méthodes et des structures sous-jacentes était plus que nécessaire. Une remarquable efficacité pédagogique, un style concis et limpide et une bibliographie très fournie en feront certainement une référence indispensable pour tout étudiant ou chercheur en probabilités intéressé par l'étude de ces processus, qui interviennent dans de très nombreux domaines appliqués.

Gregory Miermont,
Université Paris-Sud