Incomplete information and internal regret in prediction of individual sequences

Gilles Stoltz

N^o D'ORDRE : 7885

UNIVERSITÉ PARIS–XI — UFR SCIENTIFIQUE D'ORSAY

THÈSE

présentée pour obtenir le grade de

DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS–XI ORSAY

Spécialité : **Mathématiques**

présentée par

Gilles STOLTZ

INFORMATION INCOMPLÈTE ET REGRET INTERNE EN PRÉDICTION DE SUITES INDIVIDUELLES

Rapporteurs : M. Olivier **CATONI** CNRS et Université Paris–VI
M. Dean P. **FOSTER** University of Pennsylvania

Soutenue publiquement le **27 mai 2005** devant le jury composé de

M. Olivier	**CATONI**	CNRS et Université Paris–VI	Rapporteur
M. Nicolò	**CESA-BIANCHI**	Università degli Studi di Milano	Examinateur
M. Gábor	**LUGOSI**	Universitat Pompeu Fabra, Barcelone	Directeur
M. Pascal	**MASSART**	Université Paris–XI, Orsay	Co-Directeur
M. Sylvain	**SORIN**	Université Paris–VI	Président

Acknowledgements – Remerciements

I wholeheartedly thank you, Gábor, for supervising my thesis. I appreciated your constant support, your brillant ideas, your deep knowledge and understanding of statistics and machine learning. Last but not least, you always replied very quickly and very efficiently to the numerous emails I sent you during the last three years – I had the feeling that you were the guy just next door, whereas you were in Barcelona. You taught me by your example several useful elements of a researcher's life, that one should visit at least four or five different countries a year, write a book every four years, and, most importantly, that English sentences should be concise, with commas replaced by full stops and adverbs omitted. I spent some wonderful days beside you in Barcelona. In particular, I should acknowledge here the Friday evening cafés of Plaça Rius i Taulet for being the theatres of some of our mathematical exchanges.

Merci à toi Pascal de t'être proposé pour le rôle forcément un peu ingrat de conseiller en stratégie. Tu as toujours su être là pour me guider dans le monde universitaire, au Bistrot 77 pour me suggérer de me placer sous la direction de Gábor, à Orsay pour m'aider à financer mes nombreux voyages, à l'affût pour me rappeler les exigences académiques, dans l'ombre lors des différents dossiers de candidatures. Il se trouve que nous avons même parlé science à l'occasion, pour concentrer des martingales ou exhiber des bornes inférieures !

Nicolò, perhaps you have not realized how crucial your contribution was to this thesis. I learned all the theory of individual sequences from your book in progress with Gábor. This opus helped me to find my way in the numerous references and go straight to the points. Your enthusiasm and your kindness were invaluable when writing our articles.

Special thanks to my referees, Dean Foster and Olivier Catoni. Dean, I hope we will meet soon and perhaps find some common research interests. Olivier, je tenais en plus à vous remercier dans la catégorie de mes anciens enseignants, décrite plus en détails ci-dessous, pour votre cours à l'IHP lors du semestre spécial de statistiques.

Sylvain, vous avez accepté avec beaucoup de simplicité de faire partie de mon jury. Depuis quelques mois, nous faisons se parler théorie des jeux et statistiques : j'espère que ceci est le début d'une grande aventure !

Stéphane, Patricia et Vincent, vous avez été de formidables relecteurs d'introduction. Stéphane, ta connaissance du machine learning, tes blagues et ta bonne humeur m'auront été précieuses. Patricia, merci pour cette année formidable passée avec toi comme responsable adjoint des statisticiens de l'ENS (nous deux). Vincent, heureusement que le café existe, sinon nous aurions eu la gorge desséchée de refaire le monde.

Par ordre d'apparition, je tenais également à mentionner tous ceux qui m'ont ouvert la voie vers les statistiques, Thomas Duquesne, qui a réussi à m'intriguer en me parlant de cette branche des mathématiques en croissance forte, Dominique Picard et Marc Hoffman, qui m'ont initié à l'estimation lors de ma maîtrise à l'université Paris–VII, Elisabeth Gassiat, qui m'a révélé les secrets des $O_{\mathbb{P}}(1)$ et des $o_{\mathbb{P}}(1)$ et avec qui j'ai pu parler de l'inégalité de van Trees (et d'autres choses...), Sacha Tsybakov et David Pollard, qui ont ouvert mes horizons mathématiques lors du semestre spécial de statistiques à l'IHP.

Mes pensées les plus reconnaissantes vont à tous les membres de l'équipe de probabilités et statistiques d'Orsay, pour leur encadrement et leur soutien, et à ceux du DMA, qui m'ont

chaleureusement accueilli cette année. Mention spéciale aux sourires de Zaïna, Laurence, Lara et Bénédicte du DMA, Isabelle, Françoise, Marie-Christine et Valérie d'Orsay. Salutations à tous mes étudiants de l'ENS ; ensemble nous avons souvent bien ri, parfois aussi travaillé.

I would also like to send my greetings to all those I met in Pompeu Fabra, and above all, Frederic, Michael, and Fabrizio.

A nouveau par ordre chronologique, je tenais à remercier tous mes amis d'études, Stéphanie, Marielle, Aurélien, Karine, Cyril, à Cachan, ainsi que les doctorants d'Orsay, et notamment les membres et ex-membres du couloir du premier étage du bâtiment 430. Parmi eux, je voudrais citer plus particulièrement Estelle, conseillère en pots de thèse alsaciens, Céline, Béatrice, qui aime les paris sur les chevaliers de l'ordre de Malte, Reda, Ismaël, Vincent et Catherine, pour le 106, Magalie, qui a su me guider pour le choix des polices de caractères, Cédric, Marie, Antoine, le rédacteur infini, pour le 110, Laurent, Christine, Servane, Boutheina, Marc, Violaine, pour les 112–114, et Nicolas pour le rez-de-chaussée. Enfin, à l'ENS, la clique des jeunes caïmans, doctorants et chargés de recherches, Mathilde, Thierry, Mathieu, Florent, David, Sébastien, Arnaud, Raphaël. A vous tous qui avez connu mes exposés de droit du consommateur, mais avez peut-être echappé au séminaire trop court sur la vie de couple co-organisé avec Estelle et Catherine, sachez que les bavards apprécient les oreilles attentives.

Je n'ose enchaîner avec les autres amis, tous ceux qui pendant ces années m'ont permis de souffler les soirs et les fins de semaine, et d'oublier le regret interne et la prédiction de suites individuelles. Je vais plutôt conclure par une pensée pour mes parents, qui se réjouissent qu'enfin Gilles ait droit à une cérémonie, quoique moins costumée que celles de son frère. Gabriel, je suis très content que finalement, après s'être séparés un temps sur le plan académique, nos chemins convergent à nouveau vers la recherche en mathématiques – même si c'est au grand dépit d'Axelle qui doit souvent assister à des conversations insipides et techniques sur nos travaux.

Jérôme, enfin, c'est à toi que je veux dédier ce travail.

Contents

CHAPITRE 1

Prélude et vue d'ensemble des résultats

Le domaine de recherche dans lequel s'inscrit ce travail de thèse est la théorie de la prédiction des suites individuelles. Cette dernière considère les problèmes d'apprentissage séquentiel pour lesquels on ne peut ou ne veut pas modéliser le problème de manière stochastique, et fournit des stratégies de prédiction très robustes. Elle englobe aussi bien des problèmes issus de la communauté du *machine learning* que de celle de la théorie des jeux répétés. Le but de mes travaux a été de traiter un certain nombre de ces problèmes avec des méthodes statistiques, incluant par exemple les techniques de concentration de la mesure ou de l'estimation adaptative. Les résultats obtenus aboutissent, entre autres, à des stratégies d'ajustement séquentiel des prix de vente (qui correspond par exemple pour vendre des produits sur Internet), ou d'allocation séquentielle de bande passante. Des simulations sont proposées pour le problème de l'investissement dans le marché boursier.

Dans ce premier chapitre, on présente à grands traits les fondements de la prédiction des suites individuelles (section 1), puis on résume les contributions de chacune des trois grandes parties de ce manuscrit de thèse, respectivement dans les sections 2, 3 et 4, et on conclut par l'indication du plan de la thèse. Une introduction au sujet plus précise, mathématique, et contenant la présentation d'un cas d'école, est proposée au chapitre 2. Les résultats nouveaux sont décrits en détail dans le corps de la thèse, à partir du chapitre 3.

Contents

1. La théorie de la prédiction des suites individuelles

L'approche traditionnelle dans les problèmes de prédiction est de supposer que la suite des données est la réalisation d'un processus stochastique sous-jacent, dont la loi appartient à un modèle statistique, *id est*, une famille de lois possibles. Il s'agit ensuite d'étudier la possibilité, les limitations et les difficultés de la prédiction de telles suites aléatoires. Tout repose donc sur l'introduction de modèles raisonnables, ce qui dans certaines situations, comme la reconnaissance vocale, les flux de données sur Internet ou l'investissement dans le marché boursier, est une gageure.

L'objet de la théorie de la prédiction des suites individuelles est de proposer des méthodes de prédiction robustes. En particulier, on considère l'ensemble de toutes les suites de données possibles, et on ne met pas de mesure de probabilité sur ce dernier, chacune des suites possibles est prise en compte. C'est de là que vient le nom de suites individuelles (elles sont considérées individuellement).

1.1. Un cas d'école : information complète, regret externe et stratégie randomisée. On peut décrire le cadre de la théorie de la prédiction des suites individuelles dans le cas le plus simple comme suit (voir, par exemple, le cours de Lugosi [**Lug01**]). Ce cas est appelé *prédiction avec avis d'experts*. On se fixe un ensemble d'observations $\mathcal{Y}$, et on suppose que l'on a accès de manière séquentielle aux données : la suite des observations $y_1, y_2, y_3, \ldots$, n'est pas révélée d'emblée, mais pas à pas. Au t-ième pas du problème, il s'agit ainsi de prédire par $\widehat{p}_t$ ce que sera y_t, en se fondant sur les observations passées $y_1^{t-1} = (y_1, \ldots, y_{t-1})$. L'ensemble des prédictions, $\mathcal{X}$, peut être différent de l'ensemble $\mathcal{Y}$ des observations. Pour l'aider dans cette tâche, le statisticien dispose de N experts (par exemple, N estimateurs obtenus à partir de N procédures d'estimation différentes), qui eux-mêmes se fondent sur le passé observé pour former leur prédiction $f_{j,t} = f_{j,t}(y_1^{t-1}) \in \mathcal{X}$. (La manière dont le jeu se déroule et les notations seront rappelées et mises en perspective en figure 1.) Les caractéristiques de ce cas d'école, traité au chapitre 2, sont :

- **Information complète [H1] :** Les informations en notre possession à un pas t donné sont donc les conseils présents et passés des experts, de même que l'historique y_1^{t-1}. En section 3 ci-dessous et aux chapitres 5 et 6, on affaiblit cette hypothèse.
- **Regret externe [H2] :** Pour mesurer la qualité de la stratégie, on introduit une fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to [0, B]$, $B > 0$, et le but du jeu est que malgré l'obligation de prédiction séquentielle, la perte cumulée du statisticien, $\widehat{L}_n$, soit la plus proche possible de celle du meilleur expert, L_n^*, avec

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(\widehat{p}_t, y_t) \ , \qquad L_n^* = \min_{j=1,\ldots,N} L_{j,n}, \qquad \text{où } L_{j,n} = \sum_{t=1}^{n} \ell(f_{j,t}, y_t) \ .$$

  La quantité clé est ainsi la différence entre ces deux pertes cumulées, et on l'appelle le regret[1] (externe) :

$$R_n = \widehat{L}_n - L_n^* = \max_{j=1,\ldots,N} \sum_{t=1}^{n} \ell(\widehat{p}_t, y_t) - \sum_{t=1}^{n} \ell(f_{j,t}, y_t) \ . \tag{1.1}$$

  On cherche à borner le regret uniformément en les suites d'observations $y_1, y_2, \ldots$. On considère également en section 2 ci-dessous et au chapitre 3 d'autres mesures de la qualité d'une stratégie de prédiction.

[1] Voir l'appendice du chapitre 2 pour des remarques sur cette façon de mesurer la qualité d'une stratégie de prédiction ; il semble que cette mesure classique de la qualité soit la dernière trace d'un traitement stochastique des observations.

– **Fonction de perte générale [H3] :** La fonction de perte ℓ est supposée connue par le statisticien, mais elle est généralement arbitraire, c'est-à-dire non spécifiée explicitement, auquel cas sa stratégie est randomisée. Pour des fonctions de perte disposant de propriétés supplémentaires, de convexité par exemple, des stratégies de prédiction déterministes peuvent être introduites – comme c'est le cas pour l'investissement dans le marché boursier, voir la section 4.1 ci-dessous et le chapitre 7.

Une stratégie randomisée est donnée par une suite de mesures de probabilité

$$\boldsymbol{p}_t = (p_{1,t},\, \ldots,\, p_{j,t})\,, \qquad t = 1, 2, \ldots$$

sur l'ensemble des experts, calculées en fonction des observations jusqu'au tour $t-1$; on tire l'index I_t d'un expert parmi les N experts selon $\boldsymbol{p}_t$, c'est-à-dire qu'avec probabilité $p_{i,t}$, on prédit au pas t comme l'expert $I_t = i$. On pose alors

$$\widehat{p}_t = f_{I_t,t}\,.$$

On a ainsi introduit de l'aléatoire par une randomisation auxiliaire, sachant que les observations elles-mêmes $y_1,\, y_2,\, y_3, \ldots$, ne sont pas ou ne peuvent pas être modélisées comme la réalisation d'un processus stochastique sous-jacent. On peut noter $V_1, V_2, \ldots$ la suite de variables aléatoires auxiliaires dont on a besoin, et supposer qu'elles sont i.i.d. selon une loi uniforme sur $[0,1]$. Cette introduction de l'aléatoire est rendue nécessaire par la volonté d'obtenir des bornes uniformes vis-à-vis des observations, ce qui revient à considérer le pire des cas, c'est-à-dire à supposer que les observations sont choisies par un adversaire diabolique.

PRÉDICTION SÉQUENTIELLE RANDOMISÉE, AVEC AVIS D'EXPERTS À DISPOSITION

Paramètres : Un ensemble de prédictions $\mathcal{X}$, un ensemble d'observations $\mathcal{Y}$, N experts, n tours de jeu ($n = \infty$ est une valeur recevable).

A chaque tour $t = 1, 2, \ldots, n$,

(1) l'environnement choisit les prédictions $f_{1,t}, \ldots, f_{N,t} \in \mathcal{X}$ des experts, et le statisticien peut les consulter ;

(2) le statisticien choisit en secret une mesure de probabilité $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ sur les experts, tire au hasard l'indice I_t d'un expert selon $\boldsymbol{p}_t$, et forme la prédiction $\widehat{p}_t = f_{I_t,t} \in \mathcal{X}$;

(3) pendant ce temps, l'environnement choisit en secret l'observation $y_t \in \mathcal{Y}$;

(4) l'observation y_t et la prédiction $\widehat{p}_t$ sont portées à la connaissance de tous, et les pertes sont calculées.

FIG. 1. Description du cadre de prédiction séquentielle randomisée comme un jeu répété entre le statisticien et l'environnement.

Cet adversaire diabolique est omniscient (il connaît donc notre stratégie, mais comme il ne contrôle pas les dés, nous pouvons malgré tout le surprendre, grâce à la randomisation auxiliaire), de deux manières possibles – le point commun étant qu'il produit toujours une suite d'observations dont il sait que le statisticien aura du mal à la prédire. Si, comme en *machine learning*, toute la suite est choisie à l'avance, l'adversaire est dit oublieux. Ce cas correspond à toutes les situations où notre prédiction n'influe pas sur le cours des choses (typiquement, problème de prédiction météorologique). Mais l'adversaire peut également choisir y_t en fonction des prédictions passées

et de la stratégie du statisticien. Dans ce cas, le problème statistique apparaît comme un jeu répété entre deux joueurs de capacités différentes, l'adversaire diabolique ayant des informations supplémentaires sur les méchanismes de pensée du statisticien – c'est bien ce qui oblige d'ailleurs ce dernier à s'en remettre en partie à une randomisation auxiliaire.

Dans un premier temps, on s'attache souvent à l'étude de l'espérance $\overline{R}_n$ du regret d'une stratégie de prédiction donnée, c'est-à-dire aux quantités du type

$$\sup \overline{R}_n \ , \qquad \text{où } \overline{R}_n = \max_{j=1,\dots,N} \mathbb{E}\left[\widehat{L}_n - L_{j,n}\right] \ , \tag{1.2}$$

et où le supremum porte sur tous les paramètres du problème – *id est*, la donnée de $\mathcal{X}$ et $\mathcal{Y}$, la fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to [0,1]$, les experts – et sur toutes les suites $y_1, y_2, \dots$ que peut choisir l'opposant diabolique. Dans la formule ci-dessus, l'espérance[2] est prise par rapport à la randomisation auxiliaire $V_1, V_2, \dots$.

L'algorithme de pondération exponentielle dérivé des travaux de Vovk [**Vov90**] et de Littlestone et Warmuth [**LiWa94**] garantit une borne supérieure uniforme sur l'espérance du regret de l'ordre de $\sqrt{n}$, comme le rappelle le théorème ci-dessous. Cet algorithme repose sur un paramètre η, et propose de prédire par $\boldsymbol{p}_1 = (1/N, \dots, 1/N)$, puis, pour $t \geqslant 2$, par $\boldsymbol{p}_t$ défini composante par composante comme suit,

$$p_{i,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{i,s}, y_s)\right)}{\sum_{j=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell(f_{j,s}, y_s)\right)} \qquad \text{pour } i = 1, \dots, N \ . \tag{1.3}$$

Notons qu'ici, puisque nous connaissons au début du tour t tout le passé, les $f_{j,s}$ et les y_s, $s \leqslant t-1$, nous pouvons calculer toutes les pertes passées $\ell(f_{j,s}, y_s)$, de sorte que le choix de $\boldsymbol{p}_t$ est autorisé à dépendre de toutes ces quantités. Ce ne sera plus le cas dans les problèmes dits à information incomplète étudiés plus loin.

Cet algorithme de pondération exponentielle est efficace quels que soient $\mathcal{X}$, $\mathcal{Y}$, $\ell : \mathcal{X} \times \mathcal{Y} \to [0,1]$, et les experts, comme l'indique le théorème ci-dessous, qui est une version affaiblie du Théorème 2.1.

THÉORÈME 1 (voir Theorem 2.1). *L'algorithme de pondération exponentielle garantit, pour tout paramètre $\eta > 0$, que, quelle que soit la fonction de perte $\ell : \mathcal{X} \times \mathcal{Y} \to [0, B]$, quelle que soit la suite des observations, l'espérance du regret est bornée par*

$$\overline{R}_n = \max_{j=1,\dots,N} \mathbb{E}\left[\sum_{t=1}^{n} \ell(f_{I_t,t}, y_t) - \sum_{t=1}^{n} \ell(j, y_t)\right] \leqslant \frac{\ln N}{\eta} + \frac{n\eta}{8} B^2 \ .$$

En particulier, le choix $\eta = (1/B)\sqrt{8(\ln N)/n}$ conduit à la borne supérieure

$$\overline{R}_n \leqslant B\sqrt{(n/2)\ln N} \ .$$

La mise en œuvre de l'algorithme de prédiction ci-dessus requiert la connaissance de B et de n, vu le choix optimal pour η. Un des objets des chapitres 2 et 4 sera de rendre adaptatif ce choix du paramètre, ou *tuning*.

Dans un second temps, on déduit généralement des bornes sur le regret (non moyenné) $R_n = \widehat{L}_n - L_n^*$ par des inégalités de concentration des martingales, et notamment, dans les cas les plus

[2]Notons qu'en réalité, nous nous intéresserons dans les chapitres qui suivent au problème plus difficile de borner non pas simplement une espérance, mais une somme d'espérances conditionnelles, comme expliqué par (2.1) et les commentaires qui suivent cette équation.

simples, comme le cas présent, par l'inégalité d'Hoeffding-Azuma. Du théorème 1 on déduit qu'avec probabilité $1-\delta$ (par rapport à la randomisation auxiliaire), R_n est plus petit qu'une quantité de l'ordre de $\sqrt{n\ln(N/\delta)}$. Cela conclut l'étude des bornes supérieures sur le regret, et on désire alors obtenir des bornes inférieures ayant les mêmes ordres de grandeur.

Pour une borne $B=1$ sur les pertes, Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [**CeFrHaHeScWa97**] prouvent une borne inférieure sur le regret au tour n face à N experts de l'ordre de $\sqrt{n\ln N}$, et résolvent ainsi le problème minimax associé à (1.2). (Dans ce problème minimax, l'infimum est pris sur toutes les stratégies séquentielles du statisticien.) En effet, ils exhibent un cadre de prédiction, celui de la prédiction binaire $\mathcal{Y}=\mathcal{X}=\{0,1\}$, $\ell(x,y)=\mathbb{I}_{[x\neq y]}$, dans lequel aucune stratégie de prédiction séquentielle ne peut uniformément faire mieux que $\gamma\sqrt{n\ln N}$, pour tous n et N, où γ est une constante absolue. En réalité, ils obtiennent même que γ tend vers $1/\sqrt{2}$ lorsque n et N tendent vers l'infini, ce qui montre que même la constante du théorème ci-dessus est optimale. Cela est redétaillé et formalisé plus soigneusement à la fin du chapitre 2, qui est le chapitre d'introduction mathématique formelle au sujet.

1.2. Origines et fondements de la théorie. La théorie est à la croisée de deux chemins, celui de la théorie des jeux répétés (à somme nulle) et à celui de la compression séquentielle de données en théorie de l'information.

En théorie des jeux, on peut citer les travaux de Hannan [**Han57**], obtenus en 1956 et publiés l'année suivante, et ceux de Blackwell [**Bla56**]. Tous deux obtiennent des bornes supérieures uniformes sur le regret en $o(n)$, et en particulier, Hannan obtient une borne en $\sqrt{n}$, donc la bonne dépendance en n (mais pas en les autres paramètres, comme N).

En théorie de l'information, le problème considéré est la compression de suites individuelles, avec pour ensemble (dénombrable) d'experts l'ensemble des automates à nombre fini d'états et comme fonction de perte $\ell=\log$. Les pionniers sont Lempel et Ziv, avec la série d'articles [**LeZi76, ZiLe77, Ziv78**]. Ce sont eux qui pour la première fois parlent de suites individuelles et d'algorithmes universels, un algorithme universel étant simplement par définition tel que son regret est uniformément borné en $o(n)$. Feder, Merhav et Gutman [**FeMeGu92**] améliorent ces résultats en réduisant les bornes sur le regret et en complexifiant simultanément la classe des experts, donnée par l'ensemble des experts à nombre fini d'états. Des résultats encore plus forts ont été obtenus par exemple par les méthodes par arbres de contextes de Willems, Shtarkov et Tjalkens [**WiShTj95, WiShTj96**], reprises et revues également par Catoni [**Cat01**]. La classe des experts est formée ici par l'ensemble des prédicteurs markoviens à mémoire finie.

La première borne inférieure sur le regret, et partant, la première formalisation du problème minimax associé, a été obtenue par Cover [**Cov65**], pour le problème de prédiction binaire décrit ci-dessus, avec deux experts constants, l'un prédisant toujours 0, et l'autre 1. Il montre que le regret minimax dans ce cadre vaut $(1+o(1))\sqrt{n/(2\pi)}$. Il obtient la bonne dépendance en n, mais ici encore, rien n'est dit sur la dépendance en N (dont on verra qu'elle est suffisamment délicate dans de nombreux cadres de prédiction pour être source de problèmes ouverts).

Les algorithmes randomisés de prédiction et le cadre général de prédiction avec avis d'experts décrits à la section précédente ont été introduits par Vovk [**Vov90**] et Littlestone et Warmuth [**LiWa94**], et développés par Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [**CeFrHaHeScWa97**] et Vovk [**Vov98**], bien que certains ingrédients essentiels apparaissent déjà dans les travaux de De Santis, Markowski et Wegman [**SaMaWe88**] et Littlestone [**Lit88**]. Des survols sont proposés par Foster et Vohra [**FoVo99**] et Vovk [**Vov01**].

2. Regret interne et bornes du second ordre en prédiction avec avis d'experts

Les chapitres 3 et 4 présentent plusieurs résultats techniques utilisés dans les chapitres centraux de la thèse, à savoir les chapitres 5 à 8. Nous les décrivons très brièvement ci-dessous.

2.1. Regret interne. Borner le regret interne d'une stratégie correspond à se comparer à des modifications simples de la stratégie initiale, et à requérir qu'aucune d'entre elles n'apporte d'amélioration substantielle. Le critère de comparaison est désormais interne, parce qu'il est défini en fonction de la stratégie considérée. Le regret (1.1) est appelé regret externe car la classe de comparaison est indépendante de l'algorithme de prédiction.

Chacune des modifications simples est paramétrée par une fonction $\Phi : \{1, \dots, N\} \to \{1, \dots, N\}$ et prédit comme l'expert $\Phi(I_t)$ lorsque la stratégie principale prédit comme I_t. Formellement, le regret interne par rapport à Φ vaut

$$R_n^{\text{int},\Phi} = \sum_{t=1}^{n} \ell(f_{I_t,t}, y_t) - \sum_{t=1}^{n} \ell(f_{\Phi(I_t),t}, y_t) \; . \tag{1.4}$$

En particulier, la modification Φ peut être constante, $\Phi \equiv j$, et cela montre que rendre le regret interne petit par rapport à toutes les fonctions Φ possibles est plus difficile que de rendre petit le regret externe. Dans un premier temps on s'intéresse là encore à l'espérance du regret interne, et on déduit des résultats sur le regret interne par des méthodes mettant en jeu des martingales.

Le regret interne a été défini en théorie des jeux, par Foster et Vohra [**FoVo99**], Fudenberg et Levine [**FuLe99**], Lehrer [**Leh03**], et considéré ultérieurement par Hart et Mas-Colell [**HaMa00, HaMa01**] et Cesa-Bianchi et Lugosi [**CeLu03**]. Un résultat important de convergence vers un ensemble d'équilibres a été prouvé par Foster et Vohra [**FoVo99**] pour peu que tous les joueurs d'un jeu minimisent chacun leur regret interne, voir section 4.2 ci-dessous.

Au chapitre 3, nous proposons une méthode générale pour convertir les stratégies encourant un regret externe sous-linéaire en stratégies encourant un regret interne sous-linéaire. Notre méthode a vu le jour indépendamment de celle développée par Blum et Mansour [**BlMa05**]. Les deux méthodes sont comparées précisément, notamment en fonction de leurs complexités de mise en œuvre respectives, de leurs bornes théoriques et de leur extension à des situations d'information incomplète comme celles décrites en section 3. Le chapitre est conclu par une indication de la vitesse minimax du regret interne en n, qui se trouve être également $\sqrt{n}$, comme pour le regret externe. Le problème de la vitesse minimax en N n'est résolu quant à lui qu'à un facteur $\sqrt{\ln N}$ près.

2.2. Bornes plus fines sur le regret (externe). Au chapitre 4, on considère le cadre très général de jeu mixte introduit par Allenberg et Neeman [**AlNe04**], où les fonctions de perte ℓ sont à valeurs dans $[-B, B]$ (et non plus dans $[0, B]$). On cherche à améliorer le théorème 1 et sa borne générale $B\sqrt{n \ln N}$ dans le sens suivant. On cherche des procédures adaptatives en n et B, c'est-à-dire qui ne demandent pas la connaissance préalable du nombre de tours de jeu n ni de la borne B sur la valeur absolue des pertes, et qui permettent de remplacer $B\sqrt{n}$ par une quantité plus petite.

Cette quantité peut éventuellement dépendre de la suite individuelle prédite ; le remplacement par exemple par $\sqrt{BL_n^*}$ forme une amélioration pour les pertes petites, obtenue par Littlestone et Warmuth [**LiWa94**], Freund et Schapire [**FrSc97**], Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [**CeFrHaHeScWa97**], Auer, Cesa-Bianchi, et Gentile [**AuCeGe02**]. Notons cependant que ces améliorations pour les pertes petites sont établies pour des fonctions de pertes positives $\ell : \mathcal{X} \times \mathcal{Y} \to [0, B]$ et demandent la connaissance préalable de B (mais pas toujours celle de n).

En revanche, nous montrons comment, sans la connaissance préalable ni de B, ni de n, $B\sqrt{n}$ peut être remplacé par la quantité

$$\max_{t=1,\dots,n} \min_{j=1,\dots,N} \sqrt{\sum_{s=1}^{t} \ell(f_{j,s}, y_s)^2} \,, \tag{1.5}$$

comme indiqué au théorème 4.3, grâce à un nouvel algorithme de prédiction. Pour des pertes positives, il est par ailleurs facile de voir que (1.5) est plus petit que $\sqrt{BL_n^*}$, de sorte que l'on retrouve en particulier les résultats d'améliorations pour les pertes petites.

Enfin, on montre ensuite comment des bornes supérieures sur le regret plus fines que celle donnée par le théorème 1 et dépendant de quantités du second ordre sont obtenues pour une variante adaptative de l'algorithme de pondération exponentielle (1.3). Ce sont ces derniers résultats qui sont la clé pour l'analyse d'algorithmes de prédiction en situations d'information incomplète.

3. Contributions à la prédiction en situations d'information incomplète

On parle d'information incomplète dès que le statisticien n'a plus accès à l'observation y_t après avoir formé sa prédiction, mais dispose seulement d'un retour sur prédiction plus limité. Il ne peut alors plus calculer toutes les pertes $\ell(f_{j,t}, y_t)$, et le choix (1.3) n'est plus envisageable.

Un exemple de retour sur prédiction est la seule indication de la perte $\ell(f_{I_t,t}, y_t)$, et mène à une situation dite de prédiction face à des bandits manchots[3], décrite en détails en section 4 du chapitre 2, avec des références aux travaux classiques, et notamment celui d'Auer, Cesa-Bianchi, Freund et Schapire [**AuCeFrSc02**]. Notons que l'adversaire ou l'environnement n'est quant à lui pas restreint, de sorte que l'on peut interpréter le jeu de prédiction comme un jeu répété avec manque d'information d'un côté seulement.

Dans cette thèse, nous considérons deux autres situations d'information incomplète : dans la première, le nombre d'observations est limité, et dans la seconde, le contrôle est réduit, au sens où le statisticien n'a pas accès aux observations, mais à une version dégradée de ces dernières. Cette seconde situation est la plus générale, et elle englobe le problème des bandits manchots, et également, en un certain sens précisé à l'example 6.4, le problème de prédiction avec un nombre limité d'observations.

3.1. Nombre limité d'observations. Cette situation, appelée *label-efficient* dans le texte, a été introduite par Helmbold et Panizza [**HePa97**], qui ne l'ont étudiée que pour un problème de prédiction binaire, *id est*, $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $\ell(x, y) = \mathbb{I}_{[x \neq y]}$, où en outre, l'un des experts ne commet aucune erreur, $L_n^* = 0$. Elle est considérée au chapitre 5.

3.1.1. *Description du problème et état de l'art.* Dans le problème général de prédiction décrit en section 1.1 du présent chapitre, le statisticien a accès à tout le passé avant de former sa prédiction, aussi bien à l'historique des observations $y_1, y_2, \dots$, qu'à l'historique des conseils des experts. On peut cependant arguer que dans certains cas (comparaison séquentielle d'algorithmes de traitement de données sur des bases de données complexes par exemple), voir à chaque tour de prédiction l'observation y_t et calculer les différentes pertes $\ell(f_{j,t}, y_t)$ peut être très coûteux, en temps ou en argent.

On se fixe ainsi un budget m, qui est une fonction qui au pas t du problème associe un nombre maximal d'observations $m(t)$. Après avoir prédit au temps t, on choisit d'accéder (ou de ne pas accéder) à y_t et, partant, aux pertes $\ell(f_{j,t}, y_t)$ des différents experts, sachant que l'on ne peut

[3] C'est ainsi qu'on appelle en France les machines à poignées des casinos ; quant à savoir pourquoi on se réfère à une rangée de machines de casino, nous renvoyons le lecteur à l'introduction en anglais.

demander y_t que si l'on a effectué jusqu'à présent $m(t) - 1$ observations ou moins. Le problème de la section 1.1 du présent chapitre correspond au cas où $m(t) = t$ pour tout t.

La probabilité $\boldsymbol{p}_t$ selon laquelle on tire I_t ne peut alors dépendre que des observations que l'on a faites, c'est-à-dire des observations y_s associées aux tours $s \leqslant t-1$ auxquels on a précisément demandé à voir ces observations.

Pour la prédiction binaire, Helmbold et Panizza [**HePa97**] obtiennent une borne supérieure sur le regret d'un algorithme randomisé de l'ordre de $(n/m) \ln N$, et montrent une borne inférieure dans ce cas de l'ordre de n/m, en calculant récursivement la valeur exacte du jeu répété associé, sous l'hypothèse que $L_n^* = 0$.

3.1.2. *Résultats obtenus et techniques mises en œuvre.* Nous considérons des cadres de prédiction $\mathcal{X}$, $\mathcal{Y}$, $\ell : \mathcal{X} \times \mathcal{Y} \to [0,1]$ très généraux et évitons toute hypothèse sur les pertes, et notamment sur L_n^*.

Pour un nombre de pas n et un budget $m = m(n)$ donnés à l'avance, on exhibe un algorithme construit à partir de (1.3) tel que l'espérance du regret externe $\overline{R}_n$ est bornée par une quantité de l'ordre de $n\sqrt{\ln N}/\sqrt{m}$, quelque soit le cadre de prédiction (pour toute donnée des experts et de $\mathcal{X}$, $\mathcal{Y}$, $\ell : \mathcal{X} \times \mathcal{Y} \to [0,1]$). De même que l'algorithme principal d'Auer, Cesa-Bianchi, Freund et Schapire [**AuCeFrSc02**], cet algorithme emploie la technique de prédiction (1.3) sur des estimées $\widetilde{\ell}_{i,t}$ des pertes des experts ; et en outre, il demande à voir les observations de manière aléatoire, d'après une suite de variables aléatoires $Z_1, \ldots, Z_n$ i.i.d. de Bernoulli, de paramètre $\varepsilon \approx m/n$. Lorsque $Z_t = 1$, on accède à y_t et on peut calculer toutes les pertes. Pour $t = 1, 2, \ldots$ et $i = 1, \ldots, N$, l'estimée de la perte $\ell(f_{i,t}, y_t)$ est alors définie par

$$\widetilde{\ell}_{i,t} = \frac{\ell(f_{i,t}, y_t)}{\varepsilon} Z_t \, .$$

C'est bien une estimation (quand $Z_t \neq 0$, on connaît par définition $\ell(f_{i,t}, y_t)$), et elle est sans biais. Les performances en espérance de cet algorithme sont analysées grâce aux bornes du second ordre évoquées à la fin de la section précédente.

Par des inégalités de concentration des martingales, à savoir l'inégalité de Bernstein pour des accroissements de martingales (voir, e.g., Freedman, [**Fre75**]), on peut même renforcer ce résultat et obtenir des bornes, non pas en espérance par rapport à la randomisation auxiliaire, c'est-à-dire sur $\overline{R}_n$, mais directement sur le regret R_n. Cela forme le premier théorème d'importance du chapitre 5.

THÉORÈME 2 (voir Theorem 5.2). *Pour un horizon de prédiction n, un nombre d'observations m, et un niveau $\delta \in (0,1)$ donnés, on construit explicitement un algorithme de prédiction dépendant de n, m et δ, tel que, avec probabilité au moins $1-\delta$, cet algorithme ne demande pas à voir plus de m observations et encourt un regret borné par*

$$\forall\, t = 1, \ldots, n \qquad R_t = \widehat{L}_t - \min_{i=1,\ldots,N} L_{i,t} \leqslant 2n\sqrt{\frac{\ln N}{m}} + 6n\sqrt{\frac{\ln(4N/\delta)}{m}}$$

contre toute stratégie de l'adversaire diabolique (ou tout comportement de l'environnement).

Une variante simple de l'algorithme du théorème ci-dessus, dont les performances sont analysées par des techniques de martingales, encourt un regret borné par une quantité de l'ordre de $\sqrt{nL_n^* \ln(Nn)/m} + (n/m) \ln N$, ce qui forme une amélioration pour les pertes L_n^* petites. En particulier, lorsque $L_n^* = 0$, on retrouve le comportement décrit par Helmbold et Panizza [**HePa97**] pour le cadre de prédiction binaire.

On conclut par l'indication d'une borne inférieure sur le regret, qui résout le problème minimax associé à la prédiction avec un nombre limité d'observations. Elle est obtenue en utilisant les techniques de l'estimation adaptative, et notamment une version du lemme de Fano généralisée à des combinaisons convexes, déduite des travaux de Birgé [**Bir05**] (voir le lemme A.13). Mais au lieu d'avoir à utiliser le lemme de Varshamov-Gilbert pour exhiber une famille finie de suites suffisamment écartées en distance ℓ^1 et suffisamment proches en distance de Kullback-Leibler comme c'est classiquement le cas, il nous suffit de choisir N mesures de probabilité sur toutes les suites d'observations possibles, et de minorer le cas le pire par le maximum des espérances sous ces N probabilités. (Voir aussi l'utilisation de cette même technique par Auer, Cesa-Bianchi, Freund et Schapire [**AuCeFrSc02**].) Après quelques manipulations propres au cas d'un algorithme de prédiction séquentielle, l'application du lemme de Fano généralisé achève la preuve du théorème suivant.

THÉORÈME 3 (voir Theorem 5.5). *Il existe un ensemble d'observations $\mathcal{Y}$, une fonction de perte $\ell : \mathbb{N} \times \mathcal{Y} \to [0,1]$, et une constante universelle $c > 0$ tels que, pour tout $N \geqslant 2$ et tout $n \geqslant m \geqslant 20\frac{e}{1+e}\ln(N-1)$, l'espérance du regret de tout algorithme de prédiction (randomisé ou non), n'utilisant que les prédictions constantes indexées par $\{1,\dots,N\}$ et ne demandant pas à voir plus de m observations sur une suite de n d'entre elles, soit supérieure à*

$$\sup_{y_1,\dots,y_n\in\mathcal{Y}} \mathbb{E}[R_n] = \sup_{y_1,\dots,y_n\in\mathcal{Y}} \left(\mathbb{E}\left[\sum_{t=1}^{n} \ell(I_t,y_t)\right] - \min_{i=1,\dots,N} \sum_{t=1}^{n} \ell(i,y_t) \right) \geqslant c\,n\sqrt{\frac{\ln(N-1)}{m}}\,.$$

Nous prouvons en particulier le résultat pour $c = \dfrac{\sqrt{e}}{(1+e)\sqrt{5(1+e)}}$.

3.2. Contrôle réduit (jeux avec signaux). On étudie au chapitre 6 un autre cadre d'information incomplète, dit de contrôle réduit (*partial monitoring* dans le texte). Dans cette situation, la forme de la fonction de perte importera beaucoup, et on se restreint à $\mathcal{X} = \{1,\dots,N\}$, $\mathcal{Y} = \{1,\dots,M\}$. On signifie par là que les ensembles $\mathcal{X}$ et $\mathcal{Y}$ doivent être finis, et que dans ce cas, on renomme leurs éléments comme indiqué ci-avant. Les experts s'identifient simplement aux actions $j = 1,\dots,N$, au sens où pour tout t, $f_{j,t} \equiv j$. (On a donc également $\widehat{p}_t = I_t$.) Après avoir prédit y_t par I_t, le statisticien n'observe pas y_t et n'a accès qu'à une variable de contrôle $h(I_t,y_t)$, ou signal, où h est une fonction dite de feed-back, $h : \mathcal{X}\times\mathcal{Y} \to \mathcal{S}$, et $\mathcal{S}$ est un ensemble de signaux possibles. En particulier, le choix de $\boldsymbol{p}_t$ se fait en fonction uniquement des retours $h(I_s,y_s)$, $s \leqslant t-1$, et non plus en fonction des pertes passées des experts.

La situation où $h(x,y) = y$, quels que soient $x \in \mathcal{X}$ et $y \in \mathcal{Y}$, est celle décrite en section 1.1. Celle où $h = \ell$ correspond au problème des bandits manchots. Mais l'information donnée par la fonction de feed-back peut être bien plus limitée, et un des buts de l'étude de ce problème est de caractériser les paires (ℓ,h) pour lesquelles le regret peut être uniformément borné en $o(n)$, et de voir quelle est la vitesse minimax de convergence du regret sur l'ensemble de ces paires.

3.2.1. *Origine du problème et état de l'art.* La notion de contrôle réduit prend sa source dans la théorie des jeux répétés à somme nulle, voir par exemple Mertens, Sorin et Zamir [**MeSoZa94**] et Sorin [**Sor02**] pour des survols et des références aux premières formulations du problème.

Les travaux récents de Rustichini [**Rus99**] et Mannor et Shimkin [**MaSh03**] portent sur la détermination, pour chaque paire (ℓ,h), de la meilleure quantité à laquelle on peut comparer la perte cumulée $\widehat{L}_n$ du statisticien. En général, ils considèrent donc des mesures de la qualité d'une stratégie différentes du regret R_n.

Les informaticiens que sont Piccolboni et Schindelhauer [**PiSc01**] et Helmbold, Littlestone et Long [**HeLiLo00**] portent bien quant à eux leur attention sur le regret. Helmbold, Littlestone

et Long [**HeLiLo00**] étudient le problème dans le cas particulier de la prédiction binaire où l'un des experts ne commet aucune erreur, $L_n^* = 0$, pour une fonction de feed-back h définie par $h(x, y) = y$ si $x = 0$, et $h(x, y) = 1$ si $x = 1$. (Le retour sur prédiction n'est donc informatif que quand 0 est prédit.)

Piccolboni et Schindelhauer [**PiSc01**] travaillent avec des fonctions de perte ℓ et de feed-back h générales, et proposent au passage des applications en informatique, au problème d'allocation séquentielle de bande passante. Ils déterminent les paires (ℓ, h) qui permettent que l'espérance $\overline{R}_n$ du regret soit uniformément bornée en $o(n)$, et introduisent un algorithme tel que $\overline{R}_n$ est uniformément borné par une quantité de l'ordre de $n^{3/4}$ pour chacune de ces paires. Mertens, Sorin et Zamir [**MeSoZa94**] exhibent quant à eux un cadre de prédiction (ℓ, h) tel qu'aucun algorithme de prédiction ne peut encourir un regret plus petit qu'une quantité de l'ordre de $n^{2/3}$.

3.2.2. *Résultats obtenus et techniques mises en œuvre.* L'objet de nos travaux a été de combler le fossé entre la borne inférieure et les bornes supérieures, et de déterminer la vitesse minimax de convergence du regret, à savoir $n^{2/3}$, en réétudiant l'algorithme général de [**PiSc01**]. Avec des techniques similaires à celles du chapitre 5, on prouve l'alternative suivante, qui, au vu de la borne inférieure proposée par Mertens, Sorin et Zamir [**MeSoZa94**], résout le problème minimax.

THÉORÈME 4 (voir Corollary 6.2). *Pour tout problème de contrôle réduit (ℓ, h), si le regret peut être borné uniformément en $o(n)$, alors la stratégie de prédiction de la section 3 du chapitre 6 encourt un regret majoré en espérance par une quantité de l'ordre de $n^{2/3}$. De plus, le regret est également majoré avec probabilité $1 - \delta$ par une quantité de l'ordre de $n^{2/3} \ln(1/\delta)$.*

Les couples (ℓ, h) tels qu'un regret en $o(n)$ puisse être atteint uniformément sont caractérisés par [**PiSc01**] ; ce sont les couples (ℓ, h) tels qu'essentiellement, ℓ peut être reconstruite en un certain sens à partir de h, ce qui permet de définir des estimateurs des pertes $\ell(j, y_t)$, où j désigne l'un quelconque des experts, à partir des quantités observées (les $h(I_t, y_t)$).

Par ailleurs, on a introduit une nouvelle preuve de la borne inférieure, sur un exemple différent de celui de [**MeSoZa94**], en considérant une version modifiée d'un problème de prédiction avec un nombre limité d'observations comme un problème de prédiction avec contrôle réduit. Avec les mêmes techniques usuelles d'obtention des bornes inférieures en statistique minimax, déjà utilisées par [**AuCeFrSc02**], à savoir une randomisation sur les observations et l'utilisation de l'inégalité de Pinsker, on prouve le théorème suivant.

THÉORÈME 5 (voir Theorem 6.3). *Pour le problème de contrôle réduit inspiré par la prédiction avec un nombre limité d'observations, défini par $N = 3$, $M = 2$ et des fonctions ℓ et h correctement choisies, pour tout $n \geqslant 8$ et pour toute stratégie de prédiction (randomisée ou non),*

$$\sup_{y_1,\dots,y_n \in \mathcal{Y}} \mathbb{E}[R_n] = \sup_{y_1,\dots,y_n \in \mathcal{Y}} \left(\mathbb{E}\left[\sum_{t=1}^{n} \ell(I_t, y_t) \right] - \min_{i=1,2,3} \sum_{t=1}^{n} \ell(i, y_t) \right) \geqslant \frac{n^{2/3}}{5} ,$$

où $\mathbb{E}$ est l'espérance par rapport à la randomisation auxiliaire qu'utilise la stratégie.

Enfin, nous motivons le problème de prédiction avec contrôle réduit par l'exemple suivant, dit d'ajustement séquentiel du prix de vente, qui est une version modifiée d'un modèle introduit par Kleinberg et Leighton [**KlLe03**]. Il permet notamment d'obtenir des stratégies de vente de produits sur Internet.

EXEMPLE 1. A chaque pas t, un client se présente pour acheter un produit donné. Il réalise l'achat si le prix affiché par le commerçant, $I_t \in [0, 1]$, est inférieur au prix seuil $y_t \in [0, 1]$ qu'il a, consciemment ou inconsciemment, en tête. Si l'achat est effectué, le commerçant encourt une

perte, un manque à gagner, de $y_t - I_t$ (il aurait pu élever son prix et gagner plus). Si le client ressort sans acheter, la perte du commerçant est fixe et correspond aux frais de stockage, $c \in [0, 1]$. Cependant, pour améliorer sa stratégie de vente, le commerçant n'a pas accès aux y_t – que les clients eux-mêmes seraient bien en difficulté de préciser. Tout ce qu'il voit, c'est si l'achat a eu lieu ($h(I_t, y_t) = 1$) ou non ($h(I_t, y_t) = 0$). Le but du commerçant est de réaliser un chiffre d'affaires presqu'aussi élevé que le meilleur prix constant $p^* \in [0, 1]$ pour la suite donnée de clients. (Ce prix est par exemple celui qu'aurait peut-être proposé une étude de marché préalable.)

On explicite un algorithme randomisé tel que le regret, *id est*, la différence des pertes cumulées entre notre stratégie et celle du meilleur prix constant, croît à la vitesse sous-linéaire $n^{4/5}$, quelque soit la suite des clients (de sorte que le regret rapporté au nombre de clients tend vers 0). Cet algorithme repose sur celui du théorème 4, coupe l'intervalle des temps en segments de longueurs exponentiellement croissantes, et prend un nouveau départ au début de chaque segment, en discrétisant de plus en plus finement l'ensemble des prix $[0, 1]$ dans chaque segment.

4. Importation de la notion de regret interne pour des fonctions de perte générales

Nous spécifions désormais la fonction de perte, ci-dessous en (1.6), et nous pouvons nous affranchir de l'hypothèse de prédiction randomisée en exploitant sa structure, et notamment sa concavité.

4.1. Investissement dans le marché boursier et regret interne. Le chapitre 7 importe la notion de regret interne dans le cadre de l'investissement séquentiel dans le marché boursier, en montre sa pertinence, introduit des algorithmes le minimisant, et discute les résultats financiers obtenus sur des données réelles.

4.1.1. *(Absence de) modélisation du marché par suites individuelles.* Un domaine assez naturel d'application des suites individuelles est l'investissement séquentiel dans le marché boursier. Modéliser les évolutions de ce dernier étant un problème notoirement difficile, on peut être tenté de simplement décrire ses évolutions par le biais de rapports d'évolution, définis de la sorte. On considère N valeurs boursières, indexées par les entiers $\{1, \dots, N\}$. L'évolution de la j-ième valeur du jour t au jour $t + 1$ est décrite par le facteur multiplicatif $x_{j,t}$, qui représente le rapport entre le prix d'ouverture de j au jour $t + 1$ sur son prix d'ouverture de la veille. On définit $\boldsymbol{x}_t = (x_{1,t}, \dots, x_{N,t}) \in \mathbb{R}_+^N$, et on l'appelle le vecteur d'évolution du marché au jour t. Cette modélisation en forme de description, par suites individuelles, qui contraste fortement avec les modélisations stochastiques (essentiellement par des mouvements browniens), a été proposée par Cover [**Cov91**].

Il définit une stratégie d'investissement dans le marché boursier comme une suite de fonctions ; la t-ième de ces fonctions associe à un historique de vecteurs d'évolution $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1}$, un portefeuille $\boldsymbol{B}_t$, c'est-à-dire une distribution de probabilité sur les valeurs boursières, indiquant quelle proportion, chaque jour, des capitaux totaux est ré-investie dans chacune des valeurs boursières. Avec ces notations, au début du jour $t + 1$, nos capitaux sont

$$\boldsymbol{B}_t \cdot \boldsymbol{x}_t = \sum_{j=1}^{N} B_{j,t} x_{j,t}$$

fois plus importants que ceux de la veille à la même heure.

Formellement, pour retrouver le cadre du début de la section 1.1, on choisit pour ensemble $\mathcal{X}$ de prédictions le simplexe d'ordre N, et pour ensemble d'observations $\mathcal{Y}$ l'ensemble de tous les vecteurs d'évolution envisageables, à savoir $\mathcal{Y} = \mathbb{R}_+^N$. La qualité des stratégies d'investissement

est mesurée par une fonction de perte logarithmique

$$(\boldsymbol{B},\, \boldsymbol{x}) \in \mathcal{X} \times \mathbb{R}_+^N \;\; \mapsto \;\; \ell(\boldsymbol{B}, \boldsymbol{x}) = -\ln(\boldsymbol{B} \cdot \boldsymbol{x}) \tag{1.6}$$

et on se compare au meilleur portefeuille constant de placement. Un tel portefeuille ré-investit invariablement chaque jour selon la même distribution fixe $\boldsymbol{B}$. Ainsi, le regret est défini par

$$R_n = \sup_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \in \mathbb{R}_+^N} \sum_{t=1}^n \ell(\boldsymbol{B}_t, \boldsymbol{x}_t) - \min_{\boldsymbol{B}\in\mathcal{X}} \sum_{t=1}^n \ell(\boldsymbol{B}, \boldsymbol{x}_t) = \sup_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n} \max_{\boldsymbol{B}\in\mathcal{X}} \sum_{t=1}^n \ln \frac{\boldsymbol{B}\cdot\boldsymbol{x}_t}{\boldsymbol{B}_t\cdot\boldsymbol{x}_t}\,. \tag{1.7}$$

Notons que l'on considère une classe continue de stratégies de comparaison.

Borner ce regret, c'est rendre uniformément petit le $\log$-rapport entre l'argent gagné par le meilleur portefeuille constant de placement et celui obtenu par la stratégie d'investissement considérée ; or, Cover et Thomas [**CoTh91**] prouvent que les portefeuilles constants de placement forment une classe de comparaison riche, obtenant de bons résultats financiers. Différentes stratégies bornant ce regret existent, notamment le portefeuille universel de Cover [**Cov91**] et la stratégie EG de Helmbold, Schapire, Singer et Warmuth [**HeScSiWa98**].

4.1.2. *Etat de l'art.* Cover et Ordentlich [**Cov91, OrCo98**] établissent que la vitesse minimax de convergence du regret (1.7) est de $\sqrt{N \ln n}$. Ils exhibent dans [**Cov91, CoOr96**] un algorithme réalisant cette vitesse minimax, et ils l'appellent le portefeuille universel. Ce dernier est construit à partir de moyennes sur les portefeuilles constants de placement, pondérées par les performances de chacun de ces derniers. C'est l'équivalent en continu de (1.3), et en particulier, la détermination de $\boldsymbol{B}_t$ nécessite de calculer des intégrales portant sur tout le simplexe. Ce dernier point nécessite un nombre d'opérations élémentaires exponentiel en N, et est donc très gourmand en temps. Cover et Ordentlich eux-mêmes [**Cov91, OrCo98**] utilisent une discrétisation du simplexe et remplacent les intégrales par des sommes pour leurs simulations pratiques. Certes, Kalai et Vempala [**KaVe03a**] proposent des méthodes stochastiques plus fines et moins coûteuses en opération pour obtenir une mise en œuvre (d'une approximation stochastique) du portefeuille universel, mais l'avancée la plus significative semble être l'algorithme EG de Helmbold, Schapire, Singer et Warmuth [**HeScSiWa98**].

Ces derniers donnent deux versions de EG, l'une nécessitant que les rapports d'évolution du marché soient toujours compris entre deux valeurs connues $m > 0$ et M et bornant le regret par une quantité de l'ordre de $(M/m)\sqrt{n \ln N}$, et l'autre bornant le regret par rapport à *toutes* les évolutions possibles par une quantité de l'ordre de $n^{3/4}$. Cette seconde version de EG forme ainsi un algorithme universel d'investissement, et sa complexité de mise en œuvre est linéaire en n et N^2. Blum et Kalai [**BlKa99**] proposent une extension du portefeuille universel de Cover [**Cov91**] au cas d'un marché avec frais de transactions, mais pour l'instant, rien de semblable n'a pu être prouvé pour un algorithme facile à mettre en œuvre, tel EG. On peut aussi citer les travaux de Singer [**Sin97**] et Borodin, El-Yaniv et Gogan [**BoElGo00**], qui se focalisent essentiellement sur l'obtention de meilleurs résultats pratiques par des méthodes du type de celles utilisées en suites individuelles.

4.1.3. *Résultats obtenus.* Un premier travail a consisté à réétudier l'algorithme EG et à en proposer une analyse plus fine et plus simple, en lien avec le théorème 1 ci-dessus (voir section 2 du chapitre 7). Dans un deuxième temps, l'analyse a été poussée jusqu'à obtenir une vitesse de convergence pour le regret améliorée en $n^{2/3}$ pour une version universelle de EG (voir section 6.1).

Mais le travail a principalement consisté à importer la notion de regret interne, issue de la théorie des jeux répétés, dans le cadre de l'investissement séquentiel dans le marché boursier. Le regret interne R_n^{int} d'une stratégie est défini comme la différence entre les résultats financiers

obtenus par la stratégie et la meilleure de l'ensemble de ses modifications simples (linéaires), i.e.

$$R_n^{\text{int}} = \sup_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n} \sum_{t=1}^{n} \ell(\boldsymbol{B}_t, \boldsymbol{x}_t) - \min_{L\in\mathcal{L}} \sum_{t=1}^{n} \ell(L(\boldsymbol{B}_t), \boldsymbol{x}_t) = \sup_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n} \max_{L\in\mathcal{L}} \sum_{t=1}^{n} \ln \frac{L(\boldsymbol{B}_t)\cdot \boldsymbol{x}_t}{\boldsymbol{B}_t \cdot \boldsymbol{x}_t} ,$$

où $\mathcal{L}$ est l'ensemble des applications linéaires $\mathcal{X} \to \mathcal{X}$. (Remarquer la similarité avec (1.4).) La notion de regret interne correspond à l'envie du courtier de ne pas voir son travail critiqué par ses clients : ces derniers, au tour n, considèrent la stratégie d'investissement de leur courtier, et regardent quels auraient été leurs capitaux, la suite des vecteurs d'évolution étant égale par ailleurs, au vu de changements simples, comme par exemple oublier la valeur boursière i, et investir chaque jour tout ce qu'on avait mis dans i sur la valeur boursière j. Ceci correspond à un changement linéaire des portefeuilles d'investissement $\boldsymbol{B}_t$.

Après avoir prouvé que les stratégies existantes n'assuraient pas en général que ce regret interne est petit, on introduit de nouvelles stratégies, qui à la fois sont compétitives par rapport aux portefeuilles constants de placement et encourent un regret interne uniformément borné en $o(n)$. On obtient, exactement comme pour le regret externe (1.7), des stratégies assurant simultanément des bornes supérieures uniformes sur les regrets interne et externe, de l'ordre, selon la complexité de mise en œuvre des algorithmes, de $N \ln n$ (pour un algorithme semblable au portefeuille universel, non implémentable en pratique, notamment de complexité exponentielle en N) et de $\sqrt{n \ln N}$ ou $n^{2/3}$ (pour une famille d'algorithmes simples à mettre en œuvre).

Pour cette dernière, des simulations sur des données réelles ont prouvé que ces nouvelles stratégies obtenaient de bien meilleurs résultats en pratique que les stratégies pré-existantes, et ce, pour une complexité en temps similaire (voir l'appendice du chapitre 7).

4.2. Application en théorie des jeux répétés. Le chapitre 8 répond à une question naturelle survenue lors de la minimisation du regret interne dans le marché boursier, déterminer si la notion de regret interne nouvellement définie était la bonne généralisation de celle proposée en théorie des jeux.

Or, une propriété remarquable établie par Foster et Vohra [**FoVo97, FoVo99**] est que dans un jeu fini répété à N joueurs, si chaque joueur joue de telle sorte que son regret interne (1.4) est un $o(n)$, alors la suite des fréquences empiriques des profils d'actions joués converge vers un ensemble d'équilibres, celui des équilibres corrélés. Or, les résultats de la section 2 ci-dessus et ceux du chapitre 3 montrent que le regret interne peut être majoré uniformément en $o(n)$. D'autres procédures garantissant la convergence vers l'ensemble des équilibres corrélés ont été introduites par Fudenberg et Levine [**FuLe99**], Hart et Mas–Colell [**HaMa00, HaMa01, HaMa02**] et Lehrer [**Leh97, Leh03**]. Dans aucune de ces procédures les joueurs n'ont besoin de coordoner leurs mouvements, chacun se concentre sur son propre regret (interne).

La notion d'équilibre corrélé a été introduite par Aumann [**Aum74, Aum87**] dans le cadre de jeux finis, mais Hart et Schmeidler [**HaSc89**] l'ont étendue aux jeux infinis (à ensembles d'actions non finis). C'est alors que nous avons voulu prouver que dans un jeu où les joueurs disposent chacun (comme c'est le cas dans le marché financier) d'un ensemble d'actions convexe et compact, il y a également convergence des fréquences empiriques des profils d'actions joués vers l'ensemble limite des équilibres corrélés (au sens de [**HaSc89**]) du jeu originel dès que chaque joueur minimise son regret interne.

Ce résultat et des algorithmes généraux de minimisation du regret interne dans des jeux à ensembles d'actions convexes compacts reposent sur des résultats d'analyse fonctionnelle, ainsi que sur des théorèmes de point fixe de Schauder, et forment une extension au cas des jeux continus

tant des résultats que, en partie du moins, des méthodes de Foster et Vohra [**FoVo99**], et Hart et Mas-Colell [**HaMa01**].

5. Conclusion, perspectives et plan de la thèse

5.1. Conclusion. Un point de vue statistique a permis des avancées récentes en prédiction des suites individuelles. On peut le retrouver notamment dans les travaux d'Auer, Cesa-Bianchi, Freund et Schapire [**AuCeFrSc02**] et de Cesa-Bianchi et Lugosi [**Lug01, CeLu05**]. Cette thèse illustre également l'intérêt d'un tel parti pris statistique.

[**AuCeFrSc02**] a notamment permis de réaliser que des bornes sur l'espérance du regret ne sont pas suffisantes, et qu'il faut s'intéresser aux déviations à l'espérance, grâce à des inégalités de concentration des martingales. Si les déviations sont d'un ordre supérieur à l'espérance, alors l'algorithme de prédiction doit être modifié jusqu'à ce que ce ne soit plus le cas. Pour tous les algorithmes introduits aux chapitres 5 et 6, nous avons soigneusement traité les déviations à la moyenne. Au chapitre 4, le critère des déviations est utilisé à la remarque 4.5 pour choisir entre deux algorithmes de prédiction.

Par ailleurs, [**AuCeFrSc02**] a introduit une première preuve d'obtention de borne inférieure ne calculant pas récursivement (une borne inférieure sur) la valeur du jeu répété associé au problème de prédiction, comme c'est le cas par exemple dans Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [**CeFrHaHeScWa97**] (prédiction avec information complète), Mertens, Sorin et Zamir [**MeSoZa94**] (prédiction avec contrôle réduit), Helmbold et Panizza [**HePa97**] (prédiction avec un nombre limité d'observations). Ils utilisent au contraire une randomisation sur les observations et l'inégalité de Pinsker. Ils ne résolvent cependant le problème minimax associé à la prédiction dans le cadre de bandits manchots qu'à un facteur $\sqrt{\ln N}$ près. Le théorème 5 semble le premier cas où les techniques usuelles de l'estimation minimax, et notamment un lemme de Fano correctement généralisé (voir le lemme A.13), ont permis de résoudre totalement le problème minimax sans s'intéresser à la valeur du jeu.

Enfin, une contribution essentielle des statisticiens a été de mieux formaliser l'énoncé des problèmes minimax associés à la prédiction de suites individuelles. Aucune référence classique n'est encore vraiment disponible, seule la pratique mathématique a parlé pour l'instant et a montré comment les théoriciens des suites individuelles formalisaient leurs problèmes minimax. Une telle description précise et explicite est proposée à la fin du chapitre 2.

5.2. Plan de la thèse. Voici maintenant le parcours que nous allons suivre. Le plan du manuscrit est indiqué en figure 2, et il se lit de bas en haut. Un trait plein signifie que le chapitre situé en haut du trait repose sur les résultats du chapitre situé sous le trait. Un trait pointillé signifie que seuls certains résultats, le plus souvent énoncés à la fin du chapitre, nécessitent la lecture du chapitre où le trait prend sa source. C'est par exemple le cas pour l'étude du regret interne en contrôle réduit, étudié tout à la fin du chapitre 6, et qui, notons-le, réunit et fait se rencontrer en quelques pages les deux grands types de résultats considérés, ceux pour la prédiction en situations d'information incomplète (partie 2, chapitres 5 et 6) et ceux s'intéressant aux extensions de la définition du regret interne (partie 3, chapitres 7 et 8). La partie 1 est introductive, au sens où elle porte sur les fondements de la théorie des suites individuelles et résume les contributions apportées à la racine de la théorie. Le chapitre 2 introduit mathématiquement le sujet dans un cadre très formel, puis les chapitres 3 et 4 présentent respectivement les résultats fondamentaux dont on aura besoin pour les parties 3 et 2. Le manuscrit de thèse s'achève par un chapitre de rappels et d'extensions de résultats fondamentaux de statistique et théorie de l'information, et par l'indication des références bibliographiques.

