Mémoire présenté à l'Université Paris-Sud
pour l'obtention de l'habilitation à diriger des recherches

Spécialité : Mathématiques

Contributions à la prévision séquentielle de suites arbitraires :

applications à la théorie des jeux répétés et études empiriques des performances de l'agrégation d'experts

par

Gilles Stoltz

Chargé de recherche au CNRS, affecté à l'Ecole normale supérieure,
et professeur affilié à HEC Paris

Soutenu publiquement le 3 février 2011 devant le jury composé de

Elisabeth	Gassiat	Université Paris-Sud	Présidente
Gábor	Lugosi	ICREA / Universitat Pompeu Fabra	Examinateur
Pascal	Massart	Université Paris-Sud	Rapporteur
Eric	Moulines	Télécom ParisTech	Rapporteur
Sylvain	Sorin	Université Pierre et Marie Curie	Examinateur
Bruno	Sportisse	INRIA	Examinateur

et au vu du rapport également écrit par

Avrim	Blum	Carnegie Mellon University	Rapporteur

Remerciements

L'écriture des remerciements est un exercice fort délicat, infiniment plus redoutable que la rédaction du reste du manuscrit ou la préparation de la soutenance. Au moment de m'y lancer, conscient de tout ce que mes collègues, amis et famille m'ont apporté au cours des années passées, je prie tous ceux que j'aurais pu oublier dans les lignes qui suivent de bien vouloir accepter mes excuses.

Selon la tradition, je voudrais commencer par évoquer chacun des membres de mon jury, dans un ordre chronologique d'apparition.

Pascal, c'est toi dont j'ai croisé la route en premier, en l'an 2000 – déjà ! – pour mon mémoire de première année de l'ENS Cachan. Depuis, tu n'as cessé de veiller, discrètement mais efficacement, sur ma carrière, en me prodiguant des conseils tant stratégiques que scientifiques. A mes yeux, un des grands moments de nos aventures ensemble restera ces deux semaines d'oraux du concours B/L : je crois que j'ai rarement autant ri (ou dû me retenir de rire) pendant mes heures de travail.

Elisabeth, j'ai eu le plaisir de t'avoir comme enseignante au master d'Orsay, et je suis heureux, depuis quelques années, de partager avec toi un cours de ce même master. Cela nous a incité à nous pencher chacun sur les sujets de l'autre et à échanger ainsi des lectures mathématiques (nos notes de cours) et extra-mathématiques...

Gábor, le trimestre « 2001, l'Odyssée de la statistique » m'a permis de choisir comme thème de recherche l'apprentissage séquentiel, grâce à l'enthousiasme avec lequel tu présentais le domaine de la prévision des suites individuelles. J'ai déjà eu l'occasion, dans mes remerciements de thèse, de louer ta disponibilité et dresser une liste non exhaustive de tout ce que tu m'as enseigné – un certain style de rédaction, une volonté de clarté et de simplicité dans les exposés, le goût d'écrire des manuels. Pendant les années qui ont suivi la thèse, tu m'as également aidé à prendre mon envol, tout en continuant à m'adresser des recommandations bienvenues. Nous avons quand même pu écrire encore deux articles ensemble, dont celui qui est associé à mon meilleur souvenir mathématique : la preuve simple et constructive d'un théorème de Rustichini.

Sylvain, ma troisième année de thèse porte le sceau de nos rencontres, lors du groupe de travail informel que tu organisais pour tes étudiants. Grâce à toi (et à Tristan, Jérôme, Rida), j'ai pu comprendre, sur le tas, le point de vue et les résultats importants en théorie des jeux. Tu fais également partie de ces conseillers me procurant des retours périodiques sur mes travaux ; nos discussions récentes (avec Vianney, également) sur les liens entre approchabilité, caractère sans regret et calibration, ont été une source

importante d'inspiration et elles m'ouvrent de nombreuses perspectives.

Bruno, c'est sous ton égide (et grâce à la patte de Georges Oppenheim) qu'ont eu lieu mes premiers contacts avec les applications. Je me souviens encore de cette matinée et de ce repas en juillet 2005, où je vous ai présenté, à Vivien et toi, quelques stratégies simples d'agrégation d'experts. J'ai été heureux de constater peu après, dans vos travaux sur les méthodes d'ensemble en qualité de l'air (puis dans ceux que nous avons effectués avec Vivien et différents stagiaires de master à votre suite), que toutes ces stratégies que j'avais côtoyées dans un but essentiellement théorique pouvaient procurer également de réelles améliorations pratiques des performances.

Eric, je crois que mes premiers contacts avec ton humour décapant (et tes imitations désopilantes des vicissitudes de notre univers académique) ont eu lieu grâce à Aurélien, dans le cadre du groupe de travail IT-Stats. Je vous avais également présenté un jour, à Aurélien et toi, les résultats existant en minimisation du regret face à des experts composés dans le cadre des bandits à plusieurs bras. Il semblerait que dans un avenir très proche, nous nous remettions tous les trois à phosphorer sur les bandits, à continuum de bras cette fois-ci ! Mais pour l'heure, je te remercie d'avoir consacré une partie de tes congés de fin d'année à lire et commenter mon manuscrit.

Avrim, enfin, merci également d'avoir accepté le rôle de rapporteur et de vous être confronté à cette noble tradition française qu'est l'habilitation à diriger des recherches. Nos rapports ont été essentiellement par articles interposés jusqu'à présent ; cela changera peut-être dans le futur – en effet, vous n'avez pas pu venir assister à la soutenance, mais mon invitation à Paris tient toujours !

J'aimerais continuer par ceux sans qui je ne serais pas allé bien loin : mes compagnons de labeur académique (on les appelle également co-auteurs). Et je voudrais décerner quelques mentions spéciales, toujours dans un ordre chronologique...

Nicolò, cela a été un vrai plaisir, il y a quelques années, de t'avoir un mois à mes côtés comme professeur invité à l'Ecole normale supérieure. Nous avons bouclé efficacement après ma thèse nos travaux de calibration séquentielle du paramètre d'apprentissage des stratégies de pondération par poids exponentiels, et, si nous n'avons pas formellement écrit d'autres articles ensemble depuis ce moment, nous avons continué à interagir ensemble et à échanger des idées.

Vivien, le roi du `C++`, mon maître absolu en capacité de travail, j'ai beaucoup appris à tes côtés et un monde nouveau s'est ouvert à mes yeux : celui de la prévision de la qualité de l'air. Ton enthousiasme communicatif, ta rigueur, ta diligence font de toi un co-auteur (et co-encadrant de stages) de rêve ! A ce propos, j'ai beaucoup apprécié la contribution décisive des stagiaires successifs qui nous ont aidés à explorer le terrain vierge sur lequel nous nous sommes lancés : Boris, Sébastien et Karim. ... Sébastien, d'ailleurs, je salue ton courage d'avoir continué en thèse avec le petit jeune que je suis, plutôt qu'avec un grand chef expérimenté !

Shie, s'il fallait un mot pour te décrire, ce serait l'optimisme ! Il tranche avec mon scepticisme, mais cette alliance du feu et de la glace s'est révélée efficace. Tu es un hôte très attentif au confort de ses invités et c'est toujours un plaisir de te rendre visite à

Haïfa !

Vincent, c'est avec toi que j'ai partagé le projet le plus long et le plus chargé d'émotions de ma période post-thèse, ce fameux livre qui nous a tant enthousiasmé au début, qui a été notre cauchemar pendant quelques mois, et qui désormais, est un souvenir merveilleux, qui trône sur nos étagères (et que nous n'osons plus ouvrir, de peur d'y trouver une coquille). Pendant plusieurs années, nous avons pu confronter nos goûts mathématiques, nos styles de rédaction, nos arts du codage et même, nos préférences artistiques (il a bien fallu choisir une illustration de couverture !). Quelle riche et formidable aventure, scientifique et humaine ! Mais, je te l'avoue, je ne recommencerais quand même pas tous les jours...

Sébastien et Rémi, vous m'avez enseigné les bandits stochastiques : moi qui n'y connaissais rien alors, vous m'avez accueilli dans vos projets en cours. Merci beaucoup pour m'avoir fait plonger ainsi au contact de données stochastiques, elles me changent de mes suites individuelles. J'apprécie toujours mes voyages à Lille, au sein de votre équipe SequeL : on s'y sent bien !

Yannig et Marie, nous étions partis la fleur au fusil agréger des experts pour la consommation électrique et au bout de deux jours, nous avons réalisé que ce serait plus compliqué que prévu, avec ces fichus experts intermittents. Mais grâce à ton expertise métier, Yannig, à ton autonomie et ta détermination, Marie, nous avons pu les exploiter au-delà de nos espérances initiales. Des perspectives durables de collaboration s'ouvrent pour nous, Yannig, et je m'en réjouis.

Tomasz, tu été mon précepteur de macro-économie et j'ai pu plonger tout entier dans un univers totalement inconnu, avec ses propres codes et valeurs, pas toujours reliés uniquement aux théorèmes. Tout en lançant à la chaîne des tests du χ^2 d'adéquation à la loi de Benford, nous avons pu refaire le monde des dizaines de fois ; ton humour polonais m'a revigoré pendant l'hiver passé !

Enfin, j'adresse un clin d'œil à deux de mes co-auteurs en devenir : Odalric, qui a toujours un temps d'avance et que je comprends par conséquent à retardement, et Vianney, qui est imparable pour détruire d'un bon contre-exemple une théorie ou des définitions en devenir.

A vous tous, collègues chercheurs et collègues administratifs qui partagez mon quotidien dans mes deux laboratoires, à l'Ecole normale supérieure et à HEC Paris, je redis mon plaisir de travailler avec vous.

J'apprécie notamment notre ambiance détendue (mais très productive) au DMA, qui est comme une seconde famille (enfin, une seconde famille qui se recomposerait toutes les années ou presque, règle des dix ans oblige). Je voudrais avoir un mot plus particulier pour les directeurs qui s'y sont succédé, Marc Rosso, François Loeser, Olivier Debarre, et pour les responsables d'équipe de probabilités et statistique, Jean-François Le Gall, Wendelin Werner, Thierry Bodineau : vous avez tous beaucoup œuvré pour le développement de la statistique dans notre laboratoire (et avez su me réfréner parfois, toujours avec patience). Olivier, également, merci d'avoir accepté d'être le chef (comme tu détestes que je t'appelle) de notre équipe INRIA et de nous avoir rejoints !

J'adresse toutes mes félicitations à mes co-bureaux successifs, qui ont réussi à me supporter, parfois pendant plusieurs années : Thierry, un modèle de silence et d'efficacité pendant les heures de travail, Nicolas, un modèle d'énergie et d'enthousiasme, qui vous animerait un stade à lui seul, et Gérard, un modèle en relations humaines et professionnelles. Thierry et David, avant que vous ne quittiez lâchement tous les deux le DMA, nous en avons passé de bons moments à refaire le monde académique à notre bar habituel de la Contrescarpe...

Les locaux du DMA ont été ré-aménagés, et j'ai déménagé au premier étage, mais mon esprit est plein des souvenirs amassés au passage vert et des jeunes qui l'ont animé ! Patricia, tu as été l'amie des dossiers et rapports semestriels d'ANR sans fin, des organisations de colloques, de séminaires, de groupes de travail, des oraux de concours de B/L, etc. – nous en avons passé du temps à nous inquiéter ensemble de ces rouages qui font vivre un groupe de recherche, mais nous avons survécu (et bien plus). Tu es hélas partie à Nice, la vie est ainsi faite, ... mais une amusante coïncidence nous fait soutenir nos habilitations à quinze jours d'intervalle. Mathilde, Marie, Amandine : ces années passées au DMA ont bénéficié de votre touche féminine – des mots gentils et des sourires par-ci, des sapins de Noël et des dégustations de pâtisseries par-là, et hop ! vous avez pris chacune votre envol. Enfin, je termine par un mot pour l'inénarrable Philippe : j'ai bien mis deux ans à comprendre puis à apprécier ton humour au $\lfloor e^5 \rfloor$–ème degré !

J'ai eu le plaisir de découvrir depuis quelques années une ambiance différente mais tout aussi galvanisante à HEC Paris. Notre département d'économie et de sciences de la décision est un cocon très agréable dans lequel se fondre, grâce à ses membres. Parmi ces derniers, je voudrais saluer en particulier la clique des théoriciens des jeux, Nicolas, Tristan, Dinah et Marco, notre coordinateur bien-aimé, Gilles, et ma compagne d'enseignement, Veronika, avec qui nous nous sommes mutuellement accompagnés dans nos débuts (un peu à la manière de l'aveugle qui guide le boiteux), sans oublier mon co-bureau, Thomas, dont les thèmes de recherche sur le marketing de luxe me font rêver !

HEC Paris est une école plus petite et plus familiale que l'Ecole normale supérieure et il est très facile d'y nouer des liens d'amitié également avec les membres de l'administration. Ainsi, à vous toutes qui égayez mes journées et avec qui nous parlons de tout et de rien, et notamment, à Marie-Françoise, Delphine, Béatrice, Nathalie, Claudine, Stéphanie : merci ! Et à toi aussi, Jimmy, mais là, ce sont parfois également nos sorties nocturnes.

Les meilleurs pour la fin : je voudrais conclure ces longs remerciements professionnels par quelques pensées plus personnelles. Tout d'abord, pour ma famille et mes amis, que je sacrifie trop souvent sur l'autel du travail et qui l'acceptent souvent avec beaucoup de patience. Et je voudrais enfin dédier ce manuscrit à ceux qui illuminent tout particulièrement le quotidien : Bénédicte et Zaïna au DMA, Jérôme à la maison.

Liste des travaux

On présente cette liste sous trois formes : par type de publications ; par thèmes, la liste des thèmes reprenant le découpage en chapitres du présent manuscrit ; et par co-auteurs.

Classement par type de publications

Les articles issus de ma thèse sont [1, 2, 3, 4], ainsi qu'une partie substantielle de [5].

Articles publiés dans des journaux

[1] Gilles Stoltz et Gábor Lugosi : Internal regret in on-line portfolio selection. *Machine Learning*, 59:125–159, 2005.

[2] Nicolò Cesa-Bianchi, Gábor Lugosi et Gilles Stoltz : Minimizing regret with label-efficient prediction. *IEEE : Transactions on Information Theory*, 51:2152–2162, 2005.

[3] Nicolò Cesa-Bianchi, Gábor Lugosi et Gilles Stoltz : Regret minimization under partial monitoring. *Mathematics of Operations Research*, 31:562–580, 2006.

[4] Gilles Stoltz et Gábor Lugosi : Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior*, 59:187–208, 2007.

[5] Nicolò Cesa-Bianchi, Yishay Mansour et Gilles Stoltz : Improved second-order inequalities for prediction under expert advice. *Machine Learning*, 66:321–352, 2007.

[6] Gábor Lugosi, Shie Mannor et Gilles Stoltz : Strategies for prediction under imperfect monitoring. *Mathematics of Operations Research*, 33:513–528, 2008.

[7] Boris Mauricette, Vivien Mallet et Gilles Stoltz : Ozone ensemble forecast with machine learning algorithms. *Journal of Geophysical Research*, 114:D05307, 2009.

[8] Sébastien Bubeck, Rémi Munos et Gilles Stoltz : Pure exploration in finitely-armed and continuous-armed bandit problems. *Theoretical Computer Science*, 2010. A paraître.

[9] Shie MANNOR et Gilles STOLTZ : A geometric proof of calibration. *Mathematics of Operations Research*, 2010. A paraître.

[10] Gilles STOLTZ : Agrégation séquentielle de prédicteurs : méthodologie générale et applications à la prévision de la qualité de l'air et à celle de la consommation électrique. *Journal de la Société Française de Statistique*, 151(2):66–106, 2010. (Article invité suite à la réception du prix Marie-Jeanne Laurent-Duhamel).

Articles publiés dans des actes de conférences très sélectives

[11] Sébastien BUBECK, Rémi MUNOS, Gilles STOLTZ et Csaba SZEPESVÁRI : Hierarchical optimistic optimization. *In Proceedings of NIPS'08*, pages 201–208. Curran Associates, Inc., 2008. (Version longue intitulée : "$\mathcal{X}$–armed bandits").

[12] Gábor LUGOSI, Omiros PAPASPILIOPOULOS et Gilles STOLTZ : Online multi-task learning with hard constraints. *In Proceedings of COLT'09*. Omnipress, 2009.

Aux articles mentionnés ci-dessus s'ajoutent les versions préliminaires des articles [1, 2, 5, 6, 8], publiés respectivement dans les conférences COLT'03 ("Student paper award"), COLT'04, COLT'05 ("Student paper award"), COLT'07 et ALT'09.

Articles soumis pour publication

[13] Marie DEVAINE, Yannig GOUDE et Gilles STOLTZ : Forecasting the electricity consumption by aggregation of specialized experts ; application to Slovakian and French country-wide hourly predictions. 2010.

[14] Tomasz MICHALSKI et Gilles STOLTZ : Do countries falsify economic data strategically ? Some evidence that they might. 2010.

Aux articles mentionnés ci-dessus s'ajoute une version étendue de l'article de conférence [11].

Rapports techniques et thèse

Les articles [7, 13, 10] s'appuient sur les rapports techniques [MMS07, GMS08, DGS09], qui présentent chacun une étude plus fouillée relatant autant les bons que les mauvais résultats empiriques et dont les détails sont donnés dans la bibliographique générale de ce manuscrit. On peut par ailleurs retrouver dans cette dernière les références de ma thèse [Sto05].

Manuel d'enseignement

[15] Vincent RIVOIRARD et Gilles STOLTZ : *Statistique en action*. Vuibert, 2009. (Tome principal : 320 pages, auxquelles s'ajoutent 210 pages d'annexes téléchargeables librement).

Classement par thèmes

1. Etudes aux fondements de la prévision de suites individuelles : [2, 5, 12]
2. Interactions avec la théorie des jeux répétés : [1, 3, 4, 6, 9]
3. Applications des techniques d'agrégation séquentielle, avancées méthodologiques et études empiriques : [1, 7, 13, 10]
4. Bandits stochastiques à continuum de bras et autres travaux : [8, 11] et [14, 15]

Classement par co-auteurs

Exerçant dans des institutions académiques étrangères

Gábor Lugosi (ICREA et Universitat Pompeu Fabra, Barcelone) : [1, 2, 3, 4, 6, 12]

Nicolò Cesa-Bianchi (Universitá degli studi, Milan) : [2, 3, 5]

Yishay Mansour (Université de Tel Aviv) : [5]

Shie Mannor (Mc Gill University, Montréal puis Israeli Institute of Technology, Technion) : [6, 9]

Csaba Szepesvári (University of Alberta) : [11]

Omiros Papaspiliopoulos (Universitat Pompeu Fabra, Barcelone) : [12]

Exerçant dans des institutions académiques françaises ou un centre de R&D

Vivien Mallet (INRIA, équipe CLIME) et Boris Mauricette (stagiaire master 2 de l'Université Paris-Diderot) : [7]

Sébastien Bubeck et Rémi Munos (INRIA, équipe SequeL) : [8, 11]

Yannig Goude (EDF R&D) et Marie Devaine (stagiaire master 2 de l'Université Paris-Sud) : [13]

Vincent Rivoirard (Université Paris-Sud et Ecole normale supérieure) : [15]

Tomasz Michalski (HEC Paris) : [14]

Démarche de rédaction

Ce manuscrit reprend, de manière synthétique, les travaux que j'ai effectués au cours de ma thèse (septembre 2002 – mai 2005) à l'Université Paris-Sud puis comme chargé de recherche au CNRS (depuis octobre 2005). J'ai le plaisir d'être accueilli (depuis septembre 2004) par l'Ecole normale supérieure et d'effectuer également une partie de mon temps de recherche à HEC Paris (depuis septembre 2007).

Dans ce chapitre liminaire, je précise la démarche de rédaction suivie autant pour le présent manuscrit que celle adoptée pour les travaux réalisés depuis la fin de ma thèse.

Objectifs de ce manuscrit

J'ai essayé d'y décrire au moins autant le contexte dans lequel se situent mes travaux que les résultats techniques obtenus. En particulier, j'ai voulu expliquer en détails, aux statisticiens comme aux théoriciens des jeux, l'intérêt de la prévision de suites arbitraires, un domaine peu traité et même peu connu en France.

Pour les premiers, il peut apparaître comme un cadre méta-statistique de prévision, où l'on cherche à tirer le meilleur parti de plusieurs méthodes éventuellement stochastiques, elles. Pour les seconds, il s'agit non plus seulement d'étudier les équilibres d'un jeu mais de voir que les joueurs peuvent assurer la convergence des distributions empiriques des actions choisies vers un ensemble d'équilibres à préciser.

Plan d'ensemble

On s'intéresse essentiellement à la prévision séquentielle de suites arbitraires ; le chapitre 1 en présente différents modèles.

L'un porte sur des suites d'observations choisies au fil du temps par un adversaire réagissant aux prévisions et correspond ainsi à un jeu répété ; il est étudié plus en détails au chapitre 2.

Un second modèle est constitué par un cadre méta-statistique où les observations ne sont plus influencées par les prévisions du statisticien mais sont comme fixées à l'avance (on parle de suites individuelles) ; il forme le socle du chapitre 3, où les performances de l'agrégation de prévisions fondamentales procurées par des experts sont étudiées, tant sur le plan théorique qu'empirique.

Le chapitre 4 résume notamment des travaux toujours situés dans un cadre d'apprentissage séquentiel mais reposant cette fois-ci sur des hypothèses stochastiques : le problème des bandits stochastiques (à continuum de bras).

Enfin, chaque chapitre est conclu par une partie présentant des perspectives de recherche à court et moyen terme.

Chapitre 1 : Le(s) modèle(s) de la prévision séquentielle de suites arbitraires

Les deux modèles indiqués ci-dessus (jeu répété, agrégation d'experts) sont tout d'abord présentés chacun à part, avec en particulier dans les deux cas, la définition de la notion de stratégie de prévision et l'introduction d'un critère pour l'évaluation des performances de ces stratégies : le regret. Dans le cas de l'agrégation d'experts, le regret apparaît comme une difficulté d'estimation liée à la contrainte séquentielle, à équilibrer avec une erreur d'approximation, formée par la perte cumulée du meilleur expert (ou de la meilleure combinaison convexe constante des experts). Dans le cas d'un jeu répété, la notion de regret exhibée est plus discutable *a priori* et c'est pourquoi le chapitre 2 est en bonne partie consacré à sa justification.

On encapsule ensuite ces deux modèles dans un sur-modèle générique, déterministe, dans lequel il s'agit simplement de construire de manière non anticipatrice des combinaisons convexes de suites de pertes fixées à l'avance. Ce sur-modèle permet de présenter de manière unifiée différentes bornes sur le regret, dont on indique ensuite comment les instancier aux deux modèles de prévision séquentielle. On insiste notamment sur l'obtention de stratégies de prévision totalement adaptatives en l'ensemble des paramètres du problème (nombre d'échéances, étendue des pertes), qui ne requièrent plus aucune connaissance préalable ni aucune intervention humaine [5].

La majeure partie des travaux que j'ai menés ici l'ont été pendant ma thèse ou immédiatement après elle (à l'exception du retour aux sources ultérieur [12]), grâce à une dynamique portée par Gábor Lugosi et Nicolò Cesa-Bianchi. La description raisonnée des deux modèles de prévision (et du sur-modèle générique) provient de mes expériences (bonnes et mauvaises) comme orateur dans des séminaires ou conférences, de statistique ou de théorie des jeux, où aucun chercheur n'était familier du problème de prévision séquentielle de suites arbitraires.

Chapitre 2 : Interactions avec la théorie des jeux

Ce chapitre commence par rappeler les définitions fondamentales de la théorie des jeux, afin d'être également lisible par un statisticien ; il justifie ensuite assez longuement la notion de regret par l'obtention de résultats de convergence (en des sens à préciser) des distributions empiriques des actions choisies vers des ensembles d'équilibres, lorsque le regret est minimisé.

A l'issue de ma thèse, il restait un problème en suspens : dans un cadre d'observations imparfaites, Rustichini [Rus99] avait énoncé un résultat général de possibilité de minimisation du regret mais l'avait prouvé de manière abstraite et peu lisible, et surtout,

sans exhiber de stratégie explicite. Des preuves constructives et des stratégies associées avaient été fournies dans divers cas particuliers. Au cours de l'année 2006, avec Shie Mannor et Gábor Lugosi, nous avons uni nos compétences et nos résolutions partielles antérieures (par exemple [3]) pour aboutir au résultat désiré, c'est-à-dire à une stratégie simple et efficace de minimisation du regret valant pour le cas général d'observations imparfaites [6]. Nous avons alors continué avec Shie Mannor de nous intéresser aux liens entre jeux répétés et prévision séquentielle de suites arbitraires, notamment *via* l'étude de stratégies calibrées [9].

Sur tous ces sujets, il faut noter la contribution essentielle de Vianney Perchet, élève de Sylvain Sorin, qui a revisité et approfondi la plupart de nos travaux pendant sa thèse (achevée en juin 2010) ; c'est pourquoi l'énoncé de nos résultats est suivi, à la fin de chaque partie de ce chapitre, par une mise en perspective et une discussion des avancées postérieures qu'il a obtenues.

Chapitre 3 : Etudes empiriques des performances de l'agrégation convexe d'experts

Dans ma thèse, je n'avais quasiment pas exploré l'application des techniques d'agrégation d'experts à des problèmes réels et me contentais de voir le cadre de la prévision de suites arbitraires comme des mathématiques applicables. La notion d'experts n'a pris son sens que plus tard, lorsqu'il a fallu les construire avec des partenaires.

Ainsi, très peu de temps après ma soutenance, dès juillet 2005, j'ai été contacté par Vivien Mallet, de l'équipe CLIME de l'INRIA, pour étudier l'intérêt des méthodes d'agrégation pour la qualité de l'air. Nous avons avancé doucement pendant quelques mois, afin de préparer le terrain, et c'est ensuite grâce notamment aux stages de master 2 de Boris Mauricette (premier semestre 2007) et de Sébastien Gerchinovitz (premier semestre 2008) que nous sommes parvenus à une batterie d'avancées méthologiques et d'illustrations pratiques de leurs performances, publiées en partie dans [7], le reste étant en cours de rédaction pour un second article.

Parallèlement, Yannig Goude, désormais chercheur à EDF R&D, effectuait une thèse sur l'intérêt des méthodes de mélange en prévision de consommation électrique. C'est avec lui et une autre stagiaire de master 2, Marie Devaine, que avons ensuite (au premier semestre 2009) étendu et approfondi les résultats qu'il avait obtenus [13].

Le chapitre 3 résume toutes les études empiriques ainsi effectuées, de même que les avancées méthodologiques afférentes. Son écriture s'est fondée sur un article de survol [10] rédigé pour le *Journal de la Société Française de Statistique*.

Chapitre 4 : Un autre cadre d'apprentissage séquentiel

On regroupe dans le dernier chapitre du manuscrit les travaux qui ne trouvent pas leur place dans la ligne directrice de la prévision de suites arbitraires.

Parmi eux, on peut distinguer ceux qui portent sur un autre problème d'apprentissage séquentiel : les bandits stochastiques à continuum de bras. C'est notamment Sébastien Bubeck et Rémi Munos, de l'équipe SequeL de l'INRIA, qui m'ont introduit au sujet et

avec qui j'ai travaillé dessus, à partir de la fin de l'année 2008. Nos résultats [8, 11] sont présentés (assez brièvement toutefois) pas tant pour la profondeur de leur apport que pour les perspectives qu'ils ouvrent, formulées en termes d'adaptation des stratégies en les paramètres de l'environnement stochastique auquel elles font face.

Table des matières

CHAPITRE 1

Fondements de la prévision de suites arbitraires

Introduction. On se concentre sur un problème générique de prévision séquentielle, que l'on abordera sous un angle méta-statistique. Plus précisément, il s'agit de prévoir des observations $y_1, \dots, y_t$, qui viennent l'une après l'autre. On ne suppose pas que ces observations sont la réalisation d'un certain processus stochastique sous-jacent dont il faudrait estimer les caractéristiques afin d'en déduire une bonne manière de prévoir. Autrement dit, il ne s'agit pas d'un problème statistique.

Cependant, on suppose disposer d'un nombre fini de prédicteurs fondamentaux, indexés par $j = 1, \dots, N$ et qui à chaque échéance t, lorsqu'il s'agit de prévoir y_t, proposent chacun une prévision $f_{j,t}$. Ces prédicteurs peuvent, eux, reposer sur une modélisation stochastique et dériver de méthodes statistiques. L'objectif est d'agréger séquentiellement leurs prévisions $f_{j,t}$ afin d'en déduire une prévision finale $\widehat{y}_t$ la plus performante possible.

C'est en ce sens que l'on parle de cadre méta-statistique de prévision séquentielle. Ce chapitre en présente quelques résultats essentiels et tâche, autant que faire se peut, de les relier à des énoncés du cadre classique de la statistique. Cette présentation raisonnée est suivie d'un exposé des contributions mathématiques apportées aux fondements de cette théorie.

Table des matières

1.1 Présentation de deux cadres de prévision séquentielle

On commence par la présentation de la partie commune aux deux cadres ; ces derniers différeront par le fait que l'ensemble des prévisions $\mathcal{X}$ est supposé convexe ou non, ainsi que par l'existence d'une dépendance des observations en le passé ou non. Un bref historique du domaine sera proposé ensuite.

1.1.1 Socle commun aux deux cadres

L'objet du problème est la prévision séquentielle d'observations $y_1, y_2, \ldots$ issues d'un ensemble $\mathcal{Y}$, sur lequel on n'impose aucune condition particulière. Contrairement au cadre habituel de la statistique, nous ne supposerons pas que cette suite d'observations est la réalisation d'un certain processus stochastique sous-jacent ; il ne s'agit donc pas ici d'estimer au mieux les caractéristiques d'un tel processus afin d'en prévoir le comportement et d'en tirer des prévisions les précises possibles.

Ensembles d'observations et de prévisions. A chaque échéance t, le statisticien doit produire une prévision $\widehat{y}_t$ fondée sur les observations passées $y_1, \ldots, y_{t-1}$. Cette prévision appartient à un ensemble $\mathcal{X}$, qui peut être différent de $\mathcal{Y}$; un cas typique est celui où $\mathcal{X}$ est l'enveloppe convexe de $\mathcal{Y}$. Par exemple, il s'agit de prédire l'occurrence ou non d'un événement, soit $\mathcal{Y} = \{0, 1\}$, et à cet effet, le statisticien a le droit de proposer une probabilité de réalisation, soit $\mathcal{X} = [0, 1]$.

Evaluation de la qualité des prévisions. La prévision $\widehat{y}_t$ est ensuite comparée à l'observation y_t grâce à une fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}$; cette dernière est le plus souvent positive. On définit ainsi la perte cumulée du statisticien sur les T premières échéances comme

$$\sum_{t=1}^{T} \ell(\widehat{y}_t, y_t)$$

et on veut assurer que cette dernière est la plus faible possible.

Appel à des experts. Afin que ce problème de prévision non stochastique ait un sens, on considère que le statisticien est aidé dans sa tâche par des experts. Ces derniers sont des prédicteurs fondamentaux qui proposent à chaque échéance une prévision fondée sur l'observation du passé.

Plus précisément, ils sont en nombre fini N et on les indexe par $j = 1, \ldots, N$ (ou par i lorsque l'on a besoin d'une variable muette différente de j) ; l'expert j procure à l'échéance t une prévision notée $f_{j,t} \in \mathcal{X}$ et qui dépend de $y_1, \ldots, y_{t-1}$ et éventuellement d'autres informations auxquelles il aurait accès lui seul. Le statisticien peut alors former sa prévision $\widehat{y}_t$ en se fondant non seulement sur les observations passées $y_1, \ldots, y_{t-1}$ mais aussi sur les prévisions présentes et passées des experts, $f_{j,s}$ pour $1 \leqslant 1 \leqslant t$ et $j = 1, \ldots, N$. La considération des prévisions passées des experts est utile pour suivre l'intuition selon laquelle il est sage de faire d'autant plus confiance à la prévision présente d'un expert qu'il s'est montré efficace dans le passé.

Objectif et méthode. Dans la suite et pour des raisons que l'on justifiera, afin d'assurer que la perte cumulée du statisticien est faible, on voudra garantir, par exemple mais

pas seulement, qu'elle n'est pas tellement plus grande que celle du meilleur expert,

$$\min_{j=1,\dots,N} \sum_{t=1}^{T} \ell(f_{j,t}, y_t)\,;$$

on notera dès à présent que l'indice $j_T^\star$ de l'expert atteignant ce minimum change au cours du temps et qu'il ne peut être connu à l'avance en général.

Deux hypothèses sous-jacentes à fixer

Dans la description précédente, volontairement ambiguë, deux points sont à préciser : le fait que l'ensemble des observations $\mathcal{X}$ soit convexe ou non ; le fait que le processus génératif des observations y_t et les experts réagissent ou non aux prévisions du statisticien. On pourrait donc obtenir en principe quatre sous-cadres de prévision séquentielle, mais nous ne considérerons dans la suite que les deux cas extrêmes suivants : celui où $\mathcal{X}$ est convexe et où il n'y a pas de réaction aux prévisions d'une part, et d'autre part, celui où $\mathcal{X}$ est arbitraire et où processus génératif et expert tiennent compte des prévisions du statisticien.

1.1.2 Premier cadre : Agrégation convexe séquentielle

Dans ce cadre, $\mathcal{X}$ est convexe et tout se passe comme si la suite $y_1, y_2, \dots$ d'observations était fixée à l'avance mais révélée élément par élément. On impose également au statisticien de former à chaque échéance t une prévision $\widehat{y}_t$ obtenue comme combinaison convexe des prévisions $f_{j,t}$ des experts.

Experts et nature. Comme le processus génératif est indépendant du statisticien, il est identifié à la nature. Quant aux experts, ils sont appelés ainsi parce qu'en plus de pouvoir reposer sur des techniques statistiques, ils peuvent éventuellement reposer sur des informations contextuelles, utiliser des ressources numériques importantes, et même faire appel à une expertise humaine. En fait, on les traitera essentiellement comme des boîtes noires prédictives pour l'instant. Ensuite, dans la partie applicative (au chapitre 3), on explique bien sûr, pour chaque jeu de données, comment les experts ont été construits.

Notion de stratégie de prévision par agrégation convexe. Une telle stratégie $\mathcal{S}$ associe à l'information disponible au début de chaque échéance t (aux observations passées et aux prévisions passées et présentes des experts) un vecteur de mélange $\boldsymbol{p}_t = (p_{1,t}, \dots, p_{N,t})$, qui est ensuite utilisé pour former la combinaison convexe dans $\mathcal{X}$ suivante :

$$\widehat{y}_t = \sum_{j=1}^{N} p_{j,t} f_{j,t}\,;$$

Paramètres : un ensemble d'observations $\mathcal{Y}$ arbitraire, un ensemble de prévisions $\mathcal{X}$ convexe, N experts

Choix initial : la nature choisit une suite $y_1, y_2, \ldots$ d'observations de $\mathcal{Y}$

A chaque échéance $t = 1, 2, \ldots$,

1. Les experts, en se fondant sur le passé $y_1, \ldots, y_{t-1}$ et éventuellement sur des informations contextuelles propres, forment des prévisions $f_{j,t} \in \mathcal{X}$, pour $j = 1, \ldots, N$, révélées au statisticien ;
2. Le statisticien choisit un vecteur de mélange $\boldsymbol{p}_t \in \mathcal{P}$ et forme la prévision dans $\mathcal{X}$
$$\widehat{y}_t = \sum_{j=1}^{N} p_{j,t} f_{j,t} \,;$$
3. La nature révèle l'observation $y_t \in \mathcal{Y}$.

FIGURE 1.1. Le déroulement du problème d'agrégation convexe séquentielle.

on impose ici que les vecteurs de mélange soient choisis dans le simplexe $\mathcal{P}$ de $\mathbb{R}^N$, c'est-à-dire qu'ils vérifient les conditions

$$\forall i \in \{1, \ldots, N\}, \quad p_{i,t} \geqslant 0 \qquad \text{et} \qquad \sum_{j=1}^{N} p_{j,t} = 1 \,.$$

Le déroulement du problème est alors résumé à la figure 1.1.

Critère de qualité d'une stratégie : le regret

L'évaluation d'une stratégie $\mathcal{S}$ ne peut être effectuée de manière absolue : si tous les experts sont mauvais, il est vraisemblable qu'aucune stratégie de prévision par agrégation convexe ne pourra avoir de bons résultats. On retient donc un critère relatif qui quantifie la proximité de la précision de prévision d'une stratégie $\mathcal{S}$ à celle de la meilleure combinaison convexe des experts.

Définition du regret. A cet effet, on définit tout d'abord les pertes cumulées de $\mathcal{S}$ et de chaque vecteur de mélange $\boldsymbol{q} \in \mathcal{P}$ comme, respectivement

$$\widehat{L}_T(\mathcal{S}) = \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) = \sum_{t=1}^{T} \ell\left(\sum_{j=1}^{N} p_{j,t} f_{j,t},\, y_t\right)$$

$$\text{et} \qquad L_T(\boldsymbol{q}) = \sum_{t=1}^{T} \ell\left(\sum_{j=1}^{N} q_j f_{j,t},\, y_t\right) .$$

Le regret (convexe) de $\mathcal{S}$ sur les T premières échéances est alors la différence entre ces pertes cumulées,

$$R_T(\mathcal{S}) = \widehat{L}_T(\mathcal{S}) - \inf_{\boldsymbol{q}\in\mathcal{P}} L_T(\boldsymbol{q})\,.$$

Bien sûr, les quantités $\widehat{L}_T(\mathcal{S})$, $L_T(\boldsymbol{q})$ et $R_T(\mathcal{S})$ dépendent également des observations $y_1,\dots,y_T$ et des prévisions des experts même si, dans un souci d'allégement des notations, cette dépendance n'est pas explicitement rappelée.

Bornes sur le regret. Le regret $R_T(\mathcal{S})$ est de l'ordre au plus de T lorsque la fonction de perte est bornée. On recherche ici des stratégies telles que leur regret, rapporté au nombre d'échéances, tende uniformément vers 0, quelles que soient les observations et les prévisions des experts. C'est parce que l'on considère ici que toutes les suites possibles de $\mathcal{Y}$ peuvent se produire et que l'on exhibera des garanties de performance uniformes en ces suites que l'on parle de suites arbitraires ou encore, de *suites individuelles* d'une part, et d'agrégation robuste d'autre part.

Objectif 1.1. Construire des stratégies de prévision par agrégation convexe $\mathcal{S}$ minimisant le regret, c'est-à-dire telles que

$$\limsup_{T\to\infty}\ \sup\left\{\frac{R_T(\mathcal{S})}{T}\right\} \leqslant 0\,,$$

où le supremum porte sur l'ensemble des suites d'observations et de prévisions d'experts possibles.

Interprétation comme un méta-problème statistique

Equilibre entre erreur d'approximation et difficulté d'estimation. L'objectif précisé ci-dessus se rapporte à la minimisation du regret, alors que l'on rappelle que l'objectif initial est d'assurer que la perte cumulée du statisticien est petite. Or, la décomposition

$$\widehat{L}_T(\mathcal{S}) = \inf_{\boldsymbol{q}\in\mathcal{P}}\{L_T(\boldsymbol{q})\} + R_T(\mathcal{S})$$

indique que cette perte cumulée est la somme d'une erreur d'approximation, donnée par la perte cumulée de la meilleure combinaison convexe constante des experts, et d'une erreur d'estimation, donnée par le regret et qui mesure la difficulté à se rapprocher, à cause de la contrainte séquentielle, de la performance de cette meilleure combinaison convexe constante. On rappelle d'ailleurs à cet égard que la valeur du vecteur de mélange optimal (ou quasi-optimal) pour les échéances 1 à T peut fortement varier avec T, s'agissant de suites individuelles.

En pratique, il faudrait donc arbitrer entre la considération d'un nombre N suffisamment grand d'experts aux comportements suffisamment différents afin de rendre l'erreur d'approximation la plus faible possible, et le fait qu'évidemment le regret $R_T(\mathcal{S})$ croît avec N. Cependant, cette croissance est, comme on le verra, très modérée en général : elle est de l'ordre de $\sqrt{\ln N}$. On a donc souvent intérêt à considérer un nombre important d'experts.

Que sont les experts ? La question essentielle est alors de construire des experts ; pour l'instant, nous avions simplement formulé le problème en identifiant chaque expert à une boîte noire prédictive. Nous expliquerons sur les deux jeux de données considérés au chapitre 3 comment nous avons obtenu les experts mais illustrons par un exemple générique pourquoi le problème décrit dans ce manuscrit est un méta-problème statistique.

Dans un problème statistique classique où les observations (y_t) sont les réalisations d'un certain processus stochastique (Y_t), des méthodes stochastiques permettent d'obtenir des prévisions aléatoires ; on note $f_{j,t}$ la réalisation de la prévision de la j–ème méthode à l'échéance t, c'est-à-dire que l'on identifie cette méthode à un expert. Au lieu de sélectionner une méthode précise, on peut ici en considérer plusieurs et agréger leurs prévisions. Cette agrégation est effectuée, elle, de manière robuste, sans prendre en compte l'éventuel caractère stochastique des observations. L'avantage est que comme les méthodes stochastiques habituellement utilisées dépendent d'un ou plusieurs paramètres, on peut considérer pour chacune plusieurs instances obtenues avec des jeux de paramètres différents, ce qui rend le réglage précis des paramètres moins crucial.

En conclusion : On a considéré dans ce paragraphe l'agrégation robuste et non stochastique de prédicteurs fondamentaux qui, eux, peuvent reposer sur des techniques stochastiques. C'est en ce sens que l'on a affaire à un problème méta-statistique : on ne cherche pas à améliorer les performances individuelles des prédicteurs mais on vise à bien combiner leurs prévisions.

1.1.3 Second cadre : Prévision randomisée

Lorsque l'ensemble des prévisions $\mathcal{X}$ n'est pas convexe, il n'est pas toujours possible ni facile de tirer une prévision agrégée légale à partir des prévisions des experts. Une manière simple de procéder est de requérir que les stratégies de prévision choisissent un expert et suivent simplement sa prévision. On peut se persuader avec le contre-exemple suivant qu'il est nécessaire d'effectuer un tel choix d'expert au hasard : $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, une fonction de perte ℓ donnée par une distance sur $\{0, 1\}^2$ et deux experts proposant chacun, de manière constante au cours du temps, 0 ou 1. C'est pourquoi on parle de prévision randomisée.

Autre modification : réaction aux prévisions. On suppose par ailleurs que le processus génératif des observations réagit aux prévisions du statisticien : tout se passe comme si le statisticien jouait contre un adversaire qui lui aussi a une stratégie ; en fait, on peut même supposer que c'est également cet adversaire qui contrôle les experts et choisit leurs prévisions.

Tirages d'un expert au hasard. A chaque échéance, le statisticien choisit ici encore un élément $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ de $\mathcal{P}$, mais ce dernier est désormais interprété comme une probabilité et non plus comme un vecteur de mélange ; un indice d'expert est ensuite

Paramètres : un ensemble d'observations $\mathcal{Y}$ arbitraire, un ensemble de prévisions $\mathcal{X}$ également arbitraire, N experts

A chaque échéance $t = 1, 2, \ldots,$

1. L'adversaire, en se fondant sur les informations procurées par les tours passés, choisit une observation $y_t \in \mathcal{Y}$ et des prévisions $f_{j,t} \in \mathcal{X}$ pour les experts $j = 1, \ldots, N$;
2. Seules les prévisions des experts sont révélées au statisticien pour l'instant ;
3. Le statisticien détermine, en fonction d'elles et en fonction du passé, une probabilité $\boldsymbol{p}_t \in \mathcal{P}$, tire au hasard un indice d'expert I_t selon $\boldsymbol{p}_t$, et recourt à la prévision
$$\widehat{y}_t = f_{I_t,t} \,;$$
4. L'adversaire et le statisticien révèlent publiquement leurs choix, c'est-à-dire l'observation $y_t \in \mathcal{Y}$ et la prévision $\widehat{y}_t \in \mathcal{X}$ (de même que la probabilité $\boldsymbol{p}_t$ et l'indice I_t) ; ces quantités sont mémorisées par les deux protagonistes, qui pourront effectuer leurs choix futurs en se fondant sur elles.

Figure 1.2. Le déroulement du problème de prévision randomisée séquentielle.

tiré au hasard selon $\boldsymbol{p}_t$ et le statisticien prédit comme cet expert. Le déroulement plus précis du jeu de prévision est indiqué à la figure 1.2.

De nombreux aléas cachés. On note respectivement σ et τ les stratégies du statisticien et de l'adversaire. On ne les définit pas formellement dans ce cadre général mais on le fera dans un cadre de prévision randomisée simplifié au paragraphe 2.1 ; le dernier point de la figure 1.2 permet cependant, de manière certes un peu informelle, de cerner qu'elles associent à des informations passées des choix pour les échéances présentes. On se contente pour l'heure de souligner quelques difficultés à travers des exemples. Par exemple, le choix de y_t à l'échéance $t \geqslant 2$ dépend notamment de $\widehat{y}_1, \ldots, \widehat{y}_{t-1}$ et donc des variables aléatoires $I_1, \ldots, I_{t-1}$. Ainsi, même lorsque la stratégie τ est déterministe, c'est-à-dire qu'elle associe de manière déterministe à l'information disponible un élément de $\mathcal{X}^N \times \mathcal{Y}$ au premier point de la figure 1.2, les observations résultantes y_t sont des variables aléatoires. Ce raisonnement s'étend à toutes les quantités en jeu : prévisions $f_{j,t}$ des experts, probabilités de tirage $\boldsymbol{p}_t$.

Extension de la définition du regret

Ici encore, l'évaluation d'une stratégie σ ne peut être effectuée de manière absolue : si l'adversaire fait en sorte que tous les experts soient mauvais, il est vraisemblable qu'aucune stratégie de prévision randomisée ne pourra avoir de bons résultats.

Regret par rapport au meilleur expert, toutes choses égales par ailleurs. Formellement, les pertes cumulées du statisticien et de chaque expert j dépendent des stratégies σ et τ :

on les définit respectivement comme

$$\widehat{L}_T(\sigma,\tau) = \sum_{t=1}^{T} \ell(\widehat{y}_t, y_t) = \sum_{t=1}^{T} \ell(f_{I_t,t}, y_t)$$

$$\text{et} \qquad L_{j,T}(\sigma,\tau) = \sum_{t=1}^{T} \ell(f_{j,t}, y_t)\,.$$

Le regret de σ face à τ sur les T premières échéances est alors la différence entre ces pertes cumulées,

$$R_T(\sigma,\tau) = \widehat{L}_T(\sigma,\tau) - \min_{j=1,\dots,N} L_{j,T}(\sigma,\tau)\,. \tag{1.1}$$

Bornes sur le regret. Le regret $R_T(\sigma,\tau)$ est de l'ordre au plus de T lorsque la fonction de perte est bornée. On recherche ici encore des stratégies σ telles que leur regret, rapporté au nombre d'échéances, tende vers 0, quelle que soit la stratégie τ de l'adversaire. Cependant, il n'est en général pas possible d'assurer un contrôle du regret uniforme par rapport à ces stratégies τ.

Objectif 1.2. Construire des stratégies de prévision randomisées σ minimisant le regret, c'est-à-dire telles que

$$\sup_{\tau} \left\{ \limsup_{T\to\infty} \frac{R_T(\sigma,\tau)}{T} \right\} \leqslant 0 \qquad \text{p.s.},$$

où le supremum porte sur l'ensemble des stratégies possibles de l'adversaire (et où le caractère presque-sûr est relatif aux randomisations auxiliaires utilisées par le statisticien et, éventuellement, son adversaire).

Interprétation moins évidente que dans le cas de l'agrégation convexe

Problème : ce que le regret ne mesure pas. On veut souligner ici que la comparaison au meilleur expert est effectuée toutes choses égales par ailleurs, de manière rétrospective, ce qui pose un problème d'interprétabilité dans un cadre où l'adversaire réagit aux prévisions du statisticien. Formellement, on fixe un indice d'expert j. Sauf dans le cas où σ est elle-même la stratégie σ^j proposant à chaque échéance t, indépendamment du passé, la probabilité $\boldsymbol{p}_t = \delta_j$ donnée par la masse de Dirac en j, si le statisticien avait choisi à chaque tour un expert j fixé, il n'aurait en général pas obtenu la perte cumulée $L_{j,T}(\sigma,\tau)$ mais $L_{j,T}(\sigma^j,\tau)$. En particulier, ce que l'on préférerait borner, c'est l'écart entre

$$\widehat{L}_T(\sigma,\tau) \qquad \text{et} \qquad \min_{j=1,\dots,N} L_{j,T}(\sigma^j,\tau)\,;$$

mais ce n'est en général pas un objectif réalisable et c'est pourquoi l'on se contente de la définition $R_T(\sigma,\tau)$.

Solutions. [dFM03] propose de restreindre l'ensemble des stratégies τ possibles de l'adversaire à une classe de stratégies de rationalité bornée. Si en revanche on veut conserver l'ensemble des stratégies possibles, comme c'est notre cas, la justification ne sera plus intrinsèque et élémentaire, comme dans le cas de l'agrégation convexe ; c'est pourquoi nous proposons au chapitre 2 une justification de la notion de regret $R_T(\sigma, \tau)$ par des arguments de convergence vers des ensembles d'équilibres.

1.1.4 Bref historique de ces cadres de prévision séquentielle

La première mention des problèmes de prévision séquentielle de suites arbitraires (choisies par un adversaire ou non) remonte aux années 50, et plus précisément aux travaux de Hannan [Han57] et Blackwell [Bla56], deux statisticiens qui ont énoncé des résultats fondateurs en théorie des jeux dans ces articles. Cover [Cov65] propose la première étude des ordres de grandeur minimax du regret, dans le cas de prévisions de suites binaires. Il faut également citer l'étude de la compression de suites arbitraires de données en théorie de l'information, où les recherches d'avant-garde ont été menées par Ziv [Ziv78, Ziv80] et Lempel et Ziv [LZ76, ZL77] ; ils ont résolu la question de compresser une suite arbitraire de données presqu'aussi bien que le meilleur automate fini. Enfin, en théorie de l'apprentissage, l'introduction du problème de prévision séquentielle de suites arbitraires a été effectuée par Littlestone et Warmuth [LW94] et Vovk [Vov90] ; Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [CBFH$^+$97], Foster [Fos91], Freund et Schapire [FS97] et Vovk [Vov98] ont présenté quelques-un des résultats fondamentaux. Pour un exposé de l'état de l'art du domaine des années 50 jusqu'en 2006, on pourra se reporter à l'ouvrage de synthèse [CBL06].

1.2 Minimisation du regret par pondération par poids exponentiels

Ne pas tout miser sur le meilleur expert. Une stratégie naturelle mais qui échoue en général (au sens où son regret est de l'ordre de T) de est prédire à l'échéance t comme le meilleur des N experts sur les échéances 1 à $t-1$. Avec un peu de recul, on voit que le problème ici est que deux experts aux performances parfois très proches, en l'occurrence, les deux meilleurs experts sur le passé, ont des poids (ou des probabilités, selon le cadre) très différent(e)s, 0 pour le moins bon des deux et 1 pour le meilleur des deux. Une idée plus raisonnable est d'attribuer un poids (ou une probabilité) $p_{j,t}$ simplement d'autant plus grand(e) à l'expert j pour l'échéance t que ses performances ont été meilleures sur les échéances précédentes $1, \ldots, t-1$, sans qu'aucun(e) de ces poids (ou probabilités) ne soit nul(le).

1.2.1 Le résultat fondamental pour les deux cadres

Le lemme ci-dessous est l'un des résultats les plus fondamentaux et les plus connus en prévision de suites individuelles.

Cadre générique. On l'énonce dans un cadre générique, non stratégique, où l'on considère simplement des suites fixées de pertes et où l'on étudie les possibilités de contrôle d'un pseudo-regret défini en termes de vecteurs de mélange construits de manière non anticipatrice. On expliquera dans la suite, aux paragraphes 1.2.2 et 1.2.3, comment instancier les résultats présentés ici pour minimiser le regret tant dans le cadre de l'agrégation convexe séquentielle que dans celui de la prévision randomisée.

Références. Plusieurs versions en ont été données, par [LW94, Vov90, Vov98, CBFH$^+$97, FS97]. Nous reprenons ci-dessous une démonstration élémentaire suggérée par [CB99] et que l'on pourra retrouver également dans [CBL06, paragraphe 2.2]. La stratégie énoncée dans le lemme suivant est appelée la stratégie de pondération par poids exponentiels de vitesse d'apprentissage $\eta > 0$.

Lemme 1.3. On fixe deux réels $m \leqslant M$. Pour tout $\eta > 0$ et pour toute suite arbitraire d'éléments $\ell_{j,t} \in [m, M]$, où $j \in \{1, \dots, N\}$ et $t \in \{1, \dots, T\}$,

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant \frac{\ln N}{\eta} + \eta \frac{(M-m)^2}{8} T \,, \tag{1.2}$$

où pour tout $j = 1, \dots, N$, on définit $\mu_{j,1} = 1/N$ et pour $t \geqslant 2$,

$$\mu_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{i,s}\right)} \,. \tag{1.3}$$

Démonstration. La preuve repose sur le lemme de Hoeffding, dont on rappelle l'énoncé : si X est une variable aléatoire bornée, à valeurs dans $[m, M]$, alors pour tout $s \in \mathbb{R}$,

$$\ln \mathbb{E}\left[e^{sX}\right] \leqslant s\, \mathbb{E}[X] + \frac{s^2}{8}(M-m)^2 \,. \tag{1.4}$$

En particulier, pour tout $t = 1, 2, \dots$, en utilisant (pour le cas $t = 1$) la convention qu'une somme sur aucun élément est nulle,

$$-\eta \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} \geqslant \ln \frac{\sum_{j=1}^{N} \exp\left(-\eta \sum_{s=1}^{t} \ell_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{i,s}\right)} - \frac{\eta^2}{8}(M-m)^2 \,;$$

en sommant ces inégalités sur t et en divisant les deux membres par $-\eta < 0$, il vient

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} \leqslant -\frac{1}{\eta} \ln \frac{\sum_{j=1}^{N} \exp\left(-\eta \sum_{s=1}^{T} \ell_{j,s}\right)}{N} + \eta \frac{(M-m)^2}{8} T \,.$$

La preuve est alors conclue en minorant la somme de termes positifs restant dans le logarithme du membre de droite par le plus grand de ces termes. □

Meilleur choix théorique pour η à nombre d'échéances T fixé...

On verra dans la suite que le regret sera essentiellement majoré par des quantités de la forme du membre droit de (1.2). Il est donc important de pouvoir assurer que ce dernier est sous-linéaire.

Optimisation de la borne théorique. Lorsque l'échéance maximale T, de même que les bornes m et M, sont connues (le nombre d'experts N l'étant, lui, toujours), le choix de $\eta = (1/(M-m))\sqrt{(8\ln N)/T}$ minimise le membre droit de (1.2), ce dernier valant alors $(M-m)\sqrt{(T/2)\ln N}$. On note que d'une part, l'on n'a pas toujours de raison de connaître m et M et que d'autre part, T ne peut pas être fixé.

... Or, à terme, $T \to \infty$

Les objectifs 1.1 et 1.2 requièrent tous deux que le nombre T d'échéances tende vers l'infini ; mais pour tout choix fixé de $\eta > 0$, le membre droit de (1.2) est alors équivalent à une quantité linéaire et aucun de ces objectifs ne peut être rempli.

Choix adaptatif des vitesses d'apprentissage. Cela étant, il suffit d'autoriser les vitesses d'apprentissage à dépendre elles aussi des informations passées pour régler ces soucis. A cet effet, on définit les éléments $\mu_t \in \mathcal{P}$ donnés par les valeurs de leurs composantes j selon : $\mu_{j,1} = 1/N$ et pour $t \geqslant 2$,

$$\mu_{j,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_{i,s}\right)} \tag{1.5}$$

où la vitesse d'apprentissage $\eta_t > 0$ peut dépendre des éléments $\ell_{i,s}$ pour $s \in \{1,\dots,t-1\}$ et $i \in \{1,\dots,N\}$. En fait, il est même souhaitable que η_t ne dépende que de ces éléments et pas de connaissances *a priori* sur les valeurs de m ou M, ces dernières n'étant pas toujours disponibles.

Existence d'une stratégie convenable. Le résultat le plus important de [ACBG02] et [5] est formulé de manière un peu informelle ci-dessous ; on le détaille et l'énonce plus précisément au paragraphe 1.4.

Théorème 1.4. *Il existe une manière explicite de définir chacune des vitesses $\eta_t > 0$ uniquement en fonction des éléments $\ell_{i,s}$, avec $s \in \{1,\dots,t-1\}$ et $i \in \{1,\dots,N\}$, de telle sorte que la stratégie (1.5) assure le contrôle uniforme suivant. Pour tous réels $m \leqslant M$, pour toute suite arbitraire d'éléments $\ell_{j,t} \in [m,M]$, où $j \in \{1,\dots,N\}$ et $t \in \mathbb{N}^\star$, pour toute valeur de $T \in \mathbb{N}^\star$,*

$$\sum_{t=1}^{T}\sum_{j=1}^{N} \mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant 2(M-m)\sqrt{T\ln N} + 6(M-m)(1+\ln N)\,.$$

Remarque au passage. Dans toute la suite de cette partie, nous ne nous préoccuperons presque uniquement que de majorants sur les regrets. Ils seront tous optimaux en un certain sens parce que les bornes du Lemme 1.3 et du Théorème 1.4 le sont elles aussi en un certain sens. Nous précisons cet énoncé dans le cadre plus général de la prévision économique en observations, au paragraphe 1.3.1.

1.2.2 Application à la prévision randomisée

Hypothèse 1.5. On suppose ici que la fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ est bornée, à valeurs dans l'intervalle $[m, M]$ (non nécessairement connu).

Analyse à paramètres T, m et M connus

Dans ce cas, une vitesse d'apprentissage constante $\eta > 0$ suffit.

Enoncé de la stratégie. La stratégie $\mathcal{E}_\eta$, appelée stratégie de prévision randomisée par pondération par poids exponentiels des pertes cumulées, recourt à la probabilité uniforme pour $\boldsymbol{p}_1$, soit $p_{j,1} = 1/N$ pour $j = 1, \dots, N$; et pour les échéances $t \geqslant 2$, elle utilise la probabilité $\boldsymbol{p}_t$ définie par la valeur de ses composantes $j = 1, \dots, N$ selon

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{i,s}, y_s)\right)} \,. \tag{1.6}$$

Elle tire ensuite un indice d'expert I_t selon $\boldsymbol{p}_t$ et forme la prévision $\widehat{y}_t = f_{I_t,t}$.

Analyse. En appliquant le résultat du Lemme 1.3, qui vaut de manière déterministe pour toute suite d'éléments $\ell_{j,t}$, aux variables aléatoires $\ell(f_{j,t}, y_t)$, on obtient le contrôle presque-sûr

$$\sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell(f_{i,t}, y_t) \leqslant \frac{\ln N}{\eta} + \eta \frac{(M-m)^2}{8} T = (M-m)\sqrt{\frac{T}{2} \ln N} \,, \tag{1.7}$$

pour η bien choisi en fonction de m, M et T. Or, en notant $\mathbb{E}_t$ l'espérance conditionnelle par rapport aux choix effectués par le statisticien et l'adversaire aux échéances 1 à $t-1$, ainsi qu'aux choix $f_{j,t}$ et y_t effectués par l'adversaire à l'échéance t, il vient :

$$\mathbb{E}_t\Big[\ell(f_{I_t,t}, y_t)\Big] = \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) \,;$$

en effet, cette espérance conditionnelle $\mathbb{E}_t$ fixe la valeur de la probabilité $\boldsymbol{p}_t$ mais pas celle du choix aléatoire de I_t selon $\boldsymbol{p}_t$.

Remarque au passage. Lorsque l'adversaire a une stratégie τ déterministe, *id est*, qu'il ne recourt à aucune randomisation auxiliaire, alors $\mathbb{E}_t$ est exactement l'espérance conditionnelle par rapport à $I_1, \dots, I_{t-1}$.

L'inégalité de Hoeffding–Azuma assure ensuite qu'avec probabilité au moins $1-\delta$,

$$\sum_{t=1}^{T} \ell(f_{I_t,t}, y_t) - \sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) \leqslant (M-m)\sqrt{\frac{T}{2} \ln \frac{1}{\delta}}\,. \tag{1.8}$$

L'association de (1.7) et (1.8) garantit finalement que pour toute échéance T, il existe une valeur $\eta_T^\star$ pour la vitesse d'apprentissage (dépendant de T, N, m et M même si l'on ne rappelle que la dépendance en T dans la notation) telle que le regret de $\mathcal{E}_{\eta_T^\star}$ est borné, pour toute stratégie τ de l'adversaire et avec probabilité au moins $1-\delta$, par

$$R_T(\mathcal{E}_{\eta_T^\star}, \tau) \leqslant (M-m)\sqrt{\frac{T}{2}} \left(\sqrt{\ln N} + \sqrt{\ln \frac{1}{\delta}} \right). \tag{1.9}$$

Comment remplir l'objectif 1.2 de minimisation du regret

En instanciant la stratégie du Théorème 1.4 aux variables aléatoires $\ell(f_{j,t}, y_t)$, on obtient une stratégie notée $\mathcal{E}_{\text{adapt}}$ (de la même manière que le Lemme 1.3 avait mené aux stratégies $\mathcal{E}_\eta$).

Ce théorème et l'inégalité de Hoeffding–Azuma montrent que pour toute stratégie τ de l'adversaire, toute échéance T et tout niveau de confiance $1-\delta_T \in \,]0,1[$,

$$R_T(\mathcal{E}_{\text{adapt}}, \tau) \leqslant (M-m)\sqrt{T} \left(2\sqrt{\ln N} + \sqrt{\frac{1}{2} \ln \frac{1}{\delta_T}} \right) + 6(M-m)(1+\ln N)\,.$$

En particulier, en prenant $\delta_T = 1/T^2$, le lemme de Borel–Cantelli implique que pour toute stratégie τ de l'adversaire,

$$\limsup_{T\to\infty} \frac{R_T(\mathcal{E}_{\text{adapt}}, \tau)}{(M-m)\sqrt{T \ln T}} \leqslant 1 \qquad \text{p.s.},$$

ce qui prouve notamment que l'objectif 1.2 est rempli pour $\mathcal{E}_{\text{adapt}}$.

Remarque au passage. En recourant ci-dessus à une version *maximale* de l'inégalité de Hoeffding–Azuma, appliquée aux échéances de la forme $T_r = 2^r$ et alliée au lemme de Borel–Cantelli, on peut prouver qu'en fait, pour toute stratégie τ de l'adversaire,

$$\limsup_{T\to\infty} \frac{R_T(\mathcal{E}_{\text{adapt}}, \tau)}{(M-m)\sqrt{2T \ln \ln T}} \leqslant 1 \qquad \text{p.s. ;}$$

ce n'est pas sans rappeler la loi du logarithme itéré (qui montre la nécessité du terme $\sqrt{\ln \ln T}$). Ceci sera à comparer avec le cas de l'agrégation convexe, où la vitesse de convergence du regret est $\sqrt{T}$, sans facteur logarithmique.

1.2.3 Application à l'agrégation convexe séquentielle

Hypothèse 1.6. On suppose que $\mathcal{X}$ est un sous-ensemble convexe borné de $\mathbb{R}^d$ et que pour tout $y \in \mathcal{Y}$, la fonction $\ell(\,\cdot\,, y)$ est convexe et différentiable sur $\mathcal{X}$, de gradient noté $\nabla\ell(\,\cdot\,, y)$. Par ailleurs, on requiert que les gradients soient uniformément bornés lorsque y varie.

Une première idée, insuffisante. Il s'agit de rendre le regret linéaire en les vecteurs de mélange $\boldsymbol{p}_t$ afin de pouvoir appliquer le Lemme 1.3 ; mais la majoration linéaire (qui procède de la convexité de ℓ en son premier argument)

$$\sum_{t=1}^{T} \ell\left(\sum_{j=1}^{N} p_{j,t} f_{j,t},\, y_t\right) \leqslant \sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\, \ell(f_{j,t}, y_t)$$

est trop grossière pour permettre un contrôle de la perte cumulée autre que face au meilleur expert.

Une inégalité des pentes. Ce qui suit a été proposé par [KW97, CB99] et peut également être retrouvé sous une forme proche dans [CBL06, paragraphe 2.5]. Ici, on a écrit explicitement une condition de différentiabilité dans l'hypothèse 1.6, même s'il est bien connu que toute fonction convexe est sous-différentiable sur l'intérieur de son domaine de définition, une propriété qui aurait été (presque) suffisante ici. Dans tous les cas, ces propriétés donnent lieu à une inégalité des pentes : pour tous les vecteurs de mélange $\boldsymbol{p}$ et $\boldsymbol{q}$, pour toutes les prévisions $f_1, \dots, f_N \in \mathcal{X}$ et toute observation $y \in \mathcal{Y}$,

$$\ell\left(\sum_{j=1}^{N} p_j f_j,\, y\right) - \ell\left(\sum_{j=1}^{N} q_j f_j,\, y\right) \leqslant \nabla\ell\left(\sum_{j=1}^{N} p_j f_j,\, y\right) \cdot \left(\sum_{j=1}^{N} p_j f_j - \sum_{j=1}^{N} q_j f_j\right). \quad (1.10)$$

Définissons alors les pseudo-pertes suivantes, pour l'expert $j \in \{1, \dots, N\}$ à l'échéance $t \in \{1, \dots, T\}$:

$$\widetilde{\ell}_{j,t} = \nabla\ell\left(\sum_{i=1}^{N} p_{i,t} f_{i,t},\, y_t\right) \cdot f_{j,t} \quad (1.11)$$

et considérons la famille de stratégies d'agrégation de la figure 1.3, appelées $\mathcal{E}_\eta^{\mathrm{grad}}$. Le regret de $\mathcal{E}_\eta^{\mathrm{grad}}$ est par conséquent majoré selon

$$\begin{aligned} R_T(\mathcal{E}_\eta^{\mathrm{grad}}) &= \sup_{\boldsymbol{q}\in\mathcal{P}} \sum_{t=1}^{T} \left(\ell\left(\sum_{j=1}^{N} p_{j,t} f_{j,t},\, y_t\right) - \ell\left(\sum_{j=1}^{N} q_j f_{j,t},\, y_t\right)\right) \\ &\leqslant \sup_{\boldsymbol{q}\in\mathcal{P}} \sum_{t=1}^{T} \left(\sum_{j=1}^{N} p_{j,t}\widetilde{\ell}_{j,t} - \sum_{j=1}^{N} q_j \widetilde{\ell}_{j,t}\right) = \sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\widetilde{\ell}_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \widetilde{\ell}_{i,t}\,, \end{aligned}$$

où l'inégalité procède de (1.10) et la seconde égalité du fait que le majorant ainsi obtenu est linéaire en $\boldsymbol{q}$ et est donc maximisé par un vecteur de mélange égal à une masse de Dirac. Le résultat suivant découle ensuite immédiatement du Lemme 1.3.

Paramètre : vitesse d'apprentissage $\eta > 0$

Initialisation : $\boldsymbol{p}_1$ est le mélange uniforme, soit $p_{j,1} = 1/N$ pour $j = 1, \dots, N$

Pour les échéances $t = 2, 3, \dots, T$, le vecteur de mélange $\boldsymbol{p}_t$ est défini par la valeur de ses composantes $j = 1, \dots, N$ selon

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \widetilde{\ell}_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \widetilde{\ell}_{i,s}\right)} ,$$

où l'on a posé

$$\widetilde{\ell}_{j,s} = \nabla \ell \left(\sum_{i=1}^{N} p_{i,s} f_{i,s},\, y_s \right) \cdot f_{j,s} \,.$$

FIGURE 1.3. La stratégie $\mathcal{E}_\eta^{\mathrm{grad}}$ de pondération par poids exponentiels des gradients des pertes cumulées.

Théorème 1.7. *On suppose que l'hypothèse 1.6 est vérifiée, de sorte que les pseudo-pertes définies en (1.11) sont bornées, à valeurs dans un intervalle noté* $[-C, C]$*. Alors, pour tout* $\eta > 0$*,*

$$\sup\left\{ R_T(\mathcal{E}_\eta^{\mathrm{grad}}) \right\} = \sup\left\{ \widehat{L}_T(\mathcal{E}_\eta^{\mathrm{grad}}) - \inf_{\boldsymbol{q} \in \mathcal{P}} L_T(\boldsymbol{q}) \right\} \leqslant \frac{\ln N}{\eta} + \eta \frac{C^2}{2} T \,,$$

où le supremum porte sur toutes les suites possibles d'observations et de prévisions des experts. En particulier, le choix de $\eta^\star = (1/C)\sqrt{(2 \ln N)/T}$ *conduit à la majoration*

$$\sup\left\{ R_T(\mathcal{E}_{\eta^\star}) \right\} \leqslant C\sqrt{2T \ln N} \,.$$

Calibration et objectif 1.1. Ici encore, un problème de calibration se pose pour le choix de η, qui dépend de la borne C sur les pseudo-pertes, éventuellement inconnue par avance, et du nombre d'échéances T, destiné à tendre vers l'infini. Mais on peut instancier la stratégie du Théorème 1.4 sur les pseudo-pertes (1.11) afin de pallier cela, comme nous venons de le faire avec le Lemme 1.3. (On note à cet égard qu'évidemment, les pseudo-pertes dépendent fortement de la stratégie qui les produit.) L'objectif 1.1 est alors rempli, avec un contrôle uniforme sur le regret de l'ordre de $C\sqrt{T \ln N}$.

Remarque au passage. Sous des hypothèses entraînant une convexité forte (par exemple, une hypothèse additionnelle de minoration uniforme des valeurs propres des matrices hessiennes, en plus de la majoration uniforme des gradients), on peut obtenir des contrôles sur le regret plus fins, de l'ordre (en T) de $\ln T$.

1.3 Contributions à la prévision randomisée [2, 12]

Nous présentons, plutôt brièvement, deux telles contributions. Leur point commun est que du point de vue technique, elles reposent toutes deux sur l'utilisation de pondérations par poids exponentiels. On décrit notamment l'article [2] parce que les méthodes qu'il utilise sont très similaires à celles employées plus tard dans les articles [3, 6] décrits au chapitre suivant.

1.3.1 Prévision randomisée économique en observations [2]

Présentation. Dans cette variation du cadre de prévision randomisée, essentiellement introduite par [HP97], on suppose qu'accéder à l'observation y_t est coûteux pour le statisticien. Il ne dispose que d'un budget limité pour ce faire, ce budget étant modélisé par une fonction croissante $B : \mathbb{N}^\star \to \mathbb{N}^\star$ indiquant qu'à toute échéance $t \geqslant 1$, pas plus de $B(t)$ observations y_t ne peuvent avoir été révélées. On modifie ainsi le déroulement proposé à la figure 1.2 en ajoutant la fonction B, connue du statisticien, dans les paramètres et en remplaçant l'unique point 4 des itérations selon les deux points suivants :

> 4. Le statisticien révèle ses choix $\widehat{y}_t$, I_t et $\boldsymbol{p}_t$ à l'adversaire, qui les mémorise pour la suite ;
> 5. Si le statisticien a accédé à moins de $B(t)$ observations pour l'instant, et seulement dans ce cas, il peut demander à l'adversaire de révéler l'observation $y_t \in \mathcal{Y}$, qu'il mémorise pour la suite le cas échéant et qui seule lui permet de calculer sa perte et celle des experts.

Objectif. Il s'agit toujours de remplir l'objectif 1.2 ; ce n'est évidemment possible que lorsque la fonction B croît suffisamment rapidement : par exemple, lorsque B est identiquement nulle, le statisticien ne reçoit aucun retour sur prévisions et aucune stratégie ne peut remplir l'objectif fixé.

Une idée simple : estimer ce qui n'est pas observé

Hypothèses pour commencer doucement. On va se placer pour commencer dans un cadre simplifié où l'échéance maximale T est fixée, où la fonction de perte ℓ est à valeurs dans un intervalle connu et de la forme $[0, M]$ et où la fonction de budget vérifie $B(1) = B(2) = \ldots = B(T)$, la valeur commune étant notée B_T.

Estimateurs des pertes. On note $Z_1, \ldots, Z_T$ une suite de variables indépendantes (entre elles et de tout autre aléa) et identiquement distribuées selon une loi de Bernoulli de paramètre $p \in]0, 1[$; ce dernier sera choisi, *via* l'inégalité de Bernstein, de telle sorte qu'avec un niveau de confiance fixé, $Z_1 + \ldots + Z_T \leqslant B_T$. La stratégie que nous allons considérer choisira d'accéder à l'observation y_t en fonction de la randomisation

auxiliaire : si et seulement si $Z_t = 1$, et évidemment, à condition que le budget ne soit pas dépassé. Pour toute échéance $t \geqslant 1$ et indice $j \in \{1, \ldots, N\}$, on définit donc de manière légitime l'estimateur suivant de $\ell(f_{j,t}, y_t)$:

$$\widehat{\ell}_{j,t} = \begin{cases} \dfrac{\ell(f_{j,t}, y_t)}{p} & \text{si } Z_t = 1 \text{ et } 1 + \displaystyle\sum_{s=1}^{t-1} Z_s \leqslant B_T \ ; \\ 0 & \text{sinon.} \end{cases}$$

Caractère conditionnellement sans biais de ces estimateurs. On note $\mathbb{E}_t$ l'espérance conditionnelle par rapport aux informations procurées par les échéances 1 à $t-1$, y compris les $Z_1, \ldots, Z_{t-1}$, et aux choix de l'adversaire au tour t (prévisions des experts $f_{j,t}$ et observation y_t). Par construction,

$$\text{sur } \{1 + Z_1 + \ldots + Z_{t-1} \leqslant B_T\}, \qquad \mathbb{E}_t\Big[\widehat{\ell}_{j,t}\Big] = \ell(f_{j,t}, y_t)\,.$$

Cela justifie que les estimateurs exhibés sont raisonnables.

Substitution des pertes par ces estimateurs dans la stratégie de pondération. On utilise la probabilité uniforme à l'échéance $t = 1$ et aux échéances $t \geqslant 2$, la probabilité $\boldsymbol{p}_t$ de composantes définies par substitution des estimateurs aux vraies pertes dans (1.6), soit

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_{j,s}\right)}{\sum_{i=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_{i,s}\right)} \ ;$$

et comme précisé plus haut, l'accès aux observations se fait selon la suite auxiliaire des Z_t. On note $\boldsymbol{\in}_{\eta,p}$ cette stratégie (pour économique en observations) ; elle dépend des deux paramètres η et p.

Référence. [ACBFS02] a le premier proposé une telle estimation sans biais de pertes non observées, dans un cadre différent de retour sur prévisions imparfait : celui des problèmes de bandits à plusieurs bras du chapitre 4.

Analyse de cette stratégie (sous les hypothèses pour commencer doucement). Elle repose de manière assez cruciale sur le fait que les pertes soient positives (et bornées par une quantité M connue). En effet, nous modifions le Lemme 1.3 et utilisons les inégalités

$$\forall x \in \mathbb{R}_+, \qquad e^{-x} \leqslant 1 - x + \frac{x^2}{2} \qquad \text{et} \qquad \forall u > -1, \qquad \ln(1+u) \leqslant u$$

au lieu du lemme de Hoeffding pour faire intervenir les probabilités $\boldsymbol{p}_t$ dans le membre droit de (1.2) selon

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \mu_{j,t} \ell_{j,t} - \min_{i=1,\ldots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t}\, \ell_{j,t}^2 \,; \tag{1.12}$$

après quoi, il suffit de noter que cette inégalité valant de manière déterministe, elle entraîne un résultat presque-sûr lorsque les quantités $\ell_{j,t}$ sont remplacées comme ici par les variables aléatoires $\widehat{\ell}_{j,t}$. L'application de quelques résultats élémentaires de concentration de la mesure, notamment la version maximale de l'inégalité de Bernstein, permet alors d'obtenir le résultat suivant, qui vaut avec probabilité $1-\delta$ et pour toute stratégie τ de l'adversaire : pour $p^\star$ et $\eta^\star$ bien choisis en fonction de T, B_T, M et δ, la stratégie précédente assure que

$$\max_{t\leqslant T} R_t(\in_{\eta^*,p^\star},\tau) \leqslant 8MT\sqrt{\frac{\ln(4N/\delta)}{B_T}}\,. \tag{1.13}$$

Retour à l'objectif 1.2

On dit qu'on utilise la stratégie $\in$ par blocs de longueurs dyadiques lorsque l'on démarre à froid une nouvelle instance $\in_{\eta_r,p_r}$ aux échéances de la forme $T_r = 2^r-1$, pour les entiers $r \geqslant 1$, avec des paramètres p_r et η_r convenablement réglés. On peut prouver à partir d'une telle construction le résultat suivant à partir des bornes intermédiaires (1.13) ; c'est d'ailleurs ici que nous avons besoin des majorations uniformes en t qui y avaient été exhibées.

Théorème 1.8. *Lorsque $B(T) \gg \ln T \ln\ln T$, il existe une manière simple et explicite d'utiliser la stratégie $\in$ par blocs de longueurs dyadiques, fondée uniquement sur la connaissance de la fonction de budget B et l'intervalle $[0,M]$ dans lequel la fonction de perte ℓ prend ses valeurs, remplissant l'objectif 1.2 de minimisation du regret et telle qu'à l'issue d'aucune échéance $T \geqslant 1$, plus de $B(T)$ observations n'ont été requises depuis le début.*

Au passage : un résultat d'optimalité générale

On fixe dans ce paragraphe les valeurs $m = 0$ et $M = 1$, toutes les quantités en jeu (pertes cumulées, regret) étant homogènes en ces deux quantités.

Contrôles sur l'espérance du regret. Les contrôles en grande probabilité sur le regret donnés par (1.9) et (1.13) entraînent en particulier, par intégration en δ, des contrôles sur l'espérance du regret, *uniformes* en toutes les stratégies τ de l'adversaire, de l'ordre respectif de $\sqrt{T\ln N}$ et $T\sqrt{(\ln N)/B_T}$. (On retrouve également la première vitesse lorsque l'on applique directement le Théorème 1.4 et que l'on prend l'espérance des espérances conditionnelles.)

Optimalité de la borne de prévision économique en observations. En particulier, le cas $B_T = T$ redonne la vitesse $\sqrt{T\ln N}$ sans contrainte d'économie. Le résultat suivant montre qu'il suffit d'un adversaire simple et non stratège pour forcer ces ordres de grandeur sur le regret : cet adversaire, noté $\tau(y_1^T)$, choisit la suite $y_1^T = (y_1,\dots,y_T)$

des observations à l'avance et recourt à des experts procurant chacun une prévision constante, égale à leur indice : $f_{j,t} = j$ pour toute échéance t et tout expert $j = 1, \dots, N$.

Théorème 1.9. *Dans le cas $\mathcal{X} = \mathbb{N}$ et $\mathcal{Y} = [0,1]$, il existe une fonction de perte ℓ à valeurs dans $[0,1]$ telle que pour tout $N \geqslant 2$ et tout couple nombre d'échéances–budget (T, B_T) tel que $T \geqslant B_T \geqslant 15 \ln(N-1)$, le regret de toute stratégie σ du statisticien, autorisée à accéder à au plus B_T observations entre les échéances 1 et T, vérifie*

$$\sup_{y_1^T \in \mathcal{Y}^T} \left\{ \mathbb{E}\Big[R_T\big(\sigma,\, \tau(y_1^T)\big) \Big] \right\} \geqslant \frac{T}{10} \sqrt{\frac{\ln N}{B_T}} \,.$$

Techniques de preuve. Dans [2], nous utilisons comme ingrédient principal le lemme de Fano et obtenons la borne non asymptotique exhibée ci-dessus. [CBL06, Théorème 6.4] parvient à un résultat similaire avec une preuve différente fondée sur des résultats de bornes inférieures sur l'erreur de classification pour des étiquettes binaires.

Références. Les travaux sur lesquels nous nous sommes fondés sont les suivants. Dans le cas $B_T = T$, la première preuve (asymptotique) d'optimalité des ordres de grandeur $\sqrt{T \ln N}$ a été procurée par [CBFH$^+$97] ; elle repose sur le théorème de la limite centrale et le fait que l'espérance du maximum de N variables aléatoires gaussiennes standards indépendantes est équivalente à $\sqrt{\ln N}$. Enfin, une preuve non asymptotique, par emploi de l'inégalité de Pinsker, a été fournie par [ACBFS02] dans un autre cadre, celui des bandits à plusieurs bras du chapitre 4.

1.3.2 Prévisions simultanées [12]

Un problème de prévision structuré. On illustre dans ce paragraphe un exemple de problème de prévision structuré. Ces problèmes correspondent au cas où les experts sont nombreux mais peuvent être décrits de manière compacte par quelques paramètres : la classe des experts admet une certaine structure. Par exemple, chaque expert correspond à un chemin possible entre deux points d'un graphe (problème dit du chemin le plus court). Ou encore, chaque expert de la classe est formé par une suite infinie admettant pour valeurs un petit nombre d'experts fondamentaux (problème dit des experts composés, voir le paragraphe 3.5.1). Ici, on considère la prévision simultanée de plusieurs observations.

Une stratégie simple mais généralement coûteuse à mettre en œuvre. Dans tout problème de prévision structuré, une stratégie simple et naturelle se dégage : la considération de pondérations par poids exponentiels sur la classe structurée des experts, dont on rappelle qu'elle admet un cardinal N élevé ; il est aisé de contrôler son regret par application de résultats généraux, et la dépendance de ce dernier en N est de l'ordre de $\sqrt{\ln N}$, ce qui est souvent acceptable. En revanche, la complexité de mise en œuvre naïve de cette stratégie, par affectation d'un poids à chaque expert, est de l'ordre de N, ce qui en revanche est presque toujours prohibitif. Ainsi, le seul point qui demande travail est l'obtention d'un algorithme plus efficace de mise en œuvre.

Description du modèle

On considère K tâches de prévision et on les indice par $k \in \{1, \dots, K\}$. A chacune d'entre elles correspondent un ensemble de prévisions $\mathcal{X}^{(k)}$, un ensemble d'observations $\mathcal{Y}^{(k)}$ et une fonction de perte $\ell^{(k)} : \mathcal{X}^{(k)} \times \mathcal{Y}^{(k)} \to \mathbb{R}$. On considère en outre N experts, chaque expert $j \in \{1, \dots, N\}$ proposant à chaque échéance $t \geqslant 1$ une prévision $f_{j,t}^{(k)}$ pour chacune des tâches k.

Déroulement (version initiale non contrainte). On adapte le déroulement indiqué à la figure 1.2 comme suit. A chaque échéance $t \geqslant 1$, l'adversaire choisit pour chaque expert j un vecteur de prévisions $\boldsymbol{f}_{j,t}$, qui seules sont révélées au statisticien, et un vecteur d'observations $\boldsymbol{y}_t$, qui reste caché à ce dernier pour l'instant :

$$\boldsymbol{f}_{j,t} = \left(f_{j,t}^{(1)}, \dots, f_{j,t}^{(K)}\right) \qquad \text{et} \qquad \boldsymbol{y}_t = \left(y_t^{(1)}, \dots, y_t^{(K)}\right).$$

Le statisticien tire au hasard, selon une loi notée $\boldsymbol{p}_t$, un élément de $\{1, \dots, N\}^K$, noté

$$\boldsymbol{I}_t = \left(I_t^{(1)}, \dots, I_t^{(K)}\right)$$

et indiquant, pour chaque tâche k, l'indice $I_t^{(k)}$ de l'expert duquel suivre la prévision, de sorte que le statisticien forme le vecteur de prévisions

$$\widehat{\boldsymbol{y}}_t = \left(f_{I_t^{(1)},t}^{(1)}, \dots, f_{I_t^{(K)},t}^{(K)}\right). \tag{1.14}$$

Le statisticien et l'adversaire révèlent ensuite publiquement toutes les quantités décrites précédemment, qui sont mémorisées pour la suite.