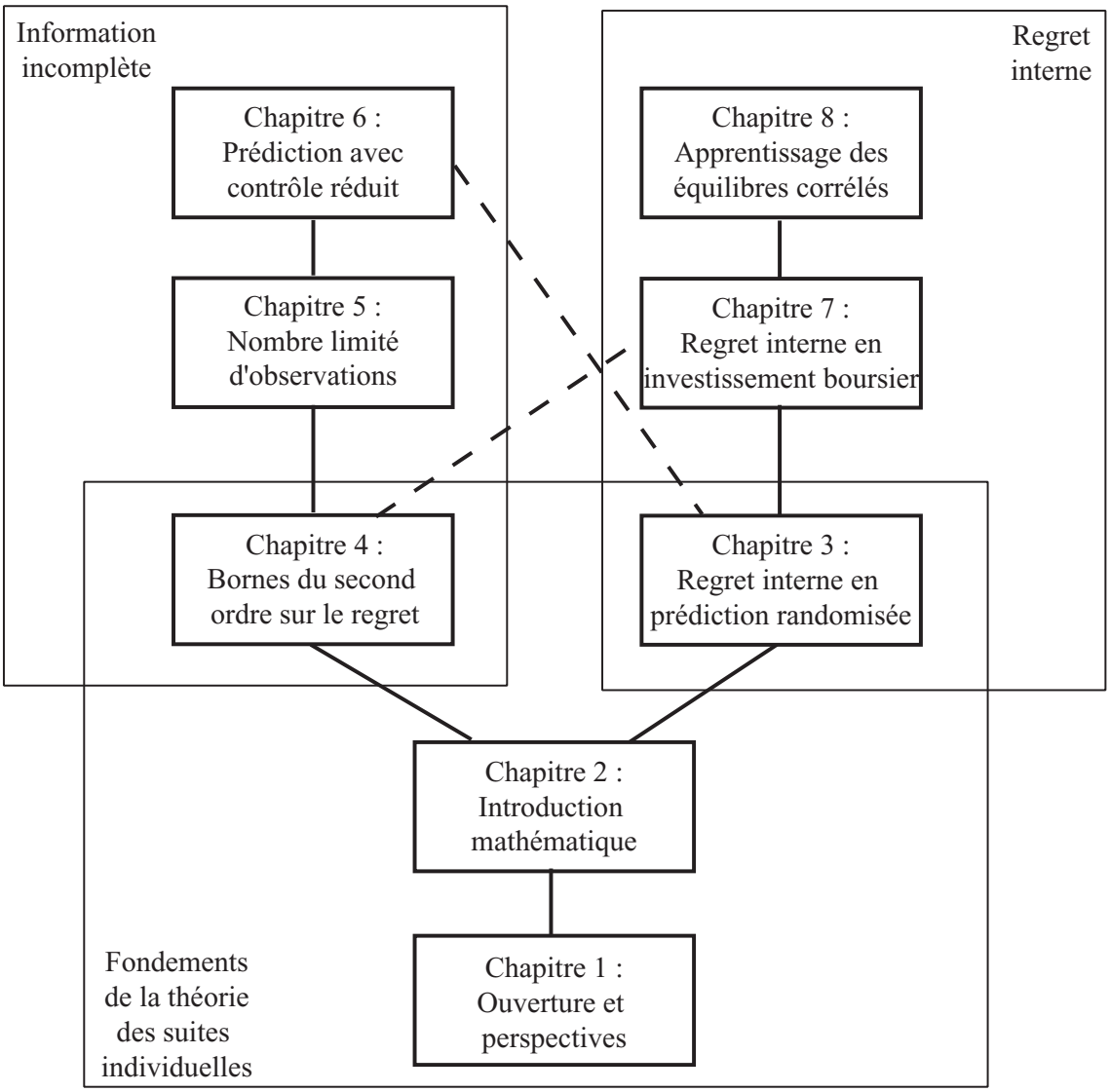

FIG. 2. Plan et organisation du manuscrit de thèse

5.3. Perspectives. Tout au long du manuscrit, les questions ouvertes sont soulignées par des environnements dédiés, de la forme suivante.

OPEN QUESTION 1.1. Avec ici l'énoncé de la question ouverte.

Une quinzaine de telles questions sont soulignées, et les plus importantes portent sur la détermination des ordres de grandeur minimax en N sur le regret interne en information complète et sur ceux du regret externe dans certains problèmes à information incomplète (bandits manchots, contrôle réduit). Toutes se situent dans la droite ligne des travaux présentés. On décrit brièvement ci-dessous deux autres axes de recherche futurs.

Le premier concerne la validation des méthodes de prédiction par suites individuelles. Hormis le cas de l'investissement séquentiel dans le marché boursier du chapitre 7, peu de simulations sur des données réelles ont été effectuées jusqu'à présent. Elles permettraient notamment de déterminer sur d'autres exemples à quel point il est souhaitable en pratique de tenir compte de tout le passé, comme le requièrent la plupart des algorithmes de prédiction par suites individuelles, et pas seulement du passé le plus proche.

Par ailleurs, un exemple d'obtention de vitesses de convergence rapides, dans le cadre des suites individuelles et pour des fonctions de perte arbitraires, est présenté au chapitre 4, en section 4.5. Rappelons que ces vitesses rapides ont été largement au centre de l'attention ces dernières années dans les problèmes de *machine learning* connus sous le nom de classification (ou apprentissage statistique). Des liens profonds existent entre classification et prédiction de suites individuelles. Dans la présentation suivante du problème de la classification, les références sont le livre de Devroye, Györfi et Lugosi [**DeGyLu96**], et le récent survol [**BoBoLu05**] des avancées récentes dans le domaine, écrit par Boucheron, Bousquet et Lugosi. En classification, on s'intéresse à une quantité appelée risque empirique, qui est l'équivalent du regret (externe) en prédiction de suites individuelles. La procédure de minimisation du risque empirique permet d'obtenir une borne en $1/\sqrt{n}$ (où n est le nombre d'exemples à classer) sur le risque empirique. Mais Mammen et Tsybakov [**MaTs99, Tsy02**] ont été les premiers à montrer que des vitesses de convergence plus rapides, entre $1/n$ et $1/\sqrt{n}$, pouvaient être obtenues, sous des conditions sur la distribution des exemples. Ses conditions ont été étendues et généralisées, notamment par Massart et Nédélec [**MaNe03**]. Récemment, Steinwart et Scovel [**StSc05**] se sont intéressés avec succès à l'obtention de vitesses rapides en classification, mais pour des procédures de SVMs [*support vector machines*], qui présentent l'avantage de pouvoir être mises en œuvre. (La procédure de minimisation du risque empirique souffre en général d'une complexité de calcul trop grande, et est essentiellement d'un intérêt théorique.) L'idée serait alors de transposer ces résultats en prédiction de suites individuelles, et d'obtenir des algorithmes de prédiction qui assurent que le regret croît strictement plus lentement que $\sqrt{n}$ sur une large classe de suites d'observations, la vitesse $\sqrt{n}$ étant au moins atteinte par l'algorithme de pondération exponentielle (1.3). La section 3.1 du chapitre 2 montre certes qu'une version de ce dernier atteint une vitesse de l'ordre de $\sqrt{L_n^*}$ pour la convergence du regret, et qu'il y a ainsi convergence rapide dès que L_n^* croît plus lentement que n. Mais une telle condition ne conduit pas à une classe suffisamment grande de suites d'observations.

Part 1

Prediction with expert advice

CHAPTER 2

Prediction of individual sequences, mathematical framework

This chapter describes the basics of sequential prediction. With the terminology introduced in the subsequent chapters, it corresponds to the minimization of external regret in a model with full information, and is meant to be a toy case for the rest of the thesis. It is partially based on the lectures [**Lug01**] that Gábor Lugosi gave at IHP four years ago. Most of the material presented here is already part of the folklore of prediction with expert advice, except maybe the discussion about unbounded losses and the concentration of refined expert bounds, but it is presented with a new viewpoint, allowing us to derive new results, and sometimes with simpler proofs.

Contents

1. Sequential prediction of individual sequences

The problem of sequential prediction may be cast as a repeated game between a decision-maker – also called below the forecaster, the statistician, the predictor, or even the prediction algorithm – and an environment – also called below the opponent player. The decision maker

has to predict an unknown (outcome) sequence $y_1, y_2, \ldots$ of elements from an *outcome space* $\mathcal{Y}$. His predictions $\widehat{p}_1, \widehat{p}_2, \ldots$ belong to a *prediction space* (also called *decision space*) $\mathcal{X}$. $\mathcal{X}$ and $\mathcal{Y}$ are usually completely arbitrary spaces, and may even be different. The forecaster computes his predictions in a sequential fashion.

The traditional approach in statistics to such problems first assumes the existence of a stochastic model for the generating mechanism of the outcome sequence and then investigates the possibilities, and limitations of the prediction of such random sequences. For example, in many applications the sequence is assumed to be a realization of some stationary process. This approach works in many cases when a tractable statistical model reasonably describes the underlying process. However, there exist situations where any statistical model is doomed to failure and more robust prediction methods are required. Typical examples of hard-to-model processes emerge, for instance, in mathematical finance or in the study of internet data streams.

The purpose of the theory of prediction of individual sequences is to provide some techniques of robust prediction and discuss their possibilities, limitations and difficulties. The robustness is in considering all possible sequences of outcomes $\mathcal{Y}^{\mathbb{N}}$. This is where the name "individual sequences" comes from.

1.1. Prediction using expert advice. Since we avoid any assumption on the sequence to be predicted, it is not immediately clear how the problem can be made meaningful. One popular possibility is to compare the predictive performance of the decision-maker to those of a set of reference forecasters which we call *experts*. We assume throughout this chapter that there is a finite number N of such experts. The sequential prediction protocol is described in Figure 1. The experts may be chosen by the opponent player, but without loss of generality we may also assume that their predictions are computed thanks to prior efficient prediction techniques, for instance, they may be given by some statistical estimators. The experts may be thought of as being misleading when the decision-maker faces a malicious opponent, or as giving helpful hints on the sequence to be predicted when they are scientific experts with worthwhile advice. We have such an open interpretation because we build below forecasting procedures competitive with respect to all possible strategies of the opponent player.

Formally, at each round $t = 1, 2, \ldots$, the decision-maker has access to the whole history of plays that consists in the past outcomes $y_1, \ldots, y_{t-1}$ and in the past experts' predictions, as well as to the present experts' predictions $f_{1,t}, \ldots, f_{N,t}$, the latter depending on the history of plays as well. (That is, the forecaster's decision must not depend on any of the future outcomes.)

The goal of the forecaster is to predict almost as well as the best expert. To make this notion mathematically precise, we introduce below a measure of the quality of the predictions formed by the decision-maker and the experts. This measure is given by a so-called loss function.

1.2. The regret as a measure of the quality of the predictions. A *loss function* is any mapping $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. Note that we often restrict our attention to bounded nonnegative losses, that is, to loss functions $\ell : \mathcal{X} \times \mathcal{Y} \to [0, B]$, where $B > 0$. We define the cumulative *regret* of the decision-maker with respect to the i-th expert of the given class of experts by

$$R_{i,n} = \sum_{t=1}^{n} \ell(\widehat{p}_t, y_t) - \sum_{t=1}^{n} \ell(f_{i,t}, y_t) = \widehat{L}_n - L_{i,n} \ ,$$

where $\widehat{L}_n$ and $L_{i,n}$ denote, respectively, the cumulative loss of the decision-maker and that of expert i,

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(\widehat{p}_t, y_t) \ , \qquad L_{i,n} = \sum_{t=1}^{n} \ell(f_{i,t}, y_t)$$

SEQUENTIAL PREDICTION

Parameters: Prediction space $\mathcal{X}$, outcome space $\mathcal{Y}$, number of experts N, number of game rounds n ($n = \infty$ is allowed).

For each round $t = 1, 2, \ldots, n$,

(1) the environment publicly chooses the experts' predictions $f_{1,t}, \ldots, f_{N,t} \in \mathcal{X}$, and the forecaster has an immediate access to them;
(2) the forecaster privately forms its prediction $\widehat{p}_t \in X$, and simultaneously, the environment privately chooses the outcome $y_t \in \mathcal{Y}$;
(3) the prediction $\widehat{p}_t$ and the outcome y_t are made public.

FIGURE 1. Sequential prediction as a repeated game between the forecaster and the environment.

(with the convention that $L_{i,0} = 0$ for all experts i). The cumulative regret with respect to the finite class of experts is simply

$$R_n = \max_{j=1,\ldots,N} R_{j,n} = \widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \, .$$

Throughout the thesis, we make the dependencies in the played actions and chosen predictions implicit.

The goal of the decision-maker is that his per-round regret goes to zero, so that he asymptotically performs almost as well as the best expert. Note that the latter may only be determined in hindsight whereas the decision-maker has to predict sequentially. We seek on-line forecasting strategies that perform almost as well as the best of those off-line strategies that correspond to predicting at each round according to the same expert.

Formally, we want to ensure

$$\frac{1}{n} \left(\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \right) \xrightarrow[n\to\infty]{} 0 \, ,$$

where the convergence is uniform over all strategies of the opponent player, that is, over all outcome sequences and all sequences of expert advice.

This ambitious goal may be achieved when the loss function ℓ has some special properties, for instance, when the decision space $\mathcal{X}$ is convex and ℓ is also convex in its first argument, see[1] [**Lug01**].

1.3. Randomized prediction using expert advice. Unless we allow some more power or some more freedom to the decision-maker, this goal is however unachievable in general. Consider, for instance, the case of 0–1 loss, $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $\ell(x, y) = \mathbb{I}_{[x \neq y]}$, which corresponds to predicting a binary sequence. Assume that the decision-maker is supplied with two experts, one of them always predicting 1, and the other one always predicting 0. It is clear that for any deterministic strategy of the predictor, there exists an outcome sequence $(y_1, \ldots, y_n) = y_1^n$ such that the predictor errs at every single time instant, that is, $\widehat{L}_n = \widehat{L}_n(y_1^n) = n$. On every outcome sequence, one of the two experts suffers a cumulative loss less than $n/2$, $\min\{L_{1,n}(y_1^n),\, L_{2,n}(y_1^n)\} \leqslant n/2$. Therefore, for all deterministic strategies of the decision-maker,

$$\sup_{y_1^n \in \mathcal{Y}^n} \left(\widehat{L}_n(y_1^n) - \min \left\{ L_{1,n}(y_1^n),\, L_{2,n}(y_1^n) \right\} \right) \geqslant \frac{n}{2} \, .$$

[1]The forecasters built therein rely on the same weighted average techniques as in Section 2.

RANDOMIZED SEQUENTIAL PREDICTION WITH EXPERT ADVICE

Parameters: Prediction space $\mathcal{X}$, outcome space $\mathcal{Y}$, number of experts N, number of game rounds n ($n = \infty$ is allowed).

For each round $t = 1, 2, \ldots, n$,

(1) the environment publicly chooses the experts' predictions $f_{1,t}, \ldots, f_{N,t} \in \mathcal{X}$, and the forecaster has an immediate access to them;

(2) the forecaster privately[a] chooses a probability distribution $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ over the set of experts, draws an expert I_t at random according to $\boldsymbol{p}_t$, and predicts as $\widehat{p}_t = f_{I_t,t}$;

(3) simultaneously, the environment privately chooses the outcome $y_t \in \mathcal{Y}$;

(4) the outcome y_t and the prediction $\widehat{p}_t$ are made public.

[a] Section 1.4 indicates that the choice of $\boldsymbol{p}_t$ could be made public, provided that I_t is drawn privately

FIGURE 2. Randomized sequential prediction with expert advice as a repeated game between the forecaster and the environment.

This is why we allow the decision-maker to randomize, and focus below on *randomized prediction using expert advice*. This problem has been studied extensively since Blackwell [**Bla56**] and Hannan [**Han57**], see the numerous references in the sections below and in the next chapters. From time to time however, we will go back to non-randomized prediction, see, for instance, Section 2.2 in Chapter 3, as well as Chapters 7 and 8. But for now, we assume that the predictor is given an i.i.d. sequence $U_1, U_2, \ldots$ of random variables with uniform law on $[0, 1]$. His forecasting strategy is given by means of probability distributions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots$ computed using the whole past history and the present experts' advice. At round t, he uses U_t to draw an expert I_t at random according to $\boldsymbol{p}_t$, see Figure 2. Then he predicts as expert I_t. His cumulative loss is thus given by

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(f_{I_t,t}, y_t)$$

and as before, we aim at comparing it to the cumulative losses $L_{j,n}$ of the experts $j = 1, \ldots, N$. However, as indicated below, we first seek uniform bounds for the expected regrets, where the expectation is taken with respect to the auxiliary randomization the forecaster has access to, and return to the general case of non-expected regret in Section 3.3. To define precisely what we mean by uniform bounds, we first describe in detail how the behavior of the opponent player is modelled.

1.4. Different models for the opponent player. Without loss of generality, we assume that the opponent player has a deterministic strategy. This we can do since we take first the supremum of the (expected) regrets over all his possible strategies before considering the infimum over all forecasting strategies for the decision-maker, see the comments after (2.9).

1.4.1. *General (game-theoretic) opponents.* Therefore, in general, a strategy for the environment is denoted by $(g,\, h_1, \ldots, h_N)$ and is given by any choice of $N + 1$ sequences of functions $g = (g_1, g_2, \ldots)$ and $h_j = (h_{j,1}, h_{j,2}, \ldots)$, $j = 1, \ldots, N$. By convention, $\mathcal{X}^0$ is the empty set. For $t \geqslant 1$, the function g_t maps $\mathcal{X}^{t-1}$ into $\mathcal{X}$, and so do also the functions $h_{1,t}, \ldots, j_{N,t}$. The experts' predictions at round t equal $f_{j,t} = h_{j,t}(\widehat{p}_1, \ldots, \widehat{p}_{t-1})$ and the outcome is $y_t = g_j(\widehat{p}_1, \ldots, \widehat{p}_{t-1})$. All these quantities are random, since the forecaster's predictions are random. The outcome y_t and the experts' advice $f_{j,t}$ are measurable with respect to the σ–algebra generated by the $U_1, \ldots, U_{t-1}$, and even with respect to the one of the $I_1, \ldots, I_{t-1}$.

Till Section 3.3, we focus on the *expected regret*

$$\overline{R}_n = \max_{i=1,\dots,N} \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \sum_{t=1}^{n} \ell_{i,t} = \max_{i=1,\dots,N} \sum_{t=1}^{n} \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) - \sum_{t=1}^{n} \ell(f_{i,t}, y_t) \,, \tag{2.1}$$

where we denoted for all $t = 1, 2, \dots$ and $i = 1, \dots, N$,

$$\ell_t(\boldsymbol{p}_t) = \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) \qquad \text{and} \qquad \ell_{i,t} = \ell(f_{i,t}, y_t) \,.$$

We call this quantity an expected regret, but it is still a random quantity. Actually, it is defined as a sum of conditional expectations, since for all $t \geqslant 1$,

$$\ell_t(\boldsymbol{p}_t) = \mathbb{E}\left[\ell(\widehat{p}_t, y_t) \,|\, U_1, \dots, U_{t-1}\right] \,.$$

Martingales inequalities allow us to deal first with these expected regrets, see Section 3.3 below.

We may now define completely the prediction problem, see Figure 3. Bounds that are uniform

AIM OF THE DECISION-MAKER

We want to design randomized forecasting strategies whose expected regrets are uniformly bounded as

$$\sup_{g,\, h_1, \dots, h_N} \overline{R}_n \leqslant \psi_n \,,$$

where $\psi_n = o(n)$ is a deterministic sequence of positive numbers, and the supremum is over all possible strategies $(g,\, h_1, \dots, h_N)$ of the opponent player.

FIGURE 3. The final statement of the problem of randomized prediction with expert advice

in the behavior of the opponent player correspond to worst-case bounds, that is, we may assume that our opponent player knows our (randomized) strategy, and reads our mind. But since our opponent does not control[2] the auxiliary randomization we use, we still have a chance to beat him, in the sense that we still may complete the plan proposed by Figure 3. This is the purpose of Section 2.

1.4.2. *Oblivious opponents.* We sometimes consider a weaker model for the opponent players, in which neither the experts' predictions nor the outcomes depend on the decision-maker's predictions. Such opponents are called *oblivious* and determine seemingly the outcome sequence and the experts' predictions before the game starts.

The model of an oblivious opponent is realistic whenever it is reasonable to believe that the actions of the forecaster do not have an effect on future outcomes of the sequence to be predicted. This is the case in many applications, such as weather forecasting or predicting a sequence of bits of a speech signal for encoding purposes. This may even be the case for sequential investment in the stock market, as long as we invest only little money and have not too large returns. However there are important cases when one cannot reasonably assume that the opponent is oblivious. The main example is when a player of a game predicts the other players' next moves and bases his action on such a prediction. In such cases the other players' future actions may depend on the

[2]Put differently, our opponent is the devil, but we may still beat him under the assumption that Somebody stronger than the devil throws the dices in this world.

action (and therefore on the forecast) of the player in any complicated way. This is the case, for instance, in the prisoner's dilemma.

2. Weighted average prediction

In this section we first study the case of the so-called *exponentially weighted average predictor* (or exponentially weighted majority predictor), and then define formally the notion of weighted average[3] prediction by indicating a large class of regret minimizing forecasters built on the same model.

2.1. The exponentially weighted majority predictor. In this section we derive regret bounds for a version of the weighted majority forecaster of Littlestone and Warmuth [**LiWa94**], see also Vovk [**Vov90**]. We consider an exponential reweighting. The resulting forecaster relies on a tuning parameter $\eta > 0$, and given this parameter, uses the distributions $\boldsymbol{p}_1 = (1/N, \ldots, 1/N)$, and $\boldsymbol{p}_t$, defined for $t \geqslant 2$ by

$$p_{i,t} = \frac{\exp\left(\eta \sum_{s=1}^{t-1} \left(\ell_s(\boldsymbol{p}_s) - \ell_{i,s}\right)\right)}{\sum_{j=1}^{N} \exp\left(\eta \sum_{s=1}^{t-1} \left(\ell_s(\boldsymbol{p}_s) - \ell_{j,s}\right)\right)} = \frac{e^{-\eta L_{i,t-1}}}{\sum_{j=1}^{N} e^{-\eta L_{j,t-1}}} \qquad \text{for } i = 1, \ldots, N \,, \tag{2.2}$$

where we used the notation introduced in the previous section. We note that this forecaster corresponds to a smoothed version of fictitious play, see [**FuLe98**], see also [**CeLu03**] and the references therein.

Versions of the following theorem appear in Cesa-Bianchi [**Ces99**], and in Cesa-Bianchi and Lugosi [**CeLu99**], see also Cesa-Bianchi, Freund, Helmbold, Haussler, Schapire, and Warmuth [**CeFrHaHeScWa97**].

THEOREM 2.1. *The exponentially weighted average forecaster with fixed tuning parameter $\eta > 0$ achieves, uniformly over all possible values of the losses $\ell_{i,t} \in [0, B]$,*

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} L_{i,n} \leqslant \frac{\ln N}{\eta} + \frac{n\eta}{8} B^2 \,.$$

In particular, with $\eta = (1/B)\sqrt{8(\ln N)/n}$, the upper bound becomes $B\sqrt{(n/2)\ln N}$.

PROOF. For $t \geqslant 1$ and $i = 1, \ldots, N$, we denote $w_{i,t} = e^{-\eta L_{i,t-1}}$, and $W_t = w_{1,t} + \ldots w_{N,t}$, so that $p_{i,t} = w_{i,t}/W_t$. Thus, on the one hand,

$$\ln \frac{W_{n+1}}{W_1} = \ln \left(\sum_{j=1}^{N} e^{-\eta L_{j,n}} \right) - \ln N \geqslant \ln \left(\max_{j=1,\ldots,N} e^{-\eta L_{j,n}} \right) - \ln N = -\eta \min_{j=1,\ldots,N} L_{j,n} - \ln N \,, \tag{2.3}$$

whereas on the other hand, for each $t = 1, \ldots, n$,

$$\ln \frac{W_{t+1}}{W_t} = \ln \frac{\sum_{j=1}^{N} e^{-\eta \ell_{j,t}} e^{-\eta L_{j,t-1}}}{\sum_{j=1}^{N} e^{-\eta L_{j,t-1}}} = \ln \left(\sum_{j=1}^{N} p_{j,t} e^{-\eta \ell_{j,t}} \right) \leqslant -\eta \sum_{j=1}^{N} p_{j,t} \ell_{j,t} + \frac{\eta^2}{8} B^2 \,,$$

where we applied Lemma A.1, due to Hoeffding [**Hoe63**]. Summing over $t = 1, \ldots, n$, we get

$$\ln \frac{W_{n+1}}{W_1} \leqslant \frac{n\eta^2}{8} B^2 - \eta \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) \,.$$

[3] In this thesis we do not consider "follow the perturbed leader" techniques, see the seminal result of Hannan [**Han57**], and, e.g., the analysis of Kalai and Vempala [**KaVe03b**], Hutter and Polland [**HuPo04**].

Combining this upper bound with the lower bound (2.3) derived above and solving for the cumulative expected loss conclude the proof. □

2.2. The doubling trick, and related on-line tunings. The above forecaster relies on the previous knowledge of the time horizon n and a bound B on the losses. In this section, we present a version of the exponentially weighted algorithm which may be computed without previous knowledge of the time length n. We deal only later with the knowledge of the bound B, in Section 3.2 and in Chapter 4. The techniques shown here, the doubling trick and the incremental update, are however the key ingredients there also.

Even if there exist predictors that already in their basic implementations do not require the knowledge of the horizon n, see Section 2.3 below, it is important to design a time-adaptive version of the exponentially weighted forecaster, for the latter is a popular method, usually achieving good results in practical situations (see, for instance, the experimental appendix of Chapter 7), and its theoretical performances may also be improved in several ways, see Section 3 below, as well as Sections 3 and 4 in Chapter 4.

2.2.1. *The doubling trick.* The doubling trick seems to be an old and well-known trick, not only in the area of on-line learning and computer science, but also, for instance, in game theory. It is not easy to trace back to the first formal statement of the trick, see perhaps [**CeFrHaHeScWa97, Vov98**] and the references therein.

The idea is to partition time into periods of exponentially increasing lengths, indexed by $r = 0, 1, \ldots$. Then in each period r the exponentially weighted average forecaster is restarted, with a parameter η_r chosen optimally depending on r. In the simplest case presented here, η_r corresponds to the optimal tuning parameter indicated by Theorem 2.1. Say that the length of period r is a^r (the popular choice $a = 2$ explains the name of the trick, since then the periods are of doubling lengths), so that the corresponding tuning parameters equal $\eta_r = (1/B)\sqrt{8(\ln N)/a^r}$, and the r-th epoch is given by the time rounds $a^r, \ldots, a^{r+1} - 1$. We get the following theorem.

THEOREM 2.2. *The doubling version of the exponentially weighted average forecaster, parametrized with $a > 1$, achieves, for all n and uniformly over all possible values of the losses $\ell_{i,t} \in [0, B]$,*

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} L_{i,n} \leqslant B\,\frac{\sqrt{a(a-1)}}{\sqrt{a}-1}\sqrt{\frac{n}{2}\ln N} + B\,(a-1)\ .$$

For $a = 2$, the bound has a leading constant equal to $1/(\sqrt{2}-1)$. This is not exactly the optimal value for a, but is very close to it.

REMARK 2.1. The proof below works simply because, provided that there are at least two regimes, that is, $R \geqslant 1$, we have $n \geqslant 1 + a + \ldots + a^{R-1}$. The factor n can be replaced by sharper bounds, as long as an inequality of the above type may be written. Denote the value of such a bound at round n by $\psi_n = \psi_n(y_1^n, \boldsymbol{p}_1^n)$. An inequality may be written whenever the sharper bounds are nondecreasing as functions of the time rounds n, see the discussion before Theorem 4.3 in Chapter 4. Depending on their precise forms, we are then able to write $\psi_n \geqslant a^{R-1}$, see the proof of Theorem 4.3 in Chapter 4, and even in some cases, $\psi_n \geqslant 1 + a + \ldots + a^{R-1}$. The latter may be written for instance when $\psi_n = n$, $\psi_n = \ell_1(\boldsymbol{p}_1) + \ldots + \ell_n(\boldsymbol{p}_n)$, or $\psi_n = L_n^*$ (where L_n^* is the cumulative loss of the best expert), see also, among others, Section 3.1 below or the proof of Theorem 4.5 in Chapter 4.

PROOF. We first use that the sum of minima is less than the minimum of the sums, and in particular, decomposing time into the above mentioned periods, indexed by $r = 0, 1, \ldots, R$, yields

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} L_{i,n} \leqslant \sum_{r=0}^{R-1} \left(\sum_{t=a^r}^{a^{r+1}-1} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} \sum_{t=a^r}^{a^{r+1}-1} \ell_{i,t} \right) + \left(\sum_{t=a^R}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} \sum_{t=a^R}^{n} \ell_{i,t} \right) .$$

We then apply Theorem 2.1 in each period r, and get

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} L_{i,n} \leqslant \sum_{r=0}^{R} \sqrt{\frac{a^r}{2} \ln N} = B \left(\frac{(\sqrt{a})^{R+1} - 1}{\sqrt{a} - 1} \right) \sqrt{\frac{\ln N}{2}} . \tag{2.4}$$

On the other hand, provided that $R \geqslant 1$,

$$n \geqslant \sum_{r=0}^{R-1} a^r = \frac{a^R - 1}{a - 1} , \quad \text{thus,} \quad (\sqrt{a})^{R+1} \leqslant \sqrt{a} \sqrt{1 + (a-1)\, n} .$$

Substituting the latter in (2.4), we get

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} L_{i,n} \leqslant B \left(\frac{\sqrt{a} \sqrt{1 + (a-1)\, n} - 1}{\sqrt{a} - 1} \right) \sqrt{\frac{\ln N}{2}} .$$

The proof is concluded by using that $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ for $x, y \geqslant 0$, and by noting that the bound of the theorem is also true when $R = 0$, that is, when $n \leqslant a - 1$. □

2.2.2. *Incremental updates of the exponentially weighted average forecaster.* The algorithm is directly inspired by the work of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**]. (We indicate here and in Section 3.1 how their bounds can be re-derived from the more general results presented in Chapter 4.)

A natural adaptive version of the optimal parameter η determined in the case of known time length is formed by defining the tuning parameter at round $t \geqslant 2$, by $\eta_t = B^{-1} \sqrt{8 \ln N / (t-1)}$. Now, the exponentially weighted average forecaster with time-varying tuning parameter predicts with $\boldsymbol{p}_1 = (1/N, \ldots, 1/N)$, and at rounds $t = 2, 3, \ldots$, with $\boldsymbol{p}_t$ defined by its i-th components, $i = 1, \ldots, N$, as

$$p_{i,t} = \frac{\exp\left(-\eta_t L_{i,t-1}\right)}{\sum_{j=1}^{N} \exp\left(-\eta_t L_{j,t-1}\right)} . \tag{2.5}$$

A simple modification of the key argument of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**] leads to Lemma 4.3. The latter, combined, for each round $t \geqslant 2$, with an application of Hoeffding's inequality A.1, which shows that with the notation of Chapter 4,

$$\Phi(\boldsymbol{p}_t,\, \eta_t,\, (-\ell_{i,t})_{i=1,\ldots,N}) \leqslant \frac{\eta_t}{8} B^2 ,$$

implies in turn the following theorem.

THEOREM 2.3. *The exponentially weighted average forecaster with time-varying tuning parameter defined above achieves, for all n and uniformly over all possible values of the losses $\ell_{i,t} \in [0, B]$,*

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\ldots,N} L_{i,n} \leqslant B \left(2 \sqrt{\frac{n}{2} \ln N} + 1 \right) .$$

Note that the main term is larger than the one of Theorem 2.1 by a factor of two. This is the usual factor we get when dealing with the dependency in the time horizon in an incremental way, see also the comments at then end of Section 3.1 below. Note also that not only the leading constant is better for the bound of Theorem 2.3 than for the one of Theorem 2.2, but above all, the forecaster takes no fresh start again and keeps on exploiting the whole past. This may result in much sharper prediction in practice.

For refined leading constants, we refer to Yaroshinsky, El-Yaniv, and Seiden [**YaElSe04**], see also Hutter and Poland [**HuPo04**].

2.3. Other functions for the reweightings. As we recall below, a whole family of predictors with performance guarantees similar to those of the exponentially weighted forecaster may be defined. See, for example, [**CeLu03**] for the details. The reweighting functions we consider in this section are often called potential functions. We focus below on the class of forecasters based on polynomial reweightings. These are of the form $\boldsymbol{p}_1 = (1/N, \ldots, 1/N)$, and, for $t \geqslant 2$,

$$p_{i,t} = \frac{\left(\sum_{s=1}^{t-1} \ell_s(\boldsymbol{p}_s) - \ell_{i,s}\right)_+^{p-1}}{\sum_{j=1}^{N} \left(\sum_{s=1}^{t-1} \ell_s(\boldsymbol{p}_s) - \ell_{j,s}\right)_+^{p-1}} \,.$$

where $p \geqslant 1$ and $(x)_+ = \max\{x,\, 0\}$ denotes the nonnegative part of the real number x. Note the similarity to (2.2), we simply replaced the exponential reweighting by a polynomial one. When $p = 2$, we recover the forecasting strategy introduced by Blackwell [**Bla56**].

These forecasters satisfy the following bound, see [**CeLu03**].

THEOREM 2.4. *The polynomial reweighted forecaster with $p \geqslant 1$ achieves, for all n and uniformly over all possible values of the losses $\ell_{i,t} \in [0, B]$,*

$$\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \leqslant B\sqrt{(p-1)nN^{2/p}} \,.$$

The upper bound is optimized for $p = 2\ln N$, and the latter choice leads to $B\sqrt{(2\ln N - 1)en}$.

Note that the forecasters using polynomial potentials do not require the previous knowledge of the time length, contrary to the basic implementation of the exponentially weighted average forecaster described above.

3. Refined bounds on the regret

3.1. Improvement for small losses. We recall in this section how the worst-case bound of Theorem 2.1 may be improved so that it depends on the cumulative loss L_n^* of the best expert rather than on the simple upper bound $Bn \geqslant L_n^*$. This comes at the cost of a worse leading constant, but results in a major improvement as soon as L_n^* is small, that is, more precisely, as soon as L_n^* grows slower than $Bn/4$ – hence the name of the improvement. The first statement of such an improvement for small losses is due to [**LiWa94**], see also [**CeFrHaHeScWa97**], who consider the absolute loss prediction setting, given by $\mathcal{X} = [0, 1]$, $\mathcal{Y} = \{0, 1\}$, and $\ell(x, y) = |x - y|$, and an improved version in [**FrSc97**].

THEOREM 2.5. *The exponentially weighted average forecaster (2.2) with fixed tuning parameter $\eta > 0$ achieves, for all n and for all possible values of the losses $\ell_{i,t} \in [0, B]$,*

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) \leqslant B\,\frac{\eta L_n^* + \ln N}{1 - e^{-\eta B}} \,,$$

where $L_n^* = \min\{L_{1,n}, \ldots, L_{N,n}\}$. *In particular, with*

$$\eta = \frac{1}{B} \ln \left(1 + \sqrt{\frac{2B \ln N}{L}} \right) , \qquad \textit{where } L > 0 ,$$

the upper bound is less than $L_n^* + \sqrt{2BL \ln N}$ *for all outcome sequences such that* $L_n^* \leqslant L$.

The proof is taken from [**FrSc97**] and is similar to that of Theorem 2.1, except that it uses the following lemma instead of Lemma A.1. (The proof of the lemma simply follows from the convexity inequality

$$e^{-\eta x} \leqslant 1 + \frac{e^{-\eta B} - 1}{B}\, x$$

for $x \in [0,1]$, and $\ln(1+u) \leqslant u$ for all $u > -1$.)

LEMMA 2.1. *For any random variable* X *with* $X \in [0, B]$ *and for any* $\eta > 0$,

$$\mathbb{E}\left[e^{-\eta X}\right] \leqslant -\frac{1 - e^{-\eta B}}{B} \mathbb{E}[X] .$$

PROOF (OF THEOREM 2.5). We simply modify the proof of Theorem 2.1, and replace the call to Hoeffding's inequality by an application of the above lemma. This leads, with the notation therein, to the upper bound

$$\ln \frac{W_{n+1}}{W_1} \leqslant -\frac{1 - e^{-\eta B}}{B} \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) ,$$

so that combining with (2.3), we get

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) \leqslant \frac{\eta B L_n^* + B \ln N}{1 - e^{-\eta B}} .$$

We recall that for all $x \geqslant 0$, $e^x - e^{-x} \geqslant 2x$, so that $x \leqslant (1 - e^{-2x})/(2e^{-x})$. With $x = \eta B$, we get

$$\begin{aligned} \frac{\eta B L_n^* + B \ln N}{1 - e^{-\eta B}} \leqslant \frac{B \ln N}{1 - e^{-\eta B}} + \frac{1 + e^{-\eta B}}{2e^{-\eta B}} L_n^* &= L_n^* + \frac{B \ln N}{1 - e^{-\eta B}} + \frac{1}{2}\left(e^{\eta B} - 1\right) L_n^* \\ &= L_n^* + \frac{1+\alpha}{\alpha} B \ln N + \frac{\alpha}{2} L_n^* , \end{aligned}$$

where we define α by $\eta B = \ln(1+\alpha)$. We upper bound the second occurrence of L_n^* by L and note that the optimal choice for α is then $\alpha = \sqrt{2B \ln N / L}$. Substituting this value concludes the proof. □

We note that the bound proposed by Theorem 2.5 is the best one at our knowledge when the value of L_n^* is known beforehand. It can be used, in combination with a doubling trick, to get an on-line algorithm, requiring only the knowledge of B and ensuring that the regret is bounded by something of the order of $\sqrt{BL_n^* \ln N}$. (See, for instance, the proof of Theorem 4.5, which is based on such an argument.)

However, to get improved constants, we need to resort to incremental updates. Similarly to Section 2.2, the best current related adaptive version of the above forecaster suffers a loss bounded, up to some constant terms, by $2\sqrt{2BL_n^* \ln N}$ – that is, we get again an extra factor of 2, see the comments after Theorem 2.3. This forecaster is introduced by Auer, Cesa-Bianchi and Gentile [**AuCeGe02**], and uses the incremental update (2.5) with $\eta_t \sim \sqrt{(\ln N)/L_{t-1}^*}$.

We note also that in Chapter 4 we give another forecaster based on the exponentially weighted average forecaster whose regret is less than $4\sqrt{BL_n^* \ln N}$, up to some constant terms, see Corollary 4.3. This forecaster uses an incremental variance-based update, and does not require previous knowledge neither of n nor of B.

3.2. Unbounded losses. One may wonder if the losses need to range between 0 and a fixed constant B, or if they could just be given by any sequence of real numbers. On the negative side, it is easy to see that there is no[4] sequential predictor such that there exists a nondecreasing sequence ψ_n, $n \geqslant 1$, such that $\psi_n = o(n)$, and uniformly over all loss sequences $\ell_{i,t} \in \mathbb{R}_+$, $i = 1, \dots, N$ and $t = 1, \dots, n$,

$$\overline{R}_n = \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\dots,N} L_{i,n} \leqslant \psi_n \, .$$

On the positive side, if the losses are bound to increase not too quickly, and provided that the rate of increase is known, something can be said. Assume that at round t, the losses $\ell_{i,t}$ are within $[0, B_t]$, where B_t is known and is non-decreasing with t. Define the tuning parameter at round t by

$$\eta_t = B_t^{-1} \sqrt{a \ln N / t} \, , \tag{2.6}$$

where the parameter $a > 0$ is determined by the analysis in each particular case. This is a natural adaptive version of the optimal parameter η determined in the case of known time length and bounded losses, see Theorems 2.3 and 2.1.

Similarly to the proof of Theorem 2.3, Lemma 4.3 combined with an application of Hoeffding's inequality A.1 implies the following theorem.

THEOREM 2.6. *Given a nondecreasing sequence $(B_t)_{n \geqslant 1}$ of positive numbers, the exponentially weighted average forecaster with time and bound-varying tuning parameter (2.6) achieves, for all n and uniformly over all possible values of the losses $\ell_{i,t} \in [0, B_t]$,*

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i=1,\dots,N} L_{i,n} \leqslant 2\sqrt{\ln N} \left(B_{n+1} \sqrt{\frac{n+1}{a}} \right) + \frac{\sqrt{a \ln N}}{8} \sum_{t=1}^{n} \frac{B_t}{\sqrt{t}} \, .$$

We obtain non-trivial bounds whenever $B_t \leqslant \gamma t^{1/2-\varepsilon}$, where γ is a constant and $\varepsilon > 0$. We could have used this theorem in Section 6 of Chapter 7. However, it turns out that in practice, we rather apply the bound proposed by Theorem 4.4, which is more general and more fundamental (since the underlying forecaster does not need to know the sequence of the B_t).

3.3. Bounds that hold with high probability. We indicated above in Section 1.4 that thanks to some martingales inequalities, we could focus on expected regrets. We now develop this argument, and explain precisely why this is so.

Bounds that hold with high probability (also referred to as non-expected bounds below) are more difficult to get than expected bounds. In fact, obtaining non-expected bounds of a larger order of magnitude than the one of expected bounds indicates that the forecaster has to be modified. This is the case, for instance, for the first forecaster described in the section about multi-armed bandit problems below, see the concerns about the deviations there. A related argument is also in Section 5 of Chapter 4, where we compare two second order forecasters, one point of comparison being given by the deviations of the regret with respect to its expectation.

[4] Whereas this is not true for sequential randomized prediction under expert advice, this may be true in other settings, like sequential investment, see Theorem 7.6 for instance.

3.3.1. *$\sqrt{n}$-rates of growth for the regret.* We denoted by $U_1, U_2, \ldots$ the i.i.d. (according, say, to a uniform law) sequence of auxiliary randomizations the forecaster has access to. His strategy is given by the sequence of probability distributions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots$ and by the chosen experts $I_1, I_2, \ldots$, which he both selects depending on the outcomes $y_1, y_2, \ldots$ chosen by the opponent and with the help of $U_1, U_2, \ldots$. We explained in Section 1.4 that y_t is measurable with respect to $I_1, \ldots, I_{t-1}$, and thus, with respect to $U_1, \ldots, U_{t-1}$. Hence, we have proved that $\ell_{I_t,t} = \ell(f_{I_t,t}, y_t)$ is measurable with respect to $U_1, \ldots, U_t$ and has conditional expectation with respect to $U_1, \ldots, U_{t-1}$ equal to

$$\ell_t(\boldsymbol{p}_t) = \sum_{j=1}^{N} p_{j,t} \ell(f_{j,t}, y_t) \; .$$

Therefore, applying the Hoeffding-Azuma inequality, see Lemma A.2, we have that for all $\delta \in]0,1[$, for all $n \geqslant 1$, and with overwhelming probability $1 - \delta \in]0,1[$,

$$\forall t = 1, \ldots, n, \qquad \sum_{s=1}^{t} \ell(f_{I_s,s}, y_s) \leqslant \sum_{s=1}^{t} \ell_s(\boldsymbol{p}_s) + B\sqrt{\frac{n}{2} \ln \frac{1}{\delta}} \; ,$$

where the loss function ℓ is bounded between 0 and B. The orders of magnitude of the typical deviations match the orders of magnitude for the expected regret obtained in Theorem 2.3.

3.3.2. *Deviations with respect to improved regret bounds.* Whenever we have sharper bounds on the expected regrets than the general $O(\sqrt{n})$ bound obtained for the exponentially weighted average predictor, like the one of Theorem 2.5, we also need a more precise concentration argument. We deal here with the case of Theorem 2.5, and carry over an analysis of the same flavour in Chapters 4 (Section 5), 5 (Section 4), and 6 (Theorem 6.1).

We introduce the sequence $X_s = \ell(f_{I_s,s}, y_s) - \ell(\boldsymbol{p}_s)$, $s = 1, \ldots, n$, which is a martingale difference sequence with respect to the filtration generated by the U_s, $s = 1, \ldots, n$. We denote $U_1^t = (U_1, \ldots, U_t)$. For all $s = 1, \ldots, n$, we note that

$$\mathbb{E}\left[X_s^2 \,|\, U_1^{s-1}\right] = \mathbb{E}\left[(\ell(f_{I_s,s}, y_s) - \ell(\boldsymbol{p}_s))^2 \mid U_1^{s-1}\right] \leqslant \mathbb{E}\left[\ell(\boldsymbol{p}_s)^2 \mid U_1^{s-1}\right] \leqslant B\ell(\boldsymbol{p}_s) \; ,$$

so that summing over s, we bounded the conditional variances as

$$V_t = \sum_{s=1}^{t} \mathbb{E}\left[X_s^2 \,|\, U_1^{s-1}\right] \leqslant B \sum_{s=1}^{t} \ell(\boldsymbol{p}_s)$$

for all $t = 1, \ldots, n$.

We now apply Corollary A.1, and get that with probability at least $1 - \delta$,

$$\forall t = 1, \ldots, n, \qquad \sum_{s=1}^{t} \ell(f_{I_s,s}, y_s) \leqslant \bar{L}_{A,n} + \sqrt{2\left(B\bar{L}_{A,n} + B^2\right) \ln(n/\delta)} + (\sqrt{2}/3) B \ln(n/\delta) \; ,$$

where we denoted the (conditional) expected cumulative loss of the forecaster by

$$\bar{L}_{A,n} = \sum_{s=1}^{t} \ell_s(\boldsymbol{p}_s) \; .$$

Substituting the bound of Theorem 2.5 and performing some simple algebra, we have proved the following corollary of it.

COROLLARY 2.1. *The exponentially weighted average forecaster (2.2) with tuning parameter*

$$\eta = \frac{1}{B} \ln\left(1 + \sqrt{\frac{2B \ln N}{L}}\right) \; , \qquad \textit{where } L > 0 \; ,$$

achieves, for all n and for all possible values of the losses $\ell_{i,t} \in [0,B]$ such that $L_n^* \leqslant L$, and with probability $1-\delta$,

$$\forall t = 1,\dots,n, \qquad \sum_{s=1}^{t} \ell(f_{I_s,s}, y_s) - \min_{i=1,\dots,N} \sum_{s=1}^{t} \ell(f_{i,s}, y_s)$$
$$\leqslant 2\sqrt{2}\sqrt{BL \ln \frac{nN}{\delta}} + 2B^{3/4} L^{1/4} \max\{1, \ln N\}^{1/4} \sqrt{\ln(n/\delta)} + 2B \ln \frac{n}{\delta} \,.$$

We may drop the condition $L_n^* \leqslant L$ and replace L by L_n^* by applying the above argument to a doubling version of the forecaster of Theorem 2.5, or to any forecaster achieving the bound of Theorem 2.5 without previous knowledge of a bound on L_n^*, like the one of [**AuCeGe02**] or the one of Corollary 4.3.

REMARK 2.2. *(Improved deviations against oblivious opponents.)* In case of an oblivious opponent, we may drop the extra $\ln n$ terms, by applying Bernstein's inequality backwards, see the techniques used in Section 4.3 of Chapter 5.

3.3.3. *Hannan consistency.* The notion of Hannan consistency is the non-expected counterpart of the uniform minimization of the expected regret studied in the previous sections. According to Hannan [**Han57**], we define a forecaster to be *(Hannan)-consistent* if for all strategies of the opponent player,

$$\limsup_{n\to\infty} \frac{1}{n} \left(\sum_{t=1}^{n} \ell(f_{I_t,t}, y_t) - \min_{j=1,\dots,N} \sum_{t=1}^{n} \ell(f_{j,t}, y_t) \right) = 0 \qquad \text{a.s.},$$

where the almost sure convergence is with respect to the auxiliary randomization the forecaster has access to. This property rules out the possibility that the regret is much larger than its expected value with a significant probability.

The Borel-Cantelli lemma, together with the martingale techniques of Subsection 3.3.1, shows that the forecaster of Theorem 2.3 is Hannan consistent.

4. Multi-armed bandit prediction

In many prediction problems the forecaster, after forming a prediction, is able to measure his loss (or reward) but he does not have access to what would have happened had he chosen another possible prediction. This is especially important in game theory, when one is forced to play an unknown game. Such prediction problems have been known as *multi-armed bandit problems*. The name refers to a gambler who plays a pool of slot machines (called "one-armed bandit" in the U.S.). The gambler places his bet each time on a possibly different slot machine and his goal is to win almost as much as if he had known in advance which slot machine would have returned the maximal total reward.

This problem has been widely studied both in a stochastic and in a worst-case setting. The worst-case or adversarial setting considered in this thesis was first investigated by Baños [**Ban68**] (see also Megiddo [**Meg80**]). Hannan consistent strategies were constructed by Foster and Vohra [**FoVo98**], Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**], and Hart and Mas Colell [**HaMa00, HaMa02**] (see also Fudenberg and Levine [**FuLe98**]). We recall below the strategy considered in Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**] and their main result (see also Auer [**Aue02**]), but start first with a simple strategy which handles only expected regret.

In almost the whole thesis, we consider loss games, in which forecasters and experts suffer losses, and do not get payoffs. While this seems only a lexical difference, in some situations, there may be an asymmetry between loss games and gain games (see Chapter 4 for a more precise

definition and comparison of the two games). In the bandit setting, the asymmetry appears at two places. First, it is easy to design simple forecasters, i.e., forecasters using no shifting on the probability distributions as is the case in step (4) in Figure 4 and that have expected regrets of the order of $\sqrt{n}$ in bandit loss games. On the contrary, it seems that we may exhibit only in bandit gain games forecasting strategies whose (non-expected) regret is with overwhelming probability of the order of $\sqrt{n}$. But of course, if one is only interested in bounds depending on n and N, it is easy to reduce a gain game to a loss game, and vice versa (see Chapter 4), and thus, Theorem 2.7 below extends to gain games, whereas Theorem 2.8 is proved by a reduction to a gain game. Problems arise only when one wants bounds that are improvements for small losses, see Section 4.4 in Chapter 4 for more details.

For expected regret in loss games, we consider the incremental update (2.5) of the exponentially weighted average forecaster, run however on the estimated losses given by

$$\widetilde{\ell}_{i,t} = \frac{\ell_{i,t}}{p_{i,t}} Z_{i,t} \;, \qquad i = 1, \dots, N \text{ and } t = 1, 2, \dots \;,$$

where $Z_{i,t} = 1$ if $I_t = i$ and 0 otherwise. These are indeed estimators since we observe $\ell_{i,t}$ if and only if $I_t = i$. Note that these estimators are unbiased,

$$\mathbb{E}\left[\widetilde{\ell}_{i,t} \,|\, U_1^{t-1}\right] = \ell_{i,t} \;.$$

More precisely, the forecaster draws its prediction at round t at random according to the probability distribution $\boldsymbol{p}_t$ given, for $t \geqslant 2$, by

$$p_{i,t} = \frac{e^{\eta_t \widetilde{L}_{i,t-1}}}{\sum_{j=1}^N e^{\eta_t \widetilde{L}_{j,t-1}}} \;, \tag{2.7}$$

where for $t \geqslant 1$, $\widetilde{L}_{i,t} = \sum_{s=1}^t \widetilde{\ell}_{i,s}$ and $\eta_t = \sqrt{\dfrac{2 \ln N}{N t}}$.

THEOREM 2.7. *Assume the loss function is bounded in* $[0,1]$. *The expected regret of the above forecaster against any opponent player is bounded by*

$$\max_{j=1,\dots,N} \mathbb{E}\left[\sum_{t=1}^n \ell(f_{I_t,t}, y_t) - \sum_{t=1}^n \ell(f_{j,t}, y_t)\right] \leqslant 2\sqrt{2}\sqrt{(n+1) N \ln N} \;.$$

PROOF. We combine Lemmas 4.3 and 4.5 to get

$$\sum_{t=1}^n \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t} - \min_{j=1,\dots,N} \sum_{t=1}^n \widetilde{\ell}_{j,t} \leqslant \frac{2 \ln N}{\eta_{n+1}} + \frac{1}{2} \sum_{t=1}^n \eta_t \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t}^2 \;, \tag{2.8}$$

and use the facts that by the definition of the $\widetilde{\ell}_{i,t}$,

$$\sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t} = \ell(f_{I_t,t}, y_t) \qquad \text{and} \qquad \sum_{i=1}^N p_{i,t} \widetilde{\ell}_{i,t}^2 = \sum_{i=1}^N \frac{\ell_{i,t}^2}{p_{i,t}} Z_{i,t} \;.$$

The expectation of the second sum is less than N, and therefore, taking expectations in both sides of (2.8) (and using that the expectation of a maximum is more than the maximum of the expectations), we get

$$\max_{j=1,\dots,N} \mathbb{E}\left[\sum_{t=1}^n \ell(f_{I_t,t}, y_t) - \sum_{t=1}^n \ell(f_{j,t}, y_t)\right] \leqslant \frac{2 \ln N}{\eta_{n+1}} + \frac{N}{2} \sum_{t=1}^n \eta_t \;.$$

Substituting the proposed values for η_t concludes the proof. □

Algorithm EXP3.P
Parameters: Positive reals β, η, γ.
Initialization: $w_{i,0} = 1$ and $p_{i,1} = 1/N$ for $i = 1, \ldots, N$.
For each round $t = 1, 2, \ldots$

(1) choose expert I_t according to the probability distribution $\boldsymbol{p}_t$;
(2) calculate the estimated gains

$$\widetilde{g}_{i,t} = \frac{1 - \ell_{i,t}}{p_{i,t}} \mathbb{I}_{[I_t = i]} + \frac{\beta}{p_{i,t}} \; ;$$

(3) update the weights $w_{i,t} = w_{i,t-1} e^{\eta \widetilde{g}_{i,t}}$;
(4) calculate the updated probability distribution

$$p_{i,t+1} = (1 - \gamma) \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}} + \frac{\gamma}{N} \; , \qquad i = 1, \ldots, N \; .$$

FIGURE 4. Algorithm EXP3.P for prediction in a multi-armed bandit setting (first introduced in [**AuCeFrSc02**]).

For the forecaster defined in (2.7), little can be said concerning the deviations of the regret with respect to its average value, since we have no obvious bound on the estimated losses $\widetilde{\ell}_{i,t}$ (as we ignore how the $p_{i,t}$ behave). In particular, the techniques of Section 3.3 fail to lower bound with overwhelming probability $\sum_{t=1}^{n} \widetilde{\ell}_{j,t}$ by its (conditional) expectation $L_{j,n}$. This is why Auer, Cesa-Bianchi, Freund and Schapire [**AuCeFrSc02**] introduce a modified forecaster, whose (non-expected) regret at round n may be bounded by a quantity of the order of $O(\sqrt{nN \log(nN)})$. Their analysis was recently improved by Cesa-Bianchi and Lugosi [**CeLu05**] and is summarized in the following theorem.

THEOREM 2.8. *Assume the loss function is bounded in* $[0, 1]$. *For any* $\delta \in]0, 1[$, *for any* $n \geqslant (1/N) \ln(N/\delta)$, *if the forecaster of Figure 4 is run with parameters*

$$\beta = \sqrt{\frac{\ln(N/\delta)}{nN}} \; , \qquad \gamma = \frac{6N\beta}{4 + \beta} \; , \qquad \eta = \frac{\gamma}{3N} \; ,$$

then with probability $1 - \delta$, *its (non-expected) regret is bounded as*

$$\sum_{t=1}^{n} \ell(f_{I_t,t}, y_t) - \min_{j=1,\ldots,N} \sum_{t=1}^{n} \ell(f_{j,t}, y_t) \leqslant 6\sqrt{nN \ln(N/\delta)} + \frac{\ln N}{2} \; .$$

The above forecaster may be turned into an on-line algorithm that does not require previous knowledge of the time horizon n by applying the techniques of Section 2.2. Auer, Cesa-Bianchi, Freund and Schapire [**AuCeFrSc02**] also propose a general lower bound on the regret. (The simpler model of constant expert predictions considered in the theorem is discussed in detail in the appendix of this chapter.)

THEOREM 2.9. *There exist an outcome space* $\mathcal{Y}$ *and a loss function* $\ell : \mathbb{N} \times \mathcal{Y} \to [0, 1]$, *such that, for all* $N \geqslant 2$ *and for all* $n \geqslant 1$, *the cumulative (expected) regret of any (randomized) forecaster that gets constant expert predictions* $f_{j,t} \equiv j$ *for all* $j = 1, \ldots, N$ *and* $t = 1, 2, \ldots$, *while predicting a sequence of* n *outcomes in a bandit setting, satisfies the inequality*

$$\sup_{y_1,\ldots,y_n \in \mathcal{Y}} \left(\mathbb{E}\left[\sum_{t=1}^{n} \ell(I_t, y_t) \right] - \min_{i=1,\ldots,N} \sum_{t=1}^{n} \ell(i, y_t) \right) \geqslant \frac{1}{20} \min\left\{ \sqrt{nN}, \, n \right\} \; .$$

OPEN QUESTION 2.1. We note here that the minimax orders of magnitude for the regret in a bandit setting (see Section 5) are not completely known yet. Theorem 2.9 only indicates that the bound of Theorem 2.8 is optimal up to the logarithmic factor $\sqrt{\ln N}$. Though, in view of Section 5.3 below and Theorem 5.5, we conjecture that this factor is necessary, it is still an open question to prove a $\Omega(\sqrt{nN \ln N})$ lower bound on the (expected) regret in a bandit setting.

5. Minimax orders of magnitude for the regret

5.1. Formal definition of the minimax value. We described the problem of sequential prediction with expert advice in Figure 1 of Section 1. With the notation therein, a *prediction setting* is formed by a prediction space $\mathcal{X}$, an outcome space $\mathcal{Y}$, and a loss function $\ell : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. For given parameters $n,\ N \geqslant 1$, we define the minimax (expected) regret of a given prediction setting $(\mathcal{X}, \mathcal{Y}, \ell)$ with N experts and till round n by

$$\mathcal{V}^{(n,N)}_{(\mathcal{X},\mathcal{Y},\ell)} = \inf \sup \mathbb{E}\left[\sum_{t=1}^{n} \ell(\widehat{p}_t, y_t) - \min_{j=1,\ldots,N} \sum_{t=1}^{n} \ell(f_{j,t}, y_t)\right] , \tag{2.9}$$

where the supremum is over all possible (deterministic) strategies of the environment, and the infimum is over all possible (randomized) strategies of the forecaster. The expectation in the above expression is with respect to the auxiliary randomization the forecaster uses. Note that the strategy of the environment consists both in the choice of the experts' predictions and the outcomes. Since we take first the supremum, we may assume without loss of generality that the strategies of the environment are deterministic. Note that the minimax (expected) regret corresponds to the value of an n stage repeated zero-sum game.

We focus below on minimax expected regrets, but we could have defined a notion of minimax $1 - \delta$ non-expected regret, where we would have taken the smallest bound on the regret over all spaces of probability at least $1 - \delta$ with respect to the auxiliary randomization (see the comments after the statement of Theorem 5.5).

In the subsequent chapters, we often consider prediction settings where there are no experts, in the sense that the forecaster is only supplied with N constant experts and the environment simply chooses the outcomes. These settings correspond to $\mathcal{X} = \{1, \ldots, N\}$, and the experts are then called actions, see the appendix of this chapter for more details. Sometimes we even simplify further the model by considering oblivious environments, which do not take the forecaster's predictions into consideration and apparently choose the outcomes in advance, see Section 1.4. The minimax expected regret in prediction settings $(\{1, \ldots, N\}, \mathcal{Y}, \ell)$ with constant actions, till round n and against oblivious opponents takes the simple form

$$\mathcal{U}^{(n,N)}_{(Y,\ell)} = \inf \sup_{y_1^n \in \mathcal{Y}^n} \left(\sum_{t=1}^{n} \sum_{i=1}^{N} p_{i,t} \ell(i, y_t) - \min_{j=1,\ldots,N} \sum_{t=1}^{n} \ell(j, y_t)\right) ,$$

where the infimum is still taken over all feasible (randomized) strategies of the environment, and where we use the notation of Section 1.3.

In this thesis, we actually work with general loss functions, and are not interested in possible refinements for specific loss functions, like the square loss, the logarithmic loss or the 0–1 loss, see [**CeFrHaHeScWa97**], [**Lug01**], [**CeLu05**] for sharper minimax bounds with these losses. The only restriction we are ready to assume on the class of prediction settings is that they correspond to bounded loss functions, say in $[0, 1]$. We focus therefore on the minimax problem given by

$$V^{(n,N)} = \sup_{(\mathcal{X},\mathcal{Y}),\, \ell:\mathcal{X}\times\mathcal{Y}\to[0,1]} \mathcal{V}^{(n,N)}_{(\mathcal{X},\mathcal{Y},\ell)} \qquad \text{and} \qquad U^{(n,N)} = \sup_{\mathcal{Y},\, \ell:\mathcal{X}\times\mathcal{Y}\to[0,1]} \mathcal{U}^{(n,N)}_{(\mathcal{Y},\ell)} .$$

Note that by their definitions, $V^{(n,N)} \geqslant U^{(n,N)}$.

REMARK 2.3. We defined above the minimax problem by a $\sup\inf\sup$. This is the only way if we want to consider all feasible forecasters, since the definition of the latter relies on the underlying decision and outcome spaces. However, if, as indicated at the beginning of the appendix of this chapter, we restrict our attention to those forecasters whose update rules are based only on the experts' losses, then we may consider the minimax problem given by $\inf\sup\sup$, where the first infimum is restricted to the loss based forecasters, the first supremum is over all prediction settings with loss functions taking values in $[0,1]$, and the second one is over all strategies of the environment. (The interpretation is that the opponent player also chooses the game.) Since $\sup\inf\sup$ is less than $\inf\sup\sup$, the second problem is actually harder (we need to find a strategy which is good for all prediction settings). It however turns out that we solve *both* minimax problems in the next section.

5.2. Definition of a solution of the minimax problem. We are interested in the orders of magnitude in n and N of $V^{(n,N)}$ (and $U^{(n,N)}$). We say that the minimax rate in n and N is $\psi_{(n,N)}$, where $(\psi_{(n,N)})_{n\geqslant 1,\, N\geqslant 1}$ is a sequence of nonnegative numbers, whenever there exists two positive constants u, v such that for all n and N sufficiently large,

$$u\,\psi_{(n,N)} \leqslant V^{(n,N)} \leqslant v\,\psi_{(n,N)}\ .$$

We seek the simplest[5] possible expressions for ψ.

We say that a forecaster is a *solution of the minimax problem* if there exists a constant $c > 0$ such that its expected loss, for all possible prediction settings with bounded losses and against all strategies of the opponent, is less than $c\,\psi_{(n,N)}$, at least for n and N sufficiently large. (The forecaster's prediction rule is loss based, see Remark 2.3.)

The usual methodology is first to get upper bounds on the rate $\psi_{(n,N)}$ by exhibiting a general forecaster. For instance, Theorem 2.3 shows that in the model considered in this chapter, $\psi_{(n,N)} \leqslant \sqrt{n\ln N}$. Lower bounds on $\psi_{(n,N)}$ may be achieved by exhibiting a precise prediction setting $\mathcal{X}$, $\mathcal{Y}$, ℓ such that in this setting, all forecasters are bound to suffer an expected regret more than $a\psi_{(n,N)}$, where $a > 0$ is a universal constant. We often take (see, e.g., Theorem 5.5) $\mathcal{X} = \mathbb{N}$ and $\mathcal{Y} = [0,1]$, as well as oblivious opponents, define precisely ℓ depending on the new model, and restrict the forecaster to use only the N first constant actions. This way, we get a lower bound for $U^{(n,N)} \leqslant V^{(n,N)}$, which is enough for our purposes. We prove in the next section that in the model considered in this chapter, $\psi_{(n,N)} \geqslant \sqrt{n\ln N}$. This shows that the exponentially weighted average predictor of Theorem 2.3 is a solution of the minimax problem, hence its optimality in a minimax sense.

5.3. The optimality of the exponentially weighted majority predictor. In this section we prove the following main result.

THEOREM 2.10. *The minimax expected regret is asymptotically lower bounded as*

$$\lim_{N\to\infty}\lim_{n\to\infty}\frac{V^{(n,N)}}{\sqrt{(n/2)\ln N}} \geqslant 1\ .$$

This theorem shows in particular that $\psi_{(n,N)} \geqslant \sqrt{n\ln N}$, as claimed above, and indicates also that the leading factor in the bound of Theorem 2.1 is optimal. We thus have not only exhibited the minimax rates in n and N, but also the asymptotically optimal constant. (Compare to Question 5.2.)

[5]for instance, we prefer $\psi_{(n,N)} = \sqrt{n\ln N}$ to $\psi_{(n,N)} = (1+1/n)\sqrt{n\ln N}$

Usual techniques for computing lower bounds on the minimax values rely on inductive arguments and somewhat tedious exact computations, see, e.g., the computations of (lower bounds on) the minimax values in Chung [**Chu94**] (see also [**CeFrHaHeScWa97**]) who considers the model of this chapter, and Helmbold and Panizza [**HePa97**] for the model of Chapter 5, Mertens, Sorin and Zamir [**MeSoZa94**] for the model of Chapter 6. These techniques may lead however to improved leading constants in the lower bounds with respect to the methods we describe next.