Mesure de la qualité des prévisions. On définit une fonction de perte globale ℓ à partir des pertes individuelles pour chacune des tâches selon une fonction de mesure globale $\psi : \mathbb{R}^K \to \mathbb{R}$; plus précisément, la qualité de la prévision donnée par un vecteur $\boldsymbol{f}_{j,t}$ contre les observations $\boldsymbol{y}_t$ est par exemple donnée par

$$\ell(\boldsymbol{f}_{j,t}, \boldsymbol{y}_t) = \psi\left(\ell^{(1)}\left(f_{j,t}^{(1)}, y_t^{(1)}\right), \ \dots, \ \ell^{(K)}\left(f_{j,t}^{(K)}, y_t^{(K)}\right)\right).$$

Des exemples de fonctions de mesure globale pour lesquelles nous avons été capables d'offrir une mise en œuvre efficace sont la somme, le minimum et le maximum des pertes associées à chaque tâches, respectivement définis par :

$$\psi(x_1, \dots, x_K) = \sum_{k=1}^{K} x_k\,, \qquad \psi(x_1, \dots, x_K) = \min\{x_1, \dots, x_K\}\,,$$
$$\text{et} \qquad \psi(x_1, \dots, x_K) = \max\{x_1, \dots, x_K\}\,.$$

Pour simplifier les écritures, pour tout K–uplet $\boldsymbol{j} = (j_1, \ldots, j_K)$ de $\{1, \ldots, N\}^K$, on notera

$$\ell_t(\boldsymbol{j}) = \psi\Big(\ell^{(1)}\Big(f^{(1)}_{j_1,t}, y^{(1)}_t\Big),\ \ldots,\ \ell^{(K)}\Big(f^{(K)}_{j_K,t}, y^{(K)}_t\Big)\Big)$$

la perte globale associée à l'échéance t au choix de l'expert j_k dans chaque tâche k, plaçant ainsi dans la notation ℓ_t les choix de l'adversaire. Par exemple, en reprenant la notation (1.14), on a

$$\ell(\widehat{\boldsymbol{y}}_t, \boldsymbol{y}_t) = \ell_t(\boldsymbol{I}_t)\,.$$

Ajout d'une contrainte reliant les tâches de prévision. Pour l'instant, rien ne lie les tâches entre elles : en un sens, il s'agit de K problèmes de prévision distincts. Cela est encore plus vrai lorsque lorsque la qualité de la prévision est évaluée par la fonction de perte donnée par la somme des pertes, car il suffit alors de mener en parallèle K stratégies fondamentales, une pour chaque tâche, ce qui a une complexité de mise en œuvre de l'ordre de NK lorsque ces stratégies fondamentales sont données par des pondérations par poids exponentielles des pertes cumulées sur chaque tâche.

C'est pourquoi on impose que les seules prévisions légales sont celles correspondant au choix à chaque échéance $t \geqslant 1$ d'un K–uplet d'experts $\boldsymbol{I}_t$ appartenant à un certain sous-ensemble strict $\mathcal{L}$ de $\{1, \ldots, N\}^K$. Les choix d'experts tant du statisticien que ceux de la classe de comparaison utilisée dans la définition du regret devront être dans $\mathcal{L}$. Ce sous-ensemble a pour objet de modéliser en un certain sens les liens entre les tâches.

Notion de regret pour la prévision simultanée. En s'inspirant de (1.1), on définit ici le regret d'une stratégie σ du statisticien contre celle, τ, de l'adversaire et au vu de la contrainte $\mathcal{L}$ comme

$$R^{\mathrm{PS}}(\sigma, \tau) = \sum_{t=1}^{T} \ell_t(\boldsymbol{I}_t) - \min_{\boldsymbol{j} \in \mathcal{L}} \sum_{t=1}^{T} \ell_t(\boldsymbol{j})\,. \tag{1.15}$$

Il s'agit encore d'exhiber une stratégie σ (avec une mise en œuvre efficace) telle que

$$\sup_{\tau} \left\{ \limsup_{T \to \infty} \frac{R^{\mathrm{PS}}(\sigma, \tau)}{T} \right\} \leqslant 0 \qquad \text{p.s.}, \tag{1.16}$$

où le supremum porte sur l'ensemble des stratégies τ de l'adversaire.

Exemple de contraintes $\mathcal{L}$. Nous proposons quatre exemples dans [12] mais n'en reprenons qu'un seul ici, le plus facile à décrire : le coût au changement. On fixe un entier $m \leqslant K - 1$ et on déclare légaux uniquement les K–uplets $\boldsymbol{j}$ tels que

$$\sum_{k=1}^{K-1} \mathbb{I}_{\{j_k \neq j_{k+1}\}} \leqslant m\,.$$

Il est facile de dénombrer au moins $(NK)^m/m!$ tels K–uplets.

Comparaison avec des travaux antérieurs

En fait, il n'est pas si facile ni si naturel de définir ce qu'est un problème de prévision simultanée pour des suites arbitraires ; deux approches en ce sens avaient été précédemment formulées par [ABR07] et [DLS07].

Le modèle de [ABR07]. A chaque échéance, le statisticien ne doit former de prévision que dans un seul problème, choisi par l'adversaire et il peut suivre l'avis de l'expert de son choix. Ainsi, la liaison entre les tâches n'apparaît pas dans la manière dont le statisticien peut former des prévisions, mais uniquement dans la classe de comparaison dans la définition du regret, ici également prise de la forme d'un sous-ensemble $\mathcal{L}'$ des suites de T indices d'expert ; c'est-à-dire qu'une définition du regret du type de (1.15) est considérée, mais que le statisticien n'est pas obligé de tenir compte des contraintes de $\mathcal{L}'$. (Or, c'est ce qui, selon nous, relie les différents problèmes et empêche le recours à des sous-stratégies fonctionnant en parallèle.)

Le modèle de [DLS07]. Il est formé de K problèmes de classification linéaire, avec des pertes chacune mesurées par la perte charnière $x \mapsto (1-x)_+$ et agrégées *via* une fonction de perte globale ψ donnée par la norme euclidienne ou la norme du supremum sur $\mathbb{R}^K$. Les différents problèmes ne sont reliés que par cette évaluation globale et la classe de comparaison dans la définition du regret, qui est formée par la considération du même hyperplan vectoriel dans tous les problèmes.

Conclusion. Notre modèle est le premier à limiter le statisticien dans sa manière de prévoir et à lui affecter les mêmes contraintes que celles mises sur la classe de comparaison dans la définition du regret.

Borne sur le regret

Stratégie. La stratégie de pondération par poids exponentiels du paragraphe 1.2.2 (avec une vitesse d'apprentissage η_t réglée selon ce qu'indique le Théorème 1.4) propose de choisir à chaque échéance $t \geqslant 2$ la loi $\boldsymbol{p}_t$ sur $\mathcal{L}$ obtenue comme la combinaison convexe de masses de Dirac $\delta_{\boldsymbol{j}}$ suivante :

$$\boldsymbol{p}_t = \sum_{\boldsymbol{j} \in \mathcal{L}} \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1} \ell_s(\boldsymbol{j})\right)}{\sum_{\boldsymbol{i} \in \mathcal{L}} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(\boldsymbol{i})\right)} \delta_{\boldsymbol{j}} \tag{1.17}$$

(et la loi uniforme sur $\mathcal{L}$ pour la première échéance). On désigne par $\mathcal{M}$ la stratégie ainsi obtenue (pour multi-tâches de prévision).

Contrôle sur le regret. Là encore, il suffit d'appliquer le Théorème 1.4 d'une manière similaire à celle utilisée au paragraphe 1.2.2 ; on suppose à cet effet que les fonctions de perte ℓ_t prennent des valeurs uniformément bornées, disons dans $[m, M]$. Alors, le

regret de $\mathcal{M}$ est borné, pour toute stratégie τ de l'adversaire et avec probabilité au moins $1-\delta$, par

$$R_T(\mathcal{M},\tau) \leqslant 2(M-m)\sqrt{T}\left(\sqrt{\ln|\mathcal{L}|}+\sqrt{\frac{1}{2}\ln\frac{1}{\delta}}\right)+6(M-m)(1+\ln|\mathcal{L}|)\,,$$

où $|\mathcal{L}|$ désigne le cardinal de $\mathcal{L}$; on notera à cet égard que le log–cardinal $\ln|\mathcal{L}|$ est petit, car toujours inférieur à $K\ln N$. Une application immédiate du lemme de Borel–Cantelli garantit que $\mathcal{M}$ atteint l'objectif (1.16).

Mise en œuvre efficace dans certains cas

Il suffit de savoir tirer selon $\boldsymbol{p}_t$. Il ne reste plus qu'à voir si l'on peut tirer efficacement un vecteur d'indices $\boldsymbol{I}_t$ selon la loi $\boldsymbol{p}_t$ de (1.17). Cela se fera évidemment sans calcul explicite de $\boldsymbol{p}_t$, puisque le coût du stockage mémoire de ce dernier serait proportionnel à $|\mathcal{L}|$, et donc en général prohibitif.

Un espace d'états cachés. Nous montrons dans [12] que ce coût est en fonction d'une structure de chaîne de Markov cachée (non homogène) que l'on peut faire apparaître sur $\mathcal{L}$, chaque élément de ce dernier étant vu comme la réalisation des K premiers états de cette chaîne. On note S le cardinal de cet espace d'états cachés.

Le tirage se fait alors de manière récursive inverse : on tire d'abord $I_t^{(K)}$ selon la dernière marginale de $\boldsymbol{p}_t$, puis $I_t^{(K-1)}$ selon la $(K-1)$–ème marginale de la loi $\boldsymbol{p}_t$ conditionnellement au tirage déjà effectué, etc. Le résultat essentiel est qu'afin de pouvoir mettre en œuvre le schéma inverse précédent, il suffit de combiner et mettre à jour échéance après échéance au plus NKS quantités. La complexité en termes d'espace mémoire est donc proportionnelle à NKS, tandis que celle en temps de calcul est légèrement supérieure, d'un facteur multiplicatif tenant compte du nombre de transitions possibles d'un état caché vers un autre.

Exemple : le cas du coût au changement. Pour cet exemple, la complexité de mise en œuvre de la procédure décrite ci-dessus est de l'ordre de NKm en espace mémoire et N^2Km en temps de calcul, à comparer aux complexités de l'ordre de $(NK)^m/m!$ pour le calcul et la représentation directes de (1.17).

1.4 Calibration : adaptation et bornes dépendant des données [5]

Un premier vœu : l'adaptation aux paramètres. Nous revenons dans cette partie sur le Théorème 1.4, qui constituait un résultat fondamental (mais formulé de manière un peu vague) : il a permis, tant dans le cadre d'agrégation convexe que dans celui de prévision randomisée, de construire des stratégies totalement automatiques vérifiant les objectifs 1.1 ou 1.2. Ce caractère totalement automatique est lié au fait que la suite (η_t) des vitesses d'apprentissage peut être construite séquentiellement de manière adéquate,

un fait que nous détaillons ici : sans connaître ni l'échéance maximale T, ni les bornes m et M sur la fonction de perte ou la borne C sur les gradients des pertes, nous avons pu obtenir les ordres de grandeur optimaux $(M-m)\sqrt{T\ln N}$ ou $C\sqrt{T\ln N}$ pour des majorants du regret.

Un second vœu : l'obtention de bornes plus fines, dépendant des données. On reproche souvent aux bornes précédentes leur caractère trop uniforme et trop pessimiste : si un des experts est bien meilleur que les autres et a une perte cumulée faible, alors le regret lui-même devrait croître bien plus lentement que $\sqrt{T}$. On cherche ainsi à remplacer les bornes générales uniformes par des bornes dépendant davantage des données (souvent, *via* une formulation en termes des suites de pertes subies).

Objectif de cette partie. Nous allons détailler les réponses apportées à ces deux vœux, par [5] et les travaux qui lui sont antérieurs. A cet effet, nous nous placerons dans le cadre générique du paragraphe 1.2.1, dont on a montré qu'il était facile d'instancier tous ses résultats pour l'agrégation convexe et la prévision randomisée. Ce cadre générique consiste à définir pour toute échéance $t \geqslant 1$ des vecteurs de mélange μ_t sur $\{1,\dots,N\}$ ne dépendant que des pertes passées $\ell_{j,s}$, pour $s \leqslant t-1$ et $j \in \{1,\dots,N\}$, et à majorer

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \tag{1.18}$$

en fonction de la suite de pertes subies, qui est déterministe et fixée à l'avance ; on pourra appeler (1.18) le regret générique d'une stratégie.

1.4.1 Résumé du contenu des travaux antérieurs à [5]

Adaptation en les paramètres inutile pour la pondération par poids polynômiaux

La stratégie de pondération par poids polynômiaux des pertes, qui est essentiellement la stratégie de [Bla56] et qui a été revisitée par [CBL03], choisit aux échéances $t \geqslant 2$ les vecteurs de mélange μ_t définis, composante par composante, selon

$$\mu_{j,t} = \frac{\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\mu_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{j,t}\right)_+^{\alpha-1}}{\sum_{k=1}^{N}\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\mu_{i,t}\ell_{i,t} - \sum_{t=1}^{T}\ell_{k,t}\right)_+^{\alpha-1}}$$

où $(\,\cdot\,)_+$ désigne la fonction partie positive et où l'exposant vérifie $\alpha \geqslant 1$. C'est le seul paramètre de cette stratégie ; par ailleurs, lorsque la suite des pertes est bornée entre les deux réels m et M, son regret générique est uniformément majoré par

$$(M-m)\sqrt{(\alpha-1)TN^{2/\alpha}} \leqslant (M-m)\sqrt{6T\ln N}$$

pour le choix $\alpha = 2\ln N$, qui ne dépend que du paramètre toujours connu N, nombre d'experts, et pas des paramètres potentiellement inconnus à l'avance que sont m, M

et T. La borne sur le regret générique obtenue étant optimale du point de vue des ordres de grandeurs en tous les paramètres, on peut se demander pourquoi on se fatigue à vouloir absolument considérer des pondérations par poids exponentiels comme (1.3).

Objection : importance des poids exponentiels en situations d'observations imparfaites. En fait, lorsque les pertes ne sont pas révélées pleinement après une échéance de prévision et qu'il s'agit par exemple de les estimer, ce qui est le cas aux paragraphes 1.3.1 et 2.2, les choses se compliquent pour la pondération par poids polynômiaux. [CBL06, Théorème 6.9] montre certes pour elle (mais au prix de pas mal d'efforts et sans obtention claire de vitesses de convergence) que son regret rapporté au nombre de tours peut, dans le contexte des bandits à nombre finis de bras du chapitre 4, être asymptotiquement négatif ou nul. Ceci est cependant à comparer à la simplicité avec laquelle le contrôle (1.13) a été obtenu *via* une pondération par poids exponentiels.

Adaptation en les paramètres par redémarrages périodiques

[CBL06, paragraphe 2.3] expose une solution (ancienne et non attribuable à un groupe de chercheurs en particulier) pour calibrer les vitesses d'apprentissage de (1.3), qui consiste à effectuer des redémarrages périodiques de la stratégie de pondération par poids exponentiels, avec des vitesses d'apprentissage η_r de plus en plus petites lorsque le nombre r de redémarrages déjà effectués augmente. On parle d'utilisation d'une stratégie par blocs et on note qu'en fait, une telle solution avait déjà été présentée plus haut pour obtenir le Théorème 1.8. Les vitesses η_r sont prises de la forme de la valeur théorique optimale indiquée par le Lemme 1.3, soit $\big(1/(M_r - m_r)\big)\sqrt{(8\ln N)/2^r}$, où m_r et M_r désignent des estimées sur les bornes des pertes calculées sur les $r-1$ instances passées. Chacun des blocs prend alors fin selon une condition d'arrêt liée à T ou au fait que les bornes estimées m_r et M_r sont violées.

Objection : perte d'information et manques périodiques d'efficacité. A chaque redémarrage, on oublie presque toute l'information procurée par le passé, à l'exception de celle résumée dans m_r et M_r, et surtout, on recommence pendant un certain temps à utiliser des vecteurs de mélange proches du vecteur uniforme, ce qui est peu efficace, voire rédhibitoire, en pratique.

Bornes plus fines sur le regret et adaptation partielle lorsque les pertes sont positives

L'analyse de [FS97] de la stratégie (1.3) assure un contrôle du regret générique (1.18) par la borne dépendant des données,

$$\sqrt{2ML_T^\star \ln N}\,, \qquad \text{où} \qquad L_T^\star = \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t}\,, \tag{1.19}$$

lorsque les pertes sont bornées à valeurs dans un intervalle $[0, M]$ et que le paramètre η de la stratégie de pondération par poids exponentiels sur les pertes cumulées est calibré

de manière rétrospective en fonction de $L_T^\star$ et de M. Une adaptation à la valeur inconnue de $L_T^\star$ est possible par utilisation par blocs. Cette borne est appelée l'amélioration pour les pertes cumulées faibles.

Une meilleure idée : la calibration séquentielle des vitesses. [ACBG02] a introduit la forme (1.5) et les ingrédients essentiels à son analyse. Cet article indique en particulier, comment, connaissant M et sous la même hypothèse de pertes positives, la considération de vitesses d'apprentissage η_t proportionnelles à

$$\sqrt{\frac{\ln N}{M \sum_{s=1}^{t-1} \sum_{i=1}^{N} \mu_{i,s} \ell_{i,s}}} \tag{1.20}$$

permettait de retrouver la borne (1.19) à, essentiellement, un facteur multiplicatif 2 près. (En particulier, l'adaptation en T découle comme cas particulier de cette méthode.) On redonne ci-dessous les grandes lignes de la preuve de ce résultat.

Ce qu'il reste à faire. Il faut cependant encore traiter l'adaptation en les bornes sur les pertes et si possible, relâcher l'hypothèse de leur positivité.

Le cas important des problèmes avec pertes signées

[ANN04] a le premier considéré le cas où les pertes $\ell_{j,t}$ ne sont plus nécessairement positives, mais simplement supposées être choisies dans un intervalle $[m, M]$, pour deux réels $m \leqslant M$. Lorsque m et M sont inconnus, on ne peut se ramener par translation au cas des pertes positives.

Importance de ce cadre. [ANN04] ne justifie pas l'intérêt des pertes signées. Nous voyons leur importance fondamentale par l'instanciation (1.11) des bornes génériques aux pseudo-pertes dans le cas de l'agrégation convexe ; ces pseudo-pertes ne sont pas nécessairement positives mais elles sont généralement bornées.

Objectif pour la suite. Il faudra en fait traiter l'adaptation en m et en M, dans le cas de pertes non nécessairement positives.

Remarque : Déviations autour de la borne de regret générique en prévision randomisée

A trop considérer le regret générique, on peut oublier qu'*in fine*, on s'intéresse notamment au regret dans le cadre de prévision randomisée. Pour contrôler ce dernier, il faut ajouter à la borne de regret générique une borne de déviation garantie par un résultat de concentration. Si l'on se contente de l'inégalité de Hoeffding–Azuma comme en (1.8), alors les déviations de l'ordre de $\sqrt{T \ln(1/\delta)}$ tuent toute amélioration dépendant des données sur le majorant du regret générique. Il vaut mieux recourir à l'inégalité de

Bernstein pour les accroissements de martingales, comme je l'ai par exemple expliqué dans ma thèse [Sto05, pages 38–39] ; elle suffit pour préserver par exemple l'amélioration pour les pertes faibles décrite ci-dessus.

1.4.2 Adaptation en les paramètres pour les pertes signées [5]

Une idée essentielle est de remplacer le dénominateur de (1.20) par une quantité ne dépendant que du passé (et plus de M), homogène au carré des pertes. En fait, ce dénominateur provenait d'une majoration au premier ordre d'une tranformée de Laplace, et ici, en poussant à l'ordre deux, nous allons obtenir directement la quantité souhaitée.

Modification de la preuve du Lemme 1.3. [ACBG02] (voir également les versions postérieures simplifiées de [CBL06, paragraphe 2.3] et [GO07, Lemme 1]) propose la borne de performance suivante pour la stratégie (1.5) : pour toute suite décroissante (η_t), éventuellement calibrée séquentiellement en fonction du passé, son regret générique est borné par

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T}\Phi\big(\mu_t,\,(\ell_{j,t})_j,\,\eta_t\big) \tag{1.21}$$

où la fonction Φ est définie en termes d'un vecteur de mélange μ, de N pertes $(\ell_1,\dots,\ell_N)$ et d'une vitesse d'apprentissage η selon

$$\Phi\big(\mu,\,(\ell_j)_j,\,\eta\big) = \frac{1}{\eta}\ln\left(\sum_{i=1}^{N}\mu_i\,e^{-\eta\left(\ell_i-\widehat{\ell}\right)}\right) \qquad \text{où} \quad \widehat{\ell} = \sum_{j=1}^{N}\mu_j\ell_j\,.$$

On suppose qu'on arrive à exhiber des majorants sur les termes en Φ de la forme $\eta_t\,z_t$, avec $z_t \geqslant 0$, soit

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T}\eta_t z_t\,.$$

Il n'est alors pas difficile de voir que pour le choix de

$$\eta_t = \sqrt{\frac{\ln N}{z_1+\ldots+z_{t-1}}}\,,$$

on a la borne sur le regret générique

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t} \leqslant 4\sqrt{\left(\sum_{t=1}^{T}z_t\right)\ln N}\,. \tag{1.22}$$

(La constante multiplicative 4 du terme de droite peut être améliorée.) Il s'agit donc d'obtenir les z_t, c'est-à-dire de contrôler des quantités de la forme

$$\Psi_\eta(X) = \frac{1}{\eta^2}\,\ln\mathbb{E}\Big[e^{-\eta\left(X-\mathbb{E}[X]\right)}\Big]$$

pour $\eta > 0$ et X une variable aléatoire prenant un nombre fini de valeurs.

Bornes uniformes (d'ordre zéro). La borne (1.2) du Lemme 1.3 correspond au choix d'une suite (η_t) constante et à une majoration des Ψ_η par lemme de Hoeffding comme indiqué en (1.4). On veut améliorer cela.

Cas des pertes positives (bornes du premier ordre). Nous avions déjà noté en (1.12) que dans le cas des pertes positives et bornées par M,

$$\Psi_\eta(X) \leqslant \frac{\eta}{2}\,\mathbb{E}\Big[X^2\Big] \leqslant \frac{\eta M}{2}\,\mathbb{E}[X]\,;$$

cela correspond, avec les notations précédentes, à

$$z_t = \frac{M}{2}\sum_{j=1}^N \mu_{j,t}\ell_{j,t}\,,$$

soit, grâce à (1.22), l'inégalité

$$\sum_{t=1}^T\sum_{j=1}^N \mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N}\sum_{t=1}^T \ell_{i,t} \leqslant 2\sqrt{2}\sqrt{M\left(\sum_{t=1}^T\sum_{j=1}^N \mu_{j,t}\ell_{j,t}\right)\ln N}\,. \tag{1.23}$$

Il suffit pour conclure de résoudre une inéquation du second degré et on parvient ainsi à une borne améliorée pour les pertes faibles comme (1.19), à facteur multiplicatif près qu'il n'est pas difficile de calculer précisément et qui représente le prix de l'adaptation en M et $L_T^\star$.

Borne du second ordre pour des pertes signées. En utilisant l'inégalité $e^x \leqslant 1+x+(e-2)\,x^2$ pour $x \leqslant 1$, il vient que pour tout couple (η, X) tel que $\eta > 0$ et $\eta X \leqslant 1$ p.s.,

$$\Psi_\eta(X) \leqslant \frac{1}{\eta}\,\ln\Big(1+(e-2)\eta^2\,\mathrm{Var}(X)\Big) \leqslant (e-2)\eta\,\mathrm{Var}(X)\,. \tag{1.24}$$

En particulier, en définissant pour toutes échéances t et T un terme de pseudo-variance des pertes à l'échéance t et un autre de pseudo-variance cumulée jusque T,

$$v_t = \sum_{j=1}^N \mu_{j,t}\left(\ell_{j,t} - \sum_{i=1}^N \mu_{i,t}\ell_{i,t}\right)^2 \qquad \text{et} \qquad V_T = v_1 + \ldots + v_T\,,$$

on s'attend à obtenir une borne sur le regret générique de l'ordre de $\sqrt{V_T \ln N}$, grâce à (1.22) et au choix $z_t = v_t$. Cependant, il faut bien garder en tête que (1.24) nécessite une condition de domination par 1, qui n'est pas toujours vérifiée ; lorsqu'elle ne l'est pas, on peut toutefois toujours appliquer la borne (1.4) du lemme de Hoeffding. Nous montrons alors le résultat fondamental suivant dans [5], qui précise le Théorème 1.4 déjà maintes fois employé précédemment. Nous le reproduisons comme il y est énoncé, même si depuis sa publication, [Ger10] a indiqué que l'on peut en fait légèrement améliorer le facteur 4 devant le terme principal et le remplacer par

$$2\sqrt{(e-2)(\sqrt{2}-1)} \leqslant 2{,}64\,.$$

Théorème 1.10. *On considère la stratégie totalement adaptative (1.5), où les vitesses d'apprentissage sont définies, pour $t \geqslant 2$, en fonction uniquement des pertes passées selon*

$$\eta_t = \min\left\{ \frac{1}{E_{t-1}}, \gamma\sqrt{\frac{\ln N}{V_{t-1}}} \right\}$$

où $\gamma = \sqrt{2(\sqrt{2}-1)\,/\,(e-2)}$ et

$$E_{t-1} = \min\left\{ 2^k : \quad k \in \mathbb{Z} \ \text{ et } \ \max_{s\leqslant t-1} \max_{i\neq j} |\ell_{i,s} - \ell_{j,s}| \leqslant 2^k \right\}.$$

Alors, pour tous réels $m \leqslant M$ non nécessairement positifs, pour toute suite arbitraire d'éléments $\ell_{j,t} \in [m, M]$, où $j \in \{1, \dots, N\}$ et $t \in \mathbb{N}^\star$, pour toute valeur de $T \in \mathbb{N}^\star$,

$$\sum_{t=1}^{T}\sum_{j=1}^{N} \mu_{j,t}\ell_{j,t} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} \leqslant 4\sqrt{V_T \ln N} + 6(M-m)(1+\ln N)\,.$$

Corollaires. Dans le théorème, le contrôle sur le regret est en termes de V_T, qui n'est pas une quantité intrinsèque mais dépend de la stratégie. On explique ici comment résoudre cette difficulté. Une première idée est de noter que les termes v_t sont des termes de variance, il sont en particulier bornés par la demi-étendue au carré, $(M-m)^2/4$. C'est en injectant cette majoration que l'on retrouve la borne uniforme initialement proposée au Théorème 1.4. Par ailleurs, une variance étant plus petite que son espérance au carré, on a également que

$$V_T \leqslant \sum_{t=1}^{T}\sum_{j=1}^{N} \mu_{j,t}\ell_{j,t}^2\,,$$

ce qui permet en particulier de retrouver une forme similaire à (1.23) lorsque les pertes sont positives, et donc une amélioration pour les pertes faibles. (On montre dans [5] qu'on peut en fait obtenir une amélioration pour les pertes positives faibles ou très grandes.)

Autres commentaires. On peut souligner que la borne sur le regret proposée au Théorème 1.10 est stable par translations additives des pertes, ce qui n'est pas le cas de nombreuses bornes de regret, notamment les améliorations pour les pertes faibles, qui requièrent une hypothèse de positivité. Par ailleurs, aussi gênant soit le terme $\sqrt{V_T}$, on peut noter que dans le cadre de la prévision randomisé, il apparaît également lors de l'application de l'inégalité de Bernstein pour les accroissements de martingales comme argument de concentration en lieu et place de l'inégalité de Hoeffding–Azuma en (1.8) ; il est donc inévitable en un certain sens, la question étant de déterminer s'il existe d'encore meilleures manières de vivre avec lui que celles exposées ci-dessus dans le paragraphe des corollaires.

1.5 Perspectives

Dans ce paragraphe, on désigne par $\square$ des constantes universelles, dont, pour la simplicité du propos et la facilité de lecture, on ne précise pas la valeur ; la valeur de $\square$ change d'occurrence en occurrence.

Revenons-en à la perte cumulée...

On cesse désormais de considérer la minimisation du regret comme le but à atteindre et on en revient au souhait originel : assurer que la perte cumulée de la stratégie est la plus faible possible.

Correspondance entre meilleures erreurs d'approximation et d'estimation. Or, les bornes exhibées précédemment sur les pertes cumulées de nos stratégies sont de la forme suivante :

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} \leqslant \underbrace{\min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t}}_{\text{erreur d'approximation}} + \underbrace{D\left(\min_{i=1,\dots,N}\sum_{t=1}^{T}\ell_{i,t}\right)}_{\text{difficulté d'estimation}} \tag{1.25}$$

où la fonction D est constante, égale à $\square\,(M-m)\sqrt{T\ln N}$ (borne uniforme) ou donnée par $x \mapsto \square\sqrt{x\ln N} + \square\, M\ln N$ (amélioration pour les pertes faibles, dans le cas de pertes positives).

Le terme donné par la fonction D correspondait à la borne sur le regret. Ici, puisque D était toujours croissante, il y avait correspondance entre l'expert atteignant la meilleure erreur d'approximation et celui atteignant la meilleure erreur d'estimation. Ceci n'est sûrement pas très réaliste et ne permet donc pas de rendre compte de la meilleure manière possible de la réalité.

Forme souhaitée de la borne sur la perte cumulée. C'est pourquoi l'on pense plutôt à des majorations de la forme

$$\sum_{t=1}^{T}\sum_{j=1}^{N}\mu_{j,t}\ell_{j,t} \leqslant \min_{i=1,\dots,N}\left\{\sum_{t=1}^{T}\ell_{i,t} + D(\ell_{i,1},\dots,\ell_{i,T})\right\} \tag{1.26}$$

où cette fois-ci la fonction $D : \mathbb{R}^T \to \mathbb{R}$ est beaucoup plus générale (mais sans doute croissante en chacun de ses T arguments). Il y a alors un compromis à faire entre erreur d'approximation et difficulté d'estimation.

Une tentative prometteuse, qui a échoué

Par exemple, la motivation initiale à [5] était de prouver (1.26) pour

$$D_{\text{CARR}}(x_1,\dots,x_T) = 2\sqrt{\sum_{t=1}^{T}x_t^2\ln N} + \square\,(M-m)\ln N$$

tandis que celle de [HK08], à la suite de nos travaux, visait

$$D_{\text{VAR}}(x_1,\dots,x_T) = \square\sqrt{\sum_{t=1}^{T}\left(x_t - \frac{1}{T}\sum_{s=1}^{T} x_s\right)^2 \ln N} + \square\,(M-m)\ln N\,.$$

Bien sûr, toutes deux auraient conduit en particulier aux contrôles (1.25) mais auraient pu procurer des améliorations substantielles.

La source d'espoir. Dans [5], le regret générique d'une stratégie appelée Prod_η, dépendant d'un paramètre $\eta > 0$ mais ne reposant pas sur des poids exponentiels (un fait rare dans ce chapitre), est contrôlé ainsi : pour toute suite arbitraire d'éléments $\ell_{j,t} \in [m,M]$, où $j \in \{1,\dots,N\}$ et $t \in \mathbb{N}^\star$, pour tout choix $0 < \eta \leqslant 1/(2M)$, pour toute valeur de $T \in \mathbb{N}^\star$,

$$\sum_{t=1}^{T}\sum_{j=1}^{N} \mu_{j,t}\ell_{j,t} \leqslant \min_{i=1,\dots,N}\left\{\sum_{t=1}^{T}\ell_{i,t} + \frac{\ln N}{\eta} + \eta\sum_{t=1}^{T}\ell_{i,t}^2\right\}. \tag{1.27}$$

Un choix rétrospectif convenable de η permet d'obtenir (1.26) avec $D = D_{\text{CARR}}$: notant

$$i_T^\star \in \operatorname*{arg\,min}_{i=1,\dots,N}\left\{\sum_{t=1}^{T}\ell_{i,t} + 2\sqrt{\sum_{t=1}^{T}\ell_{i,t}^2 \ln N}\right\},$$

il suffit de recourir au paramètre η donné par

$$\eta_T^\star = \sqrt{\frac{\ln N}{\displaystyle\sum_{t=1}^{T}\ell_{i_T^\star,t}^2}},$$

à condition que ce dernier soit plus petit que $1/(2M)$. Encore une fois, tout le jeu consiste alors à obtenir de manière séquentielle et adaptative la même borne que celle procurée par ce choix rétrospectif. On pourrait penser que c'est là un travail de routine... Or, c'est précisément là que le bât a blessé.

Ce que nous avons pu prouver. En fait, les techniques d'adaptation décrites précédemment (utilisation de stratégies par blocs ou calibration adaptative de vitesses) requièrent fondamentalement la considération de quantités croissantes en T. Ici, les quantités-clés

$$Q_T^\star = \sum_{t=1}^{T}\ell_{i_T^\star,t}^2$$

ne le sont pas nécessairement (car les valeurs de $i_T^\star$ peuvent changer au cours du temps). En revanche, leurs plus petits majorants croissants $\max_{t\leqslant T} Q_t^\star$ le sont et c'est pourquoi

les bornes finales sur le regret d'une stratégie totalement adaptative fondée sur les Prod_η sont à nouveau de la forme

$$\min_{i=1,\dots,N} \sum_{t=1}^{T} \ell_{i,t} + \square \sqrt{\max_{t\leqslant T} Q_t^\star \ln N} + \square\,(M-m)\ln N\,,$$

cette dernière ressemblant davantage à (1.25) qu'à la forme souhaitée (1.26).

Problème similaire pour [HK08]. Même si ce dernier peut obtenir (1.26) pour $D = D_{\text{VAR}}$ grâce à un choix rétrospectif convenable (et en utilisant une stratégie de pondération par poids exponentiels de pertes cumulées *pénalisées*), il souffre des mêmes soucis lors de ses tentatives de calibration séquentielle.

Enoncé du problème ouvert

Il s'agit donc soit de prouver des inégalités de la forme (1.26), soit de montrer qu'aucune stratégie ne peut les garantir. Mon intuition penche plutôt pour leur obtention, mais il semble qu'une limite conceptuelle fondamentale soit à franchir en termes d'adaptation séquentielle.

CHAPITRE 2

Interactions avec la théorie des jeux répétés

INTRODUCTION. Ce chapitre constitue une variante de la situation de prévision randomisée étudiée au chapitre précédent. En termes de terminologie, on considère désormais, au lieu d'un statisticien cherchant à prévoir l'évolution d'un environnement qui se dérobe, un couple de joueurs chacun réagissant au comportement de l'autre, et même tâchant de l'anticiper ; de plus, au lieu de pertes, on parlera de paiements et il s'agira pour chaque joueur de les maximiser. Enfin, on simplifie le cadre de prévision : il n'y a plus d'experts et chaque joueur dispose seulement d'un nombre fini d'actions.

L'heuristique sous-jacente développée ici est que si chaque joueur met en œuvre une stratégie dont les bonnes performances en termes de paiements sont garanties, au sens où son regret (à re-définir) est faible, alors on se trouve asymptotiquement dans une situation d'équilibre. On verra essentiellement trois notions de situations d'équilibre : dans le cas d'un jeu à somme nulle, la convergence, pour chaque joueur, de ses paiements moyens vers la valeur du jeu ; dans le cas général d'un jeu dans lequel peuvent d'ailleurs s'affronter plus de deux joueurs, la convergence du profil moyen des actions choisies vers l'ensemble des équilibres au sens de Hannan ou vers celui des équilibres corrélés.

Table des matières

2.1 Définition et justification de la notion de regret

On formalise le problème de la manière suivante, par un jeu à deux joueurs. On le fait par souci de simplicité, en indiquant toutefois que tous les résultats de ce chapitre, sauf ceux sur les jeux à somme nulle, s'étendent facilement au moins au cas des jeux à nombre fini de joueurs.

Notations. Les deux joueurs, appelés joueurs A et B, ont des ensembles finis d'actions respectivement notés $\mathcal{A} = \{1, \dots, N\}$ et $\mathcal{B} = \{1, \dots, M\}$. Ils disposent chacun d'une

fonction de paiement $\mathcal{A} \times \mathcal{B} \to \mathbb{R}$; on note r celle du joueur A et s celle du joueur B. Ces fonctions r et s sont étendues linéairement aux simplexes $\Delta(\mathcal{A})$ et $\Delta(\mathcal{B})$ des lois de probabilité sur $\mathcal{A}$ et $\mathcal{B}$: pour tous $\boldsymbol{p} = (p_i)_{i\in\mathcal{A}} \in \Delta(\mathcal{A})$ et $\boldsymbol{q} = (p_j)_{j\in\mathcal{B}} \in \Delta(\mathcal{B})$,

$$r(\boldsymbol{p},\boldsymbol{q}) = \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{B}} p_i\, q_j\, r(i,j) \qquad \text{et} \qquad s(\boldsymbol{p},\boldsymbol{q}) = \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{B}} p_i\, q_j\, s(i,j)\,.$$

Déroulement du jeu répété. A chaque tour $t = 1, 2, \ldots$, le joueur A choisit une action $I_t \in \mathcal{A}$ tandis que le joueur B recourt à l'action $J_t \in \mathcal{B}$; les deux actions sont alors révélées et les joueurs obtiennent les paiements respectifs $r(I_t, J_t)$ et $s(I_t, J_t)$. Les choix de I_t et J_t sont effectués à l'aide d'une randomisation auxiliaire et en se fondant sur le passé ; c'est-à-dire que ces actions sont tirées au hasard selon des probabilités $\boldsymbol{p}_t$ et $\boldsymbol{q}_t$ sur $\mathcal{A}$ et $\mathcal{B}$ qui dépendent mesurablement de l'historique des couples $(I_1, J_1), \ldots, (I_{t-1}, J_{t-1})$ joués dans le passé.

On appelle stratégies de A et B les suites d'applications qui, pour tout $t \geqslant 1$, associent respectivement à tout historique de $(\mathcal{A} \times \mathcal{B})^{t-1}$ un élément de $\Delta(\mathcal{A})$ et de $\Delta(\mathcal{B})$. Les notations [1] consacrées pour ces stratégies seront σ et τ.

On introduira dans la suite

$$\overline{\boldsymbol{p}}_T = \frac{1}{T}\sum_{t=1}^{T} \delta_{I_t} \qquad \text{et} \qquad \overline{\boldsymbol{q}}_T = \frac{1}{T}\sum_{t=1}^{T} \delta_{J_t}\,,$$

les distributions empiriques des actions choisies par chacun des deux joueurs.

Objectifs et quantités d'intérêt. Chaque joueur vise à rendre son paiement moyen le plus grand possible, en un sens asymptotique ; *id est*, les joueurs A et B s'intéressent respectivement aux quantités

$$\overline{r}_T = \frac{1}{T}\sum_{t=1}^{T} r(I_t, J_t) \qquad \text{et} \qquad \overline{s}_T = \frac{1}{T}\sum_{t=1}^{T} s(I_t, J_t)$$

lorsque $T \to \infty$.

Des quantités auxiliaires : les regrets. On définit les regrets respectifs R_T et S_T de A et B à la fin du tour T comme en (1.1), au remplacement près des pertes par des paiements, c'est-à-dire en termes de la meilleure action constante du joueur donné, les actions de son adversaire étant fixées :

$$R_T = \max_{i\in\mathcal{A}} \sum_{t=1}^{T} r(i, J_t) - \sum_{t=1}^{T} r(I_t, J_t) \qquad \text{et} \qquad S_T = \max_{j\in\mathcal{B}} \sum_{t=1}^{T} s(I_t, j) - \sum_{t=1}^{T} s(I_t, J_t)\,,$$

[1] Toutes les quantités introduites dans la suite dépendront de σ et τ, mais dans ce chapitre, dans un souci de lisibilité, nous ne répercuterons pas cette dépendance dans les notations.

ainsi que les regrets moyens (rapportés au nombre de tours) correspondants,

$$\overline{R}_T = \max_{i\in\mathcal{A}} r(i,\overline{\boldsymbol{q}}_T) - \overline{r}_T \qquad \text{et} \qquad \overline{S}_T = \max_{j\in\mathcal{B}} s(\overline{\boldsymbol{p}}_T, j) - \overline{s}_T\,. \tag{2.1}$$

Par linéarité, les maxima sur $i \in \mathcal{A}$ et $j \in \mathcal{B}$ ci-dessus peuvent être remplacés par des maxima sur $\boldsymbol{p} \in \Delta(\mathcal{A})$ et $\boldsymbol{q} \in \Delta(\mathcal{B})$.

Rendre le regret petit afin d'assurer un paiement moyen important ? L'interprétation du regret est ici moins claire que dans le cas de l'agrégation convexe du chapitre précédent, où il apparaissait comme une difficulté d'estimation qu'il fallait équilibrer avec une erreur d'approximation. En fait, dans ce cas, on peut définir ce qu'est la meilleure combinaison convexe des experts à utiliser et on pourrait exploiter sa connaissance, si un oracle nous la fournissait, et ce, parce que nous avions supposé à raison que les prévisions du statisticien n'influençaient pas l'environnement. Ici, toute l'évaluation se fait à actions fixées de l'adversaire ; mais si par exemple le joueur A avait utilisé à chaque tour l'action optimale contre la suite d'actions réalisée $J_1, \dots, J_T$, cette suite d'actions aurait été différente ! C'est une difficulté que nous avions déjà soulignée au paragraphe 1.1.3.

La justification de la notion de regret viendra ci-dessous par des arguments de convergence, en des sens à préciser, vers des ensembles d'équilibres, cette convergence ayant lieu dès que les deux joueurs assurent que leurs regrets moyens sont petits. Or, en situation d'équilibre, chaque joueur obtient un paiement moyen optimal, en un sens dépendant de la notion d'équilibre.

A cet égard, l'intérêt du regret, c'est que c'est une quantité que chaque joueur peut contrôler dans son coin, sans réellement tenir compte de la nature de son adversaire ; en particulier, aucune hypothèse sur sa rationalité ou son degré de coopération n'est requise.

Plan de cette partie introductive. Plus précisément, nous allons tout d'abord indiquer comment les joueurs peuvent rendre leurs regrets petits, avant d'étudier les conséquences d'une telle minimisation simultanée : la convergence vers l'ensemble des équilibres au sens de Hannan. Dans le cas particulier d'un jeu à somme nulle, le résultat pourra même être renforcé en l'obtention de paiements moyens optimaux égaux à la valeur du jeu et en la convergence vers l'ensemble des équilibres minimax (qui correspondront ici à l'ensemble des équilibres de Nash).

2.1.1 Contrôle du regret de chaque joueur, indépendamment de l'autre joueur

Une adaptation immédiate des résultats du chapitre précédent au cas de paiements plutôt que de pertes est la suivante. On note $\|r\|_\infty$ un majorant de $|r|$, on pose

$$\eta_t = \frac{1}{\|r\|_\infty}\sqrt{\frac{8\ln N}{t-1}}$$

pour tout $t \geqslant 2$, et on suppose que le joueur A utilise la stratégie suivante : il choisit son action I_1 au hasard selon la loi uniforme $\boldsymbol{p}_1$ sur $\mathcal{A}$, et pour $t \geqslant 2$, les actions I_t sont tirées selon la probabilité $\boldsymbol{p}_t$ dont les composantes sont

$$p_{i,t} = \frac{\exp\left(\eta_t \sum_{s=1}^{t-1} r(i, J_s)\right)}{\sum_{k\in\mathcal{A}} \exp\left(\eta_t \sum_{s=1}^{t-1} r(k, J_s)\right)} \tag{2.2}$$

pour tout $i \in \mathcal{A}$. Alors, il est facile de voir, en reprenant les calculs autour de (1.22), qu'on a le contrôle déterministe suivant, sur une quantité aléatoire : pour toute stratégie τ du joueur B,

$$\sum_{t=1}^{T} r(\boldsymbol{p}_t, J_t) \geqslant \max_{i\in\mathcal{A}} \sum_{t=1}^{T} r(i, J_t) - \|r\|_\infty \sqrt{2T \ln N}\,.$$

En particulier, le lemme de Hoeffding–Azuma assure qu'à tout tour T et pour tout niveau d'erreur $\delta > 0$, on a, avec probabilité au moins $1 - \delta$,

$$\overline{r}_T \geqslant \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big) - \|r\|_\infty \left(\sqrt{\frac{2}{T}\ln N} + \sqrt{\frac{1}{2T}\ln\frac{1}{\delta}}\right). \tag{2.3}$$

Enfin, par application du lemme de Borel–Cantelli, on a donc prouvé que contre toute stratégie τ du joueur B,

$$\liminf_{T\to\infty} \left\{\overline{r}_T - \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big)\right\} \geqslant 0 \quad \text{p.s.}, \qquad \text{soit} \qquad \limsup_{T\to\infty} \overline{R}_T \leqslant 0 \quad \text{p.s.} \tag{2.4}$$

Tous les développements ultérieurs de convergence vers des ensembles d'équilibres ne reposeront que sur des énoncés asymptotiques de la forme de (2.4), ceux du type (2.3) permettant d'avoir une idée de la vitesse de cette convergence. Dans ce chapitre, on appellera encore les stratégies assurant (2.4) les stratégies minimisant le regret.

Conclusion en termes du contrôle du regret. Le joueur A dispose d'une stratégie reposant sur les seules connaissance de sa fonction de paiement r et observation des actions J_t de son adversaire, telle que son regret soit petit au sens de (2.3). En particulier, le joueur A n'a pas à connaître en entier le jeu répété joué (il n'a pas besoin de connaître la fonction de paiement s) ni à émettre aucune hypothèse que ce soit sur τ.

La littérature parle de stratégies myopes : le joueur A ne voit que ce qui est proche, à savoir, ses propres paiements, et ne conduit pas de raisonnement quant à ce qui est loin, à savoir, la stratégie τ de son adversaire. La suite va montrer l'intérêt de telles stratégies, frustes en apparence : si tous les joueurs minimisent simultanément leur regret, une convergence vers des ensembles d'équilibres a lieu.

D'autres stratégies minimisant le regret. On citera seulement deux autres familles de stratégies assurant (2.4). Jouer à chaque tour l'action qui a le paiement cumulé pour l'instant le plus grand peut être désastreux, au sens où le regret de cette stratégie

peut être linéaire ; en revanche, Hannan [Han57] a montré qu'ajouter auparavant un terme de perturbation aléatoire adéquat au paiement cumulé permet de garantir la minimisation du regret. On parle de stratégies "follow the perturbed leader" et elles ont été ré-introduites et ré-étudiées depuis par [KV03] et d'autres à leur suite.

Une autre famille procède du théorème d'approchabilité de Blackwell [Bla56], que l'on rappelle dans l'encart ci-dessous. Là aussi, l'analyse a été re-visitée récemment par [CBL03].

Encart : Le théorème d'approchabilité**.**

Soit une fonction vectorielle $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$, dont on considère l'extension linéaire à $\Delta(\mathcal{A}) \times \Delta(\mathcal{B})$. On fait s'affronter de manière répétée les joueurs A et B et on conserve les notations I_t et J_t précédentes pour désigner les actions qu'ils choisissent.

Soit $\mathcal{C} \subset \mathbb{R}^d$ un ensemble ; $\mathcal{C}$ est dit m–approchable par le joueur A si ce dernier dispose d'une stratégie σ telle que, quelle que soit la stratégie τ du joueur B, on ait

$$\lim_{T\to\infty} \inf_{c\in\mathcal{C}} \left\| c - \frac{1}{T}\sum_{t=1}^{T} m(I_t, J_t) \right\| = 0 \qquad \text{p.s.} \tag{2.5}$$

Il découle du théorème de minimax de von Neumann (un cas particulier du lemme de Sion, voir la Définition–Théorème 2.4) la caractérisation suivante d'approchabilité dans le cas des ensembles convexes fermés. En fait, Blackwell a également donné une stratégie afin d'approcher $\mathcal{C}$; cette stratégie repose sur des projections convexes et requiert de résoudre un programme linéaire à chaque étape.

Théorème 2.1 (Référence : [Bla56, Théorème 3]). *Un sous-ensemble convexe fermé $\mathcal{C}$ de $\mathbb{R}^d$ est m–approchable si et seulement si*

$$\forall \boldsymbol{q} \in \Delta(\mathcal{B}), \quad \exists \boldsymbol{p} \in \Delta(\mathcal{A}), \qquad m(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{C}\,.$$

L'existence d'une stratégie minimisant le regret découle alors de la considération de l'orthant négatif $\mathcal{C} =]-\infty, 0]^N$ et de la fonction vectorielle m définie par

$$m(i,j) = \big(r(k,j) - r(i,j)\big)_{k\in\mathcal{A}}$$

pour tous $i \in \mathcal{A}$ et $j \in \mathcal{B}$.

2.1.2 Convergence vers l'ensemble des équilibres au sens de Hannan

En fait, la stratégie de Hannan [Han57] est antérieure à celle de Blackwell [Bla56] et c'est elle la première stratégie à avoir pu assurer (2.4). C'est pourquoi on a appelé en son honneur l'ensemble des équilibres suivants.

Il est en termes de lois jointes $\pi = \big(\pi(i,j)\big)_{(i,j)\in\mathcal{A}\times\mathcal{B}}$; on note $\Delta(\mathcal{A} \times \mathcal{B})$ l'ensemble de ces lois jointes.

Définition 2.2. L'ensemble des équilibres au sens de Hannan d'un jeu à deux joueurs est donné par l'ensemble (non vide) des lois jointes

$$\mathcal{H} = \left\{ \pi \in \Delta(\mathcal{A}\times\mathcal{B}) : \quad \forall i \in \mathcal{A},\ \sum_{k,\ell} \pi(k,\ell)\, r(k,\ell) \geqslant \sum_{k,\ell} \pi(k,\ell)\, r(i,\ell) \right.$$
$$\left. \text{et} \quad \forall j \in \mathcal{B},\ \sum_{k,\ell} \pi(k,\ell)\, s(k,\ell) \geqslant \sum_{k,\ell} \pi(k,\ell)\, s(k,j) \right\}.$$

L'interprétation en termes d'équilibre d'une telle loi jointe π est la suivante : on suppose qu'un couple d'actions (I, J) est tiré au hasard selon π par un médiateur et que les joueurs sont invités à jouer chacun l'action I ou J qui a été tirée pour lui. Alors, en espérance, si un joueur respecte cette invitation, l'autre joueur n'a pas intérêt à remplacer l'action qui lui est proposée par une autre action qu'il aurait fixée à l'avance avant d'accéder à cette proposition. (On parle de déviations unilatérales non profitables.)

Il faudra une hypothèse supplémentaire, que le jeu soit à somme nulle (que $r+s=0$), pour pouvoir assurer une convergence du couple $(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T)$. Pour l'instant, on s'intéresse à la distribution empirique des couples d'actions joués,

$$\overline{\pi}_T = \frac{1}{T}\sum_{t=1}^{T} \delta_{(I_t,J_t)}\,;$$

elle admet pour lois marginales $\overline{\boldsymbol{p}}_T$ et $\overline{\boldsymbol{q}}_T$. On suppose alors que les deux joueurs minimisent leur regret, ce qui peut se ré-écrire comme le fait que

$$\liminf_{T\to\infty} \left\{ \overline{r}_T - \max_{i\in\mathcal{A}} r\big(i, \overline{\boldsymbol{q}}_T\big) \right\} \geqslant 0 \quad \text{p.s.} \qquad \text{et} \qquad \liminf_{T\to\infty} \left\{ \overline{s}_T - \max_{j\in\mathcal{B}} s\big(\overline{\boldsymbol{p}}_T, j\big) \right\} \geqslant 0 \quad \text{p.s.}$$

Les conditions définissant $\mathcal{H}$ étant de type fermé, on déduit des inégalités asymptotiques précédentes que tout point d'adhérence π de la suite des $\overline{\pi}_T$ est un équilibre au sens de Hannan : $\pi \in \mathcal{H}$. Or, l'ensemble $\Delta(\mathcal{A}\times\mathcal{B})$ des lois jointes étant compact, un raisonnement par l'absurde montre enfin que la suite des $\overline{\pi}_T$ converge vers $\mathcal{H}$. On notera qu'il s'agit bien d'une convergence vers l'ensemble $\mathcal{H}$ et non d'une convergence vers un point de $\mathcal{H}$.

Proposition 2.3. Si les deux joueurs minimisent leur regret, alors la suite des distributions empiriques des couples d'actions joués, $(\overline{\pi}_T)$, converge presque-sûrement vers l'ensemble $\mathcal{H}$ des équilibres au sens de Hannan.

2.1.3 Cas d'un jeu à somme nulle : convergence vers l'ensemble des équilibres minimax

Ce paragraphe se place dans le cas où $r + s = 0$, soit $r = -s$, c'est-à-dire que les joueurs ont des objectifs radicalement antagonistes. Si les joueurs sont informés de ce fait, ils peuvent alors mettre chacun en œuvre une stratégie optimale, que l'on décrit maintenant.

Quelques résultats élémentaires sur les jeux à somme nulle

On aura besoin à cette effet de la notion de valeur d'un jeu ; elle découle d'un cas particulier du lemme de Sion, qu'on appelle le théorème de minimax de von Neumann.

Définition–Théorème 2.4 (von Neumann, 1928). On appelle valeur d'un jeu à somme nulle $r : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ la quantité

$$v \;=\; \max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \min_{\boldsymbol{q}\in\Delta(\mathcal{B})} r(\boldsymbol{p},\boldsymbol{q}) = \min_{\boldsymbol{q}\in\Delta(\mathcal{B})} \max_{\boldsymbol{p}\in\Delta(\mathcal{A})} r(\boldsymbol{p},\boldsymbol{q})\,.$$

Par définition même, cette notion conduit à l'ensemble des équilibres minimax (qui correspond, dans ce cas, à l'ensemble des équilibres de Nash). Dans ce qui suit, on identifie les couples de probabilités $(\boldsymbol{p},\boldsymbol{q}) \in \Delta(\mathcal{A}) \times \Delta(\mathcal{B})$ aux lois-produits qu'ils induisent dans $\Delta(\mathcal{A}\times\mathcal{B})$.

Définition–Théorème 2.5. L'ensemble $\mathcal{N}$ des équilibres minimax d'un jeu à somme nulle $r : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ est défini comme l'ensemble (non vide)

$$\begin{aligned}\mathcal{N} &= \mathcal{H} \cap (\Delta(\mathcal{A}) \times \Delta(\mathcal{B})) \\ &= \Big\{(\boldsymbol{p},\boldsymbol{q}) :\quad \forall i \in \mathcal{A},\quad r(\boldsymbol{p},\boldsymbol{q}) \geqslant r(i,\boldsymbol{q}) \quad\text{et}\quad \forall j \in \mathcal{B},\quad r(\boldsymbol{p},\boldsymbol{q}) \leqslant r(\boldsymbol{p},j)\Big\}\,.\end{aligned}$$

En particulier, tout couple $(\boldsymbol{p},\boldsymbol{q}) \in \mathcal{N}$ atteint la valeur du jeu : $r(\boldsymbol{p},\boldsymbol{q}) = v$.

Fixons pour l'étude qui suit un couple $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ de $\mathcal{N}$. Si le joueur B emploie $\boldsymbol{q}_t = \boldsymbol{q}^\star$ à tous les tours, c'est-à-dire qu'il tire ses actions J_t toutes indépendamment selon la même loi $\boldsymbol{q}^\star$, alors toute stratégie σ du joueur A admet un paiement moyen d'au plus v :

$$\limsup_{T\to\infty}\, \overline{r}_T = \limsup_{T\to\infty}\, \frac{1}{T}\sum_{t=1}^{T} r(I_t,J_t) = \limsup_{T\to\infty}\, r\big(\overline{\boldsymbol{p}}_T,\, \boldsymbol{q}^\star\big) \leqslant v \qquad \text{p.s.},$$

où l'on a appliqué l'inégalité de Hoeffding–Azuma ainsi que le lemme de Borel–Cantelli pour prouver la seconde égalité ; bien sûr, la dernière inégalité peut être une égalité, par exemple lorsque A tire ses actions de manière indépendante et identiquement distribuée selon $\boldsymbol{p}^\star$. On a évidemment un résultat similaire lorsque c'est A qui emploie $\boldsymbol{p}_t = \boldsymbol{p}^\star$ à tous les tours et que l'on s'intéresse aux stratégies τ de B : dans ce cas, le joueur A obtient au moins le paiement moyen v à la limite.

Conclusion et difficulté. Si un joueur sait que le jeu est un jeu à somme nulle, il a alors une stratégie optimale au sens où elle lui garantit un paiement moyen asymptotiquement le plus grand possible dans le cas le pire : calculer un équilibre minimax $(\boldsymbol{p}^\star,\boldsymbol{q}^\star)$ et tirer ses actions de manière indépendante et identiquement distribuée selon sa marginale de cet équilibre.

La difficulté est qu'évidemment, on s'était placé dans un cadre (le cadre myope) où les joueurs ne connaissaient que leur propre fonction de paiement ; ils ignorent donc en particulier si le jeu est à somme nulle ou pas.

Convergence vers l'ensemble des équilibres minimax

Cela dit, on va montrer que si chaque joueur emploie une stratégie myope qui minimise son regret, alors on retrouve les résultats précédents : le joueur A (respectivement, B) ne peut avoir plus de v (respectivement, plus de $-v$) comme paiement asymptotique moyen, mais par ailleurs il peut se garantir v (respectivement, $-v$). Les paiements asymptotiques moyens des joueurs A et B sont alors d'exactement v et $-v$.

De plus, on aura convergence non seulement de $(\overline{\pi}_T)$ vers $\mathcal{H}$ mais aussi de la suite $((\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T))$ de ses marginales vers $\mathcal{N}$.

Chaque joueur se garantit la valeur du jeu. Lorsque (2.4) est vérifiée, le paiement moyen du joueur A vérifie en particulier

$$\liminf_{T\to\infty} \overline{r}_T \geqslant \liminf_{T\to\infty} \max_{i\in\mathcal{A}} r(i, \overline{\boldsymbol{q}}_T) = \liminf_{T\to\infty} \max_{\boldsymbol{p}\in\Delta(\mathcal{A})} r(\boldsymbol{p}, \overline{\boldsymbol{q}}_T) \geqslant v \qquad \text{p.s.}\,; \tag{2.6}$$

une considération de symétrie montre que lorsque le joueur B minimise son regret, alors son paiement moyen vérifie

$$\liminf_{T\to\infty} -\overline{r}_T \geqslant \liminf_{T\to\infty} \max_{\boldsymbol{q}\in\mathcal{B}} -r(\overline{\boldsymbol{p}}_T, \boldsymbol{q}) \geqslant -v \qquad \text{p.s.},$$

$$\text{soit} \qquad \limsup_{T\to\infty} \overline{r}_T \leqslant \limsup_{T\to\infty} \min_{\boldsymbol{q}\in\Delta(\mathcal{B})} r(\overline{\boldsymbol{p}}_T, \boldsymbol{q}) \leqslant v \qquad \text{p.s.} \tag{2.7}$$

On a prouvé le résultat suivant.

Proposition 2.6. Si les deux joueurs d'un jeu à somme nulle de valeur v minimisent tous deux leur regret, alors $\overline{r}_T \to v$ p.s. lorsque $T \to \infty$.

Conséquence. En fait, cette proposition montre que toutes les inégalités de (2.6) et (2.7) sont des égalités et on en déduit en particulier, par un argument d'encadrement,

$$\lim_{T\to\infty} r(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T) = v \quad \text{p.s.}, \qquad \text{d'où} \qquad \lim_{T\to\infty} \overline{r}_T - r(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T) = 0 \quad \text{p.s.}$$

Par conséquent, en en revenant au fait (2.4) que les deux joueurs minimisent leurs regrets, il vient

$$\liminf_{T\to\infty} \left\{ r(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T) - \max_{i\in\mathcal{A}} r(i, \overline{\boldsymbol{q}}_T) \right\} \geqslant 0 \quad \text{p.s.}$$

$$\text{et} \qquad \limsup_{T\to\infty} \left\{ r(\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T) - \min_{j\in\mathcal{B}} r(\overline{\boldsymbol{p}}_T, j) \right\} \leqslant 0 \quad \text{p.s.}$$

En utilisant la même technique de preuve que pour la Proposition 2.3, à savoir, le fait que $\mathcal{N}$ est défini par des contraintes de type fermé, associé à un argument de compacité et à un raisonnement par l'absurde, on obtient la convergence suivante (qui, elle aussi, est en termes de convergence vers un ensemble et non vers un point de l'ensemble).

Proposition 2.7. Dans un jeu à somme nulle, si les deux joueurs minimisent leur regret, alors la suite des couples de distributions empiriques des actions jouées, $((\overline{\boldsymbol{p}}_T, \overline{\boldsymbol{q}}_T))$, converge presque-sûrement vers l'ensemble $\mathcal{N}$ des équilibres minimax.

2.2 Minimisation du regret dans les jeux avec observations imparfaites [3, 6]

On traite dans cette partie du cas où un joueur, disons le joueur A, n'observe qu'imparfaitement les actions choisies par son adversaire. Plus précisément, en plus des notations et objets précédents, on introduit un ensemble fini $\mathcal{S}$ de signaux possibles et une fonction de retour sur actions $H : \mathcal{A} \times \mathcal{B} \to \Delta(\mathcal{S})$ qui associe à chaque couple (i, j) d'actions de $\mathcal{A} \times \mathcal{B}$ une probabilité $H(i, j)$ sur $\mathcal{S}$. On étend linéairement H en une application $\Delta(\mathcal{A}) \times \Delta(\mathcal{B}) \to \Delta(\mathcal{S})$: pour tous $\boldsymbol{p} \in \Delta(\mathcal{A})$ et $\boldsymbol{q} \in \Delta(\mathcal{B})$,

$$H(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{B}} p_i \, q_j \, H(i, j) \ \in \Delta(\mathcal{S}) \, .$$

Déroulement du jeu répété avec observations imparfaites. Du point de vue du joueur A, les choses se passent ainsi. A chaque tour $t \geqslant 1$, les joueurs choisissent chacun des actions I_t et J_t, éventuellement au hasard selon des probabilités $\boldsymbol{p}_t$ et $\boldsymbol{q}_t$ pouvant dépendre des informations passées. Le joueur A obtient $r(I_t, J_t)$ comme paiement mais n'accède ni à J_t ni à la valeur de son paiement : il n'observe qu'une variable aléatoire K_t tirée indépendamment au hasard selon $H(I_t, J_t)$. Le joueur B a quant à lui un retour parfait : il observe I_t.

Objectifs. On se contente ici d'adapter les résultats du paragraphe 2.1.3 à ce nouveau cadre avec observations imparfaites ; plus précisément, on définit et justifie une extension de la notion de regret, telle que, si elle minimisée par chaque joueur et que l'on est dans un jeu à somme nulle, alors les convergences énoncées aux Propositions 2.6 et 2.7 ont encore lieu. Bien évidemment, la notion étendue de regret s'appliquerait aussi au cas des jeux finis généraux, mais par souci de simplicité, on se concentre sur le cas particulier des jeux à somme nulle.

2.2.1 Extension de la notion de regret

Indiscernabilité de certaines actions randomisées. Le joueur A ne peut pas distinguer toutes les probabilités qu'utilise le joueur B pour tirer ses actions ; pour lui, sont identiques en termes de retours sur actions deux lois $\boldsymbol{q}$ et $\boldsymbol{q}'$ telles que $H(i, \boldsymbol{q}) = H(i, \boldsymbol{q}')$ pour tout $i \in \mathcal{A}$. Ce fait se ré-écrit $H(\,\cdot\,, \boldsymbol{q}) = H(\,\cdot\,, \boldsymbol{q}')$ en notant

$$H(\,\cdot\,, \boldsymbol{q}) = \big(H(i, \boldsymbol{q})\big)_{i \in \mathcal{A}} \ \in \big(\Delta(\mathcal{S})\big)^{\mathcal{A}}$$

le vecteur des probabilités sur les signaux induit par une probabilité $\boldsymbol{q}$ sur $\mathcal{B}$. On note

$$\mathcal{V} = \Big\{ H(\,\cdot\,, \boldsymbol{q}), \quad \boldsymbol{q} \in \Delta(\mathcal{B}) \Big\}$$

l'ensemble des vecteurs que le joueur B peut engendrer en faisant varier $\boldsymbol{q}$. On désignera généralement par $\underline{h}$ un élément générique de $\mathcal{V}$.

C'est à cause de cette indiscernabilité que l'on introduit la fonction

$$\rho : \ (\boldsymbol{p}, \underline{h}) \in \Delta(\mathcal{A}) \times \mathcal{V} \longmapsto \min\Big\{ r(\boldsymbol{p}, \boldsymbol{q}) : \ \boldsymbol{q} \in \Delta(\mathcal{B}) \text{ tel que } H(\,\cdot\,, \boldsymbol{q}) = \underline{h} \Big\} \in \mathbb{R}\,.$$

Elle indique l'espérance minimale du paiement à attendre lorsque le joueur A tire son action au hasard selon $\boldsymbol{p}$ et que le joueur B le fait selon une probabilité $\boldsymbol{q}$ induisant le vecteur de probabilités sur les signaux $\underline{h}$. La fonction ρ est concave en son argument de $\Delta(\mathcal{A})$ et convexe en celui de $\mathcal{V}$.

Ré-écriture de la valeur du jeu. La première observation-clé est la suivante. Dans le cas d'un jeu à somme nulle, la valeur se ré-écrit en termes de ρ selon

$$v \ = \ \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \min_{\boldsymbol{q} \in \Delta(\mathcal{B})} \ r(\boldsymbol{p}, \boldsymbol{q}) = \ \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \min_{\underline{h} \in \mathcal{V}} \ \rho(\boldsymbol{p}, \underline{h}) = \ \min_{\underline{h} \in \mathcal{V}} \ \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \rho(\boldsymbol{p}, \underline{h})\,, \qquad (2.8)$$

où les deux premières égalités sont par définitions et où seule la troisième demande des explications, qui procèdent d'une application directe d'un théorème de minimax généralisé (mais qui peut encore être retrouvé comme cas particulier du lemme de Sion).

Une notion adéquate de regret. Il s'agit maintenant, étant donnée la ré-écriture de la valeur du jeu, de définir une notion de regret qui permette de reproduire la garantie donnée par (2.6) : le joueur A impose, en minimisant son regret, que son paiement moyen asymptotique est au moins v. Au vu de ce qui précède, il lui suffit de jouer de telle sorte que, quelle que soit la stratégie τ du joueur B,

$$\liminf_{T \to \infty} \left\{ \overline{r}_T - \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)\Big) \right\} \geqslant 0 \qquad \text{p.s. ;} \qquad (2.9)$$

ainsi, on définit le regret moyen de A au tour T en cas d'observations imparfaites comme

$$\overline{R}_T^{\text{imp}} = \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \ \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)\Big) - \overline{r}_T\,.$$

Minimiser le regret et assurer (2.9) signifieront désormais que $\limsup_{T \to \infty} \overline{R}_T^{\text{imp}} \leqslant 0$ p.s.

2.2.2 Résultats antérieurs à mes travaux

Le résultat fondamental suivant continue de justifier la notion de regret introduite à l'instant : non seulement il suffit de minimiser ce regret pour garantir la valeur du jeu comme paiement moyen asymptotique, mais en plus, il est effectivement possible de réaliser ceci. (Ce n'est par exemple pas le cas en général ici pour la notion initiale de regret $\overline{R}_T$.)

Théorème 2.8 (Référence : [Rus99]). *Il existe une stratégie pour le joueur A telle que*

$$\limsup_{T \to \infty} \overline{R}_T^{\text{imp}} \leqslant 0 \qquad \text{p.s.}$$

Comme [Rus99] l'énonce en conclusion, un prolongement de son résultat est d'exhiber une stratégie explicite assurant la minimisation du regret, afin de simplifier la démonstration ; en effet, celle qui y est présentée pour le Théorème 2.8 repose sur un théorème d'approchabilité abstrait énoncé par [MSZ94] et auquel, contrairement au Théorème 2.1, ne correspond pas de stratégie naturelle.

Par ailleurs, outre donc le mérite de procurer une preuve d'existence constructive et plus simple, cette stratégie explicite permettra d'étudier les vitesses de convergence du regret vers 0 et de déterminer les vitesses optimales possibles ; enfin, on note qu'il est souhaitable que cette stratégie admette une mise en œuvre computationnellement efficace. Ces extensions ont motivé les travaux postérieurs, qu'on passe maintenant en revue.

Recours au théorème d'approchabilité le plus simple dans un cas particulier. Dans le cas où la fonction de retour sur actions H ne dépend pas des actions du joueur A, [MS03] exhibe une stratégie explicite simple reposant sur le Théorème 2.1 ; mais dans le cas général, elle assure un résultat plus faible que (2.9).

Etude complète lorsque les retours sur actions procurent une information exhaustive

On considère dans ce cadre les situations d'observations imparfaites dans lesquelles les objectifs (2.4) et (2.9) coïncident, *id est*, celles où, pour toute loi $\boldsymbol{q}$ sur $\mathcal{B}$,

$$\max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \rho(\boldsymbol{p},\, H(\,\cdot\,,\boldsymbol{q})) = \max_{\boldsymbol{p}\in\Delta(\mathcal{A})} r(\boldsymbol{p},\boldsymbol{q}) = \max_{i\in\mathcal{A}} r(i,\boldsymbol{q})\,.$$

Notion d'information exhaustive. On parle d'information exhaustive procurée par les signaux parce que dans ce cas, il est en particulier vrai que l'égalité des lois des signaux $H(\,\cdot\,,\boldsymbol{q}) = H(\,\cdot\,,\boldsymbol{q}')$ entraîne l'égalité des quantités objectifs

$$\max_{i\in\mathcal{A}} r(i,\boldsymbol{q}) = \max_{i\in\mathcal{A}} r(i,\boldsymbol{q}')\,.$$

Cela arrive par exemple lorsque H révèle la loi choisie par le joueur B, c'est-à-dire que $H(\,\cdot\,,\boldsymbol{q}) = H(\,\cdot\,,\boldsymbol{q}')$ si et seulement si $\boldsymbol{q} = \boldsymbol{q}'$, une condition que l'on peut exprimer en termes de rang plein d'une certaine représentation matricielle de la fonction H.

Propriété de reconstruction des paiements en fonction des retours. Une autre situation importante, et qui forme presque la situation générique du cas d'information exhaustive, correspond à la possibilité de reconstruction de la fonction de paiement r à partir de la fonction de retours sur actions H. Elle a été formulée par [PS01]. On l'énonce uniquement dans le cas particulier où, pour tout couple $(i,j) \in \mathcal{A}\times\mathcal{B}$, la loi $H(i,j)$ est une masse de Dirac, en un signal $h(i,j)$. Sans perte de généralité on ré-encode l'ensemble $\mathcal{S}$ des signaux par un sous-ensemble fini de $[0,1]$: la condition de reconstruction est alors l'existence d'une fonction $f : \mathcal{A}\times\mathcal{A} \to \mathbb{R}$ telle que

$$\forall\,(i,j)\in\mathcal{A}\times\mathcal{B}, \qquad r(i,j) = \sum_{k\in\mathcal{A}} f(i,k)\,h(k,j)\,. \tag{2.10}$$

En fait, [PS01] montre que dans tous les cas d'information exhaustive on arrive à se ramener, par un algorithme effectuant des transformations élémentaires comme par exemple la duplication d'actions et de signaux, au cadre pré-cité de signaux déterministes et à une reconstruction du type (2.10).

Estimation des paiements non observés. Le joueur A peut alors estimer à chaque tour son paiement $r(I_t, J_t)$ et les paiements $r(i, J_t)$ qu'il aurait obtenus en jouant d'autres actions $i \in \mathcal{A}$, à partir de la seule information dont il dispose, le retour sur action déterministe $K_t = h(I_t, J_t)$. En effet, pour toute action $i \in \mathcal{A}$, la statistique

$$\widehat{r}_{i,t} = \frac{f(i, I_t)\, K_t}{p_{I_t,t}} = \frac{f(i, I_t)\, h(I_t, J_t)}{p_{I_t,t}}\,,$$

où l'action I_t est tirée selon la loi $\boldsymbol{p}_t$ (chargeant tout $\mathcal{A}$) et où $p_{I_t,t}$ désigne la I_t–ème composante de $\boldsymbol{p}_t$, est sans biais, conditionnellement aux variables aléatoires $\boldsymbol{p}_t$ et J_t :

$$\mathbb{E}\Big[\, \widehat{r}_{i,t} \,\Big|\, \boldsymbol{p}_t, J_t \Big] = \sum_{k \in \mathcal{A}} \frac{f(i,k)\, h(k, J_t)}{p_{k,t}}\, p_{k,t} = r(i, J_t)\,,$$

où l'on a utilisé la reconstruction (2.10).

C'est en fait [ACBFS02] qui a le premier proposé une telle estimation sans biais de paiements non observés, dans le cadre plus simple des problèmes de bandits à plusieurs bras du chapitre 4. (On pourra également jeter un œil aux techniques d'estimation des pertes proposées au paragraphe 1.3.1.)

Stratégie en découlant. La stratégie proposée par [PS01] consiste alors, essentiellement, en le choix, au tour $t \geqslant 2$, de l'équivalent suivant à la stratégie (2.2) du cas d'observations parfaites :

$$p_{i,t} = (1 - \gamma_t) \frac{\exp\Big(\eta_t \sum_{s=1}^{t-1} \widehat{r}_{i,s}\Big)}{\sum_{k \in \mathcal{A}} \exp\Big(\eta_t \sum_{s=1}^{t-1} \widehat{r}_{k,s}\Big)} + \frac{\gamma_t}{N}\,, \tag{2.11}$$

où $\eta_t > 0$ et $\gamma_t > 0$ sont deux paramètres à déterminer par l'analyse. L'interprétation est que d'une part, on remplace les paiements non observés par leurs estimateurs et que d'autre part, on impose, par mélange avec la loi uniforme, une exploration minimale de toutes les actions ; le premier terme de la définition (2.11) est quant à lui appelé terme d'exploitation et on a ici affaire à un arbitrage entre exploration et exploitation.

Du point de vue technique, la borne inférieure de γ_t/N imposée par (2.11) permet de contrôler les déviations des estimateurs $\widehat{r}_{i,t}$ autour de leurs espérances conditionnelles. [PS01] propose ainsi une majoration sur le regret $\overline{R}_T$ de l'ordre essentiellement de $T^{-1/4}$; il s'agit bien ici d'un contrôle du regret défini originellement en (2.1), ce qui est plus fort qu'un simple contrôle de $\overline{R}_T^{\mathrm{imp}}$.

2.2.3 Stratégie générale, explicite et efficace, issue de nos travaux

Nos travaux, dans la lignée des références précédentes, ont été réalisés en deux temps.

Premier temps : échauffement [3] par un approfondissement des résultats de [PS01]. Nous avons repris l'analyse de [PS01] et avons noté qu'elle pouvait être améliorée pour mener à un contrôle de l'ordre de $T^{-1/3}$ sur le regret $\overline{R}_T$; nous avons également montré par un exemple que c'était là la vitesse générale optimale dans le cadre d'information exhaustive. Il fallait cependant encore en (re)venir au cas général où seul le regret $\overline{R}_T^{\mathrm{imp}}$ peut être minimisé.

Second temps : analyse dans le cas général [6] et minimisation de $\overline{R}_T^{\mathrm{imp}}$. La remarque-clé a été que dans le cas général, il était illusoire et même dangereux de raisonner en termes des paiements $r(i, J_t)$, qu'il fallait rester concentré sur la quantité objectif

$$\max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)\Big)\,,$$

et que pour avoir une idée de cette dernière, il suffisait d'estimer $H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T\big)$. Or, cela est facile en recourant là encore à la forme d'estimateurs proposée par [ACBFS02] dans le contexte des bandits à plusieurs bras. Dans la suite, on identifie lorsque c'est nécessaire les lois sur $\mathcal{S}$ à des vecteurs de $\mathbb{R}^{\mathcal{S}}$.

On pose comme estimateur de la loi $H(i, J_t)$ sur les signaux, la statistique

$$\widehat{h}_{i,t} = \frac{\delta_{K_t}}{p_{i,t}} \mathbb{I}_{\{I_t=i\}}\,,$$

où l'on rappelle que l'on avait noté K_t le retour sur action disponible au tour t : c'est un signal tiré indépendamment au hasard selon $H(I_t, J_t)$. Cet estimateur est sans biais conditionnellement aux variables aléatoires $\boldsymbol{p}_t$ et J_t :

$$\begin{aligned}\mathbb{E}\Big[\,\widehat{h}_{i,t}\,\Big|\,\boldsymbol{p}_t, J_t\Big] = \frac{1}{p_{i,t}}\mathbb{E}\Big[\,\delta_{K_t}\mathbb{I}_{\{I_t=i\}}\,\Big|\,\boldsymbol{p}_t, J_t\Big] &= \frac{1}{p_{i,t}}\mathbb{E}\Big[\,H(I_t, J_t)\,\mathbb{I}_{\{I_t=i\}}\,\Big|\,\boldsymbol{p}_t, J_t\Big] \\ &= \frac{1}{p_{i,t}}\; p_{i,t}\, H(i, J_t) = H(i, J_t)\,,\end{aligned}$$

où l'on a pris les espérances d'abord par rapport à K_t puis par rapport à I_t.

On note Π l'opérateur de projection convexe (euclidienne) sur $\mathcal{V}$. Un argument de concentration de la mesure applicable dans des espaces hilbertiens [CW96] montre alors que pour un entier m, suffisamment grand, et pour tout entier $b \geqslant 0$,

$$\underline{\widehat{h}}^{\,b} = \Pi\left(\frac{1}{m}\sum_{t=bm+1}^{(b+1)m}\Big[\widehat{h}_{i,t}\Big]_{i\in\mathcal{A}}\right) \qquad \text{estime bien} \qquad \underline{h}^{\,b} = \frac{1}{m}\sum_{t=bm+1}^{(b+1)m} H(\,\cdot\,, J_t)\,. \tag{2.12}$$

Les deux autres ingrédients à injecter pour construire notre stratégie sont les suivants. Premièrement, on aura là aussi besoin d'imposer un compromis entre exploitation et exploration uniforme. Deuxièmement, la fonction ρ étant concave et uniformément lipschitzienne en son argument de $\Delta(\mathcal{A})$, on peut appliquer, de manière précisément contrôlée, une majoration linéaire du type (1.10). Pour tout point intérieur $\boldsymbol{p}$ de $\Delta(\mathcal{A})$

Paramètres : un entier $m \geqslant 1$, deux réels η, $\gamma > 0$

Configuration initiale : $\boldsymbol{w}^0 = (1, \dots, 1)$

A chaque itération $t = 1, 2, \dots$,

1. Calculer l'entier b tel que $bm + 1 \leqslant t \leqslant (b+1)m$;
2. Recourir à la loi $\boldsymbol{p}^b = (1-\gamma)\widetilde{\boldsymbol{p}}^b + \gamma \boldsymbol{u}$, où la probabilité $\widetilde{\boldsymbol{p}}^b$ est définie par la valeur de ses composantes selon
$$\widetilde{p}_i^b = \frac{w_i^b}{\sum_{k \in \mathcal{A}} w_k^b}, \qquad \text{pour } i \in \mathcal{A},$$
et où $\boldsymbol{u}$ désigne la loi uniforme sur $\mathcal{A}$;
3. Tirer l'action I_t au hasard selon $\boldsymbol{p}_t = \boldsymbol{p}^b$;
4. Obtenir le retour sur action K_t ;
5. Si $t = (b+1)m$, effectuer la mise à jour
$$w_i^{b+1} = w_i^b \exp\left(\eta \left(\partial\rho\Big(\boldsymbol{p}^b, \widehat{\underline{h}}^b\Big)\right)_i\right) \qquad \text{pour } i \in \mathcal{A},$$
où $\widehat{\underline{h}}^b$ est défini en (2.12) et où $\partial\rho$ désigne un sous-gradient de ρ par rapport à son premier argument.

FIGURE 2.1. Une stratégie minimisant le regret dans le cas d'un jeu avec observations imparfaites.

et tout élément $\underline{h}$ de $\mathcal{V}$, on note à cet effet $\partial\rho(\boldsymbol{p}, \underline{h})$ un sous-gradient de $\rho(\,\cdot\,, \underline{h})$ en $\boldsymbol{p}$. C'est un élément de $\mathbb{R}^N$ et on désignera sa i–ème composante en l'indexant par un i.

Forts de ces trois ingrédients, on propose ainsi la stratégie reproduite à la figure 2.1. Cette stratégie est simple, de mise en œuvre computationnellement efficace, et sa considération procure une preuve constructive au Théorème 2.8, avec l'indication de vitesses de convergence. Le théorème suivant et ses variantes reportent de manière un peu informelle les résultats obtenus dans [6] ; en particulier, nous y avions bien sûr précisé la valeur des paramètres calibrés et n'y avions formulé que des bornes de performances avec constantes explicites et valant en temps fini.

Théorème 2.9. *La stratégie de la figure 2.1, employée avec des paramètres bien calibrés, assure qu'avec probabilité* $1 - \delta$*,*

$$\overline{R}_T^{\text{imp}} \leqslant \mathcal{O}\Big(T^{-1/5}\sqrt{\ln(T/\delta)}\Big).$$

Des versions simplifiées de cette stratégie assurent par ailleurs que la vitesse de convergence vers 0 du regret est majorée, à un terme $\sqrt{\ln(T/\delta)}$ près : par $T^{-1/4}$ dans le cas où les retours sur actions sont aléatoires mais ne dépendent que de l'action du

joueur B ; par $T^{-1/3}$ dans le cas où les retours sur actions sont déterministes mais dépendent du couple d'actions choisies par les joueur A et B ; par $T^{-1/2}$ dans le cas où les retours sur actions sont déterministes et ne dépendent que de l'action du joueur B. Il découle par ailleurs des bornes minimax indiquées dans [CBFH+97] et [3] que les vitesses des cas des retours déterministes sont optimales, à des facteurs logarithmiques près.

2.2.4 Résultats postérieurs à mes travaux et perspectives

Vitesses de convergence optimales pour des stratégies elles aussi efficaces

Nous nous étions cependant demandé si les vitesses $T^{-1/5}$ et $T^{-1/4}$ que nous avions obtenues dans les cas de retours sur actions aléatoires étaient optimales ou non. [Per09c] a montré que ce n'était pas le cas, en exhibant une stratégie dont le regret est contrôlé avec grande probabilité $1-\delta$ par une quantité de l'ordre de $T^{-1/3}\ln(1/\delta)$. De même, une variante simplifiée de cette stratégie pour le cas où les retours sur actions sont aléatoires mais ne dépendent que de l'action du joueur B admet un regret au plus de l'ordre de $T^{-1/2}\ln(1/\delta)$. Ces vitesses sont optimales. De plus, les stratégies proposées sont efficaces du point de vue computationnel, car elles reposent sur la preuve de l'existence et la considération d'un sous-ensemble fini du simplexe $\Delta(\mathcal{A})$ contenant, pour tout vecteur $\underline{h} \in \mathcal{V}$, une meilleure réponse du joueur A à $\underline{h}$ au sens de ρ.

Extension du théorème d'approchabilité

On a décrit précédemment comment le théorème d'approchabilité (Théorème 2.1) entraînait l'existence de stratégies minimisant le regret dans le cas des observations parfaites. Ici, dans le cas d'observations imparfaites, on vient de discuter l'existence de stratégies minimisant le regret. A cet égard, on peut noter, comme l'énonce [Rus99], qu'assurer (2.9), c'est assurer que l'ensemble convexe fermé

$$\mathcal{C} = \left\{ (z, \boldsymbol{q}) \in \mathbb{R} \times \Delta(\mathcal{B}) : \quad z \geqslant \max_{\boldsymbol{p} \in \Delta(\mathcal{A})} \rho(\boldsymbol{p}, H(\,\cdot\,, \boldsymbol{q})) \right\}$$

est approchable pour la fonction $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R} \times \Delta(\mathcal{B})$ suivante : pour tout couple $(i, j) \in \mathcal{A} \times \mathcal{B}$,

$$m(i,j) = \begin{bmatrix} r(i,j) \\ \delta_j \end{bmatrix}.$$

[Per09a] propose la généralisation suivante du théorème d'approchabilité au cas des jeux avec observations imparfaites (et résout au passage un problème ouvert depuis une quinzaine d'années). Cette caractérisation pourrait cependant être améliorée sur deux points : elle vient sans stratégie efficace associée (des problèmes de complexité de mise en œuvre exponentielle en T se posent, qui sont décrits plus en détails au paragraphe 2.4.3) ; par ailleurs, aucune vitesse n'est précisée pour la convergence (2.5). Il serait intéressant de pallier cela.

Théorème 2.10 (Référence : [Per09a]). *On considère un ensemble convexe fermé $\mathcal{C} \subset \mathbb{R}^d$ et une fonction vectorielle $m : \mathcal{A} \times \mathcal{B} \to \mathbb{R}^d$. Alors, $\mathcal{C}$ est m–approchable si et seulement si*

$$\forall \underline{h} \in \mathcal{V}, \quad \exists \boldsymbol{p} \in \Delta(\mathcal{A}), \quad \forall \boldsymbol{q} \in \Delta(\mathcal{B}) \text{ tel que } H(\,\cdot\,, \boldsymbol{q}) = \underline{h}, \qquad m(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{C}\,.$$

La preuve de ce résultat profond repose d'une part sur diverses considérations et éléments techniques développés dans [6], et d'autre part, sur l'existence de stratégies minimisant le regret interne dans le cas d'observations imparfaites, cette existence découlant elle-même de l'existence de stratégies calibrées. Les parties suivantes de ce chapitre s'arrêteront sur ces deux notions, calibration et regret interne.