Another solution is to consider oblivious opponents, and lower bound the supremum over all possible outcome sequences by a suitable randomization on the outcomes, as suggested by Auer, Cesa-Bianchi, Freund and Schapire [**AuCeFrSc02**] in a bandit setting, see also [**Lug01**] and [**CeLu05**].

We illustrate this in the model considered in this chapter by considering the prediction setting of on-line classification, where $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and $\ell(x,y) = \mathbb{I}_{[x \neq y]}$, and propose a proof inspired from [**CeFrHaHeScWa97**] and [**Lug01**].

PROOF. We assume the environment chooses both the experts' predictions $F_{j,t}$, $j = 1, \ldots, N$, $t = 1, \ldots, n$, and the outcomes Y_t, $t = 1, \ldots, n$, all independently at random, according to a common symmetric Bernoulli law. That is, we consider the set of all possible outcome sequences $(y_1, \ldots, y_n) \in \{0,1\}^n$ and all possible sequences of experts' advices $f_{j,t} \in \{0,1\}$, $j = 1, \ldots, N$, $t = 1, \ldots, n$, and put uniform probability weights on the elements of the set thus obtained. In particular, as a worst-case bound is worse than an average bound, we get

$$V^{(n,N)} \geqslant \inf \mathbb{E} \left[\sum_{t=1}^{n} \mathbb{I}_{[\widehat{x}_t \neq Y_t]} - \min_{j=1,\ldots,N} \sum_{t=1}^{n} \mathbb{I}_{[F_{j,t} \neq Y_t]} \right] ,$$

where the expectation is with respect to the uniform probability distribution on the experts' advices and on the outcomes, as well as to the forecasters' auxiliary randomization, whereas the infimum is over all possible forecasting strategies (and the $\widehat{x}_t$, $t = 1, \ldots, n$, denote the sequence of predictions formed by each of these possible forecasters).

For all forecasting strategies, since the Y_t are i.i.d. according to a Bernoulli distribution, $\mathbb{E}[\mathbb{I}_{[\widehat{x}_t \neq Y_t]}] = 1/2$. Since the $F_{j,t}$ are independent of the Y_t, all of them with common symmetric Bernoulli law, we may see, by conditioning, that

$$\min_{j=1,\ldots,N} \sum_{t=1}^{n} \mathbb{I}_{[F_{j,t} \neq Y_t]} \stackrel{(d)}{=} \min_{j=1,\ldots,N} \sum_{t=1}^{n} F_{j,t} ,$$

where the equality means equality of the distributions. Therefore, we have a simple lower bound in closed form for $V^{(n,N)}$,

$$V^{(n,N)} \geqslant \frac{n}{2} - \mathbb{E} \left[\min_{j=1,\ldots,N} \sum_{t=1}^{n} F_{j,t} \right] = \frac{1}{2} \mathbb{E} \left[\max_{j=1,\ldots,N} \sum_{t=1}^{n} \sigma_{j,t} \right] ,$$

where the $\sigma_{j,t}$, $j = 1, \ldots, N$, $t = 1, 2, \ldots$, are i.i.d. random variables distributed according to a symmetric Rademacher law, that is, they take the values -1 and 1 with equal probabilities $1/2$.

We may now open the toolbox of probability, and conclude the proof of Theorem 2.10 by using that

$$\lim_{N \to \infty} \lim_{n \to \infty} \frac{1}{\sqrt{n} \sqrt{2 \ln N}} \mathbb{E} \left[\max_{j=1,\ldots,N} \sum_{t=1}^{n} \sigma_{j,t} \right] \geqslant 1 ,$$

see [**CeFrHaHeScWa97**, Lemma 6], see also [**Lug01**]. □

Appendix: On the pertinence of the notion of regret for small decision spaces

In the subsequent chapters, we often concentrate on prediction settings of the form $\mathcal{X} = \{1, \dots, N\}$ and where there are no experts. We explain first here why this is not a serious restriction, and then discuss the meaning of the notion of regret in view of some recent criticisms.

The constant expert model. Since we seek forecasting algorithms that make assumptions neither on the structure of the outcome space $\mathcal{Y}$ nor on the structure of the loss function ℓ, only the values of the losses of the experts $\ell_{1,t}, \dots, \ell_{N,t}$ matter. All forecasters introduced below and in the subsequent chapters rely only on these losses. Thus for all models that do not need assumptions on the loss function to be dealt with we may concentrate on the sequences of losses rather than on the sequences of experts' advice and outcomes. This will be the case for all the models in this thesis except the one of Chapter 6 (see condition (6.1)), see also Remark 5.1. Note that this way, we may also think of loss functions that change with time or that depend on an external state of Nature, see also Remark 3.3.

Consequently, we henceforth consider a simpler setting where there are no experts, in the sense that the forecaster is supplied with N constant experts. For all $t = 1, 2, \dots$ and all $j = 1, \dots, N$, $f_{j,t} \equiv j$, where the prediction space is $\mathcal{X} = \{1, \dots, N\}$, up to a relabelling. The constant experts may be identified with (constant) actions. We are then interested in performing almost as well as the best of these constant actions, that is, the regret is defined as

$$R_n = \sum_{t=1}^{n} \ell(I_t, y_t) - \min_{j=1,\dots,N} \ell(j, y_t) \; .$$

Interpretation of the regret against constant actions. The only drawback of this reduction is the interpretation of the regret. Sometimes, it may be just unreasonable or meaningless to compare the forecaster's performance to the performance of a constant action, see for instance Remark 6.1, and some other times, the comparison is interesting, see, e.g., Example 6.1 or the problem of sequential investment in the stock market described in Chapter 7.

Another (more serious) criticism is in de Farias and Megiddo [**FaMe03**]. Recall that in the simplified model, we compare the decision-maker's cumulative loss to the smallest of the

$$L_{j,n} = \ell(j, y_1) + \dots + \ell(j, y_n) \; ,$$

where j ranges over $\{1, \dots, N\}$. [**FaMe03**] points out that if we had constantly played action j, then the outcome sequence $y_1, y_2, \dots$ would have been different too. Ideally, one would compare, with the notation of Section 1.4, to

$$L'_{j,n} = \sum_{t=1}^{n} \ell\left(j, g_t(j, \dots, j)\right) \; ,$$

but these quantities are not available to the forecaster. (Therefore, [**FaMe03**] proposes new measures of the feelings of regret.) Note that this criticism is essentially grounded only in the constant action model, and to answer it we propose an argument similar to the one used later in Section 2 of Chapter 7. There, we compare the forecaster to the class of the optimal strategies for i.i.d. (or stationary) markets, and the latter is formed by the so-called constantly rebalanced portfolios. In the same way, we note that in the setting of prediction with expert advice, if the outcomes were realizations of an i.i.d. (or stationary) sequence of random variables, then the optimal strategy would be given by playing constantly one of the N possible actions. Of course, we avoid such stochastic assumptions, but interestingly enough it seems that the way we measure the regret is one of the last tracks of these widely used stochastic models.

CHAPTER 3

Internal regret in prediction with expert advice

In this chapter, we study internal regret in prediction with expert advice. The notion of internal regret plays a key role in game theory and is concerned with consistent modifications of our forecasting strategy. We show a general conversion trick to derive no-internal-regret forecasters from no-external-regret ones. This trick is also illustrated on the multi-armed bandit problem, and is extended to deal with a generalization of internal regret known as swap regret. We discuss the optimality and the pertinence of the introduced strategies by stating lower bounds on the internal regret in two prediction settings, one of them is prediction with expert advice and bounded losses, the other one is sequential probability assignment.

Contents

Though this chapter is partially based on Section 3 of [**StLu05**], most of its material is published here for the first time.

We recall in the introductory chapter that in the on-line prediction problem, the goal is to minimize the predictor's cumulative loss with respect to the best cumulative loss in a pool of experts. In a certain equivalent game-theoretic formulation of the problem, this is the same as minimizing the predictor's *external regret*, see [**FoVo99**]. External regret measures the difference between the predictor's cumulative loss and that of the best expert. However, another notion of regret, called *internal regret* in [**FoVo99**] has also been in the focus of attention mostly in the theory of playing repeated games, see [**FoVo98, FoVo99**], [**FuLe99**], [**HaMa00, HaMa01**], [**CeLu03**]. (Internal regret is often referred to as conditional regret in the game-theory community, see [**Har04**].) Roughly speaking, a predictor has a small internal regret if for each pair of experts (i, j), the predictor does not regret of not having followed expert i each time it followed expert j. It is easy to see that requiring a small internal regret is a more difficult problem since a small internal regret in the prediction problem implies small external regret as well. In this chapter, we first define precisely the notions of internal and swap regrets, introduce a general conversion trick

to design internal regret minimizing forecasters, and finish by stating lower bounds on internal regret in two prediction settings, one of them is prediction with expert advice and bounded losses, the other one is sequential probability assignment.

1. Links between external and internal regret

1.1. Definition of internal regret. We use the notation of Chapter 2. Internal regret is concerned with consistent modifications of a given forecasting strategy. Each of these possible modifications is parameterized by a departure function $\Phi : \{1, \dots, N\} \to \{1, \dots, N\}$. For *internal regret*, we restrict our attention to functions Φ that only differ from identity in one point. That is, we only consider functions for which there exists a pair $i \neq j$ such that $\Phi(i) = j$, and $\Phi(k) = k$ for all $k \neq i$.

After round n, the cumulative loss of the forecaster is compared to the cumulative loss that would have been accumulated had the forecaster chosen expert $\Phi(I_t)$ instead of expert I_t at all rounds t, $t = 1, \dots, n$. That is, for a given pair (i, j), one is interested in modifications of the predictor's strategy obtained by replacing the action of the forecaster by expert j each time it chooses expert i. If no such consistent modification results in a much smaller accumulated loss, then the strategy is said to have small internal regret (or no internal regret). Formally, we seek strategies achieving

$$\frac{1}{n} \sum_{t=1}^{n} \ell\left(f_{I_t,t}, y_t\right) - \frac{1}{n} \min_{\Phi} \sum_{t=1}^{n} \ell\left(f_{\Phi(I_t),t}, y_t\right) = o(1) \quad \text{a.s.} ,$$

where the minimization is over all functions Φ that only differ from identity in one point and $f_{j,t}$ denotes the prediction of expert j at round t. Such strategies are said Hannan consistent with respect to internal regret.

The notion of internal regret has been shown to be useful in the theory of equilibria of repeated games. Foster and Vohra [**FoVo97, FoVo99**] showed that if all players of a finite game choose a strategy that is Hannan consistent with respect to internal regret, then the joint empirical frequencies of play converge to the set of correlated equilibria of the game (see also Fudenberg and Levine [**FuLe95**], Hart and Mas-Colell [**HaMa00**]; see also the more general results of Chapter 8).

Now, to get Hannan consistency with respect to internal regret, it is enough to control uniformly the expectation of the internal regret with respect to the auxiliary randomization the forecasters uses. Martingales inequalities combined with the Borel-Cantelli lemma then show the desired Hannan consistency, just as this was the case for the (external) regret in Section 3.3 of Chapter 2. Therefore we concentrate below on expected internal regret.

Recall that the definition of external regret is based on the comparison to an external pool of strategies, the ones given by each expert, and that in the definition of the (expected) internal regret one is interested in modifications of the predictor's strategy obtained by replacing the action of the forecaster by expert j each time it chooses expert i. Because we work in expectation, this is equivalent to selecting an expert according to the distribution $\boldsymbol{p}_t^{i \to j}$ obtained from $\boldsymbol{p}_t$ by putting probability mass 0 on i and $p_{i,t} + p_{j,t}$ on j. This transformation is called the $i \to j$ *modified strategy*. Recall also that we require that none of these modified strategies is much better than the original strategy, that is, we seek strategies such that the difference between their (expected) cumulative loss and that of the best modified strategy is small. Thus,

$$\sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) - \min_{i,j \in \{1,\dots,N\}} \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t^{i \to j}) ,$$

Regimes	$\ell_{A,t}$	$\ell_{B,t}$	$\ell_{C,t}$
$1 \leqslant t \leqslant n/3$	0	1	5
$n/3+1 \leqslant t \leqslant 2n/3$	1	0	5
$2n/3+1 \leqslant t \leqslant n$	2	1	0

TABLE 1. The losses for Example 3.1.

where for all probability distributions $\boldsymbol{q} = (q_1, \ldots, q_N)$,

$$\ell_t(\boldsymbol{q}) = \sum_{k=1}^{N} q_k \ell_{k,t} = \sum_{k=1}^{N} q_k \ell(f_{k,t}, y_t) \ ,$$

should be as small as possible. This quantity is the *(expected) internal regret* of the forecaster. The internal regret may be re-written as

$$\max_{i,j \in \{1,\ldots,N\}} \sum_{t=1}^{n} r_{(i,j),t}$$

where $r_{(i,j),t} = p_{i,t}(\ell_{i,t} - \ell_{j,t})$. Thus, $r_{(i,j),t}$ expresses the predictor's regret of having put the probability mass $p_{i,t}$ on the i-th expert instead of on the j-th one, and

$$R_{(i,j),n} = \sum_{t=1}^{n} r_{(i,j),t} = \sum_{t=1}^{n} p_{i,t}(\ell_{i,t} - \ell_{j,t})$$

is the corresponding cumulative regret. Similarly to the case of the external regret, if this quantity is uniformly $o(n)$ over all possible values of the losses, then the corresponding predictor is said to exhibit no (expected) internal regret.

Now clearly, the external regret of the predictor equals

$$\max_{j=1,\ldots,N} \sum_{i=1}^{N} R_{(i,j),n} \leqslant N \max_{i,j \in \{1,\ldots,N\}} R_{(i,j),n} \ , \tag{3.1}$$

which shows that any algorithm with a small (i.e., sublinear in n) (expected) internal regret also has a small (expected) external regret. (And the same can be said to upper bound non-expected external regret by non-expected internal regret.) On the other hand, it is easy to see that a small external regret does not imply small internal regret. In fact, as it is shown in the next example, even the exponentially weighted average algorithm defined above may have a linearly growing internal regret.

EXAMPLE 3.1. *(Weighted average predictor has a large internal regret.)* Consider the following example with three experts, A, B, and C. Let n be a large multiple of 3 and assume that time is divided in three equally long regimes, characterized by a constant loss for each expert. These losses are summarized in Table 1. We claim that the regret $R_{(B,C),n}$ of B versus C grows linearly with n, that is,

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} p_{B,t} \left(\ell_{B,t} - \ell_{C,t}\right) = \gamma > 0 \ ,$$

where

$$p_{B,t} = \frac{e^{-\eta L_{B,t}}}{e^{-\eta L_{A,t}} + e^{-\eta L_{B,t}} + e^{-\eta L_{C,t}}}$$

denotes the weight assigned by the exponentially weighted average predictor to expert B, where $L_{i,t} = \sum_{s=1}^{t} \ell_{i,s}$ denotes the cumulative loss of expert i and η is chosen to minimize the external regret, that is, $\eta = (1/5)\sqrt{(8\ln 3)/n} = 1/(K\sqrt{n})$ with $K = 5/\sqrt{8\ln 3}$, see Theorem 2.1. (Note that the same argument leads to a similar lower bound for $\eta = \gamma/\sqrt{n}$, where $\gamma > 0$ is any constant.) The intuition behind this example is that at the end of the second regime the predictor quickly switches from A to B, and the weight of expert C can never recover because of its disastrous behavior in the first two regimes. But since expert C behaves much better than B in the third regime, the weighted average predictor will regret of not having followed the advice of C each time it followed B.

More precisely, we show that during the first two regimes, the number of times when $p_{B,t}$ is more than ε is of the order of $\sqrt{n}$ and that, in the third regime, $p_{B,t}$ is always more than a fixed constant ($1/3$, say). This is illustrated in Figure 1. In the first regime, a sufficient condition for

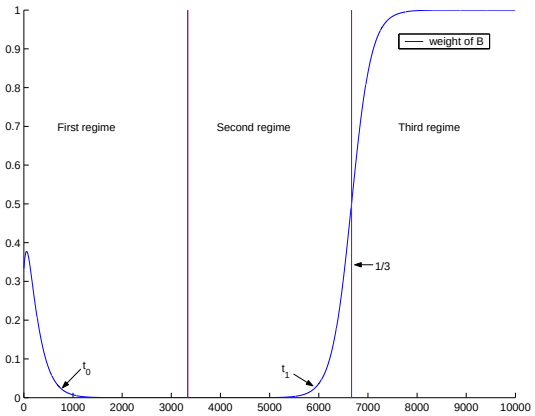

FIGURE 1. The evolution of the weight assigned to B in Example 3.1 for $n = 10000$.

$p_{B,t} \leqslant \varepsilon$ is that $e^{-\eta L_{B,t}} \leqslant \varepsilon$. This occurs whenever $t \geqslant t_0 = K\,(-\ln \varepsilon)\,\sqrt{n}$. For the second regime, we lower bound the time instant t_1 when $p_{B,t}$ gets larger than ε. To this end, note that $p_{B,t} \geqslant \varepsilon$ implies

$$(1-\varepsilon)e^{-\eta L_{B,t}} \geqslant \varepsilon \left(e^{-\eta L_{A,t}} + e^{-\eta L_{C,t}}\right) \geqslant \varepsilon e^{-\eta L_{A,t}} \ ,$$

which leads to $t_1 \geqslant \frac{2n}{3} + K\left(\ln \frac{\varepsilon}{1-\varepsilon}\right)\sqrt{n}$. Finally, in the third regime, we have at each time instant $L_{B,t} \leqslant L_{A,t}$ and $L_{B,t} \leqslant L_{C,t}$, so that $p_{B,t} \geqslant 1/3$. Putting these three steps together, we obtain the following lower bound for the internal regret of B versus C:

$$\sum_{t=1}^{n} p_{B,t}\,(\ell_{B,t} - \ell_{C,t}) \geqslant \frac{n}{9} - 5\left(\frac{2n}{3}\varepsilon + K\left(\ln \frac{1-\varepsilon}{\varepsilon^2}\right)\sqrt{n}\right) \ ,$$

which is of the order n, for a sufficiently small $\varepsilon > 0$.

1.2. A general way to design internal regret minimizing algorithms. The example above shows that special algorithms need to be designed to guarantee a small internal regret. Indeed, such predictors exist, as was shown by [**FoVo98**], see also [**FuLe99**], [**HaMa00, HaMa01**]. Here we briefly give a new insight on predictors studied in [**CeLu03**] (see the remark at the end of this section), and based on [**HaMa01**], as well as a new, simple analysis of their performance guarantees.

Consider the case of sequential prediction under expert advice, with N experts and losses bounded between 0 and B. We describe now a simple way of converting any no external regret

forecaster into a no internal regret forecaster, that is, we show how a Hannan consistent forecaster can be turned into a Hannan consistent forecaster with respect to internal regret. (Thus, in this precise sense, we can say, despite of Example 3.1, that small external regret implies small internal regret.) Such a conversion method may be defined recursively as follows.

At time $t = 1$, let $\boldsymbol{p}_1 = (1/N, \ldots, 1/N)$ be the uniform distribution over the N actions. At round $t > 2$, the forecaster has already chosen and predicted according to the probability distributions $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{t-1}$. We define $N(N-1)$ fictitious experts, indexed by pairs of integers $i \neq j$, by their losses at time instants $1 \leqslant s \leqslant t-1$, which equal $\ell_s(\boldsymbol{p}_s^{i \to j})$, where we re-used the notation of the previous section.

Define now a probability distribution Δ_t over the pairs $i \neq j$ by running one of the algorithms of Section 2 of Chapter 2, on this pool of fictitious experts, and choose $\boldsymbol{p}_t$ such that the fixed point equality

$$\boldsymbol{p}_t = \sum_{(i,j):i \neq j} \Delta_{(i,j),t} \boldsymbol{p}_t^{i \to j} \,, \tag{3.2}$$

holds. (We say that Δ_t *induces* $\boldsymbol{p}_t$.) The existence and the practical computation of such a $\boldsymbol{p}_t$ is an application of Lemma 3.1 below.

For instance, $\Delta_t = (\Delta_{(i,j),t})_{i \neq j}$ may be given by

$$\Delta_{(i,j),t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(\boldsymbol{p}_s^{i \to j})\right)}{\sum_{(k,l):k \neq l} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(\boldsymbol{p}_s^{k \to l})\right)} \,,$$

tuned, as suggested by the theory, with $\eta = 4B^{-1}\sqrt{\ln N / n}$ in case of known time horizon n.

Indeed, this choice of η and the application of the bound proposed by Theorem 2.1 (with $N(N-1)$ upper bounded by N^2) lead to

$$\sum_{t=1}^{n} \sum_{i \neq j} \Delta_{(i,j),t} \ell_t(\boldsymbol{p}_t^{i \to j}) \leqslant \min_{i \neq j} \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t^{i \to j}) + B\sqrt{n \ln N} \,,$$

that is, recalling the fixed point equality (3.2), the cumulative internal regret of the above strategy is bounded by

$$\max_{i \neq j} R_{(i,j),n} \leqslant B\sqrt{n \ln N} \,.$$

Note that this improves the bound given in Corollary 8 of [**CeLu03**], by a factor of two.

The same analysis can be carried over for the polynomial forecasters or the time-adaptive version of the exponentially weighted forecaster, using Theorems 2.3 and 2.4, and is summarized in the following theorem.

THEOREM 3.1. *The above exponentially weighted predictor achieves, uniformly over all possible values of the losses $\ell_{i,t} \in [0, B]$,*

$$\max_{i \neq j} R_{(i,j),n} \leqslant B\sqrt{n \ln N}.$$

With a time-adaptive tuning parameter the upper bound becomes

$$\max_{i \neq j} R_{(i,j),n} \leqslant B\left(2\sqrt{n \ln N} + \frac{\sqrt{\ln N}}{2}\right) \,.$$

Finally, with a polynomial predictor of order $p > 1$,

$$\max_{i \neq j} R_{(i,j),n} \leqslant B\sqrt{(p-1) n N^{4/p}} \,.$$

REMARK 3.1. The conversion trick illustrated above is a general trick which extends to any weighted average predictor, that is, to any predictor which, at each round, maintains one weight per expert. More precisely, any weighted average predictor whose external regret is small may be converted into a strategy whose internal regret remains small. This will be illustrated extensively in Chapter 7, first for convex loss functions in Sections 4.1 and 6.2 (see also Chapter 8), for exp-concave ones in Sections 5.1 and 5.2, and even for a function that is simply continuous, in Section 7. (See also the summary in Table 3 of the cited chapter.)

Note that in the case of randomized prediction under expert advice [**BlMa05**] propose a different conversion trick, with about the same algorithmic complexity, see the next two sections below. Such tricks are valuable to extend results in an effortless way from the case of external to internal regret, like the time-adaptive exponentially weighted average predictor suited for the minimization of internal regret proposed by Theorem 3.1, or the analysis of Section 6 in Chapter 7.

It only remains to see the existence and the way to compute a fixed point of the equality (3.2). The following lemma proposes a more general result, needed for subsequent analysis in Section 5.1. The meaning of this result is that each probability distribution over the expert pairs induces naturally a probability distribution over the experts.

LEMMA 3.1. *Let $\boldsymbol{q}$ be a probability distribution over the N experts. For all probability distributions Δ over the pairs of different experts $i \neq j$ and $\alpha \in [0,1]$, there exists a probability distribution $\boldsymbol{p}$ over the experts such that*

$$\boldsymbol{p} = (1-\alpha) \sum_{i \neq j} \Delta_{(i,j)} \boldsymbol{p}^{i \to j} + \alpha \boldsymbol{q} \; .$$

Moreover, $\boldsymbol{p}$ may be easily computed by a Gaussian elimination over a simple $N \times N$ matrix.

PROOF. The equality

$$\boldsymbol{p} = (1-\alpha) \sum_{i \neq j} \Delta_{(i,j)} \boldsymbol{p}^{i \to j} + \alpha \boldsymbol{q}$$

means that for all $m \in \{1, \ldots, N\}$,

$$p_m = (1-\alpha) \sum_{i \neq j} \Delta_{(i,j)} p_m^{i \to j} + \alpha q_m \left(\sum_{j=1}^{N} p_j \right) \; ,$$

or equivalently,

$$\left(\alpha \left(1 - q_m\right) + (1-\alpha) \sum_{j \neq m} \Delta_{(m,j)} \right) p_m = \sum_{i \neq m} \left((1-\alpha) \Delta_{(i,m)} + \alpha q_m \right) p_i \; ,$$

that is, $\boldsymbol{p}$ is an element of the kernel of the matrix A defined by

- if $i \neq m$, $A_{m,i} = w_{m,i}$,
- $A_{m,m} = - \sum_{j \neq m, \, 1 \leqslant j \leqslant N} w_{j,m}$,

where, for $i \neq m$,

$$w_{m,i} = (1-\alpha) \Delta_{(i,m)} + \alpha q_m \; .$$

The elements of A have a modulus less than 1. An element of the kernel of A is a fixed point of the matrix $S = A + I_N$, where I_N is the $N \times N$ identity matrix. But S is a column stochastic matrix (its columns are probability distributions), and thus admits a probability distribution $\boldsymbol{p}$ as a fixed point.

[**FoVo99**] suggest a Gaussian elimination method over A for the practical computation of $\boldsymbol{p}$. □

REMARK 3.2. [**CeLu03**] show that, writing $\mathbf{r}_t$ for the $N(N-1)$-vector with components $r_{(i,j),t}$ and $\mathbf{R}_t = \sum_{s=1}^t \mathbf{r}_s$, any predictor satisfying the so-called "Blackwell condition"

$$\nabla\Phi(\mathbf{R}_{t-1}) \cdot \mathbf{r}_t \leqslant 0 \tag{3.3}$$

for all $t \geqslant 1$, with Φ being either an exponential potential

$$\Phi(\mathbf{u}) = \sum_{i=1}^N \exp\left(\eta u_i\right) \ ,$$

with η possibly depending on t (when time-adaptive versions are considered) or a polynomial potential

$$\Phi(\mathbf{u}) = \sum_{i=1}^N \left(u_i\right)_+^p \ ,$$

has the performance guarantees given by Theorem 3.1, see also Section 3.2 of Chapter 8.

But the choice (3.2) ensures that the Blackwell condition is satisfied with an equality, as

$$\begin{aligned}
&\nabla\Phi(\mathbf{R}_{t-1}) \cdot \mathbf{r}_t \\
&\quad = \quad \sum_{i=1}^N \ell_{i,t} \left(\sum_{j=1,\dots,N,\ j\neq i} \nabla_{(i,j)}\Phi(\mathbf{R}_{t-1}) p_{i,t} - \sum_{j=1,\dots,N,\ j\neq i} \nabla_{(j,i)}\Phi(\mathbf{R}_{t-1}) p_{j,t} \right)
\end{aligned}$$

(see, e.g., [**CeLu03**] for the details), which equals 0 as soon as

$$\sum_{j=1,\dots,N,\ j\neq i} \nabla_{(i,j)}\Phi(\mathbf{R}_{t-1}) p_{i,t} - \sum_{j=1,\dots,N,\ j\neq i} \nabla_{(j,i)}\Phi(\mathbf{R}_{t-1}) p_{j,t} = 0$$

for all $i = 1, \dots, N$. The latter set of equations may be seen to be equivalent to (3.2), with the choice

$$\Delta_{(i,j),t} = \frac{\nabla_{(i,j)}\Phi(\mathbf{R}_{t-1})}{\sum_{k\neq l} \nabla_{(k,l)}\Phi(\mathbf{R}_{t-1})} \ ,$$

which was indeed the probability distribution proposed by the conversion trick introduced at the beginning of this section.

1.3. Swap regret and wide range regret. In this section, we essentially discuss and compare the conversion trick exposed in the previous section and the one proposed by Blum and Mansour [**BlMa05**]. We do this by introducing a generalization of internal regret known as swap regret.

Section 1.1 was concerned with consistent modifications of forecasting strategies parameterized by departure functions Φ that only differ from identity in one point. We now consider all possible departure functions $\Phi : \{1, \dots, N\} \to \{1, \dots, N\}$. Formally, the *swap regret* of a forecasting strategy is defined as

$$\sum_{t=1}^n \ell\left(f_{I_t,t}, y_t\right) - \min_{\Phi} \sum_{t=1}^n \ell\left(f_{\Phi(I_t),t}, y_t\right) \ ,$$

where the minimization is over all functions $\Phi : \{1, \dots, N\} \to \{1, \dots, N\}$ and $f_{j,t}$ denotes the prediction of expert j at round t.

As explained above, we may work here with expected quantities. It is easy to see that considering all departure functions Φ amounts to considering all linear departures $\varphi : \boldsymbol{p} \mapsto \varphi(\boldsymbol{p})$. The mappings $\boldsymbol{p} \mapsto \boldsymbol{p}^{i\to j}$ introduced in Section 1.1 are special cases of such linear departures. The (expected) swap regret of a forecasting strategy is then defined as

$$\sum_{t=1}^n \ell_t(\boldsymbol{p}_t) - \min_{\varphi} \sum_{t=1}^n \ell_t\left(\varphi(\boldsymbol{p}_t)\right) \ ,$$

where the minimization is over all linear mappings φ from the simplex of order N, denoted by $\mathcal{X}$, into itself. Such linear modifications were already considered by [**GrJa03**] and [**BlMa05**].

Now, by Krein-Millman theorem (see, e.g., Berger [**Ber90**]), the set of all linear mappings from the simplex into itself is the convex hull of the set of all extremal linear mappings. The latter are given by the φ associated to the Φ, and there are therefore N^2 of them. They simply transport all probability masses from each expert to another. We may apply the conversion trick of the previous section to this set of N^N fictitious experts, and get a procedure whose swap regret is bounded by a quantity of the order of $\sqrt{n \ln N^N} = \sqrt{nN \ln N}$. However, the resulting procedure has a computational complexity of the order of N^N, at least in its straightforward implementation, simply because we have to compute the losses of N^N fictitious experts. (Given the matrix with the weights computed thanks to the losses of the N^N experts, the Gaussian elimination further needed in the procedure, see Lemma 3.1, has only a computational complexity of the order of N^2.) One way around is to note that swap regret is bounded by N times internal regret, and thus, the practical forecasting scheme of the previous section (with computational complexity of the order of N^2) guarantees a bound on its swap regret of the worse order of $N\sqrt{n \ln N}$.

On the other hand, Blum and Mansour's [**BlMa05**] procedure yields a $O(\sqrt{nN \ln N})$ bound on swap regret, with a computational complexity only of the order of N^2. The only drawback of their conversion trick is that it only deals with linear loss functions. We recall that loss functions in prediction with expert advice are linear in some sense, because we consider expected losses, and these are linear in the probability distributions we use. Therefore the conversion of [**BlMa05**] does not extend to general convex losses, contrary to the one we introduced (see Remark 3.1).

REMARK 3.3. (*Wide-range regret.*) We close this section by noting, with [**BlMa05**], that the departure functions Φ could depend not only on the forecaster's played actions, but also on some side-information, such as the history, an activation function indicating which experts are asleep and which experts are active, and so on. Doing so, we would consider a finite number of those functions Lehrer [**Leh03**] uses in its definition of wide-range regret. As these are in countable number, it is easy to see that our procedures can be extended to no wide-range regret procedures thanks to classical adaptive methods like the doubling trick (see Section 2.2 in Chapter 2).

1.4. The case of limited feedback. In this section, we continue the comparison between the two conversion tricks of Remark 3.1, now from the viewpoint of prediction with limited feedback. Theorem 3.2 below shows that for multi-armed bandit prediction (see Section 4 in Chapter 2), our conversion trick yields a $O(\sqrt{nN \ln N})$ upper bound on the internal regret of a prediction scheme based on the one proposed by Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**]. Using the same methodology, it is easy to design a bandit forecaster ensuring a $O(N\sqrt{n \ln N})$ upper bound on the swap regret of a prediction scheme based on the one of [**AuCeFrSc02**]. In comparison, Blum and Mansour [**BlMa05**] can only get a $O(N\sqrt{nN \ln N})$ upper bound on swap regret. In addition, they need a very precise assumption on the no external regret algorithm to be converted (see Lemma 14 therein), and it seems that their conversion only works for a restricted class of Hannan consistent forecasting schemes. We also note that similar results were obtained by Hart and Mas-Colell [**HaMa02**].

In bandit problems, one usually estimates losses, forms weighted prediction with these estimated losses, and shifts the obtained probability distribution, so that all its components are more than a given threshold. Here, we use the conversion trick to form the weighted prediction, and then apply similarly a shifting method. To this end, we consider the setting and notation of Section 4 of Chapter 2, that is, a loss multi-armed bandit prediction game, with N experts (or actions), and losses bounded, say, in $[0, 1]$. We recall first a popular estimate of the losses in multi-armed bandit

problems. With the notation therein, we choose the unbiased estimates given by

$$\widetilde{\ell}(i,y_t) = \frac{\ell(i,y_t)}{p_{i,t}} \mathbb{I}_{[I_t=i]} \,, \tag{3.4}$$

for all $t = 1, 2, \ldots$ and $i = 1, \ldots, N$.

The forecaster we propose to minimize internal regret is formed by a sub-algorithm and a master algorithm. The parameters η_t and γ_t used below are tuned as

$$\eta_t = \alpha \sqrt{\frac{\ln N}{Nt}} \,, \quad \text{with } \alpha = \sqrt{2/3}, \qquad \text{and} \qquad \gamma_t = N\eta_t \,, \tag{3.5}$$

for $t = 1, 2, \ldots$. At each round t the sub-algorithm outputs a probability distribution

$$\boldsymbol{u}_t = \left(u_t^{i \to j} \right)_{(i,j)\,:\, i \neq j}$$

over the set of pairs of different actions; with the help of $\boldsymbol{u}_t$ the master algorithm computes a probability distribution $\boldsymbol{p}_t$ over the actions.

Consider the loss estimates $\widetilde{\ell}(i, y_t)$ defined in (3.4). For a given distribution $\boldsymbol{p}$ over $\{1, \ldots, N\}$, denote

$$\widetilde{\ell}(\boldsymbol{p}, y) = \sum_{k=1}^{N} p_k \, \widetilde{\ell}(k, y) \,.$$

Now introduce the cumulative estimated losses

$$\widetilde{L}_{t-1}^{i \to j} = \sum_{s=1}^{t-1} \widetilde{\ell}(\boldsymbol{p}_s^{i \to j}, y_s)$$

where $\boldsymbol{p}_s^{i \to j}$ denotes as above the probability distribution obtained from $\boldsymbol{p}_s$ by moving the probability mass $p_{i,s}$ from i to j; that is, we set $p_{s,i}^{i \to j} = 0$ and $p_{s,j}^{i \to j} = p_{s,j} + p_{s,i}$. The distribution $\boldsymbol{u}_t$ computed by the sub-algorithm is an exponentially weighted average associated to the cumulative losses $\widetilde{L}_{t-1}^{i \to j}$, that is,

$$u_t^{i \to j} = \frac{\exp\left(-\eta_t \widetilde{L}_{t-1}^{i \to j} \right)}{\sum_{k \neq l} \exp\left(-\eta_t \widetilde{L}_{t-1}^{k \to l} \right)} \,.$$

Now let $\widetilde{\boldsymbol{p}}_t$ be the probability distribution over the set of actions defined by the equation

$$\sum_{(i,j)\,:\, i \neq j} u_t^{i \to j} \widetilde{\boldsymbol{p}}_t^{i \to j} = \widetilde{\boldsymbol{p}}_t \,. \tag{3.6}$$

Such a distribution exists, and can be computed by a simple Gaussian elimination (see Lemma 3.1). The master algorithm then chooses, at round t, the action I_t drawn according to the probability distribution

$$\boldsymbol{p}_t = (1 - \gamma_t) \widetilde{\boldsymbol{p}}_t + \frac{\gamma_t}{N} \mathbf{1} \tag{3.7}$$

where $\mathbf{1} = (1, \ldots, 1)$.

To bound internal regret with overwhelming probability, we need the martingale inequalities (and the translation over the estimated losses) of [**AuCeFrSc02**], see also Theorem 2.8. We do not work out the straightforward details, and simply propose the following theorem.

THEOREM 3.2. *In a bandit setting with losses bounded between 0 and 1, the expected internal regret of the above forecasting scheme is bounded as*

$$\max_{i \neq j} \mathbb{E} \left[\sum_{t=1}^{n} p_{i,t} \left(\ell(i, y_t) - \ell(j, y_t) \right) \right] \leqslant 10 \sqrt{(n+1) N \ln N} \,.$$

PROOF. For a given t, the estimated losses $\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)$, $i \neq j$, fall in the interval $[0,\ N/\gamma_t]$. Since γ_t and η_t are tuned as in (3.5), $N\eta_t/\gamma_t \leqslant 1$, and we may apply Lemmas 4.3 and 4.5 below to derive

$$\sum_{t=1}^n \sum_{i\neq j} u_t^{i\to j} \widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t) - \min_{i\neq j} \sum_{t=1}^n \widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t) \leqslant \frac{2\ln N(N-1)}{\eta_{n+1}} + \sum_{t=1}^n \eta_t \sum_{i\neq j} u_t^{i\to j} \left(\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)\right)^2 . \tag{3.8}$$

For $i \neq j$, $\mathbf{1}^{i\to j}$ is the vector $\boldsymbol{v}$ such that $v_i = 0$, $v_j = 2$, and $v_k = 1$ for all $k \neq i$ and $k \neq j$. Use first (3.7) and then (3.6) to rewrite the first term of the left-hand side of (3.8) as

$$\begin{aligned}
\sum_{t=1}^n \sum_{i\neq j} u_t^{i\to j} \widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t) &= \sum_{t=1}^n \sum_{i\neq j} u_t^{i\to j} \left((1-\gamma_t)\widetilde{\ell}(\widetilde{\boldsymbol{p}}_t^{i\to j}, y_t) + \frac{\gamma_t}{N}\widetilde{\ell}(\mathbf{1}^{i\to j}, y_t)\right) \\
&= \sum_{t=1}^n (1-\gamma_t)\widetilde{\ell}(\widetilde{\boldsymbol{p}}_t, y_t) + \sum_{t=1}^n \frac{\gamma_t}{N} \sum_{i\neq j} u_t^{i\to j} \widetilde{\ell}(\mathbf{1}^{i\to j}, y_t) \\
&= \sum_{t=1}^n \widetilde{\ell}(\boldsymbol{p}_t, y_t) + \sum_{t=1}^n \frac{\gamma_t}{N} \sum_{i\neq j} u_t^{i\to j} \left(\widetilde{\ell}(\mathbf{1}^{i\to j}, y_t) - \widetilde{\ell}(\mathbf{1}, y_t)\right) \\
&= \sum_{t=1}^n \widetilde{\ell}(\boldsymbol{p}_t, y_t) + \sum_{t=1}^n \frac{\gamma_t}{N} \sum_{i\neq j} u_t^{i\to j} \left(\widetilde{\ell}(j, y_t) - \widetilde{\ell}(i, y_t)\right) .
\end{aligned}$$

Substituting into (3.8), we have

$$\max_{i\neq j} \sum_{t=1}^n p_{i,t} \left(\widetilde{\ell}(i, y_t) - \widetilde{\ell}(j, y_t)\right) \tag{3.9}$$

$$\begin{aligned}
&= \sum_{t=1}^n \widetilde{\ell}(\boldsymbol{p}_t, y_t) - \min_{i\neq j} \sum_{t=1}^n \widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t) \qquad (3.10) \\
&\leqslant \frac{4\ln N}{\eta_{n+1}} + \sum_{t=1}^n \eta_t \sum_{i\neq j} u_t^{i\to j} \left(\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)\right)^2 + \sum_{t=1}^n \frac{\gamma_t}{N} \sum_{i\neq j} u_t^{i\to j} \left(\widetilde{\ell}(i, y_t) - \widetilde{\ell}(j, y_t)\right) .
\end{aligned}$$

The crux of the proof is to handle the second sum. This is done by using the precise form (3.4) of the estimates, as well as the boundedness of ℓ in $[0, 1]$,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i\neq j} u_t^{i\to j} \left(\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)\right)^2\right] = \sum_{i\neq j} u_t^{i\to j} \sum_{k=1}^N p_{k,t} \left(p_{k,t}^{i\to j} \frac{\ell(k, y_t)}{p_{k,t}}\right)^2 \\
&\leqslant \sum_{i\neq j} u_t^{i\to j} \sum_{k=1}^N p_{k,t}^{i\to j} \frac{\ell(k, y_t)}{p_{k,t}} \qquad \text{since } 0 \leqslant p_{k,t}^{i\to j}\ell(k, y_t) \leqslant 1 \\
&= \sum_{k=1}^N \left((1-\gamma_t) \sum_{i\neq j} u_t^{i\to j} \widetilde{p}_{k,t}^{i\to j}\right) \frac{\ell(k, y_t)}{p_{k,t}} + \frac{\gamma_t}{N} \sum_{i\neq j} u_t^{i\to j} \sum_{k=1}^N 1_k^{i\to j} \frac{\ell(k, y_t)}{p_{k,t}} \qquad \text{by (3.7)} \\
&\leqslant \sum_{k=1}^N \ell(k, y_t) + N \leqslant 2N \qquad \text{by (3.6) and by using that } p_{k,t} \geqslant \gamma_t/N \quad \text{for all } k.
\end{aligned}$$

Taking expectations in (3.9), using that the expectation of a maximum is more than the maximum of the expectations, and substituting the above inequality, we get

$$\max_{i\neq j} \mathbb{E}\left[\sum_{t=1}^{n} p_{i,t}\left(\ell(i,y_t)-\ell(j,y_t)\right)\right] \leqslant \frac{4\ln N}{\eta_{n+1}} + 2N\sum_{t=1}^{n}\eta_t + \sum_{t=1}^{n}\frac{\gamma_t}{N}\,.$$

Recalling now (3.5), and using $\sum_{t=1}^{n} 1/\sqrt{t} \leqslant 2\sqrt{n}$, we conclude to

$$\max_{i\neq j} \mathbb{E}\left[\sum_{t=1}^{n} p_{i,t}\left(\ell(i,y_t)-\ell(j,y_t)\right)\right] \leqslant \left(\frac{4}{\alpha}+6\alpha\right)\sqrt{(n+1)N\ln N}\,,$$

where $\alpha = \sqrt{2/3}$. □

2. Minimax lower bounds on internal regret

2.1. A general lower bound on internal regret in an expert setting. We define the minimax order of the internal regret in the same way as we defined the minimax order of the external regret in Section 5 of Chapter 2. Recall that the external regret of a forecaster is bounded by N times its internal regret, and that the minimax order of the former is $\sqrt{n\ln N}$ (see respectively (3.1) and Section 5 in Chapter 2). From these facts, we know that the minimax order of the internal regret is at least $\sqrt{n\ln N}/N$. But tighter lower bounds may be achieved, as is shown below.

The idea is to reduce to the case of external regret with $N = 2$ actions. This reduction is not immediate, and the lower bounds on the external regret in the case $N = 2$ cannot be used directly, simply because the forecaster maintains more than two weights, and spreads the mass into N weights, one for each action. If N is large, then little can be said. Especially, the "uniform" forecaster which picks at random an action at each step, that is, $\boldsymbol{p}_t = (1/N,\dots,1.N)$ for all t, and suffers an internal regret less than n/N, may achieve a low internal regret. We provide below a rigorous reduction. The intuition is that the outcomes are such that actions 3 to N always suffer a maximal loss, and therefore are almost never played by any good forecaster. The latter thus concentrates on actions 1 and 2, which it takes some time to distinguish. The proof techniques show that then, we are basically back to the problem of lower bounding the external regret of a forecaster only supplied with two actions.

We denote the set of natural numbers by $\mathbb{N} = \{1, 2, \dots\}$.

THEOREM 3.3. *There exist an outcome space $\mathcal{Y}$, a loss function $\ell : \mathbb{N}\times\mathcal{Y} \to [0,1]$, and a universal constant $c > 0$ such that for all $N \geqslant 2$ and n such that $N \leqslant 8\sqrt{3}\sqrt{n}$, the cumulative (expected) internal regret of any (randomized) forecaster that uses actions in $\{1,\dots,N\}$ satisfies the inequality, against an oblivious opponent,*

$$\sup_{y_1,\dots,y_n\in\mathcal{Y}} \max_{i\neq j} \sum_{t=1}^{n} p_{i,t}\left(\ell(i,y_t)-\ell(j,y_t)\right) \geqslant c\sqrt{n}\,.$$

In particular, we prove the theorem for $c = 1/(64\sqrt{3})$.

Note that the technical condition we need on n and N basically ensures that we are not in the trivial case where the uniform forecaster performs betters than the general forecaster introduced in Section 1.2.

OPEN QUESTION 3.1. We note here that we still lack a factor of $\sqrt{\ln N}$ in the lower bound, or, alternatively, the bounds on internal regret derived in Section 1.2 might be improvable. Given the optimality of the weighted average forecaster with respect to external regret (see Section 5.3 in Chapter 2), and in view of the conversion trick of Section 1.2, we however conjecture that this

additional $\sqrt{\ln N}$ factor is necessary. Perhaps this can be done by applying Fano's lemma to a family of distributions over the outcomes. This family would be indexed by $(i,j)_{i\neq j}$, and the proof below introduces a distribution related to the one that would correspond to $(1,2)$.

A related lower bound on swap regret may be found in [**BlMa05**]. It is of the order of $\Omega(\sqrt{nN})$, that is, the question of the additional $\sqrt{\ln N}$ factor is also not answered there, and holds only under the additional condition that n be sub-exponential in N.

PROOF. We only sketch the proof and refer for more details to the proof of Theorem 5.5 in Section 5 of Chapter 5. We may choose $\mathcal{Y} = [0,1]$ and a loss function $\ell : \mathbb{N} \times \mathcal{Y} \to [0,1]$ such that there exist a probability space, equipped with three probability distributions $\mathbb{P}$, $\mathbb{Q}$, $\mathbb{R}$, such that there exists a sequence of random variables $Y_1, \ldots, Y_n$ defined on it and taking values in $\mathcal{Y}$ satisfying the following property. Under $\mathbb{P}$ (resp., $\mathbb{Q}$, $\mathbb{R}$), the losses

$$\widetilde{\ell}_{k,t} = \ell(k, Y_t)\,, \qquad k = 1, \ldots, N, \text{ and } t = 1, \ldots, N\,,$$

are independent random variables, equal to 1 if $k \geqslant 3$, with Bernoulli distribution with parameter $1/2$ for $k = 1$, and with parameter $1/2$ (resp., $1/2 - \varepsilon$, $1/2 + \varepsilon$) for $k = 2$. Then, denoting by $\mathbb{E}_{\mathbb{P}}$ (resp., $\mathbb{E}_{\mathbb{Q}}$, $\mathbb{E}_{\mathbb{R}}$) the expectation with respect to $\mathbb{P}$ (resp., $\mathbb{Q}$, $\mathbb{R}$), we note that it suffices to show that

$$R_n = \mathbb{E}_{\mathbb{P}}\left[\max_{i\neq j} \sum_{t=1}^{n} p_{i,t}\left(\ell(i,Y_t) - \ell(j,Y_t)\right)\right] \geqslant c\sqrt{n}\,. \tag{3.11}$$

Now,

$$\begin{aligned}
R_n &= \mathbb{E}_{\mathbb{P}}\left[\max_{i\neq j} \sum_{t=1}^{n} p_{i,t}\left(\ell(i,Y_t) - \ell(j,Y_t)\right)\right] \\
&\geqslant \frac{1}{2}\mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{n} p_{1,t}\left(\ell(1,Y_t) - \ell(2,Y_t)\right)\right] + \frac{1}{2}\mathbb{E}_{\mathbb{R}}\left[\sum_{t=1}^{n} p_{2,t}\left(\ell(2,Y_t) - \ell(1,Y_t)\right)\right] \\
&= \frac{\varepsilon}{2}\left(\mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{n} p_{1,t}\right] + \mathbb{E}_{\mathbb{R}}\left[\sum_{t=1}^{n} p_{2,t}\right]\right) \\
&= \frac{\varepsilon}{2}\left(2n - \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{n} p_{2,t}\right] - \mathbb{E}_{\mathbb{R}}\left[\sum_{t=1}^{n} p_{1,t}\right] - 2\sum_{j=3}^{N}\mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{n} p_{j,t}\right]\right)\,.
\end{aligned}$$

We may always assume that for all $j = 3, \ldots, N$,

$$\frac{1}{2}\mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{n} p_{j,t}\right] = \frac{1}{2}\mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{n} p_{j,t}\left(\ell(j,Y_t) - \ell(1,Y_t)\right)\right] \leqslant c\sqrt{n}$$

(otherwise (3.11) is true, and the proof is done), so that

$$R_n \geqslant \frac{\varepsilon}{2}\left(2n - \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{n} p_{2,t}\right] - \mathbb{E}_{\mathbb{R}}\left[\sum_{t=1}^{n} p_{1,t}\right] - 4(N-2)c\sqrt{n}\right)\,. \tag{3.12}$$

We denote by $\mathbb{P}_A$ the expectation with respect to the auxiliary randomization, and similarly to the proof of Theorem 5.5, we note that, thanks to Fubini's theorem,

$$\mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{n} p_{2,t}\right] = \mathbb{Q}\otimes\mathbb{P}_A[I_t = 2]\,. \tag{3.13}$$

We now use Pinsker's inequality (see Lemma A.6 in the Appendix) to get

$$\mathbb{Q}\otimes\mathbb{P}_A[I_t = 2] \leqslant \mathbb{P}\otimes\mathbb{P}_A[I_t = 2] + \sqrt{\frac{1}{2}\mathcal{K}\left(\mathbb{P}\otimes\mathbb{P}_A,\, \mathbb{Q}\otimes\mathbb{P}_A\right)}\,. \tag{3.14}$$

We upper bound the Kullback-Leibler divergence thanks to Lemma A.5 and (A.1),

$$\mathcal{K}\left(\mathbb{P}\otimes\mathbb{P}_A,\,\mathbb{Q}\otimes\mathbb{P}_A\right)=\mathcal{K}\left(\mathbb{B}_{1/2},\mathbb{B}_{1/2-\varepsilon}\right)\leqslant 6\,\varepsilon^2\;, \tag{3.15}$$

for all $0\leqslant\varepsilon\leqslant 1/\sqrt{6}$. We proceed similarly for the expectation under $\mathbb{R}$ in (3.12), combine (3.12)–(3.15), and perform some crude bounding,

$$\begin{aligned} R_n &\geqslant \frac{\varepsilon}{2}\left(2n-\left(\sum_{t=1}^{n}\mathbb{P}\otimes\mathbb{P}_A\left[I_t\in\{1,\,2\}\right]\right)-2n\sqrt{3\varepsilon^2}-4(N-2)c\sqrt{n}\right)\\ &\geqslant \frac{n\varepsilon}{2}\left(1-2\sqrt{3}\,\varepsilon-\frac{4cN}{\sqrt{n}}\right)\geqslant\frac{n\varepsilon}{2}\left(\frac{1}{2}-2\sqrt{3}\,\varepsilon\right)\;, \end{aligned}$$

where we used the fact that $N\leqslant 8\sqrt{3}\sqrt{n}$ and $c=1/(64\sqrt{3})$ in the last inequality. We choose $\varepsilon=1/(8\sqrt{3})$ to conclude the proof. □

2.2. Interpretation of internal regret as an extremum of performance. When a forecasting strategy suffers a small internal regret, this means that it cannot easily be improved, that is, either it was already very efficient, or it makes so poor predictions that there is no hope to improve its performances. Hopefully, for (randomized) prediction with expert advice, (3.1) shows that as external regret is upper bounded by N times internal regret, the second case of the alternative never happens. This, unfortunately, is not the case for all prediction settings (see, for instance, Example 7.1 or the example below). In Chapter 7 we derive investment strategies which, at the same time, suffer small internal and external regret, and the example below shows why we should not focus only on internal regret in general prediction settings, while the results of Chapter 7 indicate on the other hand that minimizing both regrets at the same is worthwhile.

Consider the problem of *sequential probability assignment*, described as follows (see Lugosi [**Lug01**] or Catoni [**Cat01**] for more references and background). A forecaster repeatedly has to output a probability distribution $\boldsymbol{p}_t\in\mathcal{X}$, $t=1,2,\ldots$, where $\mathcal{X}$ is the set of all probability distributions over the finite outcome space $\mathcal{Y}$. Without loss of generality, we take $\mathcal{Y}=\{1,\ldots,N\}$ and $\mathcal{X}$ is the real simplex of order N. We denote an element $\boldsymbol{p}\in\mathcal{X}$ by $\boldsymbol{p}=(p_1,\ldots,p_N)$. Now, the loss function $\ell:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}_+$ is defined by $\ell(\boldsymbol{p},y)=-\ln p(y)$. In this setting, in agreement to the definitions introduced later in Chapters 7 and 8, we define the internal regret of a forecaster by the difference between the cumulative losses of the original algorithm and of its $i\to j$ modified strategy. The latter predicts with $\boldsymbol{p}_t^{i\to j}$ instead of $\boldsymbol{p}_t$ at each time step t, where we use again the notation of Section 1.2.

It is convenient in this framework to emphasize the dependencies on the past, and denote $\boldsymbol{p}_t=\boldsymbol{p}_t(\cdot\,|y_1^{t-1})$, where $y_1^{t-1}=(y_1,\ldots,y_{t-1})$ (and y_1^0 is the empty sequence) is the history up to time $t-1$. We note that there is a one-to-one correspondence between forecasting strategies and sequences of probability distributions over the $(\mathcal{Y}^n)_{n\geqslant 1}$, given by

$$\boldsymbol{p}_n(y_1^n)=\prod_{t=1}^{n}\boldsymbol{p}_t(y_t|y_1^{t-1})\;.$$

We similarly denote, for $i\neq j$, $\boldsymbol{p}_n^{i\to j}(y_1^n)=\displaystyle\prod_{t=1}^{n}\boldsymbol{p}_t^{i\to j}(y_t|y_1^{t-1})$.

Now, the cumulative internal of a forecasting strategy over a sequence y_1^n in the setting of sequential probability assignment equals

$$\max_{i\neq j}\sum_{t=1}^{n}\ell(\boldsymbol{p}_t,y_t)-\ell\left(\boldsymbol{p}_t^{i\to j},y_t\right)=\ln\frac{\boldsymbol{p}_n^{i\to j}(y_1^n)}{\boldsymbol{p}_n(y_1^n)}\;,$$

with the usual convention that $\ln(0/0) = 0$.

PROPOSITION 3.1. *The minimax value of internal regret in the sequential probability assignment problem equals, for all $n \geqslant 2$ and $N \geqslant 2$,*

$$\inf \max_{y_1^n \in \mathcal{Y}^n} \max_{i \neq j} \ln \frac{\boldsymbol{p}_n^{i \to j}(y_1^n)}{\boldsymbol{p}_n(y_1^n)} = \ln 2 \,,$$

where the infimum is taken over all possible forecasters. Moreover, the forecaster using $\boldsymbol{p}_1 = (1/N, \ldots, 1/N)$ and $\boldsymbol{p}_t = \delta_{y_1}$, where δ_j is the Dirac mass on j, achieves this minimax value.

We omit the simple proof, and remark here that the forecasting strategy achieving the minimax value is a very poor prediction scheme, which suffers a large external regret with respect to most finite comparison classes, though its internal regret is the smallest possible. That is, this strategy suffers a small internal regret because it is already so bad that there is no possibility to improve it.

CHAPTER 4

Improved second-order bounds in prediction with expert advice

This chapter studies external regret in sequential prediction games with arbitrary payoffs (non-positive or nonnegative). External regret measures the difference between the payoff obtained by the forecasting strategy and the payoff of the best action. We focus on two important parameters: M, the largest absolute value of any payoff, and Q^*, the sum of squared payoffs of the best action. Given these parameters we derive first a simple and new forecasting strategy with regret at most order of $\sqrt{Q^*(\ln N)} + M \ln N$, where N is the number of actions. We extend the results to the case where the parameters are unknown and derive similar bounds. We then devise a refined analysis of the weighted majority forecaster, which yields bounds of the same flavour. The proof techniques we develop are finally applied to the adversarial multi-armed bandit setting, and we prove bounds on the performance of an online algorithm in the case where there is no lower bound on the probability of each action. We close the chapter with a preliminary result about fast rates of convergence in randomized prediction with expert advice. This wide range of applications demonstrates the power and generality of our methodology.

Contents

Most of this chapter is based on a joint work with Nicolò Cesa-Bianchi and Yishay Mansour. An extended abstract of these results [**CeMaSt05**] is to be presented at COLT'05.

1. Introduction

The study of online forecasting strategies in adversarial settings has received considerable attention in the last few years in the computational learning literature and elsewhere. The main focus has been on deriving simple online algorithms that have low external regret. The external

regret of an online algorithm is the difference between its expected payoff and the best payoff achievable using some strategy from a given class. Usually, this class includes a strategy, for each action, which always plays that action. In a nutshell, one can show that the average external regret per time step vanishes, and much of the research has been to both improve and refine the bounds.

Ideally, in an adversarial setting one should be able to show that the regret with respect to any action only depends on the variance of the observed payoffs for that action. In a stochastic setting such a result seems like the most natural bound, and deriving its analogue in an adversarial setting would be a fundamental result. We believe that our results make a significant step toward this goal, although, unfortunately, fall short of completely achieving it.

In order to describe our results we first set up our model and notation, and relate them to previous works. In this chapter we consider the following game-theoretic version of the prediction-with-expert-advice framework [**CeFrHaHeScWa97, LiWa94, Vov98**], see also Chapter 2. A forecaster repeatedly assigns probabilities to a fixed set of actions. After each assignment, the real payoff associated to each action is revealed and new payoffs are set for the next round. The forecaster's reward on each round is the average payoff of actions for that round, where the average is computed according to the forecaster's current probability assignment. The goal of the forecaster is to achieve, on any sequence of payoffs, a cumulative reward close to X^*, the highest cumulative payoff among all actions. As usual, we call regret the difference between X^* and the cumulative reward achieved by the forecaster on the same payoff sequence.

The special case of "one-sided games", when all payoffs have the same sign (they are either always non-positive or always nonnegative) has been considered by Freund and Schapire [**FrSc97**], and by Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**] in a related context (see also the whole Chapter 2, which deals only with losses, that is, with non-positive payoffs). These papers show that Littlestone and Warmuth's weighted majority algorithm [**LiWa94**] can be used as a basic ingredient to construct a forecasting strategy achieving a regret of $O(\sqrt{M|X^*|\ln N})$ in one-sided games, where N is the number of actions and M is a known upper bound on the size of payoffs. (If all payoffs are non-positive, then the absolute value of each payoff is called *loss* and $|X^*|$ is the cumulative loss of the best action.) By a simple rescaling of payoffs, it is possible to reduce the more general "signed game", in which each payoff might have an arbitrary sign, to either one of the one-sided games (note that this reduction assumes knowledge of M). However, the regret becomes $O(M\sqrt{n\ln N})$, where n is the number of game rounds. Recently, Allenberg and Neeman [**AlNe04**] proposed a direct analysis of the signed game avoiding this reduction. Before describing their results, we introduce some convenient notation and terminology.

Our forecasting game is played in rounds. At each time step $t = 1, 2, \ldots$ the forecaster computes an assignment $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ of probabilities over the N actions. Then the payoff vector $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{N,t}) \in \mathbb{R}^N$ for time t is revealed and the forecaster's (expected) reward is

$$\widehat{x}_t = x_{1,t}p_{1,t} + \ldots + x_{N,t}p_{N,t} \,.$$

We define the cumulative reward of the forecaster by $\widehat{X}_n = \widehat{x}_1 + \ldots + \widehat{x}_n$ and the cumulative payoff of action i by $X_{i,n} = x_{i,1} + \ldots + x_{i,n}$. For all n, let

$$X_n^* = \max_{i=1,\ldots,N} X_{i,n}$$

be the cumulative payoff of the best action up to time n. The forecaster's goal is to keep the (expected) *regret* $X_n^* - \widehat{X}_n$ as small as possible uniformly over n.

The one-sided games, mentioned above, are the *loss game*, where $x_{i,t} \leqslant 0$ for all i and t, and the *gain game*, where $x_{i,t} \geqslant 0$ for all i and t. We call *signed game* the setup in which

no assumptions are made on the sign of the payoffs. For the signed game, Allenberg and Neeman [**AlNe04**] show that weighted majority (used in conjunction with a doubling trick) achieves the following: on any sequence of payoffs there exists an action j such that the regret is at most of order $\sqrt{M(\ln N)\sum_{t=1}^{n}|x_{j,t}|}$, where $M = \max_{i,t}|x_{i,t}|$ is a known upper bound on the size of payoffs. Note that this bound does not relate the regret to the sum $|x_1^*| + \ldots + |x_n^*|$ of payoff sizes for the optimal action (i.e., the one achieving X_n^*). In particular, the bound $O(\sqrt{M|X_n^*|\ln N})$ for the one-sided games is only obtained if an estimate of X_n^* is available in advance.

In this chapter we show new regret bounds for the signed game. Our analysis has two main advantages: first, no preliminary knowledge of the payoff size M or of the best cumulative payoff X_n^* is needed; second, our bounds are expressed in terms of sums of squared payoffs, such as $x_{i,1}^2 + \ldots + x_{i,n}^2$ and related forms. These quantities replace the larger terms $M(|x_{i,1}| + \ldots + |x_{i,n}|)$ appearing in the previous bounds. As an application of our results we obtain, without any preliminary knowledge on the payoff sequence, an improved regret bound for the one-sided games of the order of $\sqrt{M\min\{Mn - |X_n^*|, |X_n^*|\}(\ln N)}$ (and even $\sqrt{(Mn - |X_n^*|)(|X_n^*|/n)(\ln N)}$).

Expressions involving squared payoffs are at the core of many analyses in the framework of prediction with expert advice, especially in the presence of limited feedback. (See, for instance, the bandit problem in Section 4 of Chapter 2 and in [**AuCeFrSc02**], and more generally prediction under partial monitoring, see Chapters 5 and 6, as well as [**CeLuSt04a, CeLuSt04b, PiSc01**]). However, to the best of our knowledge, our bounds are the first ones to explicitly include second-order information extracted from the payoff sequence. In particular, our bounds are stable under many transformations of the payoff sequence, and therefore are in some sense more "fundamental".

Some of our bounds are achieved using forecasters based on weighted majority run with a dynamic learning rate. However, we are able to obtain second-order bounds of a different flavour using a new forecaster that does not use the exponential probability assignments of weighted majority. In particular, unlike virtually all previously known forecasting schemes, the weights of this forecaster cannot be represented as the gradient of an additive potential, see Section 2.3 in Chapter 2 or [**CeLu03**].

In bandit problems and, more generally, in all incomplete-information problems like label-efficient prediction or prediction with partial monitoring, a crucial point is to estimate the unobserved losses. In such settings, a probability distribution is formed by using weighted averages of the cumulative estimated losses, and a common practice is to mix this probability distribution, so that the resulting distribution have all the probabilities above a certain value. Technically, this is important since it is common to divide by the probabilities (see [**AuCeFrSc02, CeLuSt04a, CeLuSt04b, HaMa02, PiSc01**] and the forecasting schemes of Chapters 5 and 6). We show that, for the algorithm of [**AuCeFrSc02**] and in bandit loss games,, using our proof technique one can simply use the original probability distribution, computed with the estimates, without any adjustments and get an expected bound which is an improvement for small losses.

We close the chapter with a preliminary result about fast rates of convergence in randomized prediction with expert advice.

2. A new algorithm for sequential prediction

We introduce a new forecasting strategy for the signed game. In Theorem 4.3, the main result of this section, we show that, without any preliminary knowledge of the sequence of payoffs, the regret of a variant of this strategy is bounded by a quantity defined in terms of the sums

$$Q_{i,n} = x_{i,1}^2 + \ldots + x_{i,n}^2 \,.$$

Since $Q_{i,n} \leqslant M(|x_{i,1}| + \ldots + |x_{i,n}|)$, such second-order bounds are generally better than the previously known bounds for any of the three (loss, gain, and signed) games, and in certain cases the difference can be significant.