> Remarque au passage. Historiquement, les premières stratégies calibrées reposaient sur la considération de stratégies sans regret interne pour le cas des observations parfaites, ces stratégies calibrées étant alors la pierre angulaire de la construction de stratégies sans regret interne pour le cas d'observations imparfaites. Pour garantir un enchaînement dans l'ordre de la construction des objets, il fallait donc alterner les considérations entre regret interne et calibration. Or, un de nos travaux récents montre l'existence de stratégies calibrées de manière intrinsèque, directement par approchabilité. C'est pourquoi on peut désormais présenter d'abord la calibration puis la minimisation du regret interne, en deux phases bien distinctes, ce que l'on fait dans la suite.

2.3 Obtention directe de stratégies calibrées par approchabilité [9]

Dans le jeu de calibration, le joueur A doit prévoir les actions du joueur B. Ce dernier dispose toujours d'un ensemble d'actions $\mathcal{B}$ fini mais les actions du joueur A sont désormais données par $\Delta(\mathcal{B})$, l'ensemble des probabilités sur $\mathcal{B}$. On munit pour la suite $\Delta(\mathcal{B})$ de la topologie induite par l'inclusion canonique dans $\mathbb{R}^{\mathcal{B}}$ et, en particulier, on le rend mesurable par la considération de la tribu des boréliens.

Déroulement du jeu. A chaque tour, le joueur A et le joueur B choisissent simultanément et en fonction du passé leurs actions, notées respectivement $P_t \in \Delta(\mathcal{B})$ et $J_t \in \mathcal{B}$. Ces actions sont en fait tirées au hasard selon des lois ν_t sur $\Delta(\mathcal{B})$ et $\boldsymbol{q}_t$ sur $\mathcal{B}$.

Objectif de calibration. On fixe une norme $\|\cdot\|$ sur $\Delta(\mathcal{B})$, par exemple la norme ℓ^1. L'objectif du joueur A est de former une stratégie σ aux prévisions bien calibrées, c'est-à-dire qui assure que, quelle que soit la stratégie τ du joueur B,

$$\forall \varepsilon > 0, \quad \forall \boldsymbol{p} \in \Delta(\mathcal{B}), \qquad \lim_{T \to +\infty} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\left\{\|P_t - \boldsymbol{p}\| \leqslant \varepsilon\right\}} \left(P_t - \delta_{J_t}\right) \right\| = 0 \qquad \text{p.s.} \quad (2.13)$$

On appelle erreur de calibration (au tour T) la quantité tendant vers 0 ci-dessus. L'interprétation est la suivante : le joueur A veut garantir que pour toute loi $\boldsymbol{p}$ sur le comportement du joueur B, la distribution empirique des actions de ce dernier aux tours

où le premier avait prévu un comportement proche de $\boldsymbol{p}$ est en effet proche de $\boldsymbol{p}$. C'est une question de cohérence *a posteriori* des prévisions des lois avec leurs réalisations.

L'intérêt pour la suite de ce chapitre est le suivant : en fait, on utilise de telles stratégies calibrées comme des stratégies auxiliaires. En particulier, si le joueur A dispose d'un ensemble d'actions propre $\mathcal{A}$ et d'une fonction de paiement $r : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$, alors il peut choisir son action $I_t \in \mathcal{A}$ au tour t en fonction de la prévision P_t de la loi du comportement du joueur B proposée par une stratégie auxiliaire calibrée. Puisque les prévisions P_t de cette dernière sont précises au sens de (2.13), on imagine que ce faisant, le joueur A peut garantir des propriétés intéressantes sur son paiement moyen.

Passage en revue des travaux antérieurs. Avec l'humour qui le caractérise, Foster [Fos99] notait déjà :

> "Over the past few years many proofs of the existence of calibration have been discovered. Each of the following provides a different algorithm and proof of convergence: Foster and Vohra [FV91, FV98]; Hart [Har95]; Fudenberg and Levine [FL99]; Hart and Mas-Colell [HMC00]. Does the literature really need one more? Probably not."

Et malgré tout, il a alors proposé, avec succès, une stratégie calibrée pour le cas binaire où $\mathcal{B}$ n'est composé que de deux éléments ; cette stratégie est plus directe et plus courte que les autres stratégies existant précédemment et dont il a dressé la liste. Elle repose sur le théorème d'approchabilité (Théorème 2.1). En fait, à y regarder de près et avec le confort de l'observation distanciée rétrospective, on peut noter que toutes les stratégies des références mentionnées se fondent, à des degrés divers, sur des résultats d'approchabilité, apparaissant toutefois au détour de calculs pas nécessairement pétris d'intuition. Par exemple, la stratégie de [FV98] est fondée sur une stratégie auxiliaire minimisant un certain regret interne, cette dernière étant obtenue de manière naturelle par approchabilité.

[FL99] et [HMC00] considèrent le cas d'ensembles $\mathcal{B}$ avec un nombre fini mais arbitraire d'éléments ; ils ne proposent pas de vitesses de convergence vers 0 de l'erreur de calibration (2.13). Les stratégies de [FV91, FV98, Fos99] ne sont valides quant à elles que pour le cas binaire mais elles donnent lieu à des vitesses de convergence, de l'ordre de $T^{-1/4}$, à des facteurs logarithmiques près.

Notre contribution. Nous nous intéressons au cas général d'un nombre fini d'actions pour $\mathcal{B}$ et exhibons une stratégie simple (la plus simple connue à ce jour), fondée directement sur le théorème d'approchabilité et ce faisant, capturant l'essence des preuves d'existence précédentes ; par ailleurs, nous arrivons à préciser des vitesses de convergence vers 0 de l'erreur de calibration en termes du cardinal de $\mathcal{B}$.

2.3.1 Construction préliminaire d'une stratégie ε–calibrée

La plupart des travaux précisés ci-dessus [FV91, FV98, Fos99, FL99] ne s'attachent pas à prouver (2.13) directement mais passent d'abord par la version approchée suivante, appelée critère d'ε–calibration, pour un paramètre $\varepsilon > 0$ fixé. On appelle une ε–grille de $\Delta(\mathcal{B})$ un ensemble de points $\mathcal{G}_\varepsilon = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_\varepsilon}\}$ tels que les boules de centres $\boldsymbol{p}_k$ et de rayon ε recouvrent $\Delta(\mathcal{B})$ lorsque k décrit $\{1, \ldots, N_\varepsilon\}$.

Définition 2.11. Une stratégie du joueur A est ε–calibrée si elle ne forme que des prévisions appartenant à une ε–grille $\mathcal{G}_\varepsilon$ et que, quelle que soit la stratégie τ du joueur B,

$$\limsup_{T\to+\infty} \sum_{k=1}^{N_\varepsilon} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{P_t=\boldsymbol{p}_k\}} (\boldsymbol{p}_k - \delta_{J_t}) \right\| \leqslant \varepsilon \qquad \text{p.s.} \tag{2.14}$$

On a alors le résultat suivant.

Théorème 2.12. *A toute ε–grille $\mathcal{G}_\varepsilon$ de $\Delta(\mathcal{B})$, on peut associer une stratégie ε–calibrée construite à partir du théorème d'approchabilité.*

Démonstration. On a ici affaire à un jeu fini : les actions du joueur A sont désormais indexées par l'ensemble fini $\mathcal{G}_\varepsilon$ tandis que celles du joueur B sont toujours données par $\mathcal{B}$. On définit la fonction vectorielle $m : \mathcal{G}_\varepsilon \times \mathcal{B} \to \mathbb{R}^{\mathcal{G}_\varepsilon \times \mathcal{B}}$ suivante, en identifiant les lois sur $\mathcal{B}$ à des vecteurs de $\mathbb{R}^{\mathcal{B}}$: pour tous $k \in \{1, \ldots, N_\varepsilon\}$ et $j \in \mathcal{B}$,

$$m(\boldsymbol{p}_k, j) = (\underline{0}, \ldots, \boldsymbol{p}_k - \delta_j, \underline{0}, \ldots, \underline{0}) ,$$

qui est un vecteur composé de $k-1$ premiers éléments nuls $\underline{0} \in \mathbb{R}^{\mathcal{B}}$, suivis d'un élément non nul de $\mathbb{R}^{\mathcal{B}}$, et complétés par $N_\varepsilon - k$ autres éléments nuls.

On prend alors comme ensemble convexe fermé $\mathcal{C}$ la boule de rayon ε pour la norme $\|\cdot\|$ et de centre $(\underline{0}, \ldots, \underline{0})$. Or, la condition (2.14) d'ε–calibration se ré-écrit comme le fait que

$$\frac{1}{T}\sum_{t=1}^{T} m(P_t, J_t) = \left(\frac{1}{T}\sum_{t=1}^{T} \mathbb{I}_{\{P_t=\boldsymbol{p}_1\}} (\boldsymbol{p}_1 - \delta_{J_t}),\ \ldots,\ \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}_{\{P_t=\boldsymbol{p}_{N_\varepsilon}\}} (\boldsymbol{p}_{N_\varepsilon} - \delta_{J_t}) \right)$$

converge vers $\mathcal{C}$ presque-sûrement.

L'existence d'une stratégie ε–calibrée est donc équivalente à la m–approchabilité de $\mathcal{C}$, que nous prouvons maintenant en recourant à la caractérisation du Théorème 2.1. Soit $\boldsymbol{q} \in \Delta(\mathcal{B})$ une loi sur les actions du joueur B. Par construction de la grille $\mathcal{G}_\varepsilon$, il existe $k \in \{1, \ldots, N_\varepsilon\}$ tel que $\|\boldsymbol{p}_k - \boldsymbol{q}\| \leqslant \varepsilon$, de sorte que

$$m(\boldsymbol{p}_k, \boldsymbol{q}) \in C .$$

(Ici, la loi sur $\mathcal{G}_\varepsilon$ de la condition d'approchabilité peut donc être prise égale à une masse de Dirac.) □

Complexité en temps de calcul et en espace mémoire. Nous nous sommes ensuite attachés dans [9] à énoncer la stratégie associée au Théorème 2.12, en indiquant comment calculer à chaque tour t la projection convexe prescrite par la stratégie canoniquement associée au théorème d'approchabilité et comment associer à cette projection une loi ν_t sur $\mathcal{G}_\varepsilon$, par résolution (approchée) d'un programme linéaire. A des facteurs logarithmiques près, sa complexité de mise en œuvre à chaque tour, est de l'ordre de $\varepsilon^{-|\mathcal{B}|-1}$, où $|\mathcal{B}|$ désigne le cardinal de $\mathcal{B}$.

2.3.2 Obtention d'une stratégie calibrée

En suivant une méthodologie établie par [CBL06, paragraphe 4.5 et exercice 7.23], et mettant notamment en jeu des résultats de concentration de la mesure dans les espaces hilbertiens [CW96], nous avons alors pu prouver le résultat suivant, qui est, à notre connaissance, le premier résultat de vitesses de convergence pour l'erreur de calibration lorsque $\mathcal{B}$ contient plus de deux éléments, cette vitesse étant même uniforme. (La notion d'utilisation d'une stratégie par blocs a été définie au paragraphe 1.4.1.)

Théorème 2.13. *Une stratégie jouant les stratégies du Théorème 2.12 par blocs assure que*

$$\limsup_{T\to\infty} \frac{T^{1/(|\mathcal{B}|+1)}}{\sqrt{\ln T}} \sup_{\boldsymbol{p}\in\Delta(\mathcal{B})} \sup_{\varepsilon>0} \left\| \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}_{\left\{\|P_t-\boldsymbol{p}\|\leqslant\varepsilon\right\}} \left(P_t - \delta_{J_t}\right) \right\| \leqslant \Gamma_{|\mathcal{B}|} \qquad \text{p.s.},$$

où $\Gamma_{|\mathcal{B}|}$ désigne une constante ne dépendant que de $|\mathcal{B}|$.

En fait, nous prouvons même que l'uniformité peut être plus forte : l'erreur uniforme de calibration peut être définie en termes de $P_t \in \mathcal{L}$, où $\mathcal{L}$ est un borélien quelconque, et le supremum porte alors sur l'ensemble des boréliens.

2.3.3 Comparaisons plus précises aux travaux antérieurs et postérieurs, et perspectives

Comparaisons aux travaux antérieurs et postérieurs

En termes de vitesses de convergence. Le seul résultat antérieur de vitesses de convergence apparaissant explicitement et dont nous ayons connaissance est le suivant. [CBL06, paragraphe 4.5] indique comment obtenir des vitesses de convergence (là aussi uniformes) pour l'erreur de calibration d'une stratégie fondée sur la stratégie ε–calibrée proposée par [FV98] dans le cas $|\mathcal{B}| = 2$ uniquement. Cette vitesse est de l'ordre de $T^{-1/4}$ à des facteurs logarithmiques près, à comparer à la vitesse $T^{-1/3}$ procurée par le Théorème 2.13.

Il ne nous semblait pas évident d'étendre la stratégie fondée sur [FV98] au cas non binaire où $|\mathcal{B}| > 2$. Or, [Per10] a réalisé avec succès une telle extension, à partir d'une modification des stratégies fondamentales proposées par [FV98] et en exploitant assez largement sur l'analyse développée dans [9]. (En deux mots, à l'attention des spécialistes uniquement : les stratégies modifiées ne minimisent plus leur regret interne sur l'ε–grille

que par rapport à des plus proches voisins, et non plus par rapport à tous les éléments de la grille.) Ce faisant, il a obtenu la même vitesse que celle proposée au Théorème 2.13. Il a également démontré qu'une vitesse $T^{-1/2}$, indépendante du cardinal de $\mathcal{B}$, pouvait être obtenue pour une erreur de calibration uniquement uniforme en une base de voisinages dénombrable de $\Delta(\mathcal{B})$.

En termes de complexités de mise en œuvre des stratégies ε–calibrées. Dans le cas général d'un ensemble d'actions $\mathcal{B}$ fini, la stratégie de [Per10] requiert un temps de calcul et une place mémoire au moins proportionnels à $\varepsilon^{1-|\mathcal{B}|}$ et dans tous les cas, inférieurs à une quantité proportionnelle à $\varepsilon^{2(1-|\mathcal{B}|)}$; le calcul précis de ces complexités n'y est pas effectué.

Dans le cas binaire, qui a de loin été le plus étudié, la meilleure complexité de calcul et d'espace mémoire pour une stratégie ε–calibrée est de l'ordre de $1/\varepsilon$, pour la stratégie simple et totalement explicite exhibée par [Fos99]. Ceci est à comparer aux complexités $1/\varepsilon^2$ et $1/\varepsilon^3$ obtenues respectivement par les stratégies de [FV98] et [9] dans ce cas.

Perspectives

A notre connaissance, aucun résultat de minoration n'est disponible en calibration, que soit une minoration de la vitesse de convergence des erreurs de calibration vers 0 ou une minoration des complexités (en temps de calcul et en espace mémoire) des stratégies ε–calibrées. En termes des minorants de complexité, ils apparaîtront peut-être sous la forme d'un compromis à respecter entre la complexité en temps de calcul et celle en espace mémoire. La question sous-jacente est en fait de savoir s'il existe des stratégies calibrées efficaces, dont la complexité n'augmente pas exponentiellement avec le cardinal de $\mathcal{B}$, comme c'est le cas pour les stratégies connues pour l'instant. L'objectif en termes des vitesses de convergence est peut-être plus clair : on peut espérer que les vitesses $T^{-1/(|\mathcal{B}|+1)}$ exhibées à la fois par [9] et [Per10] soient optimales.

2.4 Convergence vers l'ensemble des équilibres corrélés [1, 4]

Après ce détour (technique mais qui se montrera nécessaire) par la calibration, on en revient aux convergences vers des ensembles d'équilibres, comme celles étudiées aux paragraphes 2.1.2 et 2.1.3 et qui justifiaient la notion de regret. On en reprend d'ailleurs le cadre et les notations qui y étaient considérées, au moins dans un premier temps.

2.4.1 Cas des jeux à ensembles d'actions $\mathcal{A}$ et $\mathcal{B}$ finis

Une autre notion d'équilibre. La notion d'équilibre du paragraphe 2.1.2 est peu considérée en théorie des jeux ; on lui préfère par exemple les équilibres de Nash (dont les équilibres minimax sont un cas particulier) et les équilibres corrélés. Cependant, les premiers sont (NP–)difficiles à calculer en général : par conséquent, on ne peut pas s'attendre à ce qu'il existe des stratégies simples et efficaces pour chacun des joueurs telles que si chacun la

suit, alors il y ait convergence en un sens à préciser vers l'ensemble des équilibres de Nash.

Ce n'est pas le cas des équilibres corrélés [Aum74, Aum87], qui sont exprimés, comme les équilibres au sens de Hannan, en termes de lois jointes et à partir d'un nombre polynômial (en les nombres d'actions) de contraintes linéaires. Ces contraintes sont formalisées en termes de fonctions $\varphi : \mathcal{A} \to \mathcal{A}$ et $\psi : \mathcal{B} \to \mathcal{B}$, appelées fonctions de déviations.

Définition 2.14. L'ensemble des équilibres corrélés d'un jeu fini à deux joueurs est donné par l'ensemble (non vide) des lois jointes

$$\mathcal{E} = \left\{ \pi \in \Delta(\mathcal{A}\times\mathcal{B}) : \quad \forall \varphi : \mathcal{A} \to \mathcal{A}, \ \sum_{i,j} \pi(i,j)\, r(i,j) \geqslant \sum_{i,j} \pi(i,j)\, r\big(\varphi(i),j\big) \right.$$
$$\left. \text{et} \quad \forall \psi : \mathcal{B} \to \mathcal{B}, \ \sum_{i,j} \pi(i,j)\, s(i,j) \geqslant \sum_{i,j} \pi(i,j)\, s\big(i,\psi(j)\big) \right\}.$$

L'interprétation d'un tel équilibre est la suivante : on suppose qu'un couple d'actions (I, J) est tiré au hasard selon π par un médiateur et que les joueurs sont invités à jouer chacun l'action qui a été tirée pour eux. Alors, en espérance, si un joueur respecte cette invitation, l'autre joueur n'a pas intérêt à remplacer l'action qui lui est proposée par une autre action qu'il aurait déterminée en fonction de cette proposition (*via* une fonction de déviation). On parle ici encore de déviations unilatérales non profitables.

On comparera cette interprétation à celle formulée après la définition 2.2 : la seule différence est que pour les équilibres corrélés, la déviation peut être formulée en termes de l'action proposée, alors que pour les équilibres au sens de Hannan, la déviation était fixée à l'avance (et correspondait donc à une fonction de déviation constante). On a en particulier $\mathcal{E} \subseteq \mathcal{H}$, ce qui montre que l'objectif ci-dessous est plus ambitieux que celui poursuivi au paragraphe 2.1.2. En revanche, on peut noter que pour les jeux à somme nulle du paragraphe 2.1.3, on a l'inclusion $\mathcal{N} \subseteq \mathcal{E}$.

Objectif. On s'intéresse à la distribution empirique des couples d'actions joués,

$$\overline{\pi}_T = \frac{1}{T} \sum_{t=1}^{T} \delta_{(I_t, J_t)}\,,$$

et on veut montrer que lorsque les deux joueurs minimisent chacun indépendamment de l'autre un regret dit interne, il y a convergence de la suite $(\overline{\pi}_T)$ vers l'ensemble $\mathcal{E}$ des équilibres corrélés. C'est un résultat assez remarquable en un sens : même si chaque joueur utilise une stratégie myope et ne se préoccupe que peu de l'autre joueur, on obtient, à la limite une corrélation forte entre leurs comportements.

Définition et intérêt du regret interne

Il est facile de voir que dans la définition de $\mathcal{E}$, il suffit de se restreindre aux fonctions φ et ψ ne différant de l'identité qu'en un point, de sorte que $\mathcal{E}$ se ré-écrit

$$\mathcal{E} = \left\{ \pi \in \Delta(\mathcal{A}\times\mathcal{B}) : \quad \forall (i,k)\in\mathcal{A}^2,\ \sum_{j\in\mathcal{B}} \pi(i,j)\, r(i,j) \geqslant \sum_{j\in\mathcal{B}} \pi(i,j)\, r(k,j) \right.$$
$$\left. \text{et} \quad \forall (j,\ell)\in\mathcal{B}^2,\ \sum_{i\in\mathcal{A}} \pi(i,j)\, s(i,j) \geqslant \sum_{i\in\mathcal{A}} \pi(i,j)\, s(i,\ell) \right\}.$$

Partant de cette observation, on pourrait définir le regret interne (moyen) de la stratégie du joueur A comme

$$\max_{(i,k)\in\mathcal{A}^2} \frac{1}{T}\sum_{t=1}^{T} (r(k,J_t) - r(i,J_t))\mathbb{I}_{\{I_t=i\}}\,, \tag{2.15}$$

mais par les lemmes de Hoeffding–Azuma et de Borel–Cantelli, le comportement asymptotique de la quantité précédente est égal à celui de la quantité suivante, qui est légèrement plus simple à manipuler et qu'on retient donc pour la définition du regret interne (moyen) du joueur A :

$$\overline{R}_T^{\text{int}} = \max_{(i,k)\in\mathcal{A}^2} \frac{1}{T}\sum_{t=1}^{T} p_{i,t}\big(r(k,J_t) - r(i,J_t)\big)\,.$$

On définit bien sûr, de manière similaire, le regret interne moyen du joueur B :

$$\overline{S}_T^{\text{int}} = \max_{(j,\ell)\in\mathcal{B}^2} \frac{1}{T}\sum_{t=1}^{T} q_{j,t}\big(s(I_t,\ell) - s(I_t,j)\big)\,.$$

On cherche ici à minimiser les regrets internes, *id est*, à exhiber des stratégies telles que

$$\limsup_{T\to\infty} \overline{R}_T^{\text{int}} \leqslant 0 \quad \text{p.s.} \qquad \text{et} \qquad \limsup_{T\to\infty} \overline{S}_T^{\text{int}} \leqslant 0 \quad \text{p.s.}$$

Une adaptation immédiate des techniques de preuve menant aux Propositions 2.3 et 2.7 conduit au résultat suivant.

Proposition 2.15. Si les deux joueurs minimisent leur regret interne, alors la suite des distributions empiriques des couples d'actions joués, $(\overline{\pi}_T)$, converge presque-sûrement vers l'ensemble $\mathcal{E}$ des équilibres corrélés.

Transformation automatique de stratégies minimisant le regret externe en stratégies minimisant le regret interne [1]

Le regret dit externe correspond à la notion originelle de regret, introduite au paragraphe 2.1. On propose dans [1] une ré-interprétation de la première stratégie proposée

pour minimiser le regret interne, celle de [FV99], comme étant une stratégie minimisant un certain regret externe.

Pour toute loi $\boldsymbol{p}$ sur $\mathcal{A}$ et tout couple $(i,k) \in \mathcal{A}^2$ avec $i \neq k$, on note $\boldsymbol{p}^{i\to k}$ l'image de $\boldsymbol{p}$ par la fonction de déviation $\varphi_{i\to k} : \mathcal{A} \to \mathcal{A}$ qui ne diffère de l'identité qu'au point i, où $\varphi(i) = k$; c'est-à-dire que $\boldsymbol{p}^{i\to k}$ et $\boldsymbol{p}$ ne diffèrent elles-mêmes que pour les probabilités associées à i et k, qui valent respectivement 0 et p_i pour i, d'une part, $p_i + p_k$ et p_k d'autre part pour k.

On remarque alors qu'on peut ré-écrire le regret interne comme le regret (externe) moyen par rapport à la classe des modifications de la stratégie maître paramétrées par les $\varphi_{i\to k}$:

$$\overline{R}_T^{\text{int}} = \max_{i\neq k} \sum_{t=1}^{T} r(\boldsymbol{p}_t^{i\to k},\, J_t) - \sum_{t=1}^{T} r(\boldsymbol{p}_t, J_t)\,.$$

C'est d'ailleurs *via* cette ré-écriture qu'on comprend le mieux l'appellation regret interne : la classe de comparaison n'est plus intrinsèque et externe à la stratégie mais construite en fonction de cette dernière.

Tout se passe donc, pour cette notion de regret, comme si les actions du joueur A étaient indexées par les $\varphi_{i\to k}$. On choisit alors à chaque tour t une loi $\boldsymbol{p}_t$ vérifiant l'équation de point fixe suivante, obtenue par imitation à partir de (2.2) et avec les mêmes choix de vitesses d'apprentissages (η_t) :

$$\boldsymbol{p}_t = \sum_{i\neq k} \frac{\exp\left(\eta_t \sum_{s=1}^{t-1} r(\boldsymbol{p}_s^{i\to k},\, J_s)\right)}{\sum_{i'\neq k'} \exp\left(\eta_t \sum_{s=1}^{t-1} r(\boldsymbol{p}_s^{i'\to k'},\, J_s)\right)}\, \boldsymbol{p}_t^{i\to k}\,; \tag{2.16}$$

une telle loi existe bien comme on peut le voir en l'identifiant à une probabilité stationnaire d'une certaine chaîne de Markov finie. Il découle alors des résultats rappelés autour de (2.2) que le regret interne moyen de cette stratégie, bien qu'aléatoire, est contrôlé de manière déterministe (avec probabilité 1) selon

$$\overline{R}_T^{\text{int}} \leqslant \|r\|_\infty \sqrt{\frac{2}{T} \ln\big(N(N-1)\big)}\,.$$

Remarque au passage. L'argument précédent peut être généralisé et permet en fait de transformer toute stratégie minimisant le regret externe (pas seulement celle par pondération par poids exponentiels) en une stratégie minimisant le regret interne ; il s'applique même au cadre de la prévision par agrégation convexe. Un autre tel schéma de transformation automatique (ne valant cependant que dans le cas de la prévision randomisée) a été proposé indépendamment par [BM07].

2.4.2 Extension à des jeux à ensembles d'actions $\mathcal{A}$ et $\mathcal{B}$ convexes et compacts [4]

On se place désormais dans le cas où les ensembles d'actions $\mathcal{A}$ et $\mathcal{B}$ sont des espaces topologiques munis chacun de leur tribu borélienne et où les fonctions de paiement $r : \mathcal{A}\times\mathcal{B} \to \mathbb{R}$ et $s : \mathcal{A}\times\mathcal{B} \to \mathbb{R}$ sont des applications mesurables. On notera encore par I_t et J_t les actions choisies par les joueurs à chaque tour t et elles le seront par tirage selon des lois $\boldsymbol{p}_t$ et $\boldsymbol{q}_t$ sur $\mathcal{A}$ et $\mathcal{B}$.

Extension de la définition des équilibres corrélés

On commence par rappeler l'extension de la définition 2.14 des équilibres corrélés d'un jeu fini à ce cas plus général, telle que proposée par [HS89]. On note respectivement à cet effet $\mathcal{L}^0(\mathcal{A})$ et $\mathcal{L}^0(\mathcal{B})$ les ensembles de fonctions mesurables $\mathcal{A} \to \mathcal{A}$ et $\mathcal{B} \to \mathcal{B}$.

Définition 2.16. L'ensemble des équilibres corrélés d'un jeu à deux joueurs disposant chacun d'un ensemble d'actions donné par un espace topologique muni de sa tribu borélienne est constitué par l'ensemble (non vide) des lois jointes

$$\mathcal{E} = \Bigg\{ \pi \in \Delta(\mathcal{A}\times\mathcal{B}) : \quad \forall \varphi \in \mathcal{L}^0(\mathcal{A}), \quad \mathbb{E}_\pi\big[r(I,J)\big] \geqslant \mathbb{E}_\pi\Big[r\big(\varphi(I),J\big)\Big] \\ \text{et} \quad \forall \psi \in \mathcal{L}^0(\mathcal{B}), \quad \mathbb{E}_\pi\big[s(I,J)\big] \geqslant \mathbb{E}_\pi\Big[s\big(I,\psi(J)\big)\Big] \Bigg\}, \quad (2.17)$$

où la notation $\mathbb{E}_\pi$ signifie que le vecteur aléatoire (I,J) à valeurs dans $\mathcal{A}\times\mathcal{B}$ admet pour loi π.

Formulé ainsi, l'ensemble $\mathcal{E}$ est un sous-ensemble de $\Delta(\mathcal{A}\times\mathcal{B})$ correspondant, sauf cas particulier, à une infinité non dénombrable de contraintes ; or, pour pouvoir définir un regret interne pouvant être minimisé facilement par une stratégie, on préférerait avoir affaire à un nombre au plus dénombrable de contraintes. C'est possible sous un certain nombre de conditions de régularité précisées dans le lemme suivant. On y note $\mathcal{C}(\mathcal{A})$ et $\mathcal{C}(\mathcal{B})$ les ensembles de fonctions continues $\mathcal{A} \to \mathcal{A}$ et $\mathcal{B} \to \mathcal{B}$.

Lemme 2.17. Lorsque les ensembles $\mathcal{A}$ et $\mathcal{B}$ sont chacun un sous-ensemble convexe et compact d'un certain espace vectoriel normé, les sous-espaces vectoriels $\mathcal{C}(\mathcal{A})$ et $\mathcal{C}(\mathcal{B})$, munis de la norme du supremum, sont séparables. On fixe alors deux sous-ensembles $\mathcal{D}(\mathcal{A})$ et $\mathcal{D}(\mathcal{B})$ dénombrables et denses respectivement dans $\mathcal{C}(\mathcal{A})$ et $\mathcal{C}(\mathcal{B})$. Si en outre les fonctions r et s sont continues, alors $\mathcal{E}$ peut être défini de manière équivalente en ne considérant dans (2.17) que les fonctions $\varphi \in \mathcal{D}(\mathcal{A})$ et $\psi \in \mathcal{D}(\mathcal{B})$.

La preuve de ce lemme s'effectue en deux temps. On montre premièrement, par une version *ad hoc* du théorème de Lusin (le fait que toute fonction mesurable soit presque, en un sens à préciser, une fonction continue), qu'il suffit de considérer dans (2.17) les fonctions continues. On montre ensuite, par convergence dominée, qu'en fait, on peut même se restreindre aux deux sous-ensembles denses $\mathcal{D}(\mathcal{A})$ et $\mathcal{D}(\mathcal{B})$.

Minimisation d'un regret interne généralisé

Convergence vers l'ensemble des équilibres corrélés. On se place dans la suite sous les hypothèses du Lemme 2.17. On dit que le joueur A minimise son regret interne lorsqu'il peut assurer qu'à la limite, aucune des déviations procurées par les éléments d'un

sous-ensemble dénombrable dense $\mathcal{D}(\mathcal{A})$ de $\mathcal{C}(\mathcal{A})$ n'est profitable, *id est*,

$$\forall \varphi \in \mathcal{D}(\mathcal{A}), \qquad \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\Big(r(I_t, J_t) - r\big(\varphi(I_t), J_t\big)\Big) \geqslant 0\,; \tag{2.18}$$

et évidemment, une définition similaire pour le joueur B. Ici, en général (sauf hypothèses supplémentaires, voir [4, paragraphe 5]), on ne peut effectuer de minimisation uniforme en les fonctions de déviation et c'est pourquoi l'on ne dispose plus de résultats de vitesses de convergence.

Mais la convergence elle-même est encore garantie, comme l'énonce le lemme ci-dessous, dont la preuve est fondée sur le théorème de Prohorov. Ce dernier assure que lorsque $\mathcal{A}$ et $\mathcal{B}$ (et donc $\mathcal{A}\times\mathcal{B}$) sont des espaces métriques compacts, l'ensemble $\Delta(\mathcal{A}\times\mathcal{B})$ est également un espace métrique compact pour la topologie de la convergence faible–$\star$. Là encore, le même raisonnement qui avait conduit aux Propositions 2.3 et 2.7 (ensemble d'équilibres défini par des contraintes de type fermé, argument de compacité séquentielle et raisonnement par l'absurde) permet alors d'obtenir le résultat suivant.

Proposition 2.18. Sous les hypothèses du Lemme 2.17, si les deux joueurs minimisent leur regret interne, alors la suite des distributions empiriques des couples d'actions joués, $(\overline{\pi}_T)$, converge presque-sûrement vers l'ensemble $\mathcal{E}$ des équilibres corrélés.

Comment minimiser le regret interne. Il ne reste donc plus qu'à montrer que chaque joueur peut minimiser son regret interne ; dans ce qui suit, nous ne nous intéressons qu'au joueur A et nous plaçons encore sous les conditions du Lemme 2.17. Nous allons ici dévier légèrement de la méthode présentée dans [4], qui reposait sur la considération d'une fonction de pondération générale et qui requérait alors la définition de stratégies par blocs. Ici, nous avions restreint l'exposé, déjà dans le cas fini, à la pondération par poids exponentiels, laquelle peut en fait directement traiter le cas d'un nombre dénombrable d'experts (ici, identifiés aux fonctions de déviation).

On considère à cet effet une loi de probabilité $\mu = (\mu_\varphi)_{\varphi\in\mathcal{D}(\mathcal{A})}$ sur $\mathcal{D}(\mathcal{A})$, affectant une probabilité strictement positive à toute fonction φ, et on recourt à la stratégie qui à chaque tour t choisit une loi $\boldsymbol{p}_t \in \Delta(\mathcal{A})$ vérifiant l'équation de point fixe

$$\boldsymbol{p}_t = \sum_{\varphi\in\mathcal{D}(\mathcal{A})} \frac{\mu_\varphi \exp\Big(\eta_t \sum_{s=1}^{t-1} r\big(\boldsymbol{p}_s^\varphi,\, J_s\big)\Big)}{\sum_{\phi\in\mathcal{D}(\mathcal{A})} \mu_\phi \exp\Big(\eta_t \sum_{s=1}^{t-1} r\big(\boldsymbol{p}_s^\phi,\, J_s\big)\Big)}\; \boldsymbol{p}_t^\varphi\,, \tag{2.19}$$

où pour toute loi $\boldsymbol{p}$ sur $\mathcal{A}$ et toute fonction $\varphi \in \mathcal{L}^0(\mathcal{A})$, la loi $\boldsymbol{p}^\varphi$ désigne la loi image de $\boldsymbol{p}$ par φ, *id est*, la loi de $\varphi(I)$ lorsque la variable aléatoire I admet pour loi $\boldsymbol{p}$.

Il reste à voir qu'un tel point fixe existe toujours. Pour la topologie (métrisable) sur $\Delta(\mathcal{A})$ donnée par la topologie faible–$\star$, toutes les applications $\boldsymbol{p} \mapsto \boldsymbol{p}^\varphi$ sont continues, de sorte que le membre de gauche de (2.19) est une fonction continue de $\boldsymbol{p}_t$. L'existence de la stratégie est alors assurée par application du théorème de point fixe de Schauder–Cauty, que nous prenons plaisir à rappeler ci-dessous sous sa forme la plus générale et la plus aboutie.

Théorème 2.19 (Point de fixe de Schauder–Cauty [Cau01]). *Soit C un ensemble non vide, convexe et compact d'un espace vectoriel topologique séparé. Alors toute application continue $T : C \to C$ admet au moins un point fixe.*

Une extension (assez immédiate) des résultats autour de (2.2) au cas d'un nombre dénombrable d'actions est possible et montre que la stratégie recourant à des lois telles que les équations (2.19) soient vérifiées assure le contrôle déterministe (non uniforme) :

$$\forall \varphi \in \mathcal{D}(\mathcal{A}), \qquad \frac{1}{T}\sum_{t=1}^{T}\Big(r(\boldsymbol{p}_t, J_t) - r(\boldsymbol{p}_t^{\varphi}, J_t)\Big) \geqslant \Gamma_\varphi \sqrt{T}\,,$$

où la constante Γ_φ dépend de chaque φ. Une application conjointe des lemmes de Hoeffding–Azuma et Borel–Cantelli entraîne alors que le regret du joueur A est bien minimisé au sens de (2.18).

2.4.3 Extensions au cas des jeux finis avec observations imparfaites

On présente très brièvement dans cette partie les ingrédients nécessaires à la preuve du résultat fondamental qu'est le Théorème 2.10. En effet, on avait notamment indiqué à la suite de son énoncé qu'elle reposait sur l'utilisation de stratégies minimisant le regret interne dans le cas d'observations imparfaites. Pour ce faire, on se place à nouveau dans le cadre du paragraphe 2.2.

Présentation des résultats de [LS07, Per09a, Per09b, Per09c]

La fonction de paiement pessimiste ρ étant concave (et en général non linéaire) en son argument $\boldsymbol{p} \in \Delta(\mathcal{A})$, la meilleure réponse à un vecteur $\underline{h} \in \mathcal{V}$ est (en général) une loi de $\Delta(\mathcal{A})$ non donnée par une masse de Dirac. Partant de cette observation, les travaux mentionnés se placent dans l'extension mixte du jeu originel, où à chaque tour, les joueurs choisissent encore des lois respectives $\boldsymbol{p}_t \in \Delta(\mathcal{A})$ et $\boldsymbol{q}_t \in \Delta(\mathcal{B})$, appelées désormais actions mixtes, mais au lieu de devoir tirer des actions I_t et J_t au hasard selon ces lois, ils obtiennent directement les paiements $r(\boldsymbol{p}_t, \boldsymbol{q}_t)$ et $s(\boldsymbol{p}_t, \boldsymbol{q}_t)$. Evidemment, par application simple de résultats de concentration, les paiements moyens du jeu originel et de son extension mixtes sont asymptotiquement égaux.

La définition (2.15) du regret interne dans le cas des jeux finis est alors étendue de la manière suivante au cas des extensions mixtes ; on ne la donne que pour le joueur A. On suppose que ce dernier n'a le droit de proposer que des lois issues d'un certain sous-ensemble fini $\{\boldsymbol{p}^f,\ f \in \mathcal{F}\}$ de $\Delta(\mathcal{A})$, ce qui est une hypothèse forte mais semble raisonnable dès lors que par exemple, ce sous-ensemble forme une grille assez fine de $\Delta(\mathcal{A})$. Pour toutes les actions mixtes $\boldsymbol{p}^f$ ayant été jouées au moins une fois avant le temps T, on note

$$\overline{\boldsymbol{q}}_T(\boldsymbol{p}^f) = \frac{1}{\sum_{t=1}^{T} \mathbb{I}_{\{\boldsymbol{p}_t = \boldsymbol{p}^f\}}} \sum_{t=1}^{T} \boldsymbol{q}_t\, \mathbb{I}_{\{\boldsymbol{p}_t = \boldsymbol{p}^f\}}$$

la moyenne des actions mixtes choisies par le joueur B sur l'ensemble des tours où le joueur A a recouru à l'action mixte $\boldsymbol{p}^f$. Minimiser le regret interne, c'est alors assurer qu'aucun remplacement systématique d'une action mixte par une autre, les choix du joueur B étant fixés, n'augmente de manière significative le paiement moyen ; c'est-à-dire que quelle que soit la stratégie τ du joueur B, quel que soit $f \in \mathcal{F}$, on ait, presque-sûrement,

$$\liminf_{T\to\infty} \left(\frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{\boldsymbol{p}_t = \boldsymbol{p}^f\}} \right) \left(\max_{\boldsymbol{p}\in\Delta(\mathcal{A})} \rho\Big(\boldsymbol{p},\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T(\boldsymbol{p}^f)\big)\Big) - \rho\Big(\boldsymbol{p}^f,\, H\big(\,\cdot\,, \overline{\boldsymbol{q}}_T(\boldsymbol{p}^f)\big)\Big)\right) \;\geqslant 0\,.$$

[LS07, Per09b, Per09c] proposent des stratégies minimisant le regret interne ; les constructions de [Per09b, Per09c] requièrent l'existence de stratégies calibrées. [Per09a] déduit alors de la considération de stratégies minimisant le regret interne la partie directe de l'équivalence du Théorème 2.10.

Critiques et perspectives

A propos de la définition du regret interne. L'extension de la définition du regret interne proposée ci-dessus est effectuée par similarité à (2.15) mais elle repose sur une restriction importante du joueur A, qui est celle qu'il ne peut choisir ses actions mixtes que dans un sous-ensemble fini (ce dernier étant certes laissé à sa discrétion). Par ailleurs, dans le cas d'observations procurant une information exhaustive, cette définition ne coïncide pas nécessairement avec (2.15) ; or, nous avons montré dans [3] comment minimiser le regret interne au sens de (2.15) dans ce cas. En fait, dans les deux cas, il existe un sous-ensemble fini d'actions mixtes qui rend intrinsèque la définition du regret interne dans le cas d'observations imparfaites : c'est celui qui contient, pour tout vecteur $\underline{h} \in \mathcal{V}$, une meilleure réponse du joueur A à $\underline{h}$ au sens de ρ (son existence, déjà mentionnée plus haut, étant prouvée par [Per09c]). Par exemple, dans le cas de l'information exhaustive, ce sous-ensemble est constitué par les masses de Dirac en les actions de $\mathcal{A}$.

Cependant, une objection majeure demeure : le lien entre minimisation du regret interne et convergence vers un ensemble d'équilibres (les équilibres corrélés de l'extension mixte, par exemple) n'est pas établi ici. Or, tant pour le regret externe dans le cas d'observations imparfaites que pour le regret interne dans le cas des jeux à ensembles d'actions convexes et compacts, c'est en étudiant la structure et les contraintes définissant les ensembles d'équilibres que nous avons pu déterminer quelles étaient les bonnes notions de regret à minimiser : on comparera à cet effet (2.8) et (2.9) d'une part, et (2.18) et le Lemme 2.17 d'autre part.

L'étude et la définition de ce potentiel ensemble d'équilibres limite ne sont pas si claires ni si évidentes de prime abord ; on rappelle d'ailleurs, au passage, que [4, note de bas de page numéro 3] montre qu'il y a identité entre les équilibres corrélés d'un jeu et de son extension mixte, au sens où il y a une surjection canonique de l'ensemble de ces derniers vers celui des premiers.

En termes de complexités de mise en œuvre. Si les stratégies proposées par [LS07, Per09b] admettent une complexité de mise en œuvre exponentielle en T, ce n'est pas le cas de celle de [Per09c], qui est efficace. Ceci est à rapprocher de la stratégie efficace admettant des vitesses optimales pour le contrôle du regret externe, à laquelle nous avions fait allusion au paragraphe 2.2.4.

Cependant, aucune stratégie efficace ou naturelle n'est associée au Théorème 2.10, par opposition à ce qui se passe pour le Théorème 2.1, car la construction de [Per09a] repose sur la minimisation d'un regret interne par rapport à une grille de taille exponentielle. Une question intéressante serait d'effectuer une construction plus satisfaisante d'une stratégie canonique d'approchabilité associée au Théorème 2.10.