Our basic forecasting strategy, which we call PROD(η), has an input parameter $\eta > 0$ and maintains a set of N weights. At time $t = 1$ the weights are initialized with $w_{i,1} = 1$ for $i = 1, \ldots, N$. At each time $t = 1, 2, \ldots$, PROD(η) computes the probability assignment $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$, where $p_{i,t} = w_{i,t}/W_t$. After the payoff vector $\boldsymbol{x}_t$ is revealed, the weights are updated using the rule $w_{i,t+1} = w_{i,t}(1 + \eta x_{i,t})$. We use the notation $W_t = w_{1,t} + \ldots + w_{N,t}$.

The following simple fact plays a key role in our analysis.

LEMMA 4.1. *For all $z \geqslant -1/2$, $\ln(1 + z) \geqslant z - z^2$.*

PROOF. Let $f(z) = \ln(1 + z) - z + z^2$. Note that

$$f'(z) = \frac{1}{1+z} - 1 + 2z = \frac{z(1+2z)}{1+z} \,,$$

so that $f'(z) \leqslant 0$ for $-1/2 \leqslant z \leqslant 0$ and $f'(z) \geqslant 0$ for $z \geqslant 0$. Hence the minimum of f is achieved in 0 and equals 0, concluding the proof. □

We are now ready to state a lower bound on the cumulative reward of PROD(η) in terms of the quantities $Q_{k,n}$.

LEMMA 4.2. *Assume there exists $M > 0$ such that the payoffs satisfy $x_{i,t} \geqslant -M$ for $t = 1, \ldots, n$ and $i = 1, \ldots, N$. For any sequence of payoffs, for any action k, for any $\eta \leqslant 1/(2M)$, and for any $n \geqslant 1$, the cumulative reward of* PROD(η) *is lower bounded as*

$$\widehat{X}_n \geqslant X_{k,n} - \frac{\ln N}{\eta} - \eta\, Q_{k,n} \,.$$

PROOF. For any $k = 1, \ldots, N$, note that $x_{k,t} \geqslant -M$ and $\eta \leqslant 1/(2M)$ imply $\eta x_{k,t} \geqslant -1/2$. Hence, we can apply Lemma 4.1 to $\eta x_{k,t}$ and get

$$\begin{aligned} \ln \frac{W_{n+1}}{W_1} &= -\ln N + \ln \prod_{t=1}^{n} (1 + \eta x_{k,t}) = -\ln N + \sum_{t=1}^{n} \ln(1 + \eta x_{k,t}) \\ &\geqslant -\ln N + \sum_{t=1}^{n} \left(\eta x_{k,t} - \eta^2 x_{k,t}^2\right) = -\ln N + \eta X_{k,n} - \eta^2 Q_{k,n} \,. \end{aligned} \tag{4.1}$$

On the other hand,

$$\ln \frac{W_{n+1}}{W_1} = \sum_{t=1}^{n} \ln \frac{W_{t+1}}{W_t} = \sum_{t=1}^{n} \ln \left(\sum_{i=1}^{N} p_{i,t} \left(1 + \eta x_{i,t}\right) \right) \leqslant \eta \widehat{X}_n \tag{4.2}$$

where in the last step we used $\ln(1 + z_t) \leqslant z_t$ for all $z_t = \eta \sum_{i=1}^{N} x_{i,t} p_{i,t} \geqslant -1/2$. Combining (4.1) and (4.2), and dividing by $\eta > 0$, we get

$$\widehat{X}_n \geqslant -\frac{\ln N}{\eta} + X_{k,n} - \eta\, Q_{k,n} \,,$$

which completes the proof of the lemma. □

By choosing η appropriately, we can optimize the bound as follows.

THEOREM 4.1. *Assume there exists $M > 0$ such that the payoffs satisfy $x_{i,t} \geqslant -M$ for $t = 1, \ldots, n$ and $i = 1, \ldots, N$. For any $Q > 0$, if* PROD(η) *is run with*

$$\eta = \min \left\{ 1/(2M), \sqrt{(\ln N)/Q} \right\}$$

then for any sequence of payoffs, for any action k, and for any $n \geqslant 1$ such that $Q_{k,n} \leqslant Q$,

$$\widehat{X}_n \geqslant X_{k,n} - \max\left\{2\sqrt{Q \ln N}, 4M \ln N\right\} .$$

To achieve the bound stated in Theorem 4.1, the parameter η must be tuned using preliminary knowledge of a lower bound on the payoffs and an upper bound on the quantities $Q_{k,n}$. The next two results remove these requirements one by one. We start by introducing a new algorithm that, using a doubling trick over PROD, avoids any preliminary knowledge of a lower bound on the payoffs.

Let PROD-M(Q) be the prediction algorithm that receives a number $Q > 0$ as input parameter and repeatedly runs PROD(η_r), where $\eta_r = 1/(2M_r)$ and M_r is defined below. We call epoch r the sequence of time steps when PROD-M is running PROD(η_r). At the beginning, $r = 0$ and PROD-M(Q) runs PROD(η_0), where

$$M_0 = \sqrt{Q/(4 \ln N)} \qquad \text{and} \quad \eta_0 = 1/(2M_0) = \sqrt{(\ln N)/Q} .$$

The last step of epoch $r \geqslant 0$ is the time step $t = t_r$ when $\max_{i=1,\dots,N} |x_{i,t}| > M_r$ happens for the first time. When a new epoch $r+1$ begins, PROD is restarted with parameter $\eta_{r+1} = 1/(2M_{r+1})$, where $M_{r+1} = \max_i 2^{\lceil \log_2 |x_{i,t_r}| \rceil}$. Note that $M_1 \geqslant M_0$ and, for each $r \geqslant 1$, $M_{r+1} \geqslant 2M_r$.

THEOREM 4.2. *For any sequence of payoffs, for any action k, and for any $n \geqslant 1$ such that $Q_{k,n} \leqslant Q$, the cumulative reward of algorithm* PROD-M(Q) *is lower bounded as*

$$\widehat{X}_n \geqslant X_{k,n} - 2\sqrt{Q \ln N} - 4M\,(2 + 3 \ln N)$$

where $M = \max_{1 \leqslant i \leqslant N} \max_{1 \leqslant t \leqslant n} |x_{i,t}|$.

PROOF. We denote by R the index of the last epoch and let $t_R = n$. If we have only one epoch, then the theorem follows from Theorem 4.1 applied with a lower bound of $-M_0$ on the payoffs. Therefore, for the rest of the proof we assume $R \geqslant 1$. Let

$$X_k^r = \textstyle\sum_{s=t_{r-1}+1}^{t_r - 1} x_{k,s}, \quad Q_k^r = \sum_{s=t_{r-1}+1}^{t_r - 1} x_{k,s}^2, \quad \widehat{X}^r = \sum_{s=t_{r-1}+1}^{t_r - 1} \widehat{x}_s ,$$

where the sums are over all the time steps t in epoch r except the last one, t_r. (Here t_{-1} is conventionally set to 0.) Applying Lemma 4.1 to each epoch $r = 0, \dots, R$ we get that $\widehat{X}_n - X_{k,n}$ is equal to

$$\sum_{r=0}^{R} \left(\widehat{X}^r - X_k^r\right) + \sum_{r=0}^{R-1} (\widehat{x}_{t_r} - x_{k,t_r}) \geqslant -\sum_{r=0}^{R} \frac{\ln N}{\eta_r} - \sum_{r=0}^{R} \eta_r Q_k^r + \sum_{r=0}^{R-1} (\widehat{x}_{t_r} - x_{k,t_r}) .$$

We bound each sum separately. For the first sum note that

$$\sum_{r=0}^{R} \frac{\ln N}{\eta_r} = \sum_{r=0}^{R} 2M_r \ln N \leqslant 6M_R \ln N$$

since $M_R \geqslant 2^{R-r} M_r$ for each $r \geqslant 1$ and $M_0 \leqslant M_R$. For the second sum, using that the η_r decrease, we have

$$\sum_{r=0}^{R} \eta_r Q_k^r \leqslant \eta_0 \sum_{r=0}^{R} Q_k^r \leqslant \eta_0 Q_{k,n} \leqslant \sqrt{\frac{\ln N}{Q}}\, Q = \sqrt{Q \ln N} .$$

Finally,

$$\sum_{r=0}^{R-1} |\widehat{x}_{t_r} - x_{k,t_r}| \leqslant \sum_{r=1}^{R} 2\, M_r \leqslant 4\, M_R .$$

The resulting lower bound $2M_R(2+3\ln N)+\sqrt{Q\ln N}$ implies the one stated in the theorem by noting that, when $R \geqslant 1$, $M_R \leqslant 2M$. □

We now show a regret bound for the case when M and the $Q_{k,n}$ are both unknown. Let k_t^* be the index of the best action up to time t; that is, $k_t^* \in \operatorname{argmax}_k X_{k,t}$ (ties are broken by choosing the action k with minimal associated $Q_{k,t}$). We denote the associated quadratic penalty by

$$Q_t^* = Q_{k_t^*}^* = \textstyle\sum_{s=1}^t x_{k_t^*,s}^2 \,.$$

Ideally, our final regret bound should depend on Q_n^*. However, note that the sequence $Q_1^*, Q_2^*, \ldots$ is not necessarily monotone, as Q_t^* and Q_{t+1}^* cannot be possibly related when the actions achieving the largest cumulative payoffs at rounds t and $t+1$ are different. Therefore, we cannot use a straightforward doubling trick, as this only applies to monotone sequences. Our solution is to express the bound in terms of the smallest nondecreasing sequence that upper bounds the original sequence $(Q_t^*)_{t\geqslant 1}$. This is a general trick to handle situations where the penalty terms are not monotone. Allenberg and Neeman [**AlNe04**] faced a similar situation, and we improve their results.

We define a new (parameterless) prediction algorithm PROD-MQ in the following way. The algorithm runs in epochs using PROD-M(Q) as a subroutine. The last step of epoch r is the time step $t = t_r$ when $Q_t^* > 4^r$ happens for the first time. At the beginning of each new epoch $r = 0, 1, \ldots$, algorithm PROD-M(Q) is restarted with parameter $Q = 4^r$.

THEOREM 4.3. *For any sequence of payoffs and for any $n \geqslant 1$, the cumulative reward of algorithm* PROD-MQ *satisfies*

$$\widehat{X}_n \geqslant X_n^* - 8\sqrt{(\ln N)\max\left\{1,\, \max_{s\leqslant n} Q_s^*\right\}} - 12M\left(2+\log_4 \max_{s\leqslant n} Q_s^*\right)(1+\ln N)$$

where $M = \max_{1\leqslant i\leqslant N}\max_{1\leqslant t\leqslant n}|x_{i,t}|$.

PROOF. We denote by R the index of the last epoch and let $t_R = n$. Assume that $R \geqslant 1$ (otherwise the proof is concluded by Theorem 4.2). Similarly to the proof of Theorem 4.2, for all epochs r and actions k introduce

$$X_k^r = \textstyle\sum_{s=t_{r-1}+1}^{t_r-1} x_{k,s}\,, \quad Q_k^r = \sum_{s=t_{r-1}+1}^{t_r-1} x_{k,s}^2\,, \quad \widehat{X}^r = \sum_{s=t_{r-1}+1}^{t_r-1}\widehat{x}_s$$

where $t_{-1} = 0$. We also denote $k_r = k_{t_r-1}^*$ the index of the best overall expert up to time $t_r - 1$ (one time step before the end of epoch r). We have that $Q_{k_r}^r \leqslant Q_{k_r,t_r-1} = Q_{t_r-1}^*$. Now, by definition of the algorithm, $Q_{t_r-1}^* \leqslant 4^r$. Theorem 4.2 (applied to time steps $t_{r-1}+1, \ldots, t_r-1$) shows that $\widehat{X}^r \geqslant X_{k_r}^r - \Phi\left(M,\, 4^r\right)$, where $\Phi(M,x) = 2\sqrt{x\ln N} + 4M(2+3\ln N)$. Summing over $r = 0, \ldots, R$ we get

$$\widehat{X}_n = \sum_{r=0}^{R}\widehat{X}^r + \widehat{x}_{k_r,t_r} \geqslant \sum_{r=0}^{R}\left(\widehat{x}_{k_r,t_r} + X_{k_r}^r - \Phi\left(M,\, 4^r\right)\right)\,. \tag{4.3}$$

Now, since k_1 is the index of the expert with largest payoff up to time $t_1 - 1$, we have that $X_{k_2,t_2-1} = X_{k_2}^1 + x_{k_2,t_1} + X_{k_2}^2 \leqslant X_{k_1}^1 + X_{k_2}^2 + M$. By a simple induction, we in fact get

$$X_{k_R,t_R-1} \leqslant \sum_{r=0}^{R-1}\left(X_{k_r}^r + M\right) + X_{k_R}^R\,. \tag{4.4}$$

As, in addition, X_{k_R,t_R-1} and $X_{k_n^*,n}$ may only differ by at most M, combining (4.3) and (4.4) we have indeed proven that

$$\widehat{X}_n \geqslant X_{k_n^*,n} - \left(2(1+R)M + \sum_{r=0}^{R} \Phi\left(M,\, 4^r\right) \right) .$$

The sum over r is now bounded as follows

$$\sum_{r=0}^{R} \Phi\left(M,\, 4^r\right) \leqslant 4M(1+R)\left(2+3\ln N\right) + 2^{R+1}\left(2\sqrt{\ln N}\right) .$$

The proof is concluded by noting that, as $R \geqslant 1$, $\sup_{s\leqslant n} Q_s^* \geqslant 4^{R-1}$ by definition of the algorithm. □

As a final remark for this section, note that we may run PROD-MQ using translated payoffs $r_{k,t} = x_{k,t} - \mu_t$, where μ_t is any quantity possibly based on the past payoffs $x_{i,s}$, for $i = 1, \ldots, N$ and $s = 1, \ldots, t$. An interesting application is obtained by considering $\mu_t = \widehat{x}_t$ where $\widehat{x}_t = x_{1,t}p_{1,t} + \ldots + x_{N,t}p_{N,t}$ is the forecaster's reward at time t. As the sums $\widehat{x}_1 + \ldots + \widehat{x}_n$ cancel out in the difference $\widehat{X}_n - X_{k,n}$, we can obtain the following corollary of Theorem 4.3.

COROLLARY 4.1. *If algorithm* PROD-MQ *is run using translated payoffs $x_{k,t} - \widehat{x}_t$, then for any sequence of payoffs and for any $n \geqslant 1$,*

$$\widehat{X}_n \geqslant X_n^* - 8\sqrt{(\ln N)\max\left\{1,\, \max_{s\leqslant n} R_s^*\right\}} - 12M\left(2+\log_4 \max_{s\leqslant n} R_s^*\right)(1+\ln N)$$

where $M = 2\max_{1\leqslant i\leqslant N}\max_{1\leqslant t\leqslant n}|x_{i,t}|$ and $R_t^* = (x_{k_t^*,1} - \widehat{x}_1)^2 + \ldots + (x_{k_t^*,t} - \widehat{x}_t)^2$ for k_t^* achieving the best cumulative payoff at round t (ties broken by choosing the action k with smallest associated $R_{k,t}$).

REMARK 4.1. In a one-sided game, for instance a gain game, the forecaster always has an incentive to translate the payoffs by the minimal payoff μ_t obtained at each round t,

$$\mu_t = \min_{j=1,\ldots,N} x_{k,t} ,$$

just because for all j and t, $(x_{j,t} - \mu_t)^2 \leqslant x_{j,t}^2$ for a gain game. The matter is not so clear however for signed games, and it may be a delicate issue to determine beforehand if the payoffs should be translated, and if so, which translation rule should be used. See Section 4.2 below.

REMARK 4.2. There is one single result that can be deduced from [**AlNe04**] and which is not implied by our new forecaster. Their Theorem 3 bounds the regret at round n as follows. There exists an action j such that

$$\widehat{X}_n \geqslant X_n^* - O\left(\sqrt{MD_{j,n}\ln N}\right) ,$$

where

$$D_{j,n} = \sum_{t=1}^{n} \left(x_{j,t} - \widehat{x}_t\right)_+ .$$

This is achieved by a forecaster using weighted averages, whose basic step of update is given by

$$w_{i,t+1} = w_{i,t}\left(1 - \eta\,\mathrm{sign}\,\left(x_{i,t} - \widehat{x}_t\right)\right)^{|x_{i,t}-\widehat{x}_t|} ,$$

where $\operatorname{sign} u$ equals 1 when $u \geqslant 0$, and -1 otherwise. We can even get, with their Theorem 2 and the modified doubling trick above, the better bound

$$\widehat{X}_n \geqslant X_n^* - O\left(\sqrt{M(\ln N) \max_{1 \leqslant t \leqslant n} D_t^*}\right) ,$$

where $D_t^* = D_{k_t^*,t}$. This bound also leads to an improvement for small or large payoffs in one-sided games, see Section 4.1 for more details on such bounds. Now, in the main term of the bound proposed by Corollary 4.1, we only have

$$R_t^* \leqslant M \sum_{t=1}^{n} \left| x_{k_t^*,t} - \widehat{x}_t \right| .$$

The maximum of the D_t^* might be less than the maximum of the R_t^*, and we were not able to get the former in any bound on a PROD-MQ type algorithm.

3. Second-order bounds for weighted majority

In this section we derive new regret bounds for the weighted majority forecaster of Littlestone and Warmuth [**LiWa94**] using a time-varying learning rate. This allows us to avoid the doubling trick of Section 2 and keep the assumption that no knowledge on the payoff sequence is available in advance to the forecaster.

Similarly to the results of Section 2, the main term in the new bounds depends on second-order quantities associated to the sequence of payoffs. However, the precise definition of these quantities makes the bounds of this section generally not comparable to the bounds obtained in Section 2.

The weighted majority forecaster using the sequence $\eta_2, \eta_3, \ldots > 0$ of learning rates assigns at time t a probability distribution $\boldsymbol{p}_t$ over the N experts defined by $\boldsymbol{p}_1 = (1/N, \ldots, 1/N)$ and

$$p_{i,t} = \frac{e^{\eta_t X_{i,t-1}}}{\sum_{j=1}^{N} e^{\eta_t X_{j,t-1}}} \qquad \text{for } i = 1, \ldots, N \text{ and } t \geqslant 2 , \tag{4.5}$$

see Section 2 in Chapter 2. Note that the quantities $\eta_t > 0$ may depend on the past payoffs $x_{i,s}$, $i = 1, \ldots, N$ and $s = 1, \ldots, t-1$. The analysis of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**], for a related variant of weighted majority, is at the core of the proof of the following lemma (proof in the appendix of this chapter).

LEMMA 4.3. *Consider any nonincreasing sequence $\eta_2, \eta_3, \ldots$ of positive learning rates and any sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in \mathbb{R}^N$ of payoff vectors. Define the nonnegative function Φ by*

$$\Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) = -\sum_{i=1}^{N} p_{i,t} x_{i,t} + \frac{1}{\eta_t} \ln \sum_{i=1}^{N} p_{i,t} e^{\eta_t x_{i,t}} = \frac{1}{\eta_t} \ln \left(\sum_{i=1}^{N} p_{i,t} e^{\eta_t (x_{i,t} - \widehat{x}_t)} \right)$$

Then the weighted majority forecaster (4.5) run with the sequence $\eta_2, \eta_3, \ldots$ satisfies, for any $n \geqslant 1$ and for any $\eta_1 \geqslant \eta_2$,

$$\widehat{X}_n - X_n^* \geqslant -\left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1} \right) \ln N - \sum_{t=1}^{n} \Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) .$$

OPEN QUESTION 4.1. We show below an incremental update for a weighted-majority-based predictor, using the second-order upper bounds on the quantities $\Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t)$ given by Lemma 4.4. The parameters η_t are chosen to minimize the obtained upper bounds. If we had third-order upper bounds, or even sharper ones of a different form, then the form of the η_t would have been different too. The form of the bounds basically indicates the form of the η_t. One may wonder if there is a general way to define the η_t, in terms of the $\Phi(\boldsymbol{p}_s, \eta_s, \boldsymbol{x}_s)$, for $s \leqslant t-1$, and not in

terms of some upper bounds on these quantities. That would result in sharper and very general bounds on the regret.

Let Z_t be the random variable with range $\{x_{1,t}, \ldots, x_{N,t}\}$ and law $\boldsymbol{p}_t$. Note that $\mathbb{E}Z_t$ is the expected payoff $\widehat{x}_t$ of the forecaster using distribution $\boldsymbol{p}_t$ at time t. Introduce

$$\operatorname{Var} Z_t = \mathbb{E}Z_t^2 - \mathbb{E}^2 Z_t = \sum_{i=1}^N p_{i,t} x_{i,t}^2 - \left(\sum_{i=1}^N p_{i,t} x_{i,t} \right)^2 .$$

Hence $\operatorname{Var} Z_t$ is the variance of the payoffs at time t under the distribution $\boldsymbol{p}_t$ and the cumulative variance $V_n = \operatorname{Var} Z_1 + \ldots \operatorname{Var} Z_n$ is the main second-order quantity used in this section. The next result bounds $\Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t)$ in terms of $\operatorname{Var} Z_t$.

LEMMA 4.4. *For all payoff vectors $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{N,t})$, all probability distributions $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$, and all learning rates $\eta_t \geqslant 0$, we have*

$$\Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) \leqslant 2\, M$$

where M is such that $|x_{i,t}| \leqslant M$ for all i. If, in addition, $0 \leqslant \eta_t |x_{i,t}| \leqslant 1/2$ for all $i = 1, \ldots, N$, then

$$\Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) \leqslant (e-2) \eta_t \operatorname{Var} Z_t .$$

PROOF. The first inequality is straightforward. To prove the second one we use $e^a \leqslant 1 + a + (e-2)\, a^2$ for $|a| \leqslant 1$. Consequently, noting that $\eta_t |x_{i,t} - \widehat{x}_t| \leqslant 1$ for all i by assumption, we have that

$$\begin{aligned} \Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) &= \frac{1}{\eta_t} \ln \left(\sum_{i=1}^N p_{i,t} e^{\eta_t (x_{i,t} - \widehat{x}_t)} \right) \\ &\leqslant \frac{1}{\eta_t} \ln \left(\sum_{i=1}^N p_{i,t} \left(1 + \eta_t (x_{i,t} - \widehat{x}_t) + (e-2) \eta_t^2 (x_{i,t} - \widehat{x}_t)^2 \right) \right) . \end{aligned}$$

Using $\ln(1+a) \leqslant a$ for all $a \geqslant -1$ and some simple algebra concludes the proof of the second inequality. □

In [**AuCeFrSc02**] a very similar result is proven, except that there the variance is further bounded (up to a multiplicative factor) by the expectation $\widehat{x}_t$ of Z_t.

We now introduce a time-varying learning rate based on V_n. For any sequence of payoff vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ and for all $t = 1, 2, \ldots$ let $M_t = 2^k$, where k is the smallest *nonnegative* integer such that $\max_{s=1,\ldots,t} \max_{i=1,\ldots,N} |x_{i,s}| \leqslant 2^k$. Now let the sequence $\eta_2, \eta_3, \ldots$ be defined as

$$\eta_t = \min \left\{ \frac{1}{2\, M_{t-1}}, C \sqrt{\frac{\ln N}{V_{t-1}}} \right\} \quad \text{for } t \geqslant 2, \text{ with } \quad C = \sqrt{\frac{2}{e-2} \left(\sqrt{2} - 1 \right)} . \tag{4.6}$$

Note that η_t depends on the forecaster's past predictions. This is in the same spirit as the self-confident learning rates considered in [**AuCeGe02**].

We are now ready to state and prove the main result of this section, which bounds the regret in terms of the variances of the predictions. We show in the next section how this bound leads to more intrinsic bounds on the regret.

THEOREM 4.4. *Consider the weighted majority forecaster using the time-varying learning rate (4.6). Then, for all sequences of payoffs and for all $n \geqslant 1$,*

$$\widehat{X}_n - X_n^* \geqslant -4 \sqrt{V_n \ln N} - 16 \max\{M, 1\} \ln N - 8 \max\{M, 1\} - M^2$$

where $M = \max_{t=1,\dots,n} \max_{i=1,\dots,N} |x_{i,t}|$.

PROOF. We start by applying Lemma 4.3 using the learning rate (4.6), and setting $\eta_1 = \eta_2$ for the analysis,

$$\begin{aligned}
\widehat{X}_n - X_n^* &\geqslant -\left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1}\right) \ln N - \sum_{t=1}^{n} \Phi(\boldsymbol{p}_t,\, \eta_t,\, \boldsymbol{x}_t) \\
&\geqslant -2 \max\left\{2\, M_n \ln N,\, (1/C)\sqrt{V_n \ln N}\right\} - \sum_{t=1}^{n} \Phi(\boldsymbol{p}_t,\, \eta_t,\, \boldsymbol{x}_t) \\
&= -2 \max\left\{2\, M_n \ln N,\, (1/C)\sqrt{V_n \ln N}\right\} \\
&\quad - \sum_{t\in\mathcal{T}} \Phi(\boldsymbol{p}_t,\, \eta_t,\, \boldsymbol{x}_t) - \sum_{t\notin\mathcal{T}} \Phi(\boldsymbol{p}_t,\, \eta_t,\, \boldsymbol{x}_t)
\end{aligned}$$

where C is defined in (4.6), and $\mathcal{T}$ is the set of times rounds $t \geqslant 2$ when $\eta_t\, |x_{i,t}| \leqslant 1/2$ for all $i = 1, \dots, N$ (note that $1 \notin \mathcal{T}$ by definition).

Using the second bound of Lemma 4.4 on $t \in \mathcal{T}$ and the first bound of Lemma 4.4 on $t \notin \mathcal{T}$, which in this case reads $\Phi(\boldsymbol{p}_t,\, \eta_t,\, \boldsymbol{x}_t) \leqslant 2M_t$, we get

$$\widehat{X}_n - X_n^* \geqslant -2 \max\left\{2\, M_n \ln N,\, (1/C)\sqrt{V_n \ln N}\right\} - (e-2) \sum_{t\in\mathcal{T}} \eta_t \operatorname{Var} Z_t - \sum_{t\notin\mathcal{T}} 2\, M_t \tag{4.7}$$

(where $2M_1$ appears in the last sum). We first note that

$$\sum_{t\notin\mathcal{T}} M_t \leqslant \sum_{r=0}^{\lceil \log_2 \max\{M,1\}\rceil} 2^r \leqslant 2^{1+\lceil \log_2 \max\{M,1\}\rceil} \leqslant 4\, \max\{M, 1\}\,.$$

We now denote by T the first time step t when $V_t > M^2$. Using that $\eta_t \leqslant 1/2$ for all t and $V_T \leqslant 2M^2$, we get

$$\sum_{t\in\mathcal{T}} \eta_t \operatorname{Var} Z_t \leqslant M^2 + \sum_{t=T+1}^{n} \eta_t \operatorname{Var} Z_t\,. \tag{4.8}$$

We bound the sum using $\eta_t \leqslant C\sqrt{(\ln N)/V_{t-1}}$ for $t \geqslant 2$ (note that, for $t > T$, $V_{t-1} \geqslant V_T > M^2 > 0$). This yields

$$\sum_{t=T+1}^{n} \eta_t \operatorname{Var} Z_t \leqslant C\sqrt{\ln N} \sum_{t=T+1}^{n} \frac{V_t - V_{t-1}}{\sqrt{V_{t-1}}}\,.$$

Let $v_t = \operatorname{Var} Z_t = V_t - V_{t-1}$. Since $V_t \leqslant V_{t-1} + M^2$ and $V_{t-1} \geqslant M^2$, we have

$$\frac{v_t}{\sqrt{V_{t-1}}} = \frac{\sqrt{V_t} + \sqrt{V_{t-1}}}{\sqrt{V_{t-1}}} \left(\sqrt{V_t} - \sqrt{V_{t-1}}\right) \leqslant (\sqrt{2}+1)\left(\sqrt{V_t} - \sqrt{V_{t-1}}\right) \tag{4.9}$$

Therefore, using that $\sqrt{2}+1 = 1/(\sqrt{2}-1)$,

$$\sum_{t=T+1}^{n} \eta_t \operatorname{Var} Z_t \leqslant \frac{C\sqrt{\ln N}}{\sqrt{2}-1}\left(\sqrt{V_n} - \sqrt{V_T}\right) \leqslant \frac{C}{\sqrt{2}-1}\sqrt{V_n \ln N}\,.$$

When $\sqrt{V_n} \geqslant 2CM_n\sqrt{\ln N}$, using $M_n \geqslant M$ we have that $\widehat{X}_n - X_n^*$ is at least

$$\begin{aligned}
-\frac{2}{C}\sqrt{V_n \ln N} - \frac{C(e-2)}{\sqrt{2}-1}\sqrt{V_n \ln N} - 8\, \max\{M,\, 1\} - (e-2)M^2 \\
\geqslant -4\sqrt{V_n \ln N} - 8\, \max\{M,\, 1\} - M^2\,,
\end{aligned}$$

where we substituted the value of C and obtained a constant for the leading term equal to

$$2\frac{\sqrt{2(e-2)}}{\sqrt{\sqrt{2}-1}} \leqslant 3.75\ .$$

When $\sqrt{V_n} \leqslant 2CM_n\sqrt{\ln N}$, using $M_n \leqslant \max\{1,\, 2M\}$ we have that $\widehat{X}_n - X_n^*$ is at least

$$-8\,M\ln N - \frac{C^2 4(e-2)}{\sqrt{2}-1}\max\{1/2,\, M\}\ln N - 8\,\max\{M,\,1\} - (e-2)M^2$$
$$\geqslant -16\,\max\{M,\,1\}\ln N - 8\,\max\{M,\,1\} - M^2\ .$$

This concludes the proof. □

4. Applications

To demonstrate the usefulness of the bounds proven in Theorems 4.3 and 4.4 we show that they lead to several improvements or extensions of earlier results.

4.1. Improvements for loss games. Recall the definition of quadratic penalties Q_t^* in Section 2. In case of a loss game (i.e., all payoffs are non-positive), $Q_t^* \leqslant ML_t^*$, where L_t^* is the cumulative loss of the best action up to time t. Therefore, $\max_{s\leqslant n} Q_s^* \leqslant ML_n^*$ and the bound of Theorem 4.3 is at least as good as the family of bounds called "improvements for small losses" (see Section 3.1 in Chapter 2), whose main term is of the form $\sqrt{ML_n^*\ln N}$.

However, it is easy to exhibit examples where the new bound is far better by considering sequences of outcomes where there are some "outliers" among the $x_{i,t}$. These outliers may raise the maximum M significantly, whereas they have only little impact on the $\max_{s\leqslant n} Q_s^*$.

4.2. Using translations of payoffs. Recall that Z_t is the random variable which takes the value $x_{i,t}$ with probability $p_{i,t}$, for $i = 1, \ldots, N$. The main term of the bound stated in Theorem 4.4 contains $V_n = \operatorname{Var} Z_1 + \ldots + \operatorname{Var} Z_n$. Note that V_n is smaller than all quantities of the form

$$\sum_{t=1}^{n}\sum_{i=1}^{N} p_{i,t}\,(x_{i,t}-\mu_t)^2$$

where $(\mu_t)_{t\geqslant 1}$ is any sequence of real numbers which may be chosen in *hindsight*, as it is not required for the definition of the forecaster. (The minimal value of the expression is obtained for $\mu_t = \hat{x}_t$.) This gives us a whole family of upper bounds, and we may choose for the analysis the most convenient sequence of μ_t (see, for instance, Corollary 4.2 and Section 4.5 below).

To provide a concrete example, denote the effective range of the payoffs at time t by

$$R_t = \max_{i=1,\ldots,N} x_{i,t} - \min_{j=1,\ldots,N} x_{j,t}$$

and consider the choice $\mu_t = \min_{j=1,\ldots,N} x_{j,t} + R_t/2$. The next result improves on a result of Allenberg and Neeman [**AlNe04**] who show a regret bound, in terms of the cumulative effective range, whose main term is $5.7\sqrt{2(\ln N)M\sum_{t=1}^{n} R_t}$, for a given bound M over the payoffs. When the actual ranges are small, these bounds give a considerable advantage. Such a situation arises, for instance, in the setting of on-line portfolio selection, when we use a linear upper bound over the regrets (see, e.g., the EG strategy of Helmbold, Schapire, Singer and Warmuth [**HeScSiWa98**] with the viewpoint of Section 2 in Chapter 7).

COROLLARY 4.2. *The regret of the weighted majority forecaster run with variable learning rate (4.6) satisfies*

$$\widehat{X}_n - X_n^* \geqslant -2\sqrt{(\ln N)\sum_{t=1}^{n} R_t^2 - 16\,\max\{M,\,1\}\ln N - 8\,\max\{M,\,1\} - M^2}\,.$$

REMARK 4.3. (*About the leading constant in Corollary 4.2.*) The bound proposed by Corollary 4.2 shows that for an effective range of M, say if the payoffs all fall in $[0, M]$, the regret is lower bounded by a quantity equal to $-2M\sqrt{n \ln N}$ (a closer look at the proof of Theorem 4.4 shows that the constant factor may even be equal to 1.9). A careful modification of this proof would even bring the constant factor in the leading term as close to $2\sqrt{(e-2)}$ as wished. (The threshold at M^2 determining T by means of the V_1^{t-1} has been set quite arbitrarily. A value of aM^2, $a \geqslant 1$, would lead to a bound with a smaller constant factor in the leading term, at the cost of larger constant terms in the remainder constant term.)

The best leading constant for such bounds is, to our knowledge, $\sqrt{2}$ (see [**CeLu05**]), but the latter bound only applies to loss games or, with a simple reduction, to signed games for which we know beforehand the effective interval where the payoffs lie in. This is so because of two reasons. The first is that we lose a factor of 2 for the same reason that in Chapter 2, we lost a factor of 2 in Theorem 2.1 with respect to Theorem 2.5. We lose in addition an extra factor of $\sqrt{(e-2)/2}$ because of the difference between the bounds of Lemmas 4.4 and 4.5, since the factor $e-2$ may be improved into a $1/2$ in case of a loss game. However, as this factor $e-2$ is optimal (see Lemma A.3), we note that this second tiny gap is probably intrinsic. We do not know if the first one can be filled.

In conclusion, this shows nevertheless that the improved dependence in the bound does not come at a significant increase in the magnitude of the leading coefficient (and the same can be said when comparing the bound proposed by Corollary 4.3 below and the one of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**]).

We also note that using translations of payoffs for algorithm PROD-MQ, as suggested by Corollary 4.1, may be worthwhile as well, see Corollary 4.4 below. However, unlike the approach presented here for the weighted majority based forecaster, there the payoffs have to be explicitly translated by the forecaster, and thus, each translation rule corresponds to a different forecaster.

4.3. Improvements for one-sided games. The main drawback of V_n, used in Theorem 4.4, is that it is defined directly in terms of the forecaster's distributions $\boldsymbol{p}_t$. We now show how this dependence could be removed (see also Section 4.5 for another example). Assume $|x_{i,t}| \leqslant M$ for all t and i. The following corollary of Theorem 4.4 reveals that weighted majority suffers a small regret in one-sided games whenever $|X_n^*|$ or $Mn - |X_n^*|$ is small (where $|x_{i,t}| \leqslant M$ for all t and i); that is, whenever $|X_n^*|$ is very small or very large. Improvements of the same flavour were obtained by Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**] for loss games; however, their result cannot be converted in a straightforward manner to a corresponding useful result for gain games. Allenberg and Neeman [**AlNe04**] proved, in a gain game and for a related algorithm, a bound of the order of $11.4\sqrt{M}\min\bigl\{\sqrt{X_n^*},\,\sqrt{Mn - X_n^*}\bigr\}$. That algorithm was specifically designed to ensure a regret bound of this form, and is different from the algorithm whose performance we discussed before the statement of Corollary 4.2. Our weighted majority forecaster achieves a better bound, even though it was not directly constructed to do so.

COROLLARY 4.3 (Improvement for small or large payoffs in one-sided games). *Consider the weighted majority forecaster using the time-varying learning rate (4.6). Then, for all sequences of*

payoffs in a one-sided game (i.e., payoffs are all non-positive or all nonnegative),

$$\widehat{X}_n - X_n^* \geqslant -4\sqrt{|X_n^*|\left(M - \frac{|X_n^*|}{n}\right)\ln N} - 65\,\max\{1\,,M\}\max\{1\,,\ln N\} - 5\,M^2$$

where $M = \max_{t=1,\dots,n}\max_{i=1,\dots,N}|x_{i,t}|$.

PROOF. It suffices to give the proof for a gain game, as the bound of Theorem 4.4 is invariant under the change $\ell_{i,t} = M - x_{i,t}$, that converts bounded losses into bounded nonnegative payoffs. Since the payoffs are in $[0, M]$, we can write

$$\begin{aligned} V_n &\leqslant \sum_{t=1}^{n}\left(M\sum_{i=1}^{N}p_{i,t}x_{i,t} - \left(\sum_{i=1}^{N}p_{i,t}x_{i,t}\right)^2\right) = \sum_{t=1}^{n}(M-\widehat{x}_t)\widehat{x}_t \\ &\leqslant n\left(\frac{M\widehat{X}_n}{n} - \left(\frac{\widehat{X}_n}{n}\right)^2\right) = \widehat{X}_n\left(M - \frac{\widehat{X}_n}{n}\right) \end{aligned}$$

where we used the concavity of $x \mapsto Mx - x^2$. Assume that $\widehat{X}_n \leqslant X_n^*$ (otherwise the result is trivial). Then, Theorem 4.4 ensures that

$$\widehat{X}_n - X_n^* \geqslant -4\sqrt{X_n^*\left(M - \frac{\widehat{X}_n}{n}\right)\ln N} - \kappa$$

where $\kappa = 16\,\max\{M,\,1\}\ln N + 8\,\max\{M,\,1\} + M^2$. We solve for $\widehat{X}_n$ by using Lemma A.14, and obtain

$$\widehat{X}_n - X_n^* \geqslant -4\sqrt{X_n^*\left(M - \frac{X_n^*}{n} + \frac{\kappa}{n}\right)\ln N} - \kappa - 16\frac{X_n^*}{n}\ln N\,.$$

Using the crude upper bound $X_n^*/n \leqslant M$ and performing some simple algebra, we get the desired result. □

Quite surprisingly, a bound of the same form as the one shown in Corollary 4.3 can be derived as a consequence of Corollary 4.1, by using the payoff translation technique we discussed in the previous section.

COROLLARY 4.4 (Improvement for small or large payoffs in one-sided games). *If algorithm* PROD-MQ *is run using translated payoffs* $x_{k,t} - \widehat{x}_t$*, then for all sequences of payoffs in a one-sided game (i.e., payoffs are all non-positive or all nonnegative),*

$$\begin{aligned} \widehat{X}_n - X_n^* \geqslant -8\sqrt{2(\ln N)\max\{1,\,2M\min\{|X_n^*|,\,2Mn - |X_n^*|\}\}} \\ - 144M\,(2 + \log_4(2\,Mn))\,(1 + \ln N)\ , \end{aligned} \tag{4.10}$$

where $M = \max_{t=1,\dots,n}\max_{i=1,\dots,N}|x_{i,t}|$.

PROOF. It suffices to give the proof for a gain game, as the bound of Corollary 4.1 is invariant as well under the change $\ell_{i,t} = M - x_{i,t}$, that converts bounded losses into bounded nonnegative payoffs.

We apply the bound of Corollary 4.1, noting that, with the notation therein

$$\max_{s\leqslant n} R_s^* \leqslant \min\left\{M\left(X_n^* + \widehat{X}_n\right),\, M\left(Mn - X_n^* - \widehat{X}_n\right)\right\}\ . \tag{4.11}$$

Indeed, using that $(a-b)^2 \leqslant a^2+b^2$ for $a,\, b \geqslant 0$, we get on the one hand,

$$R_s^* \leqslant \sum_{t=1}^{s} x_{k_s^*,t}^2 + \widehat{x}_s^2 \leqslant M\left(X_{k_s^*,s} + \widehat{X}_s\right) \leqslant M\left(X_n^* + \widehat{X}_n\right) \;,$$

whereas on the other hand, the same techniques yield

$$R_s^* = \sum_{t=1}^{s} \left((M - x_{k_s^*,t}) - (M - \widehat{x}_s^2)\right)^2 \leqslant M\left((Ms - X_s^*) + \left(Ms - \widehat{X}_s\right)\right) \;.$$

Now, we note that for all s, $X_{s+1}^* \leqslant X_s^* + M$, and similarly, $\widehat{X}_{s+1} \leqslant \widehat{X}_s + M$. Thus we also have $\max_{s \leqslant n} R_s^* \leqslant M(Mn - X_n^* - \widehat{X}_n)$.

The proof is concluded by noting that we may assume that $X_n^* \geqslant \widehat{X}_n$, and therefore Corollary 4.1, combined with (4.11), yields

$$\widehat{X}_n \geqslant X_n^* - 8\sqrt{(\ln N)\max\left\{1,\, 2M\min\left\{X_n^*,\, Mn - \widehat{X}_n\right\}\right\}} - \kappa$$

where $\kappa = 12\,(2M)\,(2+\log_4(2Mn))\,(1+\ln N)$. Solving for $\widehat{X}_n$ by using Lemma A.14, and performing simple algebra concludes the proof. □

4.4. A simplified algorithm for bandit loss games. We indicate in this section a result that is not a direct consequence of Theorems 4.3 or 4.4. Rather, we derive it via an extension of Lemma 4.4, one of our key results at the core of the second-order analysis in Section 3.

Recall that payoffs $x_{i,t}$ in loss games are all non-positive. We use $\ell_{i,t} = -x_{i,t}$ to denote the loss of action i at time t. Similarly, $\widehat{\ell}_t = \ell_{1,t}p_{1,t} + \ldots + \ell_{N,t}p_{N,t}$ is the loss of the forecaster using $\boldsymbol{p}_t$ as probability assignment at time t. We make the simplifying assumption $\ell_{i,t} \in [0,1]$ for all i,t.

The bandit loss game (see Section 4 in Chapter 2 or [**AuCeFrSc02**] and references therein) is a loss game with the only difference that, at each time step t, the forecaster has no access to the loss vector $\ell_t = (\ell_{1,t}, \ldots, \ell_{N,t})$. Therefore, the loss $\widehat{\ell}_t$ cannot be computed and the individual losses $\ell_{i,t}$ cannot be used to adjust the probability assignment $\boldsymbol{p}_t$. The only information the forecaster receives at the end of each round t is the loss $\ell_{I_t,t}$, where I_t takes value i with probability $p_{i,t}$ for $i = 1, \ldots, N$.

In bandit problems and, more generally, in all incomplete information problems like label-efficient prediction or prediction with partial monitoring, a crucial point is to estimate the unobserved losses. In bandit algorithms based on weighted majority, this is usually done by shifting the probability distribution $\boldsymbol{p}_t$ so that all components are larger than a given threshold (see the forecasters proposed in Section 4 of Chapter 2, in Chapters 5 and 6, as well as those in Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**], Piccolboni and Schindelhauer [**PiSc01**], Cesa-Bianchi, Lugosi, and Stoltz [**CeLuSt04a, CeLuSt04b**] and Hart and Mas-Colell [**HaMa02**]).

Allenberg and Auer [**AlAu04**] apply the shifting technique to weighted majority obtaining, in bandit loss games, a regret bound of order $\sqrt{NL_n^* \ln N} + N\ln(n\,N)\ln n$ where L_n^* is the cumulative loss of the best action after n rounds,

$$L_n^* = \min_{j=1,\ldots,N} L_{j,n}, \qquad \text{where } \; L_{j,n} = \sum_{t=1}^{n} \ell_{j,t} \;.$$

(Note that using the results of [**AuCeFrSc02**] or Theorem 2.8), derived for gain games, one would only obtain $\sqrt{N(\ln N)n}$.)

We show here that *without any shifting*, a slight modification of weighted majority achieves an expected regret of order $N\sqrt{L_n^* \ln n} + N\ln n$, a bound which is also an improvement for small

Algorithm EXP3LIGHT.
Parameters: Real $\eta > 0$.
Initialization: $w_{i,1} = 1$ for $i = 1, \ldots, N$.
For $t = 1, 2, \ldots$

(1) draw action I_t according to the distribution $p_{i,t} = w_{i,t}/W_t$ for $i = 1, \ldots, N$, where $W_t = w_{1,t} + \ldots + w_{N,t}$, and incur loss $\ell_{I_t,t}$;
(2) let $\widetilde{\ell}_{i,t} = (\ell_{i,t}/p_{i,t})Z_{i,t}$ for $i = 1, \ldots, N$, where $Z_{i,t} = 1$ if $I_t = i$ and 0 otherwise;
(3) for each $i = 1, \ldots, N$ perform the update $w_{i,t+1} = w_{i,t}\, e^{-\eta\widetilde{\ell}_{i,t}}$.

FIGURE 1. Algorithm EXP3LIGHT for prediction in a multi-armed bandit setting.

losses. The new bound becomes better than the one by Allenberg and Auer when L_n^* is so small that $L_n^* = o((\ln n)^3)$. The bandit algorithm, which we call EXP3LIGHT, is described in Figure 1.

We start the analysis of EXP3LIGHT with a variant of Lemma 4.4 for loss games.

LEMMA 4.5. *For all losses $\ell_{i,t} \geqslant 0$, for all sets $S_t \subseteq \{1, \ldots, N\}$ and for all $\eta > 0$,*

$$\Phi(\boldsymbol{p}_t,\, \eta,\, -\ell_t) \leqslant \frac{\eta}{2} \sum_{i \in S_t} p_{i,t}\, \ell_{i,t}^2 + \sum_{i \in S_t} p_{i,t}\, \ell_{i,t}\ .$$

PROOF. We use the inequalities $e^{-x} \leqslant 1 - x + x^2/2$ for $x \geqslant 0$, and $\ln(1+u) \leqslant u$ for $u > -1$, to write

$$\begin{aligned}
\frac{1}{\eta_t} \ln\left(\sum_{i=1}^N p_{i,t} e^{-\eta_t \ell_{i,t}}\right) &\leqslant \frac{1}{\eta_t} \ln\left(\sum_{i \in S_t} p_{i,t} e^{-\eta_t \ell_{i,t}} + \sum_{i \notin S_t} p_{i,t}\right) \\
&\leqslant \frac{1}{\eta_t} \ln\left(\sum_{i \in S_t} p_{i,t}\left(1 - \eta_t \ell_{i,t} + \frac{\eta_t^2}{2}\ell_{i,t}^2\right) + \sum_{i \notin S_t} p_{i,t}\right) \\
&\leqslant -\sum_{i \in S_t} \ell_{i,t} p_{i,t} + \frac{\eta_t}{2} \sum_{i \in S_t} \ell_{i,t}^2 p_{i,t}\ ,
\end{aligned}$$

hence the result, by definition of Φ. □

Lemma 4.5 is applied as follows.

PROPOSITION 4.1. *Assume the forecaster* EXP3LIGHT *plays a bandit loss game, with losses bounded between 0 and 1. For all $\eta > 0$, the cumulative pseudo-loss of* EXP3LIGHT *satisfies*

$$\widetilde{L}_n - \widetilde{L}^* \leqslant \frac{(\ln N) + N(\ln n)}{\eta} + \frac{\eta}{2} N \widetilde{L}^* + \Delta_n$$

where $\widetilde{L}_n = \sum_{t=1}^n \sum_{i=1}^N p_{i,t}\widetilde{\ell}_{i,t}\ ,\quad \widetilde{L}_{k,n} = \sum_{t=1}^n \widetilde{\ell}_{k,t}\ ,\quad \widetilde{L}^* = \min_{k=1,\ldots,N} \widetilde{L}_{k,n}\ ,$
and Δ_n is a random variable with expectation less than $2\,N$.

PROOF. Choose $S_t = \left\{i \,:\, \widetilde{L}_{i,t} \leqslant \widetilde{L}^*\right\}$. We combine Lemmas 4.3 and 4.5 to get

$$\widetilde{L} \leqslant \widetilde{L}^* + \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^n \sum_{i \in S_t} p_{i,t}\, \widetilde{\ell}_{i,t}^2 + \sum_{t=1}^n \sum_{i \notin S_t} p_{i,t}\, \widetilde{\ell}_{i,t}\ . \tag{4.12}$$

Now, we note first that, by definition of the $\widetilde{\ell}_{i,t}$ and since $\ell_{i,t} \in [0,1]$,

$$\sum_{t=1}^{n} \sum_{i \in S_t} \widetilde{\ell}_{i,t}^2 \, p_{i,t} = \sum_{t=1}^{n} \sum_{i \in S_t} \ell_{i,t} \, \widetilde{\ell}_{i,t} \leqslant \sum_{t=1}^{n} \sum_{i \in S_t} \widetilde{\ell}_{i,t} \leqslant N \widetilde{L}^* \,,$$

where the last inequality is true by definition of the S_t.

To bound the second double sum of (4.12), note that the sets S_t are monotone decreasing; i.e., $i \notin S_t$ implies $i \notin S_r$ for all $r > t$. Let $T_i = \min \left\{ t \,:\, \widetilde{L}_{i,t} > \widetilde{L}^* \right\}$. We determine how large t can grow before $p_{i,t}$ becomes negligible. For each i such that $T_i < n$ and for each $t \geqslant T_i + 1$, we have, using $W_t \geqslant e^{-\eta \widetilde{L}^*}$,

$$p_{i,t} = \frac{w_{i,t}}{W_t} \leqslant \frac{w_{i,t}}{e^{-\eta \widetilde{L}^*}} = \exp\left(\eta \widetilde{L}^* - \eta \left(\widetilde{L}_{i,T_i} + \sum_{s=T_i+1}^{t-1} \widetilde{\ell}_{i,s} \right) \right) \,.$$

Thus, $p_{i,t} \leqslant 1/n$ whenever $\sum_{s=T_i+1}^{t-1} \widetilde{\ell}_{i,s} \geqslant (\ln n)/\eta + \widetilde{L}^* - \widetilde{L}_{i,T_i}$. Thus, let $T_i' > T_i$ be the last t such that $\sum_{s=T_i+1}^{t} \ell_{i,s} < (\ln n)/\eta + \widetilde{L}^* - \widetilde{L}_{i,T_i}$. We have

$$\begin{aligned}
\sum_{t=1}^{n} \sum_{i \notin S_t} p_{i,t} \widetilde{\ell}_{i,t} &\leqslant \sum_{i=1}^{N} p_{i,T_i} \widetilde{\ell}_{i,T_i} + \sum_{i=1}^{N} \sum_{t=T_i+1}^{T_i'} p_{i,t} \widetilde{\ell}_{i,t} + \sum_{i=1}^{N} \sum_{t=T_i'+1}^{n} p_{i,t} \widetilde{\ell}_{i,t} \\
&\leqslant \sum_{i=1}^{N} p_{i,T_i} \widetilde{\ell}_{i,T_i} + \sum_{i=1}^{N} \left(\frac{\ln n}{\eta} + \widetilde{L}^* - \widetilde{L}_{i,T_i} \right) + \sum_{i=1}^{N} \sum_{t=T_i'+1}^{n} \frac{\widetilde{\ell}_{i,t}}{n} \\
&\leqslant \Delta_n + \frac{N \ln n}{\eta}
\end{aligned}$$

where we used $\widetilde{L}^* - \widetilde{L}_{i,T_i} \leqslant 0$ by definition of T_i, and denoted

$$\Delta_n = \sum_{i=1}^{N} p_{i,T_i} \widetilde{\ell}_{i,T_i} + \sum_{i=1}^{N} \sum_{t=T_i'+1}^{n} \frac{\widetilde{\ell}_{i,t}}{n} \,.$$

The expectation of Δ_n is indeed less than $2\,N$. □

We are now ready to prove our main bandit result, thanks to a doubling trick (see Section 2.2 in Chapter 2). Note that the quantities $\mathbb{E}[L_n^*]$ may be replaced by simply L_n^* in case of an oblivious opponent, see Section 1.4 in Chapter 2.

THEOREM 4.5. *Consider the forecaster that runs algorithm* EXP3LIGHT *in epochs as follows. In each epoch $r = 0, 1 \ldots$ the algorithm uses*

$$\eta_r = \sqrt{\frac{2\left((\ln N) + N \ln n\right)}{N 4^r}}$$

and epoch r stops whenever the estimate $\widetilde{L}^*$ in this epoch is larger than 4^r. For any bandit loss game with losses bounded between 0 and 1, the expected cumulative loss of this forecaster satisfies

$$\begin{aligned}
&\max_{j=1,\ldots,N} \mathbb{E}\left[\sum_{t=1}^{n} \ell_{I_t,t} - \ell_{j,t} \right] \\
&\quad \leqslant 2\sqrt{2\left((\ln N) + N \ln n\right) N \left(1 + 3 \min_{j=1,\ldots,N} E\left[L_{j,n}\right] \right)} + (2N+1)(1 + \log_4(3n+1)) \,.
\end{aligned}$$

REMARK 4.4. Though we only prove bounds in expectation, there might be a chance that the techniques used in [**AuCeFrSc02**], namely, second-order martingale inequalities and a prior shifting of the estimated losses, lead to bounds that hold with overwhelming probability for a variant of EXP3LIGHT which still uses no shifting over the probability distributions.

PROOF. As usual, we denote by R the index of the last epoch and by t_r the last time round of each epoch r ($t_R = n$). We also denote by $\widetilde{L}^{*,r}$ the smallest cumulative estimated loss among the estimated losses of the experts. For all r, for all time intervals from t_{r-1} to $t_r - 1$, we may use Proposition 4.1 to bound the regret. We bound the instantaneous regrets at times t_r separately. Using in addition that the sum of minima is less than the minimum of the sums, we get

$$\begin{aligned}\widetilde{L}_n &\leqslant \widetilde{L}^* + \sum_{r=0}^{R} \sqrt{2\left((\ln N) + N \ln n\right) N\, 4^r} + \sum_{r=0}^{R} \left(\widetilde{\ell}_{t_r} + \Delta_r\right) \\ &\leqslant \widetilde{L}^* + 2^{R+1} \sqrt{2\left((\ln N) + N \ln n\right) N} + \sum_{r=0}^{R} \left(\widetilde{\ell}_{t_r} + \Delta_r\right) \ ,\end{aligned}$$

where the Δ_r are random variables with expectation less than $2N$. We now use that, when $R \geqslant 1$,

$$\widetilde{L}^* \geqslant \sum_{r=0}^{R-1} \widetilde{L}^{*,r} \geqslant \frac{4^R - 1}{3} \ ,$$

this inequality being still true for $R = 0$. The proof is concluded by two applications of Jensen's inequality, applied first to $2^{R+1} \leqslant 2\sqrt{3\widetilde{L}^* + 1}$. Second, we have to bound R. We note that $R \leqslant \log_4(1 + 3\widetilde{L}^*)$, so that R is in expectation less than $\log_4(3n + 1)$. □

4.5. Fast rates in prediction with expert advice. We end this application section by an example of fast rates of convergence in (randomized) prediction with expert advice. We call a fast rate of convergence any convergence rate faster than the general guaranteed $1/\sqrt{n}$ convergence rate for the sequence of per-round regrets. This issue has been under the focus of attention for several years now in classification, see the discussion at the end of Chapter 1 and the references therein (above all the survey paper by Boucheron, Bousquet and Lugosi [**BoBoLu05**], and the recent paper of Steinvart and Scovel [**StSc05**], who deal with support vector machines). Furthermore, note that fast rate results are already known in prediction of individual sequences, but only for some classes of loss functions (among them, the so-called exp-concave ones), see Haussler, Kivinen and Warmuth [**HaKiWa98, KiWa99**], Vovk [**Vov98, Vov01**]. We want to deal with arbitrary, unspecified, loss functions.

Our derivation illustrates another way to solve for the regrets, and to remove the dependency of the bound of Theorem 4.4 in the forecaster's distributions $\boldsymbol{p}_t$. We consider a loss game, for instance. Condition (4.13) stated below is the equivalent of another one in classification, asserting that fast rates are achieved as soon as the variances of the (shifted) base classifiers can be upper bounded their respective expected risks, see [**BoBoLu05**]. There the key second-order lemma is Bernstein's inequality, and we believe that our variance-based tuning (4.6) of the weighted majority algorithm is the right second-order counterpart.

PROPOSITION 4.2. *Denote by j_n^* the index of the expert $i = 1, \ldots, N$ achieving the minimal cumulative loss, i.e., such that $L_{j_n^*,n} = L_n^*$. Assume that the loss sequence is such that the forecaster behaves on it in a way such that there exists a constant γ and an integer n_0, such that for*

all $n \geqslant n_0$,

$$\sum_{t=1}^{n} \sum_{i=1}^{N} p_{i,t} \left(\ell_{i,t} - \ell_{j_n^*,t}\right)^2 \leqslant \gamma \left(\bar{L}_n - L_n^*\right) \ , \tag{4.13}$$

where

$$\bar{L}_n = \sum_{t=1}^{n} \ell_t(\boldsymbol{p}_t) = \sum_{t=1}^{n} \sum_{i=1}^{N} p_{i,t} \ell_{i,t} \ .$$

Then, the per-round regret of the forecaster on this sequence is bounded, for $n \geqslant n_0$, by

$$\frac{1}{n} \left(\bar{L}_n - L_n^*\right) \leqslant \frac{1}{n} \left(16\gamma + 20\sqrt{\gamma} + 25\right) \max\left\{\ln N,\, 1\right\} \ .$$

The proof is a simple combination of (4.13) with Theorem 4.4, together with the results of Section 4.2 (and the choice $\mu_t = \ell_{j_n^*,t}$). Solving for the regrets thanks to Lemma A.14 and over-approximating yields the result.

As we indicated above, condition (4.13) is the counterpart of the usual condition used in classification to get fast rates (see, e.g., [**BoBoLu05**, Section 5.2]). To show that (4.13) is indeed meaningful in an individual sequence setting as well, we consider the following example.

EXAMPLE 4.1. We have two experts A and B, and they suffer the losses (for integers $t \geqslant 0$) $\ell_{A,3t+1} = 1$, $\ell_{A,3t+2} = \ell_{A,3t+3} = 0$, and $\ell_{B,3t+1} = 0$, $\ell_{B,3t+2} = \ell_{B,3t+3} = 1$. j_n^* is A, and it is easy to see that condition (4.13) holds for the forecaster of Theorem 4.4, at least for integers n multiple of 3, with $\gamma = 3$. (See the proof below for the details.)

We note here that L_n^* grows linearly in n in the example above, so that the $\sqrt{L_n^* \ln N}$ upper bound for the weighted majority forecaster of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**] is of the order of $\sqrt{n}$. (Even the bounds for PROD-MQ proposed by Theorem 4.1 are of the order of $\sqrt{n}$, since $Q_{B,n} \geqslant Q_{A,n} = n/3$.) However, direct computations show that the forecaster of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**] suffers a loss bounded by a constant. The main improvement of the second-order analysis conducted for weighted majority in Section 3 is thus that the new bounds reflect in a sharper way its behavior than the previous bounds.

PROOF. We first note that for the losses of Example 4.1, (4.13) rewrites (for an integer $n = 3n'$ multiple of n) as

$$\sum_{s=1}^{n'} p_{B,3s-2} + p_{B,3s-1} + p_{B,3s} \leqslant \gamma \sum_{s=1}^{n'} -p_{B,3s-2} + p_{B,3s-1} + p_{B,3s} \ , \tag{4.14}$$

and prove now that it is indeed satisfied, with $\gamma = 3$. We take the elements of the sequence of weights of expert B three by three. Fix a positive integer s. We note that these weights are given by

$$\begin{aligned}
p_{B,3s-2} &= \frac{e^{-\eta_{3s-2}\, 2(s-1)}}{e^{-\eta_{3s-2}\, 2(s-1)} + e^{-\eta_{3s-2}\, (s-1)}} \ , \\
p_{B,3s-1} &= \frac{e^{-\eta_{3s-1}\, 2(s-1)}}{e^{-\eta_{3s-1}\, 2(s-1)} + e^{-\eta_{3s-1}\, s}} \ , \\
p_{B,3s} &= \frac{e^{-\eta_{3s}\, (2s-1)}}{e^{-\eta_{3s}\, (2s-1)} + e^{-\eta_{3s}\, s}} \ ,
\end{aligned}$$

where the $\eta_{s'}$ are defined in (4.6). In particular, we also recall that the sequence $(\eta_{s'})$ of the tuning parameters is non-increasing. This shows that both $p_{B,3s-2}$ and $p_{B,3s}$ are less than $p_{B,3s-1}$, so that $p_{B,3s-2} + p_{B,3s-1} + p_{B,3s} \leqslant 3p_{B,3s-1}$. To prove the claim (4.14), it thus suffices to show that $p_{B,3s-2} \leqslant p_{B,3s}$.

To that end, we note that $p_{B,3s-2} \leqslant p_{B,3s}$ is equivalent, by a reduction to the same denominator, to

$$e^{-\eta_{3s-2}\,2(s-1)} e^{-\eta_{3s}\,s} \leqslant e^{-\eta_{3s-2}\,(s-1)} e^{-\eta_{3s}\,(2s-1)} ,$$

or simply, $\eta_{3s-2}\,(s-1) \geqslant \eta_{3s}\,(s-1)$. But the latter is true, since the tuning parameters are non-increasing. □

5. Discussion and open problems

Though the results of Sections 2 and 3 cannot be easily compared in terms of expected regret bounds (see however Remark 4.5 below for a comparison of the non-expected regrets), the underlying algorithms work indifferently for loss games, gain games, and signed games. Note however that the bounds proposed by Corollary 4.1 and by Theorem 4.4 both lead to improvement for small or large payoffs in one-sided games, see Corollaries 4.3 and 4.4. In addition, they are both stable under many transformations, such as translations or changes of signs. Consequently, and most importantly, they are invariant under the change $\ell_{i,t} = M - x_{i,t}$, that converts bounded nonnegative payoffs into bounded losses, and vice versa. However, the occurrence of terms like $\max\{M,\, 1\}$ and M^2 makes these bounds not stable under rescaling of the payoffs. This means that if the payoffs are all multiplied by a positive number α (which may be more or less than 1), then the bounds on the regret are not necessarily multiplied by the same quantity α.

Modifying the proof of Theorem 4.4 we also obtained a regret bound equal to $-4\sqrt{V_n \ln N} - 16\, M \ln N - 8\, M - 2M \log M^2/V_1$. This bound is indeed stable under rescalings and improves on Theorem 4.4 for instance when M much smaller than 1, or even when M is large and V_1 is not too small. We hope that the unconvenient factor $1/V_1$ could be removed soon.

A practical advantage of the weighted majority forecaster is that its update rule is completely incremental and never needs to reset the weights. This in contrast to the forecaster PROD-MQ of Theorem 4.3 that uses a nested doubling trick. On the other hand, the bound proposed in Theorem 4.4 is not in closed form, as it still explicitly depends through V_n on the forecaster's rewards $\widehat{x}_t$. We therefore need to solve for the regrets, see, for instance, Corollary 4.3 or Section 4.5. Finally, we also noted in Section 4.2 that the weighted majority forecaster update is invariant under translations of the payoffs, whereas each translation rule for the payoffs leads to a different version of PROD-MQ. In practice, it may be difficult to determine beforehand what a good translation could be. Corollaries 4.1 and 4.4, as well as Remark 4.1, indicate general efficient translation rules.

OPEN QUESTION 4.2. Several issues are left open. The following list mentions some of them.

- Design and analyze incremental updates for the forecaster PROD(η) of Section 2.
- Obtain second order bounds with updates that are not multiplicative; for instance, updates based on the polynomial potentials (see Section 2.3 in Chapter 2 or [**CeLu03**]). These updates could be used as basic ingredients to get forecasters suited for bandit, label-efficient or partial monitoring prediction, and achieving the optimal rates. Note that in Section 4 of Chapter 2, as well as in Chapters 5 and 6, we thus had to use exponentially weighted averages.
- Extend the analysis of PROD-MQ to obtain an oracle inequality of the form

  $$\widehat{X}_n \geqslant \max_{k=1,\dots,N} \left(X_{k,n} - \gamma_1 \sqrt{Q_{k,n} \ln N} \right) - \gamma_2 M \ln N$$

  where γ_1 and γ_2 are absolute constants. Inequalities of this form can be viewed as game-theoretic versions of the model selection bounds in statistical learning theory.
- Obtain second-order bounds for weighted majority and PROD-MQ that are stable under rescaling.

REMARK 4.5. *(Refined bounds for non-expected regret.)* In this chapter, we focused on improved bounds for expected regret. However, recall from Chapter 2, Section 3.3.2, that in general, the non-expected cumulative regret of any forecaster is bounded by the expected cumulative regret with probability $1 - \delta$ up to deviations of the order of $\sqrt{V_n \ln(n/\delta)} + M \ln(n/\delta)$, see Corollary A.1. These deviations are of the same order of magnitude as the bound of Theorem 4.4. Unless we are able to apply a sharper concentration result than Bernstein's inequality, no further refinement of the above bounds is worthwhile. In particular, in view of the deviations from the expectations, we may prefer the results of Section 3 to those of Section 2.

Appendix: Proof of Lemma 4.3

We first note that Jensen's inequality implies that Φ is nonnegative. The proof below is a simple modification of an argument first proposed in [**AuCeGe02**]. Note that we consider real-valued (non necessarily nonnegative) payoffs in what follows. For $t = 1, \ldots, n$, we rewrite $p_{i,t} = w_{i,t}/W_t$, where $w_{i,t} = e^{\eta_t X_{i,t-1}}$ and $W_t = \sum_{j=1}^N w_{j,t}$ (the payoffs $X_{i,0}$ are understood to equal 0, and thus, η_1 may be any positive number satisfying $\eta_1 \geqslant \eta_2$). Use $w'_{i,t} = e^{\eta_{t-1} X_{i,t-1}}$ to denote the weight $w_{i,t}$ where the parameter η_t is replaced by η_{t-1}. The associated normalization factor will be denoted by $W'_t = \sum_{j=1}^N w'_{j,t}$. Finally, we use j_t^* to denote the expert with the largest cumulative payoff after the first t rounds (ties are broken by choosing the expert with smallest index). That is, $X_{j_t^*,t} = \max_{i \leqslant N} X_{i,t}$. We also make use of the following technical lemma.

LEMMA 4.6 (Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**]). *For all $N \geqslant 2$, for all $\beta \geqslant \alpha \geqslant 0$, and for all $d_1, \ldots, d_N \geqslant 0$ such that $\sum_{i=1}^N e^{-\alpha d_i} \geqslant 1$,*

$$\ln \frac{\sum_{i=1}^N e^{-\alpha d_i}}{\sum_{j=1}^N e^{-\beta d_j}} \leqslant \frac{\beta - \alpha}{\alpha} \ln N \, .$$

PROOF (OF LEMMA 4.6). We begin by writing

$$\begin{aligned} \ln \frac{\sum_{i=1}^N e^{-\alpha d_i}}{\sum_{j=1}^N e^{-\beta d_j}} &= \ln \frac{\sum_{i=1}^N e^{-\alpha d_i}}{\sum_{j=1}^N e^{(\alpha-\beta) d_j} e^{-\alpha d_j}} \\ &= -\ln \mathbb{E}\left[e^{(\alpha-\beta)D}\right] \\ &\leqslant (\beta - \alpha)\mathbb{E}\left[D\right] \end{aligned}$$

where we applied Jensen inequality to the random variable D taking value d_i with probability $e^{-\alpha d_i} / \sum_{j=1}^N e^{-\alpha d_j}$ for each $j = 1, \ldots, N$. Since D takes at most N distinct values, its entropy $H(D)$ is at most $\ln N$. Therefore

$$\begin{aligned} \ln N \geqslant H(D) &= \frac{\sum_{i=1}^N e^{-\alpha d_i}}{\sum_{j=1}^N e^{-\beta d_j}} \left(\alpha d_i + \ln \sum_{j=1}^N e^{-\beta d_j} \right) \\ &= \alpha \mathbb{E}\left[D\right] + \ln \sum_{j=1}^N e^{-\beta d_j} \geqslant \alpha \mathbb{E}\left[D\right] \end{aligned}$$

where the last inequality holds since $\sum_{i=1}^N e^{-\alpha d_i} \geqslant 1$. Hence $\mathbb{E}\left[D\right] \leqslant (\ln N)/\alpha$. As $\beta > \alpha$ by hypothesis, we can substitute the bound on $\mathbb{E}\left[D\right]$ in the upper bound above and conclude the proof. □

PROOF (OF LEMMA 4.3). As it is usual in the analysis of the exponentially weighted average predictor, we study the evolution of $\ln(W_{t+1}/W_t)$, see the proof of Theorem 2.1. However, here we need to couple this term with $\ln(w_{j_{t-1}^*,t}/w_{j_t^*,t+1})$ including in both terms the time-varying parameters η_t, η_{t+1}. Tracking the currently best expert j_t^* is used to lower bound the weight $\ln(w_{j_t^*,t+1}/W_{t+1})$. In fact, the weight of the overall best expert (after n rounds) could get arbitrarily small during the prediction process. We thus obtain the following

$$\begin{aligned} &\frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}}{W_t} - \frac{1}{\eta_{t+1}} \ln \frac{w_{j_t^*,t+1}}{W_{t+1}} \\ &= \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \ln \frac{W_{t+1}}{w_{j_t^*,t+1}} + \frac{1}{\eta_t} \ln \frac{w'_{j_t^*,t+1}/W'_{t+1}}{w_{j_t^*,t+1}/W_{t+1}} + \frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}/W_t}{w'_{j_t^*,t+1}/W'_{t+1}} \\ &= (A) + (B) + (C) \, . \end{aligned}$$

We now bound separately the three terms on the right-hand side. The term (A) is easily bounded by using $\eta_{t+1} \leqslant \eta_t$ and using the fact that j_t^* is the index of the expert with largest payoff after the first t rounds. Therefore, $w_{j_t^*,t+1}/W_{t+1}$ must be at least $1/N$. Thus we have

$$(A) = \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \ln \frac{W_{t+1}}{w_{j_t^*,t+1}} \leqslant \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \ln N \; .$$

We proceed to bounding the term (B) as follows

$$\begin{aligned}(B) &= \frac{1}{\eta_t} \ln \frac{w'_{j_t^*,t+1}/W'_{t+1}}{w_{j_t^*,t+1}/W_{t+1}} = \frac{1}{\eta_t} \ln \frac{\sum_{i=1}^N e^{-\eta_{t+1}(X_{j_t^*,t} - X_{i,t})}}{\sum_{j=1}^N e^{-\eta_t(X_{j_t^*,t} - X_{j,t})}} \\ &\leqslant \frac{\eta_t - \eta_{t+1}}{\eta_t \eta_{t+1}} \ln N = \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \ln N\end{aligned}$$

where the inequality is proven by applying Lemma 4.6 with $d_i = X_{j_t^*,t} - X_{i,t}$. Note that $d_i \geqslant 0$ since j_t^* is the index of the expert with largest payoff after the first t rounds and $\sum_{i=1}^N e^{-\eta_{t+1} d_i} \geqslant 1$ as for $i = j_t^*$ we have $d_i = 0$.
The term (C) is first split as follows,

$$(C) = \frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}/W_t}{w'_{j_t^*,t+1}/W'_{t+1}} = \frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}}{w'_{j_t^*,t+1}} + \frac{1}{\eta_t} \ln \frac{W'_{t+1}}{W_t} \; .$$

We bound separately each one of the two terms on the right-hand side. For the first one, we have

$$\frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}}{w'_{j_t^*,t+1}} = \frac{1}{\eta_t} \ln \frac{e^{\eta_t X_{j_{t-1}^*,t-1}}}{e^{\eta_t X_{j_t^*,t}}} = X_{j_{t-1}^*,t-1} - X_{j_t^*,t} \; .$$

The second term is handled by using the very definition of Φ,

$$\begin{aligned}\frac{1}{\eta_t} \ln \frac{W'_{t+1}}{W_t} = \frac{1}{\eta_t} \ln \frac{\sum_{i=1}^N w_{i,t} e^{\eta_t x_{i,t}}}{W_t} &= \frac{1}{\eta_t} \ln \sum_{i=1}^N p_{i,t} e^{\eta_t x_{i,t}} \\ &= \sum_{i=1}^N p_{i,t} x_{i,t} + \Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) \; .\end{aligned}$$

Finally, we substitute in the main equation the bounds on the first two terms (A) and (B), and the bounds on the two parts of the term (C). After rearranging we obtain

$$\begin{aligned}0 \leqslant \; & \left(X_{j_{t-1}^*,t-1} - X_{j_t^*,t}\right) + \sum_{i=1}^N p_{i,t} x_{i,t} + \Phi(\boldsymbol{p}_t, \eta_t, \boldsymbol{x}_t) \\ & - \frac{1}{\eta_{t+1}} \ln \frac{w_{j_t^*,t+1}}{W_{t+1}} + \frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}}{W_t} \\ & + 2\left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right) \ln N \; .\end{aligned}$$

We apply the above inequalities to each $t = 1, \ldots, n$ and sum up using

$$\begin{aligned}&\sum_{t=1}^n X_{j_{t-1}^*,t-1} - X_{j_t^*,t} = - \max_{j=1,\ldots,N} X_{j,n} \; , \\ &\sum_{t=1}^n \left(-\frac{1}{\eta_{t+1}} \ln \frac{w_{j_t^*,t+1}}{W_{t+1}} + \frac{1}{\eta_t} \ln \frac{w_{j_{t-1}^*,t}}{W_t}\right) \leqslant -\frac{1}{\eta_1} \ln \frac{w_{j_0^*,1}}{W_1} = \frac{\ln N}{\eta_1}\end{aligned}$$

to conclude the proof. □

Part 2

Prediction with limited feedback

CHAPTER 5

Minimizing regret with label efficient prediction

We investigate label efficient prediction, a variant, proposed by Helmbold and Panizza, of the problem of prediction with expert advice. In this variant the forecaster, after guessing the next element of the sequence to be predicted, does not observe its true value unless he asks for it, which he cannot do too often. We determine matching upper and lower bounds for the best possible excess prediction error, with respect to the best possible constant predictor, when the number of allowed queries is fixed. We also prove that Hannan consistency, a fundamental property in game-theoretic prediction models, can be achieved by a forecaster issuing a number of queries growing to infinity at a rate just slightly faster than logarithmic in the number of prediction rounds.

Contents

This chapter is a joint work with Nicolò Cesa-Bianchi and Gábor Lugosi. It is based on the article [**CeLuSt05**], which is to appear in *IEEE Transactions on Information Theory* and was first presented at COLT'04 in the extended abstract [**CeLuSt04a**].

1. Introduction

We recall from Chapter 2 that prediction with expert advice, a framework introduced about fifteen years ago in learning theory, may be viewed as a direct generalization of the theory of repeated games, a field pioneered by Blackwell and Hannan in the mid-fifties. At a certain level of abstraction, the common subject of these studies is the problem of forecasting each element y_t of an unknown "target" sequence given the knowledge of the previous elements $y_1, \ldots, y_{t-1}$. The forecaster's goal is to predict the target sequence almost as well as any forecaster forced to use the same guess all the time. We call this the sequential prediction problem. To provide a suitable parameterization of the problem, we assume that the set from which the forecaster picks

LABEL EFFICIENT PREDICTION

Parameters: number N of actions, outcome space $\mathcal{Y}$, loss function ℓ, query rate $\mu : \mathbb{N} \to \mathbb{N}$.

For each round $t = 1, 2, \ldots$

(1) the environment chooses the next outcome $y_t \in \mathcal{Y}$ without revealing it;
(2) the forecaster chooses an action $I_t \in \{1, \ldots, N\}$;
(3) each action i incurs loss $\ell(i, y_t)$;
(4) if less than $\mu(t)$ queries have been issued so far, the forecaster may issue a new query to obtain the outcome y_t; if no query is issued then y_t remains unknown.

FIGURE 1. Label efficient prediction as a game between the forecaster and the environment.

its guesses is finite, of size $N > 1$, while the set to which the target sequence elements belong may be of arbitrary cardinality. A real-valued bounded loss function ℓ is then used to quantify the discrepancy between each outcome y_t and the forecaster's guess for y_t. The pioneering results of Hannan's [**Han57**] and Blackwell [**Bla56**] showed that randomized forecasters exist whose excess cumulative loss (or regret), with respect to the loss of any constant forecaster, grows sub-linearly in the length n of the target sequence, and this holds for any individual target sequence. In particular, both Blackwell and Hannan found the optimal growth rate, $\Theta(\sqrt{n})$, of the regret as a function of the sequence length n when no assumption other than boundedness is made on the loss ℓ. Only relatively recently, Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire, and Warmuth [**CeFrHaHeScWa97**] have revealed that the correct dependence on N in the minimax regret rate is $\Theta(\sqrt{n \ln N})$.

Game theorists, information theorists, and learning theorists, who independently studied the sequential prediction model, addressed the fundamental question of whether a sub-linear regret rate is achievable in case the past outcomes $y_1, \ldots, y_{t-1}$ are not entirely accessible when computing the guess for y_t. In this work we investigate a variant of sequential prediction known as *label efficient prediction*. In this model, originally proposed by Helmbold and Panizza [**HePa97**], after choosing its guess at time t the forecaster decides whether to query the outcome y_t. However, the forecaster is limited in the number $\mu(n)$ of queries he can issue within a given time horizon n. In the case $n \to \infty$, we prove that Hannan consistency (i.e., regret growing sub-linearly with probability one) can be achieved under the only condition $\mu(n)/(\log(n) \log\log(n)) \to \infty$. Moreover, in the finite-horizon case, we show that any forecaster issuing at most $m = \mu(n)$ queries must suffer a regret of at least order $n\sqrt{(\ln N)/m}$ on some outcome sequence of length n, and we show a randomized forecaster achieving this regret to within constant factors.

The problem of label efficient prediction is closely related to other frameworks in which the forecaster has a limited access to the outcomes. Examples include prediction under partial monitoring (see Chapter 6, see also, e.g., Mertens, Sorin, and Zamir [**MeSoZa94**], Rustichini [**Rus99**], Piccolboni, and Schindelhauer [**PiSc01**], Mannor and Shimkin [**MaSh03**], Cesa-Bianchi, Lugosi, and Stoltz [**CeLuSt04b**]), the multi-armed bandit problem (see Section 4 in Chapter 2, see also Baños [**Ban68**], Megiddo [**Meg80**], Foster and Vohra [**FoVo98**], Hart and Mas Colell [**HaMa02**], Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**], and Auer [**Aue02**]), and the "apple tasting" problem proposed by Helmbold, Littlestone, and Long [**HeLiLo00**].

2. Sequential prediction and the label efficient model

We recall here the notation introduced in Chapter 2. The sequential prediction problem is parameterized by a number $N > 1$ of player actions, by a set $\mathcal{Y}$ of outcomes, and by a loss

function ℓ. The loss function has domain $\{1,\dots,N\} \times \mathcal{Y}$ and takes values in a bounded real interval, say $[0,1]$. Given an unknown mechanism generating a sequence $y_1, y_2, \dots$ of elements from $\mathcal{Y}$, a prediction strategy, or forecaster, chooses an action $I_t \in \{1,\dots,N\}$ incurring a loss $\ell(I_t, y_t)$. A crucial assumption in this model is that the forecaster can choose I_t only based on information related to the past outcomes $y_1,\dots,y_{t-1}$. That is, the forecaster's decision must not depend on any of the future outcomes. In the label efficient model, after choosing I_t the forecaster decides whether to issue a query to access y_t. If no query is issued, then y_t remains unknown. In other words, I_t does not depend on all the past outcomes $y_1,\dots,y_{t-1}$, but only on the queried ones. The label efficient model is best described as a repeated game between the forecaster, choosing actions, and the environment, choosing outcomes (see Figure 1).

The cumulative loss of the forecaster on a sequence $y_1, y_2, \dots$ of outcomes is denoted by

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(I_t, y_t) \qquad \text{for } n \geqslant 1.$$

As the forecasting strategies we consider may be randomized, each I_t is viewed as a random variable. All probabilities and expectations are understood with respect to the σ-algebra of events generated by the sequence of random choices of the forecaster.

We compare the forecaster's cumulative loss $\widehat{L}_n$ with those of the N constant forecasters $L_{i,n} = \ell(i, y_1) + \ldots + \ell(i, y_n)$, $i = 1,\dots,N$.

In this chapter we devise label efficient forecasting strategies whose expected regret

$$\max_{i=1,\dots,N} \mathbb{E}\Big[\widehat{L}_n - L_{i,n}\Big]$$

grows sub-linearly in n for any sequence $y_1, y_2, \dots$ of outcomes, that is, for any strategy of the environment whenever $\mu(n) \to \infty$. Note that the quantities $L_{1,n},\dots,L_{N,n}$ are random. Indeed, as argued in Section 3 (see also Section 1.4 in Chapter 2), in general the outcomes y_t may depend on the forecaster's past random choices. Via a more refined analysis, we also prove the stronger result

$$\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n} = o(n) \qquad \text{a.s.} \tag{5.1}$$

for any sequence $y_1, y_2, \dots$ of outcomes and whenever $\mu(n)/(\log(n)\log\log(n)) \to \infty$. The almost sure convergence is with respect to the auxiliary randomization the forecaster has access to. Property (5.1), known as *Hannan consistency* in game theory (see Section 3.3 in Chapter 2), rules out the possibility that the regret is much larger than its expected value with a significant probability.

REMARK 5.1. *(Prediction with expert advice.)* The results of this chapter extend straightforwardly to the case when the forecaster is supplied with expert advice (see the appendix of Chapter 2). The case of actions corresponds to constant experts. This is so because here, all we need is unbiased estimates of the losses, and the way we build them in the next section does not depend on the actual way the losses are computed. This is in contrast with the results for prediction with partial monitoring (see Chapter 6). There, the required assumption (6.1) prevents such an extension.

We could also apply the label efficient methodology to the sequential investment in the stock market problem described in Chapter 7, and derive label efficient variants of the EG and B1EXP strategies defined there. This is so because these strategies rely on prediction-with-expert-advice techniques, see, in particular, (7.2) and the comments after it.

Parameters: Real numbers $\eta > 0$ and $0 \leqslant \varepsilon \leqslant 1$.
Initialization: $\boldsymbol{w}_1 = (1, \ldots, 1)$.
For each round $t = 1, 2, \ldots$

(1) draw an action I_t from $\{1, \ldots, N\}$ according to the distribution

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}}\,, \qquad i = 1, \ldots, N\,;$$

(2) draw a Bernoulli random variable Z_t such that $\mathbb{P}[Z_t = 1] = \varepsilon$;
(3) if $Z_t = 1$ then obtain y_t and compute

$$w_{i,t+1} = w_{i,t}\, e^{-\eta\, \ell(i, y_t)/\varepsilon} \qquad \text{for each } i = 1, \ldots, N$$

else, let $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t$.

FIGURE 2. The label efficient exponentially weighted average forecaster.

3. A label efficient forecaster

We start by considering the finite-horizon case in which the forecaster's goal is to control the regret after n predictions, where n is fixed in advance. In this restricted setup we also assume that at most $m = \mu(n)$ queries can be issued, where μ is the query rate function. However, we do not impose any further restriction on the distribution of these m queries in the n time steps, that is, $\mu(t) = m$ for $t = 1, \ldots, n$. We introduce a simple forecaster whose expected regret is bounded by $n\sqrt{2(\ln N)/m}$. We then prove that the regret is indeed of the same order, with high probability. (Thus, if $m = n$, we recover the orders of magnitude in n and N of the optimal bound for prediction with expert advice under full monitoring, see Section 5 in Chapter 2.)

It is easy to see that in order to achieve a nontrivial performance, a forecaster must use randomization in determining whether a label should be revealed or not. It turns out that a simple biased coin is sufficient for our purpose. The strategy we propose, sketched in Figure 2, uses an i.i.d. sequence $Z_1, Z_2, \ldots, Z_n$ of Bernoulli random variables such that $\mathbb{P}[Z_t = 1] = 1 - \mathbb{P}[Z_t = 0] = \varepsilon$ and asks the label y_t to be revealed whenever $Z_t = 1$. Here $\varepsilon > 0$ is a parameter of the strategy. (Typically, we take $\varepsilon \approx m/n$ so that the number of solicited labels during n rounds is about m. Note that this way the forecaster may ask the value of more than m labels, but we ignore this detail as it can be dealt with by a simple adjustment.) Our label efficient forecaster uses the *estimated losses*

$$\widetilde{\ell}(i, y_t) \stackrel{\text{def}}{=} \begin{cases} \ell(i, y_t)/\varepsilon & \text{if } Z_t = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{N,t})$ and denote by v_1^t the prefix $(v_1, \ldots, v_t)$ of a given sequence $(v_1, v_2, \ldots)$. Then

$$\mathbb{E}[\,\widetilde{\ell}(i, y_t) \mid Z_1^{t-1}, I_1^{t-1}] \;=\; \ell(i, y_t)\,, \tag{5.2}$$

$$\mathbb{E}[\,\widetilde{\ell}(\boldsymbol{p}_t, y_t) \mid Z_1^{t-1}, I_1^{t-1}] \;=\; \ell(\boldsymbol{p}_t, y_t) = \mathbb{E}[\,\ell(I_t, y_t) \mid Z_1^{t-1}, I_1^{t-1}]\,, \tag{5.3}$$

hold for each t, where

$$\ell(\boldsymbol{p}_t, y_t) = \sum_{i=1}^{N} p_{i,t}\, \ell(i, y_t) \quad \text{and} \quad \widetilde{\ell}(\boldsymbol{p}_t, y_t) = \sum_{i=1}^{N} p_{i,t}\, \widetilde{\ell}(i, y_t)\,.$$

Note that the conditioning on Z_1^{t-1} and I_1^{t-1} is necessary because of the two following reasons: first, $\boldsymbol{p}_t$ depends both on the past realizations of the random choices of the forecaster Z_1^{t-1} (see

the third step in the algorithm of Figure 2) and on the past outcomes y_1^{t-1}; second, y_t is a function of both Z_1^{t-1} and I_1^{t-1}, as the environment is allowed to determine y_t after playing the game up to time $t-1$ (see Figure 1). For technical reasons, we sometimes consider a weaker model (which we call the *oblivious adversary*, see Section 1.4 in Chapter 2) where the sequence $y_1, y_2, \ldots$ of outcomes chosen by the environment is deterministic and independent of the forecaster random choices. This is equivalent to a game in which the environment must fix the sequence of outcomes before the game begins. Recall that the oblivious adversary model is reasonable in some scenarios, in which the forecaster's predictions have no influence on the environment. Also, any result proven in the standard model also holds in the oblivious adversary model.

The quantities $\widetilde{\ell}(i, y_t)$ may be considered as unbiased estimates of the true losses $\ell(i, y_t)$. The label efficient forecaster of Figure 2 is an exponentially weighted average forecaster using such estimates instead of the observed losses. The expected performance of this strategy may be bounded as follows.

THEOREM 5.1. *Fix a time horizon n and consider the label efficient forecaster of Figure 2 run with parameters $\varepsilon = m/n$ and $\eta = (\sqrt{2m \ln N})/n$. Then, the expected number of revealed labels equals m and*

$$\max_{i=1,\ldots,N} \mathbb{E}\left[\widehat{L}_n - L_{i,n}\right] \leqslant n\sqrt{\frac{2\ln N}{m}} \; .$$

In the sequel, for each $i = 1, \ldots, N$, we write

$$\widetilde{L}_{i,n} = \sum_{t=1}^{n} \widetilde{\ell}(i, y_t) \; .$$

PROOF. The proof is a simple adaptation of [**AuCeFrSc02**, Theorem 3.1]. The starting point is the second-order inequality below (see also [**PiSc01**, Theorem 1]). An application of Lemmas 4.3 and 4.5 to the estimated losses yields

$$\sum_{t=1}^{n} \widetilde{\ell}(\boldsymbol{p}_t, y_t) - \min_{i=1,\ldots,N} \widetilde{L}_{i,n} \leqslant \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{n} \sum_{j=1}^{N} p_{j,t} \widetilde{\ell}(j, y_t)^2 \; .$$

Since $\widetilde{\ell}(j, y_t) \in [0, 1/\varepsilon]$ for all j and y_t, the second term on the right-hand side may be bounded by $\dfrac{\eta}{2\varepsilon} \displaystyle\sum_{t=1}^{n} \sum_{j=1}^{N} p_{j,t} \widetilde{\ell}(j, y_t)$ and therefore we get, for all n,

$$\sum_{t=1}^{n} \widetilde{\ell}(\boldsymbol{p}_t, y_t) \left(1 - \frac{\eta}{2\varepsilon}\right) \leqslant \widetilde{L}_{i,n} + \frac{\ln N}{\eta} \; , \qquad i = 1, \ldots, N \; . \tag{5.4}$$

Taking expectations on both sides and substituting the proposed values of η and ε yields the desired result. □

REMARK 5.2. In the oblivious adversary model, Theorem 5.1 (and similarly Theorems 5.2 and 5.4 below) can be strengthened as follows. Consider the "lazy" forecaster of Figure 3 that keeps on choosing the same action as long as no new queries are issued. For this forecaster Theorems 5.1 and 5.2 hold with the additional statement that, with probability 1, the number of changes of an action, that is the number of steps where $I_t \neq I_{t+1}$, is at most the number of queried labels (by construction of the lazy forecaster). To prove the regret bound, note that we derive the statement of Theorem 5.1 by taking averages on both sides of (5.4), and then applying (5.2) and (5.3). Note that (5.4) holds for *every* realization of the random variables $I_1, \ldots, I_n$ and $Z_1, \ldots, Z_n$. Therefore, as the lazy forecaster differs from the forecaster of Figure 2 only in the distribution of

Parameters: Real numbers $\eta > 0$ and $0 \leqslant \varepsilon \leqslant 1$.
Initialization: $\boldsymbol{w}_1 = (1, \ldots, 1)$, $Z_0 = 1$.
For each round $t = 1, 2, \ldots$

(1) if $Z_{t-1} = 1$ then draw an action I_t from $\{1, \ldots, N\}$ according to the distribution

$$p_{i,t} = \frac{w_{i,t}}{\sum_{j=1}^{N} w_{j,t}}\,, \qquad i = 1, \ldots, N\,;$$

otherwise, let $I_t = I_{t-1}$;
(2) draw a Bernoulli random variable Z_t such that $\mathbb{P}[Z_t = 1] = \varepsilon$;
(3) if $Z_t = 1$ then obtain y_t and compute

$$w_{i,t+1} = w_{i,t}\, e^{-\eta\, \ell(i, y_t)/\varepsilon} \qquad \text{for each } i = 1, \ldots, N$$

else, let $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t$.

FIGURE 3. The lazy label efficient exponentially weighted average forecaster for the oblivious adversary model.

$I_1, \ldots, I_n$, inequality (5.4) holds for the lazy forecaster as well. In the oblivious adversary model y_t does not depend on $I_1, \ldots, I_{t-1}$; thus, by construction, $\boldsymbol{p}_t$ does not depend on $I_1, \ldots, I_{t-1}$ either. Therefore, we can take averages with respect to $I_1, \ldots, I_{t-1}$ obtaining the following version of (5.3) for the lazy forecaster,

$$\mathbb{E}\left[\widetilde{\ell}(\boldsymbol{p}_t, y_t) \mid Z_1^{t-1}\right] = \sum_{i=1}^{N} \ell(i, y_t)\, p_{i,t} = \mathbb{E}\left[\ell(I_t, y_t) \mid Z_1^{t-1}\right]\,.$$

Since (5.2) holds as well when the conditioning is limited to $Z_1, \ldots, Z_{t-1}$, we can derive for the lazy forecaster the same bounds as in Theorem 5.1 (and Theorem 5.2). Note also that the result holds even when y_t is allowed to depend on $Z_1, \ldots, Z_{t-1}$.

3.1. Bounding the regret with high probability. Theorem 5.1 guarantees that the expected per-round regret converges to zero whenever $m \to \infty$ as $n \to \infty$. The next result shows that this regret is, with overwhelming probability, bounded by a quantity proportional to $n\sqrt{(\ln N)/m}$.

THEOREM 5.2. *Fix a time horizon n and a number $\delta \in (0, 1)$. Consider the label efficient forecaster of Figure 2 run with parameters*

$$\varepsilon = \max\left\{0, \frac{m - \sqrt{2m \ln(4/\delta)}}{n}\right\} \qquad \text{and} \qquad \eta = \sqrt{\frac{2\varepsilon \ln N}{n}}\,.$$

Then, with probability at least $1 - \delta$, the number of revealed labels is at most m and

$$\forall\, t = 1, \ldots, n \qquad \widehat{L}_t - \min_{i=1,\ldots,N} L_{i,t} \leqslant 2n\sqrt{\frac{\ln N}{m}} + 6n\sqrt{\frac{\ln(4N/\delta)}{m}}\,.$$

REMARK 5.3. *(A label efficient forecaster with small internal regret.)* Remark 5.1, the conversion trick described in Section 1.2 of Chapter 3 and Theorem 5.1 lead to a label efficient forecaster with expected internal regret of the order of $n\sqrt{(\ln N)/m}$. The internal regret of this forecaster may be bounded with high probability by using the same martingale inequalities as in the proof of Theorem 5.2. The conversion of a no external regret label efficient forecaster to a no internal regret label efficient forecaster is thus straightforward, which is not the case for the conversion of forecasters suited for prediction with partial monitoring (see Section 6 of Chapter 6).

Before proving Theorem 5.2, note that if $\delta \leqslant 4Ne^{-m/8}$, then the right-hand side of the inequality is greater than n and therefore the statement is trivial. Thus, we may assume throughout the proof that $\delta > 4Ne^{-m/8}$. This ensures that

$$\varepsilon \geqslant m/(2n) > 0 \,. \tag{5.5}$$

We need a number of preliminary lemmas. The first is obtained by a simple application of Bernstein's inequality (see Lemma A.4).

LEMMA 5.1. *The probability that the strategy asks for more than m labels is at most $\delta/4$.*

PROOF. Note that the number $M = \sum_{t=1}^{n} Z_t$ of labels asked by the algorithm is binomially distributed with parameters n and ε and therefore, writing $\gamma = m/n - \varepsilon = n^{-1}\sqrt{2m\ln(4/\delta)}$, it satisfies

$$\mathbb{P}[M > m] = \mathbb{P}[M - \mathbb{E}M > n\gamma] \leqslant e^{-n\gamma^2/(2\varepsilon+2\gamma/3)} \leqslant e^{-n^2\gamma^2/2m} \leqslant \frac{\delta}{4}$$

where we used Bernstein's inequality (see Lemma A.4) in the second step and the definition of γ in the last two steps. □

LEMMA 5.2. *With probability at least $1 - \delta/4$,*

$$\forall\, t = 1, \ldots, n \qquad \sum_{s=1}^{t} \ell(\boldsymbol{p}_s, y_s) \leqslant \sum_{s=1}^{t} \widetilde{\ell}(\boldsymbol{p}_s, y_s) + \frac{4}{\sqrt{3}}\, n\sqrt{\frac{\ln(4/\delta)}{m}} \,.$$

Furthermore, with probability at least $1 - \delta/4$,

$$\forall\, i = 1, \ldots, N,\ \forall\, t = 1, \ldots, n \qquad \widetilde{L}_{i,t} \leqslant L_{i,t} + \frac{4}{\sqrt{3}}\, n\sqrt{\frac{\ln(4N/\delta)}{m}} \,.$$

PROOF. The proofs of both inequalities rely on the same techniques, namely the application of Bernstein's maximal inequality for martingales. We therefore focus on the first one, and indicate the modifications needed for the second one.

We introduce the sequence $X_s = \ell(\boldsymbol{p}_s, y_s) - \widetilde{\ell}(\boldsymbol{p}_s, y_s)$, $s = 1, \ldots, n$, which is a martingale difference sequence with respect to the filtration generated by the (Z_s, I_s), $s = 1, \ldots, n$. Defining $u = (4/\sqrt{3})n\sqrt{(1/m)\ln(4/\delta)}$ and the martingale $M_t = X_1 + \ldots + X_t$, our goal is to show that

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > u\right] \leqslant \frac{\delta}{4} \,.$$

For all $s = 1, \ldots, n$, we note that

$$\begin{aligned} \mathbb{E}\left[X_s^2 \mid Z_1^{s-1}, I_1^{s-1}\right] &= \mathbb{E}\left[(\ell(\boldsymbol{p}_s, y_s) - \widetilde{\ell}(\boldsymbol{p}_s, y_s))^2 \mid Z_1^{s-1}, I_1^{s-1}\right] \\ &\leqslant \mathbb{E}\left[\widetilde{\ell}(\boldsymbol{p}_s, y_s)^2 \mid Z_1^{s-1}, I_1^{s-1}\right] \leqslant 1/\varepsilon \,, \end{aligned}$$

so that summing over s, we have $V_t \leqslant n/\varepsilon$ for all $t = 1, \ldots, n$.

We now apply Lemma A.4 with $x = u$, $v = n/\varepsilon$, and $K = 1/\varepsilon$ (since $|X_s| \leqslant 1/\varepsilon$ with probability 1 for all s). This yields

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > x\right] = \mathbb{P}\left[\max_{t=1,\ldots,n} M_t > u \text{ and } V_n \leqslant \frac{n}{\varepsilon}\right] \leqslant \exp\left(-\frac{u^2}{2\,(n/\varepsilon + u/(3\,\varepsilon))}\right) \,.$$

Using $\ln(4/\delta) \leqslant m/8$ implied by the assumption $\delta > 4Ne^{-m/8}$, we see that $u \leqslant n$, which, combined with (5.5), shows that

$$\frac{u^2}{2\,(n/\varepsilon + u/(3\,\varepsilon))} \geqslant \frac{u^2}{(8/3)\,n/\varepsilon} \geqslant \frac{3u^2\,m}{16\,n^2} = \ln\frac{\delta}{4}$$

and this proves the first inequality.

To prove the second inequality note that, by the arguments above, for each fixed i we have

$$\mathbb{P}\left[\forall\, t=1,\dots,n \qquad \widetilde{L}_{i,t} > L_{i,t} + (4/\sqrt{3})\, n\sqrt{\frac{\ln(4N/\delta)}{m}}\right] \leqslant \frac{\delta}{4N}\,.$$

The proof is concluded by a union-of-events bound. □

PROOF (OF THEOREM 5.2). When $m \leqslant \ln N$, the bound given by the theorem is trivial, so we only need to consider the case when $m \geqslant \ln N$. Then (5.5) implies that $1-\eta/(2\varepsilon) \geqslant 0$. Thus, a straightforward combination of Lemmas 5.1 and 5.2 with (5.4) shows that, with probability at least $1-3\delta/4$, the strategy asks for at most m labels and

$$\forall\, t=1,\dots,n \qquad \sum_{s=1}^{t} \ell(\boldsymbol{p}_s, y_s)\left(1-\frac{\eta}{2\varepsilon}\right) \leqslant \min_{i=1,\dots,N} L_{i,t} + \frac{8}{\sqrt{3}}\, n\sqrt{\frac{1}{m}\ln\frac{4N}{\delta}} + \frac{\ln N}{\eta}\,,$$

which, since $\sum_{s=1}^{t} \ell(\boldsymbol{p}_s, y_s) \leqslant n$ for all $t \leqslant n$, implies

$$\begin{aligned}\forall\, t=1\dots,n \qquad \sum_{s=1}^{t} \ell(\boldsymbol{p}_s, y_s) - \min_{i=1,\dots,N} L_{i,t} &\leqslant \frac{n\eta}{2\varepsilon} + \frac{8}{\sqrt{3}}\, n\sqrt{\frac{1}{m}\ln\frac{4N}{\delta}} + \frac{\ln N}{\eta} \\ &= 2n\sqrt{\frac{\ln N}{m}} + \frac{8}{\sqrt{3}}\, n\sqrt{\frac{1}{m}\ln\frac{4N}{\delta}}\end{aligned}$$

by our choice of η and using $1/(2\varepsilon) \leqslant n/m$ derived from (5.5). The proof is finished by noting that the Hoeffding-Azuma maximal inequality (see Lemma A.2) implies that, with probability at least $1-\delta/4$,

$$\forall\, t=1,\dots,n \quad \widehat{L}_t = \sum_{s=1}^{t} \ell(I_s, y_s) \leqslant \sum_{s=1}^{t} \ell(\boldsymbol{p}_s, y_s) + \sqrt{\frac{n}{2}\ln\frac{4}{\delta}} \leqslant \sum_{s=1}^{t} \ell(\boldsymbol{p}_s, y_s) + n\sqrt{\frac{1}{2\,m}\ln\frac{4N}{\delta}}$$

since $m \leqslant n$. □

3.2. Hannan consistency. Theorem 5.1 does not directly imply Hannan consistency of the associated forecasting strategy because the regret bound does not hold uniformly over the sequence length n. However, using standard dynamical tuning techniques (such as the "doubling trick" described in [**CeFrHaHeScWa97**], see also Section 2.2 in Chapter 2) Hannan consistency can be achieved. The main quantity that arises in the analysis is the query rate $\mu(n)$, that is the number of queries that can be issued up to time n. The next result shows that Hannan consistency is achievable whenever $\mu(n)/(\log(n)\log\log(n)) \to \infty$.

In this section, we simply exhibit this small query rate for μ achieving Hannan consistency. We are not concerned with the interesting problem of finding an incremental update for our label efficient forecaster. Such an incremental update would associate to a query rate μ an on-line tuning of the weighting parameters η_t and of the instantaneous query rates ε_t (the parameters of the Bernoulli variables Z_t).

COROLLARY 5.1. *Let $\mu : \mathbb{N} \to \mathbb{N}$ be any nondecreasing integer-valued function such that*

$$\lim_{n\to\infty} \frac{\mu(n)}{\log_2(n)\log_2\log_2(n)} = \infty\,.$$

Then there exists a Hannan consistent randomized label efficient forecaster that issues at most $\mu(n)$ queries in the first n predictions, for any $n \in \mathbb{N}$.

PROOF. The algorithm we consider divides time into consecutive epochs of increasing lengths $n_r = 2^r$ for $r = 0, 1, 2, \ldots$. In the r-th epoch (of length 2^r) the algorithm runs the forecaster of Theorem 5.2 with parameters $n = 2^r$, $m = m_r$, and $\delta_r = 1/(1+r)^2$, where m_r will be determined by the analysis (without loss of generality, we assume the forecaster always asks at most m_r labels in each epoch r). Our choice of δ_r and the Borel-Cantelli lemma implies that the bound of Theorem 5.2 holds for all but finitely many epochs. Denote the (random) index of the last epoch in which the bound does not hold by $\hat{R}$. Let $L^{(r)}$ be cumulative loss of the best action in epoch r and let $\widehat{L}^{(r)}$ be the cumulative loss of the forecaster in the same epoch. Introduce $R(n) = \lfloor \log_2 n \rfloor$. Then, by Theorem 5.2 (since it proposes a maximal bound) and by definition of $\hat{R}$, for each n and for each realization of I_1^n and Z_1^n we have

$$\begin{aligned}
\widehat{L}_n - L_n^* &\leqslant \sum_{r=0}^{R(n)-1} \left(\widehat{L}^{(r)} - L^{(r)}\right) + \sum_{t=2^{R(n)}}^{n} \ell(I_t, y_t) - \sum_{t=2^{R(n)}}^{n} \min_{j=1,\ldots,N} \ell(j, y_t) \\
&\leqslant \sum_{r=0}^{\hat{R}} 2^r + 8 \sum_{r=\hat{R}+1}^{R(n)} 2^r \sqrt{\frac{\ln(4N(r+1)^2)}{m_r}} \,.
\end{aligned}$$

This, the finiteness of $\hat{R}$, and $1/n \leqslant 2^{-R(n)}$, imply that with probability 1,

$$\limsup_{n\to\infty} \frac{\widehat{L}_n - L_n^*}{n} \leqslant 8 \limsup_{R\to\infty} 2^{-R} \sum_{r=0}^{R} 2^r \sqrt{\frac{\ln(4N(r+1)^2)}{m_r}} \,.$$

Cesaro's lemma ensures that the $\limsup$ above equals zero as soon as $m_r/\ln r \to +\infty$. It remains to see that the latter condition is satisfied under the additional requirement that the forecaster does not issue more than $\mu(n)$ queries up to time n. This is guaranteed whenever $m_0 + m_1 + \ldots + m_{R(n)} \leqslant \mu(n)$ for each n. Denote by ϕ the largest nondecreasing function such that

$$\phi(t) \leqslant \frac{\mu(t)}{(1+\log_2 t)\log_2(1+\log_2 t)} \qquad \text{for all } t = 1, 2, \ldots$$

As μ grows faster than $\log_2(n)\log_2\log_2(n)$, we have that $\phi(t) \to +\infty$. Thus, choosing $m_0 = 0$, and $m_r = \lfloor \phi(2^r)\log_2(1+r) \rfloor$, we indeed ensure that $m_r/\ln r \to +\infty$. Furthermore, using that m_r is nondecreasing as a function of r, and using the monotonicity of ϕ,

$$\begin{aligned}
\sum_{r=0}^{R(n)} m_r &\leqslant (R(n)+1)\phi(2^{R(n)})\log_2(1+R(n)) \\
&\leqslant (1+\log_2 n)\phi(n)\log_2(1+\log_2 n) \leqslant \mu(n)
\end{aligned}$$

and this concludes the proof. □

4. Improvements for small losses

We now prove a refined bound in which the factors $n\sqrt{(\ln N)/m}$ of Theorem 5.2 are replaced by quantities of the order of $\sqrt{nL_n^*(\ln N)/m} + (n/m)\ln N$ in case of an oblivious adversary, and $\sqrt{nL_n^*(\ln(Nn))/m} + (n/m)\ln(Nn)$ in case of a non-oblivious one, where L_n^* is the cumulative loss of the best action,

$$L_n^* = L_n^*(y_1^n) = \min_{i=1,\ldots,N} \sum_{t=1}^{n} \ell(i, y_t) \,.$$

In particular, we recover the behavior already observed by Helmbold and Panizza [**HePa97**] for oblivious adversaries in the case $L_n^* = 0$.

Parameters: Real number $0 \leqslant \varepsilon \leqslant 1$.
Initialization: $t = 1$.
For each epoch $r = 0, 1, 2, \ldots$,

(1) let $K_r = 4^r (2 \ln N)/\varepsilon$;
(2) initialize $\widetilde{L}_i(r) = 0$ for all $i = 1, \ldots, N$;
(3) restart the forecaster of Figure 2 choosing ε and $\eta_r = \sqrt{(2\varepsilon \ln N)/K_r}$;
(4) **while** $\min_i \widetilde{L}_i(r) \leqslant K_r - 1/\varepsilon$ **do**:
  (a) denote by I_t the action chosen by the forecaster of Figure 2, and let $Z_t = 1$ if it asks for the label y_t, $Z_t = 0$ otherwise;
  (b) if $Z_t = 1$, then obtain the outcome y_t and update the estimated losses, for all $i = 1, \ldots, N$, as
$$\widetilde{L}_i(r) := \widetilde{L}_i(r) + \ell(i, y_t)/\varepsilon \; ;$$
  (c) $t := t + 1$.

FIGURE 4. A doubling version of the label efficient exponentially weighted average forecaster.

This is done by introducing a modified version of the forecaster of Figure 2, which performs a doubling trick over the estimated losses $\widetilde{L}_{i,t}$, $t = 1, \ldots, n$ (see Figure 4), and whose performance is studied below through several applications of Bernstein's lemma.

4.1. A forecaster suited for small losses. Similarly to [**AuCeFrSc02**, Section 4] and the algorithm of Theorem 4.5, we propose in Figure 4 a forecaster which uses a doubling trick based on the estimated losses of each action $i = 1, \ldots, N$. We denote the estimated accumulated loss of this algorithm by
$$\widetilde{L}_{A,n} = \sum_{t=1}^{n} \widetilde{\ell}(\boldsymbol{p}_t, y_t)$$
and prove the following inequality.

LEMMA 5.3. *For any $0 \leqslant \varepsilon \leqslant 1$, the forecaster of Figure 4 achieves, for all $n = 1, 2, \ldots$,*
$$\widetilde{L}_{A,n} \leqslant \widetilde{L}_n^* + 8\sqrt{2}\sqrt{\left(\widetilde{L}_n^* + 1/\varepsilon\right) \frac{\ln N}{\varepsilon}} + \frac{4 \ln N}{\varepsilon}$$
where
$$\widetilde{L}_n^* = \min_{i=1,\ldots,N} \widetilde{L}_{i,n} \; .$$

REMARK 5.4. *(Incremental update variant for the forecaster of Figure 4.)* Using the incremental update techniques of Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**] (see also the proof of Theorem 4.4), we note that instead of using a doubling trick in the definition of the forecaster, we could have considered an incremental update. The latter is defined by means of a sequence $\eta_1, \eta_2, \ldots$ of tuning parameters, and chooses the weights according to (4.5), computed with the estimated losses $\widetilde{\ell}_{i,t}$. The tuning parameters are of the order of
$$\eta_t \sim \sqrt{\frac{\varepsilon \ln N}{\widetilde{L}_{A,t-1}}} \; .$$
In addition, this self-confident update has the same guarantees as the forecaster of Lemma 5.3, at least as far as orders of magnitude in $\widetilde{L}_n^*$, ε, and N are concerned.

PROOF. The proof is divided in three steps. We first deal with each epoch, then sum the estimated losses over the epochs, and finally bound the total number R of different epochs (i.e., the final value of r). Let S_r and T_r be the first and last time steps completed on epoch r (where for convenience we define $T_R = n$). Thus, epoch r consists of trials $S_r, S_r + 1, \dots, T_r$. We denote the estimated cumulative loss of the forecaster at epoch r by

$$\widetilde{L}_A(r) = \sum_{t=S_r}^{T_r} \widetilde{\ell}(\boldsymbol{p}_t, y_t)$$

and the estimated cumulative losses of the actions $i = 1, \dots, N$ at epoch r by

$$\widetilde{L}_i(r) = \sum_{t=S_r}^{T_r} \widetilde{\ell}(i, y_t) \; .$$

Inequality (5.4) ensures that for epoch r, and for all $i = 1, \dots, N$,

$$\left(1 - \frac{\eta_r}{2\varepsilon}\right) \widetilde{L}_A(r) \leqslant \widetilde{L}_i(r) + \frac{\ln N}{\eta_r}$$

so dividing both terms by the quantity $1 - \eta_r/(2\varepsilon)$ (which is more than $1/2$ due to the choice of K_r), we get

$$\widetilde{L}_A(r) \leqslant \widetilde{L}_i(r) + \frac{\eta_r}{\varepsilon} \widetilde{L}_i(r) + 2\frac{\ln N}{\eta_r} \; .$$

The stopping condition now guarantees that $\min_i \widetilde{L}_i(r) \leqslant K_r$, hence, substituting the value of η_r, we have proved that for epoch r,

$$\widetilde{L}_A(r) \leqslant \min_{i=1,\dots,N} \widetilde{L}_i(r) + 2\sqrt{2}\sqrt{\frac{K_r \ln N}{\varepsilon}} \; .$$

Summing over $r = 0, \dots, R$, we get

$$\begin{aligned} \widetilde{L}_{A,n} &\leqslant \sum_{r=0}^{R} \min_{i=1,\dots,N} \widetilde{L}_i(r) + \sum_{r=0}^{R} 2\sqrt{2}\sqrt{\frac{K_r \ln N}{\varepsilon}} \\ &\leqslant \min_{i=1,\dots,N} \widetilde{L}_{i,n} + 2\sqrt{2}\sqrt{\frac{K_0 \ln N}{\varepsilon}} \left(2^{R+1} - 1\right) \; . \end{aligned} \tag{5.6}$$

It remains to bound the number R of epochs, or alternatively, to bound $2^{R+1} - 1$. Assume first that $R \geqslant 1$. In particular,

$$\begin{aligned} \widetilde{L}_n^* = \min_{i=1,\dots,N} \widetilde{L}_{i,n} &\geqslant \min_{i=1,\dots,N} \widetilde{L}_i(R-1) \\ &> K_{R-1} - 1/\varepsilon = 4^{R-1} K_0 - 1/\varepsilon \end{aligned}$$

so

$$2^{R-1} \leqslant \sqrt{\left(\widetilde{L}_n^* + 1/\varepsilon\right) \frac{1}{K_0}} \; .$$

The above is implied by

$$2^{R+1} - 1 \leqslant 1 + 4\sqrt{\left(\widetilde{L}_n^* + 1/\varepsilon\right) \frac{1}{K_0}}$$

which also holds for $R = 0$. Substituting the last inequality into (5.6) concludes the proof. □

4.2. Regret against a general opponent. We now state and prove a bound that holds in the most general (non-oblivious) adversarial model.

THEOREM 5.3. *Against any (non-oblivious) opponent, the label efficient forecaster of Figure 4, run with*

$$\varepsilon = \frac{m - \sqrt{2m\ln(4/\delta)}}{n}$$

ensures that, with probability $1-\delta$, the algorithm does not ask for more than m labels and

$$\forall\, t = 1,\dots,n \qquad \widehat{L}_t - L_t^* \leqslant U(L_n^*) + \sqrt{2\left(1 + L_n^* + U(L_n^*)\right)\ln\frac{4n}{\delta}} + \frac{1}{2}\ln\frac{4n}{\delta}$$

where

$$\begin{aligned} U(L_n^*) &= 20\sqrt{\frac{n}{m}L_n^*\ln\frac{4Nn}{\delta}} + 32\left(\frac{n}{m}\ln\frac{4Nn}{\delta}\right)^{3/4}(L_n^*)^{1/4} \\ &\quad +10\left(\frac{n}{m}\ln\frac{4Nn}{\delta}\right)^{7/8}(L_n^*)^{1/8} + 75\frac{n}{m}\ln\frac{4Nn}{\delta} \\ &\leqslant 137\times\max\left\{\sqrt{\frac{n}{m}L_n^*\ln\frac{4Nn}{\delta}}\,,\,\frac{n}{m}\ln\frac{4Nn}{\delta}\right\}. \end{aligned}$$

We remark here that the bound of the theorem is an improvement over that of Theorem 5.2 as soon as L_n^* grows slower than $n/\sqrt{\ln n}$. (For $L_n^* \sim n$ however, these bounds are worse, at least in the case of non-oblivious adversary, see Theorem 5.4 below for a refined bound for the case of an oblivious adversary.)

OPEN QUESTION 5.1. It is unclear whether this extra $\sqrt{\ln n}$ is needed or whether it is an artifact of our analysis (see also the comments before Corollary A.1).

For the proof of Theorem 5.3, we first relate $\widetilde{L}_n^*$ to L_n^*, and $\widetilde{L}_{A,n}$ to $\bar{L}_{A,n}$, where

$$\bar{L}_{A,n} = \sum_{t=1}^{n}\ell(\boldsymbol{p}_t, y_t)$$

is the sum of the conditional expectations of the instantaneous losses, and then substitute the obtained inequalities in the bound of Lemma 5.3.

LEMMA 5.4. *With probability $1-\delta/2$, the following $2n$ inequalities hold simultaneously,*

$$\begin{aligned} &\forall\, t = 1,\dots,n \qquad \widetilde{L}_t^* \leqslant L_t^* + 2\sqrt{\frac{n}{m}L_n^*\ln\frac{4Nn}{\delta}} + 4\frac{n}{m}\ln\frac{4Nn}{\delta}\,, \\ &\forall\, t = 1,\dots,n \qquad \widetilde{L}_{A,t} \geqslant \bar{L}_{A,t} - \left(2\sqrt{\frac{n}{m}\bar{L}_{A,n}\ln\frac{4n}{\delta}} + 4\frac{n}{m}\ln\frac{4n}{\delta}\right). \end{aligned}$$

PROOF. We prove that each of both lines holds with probability at least $1-\delta/4$. As the proofs are similar, we concentrate on the first one only. For all $i = 1,\dots,N$, we apply Corollary A.1 with $X_t = \widetilde{\ell}(i,y_t) - \ell(i,y_t)$, $t = 1,\dots,n$, which forms a martingale difference sequence (with respect to the filtration generated by (I_t, Z_t), $t = 1,\dots,n$). With the notation of the corollary, $K = 1/\varepsilon$, and V_n is smaller than $L_{i,n}/\varepsilon$, which shows that (for a given i), with probability at least $1-\delta/(4N)$,

$$\max_{t=1,\dots,n}\left(\widetilde{L}_{i,t} - L_{i,t}\right) \leqslant \sqrt{2\left(\frac{1}{\varepsilon^2} + \frac{L_{i,n}}{\varepsilon}\right)\ln\frac{4Nn}{\delta}} + \frac{\sqrt{2}}{3\varepsilon}\ln\frac{4Nn}{\delta}\,.$$

The proof is concluded by using $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ for $x, y \geqslant 0$, $1/\varepsilon \leqslant 2n/m$ (derived from (5.5)), $\ln(4Nn/\delta) \geqslant 1$ and the union-of-events bound. □

LEMMA 5.5. *With probability at least* $1-\delta/2$*,*

$$\forall\, t=1,\dots,n \qquad \bar{L}_{A,t} - L_t^* \leqslant U(L_n^*)\;,$$

where $U(L_n^*)$ *is as in Theorem 5.3.*

PROOF. We combine the inequalities of Lemma 5.4 with Lemma 5.3, and perform some trivial upper bounding, to get that, with probability $1-\delta/2$, for all $t=1,\dots,n$,

$$\begin{aligned}\bar{L}_{A,t} \;\leqslant\; & L_t^* + 2\sqrt{\frac{n}{m}\bar{L}_{A,n}\ln\frac{4Nn}{\delta}} + 18\sqrt{\frac{n}{m}L_n^*\ln\frac{4Nn}{\delta}} \\ & +23\,(L_n^*)^{1/4}\left(\frac{n}{m}\ln\frac{4Nn}{\delta}\right)^{3/4} + 56\frac{n}{m}\ln\frac{4Nn}{\delta}\;.\end{aligned}$$

An application of Lemma A.14 concludes the proof. □

PROOF (OF THEOREM 5.3). Lemma 5.1 shows that with probability at least $1-\delta/4$, the number of queried labels is less than m. Using the notation of Corollary A.1, we consider the martingale difference sequence formed by $X_t = \ell(I_t,y_t) - \ell(\boldsymbol{p}_t,y_t)$, with associated sum of conditional variances $V_n \leqslant \bar{L}_{A,n}$ and increments bounded by 1. Corollary A.1 then shows that with probability $1-\delta/4$,

$$\max_{t=1,\dots,n}\left(\widehat{L}_t - \bar{L}_{A,t}\right) \leqslant \sqrt{2\left(1+\bar{L}_{A,n}\right)\ln\frac{4n}{\delta}} + \frac{\sqrt{2}}{3}\ln\frac{4n}{\delta}\;.$$

We conclude the proof by applying Lemma 5.5 and a union-of-events bound. □

4.3. A refined bound for the oblivious adversary model. In the oblivious adversary model, the bound of Theorem 5.3 can be strengthened as follows.

THEOREM 5.4. *In the oblivious adversary model, the label efficient forecaster of Figure 4, run with*

$$\varepsilon = \frac{m-\sqrt{2m\ln(4/\delta)}}{n}$$

ensures that with probability $1-\delta$*, the algorithm does not ask for more than* m *labels and that*

$$\forall\, t=1,\dots,n \qquad \widehat{L}_t - L_t^* \leqslant B(L_n^*) + 2\sqrt{(L_n^* + B(L_n^*))\ln\frac{4}{\delta}}$$

where

$$\begin{aligned}B(L_n^*) \;=\; & 21\sqrt{\frac{n}{m}L_n^*\ln\frac{4N}{\delta}} + 39\left(\frac{n}{m}\ln\frac{4N}{\delta}\right)^{3/4}(L_n^*)^{1/4} \\ & +15\left(\frac{n}{m}\ln\frac{4N}{\delta}\right)^{7/8}(L_n^*)^{1/8} + 59\frac{n}{m}\ln\frac{4N}{\delta} \\ \leqslant\; & 134\,\max\left(\sqrt{\frac{n}{m}L_n^*\ln\frac{4N}{\delta}}\,,\,\frac{n}{m}\ln\frac{4N}{\delta}\right)\;.\end{aligned}$$

Observe that the order of magnitude of the bound of Theorem 5.4 is always at least as good as that of Theorem 5.2 and is better as soon as L_n^* grows slower than n.

The proof of Theorem 5.4 is based on combining Lemma 5.3 with two applications of Bernstein's inequality, but here, one of these applications is a backwards call to Bernstein's inequality: usually, one can handle the predictable quadratic variation of the studied martingale, and Bernstein's inequality is then a useful concentration result for the martingale. In the case of the second step below we know the deviations of the martingale (formed by $\widetilde{L}_{A,n}$), but we are interested in

the behavior of its predictable quadratic variation (equal to $\bar{L}_{A,n}$). The two quantities are related by a "backwards" use of Bernstein's lemma.

4.3.1. *Relating estimated losses to the cumulative loss of the best action.* We relate $\widetilde{L}_n^*$ and $\widetilde{L}_{A,n}$ to L_n^* by using Bernstein's inequality (Lemma A.4). First we point out the difference between oblivious and non-oblivious adversaries. More precisely, to apply Lemma A.4 rather than Corollary A.1, we need upper bounds K_i for all $L_{i,n} = L_{i,n}(y_1^n)$ (we exceptionally make the dependence on the played outcomes explicit) which are independent of I_1^n and Z_1^n. In case of an oblivious adversaries, the outcome sequence y_1^n is chosen in advance, and $K_i = L_{i,n}(y_1^n)$ is a suitable choice. This is not the case for non-oblivious adversaries whose behavior may take the actions of the forecaster into account (see the previous section).

Observe the similarity of the first statement of the following lemma to Lemmas 5.2 and 5.4. In particular, this first statement is an improvement on the first inequality of Lemma 5.4 in case of an oblivious opponent.

LEMMA 5.6. *When facing an oblivious adversary, with probability* $1 - \delta/4$,

$$\forall\, t = 1, \dots, n, \qquad \widetilde{L}_t^* \leqslant L_t^* + 2\sqrt{\frac{n}{m} L_n^* \ln \frac{4N}{\delta}} + \frac{n}{m} \ln \frac{4N}{\delta} \; .$$

Consequently, with probability $1 - \delta/4$,

$$\forall\, t = 1, \dots, n, \qquad \widetilde{L}_{A,t} \leqslant L_t^* + A(L_n^*) \; , \tag{5.7}$$

where

$$A(L_n^*) = 18\sqrt{\frac{n}{m} L_n^* \ln \frac{4N}{\delta}} + 23 \left(\frac{n}{m} \ln \frac{4N}{\delta} \right)^{3/4} (L_n^*)^{1/4} + 37 \frac{n}{m} \ln \frac{4N}{\delta} \; .$$

PROOF. For all $i = 1, \dots, N$, we may apply Lemma A.4 with $X_t = \widetilde{\ell}(i, y_t) - \ell(i, y_t)$, $t = 1, \dots, n$, which forms a martingale difference sequence with respect to the filtration generated by Z_t, $t = 1, \dots, n$. With the notation of Lemma A.4, $V_n \leqslant L_{i,n}/\varepsilon \leqslant 2n\, L_{i,n}/m$, which is indeed independent of the Z_t, and simple algebra and the union-of-events bound conclude the proof of the first statement. The second one follows from a combination of the first one with Lemma 5.3. □

4.3.2. *Bernstein's inequality used backwards.* Next we relate $\bar{L}_{A,n}$ to $\widetilde{L}_{A,n}$ (and thus to L_n^*, via Lemma 5.6). This is done by using Bernstein's lemma (Lemma A.4) once again, but backwards. Here again, we want to improve on the bounds yielded by Corollary A.1, which involve extra $\sqrt{\ln n}$ factors.

Relating $\widetilde{L}_{i,n}$ and $L_{i,n}$ as we did in Lemma 5.6 was straightforward, for in an oblivious setting, $L_{i,n}$ is a constant. Here, we consider the martingale $\bar{L}_{A,n} - \widetilde{L}_{A,n}$. It has $\bar{L}_{A,n}$ as an upper bound over its predictable quadratic variation, but this upper bound is not independent of the Z_t, due to the presence of the $\boldsymbol{p}_t$. Hence, Bernstein's lemma (Lemma A.4) does not apply in a direct way (and recall that we want to avoid any call to Corollary A.1). This is why we use Bernstein's lemma in a backwards sense, and get some information on the predictable quadratic variation of the martingale thanks to what we already know about its deviations.

LEMMA 5.7. *For oblivious adversaries, with probability at least* $1 - \delta/2$,

$$\forall\, t = 1, \dots, n \qquad \bar{L}_{A,t} - L_t^* \leqslant B(L_n^*) \; ,$$

where $B(L_n^*)$ *is as in Theorem 5.4.*

PROOF. Consider $A(L_n^*)$ as in Lemma 5.6 and fix a real number $x_0 > A(L_n^*)$. Consider the function ϕ_K defined in the statement of Lemma A.4. Then (5.7) and the union-of-events bound

imply that, for $\lambda > 0$ such that $\lambda - \phi_1(\lambda)/\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\max_{t=1,\dots,n}\left(\bar{L}_{A,t} - L_t^*\right) > x_0\right] \\
&\quad\leqslant \frac{\delta}{4} + \mathbb{P}\left[\max_{t=1,\dots,n}\left(\bar{L}_{A,t} - L_t^*\right) > x_0 \text{ and } \max_{t=1,\dots,n}\left(\widetilde{L}_{A,t} - L_t^*\right) \leqslant A(L_n^*)\right] \\
&\quad\leqslant \frac{\delta}{4} + \mathbb{P}\left[\max_{t=1,\dots,n}\exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)\left(\bar{L}_{A,t} - L_t^*\right) - \lambda\left(\widetilde{L}_{A,t} - L_t^*\right)\right)\right. \\
&\qquad\left. > \exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)x_0 - \lambda A(L_n^*)\right)\right] \\
&\quad\leqslant \frac{\delta}{4} + \mathbb{P}\left[\max_{t=1,\dots,n}\exp\left(\lambda\left(\bar{L}_{A,t} - \widetilde{L}_{A,t}\right) - \frac{\phi_1(\lambda)}{\varepsilon}\bar{L}_{A,t}\right)\right. \\
&\qquad\left. > \exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)x_0 - \lambda A(L_n^*) - \frac{\phi_1(\lambda)}{\varepsilon}L_n^*\right)\right] \qquad (5.8)
\end{aligned}$$

We introduce the martingale difference sequence (with increments bounded by 1) $X_t = \ell(\boldsymbol{p}_t, y_t) - \widetilde{\ell}(\boldsymbol{p}_t, y_t)$. The conditional variances satisfy

$$\mathbb{E}\left[X_t^2 \mid Z_1^{t-1}\right] \leqslant \mathbb{E}\left[\widetilde{\ell}(\boldsymbol{p}_t, y_t)^2 \mid Z_1^{t-1}\right] \leqslant \frac{\ell(\boldsymbol{p}_t, y_t)}{\varepsilon}$$

so that, using the notation of Lemma A.4, $V_n \leqslant \bar{L}_{A,n}/\varepsilon$.