2.5 Perspectives et projets de recherche

On rappelle simplement ici qu'au cours du texte de ce chapitre, plus précisément dans les paragraphes 2.2.4, 2.3.3 et 2.4.3 clôturant la mise en perspective des trois lignées de travaux présentées, nous avons formulé de manière précise des projets de recherche, tous valables plutôt à court terme et permettant d'affiner la compréhension ou l'appréciation des résultats déjà obtenus. C'est avec Shie Mannor et Vianney Perchet que je compte m'attaquer à la plupart d'entre eux.

CHAPITRE 3

Applications des techniques d'agrégation convexe séquentielle : avancées méthodologiques et études empiriques

Introduction. Les nouveaux résultats théoriques en prévision de suites individuelles sont souvent éprouvés sur des données artificielles. A vrai dire, au mieux de notre connaissance, seules peu d'études sur données réelles ont été menées jusqu'à ce jour, alors qu'en principe le cadre méta-statistique d'agrégation des prévisions d'experts a vocation à s'appliquer à tout problème pour lequel on peut construire plusieurs tels experts – autant dire, à presque tous les problèmes de prévision séquentielle.

La première lignée de tels travaux empiriques a porté sur l'investissement séquentiel dans le marché boursier et a été initiée par [Cov91] ; elle est décrite un peu plus en détails dans ce chapitre. Une seconde lignée plus récente considère la prévision de résultats sportifs, voir [DMP+06] ou [VZ08]. Les prévisions des experts y sont données par les cotes indiquées par différents bookmakers ou par celles résultant des paris de nombreux participants sur un site web de paris en ligne.

Les cadres et jeux de données que nous avons considérés sont les suivants : la prévision de consommation électrique, d'une part, où Goude [Gou08a, Gou08b] a le premier illustré l'intérêt des méthodes d'agrégation séquentielle, et la prévision de la qualité de l'air, d'autre part, pour laquelle Mallet et Sportisse [MS06] fournit des experts et une première étude de stratégies simples d'agrégation.

Table des matières

3.1 Résumé des avancées méthodologiques [10]

L'article de survol [10], dont sont tirés de larges extraits de ce chapitre, résume, à destination de mathématiciens, des avancées méthodologiques en prévision de suites individuelles formulées au passage dans les articles [7, 13], qui s'adressent en effet davantage à un public de praticiens que de théoriciens. Nous présentons ces avancées

dans la suite de cette partie. A cet effet, on reprend le cadre et les notations du paragraphe 1.1.2.

3.1.1 Calibration pratique des paramètres

Où la théorie est trop précautionneuse... Comme on le verra sur les données réelles, les valeurs optimales de η pour minimiser les bornes théoriques, par exemple celle du Théorème 1.7 ou celles obtenues par les techniques adaptatives relatées au paragraphe 1.4, conduisent en général à de mauvaises performances. On peut expliquer cela par le fait que ces bornes sont conçues par rapport à l'ensemble de toutes les suites individuelles et correspondent ainsi à des algorithmes trop précautionneux et qui ont un temps de réaction un peu trop long. Une idée naturelle consiste donc à augmenter la valeur des vitesses d'apprentissage η_t, toute la question étant d'exhiber une manière adaptative et performante de le faire.

... Mais où la théorie est utile malgré tout ! Notons, avant de continuer, que ces remarques sur un apprentissage plus rapide ne remettent pas en cause notre méthodologie d'agrégation de prédicteurs. Là encore, les résultats pratiques montreront que des stratégies d'agrégation calibrées avec des vitesses suffisamment rapides obtiennent des performances bien meilleures que celles du meilleur expert ou du meilleur vecteur de mélange constant $\boldsymbol{q}$ (qui, en outre, ne sont connus qu'à la fin de la période de prévision) et que cela se traduit, au niveau des vecteurs de mélange calculés par les stratégies, par des vecteurs $\boldsymbol{q}_t$ qui ne ressemblent absolument pas à des masses de Dirac ni ne sont tous proches d'un même vecteur de mélange.

Calibration par considération des performances empiriques d'une famille paramétrée

On présente la méthode par exemple pour la famille $\mathcal{E}_\eta^{\mathrm{grad}}$ des stratégies de pondération par poids exponentiels des gradients des pertes cumulées, décrite à la figure 1.3 ; cette méthode repose sur la considération de toutes les stratégies de cette famille (lorsque $\eta > 0$ varie) et c'est pourquoi on recourra à l'appellation de méta-stratégie de calibration. A cet effet, il nous faudra écrire explicitement la dépendance du vecteur de mélange $\boldsymbol{p}_t$ en la stratégie $\mathcal{E}_\eta^{\mathrm{grad}}$ qui le prescrit, ce que l'on fait en le notant $\boldsymbol{p}_t(\mathcal{E}_\eta^{\mathrm{grad}})$.

Enoncé. La méta-stratégie de calibration choisit, à l'échéance t, le paramètre η_t associé à la stratégie ayant obtenu les meilleurs performances dans le passé puis recourt au vecteur de mélange $\boldsymbol{p}_t(\mathcal{E}_{\eta_t}^{\mathrm{grad}})$. Formellement,

$$\eta_t \in \operatorname*{arg\,min}_{\eta>0} \ \widehat{L}_{t-1}(\mathcal{E}_\eta^{\mathrm{grad}}) \,. \tag{3.1}$$

Absence actuelle de borne théorique. Même si, comme on le verra, cette technique obtient d'excellentes performances pratiques au sens où sa perte cumulée jusqu'à l'échéance T se rapproche souvent de celle de $\mathcal{E}_{\eta_T}^{\mathrm{grad}}$, *id est*, de la meilleure stratégie de la famille

$\mathcal{E}_\eta^{\mathrm{grad}}$ sur les données, nous n'avons pour l'instant pas encore de garanties théoriques à proposer sur son regret.

Résolution des difficultés de mise en œuvre. Par ailleurs, le calcul pratique d'un antécédent du minimum, même à un facteur de tolérance près, est une opération délicate. Une idée est d'effectuer la minimisation (3.1) non pas sur tout l'espace des paramètres $\mathbb{R}_+^\star$ mais sur une grille finie de points, éventuellement construite adaptativement. Cette grille apparaît souvent sous la forme de points logarithmiquement uniformément répartis entre deux bornes. Le pas de discrétisation importe peu, comme l'ont montré nos études empiriques ; en revanche, la valeur des bornes est, elle, un peu plus cruciale et [13, 10] expliquent (sans toutefois le mettre en œuvre) comment faire varier adaptativement les bornes au cours du temps selon les performances obtenues dans le passé par les grilles considérées. La validation pratique de cette procédure reste à effectuer.

3.1.2 Deux variantes des stratégies considérant des pertes cumulées

Partons d'une critique habituelle. On reproche souvent aux stratégies des chapitres précédents, et cela n'est pas sans lien avec la mention ci-dessus de leur tempérament parfois précautionneux, de tenir trop compte du passé : en effet, les pertes cumulées utilisées à l'échéance t pour former les vecteurs de mélange accordent une importance égale aux pertes récentes et aux pertes très anciennes. Or, si le passé proche semble nécessaire et utile, l'utilisation du passé lointain paraît souvent moins profitable (notamment à tous ceux qui sont familiers du cadre stochastique, et encore plus lorsque ce dernier est non stationnaire).

Le fenêtrage : une méthode sans garantie théorique

Une variante de nos stratégies ne reposant que sur le passé proche est donnée par la considération d'une fenêtre maximale d'historique H. Par exemple, pour les stratégies $\mathcal{E}_\eta^{\mathrm{grad}}$, elle consiste en le remplacement des définitions des composantes de $\boldsymbol{p}_t$ à la figure 1.3 par

$$p_{j,t} = \frac{\exp\left(-\eta \sum_{s=\max\{1,t-H\}}^{t-1} \widetilde{\ell}_{j,s}\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=\max\{1,t-H\}}^{t-1} \widetilde{\ell}_{k,s}\right)}$$

pour tous $t \geqslant 2$ et $j = 1, \ldots, N$.

Malheureusement, il semble cependant peu vraisemblable que des bornes sur le regret uniformes en les suites individuelles soient conservées par fenêtrage.

Réconciliation des points de vue : l'escompte des pertes

Il suffit en fait de considérer que le passé est d'autant plus significatif qu'il est proche, sans toutefois décréter que le passé lointain soit inutile ; cela se traduit mathématiquement en escomptant les pertes passées par un facteur multiplicatif strictement positif mais d'autant plus petit que le passé est lointain.

Enoncé. Ici encore, on illustre la méthode sur la famille des $\mathcal{E}_\eta^{\mathrm{grad}}$. On fixe à cet effet deux suites décroissantes de réels strictement positifs, les escomptes (β_t) et les vitesses d'apprentissage (η_t) ; et on remplace alors la définition des composantes de $\boldsymbol{p}_t$ à la figure 1.3 par

$$p_{j,t} = \frac{\exp\left(-\eta_t \sum_{s=1}^{t-1}(1+\beta_{t-s})\,\widetilde{\ell}_{j,s}\right)}{\sum_{k=1}^{N}\exp\left(-\eta_t \sum_{s=1}^{t-1}(1+\beta_{t-s})\,\widetilde{\ell}_{k,s}\right)} \tag{3.2}$$

pour tous $t \geqslant 2$ et $j = 1, \ldots, N$.

Existence de garanties théoriques sur la performance. Des bornes uniformes sur le regret de cette stratégie peuvent cette fois être prouvées ; elles dépendent évidemment des suites (η_t) et (β_t), cette dernière ne devant pas décroître trop rapidement vers 0. On a plus précisément le résultat suivant.

Théorème 3.1. *La stratégie de pondération par escomptes présentée en (3.2) minimise son regret au sens de l'objectif 1.1 lorsque les vitesses d'apprentissage et les escomptes vérifient que*

$$t\,\eta_t \to \infty \qquad \text{et} \qquad \eta_t \sum_{s=1}^{t-1} \beta_s \longrightarrow 0$$

quand $t \to \infty$ et que l'hypothèse 1.6 est vérifiée.

Démonstration. Une preuve détaillée est fournie dans le rapport technique [MMS07, chapitre 6]. L'idée consiste en un schéma par approximation, où l'on quantifie les écarts, plutôt faibles, entre les vecteurs de mélange (3.2) définis avec escomptes et ceux sans escomptes (figure 1.3, avec vitesses η_t dépendant du passé) ; puis l'on ajoute ces écarts à la borne d'une version adaptative du Théorème 1.7 (voir les commentaires qui le suivent). □

Références. Il est à noter que ces techniques d'escompte ont été introduites en prévision de suites individuelles dans [CBL06, paragraphe 2.11], à ceci près qu'il est absolument crucial pour l'analyse qui y est effectuée que le nombre d'échéances T ait une valeur fixée et connue à l'avance – ce qui n'est pas une condition dont nous pouvions nous satisfaire. En théorie des jeux, les escomptes permettent de modéliser notamment les taux d'intérêts et accordent par conséquent un poids d'autant plus important aux paiements qu'ils sont anciens.

3.1.3 Régressions linéaires séquentielles avec facteurs de régularisation

Cadre de la régression linéaire séquentielle. On se limite dans cet ensemble de paragraphes au cas où les ensembles d'observations et de prévisions sont donnés par la droite réelle, $\mathcal{Y} = \mathcal{X} = \mathbb{R}$, et où la fonction de perte ℓ est la perte quadratique, $\ell(x, y) = (x - y)^2$. On autorise en outre le choix, à chaque échéance t, de vecteurs de mélange arbitraires

$\boldsymbol{u}_t = (u_{1,t}, \dots, u_{N,t}) \in \mathbb{R}^N$; on n'impose aucune contrainte de positivité ou de somme égale à 1. La prévision du statisticien est alors formée par la combinaison linéaire

$$\widehat{y}_t = \sum_{j=1}^{N} u_{j,t} f_{j,t} \,. \tag{3.3}$$

Débiaisement possible mais moindre interprétation. L'avantage de ce cadre est qu'en ôtant la condition de somme égale à 1, on peut espérer que les stratégies de prévision compensent de manière automatique un biais éventuel commun à tous les experts ; c'est d'ailleurs ce que l'on observe en pratique sur certains jeux de données. L'inconvénient est cependant qu'en général les vecteurs de mélange $\boldsymbol{u}_t$ ont de nombreuses composantes négatives, ce qui les rend bien moins facilement interprétables que leurs homologues convexes $\boldsymbol{p}_t$.

Régularisation ℓ^2 : la régression ridge

La première stratégie de prévision étudiée dans ce nouveau cadre est la régression ridge, introduite par [HK70] dans un contexte stochastique et étudiée par [AW01] et [Vov01] dans le cadre des suites individuelles.

Enoncé. Cette stratégie repose sur une régularisation ℓ^2 ; on note à cet effet

$$\|\boldsymbol{u}\|_2 = \sqrt{\sum_{j=1}^{N} u_j^2}$$

la norme euclidienne d'un vecteur $\boldsymbol{u} \in \mathbb{R}^N$. Elle dépend d'un paramètre $\lambda > 0$ et on la note $\mathcal{R}_\lambda$; elle choisit, à toute échéance $t \geqslant 1$, un vecteur de mélange $\boldsymbol{u}_t$ vérifiant

$$\boldsymbol{u}_t \in \operatorname*{arg\,min}_{\boldsymbol{v} \in \mathbb{R}^N} \left\{ \lambda \|\boldsymbol{v}\|_2^2 + \sum_{s=1}^{t-1} \left(y_s - \sum_{j=1}^{N} v_j f_{j,s} \right)^2 \right\} \tag{3.4}$$

avec la convention habituelle qu'une somme sur aucun élément est nulle (de sorte que $\boldsymbol{u}_1$ est le vecteur nul).

Garanties de performances. Pour les énoncer de manière compacte, on définit le vecteur (ligne) $\boldsymbol{f}_t = (f_{1,t}, \dots, f_{N,t})$ des prévisions des experts à l'échéance t.

Théorème 3.2 (voir la synthèse de [CBL06, paragraphe 11.7]). *On définit la matrice $M_T = \sum_{t=1}^{T} \boldsymbol{f}_t^{\mathrm{T}} \boldsymbol{f}_t$ et on note $\mu_{1,T}, \dots, \mu_{N,T}$ ses valeurs propres. Pour toutes les suites de prévisions $\boldsymbol{f}_1, \dots, \boldsymbol{f}_T$ et d'observations $y_1, \dots, y_T$, le regret de $\mathcal{R}_\lambda$ par rapport*

à tout vecteur de mélange $\boldsymbol{v} \in \mathbb{R}^N$ *est contrôlé de la manière suivante :*

$$\sum_{t=1}^{T}\left(y_t - \sum_{j=1}^{N} u_{j,t} f_{j,t}\right)^2 - \sum_{t=1}^{T}\left(y_t - \sum_{j=1}^{N} v_j f_{j,t}\right)^2 \leqslant \frac{\lambda}{2}\|\boldsymbol{v}\|_2^2 + \left(\sum_{j=1}^{N} \ln\left(1 + \frac{\mu_{j,T}}{\lambda}\right)\right) \max_{t \leqslant T}\left(y_t - \sum_{j=1}^{N} u_{j,t} f_{j,t}\right)^2 .$$

Remarques et commentaires. La restriction de $\mathcal{X}$ et $\mathcal{Y}$ à un domaine borné $[-B, B]$ permet d'ôter toute dépendance de la borne en la suite des prévisions $f_{j,t}$ des experts et des observations y_t (au prix toutefois d'une dégradation de l'ordre de grandeur de la borne sur le regret ; les détails sont omis). On conclut ce paragraphe en notant qu'ici encore, un problème de calibration de λ se pose. La détermination de la valeur théorique optimale qui minimiserait la borne du Théorème 3.2 n'est pas évidente ; en pratique, on recourra plutôt aux techniques de calibration séquentielles du paragraphe 3.1.1.

Régularisation ℓ^1 : Lasso séquentiel

Nous avons alors introduit la variante suivante, plus proche des méthodes actuellement considérées en statistique.

Enoncé. Une version plus moderne de la régression linéaire régularisée remplace la régularisation ℓ^2 utilisée dans (3.4) par une régularisation ℓ^1 : on note à cet effet

$$\|\boldsymbol{u}\|_1 = \sum_{j=1} |u_j|$$

la norme ℓ^1 d'un vecteur $\boldsymbol{u} \in \mathbb{R}^N$. Pour une constante de régularisation $\lambda > 0$ fixée, la stratégie de Lasso séquentiel choisit, à chaque échéance $t \geqslant 1$, un vecteur de mélange

$$\boldsymbol{u}_t \in \operatorname*{arg\,min}_{\boldsymbol{v} \in \mathbb{R}^N} \left\{ \lambda \|\boldsymbol{v}\|_1 + \sum_{s=1}^{t-1}\left(y_s - \sum_{j=1}^{N} v_j f_{j,s}\right)^2 \right\} . \tag{3.5}$$

On la note $\mathcal{L}_\lambda$.

Remarques et commentaires. [Tib96] a introduit et étudié le Lasso dans un cadre stochastique ; ce dernier a depuis connu un succès fulgurant et s'est révélé remarquablement adapté aux problèmes de régression en grande dimension. En effet, l'avantage de la régularisation ℓ^1 de (3.5) est qu'elle permet de retenir des vecteurs $\boldsymbol{u}_t$ ayant peu de composantes non nulles. Un inconvénient, cependant, est qu'il n'existe pas d'expression explicite pour les solutions de (3.5), même si des algorithmes, par exemple l'algorithme LARS de [EHJT04], permettent de calculer efficacement leur valeur sur des données. Ceci est à comparer au problème (3.4), pour lequel il est aisé d'exhiber une telle expression explicite de $\boldsymbol{u}_t$ en fonction de λ et du passé.

Perspective de recherche. A ce jour et au meilleur de notre connaissance, il n'existe pas de borne sur le regret de la stratégie de Lasso séquentiel en termes de suites individuelles (la forme souhaitée de la borne serait celle du Théorème 3.2).

3.2 Interlude : Plan des études empiriques

On précise ici un plan de bataille standardisé à adopter face à un nouveau jeu de données empiriques, pour lequel on veut étudier de manière *rétrospective* quel aurait été l'intérêt des méthodes d'agrégation séquentielle (quelles auraient été leurs performances opérationnelles). On suppose donc disposer de l'ensemble des observations de la période considérée.

Plan des études empiriques des performances des méthodes d'agrégation séquentielle

1. Construire un certain nombre d'experts.
2. Déterminer un critère de qualité (une fonction de perte) et évaluer les performances des experts.
3. Calculer la performance du meilleur choix rétrospectif des paramètres pour chaque famille de stratégies.
4. Mesurer le coût de l'adaptativité et déterminer les performances opérationnelles.

1. Construire un certain nombre d'experts. On veillera, si possible, à en exhiber de comportement assez différents les uns des autres, afin que les stratégies d'agrégation aient une flexibilité suffisante dans les valeurs possibles de leurs prévisions agrégées. Cette construction relève en principe du partenaire du statisticien, qui connaît bien le domaine d'application visé et les méthodes, nouvelles ou éprouvées, qui sont susceptibles d'obtenir de bonnes performances. Ces méthodes peuvent avoir été calibrées sur des données antérieures au jeu de données considéré (voir par exemple la construction des experts pour la consommation électrique au paragraphe 3.5.2).

2. Déterminer un critère de qualité et évaluer les performances des experts. On entend ici le calcul de la précision de prévision (i.e., de la perte cumulée) obtenue par quelques stratégies simples, comme la moyenne uniforme des prévisions des experts (qui est une stratégie facile à mettre en œuvre de manière séquentielle) ou les oracles suivants. Par oracles, on entend des stratégies impossibles à déterminer de manière séquentielle, sans avoir vu toutes les données : le meilleur expert global, la meilleure combinaison convexe des experts, la meilleure combinaison linéaire des experts. Enfin, la stratégie dite presciente, qui n'est contrainte que par l'obligation de choisir chaque jour une combinaison linéaire ou convexe des experts mais connaît les données, indique la borne de performance qu'aucune stratégie de prévision par agrégation ne peut améliorer.

3. Calculer la performance du meilleur choix rétrospectif des paramètres pour chaque famille de stratégies. Les stratégies d'agrégation requièrent souvent le choix par l'utilisateur

d'un petit nombre de paramètres (la plupart du temps, un ou deux). Par exemple, la famille $\mathcal{E}_\eta^{\mathrm{grad}}$ des stratégies de pondération par poids exponentiels des pertes cumulées repose sur le choix d'un paramètre η : il s'agit alors d'en tabuler les performances sur une grille fine de paramètres η et de comparer la meilleure performance obtenue aux performances des stratégies de référence de l'étape précédente. Il est absolument souhaitable d'améliorer ces dernières.

4. Mesurer le coût de l'adaptativité et déterminer les performances opérationnelles. On applique ensuite la technique de calibration du paragraphe 3.1.1 sur les familles de stratégies considérées au point précédent et on regarde combien la performance de la méta-stratégie de calibration diffère de celles des stratégies sur lesquelles elle repose. En fait, ce point de l'étude est le plus crucial parce qu'il précise la performance que l'on aurait obtenue en effectuant réellement des prévisions séquentielles fondées sur les experts construits au premier point. C'est pourquoi l'on parle de performances opérationnelles.

3.3 Investissement séquentiel dans le marché boursier [1]

La première lignée de travaux empiriques sur l'intérêt pratique des techniques de suites individuelles a porté sur l'investissement séquentiel dans le marché boursier ; elle a été initiée par [Cov91], les observations étant formées par les évolutions journalières de 36 valeurs boursières de la bourse de New-York sur la période couvrant 1963 à 1985 et les experts étant simplement identifiés à chacune de ces valeurs boursières. Une vingtaine d'articles au moins a étudié les performances obtenues par des stratégies d'agrégation séquentielle sur ce jeu de données, mais nous ne discutons dans la suite que des résultats obtenus par [HSSW98, BEYG00, GLU06, AHKS06], ainsi que par notre contribution [1].

Mise en garde. Dans cette partie, nous décrivons, brièvement, pourquoi ce cadre et ce jeu de données sont en l'état peu raisonnables et ont été beaucoup critiqués, tant par le monde académique lui-même que par les professionnels de la R&D financière que nous avons interrogés.

Résumé rapide des résultats obtenus. [BEYG00] montre que les gains résultant des stratégies d'agrégation de [Cov91] ne sont pas très éloignés de ceux correspondant à l'allocation séquentielle uniforme, où chaque jour, on redistribue les capitaux de façon à ce qu'ils soient uniformément répartis dans les 36 valeurs boursières ; cela correspond à l'utilisation d'un vecteur de mélange uniforme. [HSSW98] (par utilisation de stratégies $\mathcal{E}_\eta^{\mathrm{grad}}$), [1] (par minimisation d'un certain regret interne), puis [AHKS06] (par agrégation convexe reposant sur une optimisation d'un critère par méthode de Newton) ont amélioré graduellement les performances financières de stratégies séquentielles possédant par ailleurs des garanties en termes de contrôle de leur regret par rapport à des suites

individuelles. [GLU06] et d'autres travaux à sa suite obtiennent des résultats meilleurs de plusieurs ordres de grandeurs mais en faisant appel à des stratégies fondées sur une hypothèse stochastique sur le comportement du marché : qu'il puisse être modélisé par un processus stationnaire.

Critiques apportées. Au meilleur de notre connaissance, la plupart des études empiriques, y compris la nôtre, ne mettent pas en œuvre tout le plan d'attaque décrit au paragraphe 3.2, mais essentiellement ses points 2 et 3. En particulier, et c'est là la critique la plus importante, elles ne construisent pas d'experts (point 1), au sens où elles se limitent à identifier valeurs boursières et experts. Au contraire, il faudrait, en lien avec l'industrie financière, étudier l'agrégation de stratégies d'investissement primaires, fondées sur les techniques habituelles de mathématiques financières.

Par ailleurs, ces études se contentent souvent de tabuler les performances obtenues pour divers choix de paramètres possibles, sans étudier le passage à des résultats opérationnels *via* une calibration automatique (point 4). Malgré cela, les performances obtenues sont souvent loin de celles obtenues par le meilleur vecteur de mélange constant, ce qui contraste fortement avec les améliorations obtenues par les stratégies d'agrégation par rapport aux stratégies de référence dans d'autres cadres, décrits ci-dessous.

A ces critiques académiques sur la démarche, s'en ajoutent d'autres sur le jeu de données lui-même. D'une part, ce dernier ne comporte que des valeurs boursières ayant survécu durant toute la période considérée ; d'autre part, les chercheurs, par exemple [GLU06], ont rapidement réalisé que les plus impressionnants gains obtenus par des stratégies séquentielles étaient surtout liés à deux valeurs boursières, "Kin Ark" et "Iroquois", qui ont des comportements cycliques corrélés et fortement prévisibles.

Enfin, les liens avec l'industrie financière sont parfois un peu unilatéraux, les professionnels de cette dernière étant demandeurs de techniques nouvelles et aimant s'entretenir avec les chercheurs mais étant souvent réticents à expliquer plus en détails les procédés qu'eux emploient et à nous fournir des retours suffisamment informatifs sur les performances et l'intérêt des stratégies d'agrégation séquentielle ; cela rend difficile la construction d'experts en partenariat, un point pourtant essentiel du plan d'étude formulé au paragraphe 3.2. Par ailleurs, il semble que le cadre de travail mathématique, nécessairement idéalisé, ne soit pas assez proche des contraintes opérationnelles.

Conclusion. C'est pourquoi, après ma thèse, je ne me suis plus intéressé à ce cadre d'investissement séquentiel. Je lui en ai préféré deux autres, dans lesquels l'étape préliminaire a consisté en l'identification de partenaires solides et déjà réputés dans le milieu, pouvant notamment procurer différents experts fondamentaux de bonne qualité.

3.4 Prévision de la qualité de l'air [7]

Dans cette partie, on présente le contexte et les données relatifs au problème de prévision de qualité de l'air. On explique brièvement comment adapter les stratégies générales

(celles présentées ci-dessus et dans les chapitres précédents) à ce cadre applicatif, avant d'en détailler les performances pratiques.

Par souci de concision, on ne décrira que les grandes lignes des résultats obtenus ; davantage de détails peuvent être obtenus en consultant les articles [7, 10], de même que les rapports techniques [GMS08, MMS07] sur lesquels ces articles s'appuient.

3.4.1 Description du jeu de données et des experts utilisés

Les données étudiées correspondent en temps à la période du 28 avril au 31 août 2001, qui comporte donc $T = 126$ jours, et en espace à un ensemble de 241 sites (appelés également stations dans la suite) en France et en Allemagne : 116 sites en France métropolitaine, 81 sites en Allemagne, uniformément répartis dans chacun des deux pays. Ici, on ne discutera que des résultats obtenus pour la prévision des pics journaliers d'ozone : à chaque jour t et à chaque site s, on associe la quantité y_t^s, qui est la valeur maximale de la concentration en ozone au cours de la journée en ce site. Les indices t et s prennent respectivement leurs valeurs dans les ensembles $\{1, \dots, 126\}$ et $\mathcal{N} = \{1, \dots, 241\}$. Les mesures de concentration sont données en microgrammes par mètre cube ($\mu\mathrm{g\,m^{-3}}$), une unité généralement omise par la suite. On rappelle à cet égard que les concentrations typiques sont de l'ordre de 40 $\mu\mathrm{g\,m^{-3}}$ à 150 $\mu\mathrm{g\,m^{-3}}$ et que les seuils légaux d'information et d'alerte sont respectivement de 180 $\mu\mathrm{g\,m^{-3}}$ et 240 $\mu\mathrm{g\,m^{-3}}$.

Données manquantes. On a donc affaire ici à la prévision d'environ 30 000 pics, mais seuls environ 27 500 d'entre eux ont été effectivement mesurés au cours de la période étudiée (les valeurs manquantes étant à attribuer, notamment, à des pannes de mesure ponctuelles en certaines stations). On notera dans la suite $\mathcal{N}_t$ l'ensemble des stations actives au jour t, de sorte que pour t fixé, seules les observations y_t^s avec $s \in \mathcal{N}_t$ sont disponibles.

Autres jeux de données. D'autres jeux de données sont considérés dans [7] : l'un européen et l'autre français (fondé sur la base de données BDQA gérée par l'ADEME et regroupant des observations effectuées par une quarantaine d'associations agréées pour la surveillance de la qualité de l'air). Par ailleurs, on y étudie également la prévision horaire des concentrations, pour les trois jeux de données décrits.

Construction des experts utilisés

On dispose de $N = 48$ experts, construits précédemment dans [MS06] et intégrés dans la plateforme de prévision Polyphemus [1]. En fait, chaque expert est le résultat de trois choix : un modèle de diffusion physico-chimique des polluants atmosphériques ; un jeu de données d'entrée (notamment, des données météorologiques et d'émission de

[1] Voir `http://cerea.enpc.fr/polyphemus/`

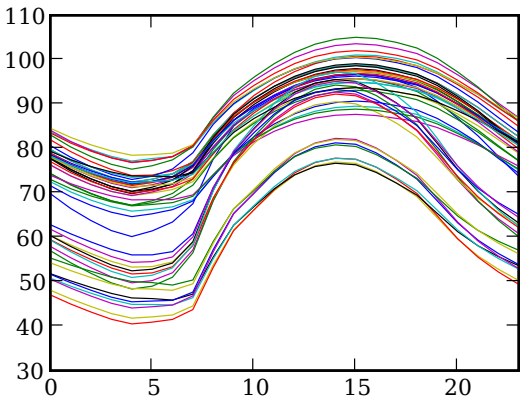

Figure 3.1. Profils de prévisions de concentration ($\mu g\,m^{-3}$) en ozone proposés par les 48 experts : moyennes des prévisions horaires en espace et en les jours de prévisions. Abscisses : heures de la journée ; ordonnées : concentrations.

polluants) ; un schéma de résolution numérique des équations aux dérivées partielles en jeu (choix d'une discrétisation spatiale et temporelle par exemple). Les choix considérés et leur combinaison pour former les 48 experts sont expliqués en détails dans [MS06, paragraphe 2.2]. De notre point de vue, cependant, les experts sont des boîtes noires prédictives, dont il s'agit d'améliorer la qualité de prévision de manière automatique. Les experts sont indexés par $j \in \{1, \dots, 48\}$ et proposent chacun, pour chaque jour t et chaque station s, une prévision de pic notée $f^s_{j,t}$. En réalité, ils proposent même un champ de prévisions sur tout l'Europe, *id est*, une prévision pour chaque point d'une grille fine de l'espace européen.

Des experts aux comportements fortement différents. La figure 3.1 montre que les prévisions des experts sont fortement dispersées : bien que l'on prenne la moyenne de leurs prévisions horaires en temps (en les jours de la période de prévision) et en espace, il y a un écart d'un facteur multiplicatif 2 entre les experts proposant les prévisions de concentration les plus fortes et les plus faibles. On notera que la forme des courbes correspond au profil typique de la concentration d'ozone au cours de la journée (creux de concentration à la fin de la nuit, pic en fin d'après-midi) mais qu'en aucun cas les experts ne sont donnés par des translations d'un expert de référence ; c'est uniquement la moyenne en temps et en espace qui concourt à produire ces profils similaires. Comme on le verra par la suite, les experts ont en effet des comportements et performances variables en temps et en espace.

Agrégation uniforme en espace mais variable en temps

On se restreint ici aux stratégies d'agrégation proposant le même vecteur de mélange convexe $\boldsymbol{p}_t$ ou linéaire $\boldsymbol{u}_t$ des prévisions des experts en tous les sites ; c'est-à-dire que

ce vecteur dépend de t uniquement, mais pas de s. C'est une contrainte que l'on peut lever afin de gagner en performance (voir [7, paragraphe A.1]) mais qui a l'avantage de faire gagner en interprétabilité et en force de prévision : en effet, les experts proposant chacun un champ de prévisions, on obtient alors un champ agrégé de prévisions, ce qui permet de proposer des prévisions même en dehors des stations (l'évaluation de la qualité de la prévision ne pouvant toutefois avoir lieu, elle, qu'en ces stations).

Critères d'évaluation de la qualité des prévisions

Avant de continuer, il nous faut définir la fonction de perte utilisée pour effectuer cette évaluation de la qualité de la prévision.

Ensembles d'observations $\mathcal{Y}$ et de prévisions $\mathcal{X}$. Les observations en chaque station sont situées dans l'intervalle $[0, 300] \cup \{\perp\}$, où le symbole $\perp$ désigne une absence d'observation correspondant à une station en panne et où la valeur 300 est prise pour fixer les idées : c'est une borne sur la concentration maximale d'ozone. De même, les prévisions individuelles en un jour donné et un site fixé sont supposées être dans l'intervalle $[0, 300]$. Les stations étant indexées par l'ensemble $\mathcal{N}$, on a donc les ensembles d'observations et de prévisions :

$$\mathcal{Y} = \left([0, 300] \cup \{\perp\}\right)^{\mathcal{N}} \qquad \text{et} \qquad \mathcal{X} = [0, 300]^{\mathcal{N}} .$$

Pertes instantanées. La fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ est égale, pour tout couple $\boldsymbol{y} = (y_s)_{s\in\mathcal{N}}$ et $\boldsymbol{x} = (x_s)_{s\in\mathcal{N}}$ d'éléments de $\mathcal{Y}$ et $\mathcal{X}$, à

$$\ell(\boldsymbol{x}, \boldsymbol{y}) = \sum_{s : y_s \neq \perp} (x_s - y_s)^2 .$$

Cette fonction vérifie bien l'hypothèse 1.6 et on peut donc dans la suite instancier les stratégies $\mathcal{E}_\eta^{\text{grad}}$ du paragraphe 1.2.3.

La perte au jour t d'une stratégie d'agrégation proposant le vecteur de mélange (convexe ou linéaire) $\boldsymbol{v}_t = (v_{1,t}, \ldots, v_{N,t})$ des prévisions des experts est alors égale, avec les notations précédentes, à une quantité que l'on note, pour simplifier, par $\ell_t(\boldsymbol{v}_t)$:

$$\ell_t(\boldsymbol{v}_t) = \sum_{s\in\mathcal{N}_t} \left(y_t^s - \sum_{j=1}^{N} v_{j,t} f_{j,t}^s \right)^2 . \tag{3.6}$$

On cache donc dans la notation ℓ_t à la fois les prévisions des experts $f_{j,t}^s$ et les observations y_t^s et on n'explicite que la dépendance en le vecteur de mélange $\boldsymbol{v}_t$ uniforme en l'espace proposé par la stratégie de prévision.

Evaluation globale : erreur quadratique moyenne. Dans cette partie, la période de prévision étant courte (126 jours), l'erreur quadratique moyenne (EQM) des experts et des stratégies de référence est calculée non pas sur toute la période de prévision mais

uniquement sur ses 96 derniers jours ; cela laisse 30 jours aux différentes stratégies comme période d'apprentissage sans évaluation. On note $\{t_0, \dots, T\} = \{31, \dots, 126\}$ les indices des jours où l'évaluation aura donc lieu. Formellement, étant donné les vecteurs de mélange (linéaires ou convexes) $\boldsymbol{v}_{t_0}, \dots, \boldsymbol{v}_T$ choisis par une stratégie d'agrégation $\mathcal{S}$, son erreur quadratique moyenne est définie comme

$$\text{EQM}(\mathcal{S}) = \sqrt{\frac{1}{\sum_{t=t_0}^{T} |\mathcal{N}_t|} \sum_{t=t_0}^{T} \ell_t(\boldsymbol{v}_t)}$$

(où $|\mathcal{N}_t|$ désigne le cardinal de $\mathcal{N}_t$).

Garantir qu'une stratégie a un regret faible est équivalent à garantir que son erreur quadratique moyenne est proche, par exemple, de celle du meilleur expert ou de la meilleure combinaison convexe constante des prévisions des experts. Dans la suite, nous reporterons les résultats uniquement en termes d'EQM plutôt que de regret, puisque c'est cette première qui est le critère de performance utilisé en pratique et que le second apparaît surtout comme un outil.

Notations pour la suite. Pour toute suite $\boldsymbol{v}_1, \dots, \boldsymbol{v}_T \in \mathbb{R}^N$, on note $\boldsymbol{v}_1^T$ la stratégie qui, quelles que soient les observations et les prévisions des experts, prescrit un mélange selon $\boldsymbol{v}_t$ à l'échéance t ; son erreur quadratique moyenne est notée $\text{EQM}(\boldsymbol{v}_1^T)$. Lorsque tous les vecteurs $\boldsymbol{v}_t$ sont égaux à la valeur commune $\boldsymbol{v}$, on note plus simplement cette erreur $\text{EQM}(\boldsymbol{v})$. On rappelle par ailleurs que l'on avait désigné par δ_j la masse de Dirac en j, qui correspond à utiliser la prévision donnée par l'expert j.

3.4.2 Performances des experts considérés et de certaines stratégies d'agrégation de référence

Une autre illustration de la diversité de comportements des experts. Le diagramme en bâtons de la figure 3.2 montre les erreurs quadratiques moyennes des experts sur le jeu de données considéré ; elles sont comprises entre 22.43 et 35.79. On pourrait penser qu'un meilleur expert ou groupe d'experts se dégage nettement, mais cela n'est pas le cas. En effet, la carte de l'Europe fournie à la figure 3.2, coloriée en fonction de l'indice de l'expert ayant la plus faible erreur quadratique moyenne sur chaque zone de l'espace, indique qu'il n'y a pas d'expert qui serait uniformément en espace le meilleur et qu'on ne peut parler au mieux que de meilleur expert local ; on note également que de nombreux experts sont meilleur expert local pour une partie de l'espace au moins. Cela illustre que tous les experts sont utiles et apportent de l'information, et qu'en outre, leur comportement et leurs performances sont variables en espace (on expliquera plus bas pourquoi on peut affirmer qu'ils sont également variables en temps).

Performances de quelques stratégies de référence. Le tableau 3.1 reporte les performances des stratégies de référence indiquées au paragraphe 3.2.

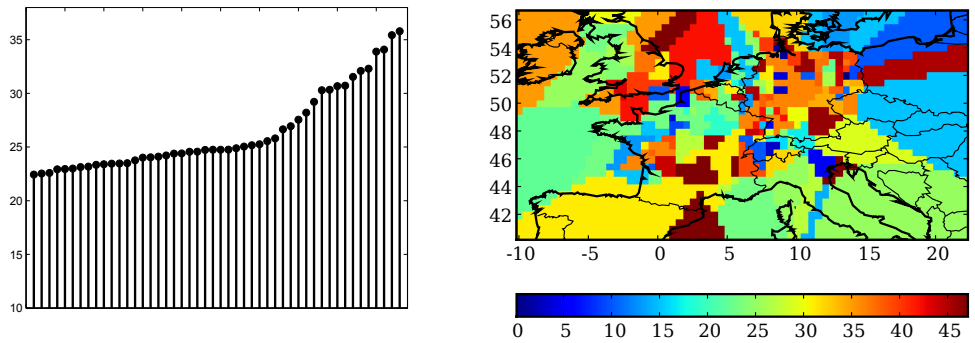

FIGURE 3.2. Représentation graphique des performances des experts : erreurs quadratiques moyennes des experts sur les données considérées, classées par ordre croissant (gauche) et coloration de la carte de toute l'Europe en fonction de l'indice du meilleur expert local (droite).

Nom de la stratégie de référence	Formule	Valeur
Moyenne uniforme	$\text{EQM}\big((1/48,\dots,1/48)\big)$	$= 24.41$
Meilleur expert	$\min_{j=1,\dots,48} \text{EQM}(\delta_j)$	$= 22.43$
Meilleure combinaison convexe	$\min_{\boldsymbol{q}\in\mathcal{P}} \text{EQM}(\boldsymbol{q})$	$= 21.45$
Meilleure combinaison linéaire	$\min_{\boldsymbol{u}\in\mathbb{R}^N} \text{EQM}(\boldsymbol{u})$	$= 19.24$
Stratégie presciente	$\min_{\boldsymbol{u}_1,\dots,\boldsymbol{u}_T\in\mathbb{R}^N} \text{EQM}\big(\boldsymbol{u}_1^T\big)$	$= 11.99$

TABLE 3.1. Performances de quelques stratégies de référence pour le jeu de données de prévision de pics d'ozone.

3.4.3 Performances de différentes stratégies de prévision

Dans [7] et [GMS08], nous avons étudié une vingtaine de stratégies : ici, nous ne reproduisons qu'un bref résumé des bonnes performances obtenues par trois stratégies (et leurs variantes).

Pondération par poids exponentiels des gradients des pertes $\mathcal{E}_\eta^{\text{grad}}$ et régression ridge $\mathcal{R}_\lambda$

On étudie dans ce paragraphe les stratégies $\mathcal{E}_\eta^{\text{grad}}$ du paragraphe 1.2.3 et $\mathcal{R}_\lambda$ du paragraphe 3.1.3. La première repose sur la considération de pseudo-pertes (1.11) associées aux gradients des pertes introduites en (3.6) et ne nécessite pas davantage de précisions. Pour la seconde en revanche, nous avons au préalable étendu la définition (3.4) et la borne de regret associée à un cas où non pas une mais $|\mathcal{N}_t|$ pertes quadratiques (celles

Valeur de η	5×10^{-7}	5×10^{-6}	2×10^{-5}	10^{-4}	Grille
EQM de $\mathcal{E}_\eta^{\text{grad}}$	22.89	21.70	<u>21.47</u>	22.10	21.77

Valeur de λ	0	100	10^4	10^6	Grille
EQM de $\mathcal{R}_\lambda$	20.79	<u>20.77</u>	21.13	21.80	20.81

Table 3.2. Performances des stratégies $\mathcal{E}_\eta^{\text{grad}}$ et $\mathcal{R}_\lambda$ pour différentes valeurs de leurs paramètres η et λ, ainsi que celles des méta-stratégies de calibration reposant sur elles. La plus faible EQM pour un paramètre constant est soulignée pour chacune des deux stratégies.

en chaque site actif) sont encourues.

Performances pour des paramètres fixés et performances opérationnelles. Les quatre premières colonnes du tableau 3.2 présentent les performances de ces deux stratégies pour diverses valeurs constantes des paramètres. La valeur 5×10^{-7} correspond approximativement à la valeur théorique optimale $\eta^\star$ précisée par le Théorème 1.7, mais elle n'est de loin pas la meilleure valeur en pratique.