By Lemma A.4, $\exp\left(\lambda\left(\bar{L}_{A,t} - \widetilde{L}_{A,t}\right) - \phi_1(\lambda)V_t\right)$ for $t = 1, 2, \dots$ is a nonnegative supermartingale. Hence, using Doob's maximal inequality, we get

$$\begin{aligned}
&\mathbb{P}\left[\max_{t=1,\dots,n}\exp\left(\lambda\left(\bar{L}_{A,t} - \widetilde{L}_{A,n}\right) - \frac{\phi_1(\lambda)}{\varepsilon}\bar{L}_{A,t}\right)\right. \\
&\qquad\left. > \exp\left(\left(\lambda - \frac{\phi_1(\lambda)}{\varepsilon}\right)x_0 - \lambda A(L_n^*) - \frac{\phi_1(\lambda)}{\varepsilon}L_n^*\right)\right] \\
&\quad\leqslant \mathbb{P}\left[\max_{t=1,\dots,n}\exp\left(\lambda\left(\bar{L}_{A,t} - \widetilde{L}_{A,t}\right) - \phi_1(\lambda)V_t\right)\right. \\
&\qquad\left. > \exp\left(\lambda\left(x_0 - A(L_n^*)\right) - \frac{\phi_1(\lambda)}{\varepsilon}\left(x_0 + L_n^*\right)\right)\right] \\
&\quad\leqslant \exp\left(\lambda\left(A(L_n^*) - x_0\right) + \frac{\phi_1(\lambda)}{\varepsilon}\left(x_0 + L_n^*\right)\right)\,. \qquad (5.9)
\end{aligned}$$

Now, choose

$$\lambda = \frac{x_0 - A(L_n^*)}{2\left(x_0 + L_n^*\right)}\,\varepsilon\,.$$

$\lambda \leqslant \varepsilon/2 \leqslant 1$, and therefore, using $\phi_1(t) \leqslant t^2$ for $t \leqslant 1$, we have proved that $\lambda - \phi_1(\lambda)/\varepsilon > 0$. Thus, (5.8) and (5.9) imply

$$\begin{aligned}
\mathbb{P}\left[\max_{t=1,\dots,n}\left(\bar{L}_{A,t} - L_t^*\right) > x_0\right] &\leqslant \frac{\delta}{4} + \exp\left(\lambda\left(A(L_n^*) - x_0\right) + \frac{\lambda^2}{\varepsilon}\left(x_0 + L_n^*\right)\right) \\
&= \frac{\delta}{4} + \exp\left(-\frac{\left(A(L_n^*) - x_0\right)^2}{4\left(x_0 + L_n^*\right)}\,\varepsilon\right)\,.
\end{aligned}$$

It suffices to find a $x_0 > A(L_n^*)$ such that

$$\frac{\left(A(L_n^*) - x_0\right)^2}{4\left(x_0 + L_n^*\right)}\,\varepsilon = \ln\frac{\delta}{4}\,.$$

One such choice is

$$x_0 = A(L_n^*) + \frac{2\ln\frac{\delta}{4}}{\varepsilon} + 2\sqrt{\frac{\ln\frac{\delta}{4}}{\varepsilon}}\sqrt{L_n^* + A(L_n^*) + \frac{\ln\frac{\delta}{4}}{\varepsilon}}\,.$$

Substituting the value of $A(L_n^*)$ yields the statement of the lemma. □

4.3.3. *Conclusion of the proof of Theorem 5.4.* Lemma 5.1 shows that, with probability at least $1 - \delta/4$, the number of queried labels is less than m. We then consider the martingale difference sequence formed by $X_t = \ell(I_t, y_t) - \ell(\boldsymbol{p}_t, y_t)$, with associated sum of conditional variances $V_n \leqslant \bar{L}_{A,n}$ and increments bounded by 1. Lemma A.4 yields

$$\mathbb{P}\left[\max_{t=1,\dots,n}\left(\widehat{L}_t - \bar{L}_{A,t}\right) > u \text{ and } \bar{L}_{A,n} \leqslant L_n^* + B(L_n^*)\right] \leqslant \exp\left(-\frac{u^2}{4\left(L_n^* + B(L_n^*)\right)}\right)$$

provided that $u \leqslant 3(L_n^* + B(L_n^*))$. Lemma 5.7 together with a union-of-events bound and the choice

$$u = 2\sqrt{(L_n^* + B(L_n^*))\ln\frac{4}{\delta}}$$

concludes the proof.

5. A lower bound for label efficient prediction

Here we show that the performance bounds proved in Section 3 for the label efficient exponentially weighted average forecaster are essentially unimprovable in the strong sense that no other label efficient forecasting strategy can have a better performance for all problems, in terms of the orders of magnitude in the parameters n, N and m.

OPEN QUESTION 5.2. *(Minimax constants.)* Theorem 5.5 solves the minimax problem (see Section 5 in Chapter 2) for the orders of magnitude in all parameters. We may now think of the best leading constant. In view of the results for prediction with expert advice with full monitoring stated in Chapter 2, see Section 5 therein, the best leading constant we may expect in Theorem 5.1 is $1/\sqrt{2}$, instead of the current $\sqrt{2}$. This gap of a factor of 2 is similar to the one between the two analyses of the no internal regret forecaster of Theorem 3.1, and is due to the two possible analysis of the performances of the exponentially weighted average algorithm. We may either use Taylor expansions (the potential approach), or Hoeffding's inequality, which is sharper. For more details on this gap, we refer to [**CeLu03**] and [**CeLu05**]. For the time being, as far as Theorem 5.1 is concerned, we do not see how to improve the constant.

Denote the set of natural numbers by $\mathbb{N} = \{1, 2, \dots\}$.

THEOREM 5.5. *There exist an outcome space $\mathcal{Y}$, a loss function $\ell : \mathbb{N} \times \mathcal{Y} \to [0,1]$, and a universal constant $c > 0$ such that, for all $N \geqslant 2$ and for all $n \geqslant m \geqslant 20\frac{e}{1+e}\ln(N-1)$, the cumulative (expected) loss of any (randomized) forecaster that uses actions in $\{1, \dots, N\}$ and asks for at most m labels while predicting a sequence of n outcomes satisfies the inequality*

$$\sup_{y_1,\dots,y_n \in \mathcal{Y}} \left(\mathbb{E}\left[\sum_{t=1}^n \ell(I_t, y_t)\right] - \min_{i=1,\dots,N}\sum_{t=1}^n \ell(i, y_t)\right) \geqslant c\, n\sqrt{\frac{\ln(N-1)}{m}}\,.$$

In particular, we prove the theorem for $c = \dfrac{\sqrt{e}}{(1+e)\sqrt{5(1+e)}}$.

Note that in the above theorem, we may take the same loss function for all N, we simply restrict the set of all possible actions to the first N ones. We also note that since the proof shows

that the opponent may be oblivious, the Hoeffding-Azuma inequality (for i.i.d. random variables, see Lemma A.2 in the Appendix) leads to a stronger result. For all forecasters using an auxiliary randomization formed by a sequence of i.i.d. variables, the lower bound also holds with high probability $1-\delta$ with respect to the auxiliary randomization (with deviations to $c\,n\sqrt{(\ln(N-1))/m}$ upper bounded by something of the order of $\sqrt{n\ln(1/\delta)}$). (For general opponents, the techniques of Devroye, Györfi, and Lugosi [**DeGyLu96**, Chapter 14] may lead to the same result.)

PROOF. First, we define $\mathcal{Y} = [0,1]$ and ℓ. Given $y \in [0,1]$, we denote by $(y_1, y_2, \ldots)$ its dyadic expansion, that is, the unique sequence not ending with infinitely many zeros such that

$$y = \sum_{k \geqslant 1} y_k \, 2^{-k} \, .$$

Now, the loss function is defined as $\ell(k,y) = y_k$ for all $y \in \mathcal{Y}$ and $k \in \mathbb{N}$.

We construct a random outcome sequence and show that the expected value of the regret (with respect both to the random choice of the outcome sequence and to the forecaster's possibly random choices) for any possibly randomized forecaster is bounded from below by the claimed quantity.

More precisely, we denote by $U_1, \ldots, U_n$ the auxiliary randomization which the forecaster has access to. Without loss of generality, this sequence can be taken as an i.i.d. sequence of uniformly distributed random variables over $[0,1]$. Our underlying probability space is equipped with the σ-algebra of events generated by the random outcome sequence $Y_1, \ldots, Y_n$ and by the randomization $U_1, \ldots, U_n$. As the random outcome sequence is independent of the auxiliary randomization, we define N different probability distributions, $\mathbb{P}_i \otimes \mathbb{P}_A$, $i = 1, \ldots, N$, formed by the product of the auxiliary randomization (whose associated probability distribution is denoted by $\mathbb{P}_A$) and one of the N different probability distributions $\mathbb{P}_1, \ldots, \mathbb{P}_N$ over the outcome sequence defined as follows.

For $i = 1, \ldots, N$, $\mathbb{Q}_i$ is defined as the distribution (over $[0,1]$) of

$$Z^* 2^{-i} + \sum_{k=1,\ldots,N,\ k \neq i} Z_k 2^{-k} + 2^{-(N+1)} U \, ,$$

where U, Z^*, $Z_1, \ldots, Z_N$ are independent random variables such that U has uniform distribution, and Z^* and the Z_k have Bernoulli distribution with parameter $1/2 - \varepsilon$ for Z^* and $1/2$ for the Z_k. Now, the randomization is such that under $\mathbb{P}_i$, the outcome sequence $Y_1, \ldots, Y_n$ is i.i.d. with common distribution $\mathbb{Q}_i$.

Then, under each $\mathbb{P}_i$ (for $i = 1, \ldots, N$), the losses $\ell(k, Y_t)$, $k = 1, \ldots, N$, $t = 1, \ldots, n$, are independent Bernoulli random variables with the following parameters. For all t, $\ell(i, Y_t) = 1$ with probability $1/2 - \varepsilon$ and $\ell(k, Y_t) = 1$ with probability $1/2$ for each $k \neq i$, where ε is a positive number specified below.

We have

$$\begin{aligned} \max_{y_1,\ldots,y_n} \left(\mathbb{E}_A \widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \right) &= \max_{y_1,\ldots,y_n} \max_{i=1,\ldots,N} \left(\mathbb{E}_A \widehat{L}_n - L_{i,n} \right) \\ &\geqslant \max_{i=1,\ldots,N} \mathbb{E}_i \left[\mathbb{E}_A \widehat{L}_n - L_{i,n} \right] \, , \end{aligned}$$

where $\mathbb{E}_i$ (resp. $\mathbb{E}_A$) denotes expectation with respect to $\mathbb{P}_i$ (resp. $\mathbb{P}_A$).

Now, we use the following decomposition lemma, which states that a randomized algorithm performs, on the average, just as a convex combination of deterministic algorithms. The simple proof is omitted.

LEMMA 5.8. *For any randomized forecaster, there exists an integer D, a point $\alpha \in \mathbb{R}^D$ in the probability simplex, $\alpha = (\alpha_1, \ldots, \alpha_D)$, and D deterministic algorithms (indexed by a superscript*

$d = 1, \ldots, D$) such that, for every t and every possible outcome sequence $y_1^{t-1} = (y_1, \ldots, y_{t-1})$,

$$\mathbb{P}_A\left[I_t = i \,|\, y_1^{t-1}\right] = \sum_{d=1}^{D} \alpha_d \, \mathbb{I}_{[I_t^d = i \,|\, y_1^{t-1}]} \,,$$

where $\mathbb{I}_{[I_t^d = i \,|\, y_1^{t-1}]}$ is the indicator function that the d-th deterministic algorithm chooses action i when the sequence of past outcomes is formed by y_1^{t-1}.

Using this lemma, we have that there exist D, α and D deterministic sub-algorithms such that

$$\begin{aligned}
\max_{i=1,\ldots,N} \mathbb{E}_i\Big[\mathbb{E}_A \widehat{L}_n - L_{i,n}\Big] &= \max_{i=1,\ldots,N} \mathbb{E}_i\left[\sum_{t=1}^{n}\sum_{d=1}^{D} \alpha_d \sum_{k=1}^{N} \mathbb{I}_{[I_t^d = k \,|\, Y_1^{t-1}]} \ell(k, Y_t) - L_{i,n}\right] \\
&= \max_{i=1,\ldots,N} \sum_{d=1}^{D} \alpha_d \, \mathbb{E}_i\left[\sum_{t=1}^{n}\sum_{k=1}^{N} \mathbb{I}_{[I_t^d = k \,|\, Y_1^{t-1}]} \ell(k, Y_t) - L_{i,n}\right]
\end{aligned}$$

Now, under $\mathbb{P}_i$ the regret grows by ε whenever an action different from i is chosen and remains the same otherwise. Hence,

$$\begin{aligned}
\max_{i=1,\ldots,N} \mathbb{E}_i\Big[\mathbb{E}_A \widehat{L}_n - L_{i,n}\Big] &= \max_{i=1,\ldots,N} \sum_{d=1}^{D} \alpha_d \, \mathbb{E}_i\left[\sum_{t=1}^{n}\sum_{k=1}^{N} \mathbb{I}_{[I_t^d = k \,|\, Y_1^{t-1}]} \ell(k, Y_t) - L_{i,n}\right] \\
&= \varepsilon \max_{i=1,\ldots,N} \sum_{d=1}^{D} \alpha_d \sum_{t=1}^{n} \mathbb{P}_i\left[I_t^d \neq i\right] \\
&= \varepsilon\, n \left(1 - \min_{i=1,\ldots,N} \sum_{d=1}^{D}\sum_{t=1}^{n} \frac{\alpha_d}{n} \mathbb{P}_i[I_t^d = i]\right) .
\end{aligned}$$

For the d-th deterministic subalgorithm, let $1 \leqslant T_1^d < \ldots < T_m^d \leqslant n$ be the times when the m queries were issued. Then $T_1^d, \ldots, T_m^d$ are finite stopping times with respect to the i.i.d. process $Y_1, \ldots, Y_n$. Hence, by a well-known fact in probability theory (see, e.g., [**ChTe88**, Lemma 2, page 138]), the revealed outcomes $Y_{T_1^d}, \ldots, Y_{T_m^d}$ are independent and identically distributed as Y_1.

Let R_t^d be the number of revealed outcomes at time t and note that R_t^d is measurable with respect to the random outcome sequence. Now, as the subalgorithm we consider is deterministic, R_t^d is fully determined by $Y_{T_1^d}, \ldots, Y_{T_m^d}$. Hence, I_t^d may be seen as a function of $Y_{T_1^d}, \ldots, Y_{T_m^d}$ rather than a function of $Y_{T_1^d}, \ldots, Y_{T_{R_t^d}^d}$ only. As the joint distribution of $Y_{T_1^d}, \ldots, Y_{T_m^d}$ under $\mathbb{P}_i$ is $\mathbb{Q}_i^m$, we have proved that

$$\mathbb{P}_i[I_t^d = i] = \mathbb{Q}_i^m[I_t^d = i] \,.$$

Consequently, the lower bound rewrites as

$$\max_{i=1,\ldots,N} \mathbb{E}_i\left[\mathbb{E}_A \widehat{L}_n - L_{i,n}\right] = \varepsilon\, n \left(1 - \min_{i=1,\ldots,N} \sum_{d=1}^{D}\sum_{t=1}^{n} \frac{\alpha_d}{n} \mathbb{Q}_i^m[I_t^d = i]\right) .$$

By Fano's inequality for convex combinations (see Lemma A.13 in the Appendix), it is guaranteed that

$$\min_{i=1,\ldots,N} \sum_{d=1}^{D}\sum_{t=1}^{n} \frac{\alpha_d}{n} \mathbb{Q}_i^m[I_t^d = i] \leqslant \max\left\{\frac{e}{1+e}, \frac{\bar{K}}{\ln(N-1)}\right\} ,$$

where

$$\bar{K} = \sum_{t=1}^{n}\sum_{d=1}^{D}\sum_{i=2}^{N} \frac{\alpha_d}{n(N-1)} \mathcal{K}(\mathbb{Q}_i^m, \mathbb{Q}_1^m) = \frac{1}{N-1}\sum_{i=2}^{N} \mathcal{K}(\mathbb{Q}_i^m, \mathbb{Q}_1^m) \,,$$

and $\mathcal{K}$ is the Kullback-Leibler divergence (or relative entropy) between two probability distributions. Moreover, $\mathbb{B}_p$ denoting the Bernoulli distribution with parameter p,

$$\mathcal{K}(\mathbb{Q}_i^m, \mathbb{Q}_1^m) = m\,\mathcal{K}(\mathbb{Q}_i, \mathbb{Q}_1) \leqslant m\,\left(\mathcal{K}\left(\mathbb{B}_{1/2-\varepsilon}, \mathbb{B}_{1/2}\right) + \mathcal{K}\left(\mathbb{B}_{1/2}, \mathbb{B}_{1/2-\varepsilon}\right)\right) \leqslant 5m\,\varepsilon^2$$

for $0 \leqslant \varepsilon \leqslant 1/10$, where the first equality holds by (A.1). For the second one, we note that the definition of the $\mathbb{Q}_i$ and Lemma A.7 imply that the considered Kullback-Leibler divergence is upper bounded by the Kullback-Leibler divergence between $(Z_1, \ldots, Z^*, \ldots, Z_n, U)$, where Z^* is in the i-th position, and $(Z^*, Z_2 \ldots, Z_n, U)$. (A.1) then shows that $\mathcal{K}(\mathbb{Q}_i, \mathbb{Q}_1) \leqslant \mathcal{K}\left(\mathbb{B}_{1/2-\varepsilon}, \mathbb{B}_{1/2}\right) + \mathcal{K}\left(\mathbb{B}_{1/2}, \mathbb{B}_{1/2-\varepsilon}\right)$, and Lemma A.5 concludes.

Therefore,

$$\max_{y_1,\ldots,y_n}\left(\mathbb{E}_A\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right) \geqslant \varepsilon\, n\,\left(1 - \max\left\{\frac{e}{1+e}, \frac{5m\,\varepsilon^2}{\ln(N-1)}\right\}\right)\,.$$

The choice

$$\varepsilon = \sqrt{\frac{e\ln(N-1)}{5(1+e)m}} \qquad (\varepsilon \leqslant 1/10)$$

yields the claimed bound. □

CHAPTER 6

Regret minimization under partial monitoring

We consider repeated games in which the player, instead of observing the action chosen by the opponent in each game round, receives a feedback generated by the combined choice of the two players. We study Hannan consistent players for these games; that is, randomized playing strategies whose per-round regret vanishes with probability one as the number n of game rounds goes to infinity. We prove a general lower bound of $\Omega(n^{-1/3})$ for the convergence rate of the regret, and exhibit a specific strategy that attains this rate for any game for which a Hannan consistent player exists.

Contents

This chapter is based on the submitted paper [**CeLuSt04b**] and is a joint work with Nicolò Cesa-Bianchi and Gábor Lugosi.

1. A motivating example

A simple yet nontrivial example of partial monitoring is the following dynamic pricing problem. A vendor sells a product to a sequence of customers whom he attends one by one. To each customer, the seller offers the product at a price he selects, say, from the interval $[0, 1]$. The customer then decides to buy the product or not. No bargaining is possible and no other information is exchanged between buyer and seller. The goal of the seller is to achieve an income almost as large as if he knew the maximal price each customer is willing to pay for the product. Thus, if the price offered to the t-th customer is p_t and the highest price this customer is willing to pay is $y_t \in [0, 1]$, then the loss of the seller is $y_t - p_t$ if the product is sold and (say) a constant $c > 0$ if the product is not sold. The first case corresponds to a loss of earnings and the second case to the fixed charges. Formally, the loss of the vendor at time t is

$$\ell(p_t, y_t) = (y_t - p_t)\mathbb{I}_{p_t \leqslant y_t} + c\,\mathbb{I}_{p_t > y_t}$$

PREDICTION WITH PARTIAL MONITORING

Parameters: number of actions N, number of outcomes M, loss function ℓ, feedback function h.

For each round $t = 1, 2 \ldots$,

(1) the environment chooses the next outcome $y_t \in \{1, \ldots, M\}$ without revealing it;
(2) the forecaster chooses a probability distribution $\boldsymbol{p}_t$ over the set of N actions and draws an action $I_t \in \{1, \ldots, N\}$ according to this distribution;
(3) the forecaster incurs loss $\ell(I_t, y_t)$ and each action i incurs loss $\ell(i, y_t)$, where none of these values is revealed to the forecaster;
(4) the feedback $h(I_t, y_t)$ is revealed to the forecaster.

where $c \in [0, 1]$. (In another version[1] of the problem the constant c may be replaced by y_t. We can even think of $y_t + c$, to take into account the loss of earnings plus the fixed charges, or of any measure of the loss.) In either case, if the seller knew in advance the empirical distribution of the y_t's then he could set a constant price $q \in [0, 1]$ which minimizes his overall loss. A natural question is whether there exists a randomized strategy for the seller such that his average regret

$$\frac{1}{n} \sum_{t=1}^{n} \ell(p_t, y_t) - \min_{q \in [0,1]} \frac{1}{n} \sum_{t=1}^{n} \ell(q, y_t)$$

is guaranteed to converge to zero as $n \to \infty$ regardless of the sequence $y_1, y_2, \ldots$ of prices. The difficulty in this problem is that the only information the seller (i.e., the forecaster) has access to is whether $p_t > y_t$ but neither y_t nor $\ell(p_t, y_t)$ are revealed. One of the main results of this chapter describes a simple strategy such that the average regret defined above is of the order of $n^{-1/5}$.

We treat such limited-feedback (or *partial monitoring*) prediction problems in a more general framework which we describe next. The dynamic pricing problem described above, which is a special case of this more general framework, has been also investigated by Kleinberg and Leighton [**KlLe03**] in a simpler setting where the reward of the seller is defined as $\rho(p_t, y_t) = p_t \, \mathbb{I}_{p_t \leqslant y_t}$. Note that, by using the feedback information (i.e., whether the customer bought the product or not), here the seller can compute the value of $\rho(p_t, y_t)$. Therefore, their game reduces to an instance of the multi-armed bandit game (see Example 6.1 below) with a continuous action space.

2. Main definitions

We adopt a learning-theoretic viewpoint and describe partial monitoring as a repeated prediction game between a *forecaster* (the player) and the *environment* (the opponent). In the same spirit, we call *outcomes* the actions taken by the environment. At each round $t = 1, 2 \ldots$ of the game, the forecaster chooses an action I_t from the set $\{1, \ldots, N\}$, and the environment chooses an action y_t from the set $\{1, \ldots, M\}$. The losses of the forecaster are summarized in the *loss matrix* $\boldsymbol{L} = [\ell(i, j)]_{N \times M}$. (This matrix is assumed to be known by the forecaster.) Without loss of generality, we rescale the losses so that they all lie in $[0, 1]$. If, at time t, the forecaster chooses an action $I_t \in \{1, \ldots, N\}$ and the outcome is $y_t \in \{1, \ldots, M\}$, then the forecaster's suffers loss $\ell(I_t, y_t)$. However, instead of the outcome y_t, the forecaster only observes the feedback $h(I_t, y_t)$, where h is a known *feedback function* that assigns, to each action/outcome pair in $\{1, \ldots, N\} \times \{1, \ldots, M\}$

[1] In this case it is easy to see that all terms depending on y_t cancel out when considering the regret, and we obtain the bandit setting analyzed by Kleinberg and Leighton [**KlLe03**]—see how the function ρ is defined below.

an element of a finite set $\mathcal{S} = \{s_1, \ldots, s_m\}$ of *signals*. The values of h are collected in a *feedback matrix* $\boldsymbol{H} = [h(i,j)]_{N \times M}$.

Note that we do not make any restrictive assumption on the power of the opponent. The environment may choose action y_t at time t by considering the whole past, that is, the whole sequence of action/outcome pairs (I_s, y_s), $s = 1, \ldots, t-1$. Without loss of generality, we assume that the opponent uses a deterministic strategy, so that the value of y_t is fixed by the sequence $(I_1, \ldots, I_{t-1})$. In comparison, the forecaster has access to significantly less information, since he only knows the sequence of past feedbacks, $(h(I_1, y_1), \ldots, h(I_{t-1}, y_{t-1}))$.

We note here that some authors consider a more general setup in which the feedback may be random. For the sake of clarity we treat the simpler model described above and return to the more general case in Section 7.

It is an interesting and complex problem to investigate the possibilities of a predictor only supplied with the limited information of the feedback. In this chapter we focus on the average regret

$$\frac{1}{n} \sum_{t=1}^{n} \ell(I_t, y_t) - \min_{i=1,\ldots,N} \frac{1}{n} \sum_{t=1}^{n} \ell(i, y_t) \; ,$$

that is, the difference between the average (per-round) loss of the forecaster and the average (per-round) loss of the best action. Forecasting strategies guaranteeing that the average regret converges to zero almost surely for all possible strategies of the environment are called *Hannan consistent* after James Hannan, who first proved the existence of a Hannan consistent strategy in the *full information* case [**Han57**] when $h(i,j) = j$ for all i, j (i.e., when the true outcome y_t is revealed to the forecaster after taking an action). The full information case has been studied extensively in the theory of repeated games, and in the fields of learning theory and information theory. A few key references and surveys include Blackwell [**Bla56**], Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire, and Warmuth [**CeFrHaHeScWa97**], Cesa-Bianchi and Lugosi [**CeLu99**], Feder, Merhav, and Gutman [**FeMeGu92**], Foster and Vohra [**FoVo99**], Hart and Mas-Colell [**HaMa01**], Littlestone and Warmuth [**LiWa94**], Merhav and Feder [**MeFe98**], and Vovk [**Vov90, Vov01**].

A natural question one may ask is under what conditions on the loss and feedback matrices it is possible to achieve Hannan consistency, that is, to guarantee that, asymptotically, the cumulative loss of the forecaster is not larger than that of the best constant action with probability one. Naturally, this depends on the relationship between the loss and feedback functions. An initial answer to this question has been provided by the work of Piccolboni and Schindelhauer [**PiSc01**]. However, since they are only concerned with expected performance, their results do not imply Hannan consistency. In addition, their bounds have suboptimal rates of convergence. Below, we extend those results by showing a forecaster that achieves Hannan consistency with optimal convergence rates.

Note that the forecaster is free to encode the values $h(i,j)$ of the feedback function by real numbers. The only restriction is that if $h(i,j) = h(i,j')$ then the corresponding real numbers should also coincide. To avoid ambiguities by trivial rescaling, we assume that $|h(i,j)| \leqslant 1$ for all pairs (i,j). Thus, in the sequel we assume that $\boldsymbol{H} = [h(i,j)]_{N \times M}$ is a matrix of real numbers between -1 and 1 and keep in mind that the forecaster may replace this matrix by $\boldsymbol{H}_\phi = [\phi_i(h(i,j))]_{N \times M}$ for arbitrary functions $\phi_i : [-1,1] \to [-1,1]$, $i = 1, \ldots, N$. Note that the set $\mathcal{S}$ of signals may be chosen such that it has $m \leqslant M$ elements, though after numerical encoding the matrix may have as many as MN distinct elements.

The problem of partial monitoring was considered by Mertens, Sorin, and Zamir [**MeSoZa94**], Rustichini [**Rus99**], Piccolboni, and Schindelhauer [**PiSc01**], and Mannor and Shimkin [**MaSh03**].

The forecaster strategy studied in Section 3 is first introduced in [**PiSc01**], where its expected regret is shown to have a sub-linear growth. Rustichini [**Rus99**] and Mannor and Shimkin [**MaSh03**] consider a more general setup in which the feedback is not necessarily a deterministic function of the pair outcome and forecaster's action, but it may be random with a distribution indexed by this pair. Based on Blackwell's approachability theorem, Rustichini [**Rus99**] establishes a general existence result for strategies with asymptotically optimal performance in this more general framework. In this chapter we answer Rustichini's question about the fastest achievable rate of convergence in the case when Hannan consistent strategies exist. Mannor and Shimkin also consider cases when Hannan consistency may not be achieved, give a partial solution, and point out important difficulties in such cases.

Before introducing a general prediction strategy and sufficient conditions for its Hannan consistency, we describe a few concrete examples of partial monitoring problems.

EXAMPLE 6.1. (*Multi-armed bandit problem.*) A well-studied special case of the partial monitoring prediction problem is the so-called multi-armed bandit problem. Recall from Chapter 2, Section 4, that here the forecaster, after taking an action, is able to measure his loss (or reward) but does not have access to what would have happened had he chosen another possible action. Here $\boldsymbol{H} = \boldsymbol{L}$, that is, the feedback received by the forecaster is just his own loss.

EXAMPLE 6.2. (*Dynamic pricing.*) Consider the dynamic pricing problem described in the introduction of the section under the additional restriction that all prices take their values from the finite set $\{0, 1/N, \ldots, (N-1)/N\}$ where N is a positive integer (see Example 6.6 for a non-discretized version). Clearly, if N is sufficiently large, this discrete version approximates arbitrarily the original problem. Now one may take $M = N$ and the loss matrix is

$$\boldsymbol{L} = [\ell(i,j)]_{N\times N} \qquad \text{where} \qquad \ell(i,j) = \frac{j-i}{N}\mathbb{I}_{i\leqslant j} + c\,\mathbb{I}_{i>j}\ .$$

The information the forecaster (i.e., the vendor) receives is simply whether the predicted value I_t is greater than the outcome y_t or not. Thus, the entries of the feedback matrix $\boldsymbol{H}$ may be taken to be $h(i,j) = \mathbb{I}_{i>j}$ or, after an appropriate re-encoding,

$$h(i,j) = a\,\mathbb{I}_{i\leqslant j} + b\,\mathbb{I}_{i>j} \quad i,j = 1,\ldots,N$$

where a and b are constants chosen by the forecaster satisfying $a, b \in [-1,1]$.

EXAMPLE 6.3. (*Apple tasting.*) This problem was considered by Helmbold, Littlestone, and Long [**HeLiLo00**] in a somewhat more restrictive setting. In this example $N = M = 2$ and the loss and feedback matrices are given by

$$\boldsymbol{L} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H} = \begin{bmatrix} a & a \\ b & c \end{bmatrix}\ .$$

Thus, the forecaster only receives feedback about the outcome y_t when he chooses the first action. (Imagine that apples are to be classified as "good for sale" or "rotten". An apple classified as "rotten" may be opened to check whether its classification was correct. On the other hand, since apples that have been checked cannot be put on sale, an apple classified "good for sale" is never checked.)

REMARK 6.1. *(On the pertinence of the regret.)* The previous example points out the limitations of the pertinence of the regret. It is not very interesting to only perform almost as well as the two strategies which consist, on the one hand, in selling all apples, and on the other hand, in throwing all apples out. However, this example may be considered as a toy example. Meaningful

situations for the regret are given, for instance, by the dynamic pricing problem stated in the introduction, or by the dynamic bandwidth allocation problem that motivated the work of Piccolboni and Schindelhauer [**PiSc01**].

EXAMPLE 6.4. (*Label efficient prediction.*) In label efficient prediction (see Helmbold and Panizza [**HePa97**], and also Cesa-Bianchi, Lugosi, and Stoltz [**CeLuSt05**] or Chapter 5) the forecaster, after choosing its prediction for round t, decides whether to query the outcome y_t, which he can only do for a limited number of times. In Chapter 5 matching upper and lower bounds are given for the regret in terms of the number of available labels, total number of rounds, and number of actions. A variant of the label efficient prediction problem may also be cast as a partial monitoring problem. Let $N = 3$, $M = 2$, and consider loss and feedback matrices of the form

$$\boldsymbol{L} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix} .$$

In this example the only times useful feedback is received are when the first action is played but in this case a maximal loss is incurred regardless of the outcome. Thus, just like in the problem of label efficient prediction, playing the "informative" action has to be limited, otherwise there is no hope for Hannan consistency.

3. General upper bounds on the regret

The purpose of this section is to derive general upper bounds for the rate of convergence of the regret achievable under partial monitoring. This will be done by analyzing a forecasting strategy inspired by Piccolboni and Schindelhauer [**PiSc01**]. This strategy is based on the exponentially weighted average forecaster, a thoroughly studied predictor in the full information case, see, for example, Auer, Cesa-Bianchi, and Gentile [**AuCeGe02**], Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire, and Warmuth [**CeFrHaHeScWa97**], Littlestone and Warmuth [**LiWa94**], Vovk [**Vov90, Vov01**]. In the special case of the multi-armed bandit problem, the forecaster reduces to the strategy of Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**] (see also Hart and Mas-Colell [**HaMa02**] and Section 4 of Chapter 2 for closely related methods).

The crucial assumption under which the strategy is defined is that there exists an $N \times N$ matrix $\boldsymbol{K} = [k(i,j)]_{N\times N}$ such that

$$\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{H} ,$$

that is,

$$\boldsymbol{H} \quad \text{and} \quad \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{L} \end{bmatrix}$$

have the same rank. In other words we may write, for all $i \in \{1,\dots,N\}$ and $j \in \{1,\dots,M\}$,

$$\ell(i,j) = \sum_{l=1}^{N} k(i,l)\, h(l,j) . \tag{6.1}$$

In this case one may define the estimated losses $\widetilde{\ell}$ by

$$\widetilde{\ell}(i,y_t) = \frac{k(i,I_t)\, h(I_t,y_t)}{p_{I_t,t}} , \quad i = 1,\dots,N . \tag{6.2}$$

(Note that, contrary to the estimates used in Chapter 5, the ones proposed above are real-valued, and may be negative.) We denote the cumulative estimated losses at round t and for action i by $\widetilde{L}_{i,t} = \sum_{s=1}^{t} \widetilde{\ell}(i,y_t)$.

Parameters: matrix $\boldsymbol{L}$ of losses, feedback matrix $\boldsymbol{H}$, matrix $\boldsymbol{K}$ such that $\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{H}$
Initialization: $\widetilde{L}_{1,0} = \cdots = \widetilde{L}_{N,0} = 0$.
For each round $t = 1, 2, \ldots$

(1) let $\eta_t = (k^*)^{-2/3}((\ln N)/N)^{2/3} t^{-2/3}$ and $\gamma_t = (k^*)^{2/3} N^{2/3} (\ln N)^{1/3} t^{-1/3}$;
(2) choose an action I_t from the set of actions $\{1, \ldots, N\}$ at random, according to the distribution $\boldsymbol{p}_t$ defined by

$$p_{i,t} = (1 - \gamma_t) \frac{e^{-\eta_t \widetilde{L}_{i,t-1}}}{\sum_{k=1}^{N} e^{-\eta_t \widetilde{L}_{k,t-1}}} + \frac{\gamma_t}{N} \; ;$$

(3) let $\widetilde{L}_{i,t} = \widetilde{L}_{i,t-1} + \widetilde{\ell}(i, y_t)$ for all $i = 1, \ldots, N$.

FIGURE 1. The randomized forecaster for prediction under partial monitoring.

Consider the forecaster defined in Figure 1. Roughly speaking, the two terms in the expression of $p_{i,t}$ correspond to "exploitation" and "exploration". The first term assigns exponentially decreasing weights to the actions depending on their estimated cumulative losses, while the second term ensures sufficient exploration to guarantee accurate estimates of the losses.

A key property of the loss estimates is their unbiasedness in the following sense. Denoting by $\mathbb{E}_t$ the conditional expectation given $I_1, \ldots, I_{t-1}$ (i.e., the expectation with respect to the distribution $\boldsymbol{p}_t$ of the random variable I_t), observe that this conditioning fixes the value of y_t, and thus,

$$\begin{aligned} \mathbb{E}_t \widetilde{\ell}(i, y_t) &= \sum_{k=1}^{N} \frac{k(i,k)\, h(k, y_t)}{p_{k,t}} p_{k,t} \\ &= \sum_{k=1}^{N} k(i,k)\, h(k, y_t) = \ell(i, y_t) \; , \quad i = 1, \ldots, N \; , \end{aligned}$$

and therefore $\widetilde{\ell}(i, y_t)$ is an unbiased estimate of the loss $\ell(i, y_t)$.

The main performance bound of this section is summarized in the next theorem. Note that the average regret

$$\frac{1}{n} \left(\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{i=1,\ldots,N} \sum_{t=1}^{n} \ell(i, y_t) \right)$$

decreases to zero at a rate $n^{-1/3}$. This is significantly slower than the best rate $n^{-1/2}$ obtained in the "full information" case. In the next section we show that this rate cannot be improved in general. Thus, the price paid for having access only to some feedback except for the actual outcomes is the deterioration in the rate of convergence. However, Hannan consistency is still achievable whenever the conditions of the theorem are satisfied.

THEOREM 6.1. *Consider any partial monitoring problem such that the loss and feedback matrices satisfy $\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{H}$ for some $N \times N$ matrix $\boldsymbol{K}$ with $k^* = \max\{1, \max_{i,j} |k(i,j)|\}$, and consider the forecaster of Figure 1. Let $\delta \in (0,1)$. Then, for all strategies of the opponent, for all*

n, with probability at least $1-\delta$,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,y_t) - \min_{i=1,\dots,N}\frac{1}{n}\sum_{t=1}^{n}\ell(i,y_t) \\ \leqslant \quad 5\left(\frac{(k^*N)^2\ln N}{n}\right)^{1/3}\left(1+\sqrt{\frac{3}{2}\frac{\ln((N+4)/\delta)}{\ln N}}\right) \\ +\sqrt{\frac{1}{2n}\ln\frac{N+4}{\delta}} \;+\; 5(k^*N)^{4/3}n^{-2/3}(\ln N)^{-1/3}\ln\frac{N+4}{\delta} \\ +\frac{1}{n}\left(1+((k^*N)^2\ln N)^{1/3}+k^*N\right)\ln\frac{N+4}{\delta}\;.\end{aligned}$$

The main term in the performance bound has the order of magnitude $n^{-1/3}(k^*N)^{2/3}(\ln N)^{1/3}$. Observe that this theorem directly implies Hannan consistency, by a simple application of the Borel-Cantelli lemma. We also remark that the bound proposed by the theorem could be strengthened in a maximal version of the same flavor as the statement of Theorem 5.2 by a more careful way of writing the proof, by exploiting the fact that we apply a maximal version of Bernstein's inequality (see Lemma A.4). We do not do so for the sake of simplicity and readability.

REMARK 6.2. *(Improvement for small losses.)* We may design a forecaster suited for small losses under partial monitoring, in the same spirit as we did for the label efficient prediction setting in Section 4 of Chapter 5. Denoting by $\bar{L}_{A,n}$ the (conditional) expected cumulative loss of the forecaster, and by L_n^* the cumulative loss of the best action, we may prove that

$$\bar{L}_{A,n} \leqslant L_n^* + \gamma(n\bar{L}_{A,n})^{1/3}\;,$$

for an absolute constant γ. Solving shows that the (expected) regret $\bar{L}_{A,n}-L_n^*$ is bounded by a quantity of the order of $(nL_n^*)^{1/3}$, and using the same martingale inequalities as in Section 4 of Chapter 5, we may prove that this is still the order of magnitude of the non-expected regret. This improves on the general $n^{2/3}$ upper bound on the regret proposed by Theorem 6.1. We do not work out the tedious details.

PROOF (OF THEOREM 6.1). The starting point of the proof of the theorem is an application of Lemmas 4.3 and 4.5 to the estimated losses (see also the proof of Theorem 5.1). Since $\widetilde{\ell}_{i,t}$ lies between $-B_t$ and B_t, where $B_t = k^*N/\gamma_t$, the proposed values of γ_t and η_t imply that $\eta_t B_t \leqslant 1$ if and only if $t \geqslant (\ln N)/(Nk^*)$, that is, for all $t \geqslant 1$. Therefore, defining for $t=1,\dots,n$, the probability vector $\widetilde{\boldsymbol{p}}_t$ by its components

$$\widetilde{p}_{i,t} = \frac{e^{-\eta_t\widetilde{L}_{i,t-1}}}{\sum_{k=1}^{N}e^{-\eta_t\widetilde{L}_{k,t-1}}} \qquad i=1,\dots,N\;,$$

we may apply Lemmas 4.3 and 4.5 (and use $e-2\leqslant 1$) to obtain

$$\sum_{t=1}^{n}\sum_{i=1}^{N}\widetilde{p}_{i,t}\widetilde{\ell}(i,y_t) - \min_{j=1,\dots,N}\widetilde{L}_{j,n} \leqslant \frac{2\ln N}{\eta_{n+1}} + \sum_{t=1}^{n}\eta_t\sum_{i=1}^{N}\widetilde{p}_{i,t}\widetilde{\ell}(i,y_t)^2\;.$$

Since $p_{i,t} = (1-\gamma_t)\widetilde{p}_{i,t} + \gamma_t/N$, the inequality above yields, after some simple bounding,

(6.3)
$$\sum_{t=1}^{n}\sum_{i=1}^{N}p_{i,t}\widetilde{\ell}(i,y_t) - \min_{j=1,\dots,N}\widetilde{L}_{j,n} \leqslant \frac{2\ln N}{\eta_{n+1}} + \sum_{t=1}^{n}\eta_t\sum_{i=1}^{N}\widetilde{p}_{i,t}\widetilde{\ell}(i,y_t)^2 + \sum_{t=1}^{n}\gamma_t\sum_{i=1}^{N}\frac{1}{N}\widetilde{\ell}(i,y_t)\;.$$

Introduce the notation

$$\widehat{L}_n = \sum_{t=1}^{n} \ell(I_t, y_t) \quad \text{and} \quad L_{j,n} = \sum_{t=1}^{n} \ell(j, y_t), \;\; j = 1, \ldots, N \; .$$

Next we show that, with an overwhelming probability, the right-hand side of the inequality (6.3) is less than something of the order $n^{2/3}$, and that the left-hand side is close to the actual regret

$$\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{j=1,\ldots,N} L_{j,n} \; .$$

Our main tool is Bernstein's inequality for martingales, see Lemma A.4 in the Appendix. This inequality implies the following four lemmas, whose proofs are similar, so we omit some of them.

LEMMA 6.1. *With probability at least* $1 - \delta/(N+4)$,

$$\sum_{t=1}^{n} \sum_{i=1}^{N} p_{i,t} \ell(i, y_t) \leqslant \sum_{t=1}^{n} \sum_{i=1}^{N} p_{i,t} \widetilde{\ell}(i, y_t) + \sqrt{2(k^*N)^2 \left(\sum_{t=1}^{n} \frac{1}{\gamma_t} \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} \left(1 + \frac{k^*N}{\gamma_n} \right) \ln \frac{N+4}{\delta} \; .$$

PROOF. Define $Z_t = -\sum_{i=1}^{N} p_{i,t} \widetilde{\ell}(i, y_t)$ so that $\mathbb{E}_t[Z_t] = -\sum_{i=1}^{N} p_{i,t} \ell(i, y_t)$, and consider $X_t = Z_t - \mathbb{E}_t[Z_t]$. We note that

$$\begin{aligned} \mathbb{E}_t[X_t^2] \leqslant \mathbb{E}_t[Z_t^2] &= \sum_{i,j} p_{i,t} p_{j,t} \mathbb{E}_t \left[\widetilde{\ell}(i, y_t) \widetilde{\ell}(j, y_t) \right] \\ &= \sum_{i,j} p_{i,t} p_{j,t} \sum_{k=1}^{N} p_{k,t} \frac{k(i,k) k(j,k) h(k, y_t)^2}{p_{k,t}^2} \leqslant \frac{(k^*N)^2}{\gamma_t} \; , \end{aligned}$$

and therefore,

$$V_n = \sum_{t=1}^{n} \mathbb{E}_t[X_t^2] \leqslant (k^*N)^2 \sum_{t=1}^{n} \frac{1}{\gamma_t} \; .$$

On the other hand, $|X_t|$ is bounded by $K = 1 + (k^*N)/\gamma_n$. Bernstein's inequality (see Lemma A.4) thus concludes the proof. □

LEMMA 6.2. *For each fixed* j*, with probability at least* $1 - \delta/(N+4)$,

$$\widetilde{L}_{j,n} \leqslant L_{j,n} + + \sqrt{2(k^*N)^2 \left(\sum_{t=1}^{n} \frac{1}{\gamma_t} \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} \left(1 + \frac{k^*N}{\gamma_n} \right) \ln \frac{N+4}{\delta} \; .$$

LEMMA 6.3. *With probability at least* $1 - \delta/(N+4)$,

$$\sum_{t=1}^{n} \eta_t \sum_{i=1}^{N} \widetilde{p}_{i,t} \widetilde{\ell}(i, y_t)^2 \leqslant \sum_{t=1}^{n} \eta_t \frac{(k^*N)^2}{\gamma_t} + \sqrt{2(k^*N)^4 \left(\sum_{t=1}^{n} \frac{\eta_t^2}{\gamma_t^3} \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} \ln \frac{N+4}{\delta} \; .$$

PROOF. Let $Z_t = \eta_t \sum_{i=1}^{N} \widetilde{p}_{i,t} \widetilde{\ell}(i, y_t)^2$, and $X_t = Z_t - \mathbb{E}_t[Z_t]$. All $|X_t|$ are bounded by

$$K = \max_{t=1,\ldots,n} \eta_t \frac{(k^*N)^2}{\gamma_t^2} = 1 \; .$$

On the other hand,

$$V_n = \sum_{t=1}^{n} \mathbb{E}_t[X_t^2] \leqslant (k^*N)^4 \sum_{t=1}^{n} \frac{\eta_t^2}{\gamma_t^3} \; .$$

Lemma A.4 now concludes the proof, together with the inequality

$$\mathbb{E}_t[Z_t] \leqslant \eta_t \frac{(k^*N)^2}{\gamma_t} \,.$$

□

LEMMA 6.4. *With probability at least* $1 - \delta/(N+4)$,

$$\sum_{t=1}^{n} \gamma_t \sum_{i=1}^{N} \frac{1}{N} \widetilde{\ell}(i, y_t) \leqslant \sum_{t=1}^{n} \gamma_t + \sqrt{2(k^*N)^2 \left(\sum_{t=1}^{n} \gamma_t \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} (k^*N + \gamma_1) \ln \frac{N+4}{\delta} \,.$$

The next lemma is an easy consequence of the Hoeffding-Azuma inequality for sums of bounded martingale differences (see Lemma A.2 in the Appendix).

LEMMA 6.5. *With probability at least* $1 - \delta/(N+3)$,

$$\sum_{t=1}^{n} \ell(I_t, y_t) \leqslant \sum_{t=1}^{n} \sum_{i=1}^{N} p_{i,t} \ell(i, y_t) + \sqrt{\frac{n}{2} \ln \frac{N+4}{\delta}} \,.$$

The proof of the main result follows now from a combination of Lemmas 6.1 to 6.5 with (6.3) (where Lemma 6.2 is applied N times). Using a union-of-event bound, we see that, with probability $1 - \delta$,

$$\begin{aligned}
&\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{j=1,\dots,N} L_{j,n} \\
&\quad \leqslant \quad \frac{2 \ln N}{\eta_{n+1}} \\
&\quad + \quad 2 \left(\sqrt{2(k^*N)^2 \left(\sum_{t=1}^{n} \frac{1}{\gamma_t} \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} \left(1 + \frac{k^*N}{\gamma_n} \right) \ln \frac{N+4}{\delta} \right) \\
&\quad + \quad \sum_{t=1}^{n} \eta_t \frac{(k^*N)^2}{\gamma_t} + \sqrt{2(k^*N)^4 \left(\sum_{t=1}^{n} \frac{\eta_t^2}{\gamma_t^3} \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} \ln \frac{N+4}{\delta} \\
&\quad + \quad \sum_{t=1}^{n} \gamma_t + \sqrt{2(k^*N)^2 \left(\sum_{t=1}^{n} \gamma_t \right) \ln \frac{N+4}{\delta}} + \frac{\sqrt{2}}{3} (k^*N + \gamma_1) \ln \frac{N+4}{\delta} \\
&\quad + \quad \sqrt{\frac{n}{2} \ln \frac{N+4}{\delta}} \,.
\end{aligned}$$

Substituting the proposed values of γ_t and η_t, and using that for $-1 < \alpha \leqslant 0$

$$\sum_{t=1}^{n} t^{\alpha} \leqslant \frac{1}{\alpha+1} n^{\alpha+1} \,,$$

we obtain the claimed result with a simple calculation. □

We close this section by considering the implications of Theorem 6.1 to the special cases mentioned in the introduction.

EXAMPLE 6.5. (*Multi-armed bandit problem.*) Recall that in the case of the multi-armed bandit problem $\boldsymbol{H} = \boldsymbol{L}$ and the condition of the theorem is trivially satisfied. Indeed, one may take $\boldsymbol{K}$ to be the identity matrix so that $k^* = 1$. Thus, Theorem 6.1 implies a bound of the order of $((N^2 \ln N)/n)^{1/3}$. Even though, as it is shown in the next section, the rate $O(n^{-1/3})$

cannot be improved in general, faster rates of convergence are achievable for the special case of the bandit problem. Indeed, for the bandit problem Cesa-Bianchi and Lugosi [**CeLu05**] describe careful modifications of the forecaster of Theorem 6.1 that achieve an upper bound of the order of $\sqrt{N(\ln N)/n}$, see Theorem 2.8. It remains a challenging problem to characterize the class of problems that admit rates of convergence faster than $O(n^{-1/3})$, see Question 6.3.

EXAMPLE 6.6. (*Dynamic pricing.*) In the discretized version of the dynamic pricing problem (i.e., when all prices are restricted to the set $\{0, 1/N, \ldots, (N-1)/N\}$), the feedback matrix is given by $h(i,j) = a\,\mathbb{I}_{i \leqslant j} + b\,\mathbb{I}_{i>j}$ for some arbitrarily chosen values of a and b. By choosing, for example, $a = 1$ and $b = 0$, it is clear that $\boldsymbol{H}$ is an invertible matrix and therefore one may choose $\boldsymbol{K} = \boldsymbol{L}\,\boldsymbol{H}^{-1}$ and obtain a Hannan-consistent strategy with average regret of the order of $n^{-1/3}$. Thus, the seller has a way of selecting the prices I_t such that his loss is not much larger than what he could have achieved had he known the values y_t of all costumers and offered the best constant price. Note that with this choice of a and b, the value of k^* equals 1 (i.e., does not depend on N) and therefore the upper bound has the form $C((N^2 \log N)/n)^{1/3}\sqrt{\ln(1/\delta)}$ for some constant C. By choosing $N \approx n^{1/5}$ and running the forecaster into stages of doubling lengths the effect of discretization sums up to n/N and decreases at about the same rate as the average regret, so that for the original problem with unrestricted price range one may obtain a regret bound of the form

$$\frac{1}{n}\sum_{t=1}^{n} \ell(p_t, y_t) - \min_{q \in [0,1]} \frac{1}{n}\sum_{t=1}^{n} \ell(q, y_t) = O(n^{-1/5} \ln n)\,.$$

We leave out the simple but tedious details of the proof, except the precise way we should discretize. We show below that the discretization $Y_N(y_t) = \lfloor N y_t \rfloor / N$ ensures that

$$\ell(p,y) - \ell\left(p, Y_N(y)\right) \leqslant \frac{1}{N}$$

for all p and y in $[0,1]$. To this end, we note that only three cases may happen, $p > y$ (in which case, $p > Y_N(y)$), $p \leqslant Y_N(y)$ (in which case, $p \leqslant y$), and $Y_N(y) < p \leqslant y$. In these cases, $\ell(p,y) - \ell\left(p, Y_N(y)\right)$ respectively equals 0, $y - Y_N(y) \leqslant 1/N$ and $(y-p) - c \leqslant y - p \leqslant 1/N$. Thus, the cumulated effect of the discretization in one stage may be bounded by n/N as claimed, provided[2] that the discretization is given by $Y_N(y_t) = \lfloor N y_t \rfloor / N$.

EXAMPLE 6.7. (*Apple tasting.*) In the apple tasting problem described above, one may choose the feedback values $a = b = 1$ and $c = 0$. Then, the feedback matrix is invertible and, once again, Theorem 6.1 applies.

EXAMPLE 6.8. (*Label efficient prediction.*) Recall next the variant of the label efficient prediction problem described in the previous section. Here the rank of $\boldsymbol{L}$ equals two, so it is necessary (and sufficient) to encode the feedback matrix such that its rank equals two. One possibility is to choose $a = 1$, $b = 1/2$, and $c = 1/4$. Then we have $\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{H}$ for

$$\boldsymbol{K} = \begin{bmatrix} 0 & 2 & 2 \\ 2 & -2 & -2 \\ -2 & 4 & 4 \end{bmatrix}\,.$$

The obtained rate of convergence $O(n^{-1/3})$ may be shown to be optimal. In fact, it is this example that we use in Section 5 to show that this rate of convergence cannot be improved in general.

REMARK 6.3. It is interesting to point out that the bound of Theorem 6.1 does not depend explicitly on the value of the cardinality M of the set of outcomes. Of course, in some problems

[2]Other choices for the discretization may only lead to a worse $\max\{1/N,\, c\}$ upper bound.

the value k^* may depend on M. However, in some important special cases, such as the multi-armed bandit problem for which $k^* = 1$, this value is independent of M. In such cases the result extends easily to an infinite set of outcomes. In particular, the case when the loss matrix may change with time can be encoded this way.

4. Other regret-minimizing strategies

In the previous section we saw a forecasting strategy that guarantees that the average regret is of the order of $n^{-1/3}$ whenever the loss matrix $\boldsymbol{L}$ can be expressed as $\boldsymbol{K}\,\boldsymbol{H}$ for some matrix $\boldsymbol{K}$. In this section we discuss some alternative strategies that yield small regret under different conditions.

First note that it is not true that the existence of a Hannan consistent predictor is guaranteed if and only the loss matrix $\boldsymbol{L}$ can be expressed as $\boldsymbol{K}\,\boldsymbol{H}$. The following example describes such a situation.

EXAMPLE 6.9. Let $N = M = 3$,

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H} = \begin{bmatrix} a & b & c \\ d & d & d \\ e & e & e \end{bmatrix} .$$

Clearly, for all choices of the numbers a, b, c, d, e, the rank of the feedback matrix is at most two and therefore there is no matrix $\boldsymbol{K}$ for which $\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{H}$. However, note that whenever the first action is played, the forecaster has full information about the outcome y_t. Formally, an action $i \in \{1, \dots, N\}$ is said to be *revealing* for a feedback matrix $\boldsymbol{H}$ if all entries in the i-th row of $\boldsymbol{H}$ are different. Below we prove the existence of a Hannan consistent forecaster for all problems in which there exists a revealing action.

THEOREM 6.2. *Consider an arbitrary partial monitoring problem $(\boldsymbol{L}, \boldsymbol{H})$ such that $\boldsymbol{L}$ has a revealing action. Let $\delta \in (0, 1)$. If the randomized forecasting strategy of Figure 2 is run with parameters*

$$\varepsilon = \max\left\{0, \frac{m - \sqrt{2m\ln(4/\delta)}}{n}\right\} \qquad \text{and} \qquad \eta = \sqrt{\frac{2\varepsilon \ln N}{n}}$$

where $m = (4n)^{2/3}(\ln(4N/\delta))^{1/3}$, then

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{i=1,\dots,N} L_{1,n}\right) \leqslant 8n^{-1/3}\left(\ln\frac{4N}{\delta}\right)^{1/3}$$

holds with probability at least $1 - \delta$ for any strategy of the opponent.

PROOF. The forecaster of Figure 2 chooses at each round a revealing action with a small probability $\varepsilon \approx m/n$ (of the order of $n^{-1/3}$). At these m stages where a revealing action is chosen, the forecaster suffers a total loss of about $m = O(n^{2/3})$ but gets full information about the outcome y_t. This situation is a modification of the problem of *label efficient prediction* studied in Helmbold and Panizza [**HePa97**], and in Chapter 5 (see also Cesa-Bianchi, Lugosi, and Stoltz [**CeLuSt05**]). In particular, the algorithm proposed in Figure 2 coincides with that of of Theorem 5.2 –except maybe at those rounds when $Z_t = 1$. Indeed, Theorem 5.2 ensures that, with probability at least $1 - \delta$, not more than m among the Z_t have value 1, and that

$$\sum_{t=1}^{n} \ell(J_t, y_t) - \min_{j=1,\dots,N} \sum_{t=1}^{n} \ell(j, y_t) \leqslant 8n\sqrt{\frac{\ln(4N/\delta)}{m}} .$$

Parameters: $0 \leqslant \varepsilon \leqslant 1$ and $\eta > 0$. Action r is revealing.
Initialization: $w_{1,0} = \cdots = w_{N,0} = 1$.
For each round $t = 1, 2, \ldots$

(1) draw an action J_t from $\{1, \ldots, N\}$ according to the distribution

$$p_{i,t} = \frac{w_{i,t-1}}{\sum_{j=1}^{N} w_{j,t-1}} , \qquad i = 1, \ldots, N ,$$

(2) draw a Bernoulli random variable Z_t such that $\mathbb{P}[Z_t = 1] = \varepsilon$;
(3) if $Z_t = 1$ then play a revealing action, $I_t = r$, observe y_t, and compute

$$w_{i,t} = w_{i,t-1} e^{-\eta\, \ell(i,y_t)/\varepsilon} \qquad \text{for each } i = 1, \ldots, N ;$$

(4) otherwise, if $Z_t = 0$, play $I_t = J_t$ and let $w_{i,t} = w_{i,t-i}$ for each $i = 1, \ldots, N$.

FIGURE 2. The randomized forecaster for feedback matrices with a revealing action.

This in turn implies that

$$\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{j=1,\ldots,N} \sum_{t=1}^{n} \ell(j, y_t) \leqslant m + 8n\sqrt{\frac{\ln(4N/\delta)}{m}} ,$$

and substituting the proposed value for the parameter m concludes the proof. □

REMARK 6.4. (*Dependence on* N.) Observe that, even when the condition of Theorem 6.1 is satisfied, the bound of Theorem 6.2 is considerably tighter. Indeed, even though the dependence on the time horizon n is identical in both bounds (of the order of $n^{-1/3}$), the bound of Theorem 6.2 depends on the number of actions N in a logarithmic way only. As an example, consider the case of the multi-armed bandit problem. Recall that here $\boldsymbol{H} = \boldsymbol{L}$ and there is a revealing action if and only if the loss matrix has a row whose elements are all different. In such a case Theorem 6.2 provides a bound of the order of $((\ln N)/n)^{1/3}$. On the other hand, there exist bandit problems for which, if $N \leqslant n$, it is impossible to achieve a regret smaller than $(1/20)(N/n)^{1/2}$ (see [**AuCeFrSc02**]). If N is large, the logarithmic dependence of Theorem 6.2 gives a considerable advantage.

Interestingly, even if $\boldsymbol{L}$ cannot be expressed as $\boldsymbol{K}\,\boldsymbol{H}$, if a revealing action exists, the strategy of Section 3 may be used to achieve a small regret. This may be done by using a trick of Piccolboni and Schindelhauer [**PiSc01**] to first convert the problem into another partial-monitoring problem for which the strategy of Section 3 can be used. The basic step of this conversion is to replace the pair of $N \times M$ matrices $(\boldsymbol{L}, \boldsymbol{H})$ by a pair of $mN \times M$ matrices $(\boldsymbol{L}', \boldsymbol{H}')$ where $m \leqslant M$ denotes the cardinality of the set $\mathcal{S} = \{s_1, \ldots, s_m\}$ of signals (i.e., the number of distinct elements of the matrix $\boldsymbol{H}$). In the obtained prediction problem the forecaster chooses among mN actions at each time instance. The converted loss matrix $\boldsymbol{L}'$ is obtained simply by repeating each row of the original loss matrix m times. The new feedback matrix $\boldsymbol{H}'$ is binary and is defined by

$$H'(m(i-1)+k, j) = \mathbb{I}_{h(i,j)=s_k} , \quad i = 1, \ldots, N,\ k = 1, \ldots, m,\ j = 1, \ldots, M .$$

Note that this way we get rid of the inconvenient problem of how to encode in a natural way the feedback symbols. We also propose the following interpretation for this first step. Before taking an action, the forecaster has a belief about the nature of the feedback he will get. He then is only interested in knowing whether he was right or not. If the matrices

$$\boldsymbol{H}' \quad \text{and} \quad \begin{bmatrix} \boldsymbol{H}' \\ \boldsymbol{L}' \end{bmatrix}$$

have the same rank, then there exists a matrix $\boldsymbol{K}'$ such that $\boldsymbol{L}' = \boldsymbol{K}'\,\boldsymbol{H}'$ and the forecaster of Section 3 may be applied to obtain a forecaster that has an average regret of the order of $n^{-1/3}$ for the converted problem. However, it is easy to see that any forecaster A with such a bounded regret for the converted problem may be trivially transformed into a forecaster A' for the original problem with the same regret bound: A' simply takes an action i whenever A takes an action of the form $m(i-1)+k$ for any $k = 1, \dots, m$.

The above conversion procedure guarantees Hannan consistency for a large class of partial monitoring problems. For example, if the original problem has a revealing action i, then $m = M$ and the $M \times M$ sub-matrix formed by the rows $M(i-1)+1, \dots, Mi$ of $\boldsymbol{H}'$ is the identity matrix (up to some permutations over the rows), and therefore has full rank. Then obviously a matrix $\boldsymbol{K}'$ with the desired property exists and the procedure described above leads to a forecaster with an average regret of the order of $n^{-1/3}$. This forecaster is similar to the one considered in Theorem 6.2 in the sense that it may build its predictions only on feedbacks received when playing the (original) revealing action i. This is so because the matrix $\boldsymbol{K}'$ may be taken equal to $[0\ \boldsymbol{L}'\ 0]$, where $\boldsymbol{L}'$ lies in the columns $M(i-1)+1, \dots, Mi$ of $\boldsymbol{K}'$.

This last statement may be generalized, in a straightforward way, to an even larger class of problems as follows.

COROLLARY 6.1 (Distinguishing actions). *Assume that the feedback matrix $\boldsymbol{H}$ is such that for each outcome $j = 1, \dots, M$ there exists an action $i \in \{1, \dots, N\}$ such that for all outcomes $j' \neq j$, $h(i,j) \neq h(i,j')$. Then the conversion procedure described above leads to a Hannan consistent forecaster with an average regret of the order of $n^{-1/3}$.*

The rank of $\boldsymbol{H}'$ may be considered as a measure of the information provided by the feedback. The highest possible value is achieved by matrices $\boldsymbol{H}'$ with rank M. For such feedback matrices, Hannan consistency may be achieved for all associated loss matrices $\boldsymbol{L}'$.

Even though the above conversion strategy applies to a large class of problems, the associated condition fails to characterize the set of pairs $(\boldsymbol{L}, \boldsymbol{H})$ for which a Hannan consistent forecaster exists. Thus, for matrices $\boldsymbol{H}'$ with rank strictly less then M, the precise form of $\boldsymbol{L}'$ matters. Consider the following two examples, which we already encoded (and simplified by deleting redundant lines).

EXAMPLE 6.10. Consider an example proposed by Piccolboni and Schindelhauer [**PiSc01**], with $N = M = 4$,

$$\boldsymbol{L} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

They also consider the modified version given by

$$\boldsymbol{L}_h = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H}_h = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The problem with $(\boldsymbol{L}_h, \boldsymbol{H}_h)$ is more difficult than the one with $(\boldsymbol{L}, \boldsymbol{H})$ in the following sense. The losses of actions 3 and 4 increased, but since these are dominated by actions 1 and 2, none of them will achieve the argmin in the cumulative losses of the constant actions. Therefore, for all sequences of played actions $I_1, I_2, \dots$ and obtained outcomes $y_1, y_2, \dots$, the regret of any forecasting strategy is larger for the problem with $(\boldsymbol{L}_h, \boldsymbol{H}_h)$ than for the problem with $(\boldsymbol{L}, \boldsymbol{H})$.

Consequently, any Hannan-consistent strategy for $(\boldsymbol{L}_h, \boldsymbol{H}_h)$ is a Hannan-consistent strategy for $(\boldsymbol{L}, \boldsymbol{H})$. Now,

$$\boldsymbol{H}_h \quad \text{and} \quad \begin{bmatrix} \boldsymbol{H}_h \\ \boldsymbol{L}_h \end{bmatrix}$$

have the same rank, and we may therefore construct explicitly a Hannan-consistent strategy with the above techniques. Note that on the contrary

$$\boldsymbol{H} \quad \text{and} \quad \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{L} \end{bmatrix}$$

do not have the same ranks.

EXAMPLE 6.11. Consider a case with $N = M = 3$,

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} .$$

In this example, when the second and third actions are chosen, the feedback is identical, independently of the outcome, so Hannan consistency is impossible to achieve in this case. (This is because in this example all three actions may achieve the argmin in the cumulative losses of the constant actions, for suitable outcome sequences.) However, it is easy to construct a strategy for which

$$\frac{1}{n}\left(\sum_{t=1}^{n} \ell(I_t, y_t) - \min\left(L_{1,n}, \frac{L_{2,n} + L_{3,n}}{2}\right)\right) = o(1)$$

with probability 1. Rustichini [**Rus99**] and Mannor and Shimkin [**MaSh03**] determine more generally the asymptotically optimal performance that a strategy can get given the matrices $\boldsymbol{L}$ and $\boldsymbol{H}$. This may be in some cases much worse than what the best constant action achieves, that is, Hannan consistency is not always achievable.

Following the techniques used in Example 6.10, Piccolboni and Schindelhauer [**PiSc01**] show a second simple conversion of the pair $(\boldsymbol{L}', \boldsymbol{H}')$ that can be applied in situations when there is no matrix $\boldsymbol{K}'$ with the property $\boldsymbol{L}' = \boldsymbol{K}' \boldsymbol{L}'$. This second conversion step basically deals with some actions which they define as "non-exploitable" and which correspond to Pareto-dominated actions. These actions are not erased, for they may be associated with worthwhile feedbacks, but their losses are set to 1 on all outcomes. A third conversion step shifting in a certain sense the losses associated to exploitable actions follows, and ends up with a matrix pair $(\boldsymbol{L}'', \boldsymbol{H}'')$ such that any Hannan-consistent forecaster for the problem with $(\boldsymbol{L}'', \boldsymbol{H}'')$ is also Hannan-consistent for the problem with $(\boldsymbol{L}, \boldsymbol{H})$. Now, a Hannan consistent procedure may be constructed based on the forecaster of Section 3, provided that $\boldsymbol{L}''$ may be expressed as $\boldsymbol{L}'' = \boldsymbol{K}'' \boldsymbol{H}''$.

In addition, Piccolboni and Schindelhauer show that if this condition is not satisfied, then there exists an external randomization over the sequences of outcomes such that the sequence of expected regrets grows at least as n, where the expectations are understood with respect to the forecaster's auxiliary randomization and the external randomization. (An external randomization over the outcomes corresponds to the case of an oblivious adversary.) Thus, a proof by contradiction using the dominated-convergence theorem (thanks to the boundedness of the losses) shows that Hannan consistency is impossible to achieve in these cases. This result combined with Theorem 6.1 implies the following gap theorem (see also Theorem 3 of [**PiSc01**] for a similar, though weaker, statement, for expected regrets).

COROLLARY 6.2. *Consider a partial monitoring forecasting problem with loss and feedback matrices $\boldsymbol{L}$ and $\boldsymbol{H}$. If Hannan consistency can be achieved for this problem, then there exists a Hannan consistent forecaster based on the results of Section 3 whose average regret vanishes at rate $n^{-1/3}$.*

Thus, whenever it is possible to force the average regret to converge to zero, a convergence rate of the order of $n^{-1/3}$ is also possible. In some special cases, such as the multi-armed bandit problem, even faster rates of the order of $n^{-1/2}$ may be achieved (see Auer, Cesa-Bianchi, Freund, and Schapire [**AuCeFrSc02**] and Auer [**Aue02**]). However, as it is shown in Section 5 below, for certain problems in which Hannan consistency is achievable, it can only be achieved with rate of convergence not faster than $n^{-1/3}$.

OPEN QUESTION 6.1. We still lack a concise (and more intrinsic) characterization of the problems $(\boldsymbol{L}, \boldsymbol{H})$ for which an Hannan-consistent forecaster may be constructed. Furthermore, we also lack a characterization of the problems $(\boldsymbol{L}, \boldsymbol{H})$ for which convergence rates faster than $n^{-1/3}$ may be achieved.

5. A lower bound on the regret

Next we show that the order of magnitude (in terms of the length of the play n) of the bound of Theorem 6.1 is, in general, not improvable. A closely related idea in a somewhat different context appears in Mertens, Sorin and Zamir [**MeSoZa94**, page 290]. They introduce a zero-sum game, whose first player has full monitoring and whose second player has only partial monitoring. They compute by induction a lower bound on the minimax value of the game, and are able to further lower bound it by a quantity of the order of $n^{-1/3}$ thanks to repeated applications of the game-theoretic minimax theorem.

THEOREM 6.3. *Consider the partial monitoring problem of label efficient prediction introduced in Example 6.4 and defined by the pair of loss and feedback matrices*

$$\boldsymbol{L} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix} .$$

Then, for any $n \geqslant 8$ and for any (randomized) forecasting strategy there exists a sequence $y_1, \ldots, y_n$ of outcomes such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, y_t)\right] - \min_{i=1,2,3} \frac{1}{n}\sum_{t=1}^{n} \ell(i, y_t) \geqslant \frac{n^{-1/3}}{5} ,$$

where $\mathbb{E}$ denotes the expectation with respect to the auxiliary randomization of the forecaster.

The proof is inspired by the proof techniques of Section 5 of Chapter 5. Here however, we need to take into account that the number of asked labels, that is, the number of times when the informative action is played, may not be limited a priori by a fixed integer. Rather, different outcome sequences may lead to different numbers of asked labels. We also note, similarly to the comments after the statement of Theorem 5.5, that the proof shows, thanks to the Hoeffding-Azuma inequality, that the lower bound also holds with high probability with respect to the auxiliary randomization for all forecasters using an auxiliary i.i.d. sequence of random variables to draw their predictions.

OPEN QUESTION 6.2. *(Minimax orders in N and M.)* We consider here the minimax problem restricted to the prediction settings $(\boldsymbol{L}, \boldsymbol{H})$ where Hannan-consistency is achievable. Theorems 6.3

and Corollary 6.2 show that our general forecaster solves the minimax problem as for the orders of magnitude in n.

Using the techniques of Section 5 of Chapter 5 (namely Fano's inequality, see Lemma A.13), it is easy to extend the theorem above to get a lower bound of the order of $n^{-1/3}(\ln N)^{1/3}$, by considering suitable $(N+1) \times N$ matrices $\boldsymbol{L}$ and $\boldsymbol{H}$. The latter is the best possible lower bound for label efficient prediction as a special case of prediction with partial monitoring, in view of the upper bound obtained in Theorem 6.2. However, the order of magnitude in N of this lower bound still does not match the one of the bound proposed by Theorem 6.1, as we still lack at least a factor of $N^{2/3}$. We do not know yet if this is because Theorem 6.1 has to be improved, or because somewhat harder examples have to be found (perhaps by considering some general examples of prediction with distinguishing actions). However, we favor the second option and conjecture that in many cases significantly larger lower bounds (as a function of N) hold.

In addition, the dependencies of k^* on N and M should be studied and made more explicit. In conclusion, we solve in this section the problem of the minimax order in n but leave open the delicate issue of the minimax orders in N and M.

OPEN QUESTION 6.3. *(Optimal order in n for a given prediction setting.)* We recalled in Sections 4 and 5 of Chapter 2 that the optimal orders of magnitude in n for the regret are $\sqrt{n}$ with full information, and even in a bandit setting. Theorem 6.3 shows that this optimal order is $n^{2/3}$ for the setting of label efficient prediction. We do not know if there exist orders in between that are optimal for a certain prediction setting, i.e., for a certain pair $(\boldsymbol{L}, \boldsymbol{H})$. (See also Question 6.1.)

PROOF. The proof proceeds by constructing a random sequence of outcomes and showing that, for any (possibly randomized) forecaster, the expected value of the regret with respect both to the random choice of the outcome sequence and to the forecaster's random choices is bounded from below by the claimed quantity.

More precisely, fix $n \geqslant 8$ and denote by $U_1, \ldots, U_n$ the auxiliary randomization which the forecaster has access to. Without loss of generality, it can be taken as an i.i.d. sequence of uniform random variables in $[0,1]$. The underlying probability space is equipped with the σ-algebra of events generated by the random sequence of outcomes $Y_1, \ldots, Y_n$ and by the randomization $U_1, \ldots, U_n$. The random sequence of outcomes is independent of the auxiliary randomization, whose associated probability distribution is denoted by $\mathbb{P}_A$.

We define three different probability distributions, $\mathbb{P} \otimes \mathbb{P}_A$, $\mathbb{Q} \otimes \mathbb{P}_A$, and $\mathbb{R} \otimes \mathbb{P}_A$, formed by the product of the auxiliary randomization and one of the three probability distributions $\mathbb{P}$, $\mathbb{Q}$, and $\mathbb{R}$ over the sequence of outcomes defined as follows. Under $\mathbb{P}$ the sequence $Y_1, Y_2, \ldots, Y_n$ is formed by independent, identically distributed $\{1,2\}$-valued random variables with parameter $1/2$. Under $\mathbb{Q}$ (respectively $\mathbb{R}$) the Y_i are also i.i.d. and $\{1,2\}$-valued but with parameter $1/2 - \varepsilon$ (respectively $1/2 + \varepsilon$), where $\varepsilon > 0$ is chosen below.

We denote by $\mathbb{E}_{\mathbb{A}}$ (respectively, $\mathbb{E}_{\mathbb{P}}$, $\mathbb{E}_{\mathbb{Q}}$, $\mathbb{E}_{\mathbb{R}}$, $\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}$, $\mathbb{E}_{\mathbb{Q}\otimes\mathbb{P}_A}$, $\mathbb{E}_{\mathbb{R}\otimes\mathbb{P}_A}$) the expectation with respect to $\mathbb{P}_{\mathbb{A}}$ (respectively, $\mathbb{P}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{P} \otimes \mathbb{P}_A$, $\mathbb{Q} \otimes \mathbb{P}_A$, $\mathbb{R} \otimes \mathbb{P}_A$). Obviously,

$$\sup_{y_1^n} \left(\mathbb{E}_{\mathbb{A}} \left[\widehat{L}_n \right] - \min_{j=1,2,3} L_{j,n} \right) \geqslant \mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{\mathbb{A}} \left[\widehat{L}_n \right] - \min_{j=1,2,3} L_{j,n} \right] . \tag{6.4}$$

Now,

$$\mathbb{E}_{\mathbb{Q}} \left[\min_{j=1,2,3} L_{j,n} \right] \leqslant \min_{j=1,2,3} \mathbb{E}_{\mathbb{Q}} \left[L_{j,n} \right] = \frac{n}{2} - n\varepsilon \, ,$$

whereas

$$\mathbb{E}_{\mathbb{Q}} \left[\widehat{L}_n \right] = \frac{n}{2} + \frac{1}{2} \mathbb{E}_{\mathbb{Q}} \left[N_1 \right] + \varepsilon \mathbb{E}_{\mathbb{Q}} \left[N_3 \right] - \varepsilon \mathbb{E}_{\mathbb{Q}} \left[N_2 \right] \, ,$$

where N_j is the random variable denoting the number of times the forecaster chooses the action j over the sequence $Y_1, \dots, Y_n$, given the state $U_1, \dots, U_n$ of the auxiliary randomization, for $j = 1, 2, 3$. Thus, using Fubini's theorem,

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{A}}\left[\widehat{L}_n\right] - \min_{j=1,2,3} L_{j,n}\right] \geqslant \frac{1}{2}\mathbb{E}_{\mathbb{Q}\otimes\mathbb{P}_A}[N_1] + \varepsilon\left(n - \mathbb{E}_{\mathbb{Q}\otimes\mathbb{P}_A}[N_2]\right) .$$

A similar argument shows that

$$\mathbb{E}_{\mathbb{R}}\left[\mathbb{E}_{\mathbb{A}}\left[\widehat{L}_n\right] - \min_{j=1,2,3} L_{j,n}\right] \geqslant \frac{1}{2}\mathbb{E}_{\mathbb{R}\otimes\mathbb{P}_A}[N_1] + \varepsilon\left(n - \mathbb{E}_{\mathbb{R}\otimes\mathbb{P}_A}[N_3]\right) .$$

Averaging the two inequalities we get

(6.5)
$$\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{A}}\left[\widehat{L}_n\right] - \min_{j=1,2,3} L_{j,n}\right] \geqslant \frac{1}{2}\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_1] + \varepsilon\left(n - \frac{1}{2}\left(\mathbb{E}_{\mathbb{Q}\otimes\mathbb{P}_A}[N_2] + \mathbb{E}_{\mathbb{R}\otimes\mathbb{P}_A}[N_3]\right)\right) .$$

Consider first a *deterministic* forecaster. Denote by $T_1, \dots, T_{N_1} \in \{1, \dots, n\}$ the times when the forecaster chose action 1. Since action 1 is revealing, we know the outcomes at these times, and denote them by $Z_{n+1} = (Y_{T_1}, \dots, Y_{T_{N_1}})$. Denote by K_t the (random) index of the largest integer j such that $T_j \leqslant t-1$. Each action I_t of the forecaster is determined by the random vector (of random length) $Z_t = \left(Y_1, \dots, Y_{T_{K_t}}\right)$. Since the forecaster we consider is deterministic, K_t is fully determined by Z_{n+1}. Hence, I_t may be seen as a function of Z_{n+1} rather than a function of Z_t only. This implies that, denoting by $\mathbb{P}_n$ (respectively $\mathbb{Q}_n$) the distribution of Z_{n+1} under $\mathbb{P}$ (respectively $\mathbb{Q}$), we have $\mathbb{Q}[I_t = 2] = \mathbb{Q}_n[I_t = 2]$ and $\mathbb{P}[I_t = 2] = \mathbb{P}_n[I_t = 2]$. Pinsker's inequality (see Lemma A.6 in the Appendix) then ensures that, for all t,

$$\mathbb{Q}[I_t = 2] \leqslant \mathbb{P}[I_t = 2] + \sqrt{\frac{1}{2}\mathcal{K}(\mathbb{P}_n, \mathbb{Q}_n)} , \tag{6.6}$$

where $\mathcal{K}$ denotes the Kullback-Leibler divergence. The right-hand side may be further bounded using the following lemma.

LEMMA 6.6. *Consider a deterministic forecaster. For* $0 \leqslant \varepsilon \leqslant 1/\sqrt{6}$,

$$\mathcal{K}(\mathbb{P}_n, \mathbb{Q}_n) \leqslant 6\mathbb{E}_{\mathbb{P}}[N_1]\,\varepsilon^2 .$$

PROOF. We note that $Z_{n+1} = Z_n$, except when $I_n = 1$. In this case, $Z_{n+1} = (Z_n, Y_n)$. Therefore, using the chain rule for relative entropy (see Lemma A.8 in the Appendix), as well as the first bound of Lemma A.5, we get

$$\begin{aligned}\mathcal{K}(\mathbb{P}_n, \mathbb{Q}_n) &\leqslant \mathcal{K}(\mathbb{P}_{n-1}, \mathbb{Q}_{n-1}) + \mathbb{P}[I_n = 1]\,\mathcal{K}\left(\mathbb{B}_{1/2}, \mathbb{B}_{1/2-\varepsilon}\right) \\ &\leqslant \mathcal{K}(\mathbb{P}_{n-1}, \mathbb{Q}_{n-1}) + 6\,\mathbb{P}[I_n = 1]\,\varepsilon^2 ,\end{aligned}$$

where $\mathbb{B}_p$ denotes the Bernoulli distribution with parameter p. We conclude by iterating the argument. □

Summing (6.6) over $t = 1, \dots, n$, we have proved that

$$\mathbb{E}_{\mathbb{Q}}[N_2] \leqslant \mathbb{E}_{\mathbb{P}}[N_2] + n\varepsilon\sqrt{3\mathbb{E}_{\mathbb{P}}[N_1]} ,$$

and this holds for any deterministic strategy. (Note that considering a deterministic strategy amounts to conditioning on the auxiliary randomization $U_1, \dots, U_n$.)

Consider now an arbitrary (possibly randomized) forecaster. Using Fubini's theorem and Jensen's inequality, we get

$$\mathbb{E}_{\mathbb{Q}\otimes\mathbb{P}_A}[N_2] \leqslant \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_2] + n\varepsilon\sqrt{3\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_1]} . \tag{6.7}$$

Symmetrically,

$$\mathbb{E}_{\mathbb{R}\otimes\mathbb{P}_A}[N_3] \leqslant \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_3] + n\varepsilon\sqrt{3\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_1]}\,. \tag{6.8}$$

Using $\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_2] + \mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_3] \leqslant n$, and substituting (6.7) and (6.8) into (6.5) yield

$$\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{A}}\left[\widehat{L}_n\right] - \min_{j=1,2,3} L_{j,n}\right] \geqslant \frac{1}{2}m_0 + n\varepsilon\left(\frac{1}{2} - \varepsilon\sqrt{3m_0}\right)\,, \tag{6.9}$$

where m_0 denotes $\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_A}[N_1]$. If $m_0 \leqslant 1/8$ then for $\varepsilon = 1/\sqrt{6}$ the right-hand side of (6.9) is at least $n/10$, which is greater than $n^{2/3}/5$ for $n \geqslant 8$. Otherwise, if $m_0 \geqslant 1/8$, we set $\varepsilon = \left(4\sqrt{3m_0}\right)^{-1}$, which still satisfies $0 \leqslant \varepsilon \leqslant 1/\sqrt{6}$. The lower bound then becomes

$$\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\mathbb{A}}\left[\widehat{L}_n\right] - \min_{j=1,2,3} L_{j,n}\right] \geqslant \frac{1}{2}m_0 + \frac{n}{16\sqrt{3m_0}}$$

and the right-hand side may be seen to be always bigger than $n^{2/3}/5$. An application of (6.4) concludes the proof. □

6. Internal regret

In this section we deal with the stronger notion of swap regret, see Chapter 3. For simplicity, we no longer distinguish between swap and internal regret, and refer to the former by the latter. We briefly recall that internal regret is concerned with consistent modifications of the forecasting strategy. Each of these possible modifications is parameterized by a departure function $\Phi : \{1,\dots,N\} \to \{1,\dots,N\}$. After round n, the cumulative loss of the forecaster is compared to the cumulative loss that would have been accumulated had the forecaster chosen action $\Phi(I_t)$ instead of action I_t at round t, $t = 1,\dots,n$. If such a consistent modification does not result in a much smaller accumulated loss, then the strategy is said to have small internal regret. Formally, we seek strategies achieving

$$\frac{1}{n}\sum_{t=1}^{n}\ell(I_t,y_t) - \frac{1}{n}\min_{\Phi}\sum_{t=1}^{n}\ell(\Phi(I_t),y_t) = o(1) \qquad \text{with probability } 1\,,$$

where the minimization is over all possible functions Φ. Such strategies are called Hannan consistent for the internal regret.

We recalled in Chapter 3 some internal regret minimizing strategies for the full-information case, and indicated how to extend them to the multi-armed bandit setting. We design here such a procedure in the setting of partial monitoring. The key tool is the conversion trick described in Sections 1.2 and 1.4 of Chapter 3 (see also Blum and Mansour [**BlMa05**], for a related procedure, which is however by far less convenient in an incomplete information setting). This trick converts external regret minimizing strategies into internal regret minimizing strategies, under full information, as well as in multi-armed bandit settings.

We extend it here to prediction under partial monitoring as follows. The forecaster we use is the one of Section 1.4, run however with new parameters γ_t, η_t, and new estimates of the losses. The parameters η_t and γ_t used below are tuned as in Section 3, and we consider the loss estimates $\widetilde{\ell}(i,y_t)$ defined in (6.2).

REMARK 6.5. Just like for the multi-armed bandit setting, the conversion trick of Section 1.2 of Chapter 3 does not apply directly to prediction under partial monitoring and has to be extended because of the shifting we perform on the probability distribution computed with the estimated losses (see step (2) in Figure 1, and similarly, (3.7)). We did not need such a shifting when

designing our label efficient forecasters, and consequently, the conversion was straightforward in that setting, see Remark 5.3.

THEOREM 6.4. *Consider any partial monitoring problem such that the loss and feedback matrices satisfy $\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{H}$ for some $N \times N$ matrix $\boldsymbol{K}$ with $k^* = \max\{1, \max_{i,j} |k(i,j)|\}$, and consider the forecaster described above. Let $\delta \in (0,1)$. Then, for all n, with probability at least $1-\delta$, the cumulative internal regret is bounded as*

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{\Phi} \frac{1}{n}\sum_{t=1}^{n} \ell(\Phi(I_t), y_t) \\
&\quad\leqslant\quad 9\left(\frac{(k^*)^2 N^5 \ln N}{n}\right)^{1/3}\left(1+\sqrt{\frac{3}{2}\frac{\ln(2N^2)/\delta)}{\ln N}}\right) \\
&\qquad + N\sqrt{\frac{1}{2n}\ln\frac{2N^2}{\delta}} + 4(k^*N)^{4/3} n^{-2/3} (\ln N)^{-1/3} \ln\frac{2N^2}{\delta} \\
&\qquad + \frac{1}{n}\left(2N + ((k^*N)^2 \ln N)^{1/3} + k^* N\right)\ln\frac{2N^2}{\delta}
\end{aligned}$$

where the minimum is taken over all functions $\Phi : \{1,\dots,N\} \to \{1,\dots,N\}$.

Note that with the help of Borel-Cantelli lemma, Theorem 6.4 shows that, under the same conditions on $\boldsymbol{L}$ and $\boldsymbol{H}$, the forecaster decribed above achieves Hannan consistency with respect to internal regret. Consequently, recalling that a small internal regret also implies a small external regret, we may see that the discussion before Corollary 6.2 indicates that for a given prediction problem $(\boldsymbol{L}, \boldsymbol{H})$, Hannan-consistency with respect to external regret can be achieved if and only if it can be achieved with respect to internal regret.

PROOF. First observe that it suffices to consider departure functions Φ that differ from the identity function in only one point of their domain. This follows simply from

$$\sum_{t=1}^{n} \ell(I_t, y_t) - \min_{\Phi} \sum_{t=1}^{n} \ell(\Phi(I_t), y_t) \leqslant N\left(\max_{i\neq j} \sum_{t=1}^{n} \mathbb{I}_{I_t=i}\left(\ell(i, y_t) - \ell(j, y_t)\right)\right) .$$

We now bound the right-hand side of the latter inequality.

For a given t, the estimated losses $\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)$, $i \neq j$, fall in the interval $[-k^*N/\gamma_t,\ k^*N/\gamma_t]$. Since γ_t and η_t are tuned as in Theorem 6.1, $k^*N\eta_t/\gamma_t \leqslant 1$, and we may apply Lemmas 4.3 and 4.5 to derive

$$\begin{aligned}
&\sum_{t=1}^{n}\sum_{i\neq j} u_t^{i\to j}\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t) - \min_{i\neq j}\sum_{t=1}^{n}\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t) \\
&\qquad\qquad \leqslant \frac{2\ln N(N-1)}{\eta_{n+1}} + \sum_{t=1}^{n}\eta_t \sum_{i\neq j} u_t^{i\to j}\left(\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)\right)^2 .
\end{aligned}$$

We then proceed as in the proof of Theorem 3.2, and get exactly (3.9),

$$\begin{aligned}
&\max_{i\neq j}\sum_{t=1}^{n} p_{i,t}\left(\widetilde{\ell}(i, y_t) - \widetilde{\ell}(j, y_t)\right) \\
&\quad\leqslant\quad \frac{4\ln N}{\eta_{n+1}} + \sum_{t=1}^{n}\eta_t\sum_{i\neq j} u_t^{i\to j}\left(\widetilde{\ell}(\boldsymbol{p}_t^{i\to j}, y_t)\right)^2 + \sum_{t=1}^{n}\frac{\gamma_t}{N}\sum_{i\neq j} u_t^{i\to j}\left(\widetilde{\ell}(i, y_t) - \widetilde{\ell}(j, y_t)\right) .
\end{aligned} \tag{6.10}$$

Now, we apply Bernstein's inequality (Lemma A.4) several times again and mimic the proofs of Lemmas 6.1 and 6.2. For all pairs $i \neq j$, with probability at least $1 - \delta/(2N(N-1)+2)$,

$$\sum_{t=1}^{n} p_{i,t}\left(\widetilde{\ell}(i,y_t) - \widetilde{\ell}(j,y_t)\right) \geqslant \sum_{t=1}^{n} p_{i,t}\left(\ell(i,y_t) - \ell(j,y_t)\right) - \left(\sqrt{4(k^*N)^2\left(\sum_{t=1}^{n}\frac{1}{\gamma_t}\right)\ln\frac{2N(N-1)+2}{\delta}} + \frac{2\sqrt{2}}{3}\left(1+\frac{k^*N}{\gamma_n}\right)\ln\frac{2N(N-1)+2}{\delta}\right). \tag{6.11}$$

Similarly to Lemma 6.3, we also have, with probability at least $1 - \delta/(2N(N-1)+2)$,

$$\sum_{t=1}^{n}\eta_t\sum_{i\neq j}u_t^{i\to j}\left(\widetilde{\ell}(\boldsymbol{p}_t^{i\to j},y_t)\right)^2 \leqslant \sum_{t=1}^{n}\eta_t\frac{(k^*N)^2}{\gamma_t} + \sqrt{2(k^*N)^4\left(\sum_{t=1}^{n}\frac{\eta_t^2}{\gamma_t^3}\right)\ln\frac{2N(N-1)+2}{\delta}} + \frac{\sqrt{2}}{3}\ln\frac{2N(N-1)+2}{\delta} \tag{6.12}$$

whereas, similarly to Lemma 6.4, with probability at least $1 - \delta/(2N(N-1)+2)$,

$$\sum_{t=1}^{n}\frac{\gamma_t}{N}\sum_{i\neq j}u_t^{i\to j}\left(\widetilde{\ell}(i,y_t) - \widetilde{\ell}(j,y_t)\right) \leqslant \frac{1}{N}\sum_{t=1}^{n}\gamma_t + \sqrt{4(k^*)^2\left(\sum_{t=1}^{n}\gamma_t\right)\ln\frac{2N(N-1)+2}{\delta}} + \frac{\sqrt{2}}{3}\left(k^*+\frac{\gamma_1}{N}\right)\ln\frac{2N(N-1)+2}{\delta}. \tag{6.13}$$

We then use the Hoeffding-Azuma inequality (see Lemma A.2) $N(N-1)$ times to show that for every pair $i \neq j$, with probability at least $1 - \delta/(2N(N-1)+2)$,

$$\sum_{t=1}^{n} p_{i,t}\left(\ell(i,y_t) - \ell(j,y_t)\right) \geqslant \sum_{t=1}^{n}\mathbb{I}_{I_t=i}\left(\ell(i,y_t) - \ell(j,y_t)\right) - \sqrt{2n\ln\frac{N(N-1)+3}{\delta}}. \tag{6.14}$$

Finally, we substitute inequalities (6.11)–(6.14) into (6.10) and use a union-of-event bound to obtain that, with probability at least $1-\delta$,

$$\begin{aligned}
&\max_{i\neq j}\sum_{t=1}^{n}\mathbb{I}_{I_t=i}\left(\ell(i,y_t) - \ell(j,y_t)\right) \\
&\leqslant \frac{4\ln N}{\eta_{n+1}} \\
&\quad + \sqrt{4(k^*N)^2\left(\sum_{t=1}^{n}\frac{1}{\gamma_t}\right)\ln\frac{1}{\delta'}} + \frac{2\sqrt{2}}{3}\left(1+\frac{k^*N}{\gamma_n}\right)\ln\frac{1}{\delta'} \\
&\quad + \sum_{t=1}^{n}\eta_t\frac{(k^*N)^2}{\gamma_t} + \sqrt{2(k^*N)^4\left(\sum_{t=1}^{n}\frac{\eta_t^2}{\gamma_t^3}\right)\ln\frac{1}{\delta'}} + \frac{\sqrt{2}}{3}\ln\frac{1}{\delta'} \\
&\quad + \frac{1}{N}\sum_{t=1}^{n}\gamma_t + \sqrt{4(k^*)^2\left(\sum_{t=1}^{n}\gamma_t\right)\ln\frac{1}{\delta'}} + \frac{\sqrt{2}}{3}\left(k^*+\frac{\gamma_1}{N}\right)\ln\frac{1}{\delta'} \\
&\quad + \sqrt{2n\ln\frac{1}{\delta'}},
\end{aligned}$$

where we used the notation $\delta' = \delta/(2N(N-1)+2)$, with $\delta' \geqslant \delta/(2N^2)$ when $N \geqslant 2$. The proof is now concluded as that of Theorem 6.1. □

7. Random feedback

Several authors consider an extended setup in which the feedbacks are random variables. See Rustichini [**Rus99**], Mannor and Shimkin [**MaSh03**], Weissman and Merhav [**WeMe01**], Weissman, Merhav and Somekh-Baruch [**WeMeSo01**] for examples. In this section we briefly point out that most of the results of this chapter extend effortlessly to this more general case.

To describe the model, denote by $\Delta(\mathcal{S})$ the set of all probability distributions over the set of signals $\mathcal{S}$. The signaling structure is formed by a collection of NM probability distributions $\mu_{(i,j)}$ over $\mathcal{S}$, for $i = 1, \ldots, N$ and $j = 1, \ldots, M$. At each round, the forecaster now observes a random variable $H(I_t, y_t)$, drawn independently from all the other random variables, with distribution $\mu_{(I_t,y_t)}$. More precisely, we assume without loss of generality that $H(I_t, y_t)$ is a function of $\mu_{(I_t,y_t)}$ and V_t, where $(V_1, V_2, \ldots)$ is an i.i.d. sequence of random variables with uniform law over $[0,1]$, independent of all the other random variables. This sequence is called the external randomization. All expectations and probabilities here are understood with respect to the probability space formed by the product of the external randomization and the forecaster's randomization.

We may easily generalize the results of Theorems 6.1 and 6.4 to the case of random feedbacks. As above, each element of $\mathcal{S}$ is encoded by a real number in $[-1,1]$. Let $\boldsymbol{E}$ be the $N \times M$ matrix whose elements are given by the expectations of the random variables $H(i,j)$. Theorems 6.1 and 6.4 remain true under the condition that there exists a matrix $\boldsymbol{K}$ such that $\boldsymbol{L} = \boldsymbol{K}\,\boldsymbol{E}$. The only necessary modification is how the losses are estimated. Here the forecaster uses the estimates

$$\breve{\ell}(i, y_t) = \frac{k(i, I_t) H(I_t, y_t)}{p_{I_t,t}} \quad i = 1, \ldots, N$$

instead of the estimates defined in Section 3. Conditioned on $I_1, \ldots, I_{t-1}$, the expectation of $\breve{\ell}(i, y_t)$ is the loss $\ell(i, y_t)$. Since this, together with boundedness, are the only conditions that were needed in the proofs, the extension of the results to this more general framework is immediate.

The results of Section 4 may be generalized to the case of random feedbacks as well. For example, to construct $\boldsymbol{H}'$ when $\boldsymbol{H}$ is a matrix of probability distributions over $\mathcal{S}$, we proceed as follows: for $1 \leqslant i \leqslant N$ and $s \in \mathcal{S}$, denote by $H_{(i,s)}$ the row vector of elements in $[0,1]$, such that the k-th element of $H_{(i,s)}$ is $\mu_{(i,k)}(s)$. Now, the $((k_1-1)m + k_2)$-th row of $\boldsymbol{H}'$, $1 \leqslant k_1 \leqslant N$, $1 \leqslant k_2 \leqslant m$, is $H_{(k_1, s_{k_2})}$. All the other details of the construction and the proofs go through.

Part 3

Internal regret for general convex loss functions

CHAPTER 7

Internal regret in on-line portfolio selection

This chapter extends the game-theoretic notion of internal regret to the case of on-line potfolio selection problems. New sequential investment strategies are designed to minimize the cumulative internal regret for all possible market behaviors. Some of the introduced strategies, apart from achieving a small internal regret, achieve an accumulated wealth almost as large as that of the best constantly rebalanced portfolio. It is argued that the low-internal-regret property is related to stability and experiments on real stock exchange data demonstrate that the new strategies achieve better returns compared to some known algorithms.

Contents

This chapter is based on the article [**StLu05**], invited by *Machine Learning* after the extended abstract [**StLu03**] was awarded at COLT'03. Sections 6 and 7 are however published here for the first time.

1. Introduction

The problem of sequential portfolio allocation is well-known to be closely related to the on-line prediction of individual sequences under expert advice, see, for example, [**Cov91**], [**CoOr96**], [**HeScSiWa98**], [**OrCo98**], [**BlKa99**], [**CeLu00**]. The goal in the sequential investment problem is to distribute one's capital in each trading period among a certain number of stocks such that the total achieved wealth is almost as large as the wealth of the largest in a certain class of investment strategies. This problem, known as the minimization of the worst-case logarithmic wealth ratio, is easily seen to be the generalization of an external regret minimization problem in the "expert" setting under the logarithmic loss function. The main purpose of this chapter is to extend the notion of internal regret to the sequential investment problem, understand its relationship to the worst-case logarithmic wealth ratio, and design investment strategies minimizing this new notion of regret. The definition of internal regret given here has a natural interpretation and the investment strategies designed to minimize it have several desirable properties both in theory and in the experimental study described in the Appendix.

This chapter is organized as follows. In Section 2 the sequential portfolio selection problem is described, and basic properties of Cover's universal portfolio and the EG investment strategy are discussed. In Section 3 we introduce the notion of internal regret for sequential portfolio selection, and describe some basic properties. In Section 4 new investment strategies are presented aiming at the minimization of the internal regret (and these strategies are further investigated in Section 6). In Section 5 the notion of internal regret is generalized for an uncountable class of investment strategies and an algorithm inspired by Cover's universal portfolio is proposed which minimizes the new notion of internal regret. Section 7 explains the modifications needed for the algorithms of Section 5 to be competitive in a market with transaction costs.

2. Sequential portfolio selection

In this section we describe the problem of sequential portfolio selection, recall some previous results, and take a new look at the EG strategy of [**HeScSiWa98**].

A *market vector* $\boldsymbol{x} = (x_1, \ldots, x_N)$ for N assets is a vector of nonnegative numbers representing price relatives for a given trading period. In other words, the quantity $x_i \geqslant 0$ denotes the ratio of closing to opening price of the i-th asset for that period. Hence, an initial wealth invested in the N assets according to fractions $Q_1, \ldots, Q_N$ multiplies by a factor of $\sum_{i=1}^{N} x_i Q_i$ at the end of period. The market behavior during n trading periods is represented by a sequence $\boldsymbol{x}_1^n = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ of market vectors. $x_{j,t}$, the j-th component of $\boldsymbol{x}_t$, denotes the factor by which the wealth invested in asset j increases in the t-th period. We denote the probability simplex in $\mathbb{R}^N$ by $\mathcal{X}$.

An *investment strategy* Q for n trading periods consists in a sequence $\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_n$ of vector-valued functions $\boldsymbol{Q}_t : (\mathbb{R}_+^N)^{t-1} \to \mathcal{X}$, where the i-th component $Q_{i,t}(\boldsymbol{x}_1^{t-1})$ of the vector $\boldsymbol{Q}_t(\boldsymbol{x}_1^{t-1})$ denotes the fraction of the current wealth invested in the i-th asset at the beginning of the t-th period based on the past market behavior $\boldsymbol{x}_1^{t-1}$. We use

$$S_n(Q, \boldsymbol{x}_1^n) = \prod_{t=1}^{n} \left(\sum_{i=1}^{N} x_{i,t} Q_{i,t}(\boldsymbol{x}_1^{t-1}) \right)$$

to denote the wealth factor of strategy Q after n trading periods, and often omit the dependency in Q and $\boldsymbol{x}_1^n$ whenever both are understood. In this case, we simply write S_n to refer to $S_n(Q, \boldsymbol{x}_1^n)$.

The simplest examples of investment strategies are the so called *buy-and-hold* strategies. A buy-and-hold strategy simply distributes its initial wealth among the N assets according to some

distribution $\boldsymbol{Q}_1 \in \mathcal{X}$ before the first trading period, and does not trade anymore, which amounts to investing, at day t and for $1 \leqslant i \leqslant N$, as

$$Q_{i,t}(\boldsymbol{x}_1^{t-1}) = \frac{Q_{i,1} \prod_{s=1}^{t-1} x_{i,s}}{\sum_{k=1}^{N} Q_{k,1} \prod_{s=1}^{t-1} x_{k,s}} \, .$$

The wealth factor of such a strategy, after n periods, is simply

$$S_n(Q, \boldsymbol{x}_1^n) = \sum_{j=1}^{N} Q_{j,1} S_n(j) \, ,$$

where

$$S_n(j) = \prod_{t=1}^{n} x_{j,t}$$

is the accumulated wealth of stock j. Clearly, the wealth factor of any buy-and-hold strategy is at most as large as the gain $\max_{j=1,\ldots,N} S_n(j)$ of the best stock over the investment period, and achieves this maximal wealth if $\boldsymbol{Q}_1$ concentrates on the best stock.

Another simple and important class of investment strategies is the class of *constantly rebalanced portfolios*. Such a strategy B is parametrized by a probability vector $\boldsymbol{B} = (B_1, \ldots, B_N) \in \mathcal{X}$, and simply $\boldsymbol{Q}_t(\boldsymbol{x}_1^{t-1}) = \boldsymbol{B}$ regardless of t and the past market behavior $\boldsymbol{x}_1^{t-1}$. Thus, an investor following such a strategy rebalances, at every trading period, his current wealth according to the distribution $\boldsymbol{B}$ by investing a proportion B_1 of his wealth in the first stock, a proportion B_2 in the second stock, etc. The wealth factor achieved after n trading periods is

$$S_n(\boldsymbol{B}, \boldsymbol{x}_1^n) = \prod_{t=1}^{n} \left(\sum_{i=1}^{N} x_{i,t} B_i \right) \, .$$

In [**CoTh91**], it is shown that the constantly rebalanced portfolios are the optimal investment strategies in an i.i.d. market.

Now given a class $\mathcal{Q}$ of investment strategies, we define the *worst-case logarithmic wealth ratio* of strategy P by

$$W_n(P, \mathcal{Q}) = \sup_{\boldsymbol{x}_1^n} \sup_{Q \in \mathcal{Q}} \ln \frac{S_n(Q, \boldsymbol{x}_1^n)}{S_n(P, \boldsymbol{x}_1^n)} \, .$$

The worst-case logarithmic wealth ratio is the analog of the external regret in the sequential portfolio selection problem. $W_n(P, \mathcal{Q}) = o(n)$ means that the investment strategy P achieves the same exponent of growth as the best reference strategy in the class $\mathcal{Q}$ for all possible market behaviors.

For example, it is immediate to see that if $\mathcal{Q}$ is the class of all buy-and-hold strategies, then if P is chosen to be the buy-and-hold strategy based on the uniform distribution $\boldsymbol{Q}_1$, then $W_n(P, \mathcal{Q}) \leqslant \ln N$.

The class of constantly rebalanced portfolios is significantly richer and achieving a small worst-case logarithmic wealth ratio is a greater challenge. Cover's *universal portfolio* [**Cov91**] was the first example to achieve this goal. The universal portfolio strategy P is defined by

$$P_{j,t}(\boldsymbol{x}_1^{t-1}) = \frac{\int_{\mathcal{X}} B_j S_{t-1}(\boldsymbol{B}, \boldsymbol{x}_1^{t-1}) \phi(\boldsymbol{B}) \, \mathrm{d}\boldsymbol{B}}{\int_{\mathcal{X}} S_{t-1}(\boldsymbol{B}, \boldsymbol{x}_1^{t-1}) \phi(\boldsymbol{B}) \, \mathrm{d}\boldsymbol{B}} \, , \quad j = 1, \ldots, N, t = 1, \ldots, n$$

where ϕ is a density function on $\mathcal{X}$. In the simplest case ϕ is the uniform density over $\mathcal{X}$. In that case, the worst-case logarithmic wealth ratio of P with respect to the class $\mathcal{Q}$ of all universal portfolios satisfies

$$W_n(P, \mathcal{Q}) \leqslant (N-1) \ln(n+1) \, .$$

If the universal portfolio is defined using the Dirichlet$(1/2,\dots,1/2)$ density ϕ, then the bound improves to

$$W_n(P,\mathcal{Q}) \leqslant \frac{N-1}{2}\ln n + \ln\frac{\Gamma(1/2)^N}{\Gamma(N/2)} + \frac{N-1}{2}\ln 2 + o(1)\;,$$

see [**CoOr96**]. The worst-case performance of the universal portfolio is basically unimprovable (see [**OrCo98**]) but it has some practical disadvantages, including computational difficulties for not very small values of N. [**HeScSiWa98**] suggest their EG strategy to overcome these difficulties.

The EG strategy is defined by

$$P_{i,t+1} = \frac{P_{i,t}\exp\left(\eta x_{i,t}/\boldsymbol{P}_t\cdot\boldsymbol{x}_t\right)}{\sum_{j=1}^N P_{j,t}\exp\left(\eta x_{j,t}/\boldsymbol{P}_t\cdot\boldsymbol{x}_t\right)}\;. \tag{7.1}$$

[**HeScSiWa98**] prove that if the market values $x_{i,t}$ all fall between the positive constants m and M, then the worst-case logarithmic wealth ratio of the EG investment strategy is bounded by

$$\frac{\ln N}{\eta} + \frac{n\eta}{8}\frac{M^2}{m^2} = \frac{M}{m}\sqrt{\frac{n}{2}\ln N}\;,$$

where the equality holds for the choice $\eta = (m/M)\sqrt{(8\ln N)/n}$. Here we give a simple new proof of this result, mostly because the main idea is at the basis of other arguments that follow. Recall that the worst-case logarithmic wealth ratio is

$$\max_{\boldsymbol{x}_1^n}\max_{\mathbf{B}\in\mathcal{X}}\ln\frac{\prod_{t=1}^n \mathbf{B}\cdot\boldsymbol{x}_t}{\prod_{t=1}^n \boldsymbol{P}_t\cdot\boldsymbol{x}_t}$$

where in this case the first maximum is taken over market sequences satisfying the boundedness assumption. By using the elementary inequality $\ln(1+u)\leqslant u$, we obtain

$$\begin{aligned}
\ln\frac{\prod_{t=1}^n \mathbf{B}\cdot\boldsymbol{x}_t}{\prod_{t=1}^n \boldsymbol{P}_t\cdot\boldsymbol{x}_t} &= \sum_{t=1}^n \ln\left(1+\frac{(\mathbf{B}-\boldsymbol{P}_t)\cdot\boldsymbol{x}_t}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t}\right)\\
&\leqslant \sum_{t=1}^n\sum_{i=1}^N \frac{(B_i - P_{i,t})\,x_{i,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t}\\
&= \sum_{t=1}^n\left(\sum_{j=1}^N\sum_{i=1}^N P_{i,t}\frac{B_j x_{j,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t} - \sum_{i=1}^N\sum_{j=1}^N B_j\frac{P_{i,t}x_{i,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t}\right)\\
&= \sum_{j=1}^N B_j\left(\sum_{t=1}^n\sum_{i=1}^N P_{i,t}\left(\frac{x_{j,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t} - \frac{x_{i,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t}\right)\right)\;. \qquad (7.2)
\end{aligned}$$

Under the boundedness assumption $0 < m \leqslant x_{i,t} \leqslant M$, the quantities

$$\ell_{i,t} = M/m - x_{i,t}/(\boldsymbol{P}_t\cdot\boldsymbol{x}_t)$$

are within $[0, M/m]$ and can therefore be interpreted as bounded loss functions. Thus, the minimization of the above upper bound on the worst-case logarithmic wealth ratio may be cast as a sequential prediction problem as described in Chapter 2. Observing that the EG investment algorithm is just the exponentially weighted average predictor for this prediction problem, and using the performance bound of Theorem 2.1 we obtain the cited inequality of [**HeScSiWa98**].

Note that in (7.1), we could replace the fixed η by a time-adaptive $\eta_t = (m/M)\sqrt{(8\ln N)/t}$. Applying Theorem 2.3 to the linear upper bound (7.2), we may prove that this still leads to a worst-case logarithmic wealth ratio less than something of the order of $(M/m)\sqrt{n\ln N}$.

REMARK 7.1. *(Sub-optimality of the* EG *investment strategy.)* Using the approach of bounding the worst-case logarithmic wealth ratio linearly as above is inevitably suboptimal. Indeed, the right-hand side of the linear upper bounding

$$\sum_{j=1}^{N} B_j \left(\sum_{t=1}^{n} \left(\sum_{i=1}^{N} P_{i,t} \ell_{i,t} \right) - \ell_{j,t} \right) = \sum_{j=1}^{N} B_j \sum_{i=1}^{N} \left(\sum_{t=1}^{n} P_{i,t} \left(\ell_{i,t} - \ell_{j,t} \right) \right)$$

is maximized for a constantly rebalanced portfolio $\mathbf{B}$ lying in a corner of the simplex $\mathcal{X}$, whereas the left-hand side is concave in $\mathbf{B}$ and therefore is possibly maximized in the interior of the simplex. Thus, no algorithm trying to minimize (in a worst-case sense) the linear upper bound on the external regret can be minimax optimal. However, as it is shown in [**HeScSiWa98**], on real data good performance may be achieved.

Note also that the bound obtained for the worst-case logarithmic wealth ratio of the EG strategy grows as $\sqrt{n}$ whereas that of Cover's universal portfolio has only a logarithmic growth. In [**HeScSiWa98**] it is asked whether the suboptimal bound for the EG strategy is an artifact of the analysis or it is inherent in the algorithm. The next simple example shows that no bound of a smaller order than $\sqrt{n}$ holds. Consider a market with two assets and market vectors $\boldsymbol{x}_t = (1, 1-\varepsilon)$, for all t. Then every wealth allocation $\boldsymbol{P}_t$ satisfies $1 - \varepsilon \leqslant \boldsymbol{P}_t \cdot \boldsymbol{x}_t \leqslant 1$. Now, the best constantly rebalanced portfolio is clearly $(1, 0)$, and the worst-case logarithmic wealth ratio is simply

$$\sum_{t=1}^{n} \ln \frac{1}{1 - P_{2,t}\varepsilon} \geqslant \sum_{t=1}^{n} P_{2,t}\varepsilon \; .$$

In the case of the EG strategy,

$$\begin{aligned} P_{2,t} &= \frac{\exp\left(\eta \sum_{s=1}^{t-1} \frac{(1-\varepsilon)}{\boldsymbol{P}_s \cdot \boldsymbol{x}_s}\right)}{\exp\left(\eta \sum_{s=1}^{t-1} \frac{1}{\boldsymbol{P}_s \cdot \boldsymbol{x}_s}\right) + \exp\left(\eta \sum_{s=1}^{t-1} \frac{(1-\varepsilon)}{\boldsymbol{P}_s \cdot \boldsymbol{x}_s}\right)} \\ &= \frac{\exp\left(-\eta\varepsilon \sum_{s=1}^{t-1} \frac{1}{\boldsymbol{P}_s \cdot \boldsymbol{x}_s}\right)}{1 + \exp\left(-\eta\varepsilon \sum_{s=1}^{t-1} \frac{1}{\boldsymbol{P}_s \cdot \boldsymbol{x}_s}\right)} \\ &\geqslant \frac{\exp\left(-\eta \left(\varepsilon/(1-\varepsilon)\right)(t-1)\right)}{2} \; . \end{aligned}$$

Thus, the logarithmic wealth ratio of the EG algorithm is lower bounded by

$$\begin{aligned} \sum_{t=1}^{n} \varepsilon \frac{\exp\left(-\eta \left(\varepsilon/(1-\varepsilon)\right)(t-1)\right)}{2} &= \frac{\varepsilon}{2} \frac{1 - \exp\left(-\eta \left(\varepsilon/(1-\varepsilon)\right) n\right)}{1 - \exp\left(-\eta \left(\varepsilon/(1-\varepsilon)\right)\right)} \\ &= \frac{1}{2}\sqrt{\frac{n}{8 \ln N}} + o(\sqrt{n}) \; . \end{aligned}$$

3. Internal regret of investment strategies

The aim of this section is to introduce the notion of internal regret to the sequential investment problem. In the latter, the loss function we consider is defined by $\ell'(\boldsymbol{Q}, \boldsymbol{x}) = -\ln(\boldsymbol{Q} \cdot \boldsymbol{x})$ for a portfolio $\boldsymbol{Q}$ and a market vector $\boldsymbol{x}$. This is no longer a linear function of $\boldsymbol{Q}$ (as this was the case, for instance, in Chapters 2 and 4 for the expected loss of the predictor in the setting of randomized prediction under expert advice).

Recall that in the framework of sequential prediction described in Chapter 3, the cumulative internal regret $R_{(i,j),n}$ for the pair of experts (i, j) may be interpreted as how much the predictor would have gained, had he replaced all values $P_{i,t}$ ($t \leqslant n$) by zero and all values $P_{j,t}$ by $P_{i,t} + P_{j,t}$.

Analogously, given an investment strategy $P = (\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots)$, we may define the *internal regret of P with respect to the pair of assets (i, j) at day t* (where $1 \leqslant i, j \leqslant N$) by

$$\widetilde{r}_{(i,j),t} = \ell'(\boldsymbol{P}_t, \boldsymbol{x}_t) - \ell'(\boldsymbol{P}_t^{i\to j}, \boldsymbol{x}_t) = \ln \frac{\boldsymbol{P}_t^{i\to j} \cdot \boldsymbol{x}_t}{\boldsymbol{P}_t \cdot \boldsymbol{x}_t}$$

where the probability vector $\boldsymbol{P}_t^{i\to j}$ is defined such that its i-th component equals zero, its j-th component equals $P_{j,t} + P_{i,t}$, and all other components are equal to those of $\boldsymbol{P}_t$. $\widetilde{r}_{(i,j),t}$ expresses the regret the investor using strategy P suffers after trading day t of not having invested all the capital he invested in stock i in stock j instead. The *cumulative internal regret of P with respect to the pair* (i, j) after n trading periods is simply

$$\widetilde{R}_{(i,j),n} = \sum_{t=1}^{n} \widetilde{r}_{(i,j),n} \,.$$

This notion of internal regret in on-line portfolio selection may be seen as a special case of the definition of internal regret for general loss functions proposed in Chapter 8, with the class of departure functions given by those functions that move all probability mass from a given component to another one. In Section 5.2, we study internal regret with respect to a much larger class, whose size is of the power of the continuum. It is a desirable property of an investment strategy that its cumulative internal regret grows sub-linearly for all possible pairs of assets, independently of the market outcomes. Indeed, otherwise the owners of the portfolio could exhibit simple[1] modifications of the betting strategy which would have led to exponentially larger wealth. In this sense, the notion of internal regret is a measure of the efficiency of the strategy: the aim of the broker is not that the owner of the portfolio gets rich, but that he cannot criticize easily the chosen strategy. Note that the worst-case logarithmic wealth ratio corresponds to the case when the owner compares his achieved wealths to those obtained by others who have different brokers. Based on this, we define the *internal regret* of the investment strategy P by

$$\widetilde{R}_n = \max_{1\leqslant i,j\leqslant N} \widetilde{R}_{(i,j),n}$$

and ask whether it is possible to guarantee that $\widetilde{R}_n = o(n)$ for all possible market sequences. Thus, an investor using a strategy with a small internal regret is guaranteed that for any pair of stocks the total regret of not investing in one stock instead of the other becomes negligible. (Note that in Section 5.2 we introduce a richer class of possible departures from the original investment strategies.)

The next two examples show that it is not trivial to achieve a small internal regret. Indeed, the buy-and-hold and EG investment strategies have linearly increasing internal regret for some bounded market sequences. (We do not know whether Cover's [**Cov91**] universal portfolio suffers from this drawback or not, but guess it does so.)

EXAMPLE 7.1. *(Buy-and-hold strategies may have large internal regret.)* Consider a market with $N = 3$ assets which evolves according to the following repeated scheme:

$$(1-\varepsilon, \varepsilon, \varepsilon),\ (\varepsilon, 1-\varepsilon, 1-\varepsilon),\ (1-\varepsilon, \varepsilon, \varepsilon),\ (\varepsilon, 1-\varepsilon, 1-\varepsilon),\ \ldots$$

where $\varepsilon < 1$ is a fixed positive number.

[1] We assume here that the broker's customers only think of simple modifications, such as putting all the wealth from one stock to another one.

The buy-and-hold strategy, which distributes its initial wealth uniformly among the assets invests, at odd t's, with

$$\boldsymbol{P}_t = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) , \quad \text{so that } \boldsymbol{P}_t^{2\to 1} = \left(\frac{2}{3}, 0, \frac{1}{3}\right) ,$$

and at even t's, with

$$\boldsymbol{P}_t = \left(\frac{1-\varepsilon}{1+\varepsilon}, \frac{\varepsilon}{1+\varepsilon}, \frac{\varepsilon}{1+\varepsilon}\right) , \quad \text{so that } \boldsymbol{P}_t^{2\to 1} = \left(\frac{1}{1+\varepsilon}, 0, \frac{\varepsilon}{1+\varepsilon}\right) .$$

Straightforward calculation now shows that for an even n, the cumulative internal regret $\widetilde{R}_{(2,1),n}$ of this strategy equals

$$\frac{n}{2}\left(\ln \frac{(2-\varepsilon)^2}{3(1-\varepsilon)(1+\varepsilon)}\right) ,$$

showing that even for bounded markets, the naive buy-and-hold strategy may incur a large internal regret. Later we will see a generalization of buy-and-hold with small internal regret.

EXAMPLE 7.2. *(The* EG *strategy may have large internal regret.)* The next example, showing that for some market sequence the EG algorithm of [**HeScSiWa98**] has a linearly growing internal regret, is inspired by Example 3.1 above. Consider a market of three stocks A, B, and C. Divide the n trading periods into three different regimes of lengths n_1, n_2, and n_3. The wealth ratios (which are constant in each regime) are summarized in Table 1. We show that it is possible to set

Regimes	$x_{A,t}$	$x_{B,t}$	$x_{C,t}$
$1 \leqslant t \leqslant T_1 = n_1$	2	1	0.5
$T_1 + 1 \leqslant t \leqslant T_2 = n_1 + n_2$	1	2	0.5
$T_2 + 1 \leqslant t \leqslant T_3 = n$	1	2	2.05

TABLE 1. The market vectors for Example 7.2.

n_1, n_2, and n_3 in such a way that the cumulative internal regret $R_{(B,C),n}$ is lower bounded by a positive constant times n for n sufficiently large.

The internal regret of B versus C can be lower bounded by using the inequality $\ln(1+u) \leqslant u$:

$$\sum_{t=1}^{n} \ln \frac{\mathbf{Q}_t^{B\to C} \cdot \mathbf{x}_t}{\mathbf{Q}_t \cdot \mathbf{x}_t} \geqslant \sum_{t=1}^{n} Q_{B,t} \left(\frac{x_{C,t}}{\mathbf{Q}_t^{B\to C} \cdot \mathbf{x}_t} - \frac{x_{B,t}}{\mathbf{Q}_t^{B\to C} \cdot \mathbf{x}_t}\right),$$

where the difference in the parenthesis is larger than -1 in the first regime, -3 in the second one and $0.05/2.05$ in the third one. It suffices now to estimate $Q_{B,t}$:

$$Q_{B,t} = \frac{e^{\eta G_{B,t}}}{e^{\eta G_{A,t}} + e^{\eta G_{B,t}} + e^{\eta G_{C,t}}}, \tag{7.3}$$

where

$$\eta = 4.1\sqrt{\frac{8\ln 3}{n}} \hat{=} \frac{1}{C_\eta \sqrt{n}} \quad \text{and} \quad G_{B,t} = \sum_{s=1}^{t} \frac{x_{B,s}}{\mathbf{Q}_s \cdot \mathbf{x}_s}$$

(and similarly for the two other stocks).

We take $n_1 = dn$, where $d > 0$ will be determined later. In the first regime, a sufficient condition for $Q_{B,t} \leqslant \varepsilon$ is that $e^{\eta G_{B,t}}/e^{\eta G_{A,t}} \leqslant \varepsilon$, which can be ensured by

$$G_{A,t} - G_{B,t} = \sum_{s=1}^{t} \frac{1}{\mathbf{Q}_s \cdot \mathbf{x}_s} \geqslant \frac{-\ln \varepsilon}{\eta},$$

which is implied, since $\mathbf{Q}_s \cdot \mathbf{x}_s \leqslant 2$, by

$$t \geqslant t_0 = 2C_\eta \left(-\ln \varepsilon\right) \sqrt{n}.$$

In the second regime, the $Q_{B,t}$'s increase. Let T_2 denote the first time instant t when $Q_{B,t} \geqslant 1/2$, and denote by $n_2 = T_2 - T_1$ the length of this second regime. Now, it is easy to see that $n_2 \geqslant n_1/4$ and $n_2 \leqslant 4n_1 + (2\ln 2)C_\eta\sqrt{n} \leqslant 5dn$, for n sufficiently large. Moreover, the number of times that $Q_{B,t}$ is larger than ε in this regime is less than

$$C_\eta \left(\ln\left(2\frac{1-\varepsilon}{\varepsilon}\right)\right)\sqrt{n}.$$

At the beginning of the third regime, we then have $Q_{B,t} \geqslant 1/2$, which means that $G_{A,t} \leqslant G_{B,t}$ and $G_{C,t} \leqslant G_{B,t}$. The first inequality remains true during the whole regime and we set n_3 such that the second one also remains true. This will imply that $Q_{B,t} \geqslant 1/3$ during the third regime. Now by the bounds on $\mathbf{Q}_s \cdot \mathbf{x}_s$ in the different regimes, a sufficient condition on n_3 is

$$0.05n_3 \leqslant \frac{n_1}{4} + \frac{3n_2}{4},$$

which, recalling the lower bound $n_2 \geqslant n_1/4$, is implied by

$$n_3 \leqslant \frac{35}{4}dn.$$

It remains to set the value of d. We have to ensure that n_3 is not larger than $35dn/4$ and that it is larger than γn, where γ is a universal constant denoting the fraction of time spent in the third regime. That is, we have to find d and γ such that

$$\begin{cases} d + 5d + \gamma & \leqslant & 1 \\ d + \frac{1}{4}d + \frac{35}{4}d & \geqslant & 1\,, \end{cases}$$

where we used $n_1/n + n_2/n + n_3/n = 1$ and the various bounds and constraints described above. $\gamma = 1/7$ and $d = 1/7$ are adequate choices.

Summarizing, we have proved the following lower bound on the internal regret

$$\sum_{t=1}^{n} \ln \frac{\mathbf{Q}_t^{B\to C} \cdot \mathbf{x}_t}{\mathbf{Q}_t \cdot \mathbf{x}_t} \geqslant \frac{1}{3}\gamma\frac{0.05}{2.05}n - \varepsilon\left(3(1-\gamma)\right)n + \Omega\left(\left(-\ln\varepsilon\right)\sqrt{n}\right),$$

and the proof that the EG strategy has a large internal regret is concluded by choosing $\varepsilon > 0$ small enough (for instance, $\varepsilon = 1/5000$).

REMARK 7.2. A mixture in the buy-and-hold sense of no-internal-regret investment strategies is still a no-internal-regret minimizing strategy. Its internal regret is less than the maximum of the internal regrets of the original strategies.

4. Investment strategies with small internal regret

The investment algorithm introduced in the next section has the surprising property that, apart from a guaranteed sublinear internal regret, it also achieves a sublinear worst-case logarithmic wealth ratio not only with respect to the class of buy-and-hold strategies, but also with respect to the class of all constantly rebalanced portfolios.

4.1. A strategy with small internal and external regrets. The investment strategy introduced in this section – which we call B1EXP – is based on the same kind of linear upper bound on the internal regret as the one that was used in our proof of the performance of the EG strategy in Section 2. This strategy may be seen as the algorithm that results from an application of the conversion trick explained in Section 1.2 of Chapter 3 to the EG strategy. However, this only proves the no-internal-regret property. Since the worst-case logarithmic wealth ratio is also minimized, we provide a detailed analysis below.

The same argument as for the EG strategy may be used to upper bound the cumulative internal regret as

$$\begin{aligned}\widetilde{R}_{(i,j),n} &= \sum_{t=1}^{n} \ln\left(\boldsymbol{P}_t^{i\to j}\cdot\boldsymbol{x}_t\right) - \ln\left(\boldsymbol{P}_t\cdot\boldsymbol{x}_t\right) \\ &\leqslant \sum_{t=1}^{n} P_{i,t}\left(\frac{x_{j,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t} - \frac{x_{i,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t}\right) .\end{aligned}$$

Introducing again

$$\ell_{i,t} = -\frac{x_{i,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t} ,$$

we may use the internal-regret minimizing prediction algorithm of Section 1.2 of Chapter 3. For simplicity, we use exponential weighting. This definition, of course, requires the boundedness of the values of $\ell_{i,t}$. This may be guaranteed by the same assumption as in the analysis of the EG investment strategy, that is, by assuming that the returns $x_{i,t}$ all fall in the interval $[m, M]$ where $m < M$ are positive constants. Then the internal regret of the algorithm B1EXP may be bounded by the result of Theorem 3.1. An important additional property of the algorithm is that its worst-case logarithmic wealth ratio, with respect to the class of all constantly rebalanced portfolios, may be bounded similarly as that of the EG algorithm. These main properties are summarized in the following theorem.

THEOREM 7.1. *Assume that $m \leqslant x_{i,t} \leqslant M$ for all $1 \leqslant i \leqslant N$ and $1 \leqslant t \leqslant n$. Then the cumulative internal regret of the* B1EXP *strategy P over such bounded market evolutions is less than*

$$\widetilde{R}_n \leqslant \frac{\ln N(N-1)}{\eta} + \frac{n\eta}{8}\frac{M^2}{m^2} = \frac{M}{m}\sqrt{n\ln N} ,$$

where we set $\eta = 4(m/M)\sqrt{(\ln N)/n}$. In addition, if $\mathcal{Q}$ denotes the class of all constantly rebalanced portfolios, then the worst-case logarithmic wealth ratio (restricted to all those sequences of market vectors bounded between m and M) of P is less than

$$W_n(P,\mathcal{Q}) \leqslant N\frac{M}{m}\sqrt{n\ln N} .$$

PROOF. The bound for the internal regret $\widetilde{R}_n$ follows from the linear upper bound described above and Theorem 3.1.

To bound the worst-case logarithmic wealth ratio $W_n(P,\mathcal{Q})$, recall that by inequality (7.2), for any constantly rebalanced portfolio $\boldsymbol{B}$,

$$\begin{aligned}W_n(P,\mathcal{Q}) &\leqslant \sum_{j=1}^{N} B_j \sum_{i=1}^{N}\left(\sum_{t=1}^{n} P_{i,t}\left(\ell_{i,t}-\ell_{j,t}\right)\right) \\ &\leqslant N \max_{1\leqslant i,j\leqslant N}\sum_{t=1}^{n} P_{i,t}\left(\frac{x_{j,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t} - \frac{x_{i,t}}{\boldsymbol{P}_t\cdot\boldsymbol{x}_t}\right)\end{aligned}$$

which is not larger than N times the upper bound obtained on the cumulative internal regret $\widetilde{R}_n$ which completes the proof. □

REMARK 7.3. The computation of the investment strategy requires the inversion of an $N \times N$ matrix at each trading period (see Lemma 3.1). This is quite feasible even for large markets in which N may be as large as about 100.

REMARK 7.4. Recalling Section 1.2 of Chapter 3 we observe that the B1EXP strategy may be considered as an instance of the exponentially weighted average predictor, which uses the fictitious strategies $\boldsymbol{P}_t^{i\to j}$ as experts. Thus, instead of considering single stocks, as EG, B1EXP considers pairs of stocks and their relative behaviors. This may explain the greater stability observed on real data (see the Appendix).

REMARK 7.5. Just like in the case of the sequential prediction problem, exponential weighting may be replaced by others such as polynomial weighting. In that case Theorem 3.1 shows that the cumulative internal regret is bounded by $\frac{M}{m}\sqrt{n(p-1)}N^{2/p}$ which is approximately optimized by the choice $p = 4\ln N$. We call this investment strategy B1POL. Even though this strategy has comparable theoretical guarantees to those of B1EXP, our experiments show a clear superiority of the use of exponential weighting. This and other practical issues are discussed in the Appendix.