C'est pourquoi, ainsi qu'expliqué au paragraphe 3.1.1, on recourt à une méthode de calibration par grille. Ici, on utilise une version simplifiée de cette calibration, où la grille est fixée et immuable ; celle pour la famille des stratégies $\mathcal{E}_\eta^{\text{grad}}$, respectivement, $\mathcal{R}_\lambda$, consiste en 11 points logarithmiquement uniformément répartis entre 10^{-8} et 10^{-4}, respectivement, entre 1 et 10^6. Les résultats obtenus par calibration sur cette grille sont précisés dans la dernière colonne du tableau 3.2. On voit notamment que l'adaptation a un coût assez réduit par rapport aux meilleurs choix rétrospectifs de paramètres.

Agréger, ce n'est pas suivre un seul expert. On conclut ce paragraphe en notant que les stratégies d'agrégation considérées ne se concentrent pas sur un seul expert et qu'au contraire, les poids attribués aux experts dans les vecteurs de mélange retenus peuvent changer rapidement et de manière significative au cours du temps. Ceci est illustré à la figure 3.3, où l'on a considéré les paramètres η et λ rétrospectivement optimaux, et est à attribuer au fait que les performances des experts changent au cours du temps (le paragraphe 3.4.2 avait déjà insisté sur le fait qu'elles variaient également en espace).

Variantes des deux familles de stratégies précédentes : fenêtrage et escompte

On applique ici les variantes présentées au paragraphe 3.1.2 aux deux familles de stratégies étudiées ci-dessus, en précisant au lecteur que bien sûr, ces variantes uniquement décrites dans le cas des stratégies de pondération par poids exponentiels s'étendent

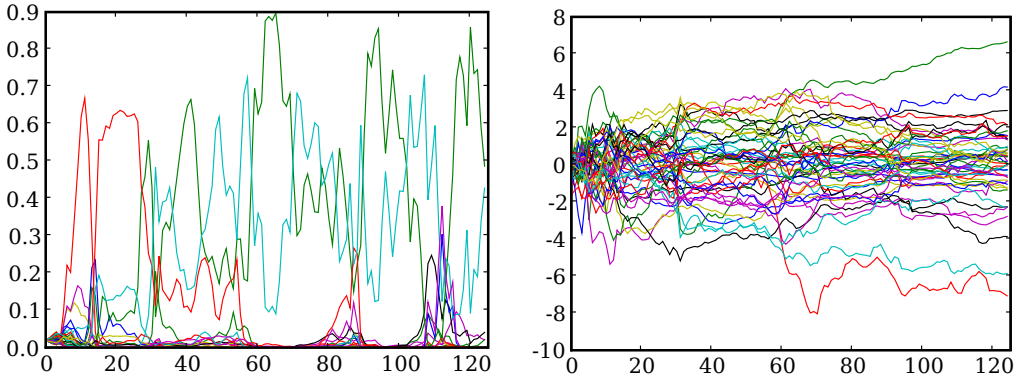

FIGURE 3.3. Représentation graphique des vecteurs de mélange convexes choisis par $\mathcal{E}^{\text{grad}}_{2\times 10^{-5}}$ (à gauche) et linéaires choisis par $\mathcal{R}_{100}$ (à droite), en fonction du temps.

Famille	Originelle	Fenêtrée	Escomptée
$\mathcal{E}^{\text{grad}}_{\eta}$	21.47	21.37	21.31
$\mathcal{R}_{\lambda}$	20.77	20.03	19.45

TABLE 3.3. EQM de différentes familles de stratégies, calibrées chacune avec le(s) meilleur(s) paramètre(s) rétrospectifs : versions originelles, fenêtrées et escomptées.

naturellement au cas des régressions séquentielles. Le tableau 3.3 reporte les résultats obtenus. On y entend par versions originelles celles qui avaient été décrites au tableau 3.2 ; on reporte la plus faible EQM qui y avait été obtenue pour un choix constant des paramètres. De même, pour les versions fenêtrées et escomptées, les valeurs reportées correspondent à une optimisation rétrospective des paramètres (pour laquelle on peut retrouver davantage de détails dans [10]).

Conclusion. On constate que tenir moins compte du passé par fenêtrage ou escompte améliore les performances, comme nous le suggéraient les praticiens. Cela étant, les résultats obtenus par la considération d'escomptes étant meilleurs que ceux par fenêtrage, on en conclut qu'il ne faut pas pour autant totalement oublier le passé lointain.

Régression ridge escomptée : propriétés de robustesse et de correction automatique du biais

Robustesse. On effectue dans [7, paragraphes 4.3.2 et 4.3.3] une étude de robustesse de la meilleure stratégie obtenue jusque-là, la régression ridge escomptée. On vérifie que l'excellente performance globale moyenne, sur l'ensemble des stations et des jours de prévision, ne vient pas au prix de quelques catastrophes localisées.

Expert	EQM originel	EQM après débiaisement
Meilleur	22.43	21.66
De référence	24.01	22.43
Pire	35.79	24.78

Table 3.4. Réductions d'EQM par application du pré-traitement de débiaisement consistant à lancer la régression ridge escomptée optimale du tableau 3.3 sur l'expert seul.

Correction automatique du biais. Par ailleurs, la régression ridge et plus encore, la régression ridge escomptée, peut être utilisée comme un pré-traitement de correction automatique du biais des experts. Cette faculté à corriger le biais était d'ailleurs la motivation à recourir à des vecteurs de mélange linéaires plutôt que convexes et elle devait compenser leur moindre interprétabilité. Formellement, on fixe un expert k et on propose, à chaque échéance t, les prévisions $b_t f^s_{k,t}$ en lieu et place des $f^s_{k,t}$, où b_t est le scalaire de débiaisement

$$b_t \in \operatorname*{arg\,min}_{b\in\mathbb{R}} \left\{ \lambda|b|^2 + \sum_{t'=1}^{t-1} (1+\beta_{t-t'}) \sum_{s\in\mathcal{N}_{t'}} \left(y^s_{t'} - b f^s_{j,t'}\right)^2 \right\} ;$$

évidemment, dans ce cas, b_t sera toujours positif et tendra à être d'autant plus proche de 1 que l'expert k a un biais originel faible. L'objectif est en effet, comme l'indique le Théorème 3.2, de faire presqu'aussi bien que le meilleur des méta-experts proposant $b f^s_{j,t}$ en chaque site et à chaque échéance, pour un paramètre scalaire positif b représentant un facteur multiplicatif constant de débiaisement.

Le tableau 3.4 illustre l'intérêt de ce pré-traitement sur trois experts parmi les 48 : le meilleur et le pire expert au sens de leur EQM, ainsi qu'un expert de référence formé par les valeurs les plus courantes des choix du paragraphe 3.4.1 (voir [MS06, paragraphe 2.2] pour plus de détails). Dans tous les cas, une réduction d'EQM est obtenue. Une idée que nous avons eue mais n'avons pas encore exploitée serait d'appliquer ce pré-traitement à tous les experts avant d'utiliser des stratégies de prévision par agrégation.

Stratégie de Lasso séquentiel escomptée

Afin de préparer des avancées futures qui consisteraient par exemple en la considération d'un grand nombre d'experts, nous avons voulu voir dans [GMS08] s'il était possible de réaliser agrégation et sélection d'experts simultanément, *id est*, d'agréger les prévisions uniquement d'un sous-ensemble d'experts de cardinal petit. Ce sous-ensemble a bien sûr vocation à changer au cours du temps. Pour cela, nous avons recouru à (une version escomptée de) la stratégie de Lasso séquentiel du paragraphe 3.1.3. Sa performance en termes d'EQM est de 19.31 (après optimisation rétrospective des paramètres, voir [10]) ;

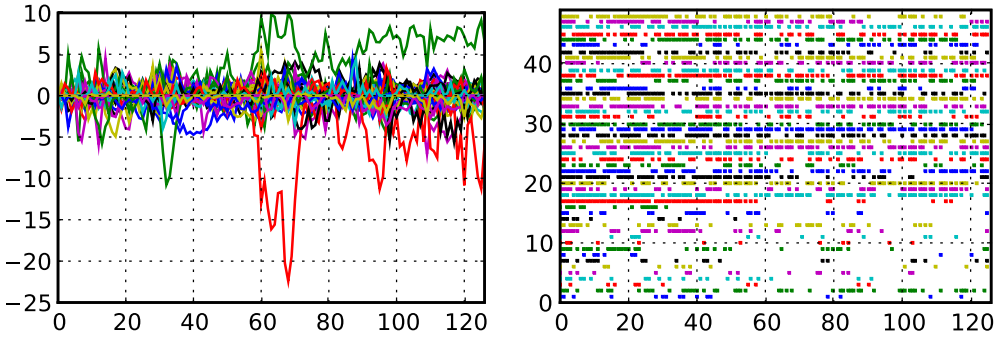

FIGURE 3.4. Représentation graphique des vecteurs de mélange linéaires choisis par la version escomptée optimale de la stratégie de Lasso séquentiel (à gauche) et des experts sélectionnés (à droite : un carré plein signifie que l'expert est absent du mélange).

elle est donc comprise entre l'EQM de l'oracle linéaire, qui est 19.24, et celle de la version escomptée de la régression ridge, qui est 19.45. La sélection–agrégation qu'effectue cette stratégie est décrite quant à elle à la figure 3.4. On note que typiquement, à chaque échéance une vingtaine de modèles est éliminée et que l'agrégation ne s'effectue plus que sur un sous-ensemble d'une trentaine d'experts.

3.5 Prévision de consommation électrique [13]

Dans cette partie, on s'intéresse à la prévision demi-horaire de la consommation électrique des clients d'EDF, le fournisseur historique français. Les résultats présentés sont tirés de l'article soumis [13] ainsi que du rapport technique correspondant [DGS09] ; tous deux étudient également la prévision horaire de la consommation des clients de la filiale slovaque d'EDF.

Utilisation d'experts spécialisés. La principale difficulté ici (mais qui est également une chance), c'est que l'on a affaire à des experts spécialisés à certains contextes et qui ne fournissent par conséquent que des prévisions par intermittence. Par exemple, certains experts peuvent être construits pour fournir de bonnes prévisions en hiver et sont inactifs en été ; il peut également y avoir des experts de jours travaillés ou des experts de fin de semaine. C'est une chance parce que ces experts spécialisés sont susceptibles d'être bien plus précis ; cela a constitué une difficulté à première vue parce qu'il a fallu étudier comment adapter les définitions et résultats du chapitre 1.

Plan de l'étude. On commence par expliquer brièvement comment traiter mathématiquement ce nouveau cadre, avant de présenter le jeu de données et les experts retenus, puis de décrire les performances obtenues par les stratégies de prévision par agrégation considérées.

3.5.1 Comment tirer parti d'experts spécialisés

Références bibliographiques. A notre connaissance, ce cadre a été peu considéré dans le domaine de la prévision de suites individuelles et nous ne pouvons citer comme références que les articles fondateurs [Blu97] et [FSSW97], l'un proposant le cadre, l'autre le formalisant ; de même que deux autres articles principalement consacrés à d'autres sujets mais mentionnant des résultats pour les experts spécialisés en passant, à savoir [BM07, paragraphes 6–8] et [CBL03, paragraphe 6.2]. Tous ont pour objet l'agrégation par vecteurs de mélange convexes ; il ne semble pas exister à ce jour de formulation et de résultats théoriques pour celle par vecteurs linéaires. Des tentatives en ce sens ont été considérées dans [DGS09] mais elles ne sont pas satisfaisantes en l'état et nécessitent au moins un sérieux travail d'approfondissement.

Formalisation mathématique. On reprend les notations du chapitre 1. L'ensemble des prévisions $\mathcal{X}$ est étendu pour contenir le point $\perp$, qui a la signification suivante. Lorsqu'à l'échéance t, l'expert $j \in \{1, \dots, N\}$ propose $f_{j,t} = \perp$, c'est que les conditions externes liées à sa spécialisation ne sont pas remplies et qu'il s'abstient de former une prévision. On dira qu'il est inactif. Par opposition, les experts proposant comme prévision un autre élément de $\mathcal{X}$ sont dits actifs. On suppose qu'à chaque échéance t, au moins un expert est actif et on notera E_t l'ensemble non vide de ces experts actifs. Comme indiqué plus haut, on se restreint au choix de vecteurs de mélange convexes. En particulier, une stratégie de prévision $\mathcal{S}$ choisit à l'échéance t un vecteur de mélange convexe $\boldsymbol{p}_t$ de support inclus dans E_t et forme la prévision

$$\widehat{y}_t = \sum_{j \in E_t} p_{j,t} f_{j,t} \,.$$

Critères d'évaluation. On considère ici encore la perte quadratique : on définit alors la perte cumulée et l'EQM d'une stratégie $\mathcal{S}$ sur les T premières échéances de la même manière que précédemment, c'est-à-dire selon

$$\widehat{L}_T(\mathcal{S}) = \sum_{t=1}^{T} \big(\widehat{y}_t - y_t\big)^2 \qquad \text{et} \qquad \mathrm{EQM}(\mathcal{S}) = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \big(\widehat{y}_t - y_t\big)^2} \,.$$

Il est à noter que dans cette partie, contrairement à la précédente, l'évaluation se fera bien sur toute la période de prévision, sans période d'entraînement, parce que cette période de prévision a une longueur T suffisamment grande.

Comparaison au meilleur expert ou à la meilleure combinaison convexe constante des experts

Les choses se compliquent au moment de définir les quantités correspondantes pour les experts et leurs combinaisons convexes constantes ; on reproduit ici la démarche proposée par [FSSW97].

Pour un expert fixé. La perte cumulée d'un expert j donné a peu de sens, mais son EQM est facile à définir :

$$\text{EQM}(j) = \sqrt{\frac{1}{\sum_{t=1}^{T} \mathbb{I}_{\{j \in E_t\}}} \sum_{t \leqslant T : j \in E_t} (f_{j,t} - y_t)^2} \,.$$

Il est également facile d'introduire une notion de regret, qui sera cette fois-ci fortement dépendante de l'expert j auquel on compare la stratégie $\mathcal{S}$ et c'est pourquoi on indexera ce regret par T, $\mathcal{S}$ mais aussi j ; en fait, pour que la comparaison entre j et $\mathcal{S}$ soit honnête, on ne l'effectue que sur les échéances où j était actif :

$$R_T(\mathcal{S}, j) = \sum_{t \leqslant T : j \in E_t} \left(\left(\widehat{y}_t - y_t\right)^2 - (f_{j,t} - y_t)^2 \right) .$$

Pour une combinaison convexe fixée. La dernière difficulté est maintenant d'étendre ces définitions au cas non pas d'un expert mais d'une combinaison convexe constante de ces experts, donnée par le vecteur de mélange convexe $\boldsymbol{q}$, de telle manière que lorsque $\boldsymbol{q} = \delta_j$, la masse de Dirac en j, on retrouve exactement les définitions précédentes. A cet effet, on introduit la renormalisation $\boldsymbol{q}^E$ de $\boldsymbol{q}$ à un sous-ensemble E de $\{1, \dots, N\}$ en définissant d'abord le poids de E sous $\boldsymbol{q}$,

$$\boldsymbol{q}(E) = \sum_{j \in E} q_j \,,$$

puis

$$\boldsymbol{q}^E = \begin{cases} (0, \dots, 0) & \text{lorsque } \boldsymbol{q}(E) = 0 \,; \\ \left(\dfrac{q_1 \mathbb{I}_{\{1 \in E\}}}{\boldsymbol{q}(E)}, \dots, \dfrac{q_N \mathbb{I}_{\{N \in E\}}}{\boldsymbol{q}(E)} \right) & \text{lorsque } \boldsymbol{q}(E) > 0. \end{cases}$$

On propose alors les extensions

$$\text{EQM}(\boldsymbol{q}) = \sqrt{\frac{1}{\sum_{t=1}^{T} \boldsymbol{q}(E_t)} \sum_{t=1}^{T} \left(\sum_{j \in E_t} q_j^{E_t} f_{j,t} - y_t \right)^2 \boldsymbol{q}(E_t)}$$

et

$$R_T(\mathcal{S}, \boldsymbol{q}) = \sum_{t=1}^{T} \left(\left(\widehat{y}_t - y_t\right)^2 - \left(\sum_{j \in E_t} q_j^{E_t} f_{j,t} - y_t \right)^2 \right) \boldsymbol{q}(E_t) \,.$$

Garanties théoriques de certaines stratégies d'agrégation. [13, paragraphe 2.3] effectue un survol de la littérature montrant qu'il est possible d'assurer que le regret face aux vecteurs de mélange convexes $\boldsymbol{q}$ soit uniformément borné par une quantité de l'ordre de $\sqrt{T}$, où l'uniformité porte sur $\boldsymbol{q}$ mais aussi sur les suites d'observations y_t et de prévisions des experts $f_{j,t}$. Les stratégies assurant cela sont obtenues par une extension des stratégies de pondération par poids exponentiels des gradients des pertes du paragraphe 1.2.3 ; elles reposent sur le même paramètre $\eta > 0$. Par souci de concision, nous ne détaillons pas davantage cela et nous contentons de les noter $\mathcal{W}_\eta^{\text{grad}}$.

Comparaison à des experts composés

Les travaux [Gou08a, Gou08b] ont montré l'intérêt, en prévision de consommation électrique, de stratégies d'agrégation ayant pour objet d'obtenir des performances presqu'aussi bonnes non pas que celles d'un expert donné, mais que celles d'une suite d'experts.

Formalisation (dans notre cadre). [HW98] a introduit la classe des experts composés. Pour un nombre d'échéances T, cette classe peut être identifiée, s'agissant d'experts par ailleurs spécialisés, à $\mathcal{C}'_T = E_1 \times \ldots \times E_T$. Pour tout élément $j_1^T = (j_1, \ldots, j_T)$ de $\mathcal{C}'_T$, on note

$$L_T(j_1^T) = \sum_{t=1}^{T} \ell(f_{j_t,t}, y_t)$$

la perte cumulée de l'expert composé lui correspondant. Bien évidemment, aucune stratégie ne peut être compétitive face au meilleur expert composé : cela reviendrait essentiellement à connaître l'indice du meilleur expert pour la prochaine échéance. Ainsi, il faut contraindre un peu les experts composés, en requérant par exemple que leur nombre de ruptures ne soit pas trop grand, où ce nombre est défini, pour un expert composé j_1^T, par

$$s(j_1^T) = \sum_{t=2}^{T} \mathbb{I}_{\{j_{t-1} \neq j_t\}} \, .$$

On parle également de sauts (d'où la définition par une fonction notée s).

Regret face à des suites d'experts avec un nombre pas trop élevé de ruptures. On note $\mathcal{C}'_{T,m}$ le sous-ensemble d'experts composés de $\mathcal{C}'_T$ contenant au plus m ruptures. Bien évidemment, pour m petit, ces sous-ensembles peuvent être vides ici. Les notions de regret d'une stratégie $\mathcal{S}$ par rapport à un expert composé j_1^T ou d'EQM d'un tel expert composé sont claires, puisque tant la stratégie $\mathcal{S}$ que l'expert j_1^T forment une prévision à chaque tour.

Contrôle de ce regret dans le cas d'experts non spécialisés. [HW98] (mais on pourra également se référer à [CBL06, paragraphe 5.2]) présente une stratégie effectuant essentiellement une mise en œuvre efficace et séquentielle de la stratégie de pondération par poids exponentiels sur l'ensemble des experts composés, chacun de ces derniers étant affecté d'un poids initial non pas uniforme mais fonction du nombre de ses sauts. Cette stratégie a été originellement appelée "fixed share" (stratégie par redistribution des poids, dans ce texte) et repose sur deux paramètres $\eta > 0$, vitesse d'apprentissage, et $\alpha \in [0, 1]$, vitesse de mélange. Lorsque ces deux paramètres sont bien calibrés (notamment en fonction de T et de m), son regret face à l'ensemble des experts admettant au plus m ruptures est uniformément borné par une quantité de l'ordre de $\sqrt{m\,T \ln N}$.

Extension au cas d'experts spécialisés. Là encore, il est facile de modifier cette stratégie afin qu'elle tienne compte de la spécialisation ; on peut également la faire travailler sur les (sous-)gradients des pertes plutôt que sur les pertes elles-mêmes. Les détails de l'adaptation sont omis et on renvoie le lecteur à [13, paragraphe 2.3]. On obtient ainsi deux familles de stratégies par redistribution des poids pour notre cadre, notées $\mathcal{G}_{\eta,\alpha}$ et $\mathcal{G}^{\text{grad}}_{\eta,\alpha}$ selon qu'elles utilisent les pertes ou les (sous-)gradients des pertes.

3.5.2 Description du jeu de données et des experts utilisés

On utilise un jeu de données couramment employé à EDF pour la calibration des modèles de prévision à court terme de la consommation électrique. Il est décrit en détails dans [DKO$^+$08] et [13] et nous ne résumons qu'à très grands traits ses caractéristiques.

Il est constitué d'une part, de données de consommation à pas demi-horaire, et d'autre part, de prévisions météorologiques (température et couverture nuageuse) effectuées sur l'ensemble du territoire français. Les données de consommation sont construites par EDF à partir des mesures procurées par une entité filiale responsable de la distribution d'électricité, RTE, tandis que les prévisions météorologiques sont fournies par Météo-France.

Ensemble d'apprentissage et ensemble de validation. Ce jeu de données est divisé en deux parties, la première couvrant la période entre le 1$^{\text{er}}$ septembre 2002 et le 31 août 2007 (l'ensemble d'apprentissage) et la seconde, celle entre le 1$^{\text{er}}$ septembre 2007 et le 31 août 2008 (l'ensemble de validation). Les experts construits ci-dessous le seront sur l'ensemble d'apprentissage, après quoi ils procureront des prévisions tout au long de la période correspond à l'ensemble de validation. En réalité, pour être tout à fait précis, nous excluons certains jours particuliers de l'ensemble de validation : des 366 jours que couvre sa période nous n'en conservons finalement que 320. Ces jours particuliers sont les jours fériés, ainsi que les jours situés immédiatement avant ou après eux ; les deux jours de changement d'heure ; et les vacances de Noël (du 21 décembre 2007 au 4 janvier 2008). Nous conservons en revanche les grandes vacances (et notamment, août 2008) parce que cette période étant suffisamment longue, il est possible de construire des experts dédiés à elle. Les caractéristiques des consommations y_t sur l'ensemble de validation sont précisées au tableau 3.5.

Unités. Dans cette partie également, nous omettons l'unité (GW, gigawatt) des observations et des prévisions de la consommation, de même que celle de leur EQM correspondante.

Construction de trois familles d'experts

Les trois familles d'experts que nous avons construites sont issues des trois grandes familles de modèles statistiques : paramétriques, semi-paramétriques et non-paramétriques ; le but recherché en faisant varier les méthodes de construction est d'obtenir des experts

Echéances	Toutes les 30 minutes
Nombre de jours D	320
Nombre d'échéances T	15 360
Nombre d'experts N	24 $(= 15 + 8 + 1)$
Médiane des y_t	56.33
Maximum B des y_t	92.76

Table 3.5. Quelques caractéristiques des consommations électriques y_t sur l'ensemble de validation.

aussi dissemblables que possible. Nous décrivons ci-dessous les traits principaux de leur création et adressons le lecteur à [DGS09, paragraphe 4.1] et [13] pour plus de détails.

Modèle paramétrique. Le modèle paramétrique utilisé pour engendrer le premier groupe d'experts est décrit dans [BDR05] et est mis en œuvre dans le logiciel de prévision d'EDF appelé «Eventail». C'est en faisant varier les différents paramètres nécessaires que nous avons défini 15 experts, que nous appellerons les experts Eventail.

Modèle semi-paramétrique. Le second groupe d'experts a été engendré par un modèle additif généralisé (abrégé en MAG dans la suite) ; ici encore, c'est en faisant varier les différents paramètres du modèle additif généralisé que nous avons créé 8 experts, les experts dits MAG.

Modèle non-paramétrique. Enfin, nous avons considéré un dernier expert, issu d'une modélisation non-paramétrique proposée par [ABCP10] et qui consiste à voir la consommation électrique au cours de la journée comme la réalisation d'une certaine courbe stochastique sous-jacente, dont on dispose chaque jour d'une discrétisation de pas demi-horaire. L'expert ainsi obtenu sera appelé l'expert non-paramétrique.

Illustration des taux d'activité. Les performances des experts présentés ci-dessus sont résumées à la figure 3.5. Le diagramme en bâtons montre leurs EQM, classées dans l'ordre croissant, tandis que le diagramme bi-dimensionnel relie ces EQM aux fréquences d'activité, *id est*, représente les couples

$$\left(\mathrm{EQM}(j),\ \frac{\sum_{t=1}^{T} \mathbb{I}_{\{j \in E_t\}}}{T} \right)$$

pour tous les experts j. On voit que trois experts Eventail sur les quinze sont actifs sur toute la période, dont celui correspondant au modèle de prévision opérationnelle.

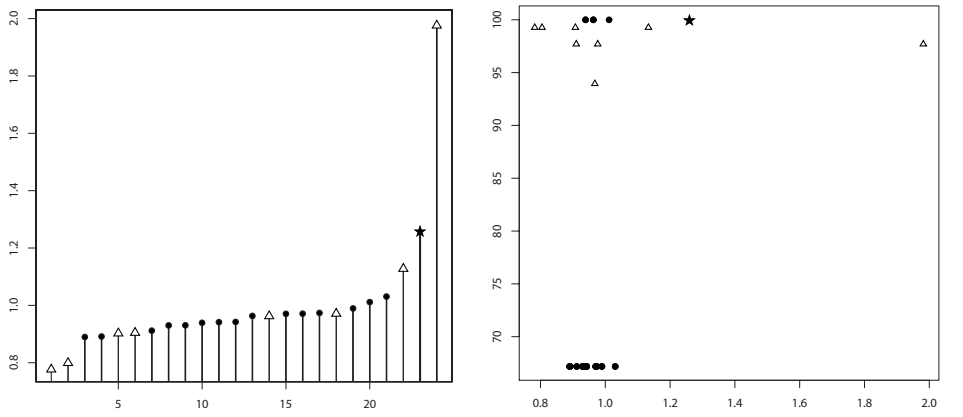

FIGURE 3.5. Représentations graphiques des performances des experts : EQM classées par ordre croissant (à gauche) et couples EQM–fréquence d'activité (à droite) ; la correspondance entre les experts et les symboles est la suivante : experts Eventail (•), experts MAG (△) et expert non-paramétrique (⋆).

Les douze autres experts Eventail sont inactifs durant l'été : en effet, leurs prévisions sont redondantes avec le modèle opérationnel car ils sont obtenus par variation d'un paramètre (le gradient de température) qui n'est lié qu'aux prévisions de la période hivernale. Les experts MAG sont actifs la plupart du temps, sauf à certaines périodes (par exemple, dans les semaines avec des jours fériés) où par expérience, les services d'EDF R&D savent qu'ils seront peu précis ; l'importance de ces périodes dépend par ailleurs des paramètres précis, c'est pour cela que tous ces experts MAG n'ont pas le même taux d'activité. Enfin, l'expert non-paramétrique est actif en permanence.

Ajout d'une contrainte opérationnelle

On requiert que les stratégies de prévision proposent chaque jour à midi des prévisions pour les 24 prochaines heures, c'est-à-dire, pour les 48 prochaines échéances demi-horaires. On suppose à cet effet que les prévisions des experts sont elles aussi disponibles pour toutes ces échéances futures. En revanche, on n'impose pas de contrainte d'égalité à une valeur commune pour les 48 vecteurs de mélanges convexes qu'une stratégie doit ainsi proposer. (En clair : on peut mélanger différemment les prévisions des experts selon le moment de la journée ; c'est même nécessaire lorsque certains d'entre eux deviennent actifs ou inactifs en cours de journée.)

3.5.3 Performances des stratégies de prévision

Performances de certaines stratégies de référence

Le tableau 3.6 révèle que les experts construits sont très bons, au vu des ordres de grandeur typiques des y_t précisés au tableau 3.5. Quelques stratégies qui y sont

Nom de la stratégie de référence	Formule	Valeur
Stratégie d'agrégation uniforme	$\textsc{eqm}(\mathcal{U})$	$= 0.724$
Combinaison convexe uniforme	$\textsc{eqm}\big((1/24,\dots,1/24)\big)$	$= 0.748$
Meilleur expert	$\min_{j=1,\dots,24} \textsc{eqm}(j)$	$= 0.782$
Meilleur combinaison convexe	$\min_{\boldsymbol{q}\in\mathcal{P}} \textsc{eqm}(\boldsymbol{q})$	$= 0.683$
Meilleur expert composé		
Au plus $m = 50$ ruptures	$\min_{j_1^T\in\mathcal{C}'_{T,50}} \textsc{eqm}\big(j_1^T\big)$	$= 0.534$
Au plus $m = 100$ ruptures	$\min_{j_1^T\in\mathcal{C}'_{T,100}} \textsc{eqm}\big(j_1^T\big)$	$= 0.474$
Stratégie presciente	$\min_{j_1^T\in E_1\times E_2\times\dots\times E_T} \textsc{eqm}\big(j_1^T\big)$	$= 0.223$

Table 3.6. Définition et performances de différentes stratégies de référence pour le jeu de données de consommation électrique.

introduites appellent des commentaires.

Agrégation uniforme et combinaison convexe uniforme. Dans le cadre des experts spécialisés, il y a une différence subtile entre l'utilisation de la combinaison convexe uniforme $\boldsymbol{q} = (1/24,\dots,1/24)$ et la stratégie d'agrégation uniforme $\mathcal{U}$. Cette dernière est en effet définie comme employant, à chaque échéance t, le vecteur de mélange convexe donné par la loi uniforme sur l'ensemble E_t des experts actifs, de sorte que

$$\textsc{eqm}(\mathcal{U}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\sum_{j\in E_t} f_{j,t}}{|E_t|} - y_t\right)^2}$$

$$\text{tandis que} \qquad \textsc{eqm}\big((1/24,\dots,1/24)\big) = \sqrt{\frac{1}{\sum_{t=1}^{T}|E_t|}\sum_{t=1}^{n}|E_t|\left(\frac{\sum_{j\in E_t} f_{j,t}}{|E_t|} - y_t\right)^2}.$$

Ainsi, pour l'évaluation des performances de la combinaison convexe uniforme, les pertes associées à des échéances pour lesquelles de nombreux experts sont actifs comptent davantage que celles pour lesquelles peu d'experts sont actifs, tandis que pour $\mathcal{U}$ toutes les pertes instantanées ont le même poids.

		Meilleur(s) paramètre(s) fixé(s)	Calibration sur la grille
EQM des	$\mathcal{W}_\eta^{\text{grad}}$	0.650	0.654
	$\mathcal{G}_{\eta,\alpha}$	0.632	0.644
	$\mathcal{G}_{\eta,\alpha}^{\text{grad}}$	0.598	0.599

TABLE 3.7. EQM de trois familles de stratégies de prévision par agrégation pour les données de consommation électrique française, calculées sur les grilles indiquées au paragraphe 3.5.3 : pour le meilleur paramètre fixé (colonne de gauche) et pour l'adaptation sur la grille (droite).

Valeurs de référence. Ici, la combinaison convexe uniforme a une plus grande EQM que la stratégie $\mathcal{U}$, ce qui indique que les experts ont tendance à être davantage actifs dans les situations de prévision difficile. C'est un avantage dont sauront tirer parti les stratégies d'agrégation même si, pour l'instant, cela se traduit par le fait que le meilleur expert a une EQM plus grande que celle de la stratégie naïve $\mathcal{U}$. On retient donc du tableau 3.6 qu'il serait bon que les performances de nos stratégies d'agrégation sophistiquées dépassent de loin celles de la stratégie $\mathcal{U}$ (EQM de 0.724). Les performances de la meilleure combinaison convexe constante sont déjà un peu meilleures (EQM de 0.683) mais les EQM obtenues pour les experts composés montrent qu'effectivement, un fort gain de performance est possible par rapport à $\mathcal{U}$.

La fin de cette partie va montrer que c'est également le cas pour la stratégie de pondération par poids exponentiels des gradients des pertes $\mathcal{W}_\eta^{\text{grad}}$, de même que pour celles par redistribution des poids $\mathcal{G}_{\eta,\alpha}$ et $\mathcal{G}_{\eta,\alpha}^{\text{grad}}$, auxquelles on a fait allusion à la fin du paragraphe 3.5.1.

Performances et robustesse des stratégies d'agrégation étudiées

Pour la tabulation des performances de la famille de stratégies $\mathcal{W}_\eta^{\text{grad}}$, nous avons recouru à une grille de 19 paramètres η donnée par des points logarithmiquement uniformément répartis entre 10^{-6} et 1. Pour les familles $\mathcal{G}_{\eta,\alpha}$ et $\mathcal{G}_{\eta,\alpha}^{\text{grad}}$, il s'est agi d'une grille finie de $\mathbb{R}_+ \times [0,1]$, contenant 22×6 points. Les performances obtenues sont résumées au tableau 3.7, où l'on ne reporte pour chaque famille que les EQM correspondant ou au meilleur choix constant d'un point des grilles ou à l'adaptation sur les grilles.

Commentaires. Toutes les familles de stratégies obtiennent des résultats satisfaisants voire très bons, au sens des valeurs de références indiquées après le tableau 3.6. Ici encore, on observe que les versions fondées sur les gradients sont plus efficaces en pratique que les versions initiales, ce qui était attendu au vu des résultats généraux du paragraphe 1.2.3. En fait, nous avons été agréablement surpris (et même un peu intrigués) par les performances de la famille $\mathcal{G}_{\eta,\alpha}^{\text{grad}}$, comme le soulignent les remarques

de robustesse suivantes.

Etude de robustesse. Cette étude, que nous décrivons ici pour les données de consommation électrique alors que nous l'avions omise au paragraphe 3.4.3 pour celles de pics d'ozone, consiste à comparer les performances des stratégies par agrégation à celles du meilleur expert ou de la meilleure combinaison convexe constante des experts, non pas de manière globable comme c'est le cas pour le critère d'EQM mais de manière plus locale. A cet effet, nous avons d'une part découpé le jeu de données en les 48 sous-jeux correspondant à chaque demi-heure ; d'autre part, pour chacune de ces demi-heures, nous reportons non seulement l'EQM encourue mais aussi une idée de la dispersion des valeurs absolues des résidus de prévision.

Ces dernières sont définies comme $|\widehat{y}_t - y_t|$, où y_t mesure la consommation réelle à l'échéance t et $\widehat{y}_t$ la prévision qui en avait été faite. L'étude des quantiles des résidus permet de déterminer si les bonnes performances globales des stratégies de prévision par agrégation viennent ou non au prix de quelques ratés spectaculaires ; en particulier, ce sont les queues de distributions empiriques (quantiles à 75 % ou à 90 %) qui nous intéressent le plus. La figure 3.6 étudie les méta-stratégies de calibration formées respectivement à partir des familles $\mathcal{W}_{\eta}^{\mathrm{grad}}$ et $\mathcal{G}_{\eta,\alpha}^{\mathrm{grad}}$. Les performances de la première collent exactement à celles de la meilleure combinaison convexe constante des experts ou les améliorent un peu, tant au niveau des EQM que des quantiles demi-horaires. Le comportement de la seconde méta-stratégie peut intriguer : les performances sont nettement améliorées dans la période entre 12 heures et 21 heures mais peut-être un peu dégradées entre 6 heures et 12 heures par rapport à celles de la meilleure combinaison convexe constante – la période d'amélioration compensant de loin celle de légère dégradation. C'est pour nous une question ouverte que de mieux comprendre ce comportement et de tirer mieux parti des excellentes performances sur la fenêtre située juste après la mise à jour des poids effectuée à midi.

3.6 Conclusions et perspectives

3.6.1 Conclusions à l'attention des praticiens

Dans ce chapitre portant sur les (bonnes) performances pratiques de la prévision séquentielle par agrégation de prévisions d'experts, on a tout d'abord décrit un cadre méthodologique et des stratégies de prévision générales. On a ensuite montré que ces stratégies pouvaient être appliquées avec succès, au prix d'adaptations mineures, à deux cadres applicatifs : la prévision de pics journaliers d'ozone et la prévision de consommation électrique à pas demi-horaire. Ce faisant, on encourage le lecteur à employer les stratégies décrites ici dans tout cadre de prévision séquentielle où il disposerait d'un ensemble d'experts dont il ne peut dire à l'avance qui sera le meilleur. En particulier, ces experts peuvent être donnés par des méthodes issues de modélisations stochastiques nécessitant le réglage de différents paramètres : une alternative au réglage est formée par la considération de plusieurs instances de la méthode correspondant à

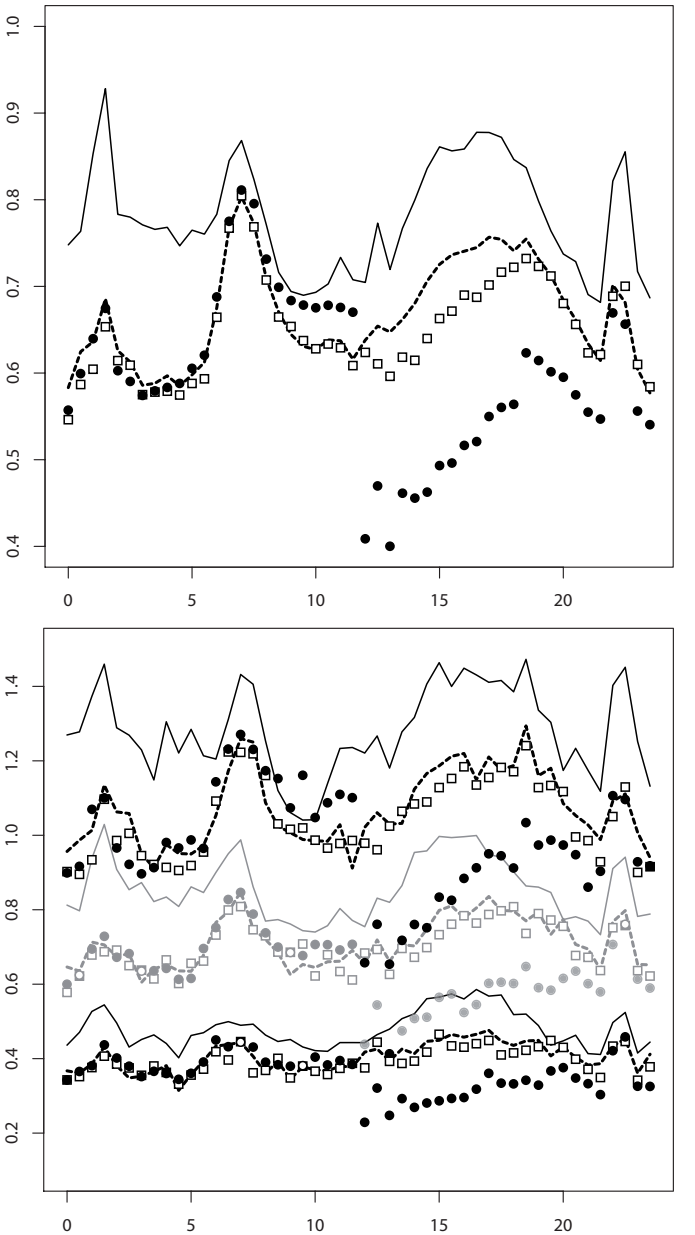

FIGURE 3.6. Mesures des performances demi-horaires du meilleur expert global (trait plein) et de la meilleure combinaison convexe globale (trait en pointillés), ainsi que des méta-stratégies de calibration fondées sur les familles $\mathcal{W}_{\eta}^{\mathrm{grad}}$ (symbole : •) et $\mathcal{G}_{\eta,\alpha}^{\mathrm{grad}}$ (symbole : □) ; EQM (en haut) et quantiles à 50 % (noir), 75 % (gris) et 90 % (noir) des valeurs absolues des résidus (en bas). Abscisses : demi-heures de la journées ; ordonnées : EQM (en haut) et valeurs absolues des résidus (en bas).

différents jeux de paramètres suffisamment éloignés. Ou encore, on peut toujours ajouter des experts supplémentaires correspondant à des prédicteurs sans garantie théorique formelle mais dont l'intuition soutient qu'ils devraient bien se comporter. Les deux études empiriques ont démontré une variante de l'adage bien connu "Garbage in, garbage out" : lorsqu'il existe quelques bons experts, les stratégies d'agrégation auront elles aussi de bonnes performances. En particulier, il n'est pas besoin de ne construire que de bons experts, mais il faut pouvoir assurer qu'un petit nombre eux le sera (sans avoir besoin pour autant de savoir à l'avance lesquels).

3.6.2 Perspectives de recherche

Au niveau méthodologique

Nous avons noté, dans la partie 3.1, que pour deux procédures aux bonnes performances pratiques les garanties théoriques n'étaient pour l'instant pas encore établies : il s'agit de la méthode de méta-calibration du paragraphe 3.1.1 et de la stratégie de Lasso séquentiel du paragraphe 3.1.3.

Pour la prévision des pics d'ozone

Les projets à court terme sont d'évaluer l'impact de pré-traitement de débiaisement automatique sur les experts et d'étudier les performances lorsque la période de prévision est plus longue, de l'ordre d'un an, avec ou sans augmentation du nombre d'experts. Des résultats préliminaires nous ont montré que, comme attendu, les gains de performances par rapport au meilleur expert ou à la meilleure combinaison convexe constante des experts sont encore plus importants dans ce cas.

A moyen terme, il faudra s'intéresser à la prévision des dépassements de seuils réglementaires. Nous avons lancé là aussi quelques études préliminaires mais n'avons pu faire mieux pour l'instant que la procédure qui consiste à comparer les valeurs prévues par des stratégies d'agrégation aux seuils – ce qui est surprenant malgré tout, s'agissant d'un problème de classification, généralement censé être plus facile (parce que plus direct) qu'un problème de prévision pure.

On mentionne également un travail récent de Vivien Mallet sur les liens entre assimilation de données et agrégation d'experts [Mal10].

Pour la prévision de consommation électrique

Il s'agit à court terme de mieux comprendre pourquoi les stratégies de prévision par redistribution des poids sont si précises à horizon très court mais obtiennent des performances un peu plus décevantes juste avant la remise à jour des poids. Il faudra également tirer encore plus parti de la spécialisation en augmentant l'ensemble des experts et en affinant leur construction. Enfin, nous aimerions, sur un plan plus méthodologique, pouvoir étendre les stratégies de régressions linéaires séquentielles régularisées du paragraphe 3.1.3 au cas des experts spécialisés. Cela semble un problème

assez redoutable, absolument pas traité dans la littérature, et divers essais proposés dans [DGS09] se sont soldés par de retentissants échecs en termes de performances pratiques : on attend d'une telle extension qu'elle améliore grandement les performances ; or nous y avions observé des dégradations très nettes.

Par ailleurs, il serait intéressant, ici comme en prévision de pics d'ozone, de pouvoir fournir une quantification des incertitudes sur la prévision agrégée. Celle-ci pourrait s'appuyer soit seulement sur la dispersion des prévisions des experts (on s'attend à une incertitude d'autant plus forte que cette dispersion est grande), soit également sur l'indication par les experts eux-mêmes d'une incertitude sur leurs prévisions.

Autres jeux de données et création d'un logiciel

L'objectif, à moyen terme, est d'étudier d'autres jeux de données, afin d'encore mieux cerner la méthodologie générale et de gagner en expérience ; on peut penser, en économie, à la prévision de taux de change (en prenant pour experts les prévisions fournies par différents analystes monétaires) ou, en hydrogéologie, à la prévision des hauteurs de rivières, l'étape préliminaire consistant, dans les deux cas, à identifier un partenaire fiable et compétent dans le domaine applicatif visé.

Cette diversification a un but avoué : parvenir à un logiciel intégrant les stratégies de prévisions décrites dans ce chapitre, sachant les calibrer efficacement, traitant automatiquement les cas de données manquantes (dans les observations et/ou les prévisions des experts), le tout, avec d'excellentes performances pratiques et une grande facilité d'utilisation. Ce but est plutôt énoncé avec un horizon de quelques années et requerra sans doute l'aide d'un petit groupe de chercheurs et d'ingénieurs.

CHAPITRE 4

Bandits stochastiques à continuum de bras et autres travaux

INTRODUCTION. Ce dernier chapitre regroupe divers travaux qui ne sont pas liés à la prévision séquentielle de suites arbitraires. Deux d'entre eux illustrent un domaine de recherche dans lequel je viens de me lancer récemment, les bandits stochastiques à continuum de bras, et dans lequel je compte m'investir dans les années à venir.

Table des matières

4.1 Apprentissage de bandits stochastiques à continuum de bras [8, 11]

On commence par décrire un modèle général dans lequel les bras sont indexés par un ensemble $\mathcal{X}$ arbitraire (fini ou infini) et qui sont chacun associés à une probabilité. On précise ensuite les résultats nouveaux obtenus, essentiellement pour le cas où $\mathcal{X}$ est un espace métrique (en passant ainsi sous le boisseau la première moitié de [8], consacrée au cas où $\mathcal{X}$ est fini).

Cadre stochastique. Comme on le notera ci-dessous dans le passage en revue des travaux antérieurs, il existe une version du problème de bandits à nombre fini (ou même dénombrable) de bras pour le problème de prévision séquentielle randomisée du paragraphe 1.1.3, qui utilise des techniques similaires à celles du paragraphe 1.3.1. Mais, une fois n'est pas coutume, ce qui suit est dévolu à un cadre stochastique plus classique.

4.1.1 Description du modèle mathématique (cas de paiements bornés)

Un statisticien dispose de bras indexés par un ensemble $\mathcal{X}$ et joue contre un environnement stochastique E selon le déroulement décrit à la figure 4.1. Il reçoit en particulier à chaque tour un paiement borné, à valeurs dans un intervalle connu ; pour fixer les idées, on suppose que cet intervalle est $[0,1]$ et on note $\Delta([0,1])$ l'ensemble des lois sur $[0,1]$.

Paramètres : un ensemble de bras $\mathcal{X}$ connu (fini ou infini, muni d'une topologie), un environnement $E : \mathcal{X} \to \Delta\big([0,1]\big)$ inconnu

A chaque échéance $t = 1, 2, \ldots$,

1. Le statisticien choisit une probabilité $\nu_t \in \Delta(\mathcal{X})$ et tire un bras $I_t \in \mathcal{X}$ selon ν_t ;
2. L'environnement tire le paiement Y_t du statisticien indépendamment au hasard selon la loi $E(I_t)$ associée au bras choisi par ce dernier ;
3. Le statisticien n'accède qu'à Y_t, qui forme la seule information procurée par cette échéance et dont il mémorise la valeur pour la suite.

FIGURE 4.1. Le déroulement du problème de bandits à ensemble de bras indexé par $\mathcal{X}$.

Notion d'environnement stochastique. Un environnement stochastique E est défini comme une application $\mathcal{X} \to \Delta\big([0,1]\big)$. Il associe à chaque élément $x \in \mathcal{X}$ une loi $E(x)$ sur $[0,1]$. Lorsque le statisticien choisit le bras $I_t \in \mathcal{X}$ à l'échéance t, l'environnement tire un paiement Y_t indépendamment au hasard selon la loi $E(I_t)$. On note $\mu_E : \mathcal{X} \to [0,1]$ la fonction précisant les espérances des paiements : pour tout $x \in \mathcal{X}$, l'espérance de la loi $E(x)$ est notée $\mu_E(x)$.

Notion de stratégie du statisticien. Le statisticien ne connaît pas l'environnement E. Les seules informations dont il dispose sont les paiements associés aux bras choisis dans le passé. Ainsi, à l'échéance $t \geqslant 2$, il détermine le bras I_t à tirer en fonction des bras $I_1, \ldots, I_{t-1}$, des paiements $Y_1, \ldots, Y_{t-1}$ qui leur ont été associés par l'environnement, et, éventuellement, d'une randomisation auxiliaire. A cet effet, on suppose que $\mathcal{X}$ est un espace topologique et on le munit de la tribu des boréliens ; on note alors $\Delta(\mathcal{X})$ l'ensemble des lois de probabilité sur $\mathcal{X}$.

Ainsi, une stratégie Ψ est une collection d'applications mesurables Ψ_t, avec $t \geqslant 2$, complétée par une loi initiale $\Psi_1 \in \Delta(\mathcal{X})$ fixe. Pour $t \geqslant 2$, l'application Ψ_t est définie sur $\mathcal{X}^{t-1} \times [0,1]^{t-1}$ et prend ses valeurs dans $\Delta(\mathcal{X})$: avec les notations de la figure 4.1, le statisticien choisit loi

$$\nu_t = \Psi_t\Big(I_1, \ldots, I_{t-1}, Y_1, \ldots, Y_{t-1}\Big)$$

sur $\mathcal{X}$ et tire à l'échéance t un bras I_t selon ν_t.

Randomisations auxiliaires. Le statisticien et l'environnement utilisent tous deux une suite de randomisations auxiliaires. Les probabilités $\mathbb{P}$ et espérances $\mathbb{E}$ seront relatives à ces randomisations uniquement (elles ne dépendent ni de E ni de Ψ).

Objectif et difficultés pour l'atteindre

L'objectif du statisticien est de rendre son paiement cumulé $Y_1 + \ldots + Y_T$ le plus grand possible.

Dilemme entre exploration et exploitation. A cet effet et à cause du retour sur action aléatoire, il doit arbitrer entre l'exploration des performances des différents bras (ce qui nécessite de les jouer chacun un nombre significatif de fois, afin de pouvoir en estimer la loi du paiement) et l'exploitation des informations ainsi collectées : il s'agit qu'il puisse activer plus souvent les bras correspondant aux meilleurs paiements. On parle de dilemme (ou d'équilibre) entre exploration et exploitation.

Notion de regret et objectif de minimisation de ce regret

Ici encore, assurer qu'un certain regret est faible garantira un paiement cumulé important. En fait, dans le cadre stochastique que l'on considère, tous les résultats de la littérature sont formulés en termes d'espérances des paiements cumulés, de sorte que le regret lui-même est une quantité déterministe, correspondant à une espérance.

Formellement, le regret d'une stratégie Ψ contre un environnement E est égal à

$$R_T(\Psi, E) = T\mu_E^\star - \sum_{t=1}^{T} Y_t \qquad \text{où} \quad \mu_E^\star = \sup_{x\in\mathcal{X}} \mu_E(x)\,;$$

et on étudiera généralement les quantités $\mathbb{E}\big[R_T(\Psi, E)\big]$. Minimiser le regret revient bien à maximiser l'espérance du paiement cumulé. On explique maintenant pourquoi on considère cette définition du regret plutôt que d'autres qui auraient été possibles.

Raison d'être du terme $T\mu_E^\star$. On note qu'en prenant l'espérance d'espérances conditionnelles (par rapport aux I_t), il vient

$$\mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} \mu_E(I_t)\right] \geqslant T \sup_{x\in\mathcal{X}} \mu_E(x) = T\mu_E^\star\,.$$

Ainsi, on a toujours $\mathbb{E}\big[R_T(\Psi, E)\big] \geqslant 0$, ce qui est une propriété agréable et qui admet une interprétation claire. Par ailleurs, grâce aux remarques précédentes, on définit alors le pseudo-regret d'une stratégie Ψ contre un environnement E comme la quantité non observable

$$R'_T(\Psi, E) = \sum_{t=1}^{T} (\mu_E^\star - \mu_E(I_t))\,;$$

on a bien l'égalité $\mathbb{E}\big[R_T(\Psi, E)\big] = \mathbb{E}\big[R'_T(\Psi, E)\big]$. Or, dans les démonstrations de bornes sur l'espérance du regret, il est souvent plus facile de considérer ce pseudo-regret.

Pourquoi on ne remplace par $T\mu_E^\star$ par un supremum de processus empiriques. Dans ce cadre, on n'impose pas qu'à chaque tour, une réalisation de chacune des lois $E(x)$ soit tirée, pour tout $x \in \mathcal{X}$; seule une réalisation associée au bras choisi est déterminée. L'ensemble $\mathcal{X}$ étant arbitraire et parfois non dénombrable, c'est une manière assez sage de procéder afin d'éviter des problèmes de mesurabilité. Cependant, cela empêche de remplacer chacun des termes $T\mu_E(x)$ par une somme empirique et par conséquent, de

remplacer $T\mu_E^\star$ par un supremum de processus empiriques. La considération de $T\mu_E^\star$ se fait, en résumé, faute de mieux.

Reformulation de l'objectif en termes de regret. On suppose que le statisticien a connaissance d'un modèle sur l'environnement E contre lequel il joue : il sait que E appartient à une certaine famille (éventuellement non paramétrique) $\mathcal{F}$ d'applications $\mathcal{X} \to \Delta([0,1])$. Comme aux chapitres précédents, on requiert que le regret rapporté au nombre de tours soit asymptotiquement négatif ou nul : nul ici, donc, vu la propriété de positivité.

Ainsi, on définit qu'une stratégie Ψ minimise son regret par rapport à une famille $\mathcal{F}$ si

$$\forall\, E \in \mathcal{F}, \qquad \lim_{T\to\infty} \frac{\mathbb{E}\big[R_T(\Psi, E)\big]}{T} = 0\,.$$

Dans certains cas, les convergences ci-dessus pourront être garanties de manière uniforme sur $\mathcal{F}$.

Passage en revue (très bref) des travaux antérieurs

La première mention du problème remonte à Robbins [Rob52]. On distinguera deux grandes catégories pour les références aux travaux antérieurs : celles qui portent sur le cas où $\mathcal{X}$ est un ensemble fini ou dénombrable (avec deux sous-catégories associées, selon que les paiements sont stochastiques ou formés par des suites arbitraires) et celles qui traitent le cas d'un nombre infini non dénombrable de bras.

Bandits à nombre fini (ou dénombrable) de bras. Dans le cas fini on note généralement $|\mathcal{X}| = K$. Les résultats essentiels pour le cadre stochastique présenté ci-dessus sont que le regret peut être minimisé face à tous les environnements E, c'est-à-dire, face à toutes les familles de K lois sur $[0,1]$, avec les vitesses suivantes de convergence vers 0 : une constante dépendant de E fois $(\ln T)/T$ pour les stratégies proposées dans [LR85, BK96, ACBF02, AB09, HT10] et des vitesses uniformes de l'ordre de $\sqrt{K(\ln T)/T}$ pour [ACBF02], $\sqrt{K(\ln K)/T}$ pour [ACBFS02], et $\sqrt{K/T}$ pour [AB09] ; cette dernière vitesse uniforme étant la vitesse uniforme optimale ainsi qu'il découle de la borne inférieure sur le regret de toute stratégie prouvée par [ACBFS02]. En jouant par blocs, on étend les résultats de convergence précédents au cas d'un nombre dénombrable de bras (sans plus pouvoir garantir aucune convergence uniforme, cependant).

On peut également définir une version du problème pour des suites arbitraires (selon les grandes lignes du paragraphe 1.1.3) ; [ACBFS02, AB09] s'intéressent à ce cadre plus difficile.

Bandits à nombre non dénombrable de bras. Ce cas a été considéré en premier lieu par [Agr95, Kle04] et approfondi par [Cop09, AOS07, KSU08]. Les stratégies exhibées ne minimisent le regret que sous des hypothèses topologiques sur $\mathcal{X}$ et face à des classes d'environnements suffisamment réguliers, les conditions de régularité portant en fait sur μ_E.

Par exemple, [KSU08] suppose que $\mathcal{X}$ est un espace métrique et construit en particulier des stratégies minimisant le regret face à la classe des environnements E admettant une fonction d'espérances des paiements μ_E qui soit L–lipschitzienne, pour une constante de Lipschitz L plus petite qu'une borne fixée L_0 et connue par les dites stratégies ; on s'intéresse donc ici à la régularité globale des fonctions μ_E.

Les hypothèses de [AOS07] portent quant à elles uniquement sur le comportement de μ_E autour de ses maxima globaux.

4.1.2 Caractérisation des cas où la minimisation du regret est possible [8]

La seconde partie de [8] s'intéresse aux deux familles d'environnements suivantes :

$$\mathcal{F}_{\text{tous}} = \Delta([0,1])^{\mathcal{X}} \qquad \text{et} \qquad \mathcal{F}_{\text{cont}} = \mathcal{C}\Big(\Delta([0,1])^{\mathcal{X}}\Big),$$

qui sont respectivement l'ensemble de tous les environnements possibles et l'ensemble des environnements E admettant une fonction d'espérances des paiements μ_E continue ; elle caractérise l'existence de stratégies minimisant le regret par rapport à ces deux familles.

Théorème 4.1. *Lorsque $\mathcal{X}$ est un espace métrique, le regret peut être minimisé par rapport à la famille $\mathcal{F}_{\text{cont}}$ si et seulement si $\mathcal{X}$ est séparable.*

Eléments de preuve. La preuve n'est qu'une formalisation des idées naturelles suivantes, relatives à la possibilité ou à l'impossibilité d'une exploration uniforme de $\mathcal{X}$.

D'une part, lorsque $\mathcal{X}$ est séparable, tout sous-espace dénombrable dense fixé représente bien $\mathcal{X}$ face à tous les environnements E avec une fonction d'espérances des paiements μ_E continue. Or, on sait que dans le cas d'un nombre dénombrable de bras, on peut minimiser le regret face à tout environnement. Pour ce faire, on peut procéder en phases successives et répétées, d'une part d'exploration selon une loi chargeant tous les points du sous-ensemble dénombrable dense, et d'autre part, d'exploitation des résultats de l'exploration.

Réciproquement, lorsqu'un espace métrique n'est pas séparable, il contient une infinité non dénombrable de boules disjointes $\mathcal{B}(a,\rho)$ de rayon fixé $\rho > 0$, indexées par $a \in A$; à chacune de ces boules, on associe un environnement E_a de fonction μ_{E_a} admettant un maximum égal à 1 et de support inclus dans $\mathcal{B}(a,\rho)$. Or, une loi de probabilité charge au plus un nombre dénombrable d'ouverts disjoints. Ainsi, une stratégie fixée obtiendra des paiements nuls à chaque échéance contre tous les environnements E_a, lorsque $a \in A$, sauf pour au plus un nombre dénombrable d'entre eux, pour lesquels elle aura eu une probabilité strictement positive d'explorer leur support. Son regret sera alors égal, contre la plupart des environnements, au nombre d'échéances et ne sera pas sous-linéaire.

Corollaire. On adopte une démarche bourbakiste et on déduit le résultat suivant du Théorème 4.1.

Corollaire 4.2. Soit $\mathcal{X}$ un ensemble quelconque. Le regret peut être minimisé par rapport à la famille $\mathcal{F}_{\text{tous}}$ de tous les environnements si et seulement si $\mathcal{X}$ est dénombrable.

C'est évidemment la nécessité de la dénombrabilité de $\mathcal{X}$ pour la minimisation du regret qui est d'intérêt ici (son caractère suffisant ayant déjà été noté plus haut). Pour établir ce corollaire à partir du Théorème 4.1, il suffit de munir $\mathcal{X}$ de sa topologie discrète (qui correspond à la distance de Hamming) ; dans ce cas, toute application $\mathcal{X} \to [0,1]$ est continue, de sorte que $\mathcal{F}_{\text{cont}} = \mathcal{F}_{\text{tous}}$.

Conclusion. Des hypothèses minimales sur la topologie de $\mathcal{X}$ et la régularité des fonctions d'espérances des paiements μ_E sont nécessaires afin de garantir l'existence de stratégies minimisant le regret. Dans la suite, afin de pouvoir exhiber des stratégies simples et efficaces, nous aurons besoin de renforcer ces hypothèses.

4.1.3 Obtention de vitesses explicites par une stratégie efficace et naturelle [11]

On introduit une stratégie notée HOO (pour "hierarchical optimistic optimization") et qui repose sur trois paramètres.

Paramètres de HOO. Ce sont deux nombres réels $\nu_1 > 0$ et $\rho \in]0,1[$, ainsi qu'un arbre de recouvrements $\mathcal{T} = (\mathcal{T}_{h,i})$, c'est-à-dire une collection de sous-ensembles de $\mathcal{X}$, non nécessairement disjoints, indexés par $h \in \mathbb{N}$ et $1 \leqslant i \leqslant 2^h$, et vérifiant

$$\begin{aligned} &\mathcal{T}_{0,1} = \mathcal{X}\,, \\ &\mathcal{T}_{h,i} = \mathcal{T}_{h+1,2i-1} \cup \mathcal{T}_{h+1,2i} \qquad \text{pour tous } h \geqslant 0 \text{ et } 1 \leqslant i \leqslant 2^h. \end{aligned}$$

Pour toute profondeur $h \geqslant 0$, les sous-ensembles $\mathcal{T}_{h,i}$ recouvrent $\mathcal{X}$ lorsque i décrit $\{1, \dots, 2^h\}$, ce qui justifie l'appellation d'arbre de recouvrements pour $\mathcal{T}$.

Principe de HOO. On se contente de décrire HOO de manière informelle. A chaque nœud (h,i) de $\mathcal{T}$, on associe un estimateur du supremum de μ_E sur le sous-ensemble $\mathcal{T}_{h,i}$. Cet estimateur est défini de manière récursive à partir de ceux situés aux nœuds-fils $(h+1, 2i-1)$ et $(h+1, 2i)$. A chaque échéance t, la stratégie choisit le chemin le plus prometteur : elle suit le parcours qui, à chaque nœud, sélectionne le fils dont l'estimateur admet la plus grande réalisation. Quand elle arrive à un nœud (H_t, J_t) qui n'avait jamais été exploré (et à qui aucun estimateur n'est donc associé pour le moment), elle tire un bras I_t au hasard dans le sous-ensemble $\mathcal{T}_{H_t,J_t}$ et peut dès lors associer un estimateur à ce nœud (même s'il ne sera sans doute pas très précis pour l'instant).

Références. Cette stratégique hiérarchique reposant sur un arbre est inspiré des techniques et algorithmes exposés dans [KS06, GWMT06, CM07].

Notion de pseudo-divergence et hypothèses sur les paramètres de HOO. Une pseudo-divergence ℓ est une application $\mathcal{X}^2 \to \mathbb{R}_+$ telle que $\ell(x,x) = 0$ pour tout $x \in \mathcal{X}$ (mais qui ne vérifie pas nécessairement ni l'axiome de symétrie, ni celui de séparation, ni l'inégalité triangulaire). On note $\mathcal{B}(x,r)$ la boule de centre x et de rayon r pour ℓ et on considère l'hypothèse suivante.

Hypothèse 4.3. Les paramètres de HOO ont été choisis de telle sorte qu'il existe une pseudo-divergence ℓ et un réel $\nu_2 > 0$ tels que, pour tout entier $h \geqslant 0$,

(a) pour tout $1 \leqslant i \leqslant 2^h$, le diamètre de $\mathcal{T}_{h,i}$ vérifie $\sup_{x,y \in \mathcal{T}_{h,i}} \ell(x,y) \leqslant \nu_1 \rho^h$;

(b) pour tout $1 \leqslant i \leqslant 2^h$, il existe $x_{h,i} \in \mathcal{T}_{h,i}$ tel que $\mathcal{B}_{h,i} \stackrel{\text{def}}{=} \mathcal{B}(x_{h,i},\, \nu_2 \rho^h) \subseteq \mathcal{T}_{h,i}$;

(c) pour tous $1 \leqslant i < j \leqslant 2^h$, les boules $\mathcal{B}_{h,i}$ et $\mathcal{B}_{h,j}$ sont disjointes.

Environnements de fonction d'espérances faiblement $(1,\ell)$–lipschitzienne. Pour toute pseudo-divergence ℓ, on note $\mathcal{F}_{1,\ell}$ la classe des environnements E tels que leur fonction d'espérances des paiements μ_E vérifie

$$\forall\, (x,y) \in \mathcal{X}^2, \qquad \mu_E^\star - \mu_E(y) \leqslant \mu_E^\star - \mu_E(x) + \max\Big\{\mu_E^\star - \mu_E(x),\ \ell(x,y)\Big\}$$
$$\text{où} \quad \mu_E^\star = \sup_{x \in \mathcal{X}} \mu_E(x)\,.$$

On dit d'une telle fonction μ_E qu'elle est faiblement lipschitzienne par rapport à ℓ, de constante de Lipschitz faible égale à 1. En effet, lorsque μ_E est 1–lipschitzienne (au sens ordinaire) sur $\mathcal{X}$ par rapport à ℓ, elle est en particulier faiblement lispchitzienne.

Un exemple de contrôle uniforme du regret. L'essentiel de [11] porte sur l'amélioration des ordres de grandeur des bornes sur le regret sous certaines conditions sur $\mathcal{X}$ et sur μ_E, plus faibles que celles de [KSU08]. Pour la simplicité du propos, on ne présente ci-dessous qu'un cas très particulier de ces bornes, où grâce à une hypothèse topologique suffisamment forte sur $\mathcal{X}$, on obtient un contrôle uniforme du regret face à une classe assez grande d'environnements ; ce résultat est prouvé là aussi sous des conditions un peu plus faibles que celles de [KSU08], qui requérait, par exemple, le caractère lipschitzien (au sens ordinaire) des fonctions μ_E par rapport à une distance d sur $\mathcal{X}$. L'hypothèse topologique porte sur la dimension de remplissage de $\mathcal{X}$ par ℓ.

Définition 4.4. Pour toute pseudo-divergence ℓ et tout $\varepsilon > 0$, on note $\mathcal{N}(\mathcal{X}, \ell, \varepsilon)$ le nombre maximal de boules disjointes de rayon ε pour ℓ que contient $\mathcal{X}$ (c'est le nombre d'ε–remplissage). On définit alors la ℓ–dimension de remplissage d'un ensemble $\mathcal{X}$ comme la quantité

$$D_{\mathcal{X},\ell} = \limsup_{\varepsilon \to 0} \frac{\ln \mathcal{N}(\mathcal{X}, \ell, \varepsilon)}{\ln(1/\varepsilon)}\,.$$

Théorème 4.5. *On considère la stratégie* HOO *associée aux paramètres fixés ν_1, ρ et $\mathcal{T}$. Alors, pour toute pseudo-divergence ℓ vérifiant l'hypothèse 4.3 et pour tout réel $D > D_{\mathcal{X},\ell}$,*

$$\limsup_{T\to\infty} \frac{\sup_{E\in\mathcal{F}_{1,\ell}} \mathbb{E}\big[R_T(\text{HOO}, E)\big]}{T^{(D+1)/(D+2)}(\ln T)^{1/(D+2)}} < \infty\,.$$

Optimalité de cette borne. On montre ensuite dans [11] que lorsque la pseudo-divergence ℓ est une distance, l'ordre de grandeur en T de la borne uniforme obtenue ci-dessus sur le regret est optimale. Cette démonstration est effectuée en exhibant tout d'abord une réduction au cas d'un nombre fini K suffisamment grand de bras et en recourant ensuite aux résultats de bornes inférieures de [ACBFS02]. Il est à noter que [KSU08] propose également un résultat d'optimalité très similaire, fondé sur des techniques différentes.

4.1.4 Perspectives : adaptation aux paramètres de régularité inconnus

Pour l'heure, la littérature sur les problèmes de bandits à continuum de bras est encore assez exploratoire et formule des résultats seulement à moitié satisfaisants. Le Théorème 4.5 est typique de ceux exhibés pour l'instant : on part d'une stratégie, calibrée avec certains paramètres, et on montre qu'elle minimise son regret au moins par rapport à une certaine classe d'environnements. Cependant, cette dernière, bien que souvent assez massive, est généralement définie en fonction des paramètres de la stratégie et/ou de quelques paramètres additionnels inconnus.

C'est pourquoi nous avons précisé ici de manière explicite ces paramètres dans l'indexation des classes d'environnements : il s'agit de la pseudo-divergence ℓ et de la constante de Lipschitz faible, qui vaut 1. En particulier, la stratégie HOO exploite le fait que cette constante vaille 1 et ne cherche pas à l'estimer (ni, de manière plus générale, à estimer la régularité de l'environnement).

Un problème de point du vue. Evidemment, un statisticien dirait que ce qui pré-existe, c'est l'environnement E et qu'il faut construire la stratégie en fonction de E, et non pas le contraire. Toutefois, on peut éventuellement avoir défini un modèle statistique et savoir que l'environnement admet une fonction μ_E par exemple faiblement lipschitzienne par rapport à une pseudo-divergence ℓ et pour une constante de Lipschitz L. Ces paramètres ℓ et L étant tous deux inconnus, le but est alors de calibrer séquentiellement les paramètres de la stratégie de sorte que son regret soit minimisé.

Idéalement, il faudrait retrouver pour le regret de cette stratégie adaptative un contrôle uniforme similaire à celui du Théorème 4.5 (avec cependant des ordres de grandeurs différents, ce qui permet de mesurer le coût de l'adaptation). Un problème encore plus difficile serait de même ignorer l'intervalle dans lequel se trouvent les paiements (ici, il s'agissait de $[0,1]$).

Lien entre apprentissage et statistique adaptative. Ce renversement de point de vue et l'adaptation aux paramètres de régularité inconnus de l'environnement sont une

occasion de jeter un pont entre les techniques et cadres considérés en apprentissage et ceux de la statistique adaptative. Une idée préliminaire en ce sens nous a été proposée par Pascal Massart : utiliser les résultats bien connus d'approximation de fonctions régulières par histogrammes ; et à cet effet, transposer les résultats du cadre statistique classique d'estimation d'histogrammes (par exemple, par techniques de sélection de modèles) dans le cadre des bandits stochastiques.

4.2 Autres travaux [14, 15]

Détection de la manipulation de données macro-économiques [14]

Ce travail est essentiellement une étude empirique des données macro-économiques publiées chaque trimestre par les Etats à destination des investisseurs. Pour des raisons que nous rappelons et détaillons, le premier chiffre significatif de ces données a tendance à suivre la loi dite de Benford, qui précise la répartition, non uniforme, attendue pour les différentes occurrences possibles, à savoir $\{1, 2, \ldots, 9\}$. Nous commençons par vérifier ce fait pour l'ensemble des données, ainsi que pour différents sous-jeux de données déterminés par des conditions géographiques ou des conditions économiques favorables. Nous exhibons alors des conditions financières et économiques défavorables telles que les Etats qui les rencontrent ont, d'un point de vue théorique, des incitations à publier des données inexactes (embellies) et montrons que c'est bien ce qui semble observé en pratique.

Contenu mathématique. Hormis l'introduction d'une méthodologie destinée à rendre moins sensibles des problèmes de dépendance entre les données d'un trimestre à un autre, le travail effectué consiste essentiellement en un travail de passage en revue de travaux théoriques antérieurs (en probabilités ou en économie) et en l'application d'un nombre important de tests d'ajustement du χ^2.

Rédaction d'un manuel de niveaux master / agrégation [15]

L'ouvrage *Statistique en action*, co-écrit avec Vincent Rivoirard, propose un cours condensé de statistique suivi de huit problèmes corrigés. L'essentiel du texte consiste en un exposé raisonné des fondements de la statistique et en leur illustration et mise en action sur différents thèmes de modélisation. Du point de vue mathématique, l'accent a été mis sur l'obtention de bornes non asymptotiques et sur la calibration des paramètres des algorithmes d'estimation (autant que cela est possible à ce niveau), ainsi que sur les liens et prolongements des thèmes traités avec des questions de recherche actuelles.

Bibliographie

[AB09] J.-Y. Audibert et S. Bubeck : Minimax policies for adversarial and stochastic bandits. *In Proceedings of the Twenty-Second Annual Conference on Learning Theory (COLT)*, 2009.

[ABCP10] A. Antoniadis, X. Brossat, J. Cugliari et J.M. Poggi : Clustering functional data using wavelets. *In Proceedings of the Nineteenth International Conference on Computational Statistics (COMPSTAT)*, 2010.

[ABR07] J. Abernethy, P.L. Bartlett et A. Rakhlin : Multitask learning with expert advice. *In Proceedings of the Twentieth Annual Conference on Learning Theory (COLT)*, pages 484–498, 2007.

[ACBF02] P. Auer, N. Cesa-Bianchi et P. Fischer : Finite-time analysis of the multiarmed bandit problem. *Machine Learning Journal*, 47:235–256, 2002.

[ACBFS02] P. Auer, N. Cesa-Bianchi, Y. Freund et R. Schapire : The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32:48–77, 2002.

[ACBG02] P. Auer, N. Cesa-Bianchi et C. Gentile : Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64:48–75, 2002.

[Agr95] R. Agrawal : The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33:1926–1951, 1995.

[AHKS06] A. Agarwal, E. Hazan, S. Kale et R.E. Schapire : Algorithms for portfolio management based on the Newton method. *In Proceedings of the Twenty-Third International Conference on Machine Learning*, 2006.

[ANN04] C. Allenberg-Neeman et B. Neeman : Full information game with gains and losses. *In Proceedings of the Fifteenth International Conference on Algorithmic Learning Theory (ALT)*, pages 264–278, 2004.

[AOS07] P. Auer, R. Ortner et C. Szepesvári : Improved rates for the stochastic continuum-armed bandit problem. *In Proceedings of the Twentieth Annual Conference on Learning Theory (COLT)*, pages 454–468, 2007.

[Aum74] R.J. AUMANN : Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.

[Aum87] R.J. AUMANN : Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55:1–18, 1987.

[AW01] K.S. AZOURY et M. WARMUTH : Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, 43:211–246, 2001.

[BDR05] A. BRUHNS, G. DEURVEILHER et J.-S. ROY : A non-linear regression model for mid-term load forecasting and improvements in seasonnality. *In Proceedings of the Fifteenth Power Systems Computation Conference (PSCC)*, 2005.

[BEYG00] A. BORODIN, R. EL-YANIV et V. GOGAN : On the competitive theory and practice of portfolio selection. *In Proceedings of the Fourth Latin American Symposium on Theoretical Informatics (LATIN)*, pages 173–196, 2000.

[BK96] A.N. BURNETAS et M.N. KATEHAKIS : Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17:122–142, 1996.

[Bla56] D. BLACKWELL : An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[Blu97] A. BLUM : Empirical support for winnow and weighted-majority algorithms : Results on a calendar scheduling domain. *Machine Learning*, 26:5–23, 1997.

[BM07] A. BLUM et Y. MANSOUR : From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

[Cau01] R. CAUTY : Solution du problème de point fixe de Schauder. *Fundamenta Mathematicæ*, 170:231–246, 2001.

[CB99] N. CESA-BIANCHI : Analysis of two gradient-based algorithms for on-line regression. *Journal of Computer and System Sciences*, 59(3):392–411, 1999.

[CBFH+97] N. CESA-BIANCHI, Y. FREUND, D. HAUSSLER, D.P. HELMBOLD, R. SCHAPIRE et M. WARMUTH : How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[CBL03] N. CESA-BIANCHI et G. LUGOSI : Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51:239–261, 2003.

[CBL06] N. CESA-BIANCHI et G. LUGOSI : *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[CM07] P.-A. COQUELIN et R. MUNOS : Bandit algorithms for tree search. *In Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 67–74, 2007.

[Cop09] E. COPE : Regret and convergence bounds for immediate-reward reinforcement learning with continuous action spaces. *IEEE Transactions on Automatic Control*, 54(6):1243–1253, 2009.

[Cov65] T. COVER : Behavior of sequential predictors of binary sequences. *In Proceedings of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 263–272. Maison d'édition de l'Académie des sciences de Tchécoslovaquie, Prague, 1965.

[Cov91] T.M. COVER : Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

[CW96] X. CHEN et H. WHITE : Laws of large numbers for Hilbert space-valued mixingales with applications. *Econometric Theory*, 12(2):284–304, 1996.

[dFM03] D.P. de FARIAS et N. MEGIDDO : How to combine expert (or novice) advice when actions impact the environment. *In Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems (NIPS)*, 2003.

[DGS09] M. DEVAINE, Y. GOUDE et G. STOLTZ : Aggregation of sleeping predictors to forecast electricity consumption. Rapport technique, EDF R&D et École normale supérieure, Paris, août 2009. Voir `http://www.math.ens.fr/%7stoltz/DeGoSt-report.pdf`.

[DKO+08] V. DORDONNAT, S.J. KOOPMAN, M. OOMS, A. DESSERTAINE et J. COLLET : An hourly periodic state space model for modelling French national electricity load. *International Journal of Forecasting*, 24:566–587, 2008.

[DLS07] O. DEKEL, P.M. LONG et Y. SINGER : Online learning of multiple tasks with a shared loss. *Journal of Machine Learning Research*, 8:2233–2264, 2007.

[DMP+06] V. DANI, O. MADANI, D. PENNOCK, S. SANGHAI et B. GALEBACH : An empirical comparison of algorithms for aggregating expert predictions. *In Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.

[EHJT04] B. EFRON, T. HASTIE, I. JOHNSTONE et R. TIBSHIRANI : Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.

[FL99] D. FUDENBERG et D. LEVINE : An easier way to calibrate. *Games and Economic Behavior*, 29:131–137, 1999.

[Fos91] D. FOSTER : Prediction in the worst-case. *Annals of Statistics*, 19:1084–1090, 1991.

[Fos99] D. FOSTER : A proof of calibration via Blackwell's approachability theorem. *Games and Economic Behavior*, 29:73–78, 1999.

[FS97] Y. FREUND et R.E. SCHAPIRE : A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[FSSW97] Y. FREUND, R. SCHAPIRE, Y. SINGER et M. WARMUTH : Using and combining predictors that specialize. *In Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 334–343, 1997.

[FV91] D. FOSTER et R. VOHRA : Asymptotic calibration. Rapport technique, Graduate School of Business, University of Chicago, 1991.

[FV98] D. FOSTER et R. VOHRA : Asumptotic calibration. *Biometrika*, 85:379–390, 1998.

[FV99] D. FOSTER et R. VOHRA : Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–36, 1999.

[Ger10] S. GERCHINOVITZ : Communication personnelle, 2010.

[GLU06] L. GYÖRFI, G. LUGOSI et F. UDINA : Nonparametric kernel-based sequential investment strategies. *Mathematical Finance*, 16:337–358, 2006.

[GMS08] S. GERCHINOVITZ, V. MALLET et G. STOLTZ : A further look at sequential aggregation rules for ozone ensemble forecasting. Rapport technique, INRIA Paris-Rocquencourt et École normale supérieure, Paris, septembre 2008. Voir `http://www.math.ens.fr/%7Estoltz/GeMaSt-report.pdf`.

[GO07] L. GYÖRFI et G. OTTUCSÁK : Sequential prediction of unbounded stationary time series. *IEEE Transactions on Information Theory*, 53(5):1866–1872, 2007.

[Gou08a] Y. GOUDE : *Mélange de prédicteurs et application à la prévision de consommation électrique*. Thèse de doctorat, Université Paris-Sud, janvier 2008. Effectuée en convention avec EDF R&D.

[Gou08b] Y. GOUDE : Tracking the best predictor with a detection based algorithm. *In Proceedings of the Joint Statistical Meetings*. American Statistical Association, 2008. Voir la section de "Statistical Computing".

[GWMT06] S. GELLY, Y. WANG, R. MUNOS et O. TEYTAUD : Modification of UCT with patterns in Monte-Carlo go. Rapport technique RR-6062, INRIA, 2006.

[Han57] J. HANNAN : Approximation to Bayes risk in repeated play. *In* M. DRESHER, A. TUCKER et P. WOLFE, éditeurs : *Contributions to the Theory of Games*, volume III, pages 97–139. Princeton University Press, 1957.

[Har95] S. HART : Personal communication to Dean P. Foster, 1995.

[HK70] A.E. HOERL et R.W. KENNARD : Ridge regression : Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.

[HK08] E. HAZAN et S. KALE : Extracting certainty from uncertainty : Regret bounded by variation in costs. *In Proceedings of the Twenty-First Annual Conference on Learning Theory (COLT)*, 2008.

[HMC00] S. HART et A. MAS-COLELL : A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[HP97] D.P. HELMBOLD et S. PANIZZA : Some label efficient learning results. *In Proceedings of the Tenth Annual Conference on Computational Learning Theory (COLT)*, pages 218–230, 1997.

[HS89] S. HART et D. SCHMEIDLER : Existence of correlated equilibria. *Mathematics of Operations Research*, 14:18–25, 1989.

[HSSW98] D.P. HELMBOLD, R.E. SCHAPIRE, Y. SINGER et M.W. WARMUTH : On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8:325–344, 1998.

[HT10] J. HONDA et A. TAKEMURA : An asymptotically optimal bandit algorithm for bounded support models. *In Proceedings of the Twenty-Third Annual Conference on Learning Theory (COLT)*, 2010.

[HW98] M. HERBSTER et M. WARMUTH : Tracking the best expert. *Machine Learning*, 32:151–178, 1998.

[Kle04] R. KLEINBERG : Nearly tight bounds for the continuum-armed bandit problem. *In Proceedings of the Eighteenth Annual Conference on Neural Information Processing Systems (NIPS)*, 2004.

[KS06] L. KOCSIS et C. SZEPESVARI : Bandit based Monte-Carlo planning. *In Proceedings of the Fifteenth European Conference on Machine Learning (ECML)*, pages 282–293, 2006.

[KSU08] R. KLEINBERG, A. SLIVKINS et E. UPFAL : Multi-armed bandits in metric spaces. *In Proceedings of the Fortieth Annual ACM Symposium on the Theory of Computing (STOC)*, 2008.

[KV03] A. KALAI et S. VEMPALA : Efficient algorithms for the online decision problem. *In Proceedings of the Sixteenth Annual Conference on Learning Theory (COLT)*, pages 26–40. Springer, 2003.

[KW97] J. KIVINEN et M. WARMUTH : Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

[LR85] T.L. LAI et H. ROBBINS : Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

[LS07] E. LEHRER et E. SOLAN : Learning to play partially-specified equilibrium. Article soumis, 2007.

[LW94] N. LITTLESTONE et M. WARMUTH : The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[LZ76] A. LEMPEL et J. ZIV : On the complexity of an individual sequence. *IEEE Transactions on Information Theory*, 22:75–81, 1976.

[Mal10] V. MALLET : Ensemble forecast of analyses : Coupling data assimilation and sequential aggregation. Article soumis, 2010.

[MMS07] V. MALLET, B. MAURICETTE et G. STOLTZ : Description of sequential aggregation methods and their performance for ozone ensemble forecasting. Rapport technique DMA-07-08, École normale supérieure, Paris, 2007.

[MS03] S. MANNOR et N. SHIMKIN : On-line learning with imperfect monitoring. *In Proceedings of the Sixteenth Annual Conference on Learning Theory*, pages 552–567. Springer, 2003.

[MS06] V. MALLET et B. SPORTISSE : Ensemble-based air quality forecasts : A multimodel approach applied to ozone. *Journal of Geophysical Research*, 111(D18), 2006.

[MSZ94] J.-F. MERTENS, S. SORIN et S. ZAMIR : Repeated games. Rapport technique 9420, 9421, 9422, Université catholique de Louvain, 1994.

[Per09a] V. PERCHET : Approachability of convex sets in games with partial monitoring. Article soumis, 2009.

[Per09b] V. PERCHET : Calibration and internal no-regret with random signals. *In Proceedings of the Twentieth International Conference on Algorithmic Learning Theory (ALT)*, pages 68–82, 2009.

[Per09c] V. PERCHET : No-regret with partial monitoring calibration-based optimal algorithms. Article soumis, 2009.

[Per10] V. PERCHET : *Approchabilité, calibration et regret dans les jeux à observations partielles*. Thèse de doctorat, Université Paris VI Pierre-et-Marie-Curie, 2010.

[PS01] A. PICCOLBONI et C. SCHINDELHAUER : Discrete prediction games with arbitrary feedback and loss. *In Proceedings of the Fourteenth Annual Conference on Computational Learning Theory*, pages 208–223, 2001.

[Rob52] H. ROBBINS : Some aspects of the sequential design of experiments. *Bulletin of the American Mathematics Society*, 58:527–535, 1952.

[Rus99] A. RUSTICHINI : Minimizing regret : The general case. *Games and Economic Behavior*, 29:224–243, 1999.

[Sto05] G. STOLTZ : *Information incomplète et regret interne en prédiction de suites individuelles*. Thèse de doctorat, Université Paris-Sud, mai 2005.

[Tib96] R. TIBSHIRANI : Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.

[Vov90] V. VOVK : Aggregating strategies. *In Proceedings of the Third Annual Workshop on Computational Learning Theory (COLT)*, pages 372–383, 1990.

[Vov98] V. VOVK : A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[Vov01] V. VOVK : Competitive on-line statistics. *International Statistical Review*, 69:213–248, 2001.

[VZ08] V. VOVK et F. ZHDANOV : Prediction with expert advice for the Brier game. *In Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML)*, 2008.

[Ziv78] J. ZIV : Coding theorems for individual sequences. *IEEE Transactions on Information Theory*, 24:405–412, 1978.

[Ziv80] J. ZIV : Distortion-rate theory for individual sequences. *IEEE Transactions on Information Theory*, 26:137–143, 1980.

[ZL77] J. ZIV et A. LEMPEL : A universal algorithm for sequential data-compression. *IEEE Transactions on Information Theory*, 23:337–343, 1977.