REMARK 7.6. Similarly to EG, the strategy B1EXP requires the knowledge of the time horizon n and the ratio M/m of the bounds assumed on the market. This first disadvantage may be avoided by either using the well-known "doubling trick" or considering a time-varying value of η and applying the second bound of Theorem 3.1. Both methods lead to internal regret and worst-case logarithmic wealth ratios bounded by quantities of the order of $\sqrt{n}$. To deal with the boundedness assumption however, we need more sophisticated techniques introduced in [**HeScSiWa98**], see Section 6.2.

4.2. Another strategy with small internal regret. In this section we introduce a new algorithm, called B2POL. We use polynomial weighting and assume bounded market evolutions. The Blackwell condition (3.3) is sufficient to ensure the property of small internal regret. It may be written as

$$\sum_{i\neq j} \Delta_{(i,j),t}\widetilde{r}_{(i,j),t} \leqslant 0 \,,$$

where

$$\Delta_{(i,j),t} = \frac{\left(\widetilde{R}_{(i,j),t-1}\right)_+^{p-1}}{\sum_{a\neq b}\left(\widetilde{R}_{(a,b),t-1}\right)_+^{p-1}} \,.$$

Note that the $\Delta_{(i,j),t}$'s are nonnegative and sum up to one. The concavity of the logarithm and the definition of $\widetilde{r}_{(i,j),t}$ lead to

$$\begin{aligned}\sum_{i\neq j} \Delta_{(i,j),t}\widetilde{r}_{(i,j),t} &= \left(\sum_{i\neq j} \Delta_{(i,j),t} \ln\left(\boldsymbol{P}_t^{i\to j}\cdot\boldsymbol{x}_t\right)\right) - \ln\left(\boldsymbol{P}_t\cdot\boldsymbol{x}_t\right)\\ &\leqslant \ln\left(\sum_{i\neq j}\Delta_{(i,j),t}\boldsymbol{P}_t^{i\to j}\cdot\boldsymbol{x}_t\right) - \ln\left(\boldsymbol{P}_t\cdot\boldsymbol{x}_t\right).\end{aligned}$$

It is now obvious that the Blackwell condition (3.3) is satisfied whenever

$$\boldsymbol{P}_t = \sum_{i\neq j}\Delta_{(i,j),t}\boldsymbol{P}_t^{i\to j}. \tag{7.4}$$

Lemma 3.1 shows that such a portfolio $\boldsymbol{P}_t$ indeed exists for all t. This defines a strategy which we call B2POL. The following theorem is an immediate consequence of Corollary 1 of [**CeLu03**] (see also Section 3.2 of Chapter 8).

THEOREM 7.2. *Assume that $m \leqslant x_{i,t} \leqslant M$ for all $1 \leqslant i \leqslant N$ and $1 \leqslant t \leqslant n$. Then the cumulative internal regret of the* B2POL *strategy P is bounded by*

$$\widetilde{R}_n \leqslant \left(\ln \frac{M}{m}\right) \sqrt{n(p-1)} N^{2/p} \, .$$

The above bound is approximately minimized for $p = 4 \ln N$. Note also that it only differs from the bound on the cumulative internal regret of the B1POL strategy by a constant factor which is smaller here ($\ln(M/m)$ instead of M/m).

5. Generalizations

5.1. Generalized buy-and-hold strategy. The GBH strategy performs buy-and-hold on the $N(N-1)$ fictitious modified strategies, using the conversion trick explained in Section 1.2 of Chapter 3 (and, in the particular case of $N = 2$ assets, it reduces to the simple buy-and-hold strategy–hence its name). The main property of this investment strategy is that its internal regret is bounded by a constant, as stated by the theorem below.

More precisely, the GBH strategy is defined such that at each round t, we have the fixed point equality

$$\boldsymbol{P}_t = \sum_{i \neq j} \frac{S_{t-1}^{i \to j}}{\sum_{k \neq l} S_{t-1}^{k \to l}} \boldsymbol{P}_t^{i \to j} \, , \tag{7.5}$$

where $S_t = \prod_{s=1}^t \boldsymbol{P}_s \cdot \boldsymbol{x}_s$ is the wealth achieved by the investment strategy we consider and $S_t^{i \to j} = \prod_{s=1}^t \boldsymbol{P}_s^{i \to j} \cdot \boldsymbol{x}_s$ is the fictitious wealth obtained by the $i \to j$ modified version of it. The existence and the practical computation of such a portfolio $\boldsymbol{P}_t$ are given by Lemma 3.1.

We note here the similarity of (7.5) to (7.4). In these two fixed point equalities, only the potential functions differ (see Section 2.3 in Chapter 2). (7.5) corresponds to an exponential potential, tuned with $\eta = 1$. The GBH strategy could thus have been called B2EXP, in reference to B2POL. We used this similarity to prove the following theorem in [**StLu03**]. But as indicated below, the latter may be proved in a much simpler way.

THEOREM 7.3. *For all n and all sequences of market vectors, the* GBH *investment strategy incurs a cumulative internal regret $\widetilde{R}_n \leqslant \ln N(N-1)$.*

PROOF. The proof is done by a simple telescoping argument:

$$S_n = \prod_{t=1}^n \mathbf{P}_t \cdot \mathbf{x}_t = \prod_{t=1}^n \sum_{i \neq j} \frac{S_{t-1}^{i \to j} \boldsymbol{P}_t^{i \to j} \cdot \boldsymbol{x}_t}{\sum_{k \neq l} S_{t-1}^{k \to l}} = \frac{\sum_{i \neq j} S_n^{i \to j}}{N(N-1)} \, .$$

□

The advantage of this algorithm is that its performance bounds do not depend on the market. We also note that the proof indicates that the internal regret of the GBH strategy is always non-negative, $\widetilde{R}_n \geqslant 0$.

REMARK 7.7. *(The worst-case logarithmic wealth ratio is not linked to the internal regret of an investment strategy.)* Unlike in the sequential prediction problem described in Section 1.2 of Chapter 3, a small internal regret in the problem of sequential portfolio selection does not necessarily imply a small worst-case logarithmic wealth ratio, not even with respect to the class of all

buy-and-hold strategies. This may be seen by considering the following numerical counterexample. Let the market be formed by three stocks and let it be cyclic such that at odd-indexed rounds the wealth ratios are respectively $1/2$, 1, 2 and at even ones they equal 2, 1.1, $1/2$. The accumulated wealth of the best stock increases exponentially fast whereas the one of the GBH strategy is bounded.

The reason is that the loss function ℓ' associated to this problem is no longer linear, and therefore, the argument of Equation (3.1) does not extend to it.

However, there is a simple modification of the GBH strategy leading to internal regret less than $2\ln N$ and external regret with respect to buy-and-hold strategies less than $2\ln N$. We call this modification the GBH2 algorithm.

Instead of (7.5), the GBH2 strategy is such that

$$\boldsymbol{P}_t = \frac{\sum_{1\leqslant k\leqslant N} S_{t-1}(k)\mathbf{e}_k + \sum_{i\neq j} S_{t-1}^{i\to j}\boldsymbol{P}_t^{i\to j}}{\sum_{1\leqslant k\leqslant N} S_{t-1}(k) + \sum_{i\neq j} S_{t-1}^{i\to j}} \,, \tag{7.6}$$

for every t, where $\mathbf{e}_k$ denotes the portfolio that invests all its wealth in the k-th stock. Now a telescoping argument similar to that of the proof of Theorem 7.3 shows that the final wealth equals

$$S_n = \frac{1}{N^2}\left(\sum_{1\leqslant k\leqslant N} S_n(k) + \sum_{i\neq j} S_n^{i\to j}\right) \,,$$

thus ensuring that both regrets are less than the claimed upper bound $2\ln N$. Lemma 3.1 shows that (7.6) can be satisfied and how the portfolios $\boldsymbol{P}_t$ are computed.

The next section is an extension of GBH and GBH2 strategies to a continuum of fictitious experts.

5.2. A generalized universal portfolio. Next we extend the notion of internal regret for investment strategies, similarly to what we did for internal regret in prediction with expert advice in Section 1.3 of Chapter 3. Recall that the definition of internal regret $\widetilde{R}_n$ considers the regret suffered by not moving one's capital from one stock to another. Moving the capital from one stock to another may be considered as a simple linear function from the probability simplex $\mathcal{X}$ to $\mathcal{X}$. A more exigent definition is obtained by considering all linear functions $g : \mathcal{X} \to \mathcal{X}$. Clearly, any such function may be written as $g(\boldsymbol{P}_t) = \boldsymbol{A}\boldsymbol{P}_t$ where $\boldsymbol{A}$ is a column-stochastic matrix. Denote the set of all column-stochastic matrices of order N by $\mathcal{A}$ and let the linear modifications $\boldsymbol{A}\boldsymbol{P}_t$ of the master strategy be denoted by $\boldsymbol{P}_t^{\boldsymbol{A}}$. The generalized internal regret (or swap regret for investment strategies, see Section 1.3 of Chapter 3) is defined as

$$\max_{\boldsymbol{A}\in\mathcal{A}} \ln \frac{S_n^{\boldsymbol{A}}}{S_n}$$

where $S_n^{\boldsymbol{A}} = \prod_{t=1}^n \sum_{i=1}^N P_{i,t}^{\boldsymbol{A}} x_{i,t}$.

Linear modifications were already considered (in finite number) by [**GrJa03**] in the case of sequential prediction. In that case, due to the linearity of the loss function $\ell(\boldsymbol{P}_t)$, it is not more difficult to have a low generalized internal regret than the usual internal regret, see Section 1.3 of Chapter 3. On the contrary here, due to the concavity of the logarithm, minimizing the generalized internal regret turns out to be a greater challenge. Since the algorithms B1EXP and B1POL are based on a linear upper bounding of the internal regret, it is easy to see that their generalized internal regret is bounded by N times the bounds derived for the internal regret in Sections 4.1, leading to upper bounds both of the order of $N\sqrt{n\ln N}$.

THEOREM 7.4. *The generalized internal regret of the* B1EXP *strategy Q over sequences of market vectors bounded between m and M is less than*

$$\max_{\boldsymbol{A}\in\mathcal{A}} \ln \frac{S_n^{\boldsymbol{A}}}{S_n} \leqslant \frac{M}{m}\, N\sqrt{n \ln N}\,,$$

where the strategy is tuned with $\eta = 4(m/M)\sqrt{(\ln N)/n}$.

The main result of this section is that there exist investment strategies that achieve a much smaller generalized internal regret. The proof below is inspired by Theorem 7.3 and uses some techniques introduced by [**BlKa99**]. The investment strategy presented above may be seen as a modification of Cover's universal portfolio [**Cov91**] through a conversion trick to deal with generalized internal regret of the same flavor as the one explained in Section 1.2 of Chapter 3.

THEOREM 7.5. *There exists an investment strategy P such that for all sequences of market vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ in $\mathbb{R}^N_+$,*

$$\max_{\boldsymbol{A}\in\mathcal{A}} \ln \frac{S_n^{\boldsymbol{A}}}{S_n} \leqslant N(N-1)\ln(n+1) + 1\,.$$

REMARK 7.8. The algorithm given in the proof has a computational complexity exponential in the number of stocks (at least in its straightforward implementation). However, it provides a theoretical bound which is likely to be of the best achievable order. The techniques of Kalai and Vempala [**KaVe03a**] may be used to implement it more efficiently.

The algorithm could also be easily modified, using the techniques of Section 5.1, to be competitive with respect to the best constantly rebalanced portfolio as well as to suffer a low generalized internal regret, with associated performance bounds for both of the order $N^2 \ln n$.

PROOF. Denote a column-stochastic matrix $\boldsymbol{A}$ by $[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, where the $\boldsymbol{a}_j$'s are the columns of $\boldsymbol{A}$. Let μ be the uniform measure over the simplex and let ν be the measure over $\mathcal{A}$ given by the product of N independent instances of μ:

$$\nu(\boldsymbol{A}) = \prod_{j=1}^{N} \mu(\boldsymbol{a}_j)\,.$$

If the investment strategy, at each time instant t, satisfied the equality

$$\boldsymbol{P}_t = \frac{\int_{\boldsymbol{A}\in\mathcal{A}} S_{t-1}^{\boldsymbol{A}} \boldsymbol{P}_t^{\boldsymbol{A}}\, \mathrm{d}\nu(\boldsymbol{A})}{\int_{\boldsymbol{A}\in\mathcal{A}} S_{t-1}^{\boldsymbol{A}}\, \mathrm{d}\nu(\boldsymbol{A})}, \tag{7.7}$$

then the final wealth would be given by an average over all modified strategies, that is,

$$S_n = \int_{\boldsymbol{A}\in\mathcal{A}} S_n^{\boldsymbol{A}}\, \mathrm{d}\nu(\boldsymbol{A}). \tag{7.8}$$

Fix a matrix $\boldsymbol{A}$ and consider the set $\chi_{\alpha,\boldsymbol{A}}$ of column-stochastic matrices of the form $(1-\alpha)\boldsymbol{A} + \alpha\mathbf{z}$, $\mathbf{z} \in \mathcal{A}$. Similarly, denote by $\chi_{\alpha,\boldsymbol{a}_j}$ the set of probability vectors of the form $(1-\alpha)\boldsymbol{a}_j + \alpha\mathbf{z}_j$, $\mathbf{z}_j \in \mathcal{X}$. It is easy to see that (with a slight abuse of notation)

$$\chi_{\alpha,\boldsymbol{A}} = \prod_{j=1}^{N} \chi_{\alpha,\boldsymbol{a}_j}\,. \tag{7.9}$$

Any element $\boldsymbol{A}'$ of $\chi_{\alpha,\boldsymbol{A}}$ may be seen to satisfy (component-wise)

$$P_t^{\boldsymbol{A}'} \geqslant (1-\alpha) P_t^{\boldsymbol{A}}\,,$$

for all t and therefore

$$S_n^{\boldsymbol{A}'} \geqslant (1-\alpha)^n S_n^{\boldsymbol{A}}\,.$$

Finally, using equality (7.9), we have

$$\nu\left(\chi_{\alpha,\boldsymbol{A}}\right) = \prod_{j=1}^{N} \mu\left(\chi_{\alpha,\boldsymbol{a}_j}\right) = \left(\alpha^{N-1}\right)^N ,$$

implying

$$\int_{\mathbf{A}' \in \chi_{\alpha,\boldsymbol{A}}} S_n^{\mathbf{A}'} \, \mathrm{d}\nu(\mathbf{A}') \geqslant (1-\alpha)^n \, \alpha^{N(N-1)} S_n^{\boldsymbol{A}} .$$

Taking $\alpha = 1/(n+1)$, recalling that

$$(1-\alpha)^n \, \alpha^{N(N-1)} \geqslant \frac{e^{-1}}{(n+1)^{N(N-1)}} ,$$

and combining this with 7.8, we obtain the theorem.

Thus, it suffices to see that one may satisfy the set of linear equations (7.7). We denote an element $\boldsymbol{A} \in \mathcal{A}$ by $\boldsymbol{A} = [A_{(i,j)}]$. Writing only the equality for the ith components of both sides,

$$\begin{aligned} &\left(\int_{\boldsymbol{A} \in \mathcal{A}} S_{t-1}^{\boldsymbol{A}} \, \mathrm{d}\nu(\mathbf{A}) \right) P_{i,t} \\ &\quad = \int_{\boldsymbol{A} \in \mathcal{A}} S_{t-1}^{\boldsymbol{A}} \left(\sum_{k=1}^{N} A_{(i,k)} P_{k,t} \right) \mathrm{d}\nu(\mathbf{A}) , \end{aligned}$$

we see that $\boldsymbol{P}_t$ has to be an element of the kernel of the matrix T defined by

- if $i \neq k$, $T_{i,k} = w_{i,k}$,
- $T_{i,i} = -\sum_{j \neq i,\ 1 \leqslant j \leqslant N} w_{j,i}$,

where

$$w_{i,k} = \int_{\boldsymbol{A} \in \mathcal{A}} S_{t-1}^{\boldsymbol{A}} A_{(i,k)} \, \mathrm{d}\nu(\mathbf{A}).$$

The same argument as in the proof of Lemma 3.1 shows that such a vector exists (and the computability of the latter depends on how easy it is to compute the elements of the matrix T). □

6. Universal versions of EG and B1EXP

The EG and B1EXP strategies rely on the prior knowledge of the total number n of trading periods, and also on the bounds m and M on the market values. Since these values may not be known in advance in practice or since the market evolutions may be unbounded, appropriate modifications are required. The purpose of this section is to introduce "universal" variants of these two strategies which do not assume the prior knowledge of any of these parameters. The proposed adaptive strategies are based on a combination of Lemmas 4.3 and 4.4 with an argument of Helmbold, Schapire, Singer, and Warmuth [**HeScSiWa98**].

6.1. A universal version for the EG strategy. Observe first that all regrets are defined in terms of ratios, so that the investment strategy may always renormalize the past market vectors $\boldsymbol{x}_t$ so that $\max_{i \leqslant N} x_{i,t} = 1$. Our "universal" version of the EG strategy is then defined in Figure 1, and is called EG-UNIV. It is competitive with respect to the class of all constantly rebalanced portfolios, as revealed by the following result.

THEOREM 7.6. *Consider a market with $N \geqslant 2$ assets. If $\mathcal{Q}$ denotes the class of all constantly rebalanced portfolios, then the worst-case logarithmic wealth ratio of* EG-UNIV *strategy P (for* all *possible behaviors of the market) is bounded by*

$$W_n(P, \mathcal{Q}) \leqslant 10 N n^{2/3} .$$

Algorithm EG-UNIV

Initialization: $\boldsymbol{P}_1 = (1/N, \ldots, 1/N)$, and $\widetilde{L}_{i,0} = 0$ for all $i = 1, \ldots, N$.

For each round $t = 1, 2, \ldots$,

(1) invest in the stock market with portfolio $\boldsymbol{P}_t$, and get the market vector $\boldsymbol{x}_t$ of day t; renormalize $\boldsymbol{x}_t$ so that $\max_{i \leqslant N} x_{i,t} = 1$;

(2) let $\alpha_t = t^{-1/3}/2$ and $\widetilde{\boldsymbol{x}}_t = (1 - \alpha_t/N)\boldsymbol{x}_t + (\alpha_t/N)\mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)$;

(3) for $i = 1, \ldots, N$, let

$$\widetilde{\ell}_{i,t} = -\frac{\widetilde{x}_{i,t}}{\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t} \, ,$$

and $\widetilde{L}_{i,t} = \widetilde{L}_{i,t-1} + \widetilde{\ell}_{i,t}$;

(4) let $\eta_t = t^{-2/3}/4$, and define the portfolio $\widetilde{\boldsymbol{P}}_{t+1}$ by its components, for $i = 1, \ldots, N$,

$$\widetilde{P}_{i,t+1} = -\frac{e^{-\eta_t \widetilde{L}_{i,t}}}{\sum_{j=1}^{N} e^{-\eta_t \widetilde{L}_{j,t}}} \, ;$$

(5) let the next round portfolio be

$$\boldsymbol{P}_{t+1} = (1 - \alpha_t)\widetilde{\boldsymbol{P}}_{t+1} + (\alpha_t/N)\mathbf{1} \, .$$

FIGURE 1. A universal version of the EG algorithm.

REMARK 7.9. Helmbold, Schapire, Singer, and Warmuth [**HeScSiWa98**] were the first to define a universal version of the EG strategy, based on a "doubling trick" which requires to periodically reset the algorithm by forgetting everything learnt up to that point. We feel that modifying the parameter η "smoothly" as in the version introduced above is more natural. Moreover, the bound obtained for the new algorithm EG-UNIV is of the order of $n^{2/3}$, whereas the one for the universal investment strategy of [**HeScSiWa98**] is of the worse order of $n^{3/4}$.

The proof below is a straightforward extension of the methodology originally proposed in [**HeScSiWa98**], and we first recall a lemma proved therein.

LEMMA 7.1. *Whenever* $\alpha_t \in [0, 1/2]$,

$$\ln (\boldsymbol{P}_t \cdot \boldsymbol{x}_t) \geqslant \ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) - 2\alpha_t \, .$$

PROOF (OF THEOREM 7.6). We decompose the quantity of interest into three sums,

$$\begin{aligned} \sum_{t=1}^{n} \left(\ln (\boldsymbol{B} \cdot \boldsymbol{x}_t) - \ln (\boldsymbol{P}_t \cdot \boldsymbol{x}_t) \right) \leqslant & \sum_{t=1}^{n} \left(\ln (\boldsymbol{B} \cdot \boldsymbol{x}_t) - \ln (\boldsymbol{B} \cdot \widetilde{\boldsymbol{x}}_t) \right) \\ & + \sum_{t=1}^{n} \left(\ln (\boldsymbol{B} \cdot \widetilde{\boldsymbol{x}}_t) - \ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) \right) \\ & + \sum_{t=1}^{n} \left(\ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) - \ln (\boldsymbol{P}_t \cdot \boldsymbol{x}_t) \right) \, . \end{aligned}$$

The first sum in the right-hand side is non-positive, as the $\boldsymbol{x}_t$ are renormalized such that all their components are less than 1, and thus $\boldsymbol{x}_t \leqslant \widetilde{\boldsymbol{x}}_t$ pointwise. The third sum is less than $2(\alpha_1 + \ldots + \alpha_n)$ by Lemma 7.1. We simply have to deal with the second sum.

We note that the portfolios $\widetilde{\boldsymbol{P}}_t$ correspond to exponential reweighting over the losses $\widetilde{\ell}_{i,t}$. The analysis of Section 2 leads to the linear upper bound, for all constantly rebalanced portfolio $\boldsymbol{B}$,

and all market sequences,

$$\sum_{t=1}^{n} \left(\ln \left(\boldsymbol{B} \cdot \widetilde{\boldsymbol{x}}_t \right) - \ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) \right) \leqslant \sum_{j=1}^{N} B_j \left(\sum_{t=1}^{n} \sum_{i=1}^{N} \widetilde{P}_{i,t} \widetilde{\ell}_{i,t} - \widetilde{\ell}_{j,t} \right) .$$

As the η_t are non-increasing, Lemma 4.3, combined with the definition of the $\widetilde{\boldsymbol{P}}_t$, then guarantees that, with the notation of this lemma,

$$\sum_{t=1}^{n} \left(\ln \left(\boldsymbol{B} \cdot \widetilde{\boldsymbol{x}}_t \right) - \ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) \right) \leqslant \left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1} \right) \ln N + \sum_{t=1}^{n} \Phi(\widetilde{\boldsymbol{P}}_t, \, \eta_t, \, -\widetilde{\ell}_t) .$$

The definition of $\widetilde{\boldsymbol{x}}_t$ ensures that $\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \geqslant \alpha_t / N$, and thus that all $\widetilde{\ell}_{j,t}$ lie in $[-N/\alpha_t, 0]$. The choice of η_t, combined with Lemma 4.4, leads to $N\eta_t/\alpha_t \leqslant 1$, and

$$\sum_{t=1}^{n} \left(\ln \left(\boldsymbol{B} \cdot \widetilde{\boldsymbol{x}}_t \right) - \ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) \right) \leqslant \left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1} \right) \ln N + (e-2) \sum_{t=1}^{n} \eta_t \sum_{i=1}^{N} \widetilde{P}_{i,t} \frac{\widetilde{x}_{i,t}^2}{\left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right)^2} .$$

The renormalization of the $\boldsymbol{x}_t$ is such that all $x_{i,t} \leqslant 1$, and this is thus also the case for the $\widetilde{x}_{i,t}$. Using in addition that $\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \geqslant \alpha_t / N$, we get

$$\sum_{t=1}^{n} \left(\ln \left(\boldsymbol{B} \cdot \widetilde{\boldsymbol{x}}_t \right) - \ln \left(\widetilde{\boldsymbol{P}}_t \cdot \widetilde{\boldsymbol{x}}_t \right) \right) \leqslant \left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1} \right) \ln N + (e-2) N \sum_{t=1}^{n} \frac{\eta_t}{\alpha_t} .$$

In conclusion, we have shown that

$$\sum_{t=1}^{n} \left(\ln \left(\boldsymbol{B} \cdot \boldsymbol{x}_t \right) - \ln \left(\boldsymbol{P}_t \cdot \boldsymbol{x}_t \right) \right) \leqslant \left(\frac{2}{\eta_{n+1}} - \frac{1}{\eta_1} \right) \ln N + (e-2) N \sum_{t=1}^{n} \frac{\eta_t}{\alpha_t} + 2 \sum_{t=1}^{n} \alpha_t ,$$

provided that for all t, $\alpha_t \leqslant 1/2$ and $N\eta_t/\alpha_t \leqslant 1$. Substituting the proposed values for η_t and α_t and performing simple algebra conclude the proof. (We note that these values do not optimize the order of magnitude in terms of N: the choices $\eta_t \sim N^{-1/3} (\ln N)^{2/3} t^{-2/3}$ and $\alpha_t \sim (N \ln N)^{1/3} t^{-1/3}$ would lead to a $(1 + o(1))(N \ln N)^{1/3} n^{2/3}$ upper bound.) □

6.2. A universal version for the B1EXP **strategy.** The universal variant of the B1EXP strategy is designed by applying the conversion trick described in Remark 3.1 to the EG-UNIV strategy introduced above. Here however, as EG-UNIV does not simply minimize a linearized upper bound over the regrets, we need to apply the conversion trick to a set of N^2 fictitious assets, $N(N-1)$ given by the $i \to j$ modified strategies of the master strategy, and the N other given by the single stocks. (See also how we defined the GBH2 strategy, due to the lack of linearity noted in Remark 7.7, we had to use these N^2 fictitious assets as well.)

More precisely, consider the sequence of market vectors $\boldsymbol{y}_t$, $t = 1, 2, \ldots$, with N^2 components given by

$$\boldsymbol{y}_t = \left(\boldsymbol{x}_t, \, \left(\boldsymbol{P}_t^{i \to j} \cdot \boldsymbol{x}_t \right)_{i \neq j} \right) ,$$

and denote by $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots$ the sequence of portfolios associated to it by EG-UNIV. Define $\boldsymbol{P}_t$ as the portfolio (over the N initial assets) such that the fixed-point equality $\boldsymbol{Q}_t \cdot \boldsymbol{y}_t = \boldsymbol{P}_t \cdot \boldsymbol{x}_t$ is satisfied. Existence and practical computation of $\boldsymbol{P}_t$ are indicated by Lemma 3.1.

Theorem 7.6 guarantees that for all fixed probability distributions $\boldsymbol{B}'$ from the simplex of order N^2, and for all n, we have

$$\sup_{\boldsymbol{x}_1^n \in (\mathbb{R}_+^N)^n} \sum_{t=1}^{n} \ln \left(\boldsymbol{B}' \cdot \boldsymbol{y}_t \right) - \ln \left(\boldsymbol{P}_t \cdot \boldsymbol{x}_t \right) \leqslant 10 N^2 n^{2/3} . \tag{7.10}$$

Choosing a probability distribution $\boldsymbol{B}'$ concentrated on the first N components, or putting probability mass 1 over the k-th component, $k > N$, we get the following result.

THEOREM 7.7. *The cumulative internal regret of the* B1EXP-UNIV *strategy P is bounded by*

$$\widetilde{R}_n \leqslant 10N^2 n^{2/3} .$$

In addition, if $\mathcal{Q}$ denotes the class of all constantly rebalanced portfolios, then the worst-case logarithmic wealth ratio of P is bounded by

$$W_n(P, \mathcal{Q}) \leqslant 10N^2 n^{2/3} .$$

REMARK 7.10. As indicated by (7.10), the B1EXP-UNIV strategy minimizes its internal regret with respect to a class of deviations larger than simply the ones of the form $i \to j$. This class is the convex hull formed by the Dirac masses $\mathbf{e}_k$ introduced in Section 5.1, and the applications which associate to a portfolio $\boldsymbol{P}$ its modification $\boldsymbol{P}^{i\to j}$. This yields a class which, on the one hand, contains the simple $N(N-1)$ modifications, and on the other hand, is strictly contained in the class of all linear departures introduced in Section 5.2. (This is the class considered in Section 7 of [**StLu03**].)

To get a version of the B1EXP strategy minimizing its generalized internal regret, we would need to apply the conversion trick to EG-UNIV run on a set of N^N fictitious strategies, corresponding to the N^N extremal points of the convex hull of all linear departures from the simplex into itself. These extremal points are given by the column-stochastic matrices with 0 and 1 only, and generate all linear departures, according to the Krein-Millman (see, e.g., Berger [**Ber90**]) theorem. Unfortunately, this version suited for the minimization of the generalized internal regret has a computational complexity of the order of N^N, that is, more than exponential in the number of stocks N. In comparison, the complexity of B1EXP-UNIV is of the order of N^2.

7. On-line investment with transaction costs

We indicate how some of the investment strategies introduced above, namely those defined only by means of fixed-point equalities, may be modified to be competitive in presence of transaction costs. We recall below the model considered by Blum and Kalai [**BlKa99**] in which, without loss of generality, transaction fees are paid at purchase only. The model is best described by a function $\mathrm{TC}(\boldsymbol{P}, \boldsymbol{Q})$, which indicates the cost of rebalancing the investor's wealth distributed according to $\boldsymbol{P}$ to $\boldsymbol{Q}$. To perform such a rebalancing, the investor first has to sell a certain amount of some assets to be able to pay for the transaction fees when purchasing the needed quantities of the other assets. To buy a quantity w of a given asset, he has to pay $(1+c)w$, where c is called the *commission rate*. TC indicates that 1 euro distributed according to $\boldsymbol{P}$ leads to $\alpha = \mathrm{TC}(\boldsymbol{P}, \boldsymbol{Q})$ euros distributed according to $\boldsymbol{Q}$. (That is, TC is a multiplicative factor.) The precise way of rebalancing optimally, as well as an implicit formula for TC, is indicated in [**BlKa99**]. We denote by

$$\boldsymbol{P}(\boldsymbol{x}) = \left(\frac{P_{k,t} x_{k,t}}{\boldsymbol{P}_t \cdot \boldsymbol{x}_t} \right)_{k=1,\ldots,N}$$

the distribution of the investor's wealth, when the latter was originally distributed according to $\boldsymbol{P}$ and the market evolved according to the wealth ratio $\boldsymbol{x}$. In particular, the investment strategy has to rebalance at the beginning of each day $t+1$ from $\boldsymbol{P}_t(\boldsymbol{x}_t)$ to $\boldsymbol{P}_{t+1}$, and has to pay a fraction $\mathrm{TC}(\boldsymbol{P}_t(\boldsymbol{x}_t), \boldsymbol{P}_{t+1})$ of the wealth it owned at the end of day t to do so.

More precisely, Blum and Kalai [**BlKa99**] show that $\mathrm{TC}(\boldsymbol{P},\boldsymbol{Q})$ is the number α satisfying the equation

$$\alpha = 1 - c \sum_{j=1,\ldots,N} (\alpha Q_j > P_j)_+ \ ,$$

and list some basic properties of TC. One of them is that joint rebalancing is more efficient than the weighted combination of the separate rebalancings: investors may occasionally save in commission cost by trading among themselves without commission instead of trading in the stock exchange. Formally, this means that for any convex combination $\alpha_1,\ldots,\alpha_m$, any m portfolio couples $(\boldsymbol{P}_r,\boldsymbol{Q}_r)$, $r=1,\ldots,m$, and any market vector $\boldsymbol{x}$, we have

$$\sum_{r=1}^{m} \alpha_r \left(\boldsymbol{P}_r\cdot\boldsymbol{x}\right) \mathrm{TC}\left(\boldsymbol{P}_r(\boldsymbol{x}),\boldsymbol{Q}_r\right) \leqslant \left(\left(\sum_{r=1}^{m}\alpha_r\boldsymbol{P}_r\right)\cdot\boldsymbol{x}\right) \mathrm{TC}\left(\sum_{r=1}^{m}\alpha_r\boldsymbol{P}_r(\boldsymbol{x}),\boldsymbol{Q}'\right) \ , \tag{7.11}$$

where

$$\boldsymbol{Q}' = \frac{\sum_{r=1}^{m}\alpha_r\left(\boldsymbol{P}_r\cdot\boldsymbol{x}\right)\mathrm{TC}\left(\boldsymbol{P}_r(\boldsymbol{x}),\boldsymbol{Q}_r\right)\boldsymbol{Q}_r}{\sum_{r=1}^{m}\alpha_r\left(\boldsymbol{P}_r\cdot\boldsymbol{x}\right)\mathrm{TC}\left(\boldsymbol{P}_r(\boldsymbol{x}),\boldsymbol{Q}_r\right)}$$

is the final distribution of the separate rebalancings. (A related, though different, property is that TC, as a function of the couple $(\boldsymbol{P},\boldsymbol{Q})$, is concave. This may seen by direct computation with the implicit definition of TC.) Furthermore, we note here that the implicit definition of TC shows that, for a fixed $\boldsymbol{P}$, the map $\boldsymbol{Q}\mapsto\mathrm{TC}(\boldsymbol{P},\boldsymbol{Q})$ is continuous.

7.1. The extension of the GBH strategy to a market with commission rates. We now describe the variant GBH_c of the GBH strategy suited for a market with a commission rate c. The idea is to divide (fictitiously) the capital at the beginning of day t among the $i\to j$ modified strategies and to force them to rebalance (separately) from $\boldsymbol{P}_t^{i\to j}(\boldsymbol{x}_t)$ to $\boldsymbol{P}_{t+1}^{i\to j}$ at the beginning of day $t+1$. The trick is to obtain the wealth allocation $\boldsymbol{P}_{t+1}$ once each fictitious strategy has rebalanced to $\boldsymbol{P}_{t+1}^{i\to j}$.

Formally, denote by $S_{n,c}$ (respectively, $S_{n,c}^{i\to j}$) the wealth obtained by the GBH_c strategy (respectively, by the fictitious $i\to j$ modified strategy) at the end of day n, after rebalancing to the distribution prescribed for day $n+1$,

$$\begin{aligned} S_{n,c} &= \prod_{t=1}^{n}\left(\boldsymbol{P}_t\cdot\boldsymbol{x}_t\right)\ \mathrm{TC}\left(\boldsymbol{P}_t(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}\right) \ , \\ S_{n,c}^{i\to j} &= \prod_{t=1}^{n}\left(\boldsymbol{P}_t^{i\to j}\cdot\boldsymbol{x}_t\right)\ \mathrm{TC}\left(\boldsymbol{P}_t^{i\to j}(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}^{i\to j}\right) \ . \end{aligned}$$

Now, we choose $\boldsymbol{P}_1$ as the uniform wealth allocation, and for $t=1,2,\ldots$, $\boldsymbol{P}_{t+1}$ is chosen such that the fixed point equality

$$\boldsymbol{P}_{t+1} = \frac{\sum_{i\neq j} S_{t-1,c}^{i\to j}\left(\boldsymbol{P}_t^{i\to j}\ \cdot\ \boldsymbol{x}_t\right)\ \mathrm{TC}\left(\boldsymbol{P}_t^{i\to j}(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}^{i\to j}\right)\boldsymbol{P}_{t+1}^{i\to j}}{\sum_{i\neq j} S_{t-1,c}^{i\to j}\left(\boldsymbol{P}_t^{i\to j}\ \cdot\ \boldsymbol{x}_t\right)\ \mathrm{TC}\left(\boldsymbol{P}_t^{i\to j}(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}^{i\to j}\right)} = \frac{\sum_{i\neq j} S_{t,c}^{i\to j}\boldsymbol{P}_{t+1}^{i\to j}}{\sum_{i\neq j} S_{t,c}^{i\to j}} \tag{7.12}$$

is satisfied. Such a portfolio $\boldsymbol{P}_{t+1}$ indeed exists by Brouwer's theorem, as the middle term of (7.12) is a continuous function of $\boldsymbol{P}_{t+1}$, and thus the GBH_c strategy is well-defined.

THEOREM 7.8. *The GBH_c investment strategy incurs a cumulative internal regret $\widetilde{R}_n \leqslant \ln N(N-1)$ for all n.*

PROOF. The defining expression (7.12) and property (7.11) directly show that, for $t \geqslant 1$,

$$(\boldsymbol{P}_t \cdot \boldsymbol{x}_t) \,\mathrm{TC}\, (\boldsymbol{P}_t(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}) \geqslant \frac{\sum_{i \neq j} S_{t-1,c}^{i \to j} \left(\boldsymbol{P}_t^{i \to j} \cdot \boldsymbol{x}_t\right) \mathrm{TC} \left(\boldsymbol{P}_t^{i \to j}(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}^{i \to j}\right)}{\sum_{i \neq j} S_{t-1,c}^{i \to j}} \,.$$

A telescoping argument finally yields

$$S_{n,c} \geqslant \frac{1}{N(N-1)} \sum_{i \neq j} S_{n,c}^{i \to j} \,.$$

□

The extension GBH2$_c$ of GBH2 to a market with transaction costs is defined is a similar way, and still ensures that both the internal regret and the external regret with respect to the class of all buy-and-hold strategies are less than $2 \ln N$. We omit the details and concentrate rather on the extension of the generalized universal portfolio described in Section 5.2. This is done by combining the argument of the present section with those of Section 5.2.

7.2. A modification of the generalized universal portfolio. We extend the notation of Section 5.2. For any column-stochastic matrix $\boldsymbol{A}$, we denote by

$$S_{n,c}^{\boldsymbol{A}} = \prod_{t=1}^{n} \left(\boldsymbol{P}_t^{\boldsymbol{A}} \cdot \boldsymbol{x}_t\right) \,\mathrm{TC}\, \left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}^{\boldsymbol{A}}\right)$$

the wealth achieved by consistent modifications of the master strategy according to $\boldsymbol{A}$ in a market with a transaction commission c.

We choose $\boldsymbol{P}_1$ as the uniform wealth allocation, and for $t = 1, 2, \ldots$, $\boldsymbol{P}_{t+1}$ is chosen such that the fixed point equality

(7.13)
$$\boldsymbol{P}_{t+1} = \frac{\int_{\mathcal{A}} S_{t-1,c}^{\boldsymbol{A}} \left(\boldsymbol{P}_t^{\boldsymbol{A}} \cdot \boldsymbol{x}_t\right) \,\mathrm{TC}\, \left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}^{\boldsymbol{A}}\right) \boldsymbol{P}_{t+1}^{\boldsymbol{A}} \,\mathrm{d}\nu(\boldsymbol{A})}{\int_{\mathcal{A}} S_{t-1,c}^{\boldsymbol{A}} \left(\boldsymbol{P}_t^{\boldsymbol{A}} \cdot \boldsymbol{x}_t\right) \,\mathrm{TC}\, \left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}^{\boldsymbol{A}}\right) \,\mathrm{d}\nu(\boldsymbol{A})} = \frac{\int_{\mathcal{A}} S_{t,c}^{\boldsymbol{A}} \boldsymbol{P}_{t+1}^{\boldsymbol{A}} \,\mathrm{d}\nu(\boldsymbol{A})}{\int_{\mathcal{A}} S_{t,c}^{\boldsymbol{A}} \,\mathrm{d}\nu(\boldsymbol{A})}$$

is satisfied. This defining expression is the exact counterpart of (7.12) for a continuum of deviations, and is valid thanks to Brouwer's theorem.

THEOREM 7.9. *The investment strategy defined above ensures that in a market with a commission rate c, for all n and all market sequences,*

$$\max_{\boldsymbol{A} \in \mathcal{A}} \ln \frac{S_{n,c}^{\boldsymbol{A}}}{S_{n,c}} \leqslant N(N-1) \ln\left((1+c)n + 1\right) + 1 \,.$$

Note that the orders of magnitude of the above upper bound in terms of c, n, and N are the same as those for the worst-case logarithmic wealth ratio of Blum and Kalai's [**BlKa99**] generalization of Cover's [**Cov91**] universal portfolio.

PROOF. Property (7.11) extends to a continuous weighted average, and thus, similarly to the analysis of GBH$_c$, we get that

$$S_{n,c} \geqslant \int_{\mathcal{A}} S_{n,c}^{\boldsymbol{A}} \,\mathrm{d}\nu(\boldsymbol{A}) \,.$$

We conclude the proof by the same kind of argument as in the proof of Theorem 7.5, and use the notation introduced there. We fix a matrix $\boldsymbol{A}$ and a number $\alpha \in]0,1[$, and consider their associated set $\chi_{\alpha,\boldsymbol{A}}$. The elements $\boldsymbol{A}'$ of $\chi_{\alpha,\boldsymbol{A}}$ are such that there exists $\boldsymbol{z} \in \mathcal{A}$ with $\boldsymbol{P}_t^{\boldsymbol{A}'} = (1-\alpha)\boldsymbol{P}_t^{\boldsymbol{A}} + \alpha \boldsymbol{P}_t^{\boldsymbol{z}}$, and

$$\left(\boldsymbol{P}_t^{\boldsymbol{A}'} \cdot \boldsymbol{x}_t\right) \mathrm{TC} \left(\boldsymbol{P}_t^{\boldsymbol{A}'}(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}^{\boldsymbol{A}'}\right) \geqslant (1-\alpha)^{1+c} \left(\boldsymbol{P}_t^{\boldsymbol{A}} \cdot \boldsymbol{x}_t\right) \mathrm{TC} \left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t), \boldsymbol{P}_{t+1}^{\boldsymbol{A}}\right) \,. \tag{7.14}$$

To see this, we lower bound the left-hand side by

$$(1-\alpha)\left(\boldsymbol{P}_t^{\boldsymbol{A}}\cdot\boldsymbol{x}_t\right)\,\mathrm{TC}\left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}^{\boldsymbol{A}'}\right)\ ,$$

by ignoring the fraction α of the wealth not distributed according to $\boldsymbol{P}_t^{\boldsymbol{A}}$ in $\boldsymbol{P}_t^{\boldsymbol{A}'}$. Now,

$$\mathrm{TC}\left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}^{\boldsymbol{A}'}\right)\geqslant\mathrm{TC}\left(\boldsymbol{P}_t^{\boldsymbol{A}}(\boldsymbol{x}_t),\boldsymbol{P}_{t+1}^{\boldsymbol{A}}\right)\,\mathrm{TC}\left(\boldsymbol{P}_{t+1}^{\boldsymbol{A}},\boldsymbol{P}_{t+1}^{\boldsymbol{A}'}\right)\ ,$$

and since rebalancing from $\boldsymbol{P}_{t+1}^{\boldsymbol{A}}$ to $\boldsymbol{P}_{t+1}^{\boldsymbol{A}'}$ involves moving at most a fraction α of the wealth,

$$\mathrm{TC}\left(\boldsymbol{P}_{t+1}^{\boldsymbol{A}},\boldsymbol{P}_{t+1}^{\boldsymbol{A}'}\right)\geqslant 1-\alpha+\frac{\alpha}{1+c}=1-\alpha\frac{c}{1+c}\geqslant 1-\alpha c\geqslant(1-\alpha)^c\ ,$$

where the last inequality is recalled in [**BlKa99**].

The proof is concluded by multiplying (7.14) over $t=1,\ldots,n$ and using the same argument as in the end of the proof of Theorem 7.5, with n replaced by $(1+c)n$. □

OPEN QUESTION 7.1. We extended above all the algorithms which do not use the linearized upper bounds and only rely on fixed point theorems. The obtained generalizations are of theoretical interest, as the different calls to Brouwer's theorem do not provide any practical method to implement the investment strategies. We do not mention any extension for EG nor B1EXP. This is because even if TC is (jointly) concave, the function ψ that maps $(\boldsymbol{P},\boldsymbol{Q},\boldsymbol{x})$ to

$$\psi(\boldsymbol{P},\boldsymbol{Q},\boldsymbol{x})=-\ln\left((\boldsymbol{P}\cdot\boldsymbol{x})\,\mathrm{TC}\,(\boldsymbol{P}(\boldsymbol{x}),\boldsymbol{Q})\right)$$

is not necessarily convex in $(\boldsymbol{P},\boldsymbol{Q})$ for a fixed $\boldsymbol{x}$ (essentially, because $\boldsymbol{P}\mapsto\boldsymbol{P}(\boldsymbol{x})$ is not linear). The proofs above show that in presence of transaction costs, ψ is the loss function of interest, and we may only linearize convex losses. (See the appendix of Chapter 8.)

In particular, even finding an equivalent of EG, easily computable and competitive in presence of transaction costs, is still an open question. Blum and Kalai's [**BlKa99**] extension of Cover's [**Cov91**] universal portfolio has indeed the same computational drawbacks as the latter.

Appendix: Experimental results

In this appendix we present an experimental comparison of the performance of the new algorithms with existing ones. In the experiments we used a data set of daily wealth ratios of 36 stocks of the New York Stock Exchange that has been used by various authors including [**Cov91**], [**CoOr96**], [**HeScSiWa98**], [**BlKa99**], [**Sin97**], and [**BoElGo00**]. The data set is formed by 5651 daily prices covering the 22-year period from July 3rd, 1962, to December 31st, 1984. The behaviors of some selected stocks is plotted in Figure 2. We also considered monthly wealth ratios

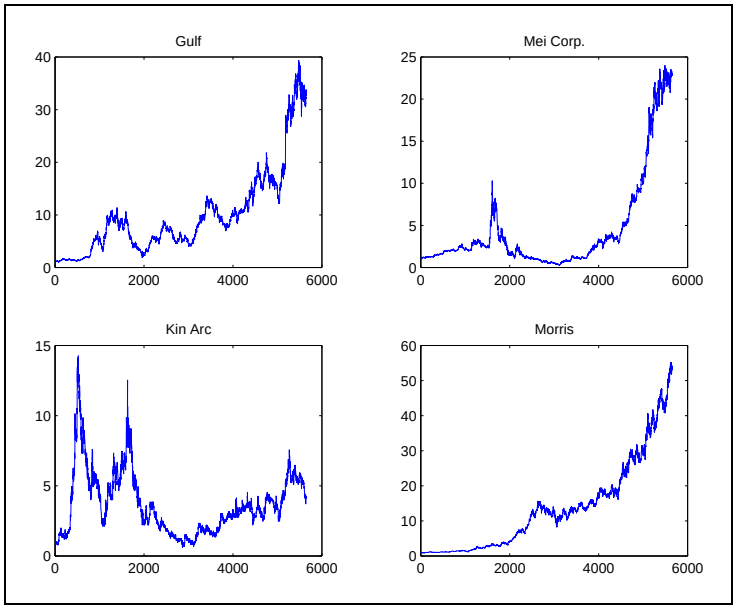

FIGURE 2. Evolution of some selected stocks over the 22-year period of study.

(taking 20 trading days for a month).

Of course, as all these stocks survived during this period, they performed well. There is therefore a "survivor bias", which implies that any investment strategy using these stocks will do fine. However, we compare below our investment strategies to other ones, and both the new and the existing ones benefit from the bias.

We begin this appendix with an overview of the strategies introduced in this chapter.

Overview of the investment strategies. We give two overviews of the methodology we used to derive our investment algorithms.

A strategy is given by the choice of a measure of the regret $\mathbf{r}_t$ and of a potential function Φ (see Section 2.3 in Chapter 2). We consider three ways of measuring the regrets:

(1) Linear approximation to the instantaneous external regret (see Section 2):

$$r_{i,t} = -\frac{x_{i,t}}{\boldsymbol{P}_t \cdot \boldsymbol{x}_t} ,$$

(2) Instantaneous internal regret (see Sections 4.2 and 5.1):

$$\widetilde{r}_{(i,j),t} = \left(\boldsymbol{P}_t^{i\to j} \cdot \boldsymbol{x}_t\right) - \ln\left(\boldsymbol{P}_t \cdot \boldsymbol{x}_t\right) ,$$

(3) Linear approximation to the instantaneous external regret (see Section 4.1):

$$r_{(i,j),t} = P_{i,t}\left(\frac{x_{j,t}}{\boldsymbol{P}_t \cdot \boldsymbol{x}_t} - \frac{x_{i,t}}{\boldsymbol{P}_t \cdot \boldsymbol{x}_t}\right) .$$

Also, both the exponential and the polynomial potentials are used. Each combination of $\mathbf{r}_t$ and Φ induces an investment strategy as summarized in Table 2.

Φ	$r_{i,t}$	$\widetilde{r}_{(i,j),t}$	$r_{(i,j),t}$
Exp	EG	GBH	B1EXP
Pol	–	B2POL	B1POL

TABLE 2. A first summary of the investment strategies.

The second overview indicates the external–internal regret minimizing pairs we used. For instance, the algorithm of Section 5.2 is the no internal regret counterpart of Cover's [**Cov91**] universal portfolio, and the algorithm of Section 7.2 is the no-internal-regret counterpart of the algorithm proposed in [**BlKa99**]. The other pairs are indicated in Table 3. The uniform buy-and-hold strategy of Section 2 is denoted by UBH.

external	EG	EG-UNIV	UBH
internal	B1EXP	B1EXP-UNIV	GBH

TABLE 3. A second summary of the investment strategies: the bottom line corresponds to the no internal regret counterparts of the algorithms of the top line.

The tuning of the EG **and** B1EXP **strategies.** The first experiment compares the behavior of the B1EXP and EG strategies whose results are summarized in Tables 4 and 5 and Figure 3. We compared the strategies EG and B1EXP for various choices of the tuning parameter η. We used the parameters suggested by theory $\eta^* = \alpha\sqrt{8\ln N/n}$ and $\eta^* = 4\alpha\sqrt{\ln N/n}$, respectively, in case of known time horizon n, and also the time varying versions $\eta_t^* = \alpha\sqrt{8\ln N/t}$ and $\eta_t^* = 4\alpha\sqrt{\ln N/t}$ where the ratio $\alpha = m/M$ is taken to be 0.5 for daily rebalancing and 0.3 for monthly rebalancing. (These values are estimated on the data.)

Tables 4 and 5 show the arithmetic averages of the wealths achieved on random samples of size 100. For example, the numbers in the columns "ten stocks" have been obtained by choosing ten of the 36 stocks randomly to form a market of $N = 10$ assets. This experiment was repeated 100 times and the averages of the achieved wealth factors appear in the table. The column "Freq." contains the number of times B1EXP outperformed EG of these 100 experiments. The average wealth ratios for both strategies were calculated for different fixed and time varying parameters. One of the interesting conclusions is that time varying updating never affects the performance of B1EXP while that of EG drops in case of monthly rebalancing or when the number of stocks is large.

In the rest of this experimental study both algorithms are used with their respective time varying theoretical optimal parameter η_t^*. It is also seen in Tables 4 and 5 that EG is less robust against a bad choice of η. Its performance degrades faster when η or η_t is increased.

Interestingly, the increase of the external regret when the tuning parameter is increased corresponds to an increase in the internal regret, as shown in Figure 3. The increase of the internal regret is far larger for the EG strategy. This suggests that minimizing internal regret results in more stability.

η	Monthly rebalancing					
(para-	Three stocks			Ten stocks		
meter)	EG	B1EXP	Freq.	EG	B1EXP	Freq.
2	14.7	15.5	73	12.8	19.2	95
1.5	15.1	16.0	76	14.0	19.9	96
1	15.9	16.7	80	16.0	20.6	97
0.5	17.3	18.0	84	18.8	21.3	97
0.2	18.7	19.0	84	20.7	21.6	97
0.15	18.9	19.2	84	21.0	21.7	95
0.1	19.2	19.4	84	21.3	21.8	94
0.05	19.5	19.6	82	21.6	21.8	94
0.03	19.6	19.7	82	21.7	21.8	94
0.02	19.7	19.7	82	21.8	21.9	94
0.01	19.7	19.7	82	21.8	21.9	94
η^*	19.5	19.5	80	21.4	21.8	94
η_t^*	19.3	19.4	80	21.2	21.7	95
$0.1\,\eta_t^*$	19.7	19.7	81	21.8	21.9	95
$0.2\,\eta_t^*$	19.7	19.7	80	21.7	21.8	95
$0.5\,\eta_t^*$	19.6	19.6	79	21.5	21.8	95
$2\,\eta_t^*$	18.9	19.0	81	20.5	21.5	95
$5\,\eta_t^*$	17.8	17.9	77	18.7	20.8	97
$10\,\eta_t^*$	16.5	16.7	71	16.1	19.8	94
$25\,\eta_t^*$	14.7	15.4	61	12.5	17.8	92

TABLE 4. Evolution of the achieved wealths according to the tuning parameter of EG and B1EXP both for fixed and time varying parameters. Computations are realized on random samples of size 100, arithmetic means are displayed. Monthly rebalancing.

Tuning of B1POL **and** B2POL**.** Table 6 shows that for B1POL and B2POL the theoretically (almost) optimal parameter $p = 4 \ln N$ performs quite poorly in our experiments, for it leads to too fast wealth reallocations. The values of p with better numerical performance are usually far smaller than the ones prescribed by theory. Thus, for the rest of this experimental study and the subsequent simulations, we choose $p = 2$, as it was originally suggested by [**Bla56**]. (Note that in Table 6 we show the geometric averages instead of the arithmetic ones, to take into account the huge dispersion of the wealths achieved by these two investment strategies – see also Table 9 and the related comments.)

Global comparison. In the next experiment various different investment strategies are compared, which we denominate by EG, B1EXP, B1POL, GBH, GBH2, B2POL, Cover's, UBH, B-CRP, and U-CRP. For the first six strategies we have already described how to tune (some of them do not require any tuning). The algorithm "Cover's" stands for Cover's universal portfolio based on the uniform density. To compute the universal portfolio, we drew at random 10^3 different constantly rebalanced portfolios and took the average on the wealth ratio sequences to compute each instance of Cover's algorithm. (The value 10^3 may seem to be too small in view of the 10^8 used in [**HeScSiWa98**] but calculations using the Chebyshev bound of [**BlKa99**] indicate that this value is sufficient to have a good idea of the order of the wealth achieved by the universal portfolio.) To

η	Daily rebalancing					
(para-	Three stocks			Ten stocks		
meter)	EG	B1EXP	Freq.	EG	B1EXP	Freq.
2	13.2	14.5	77	12.4	21.7	93
1.5	14.1	15.6	80	14.0	23.2	95
1	15.7	17.4	86	17.0	24.7	95
0.5	18.8	20.4	89	22.0	25.8	94
0.2	22.1	23.1	89	25.2	26.3	92
0.15	22.8	23.6	89	25.6	26.3	91
0.1	23.6	24.2	89	26.0	26.4	88
0.05	24.5	24.8	88	26.3	26.4	83
0.03	24.8	25.0	88	26.4	26.5	82
0.02	25.0	25.1	88	26.4	26.5	82
0.01	25.2	25.3	88	26.4	26.5	82
η^*	25.0	25.0	89	26.4	26.5	82
η_t^*	24.8	24.8	86	26.2	26.4	94
$0.1\,\eta_t^*$	25.3	25.3	88	26.5	26.5	91
$0.2\,\eta_t^*$	25.3	25.3	88	26.4	26.5	91
$0.5\,\eta_t^*$	25.1	25.1	87	26.3	26.4	92
$2\,\eta_t^*$	24.2	24.3	86	25.8	26.3	94
$5\,\eta_t^*$	22.6	22.7	85	24.5	26.0	98
$10\,\eta_t^*$	20.4	20.5	82	22.0	25.2	98
$25\,\eta_t^*$	16.2	16.4	72	15.2	22.3	99

TABLE 5. Evolution of the achieved wealths according to the tuning parameter of EG and B1EXP both for fixed and time varying parameters. Computations are realized on random samples of size 100, arithmetic means are displayed. Daily rebalancing.

compute the best constantly rebalanced portfolio (called B-CRP) we used a technique described in [**Cov84**], with (according to the notation therein) $\varepsilon = 10^{-4}$ for daily rebalancing and $\varepsilon = 10^{-5}$ for monthly rebalancing. This guarantees an estimate within a multiplicative factor of 1.0028 of the wealth achieved by the best constantly rebalanced portfolio in case of a monthly rebalancing and 1.7596 in case of a daily rebalancing. Nevertheless, the values thus obtained are often even closer to the optimal, despite the weak guarantees in case of daily rebalancing. We also considered the uniform buy-and-hold strategy UBH and, following [**BoElGo00**], the uniform constantly rebalanced portfolio (U-CRP).

Transaction costs were also taken into account (whose amount is indicated in the column TC of the tables) according to the model defined in [**BlKa99**], and recalled in Section 7. We implemented Blum and Kalai's optimal rebalancing algorithm, using different transaction costs. To be fair, we considered all algorithms in their no-transaction-cost definition – that is, we consider GBH and GBH2 instead of GBH_c and GBH2_c. Here, we summarize the results for zero transaction cost and a heavy 2% at-purchase transaction cost in case of monthly rebalancing and a milder 1% transaction cost when the rebalancing occurs daily.

All these algorithms were run on randomly chosen sets of stocks. The number of selected stocks is shown in the first column of Tables 7 and 8. These tables indicate the arithmetic averages of the wealths achieved. In each line, the results of the algorithm which outperformed its

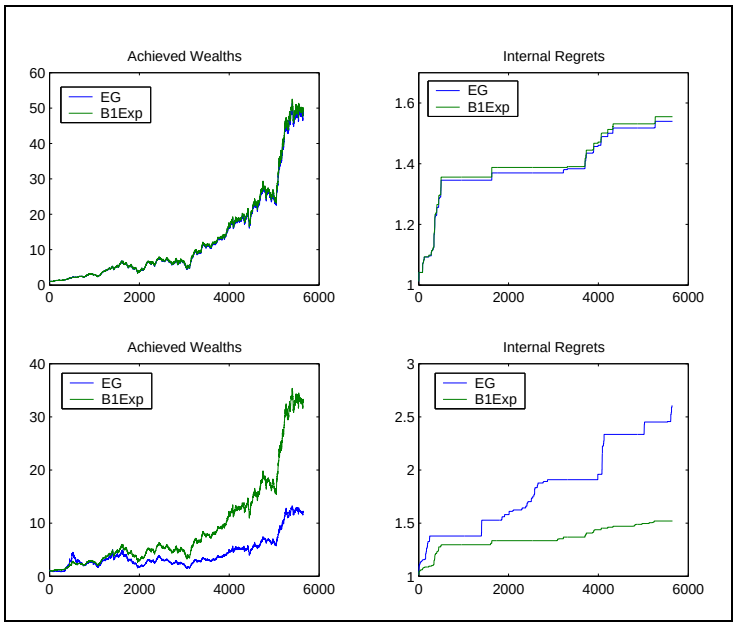

FIGURE 3. Evolution of both external and internal regrets for the optimal time varying tuning parameter (top) and a 25 times too large one (bottom). Stocks used: Dow Chemical, Coke, GTE, Mei Corp., Gulf, Iroquois, Kin Arc, Amer Brands, Fischbach, Lukens.

p	Monthly rebalancing				Daily rebalancing			
(parameter)	Three stocks		Ten stocks		Three stocks		Ten stocks	
	B1POL	B2POL	B1POL	B2POL	B1POL	B2POL	B1POL	B2POL
p^*	11.5	9.5	15.7	12.4	9.1	7.3	11.1	9.7
1.1	13.3	10.9	16.2	13.5	12.7	9.5	16.5	13.5
1.2	13.1	10.9	16.0	13.9	12.3	9.5	16.4	13.5
1.3	13.0	10.9	16.0	13.8	12.1	9.3	16.4	13.8
1.5	12.9	11.0	16.5	14.1	11.5	8.9	16.5	13.5
2	12.3	10.4	16.9	13.5	10.7	8.5	15.9	13.5
2.5	12.0	10.1	16.1	14.4	10.3	8.1	15.6	13.2
3	11.8	9.9	16.9	15.4	9.9	7.8	15.4	12.9
3.5	11.7	9.8	17.0	15.2	9.5	7.5	15.0	12.6
4	11.5	9.7	17.8	14.3	9.3	7.4	14.8	12.5
4.5	11.5	9.5	17.1	14.5	9.1	7.3	14.7	12.0
5	11.5	9.4	17.1	14.6	9.1	7.3	14.5	11.7
6	11.5	9.4	16.2	14.0	8.6	7.1	13.8	11.8
8	11.2	9.4	14.5	11.9	8.1	7.0	12.3	9.8
10	10.4	9.0	14.3	11.9	7.8	6.8	10.7	8.4

TABLE 6. Evolution of the achieved wealths according to the tuning parameter of B1POL and B2POL. Computations are realized on random samples of size 100, geometric means are displayed.

competitors the more often are set in bold face. Globally, B1EXP seems to have the best results in terms of accumulated wealth, but there are some fine variations which should be mentioned.

ST.	EG	B1EXP	B1POL	GBH	GBH2	B2POL	Cover's	UBH	B-CRP
2	**16.2**	16.2	12.4	13.6	13.6	12.0	15.5	13.6	21.0
3	19.3	**19.4**	15.6	16.1	15.5	13.8	18.4	14.9	30.2
5	20.0	**20.3**	16.6	18.0	16.3	13.2	19.6	14.9	39.6
8	21.3	**21.7**	20.9	20.2	17.6	17.4	21.2	15.4	53.9
10	21.2	**21.7**	19.3	20.6	17.7	15.3	21.3	15.2	61.2
12	20.9	**21.5**	18.1	20.5	17.3	16.0	21.1	14.6	62.4
15	21.9	**22.5**	20.4	21.8	18.3	17.6	22.2	15.3	72.3
18	21.0	**21.6**	17.8	21.1	17.9	16.0	21.4	15.0	76.3
20	21.3	**21.9**	19.7	21.5	18.1	17.5	21.8	15.2	80.3
25	21.4	**22.0**	20.5	21.6	18.2	17.1	21.9	15.2	85.9
2	**14.9**	14.9	10.6	13.7	13.7	10.6	14.5	13.7	20.2
3	16.8	**16.8**	11.1	14.9	14.5	9.9	16.2	14.2	26.9
5	18.5	**18.6**	11.5	17.2	16.1	9.6	18.1	15.0	36.3
8	17.8	**17.9**	9.6	17.2	15.9	9.3	17.6	14.7	46.1
10	18.9	**19.1**	10.3	18.3	16.5	8.6	18.8	14.9	51.2
12	19.0	**19.2**	10.4	18.7	17.0	9.4	19.0	15.4	57.4
15	19.9	**20.1**	10.2	19.7	17.6	9.0	19.9	15.7	65.1
18	19.1	**19.3**	8.9	19.0	17.0	7.7	19.2	15.1	67.3
20	18.5	**18.7**	9.2	18.5	16.6	7.7	18.6	14.9	68.1
25	19.1	**19.3**	10.0	19.2	17.2	7.7	19.3	15.3	75.8

TABLE 7. Arithmetic means of the wealths achieved on randomly selected sets of stocks, repeated 100 times. Monthly rebalancing. A different sample was drawn for each line of this table. Top lines correspond to a no transaction cost setting, whereas the bottom lines consider the case of 2% transaction costs.

First, EG is better than B1EXP when the portfolio is reduced to two stocks only. The reason that in this case the internal regret is nothing else than the external regret and the exponential weighted algorithm on which EG is based is known to be optimal for the minimization of the external regret. Second, in the presence of transaction costs and for a daily rebalancing, GBH performs well. This is due to its closeness to buy-and-hold. Interestingly enough, it performs considerably better than buy-and-hold, which is known to be valuable in the presence of such heavy transaction costs. Surprisingly enough, GBH2, which was designed to be a modification of GBH suffering a low external regret with respect to buy-and-hold, performs quite poorly compared to GBH. Actually, the wealths achieved by GBH2 seem to interpolate those of GBH and the uniform buy-and-hold strategy. Finally, the at first sight naive U-CRP strategy seems to have interesting results, as already noted in [**BoElGo00**], even though there are no theoretical guarantees for its universality (see for instance Table 12).

Finer comparison. After this global comparison, we compare B1EXP more carefully with the best opponents in case of no transaction costs, which are EG and B1POL. The comparison to EG is done in Table 10 which shows the geometric and arithmetic averages obtained, as well as the number of times B1EXP won and also by how much each algorithm outperformed the other. The value of Δ^+ indicates the maximal gap between B1EXP and EG (in the favour of the former) on the 100 elements of the randomly selected sample and Δ^- is in favour of the latter. We conclude from this table that (in case of no transaction costs) B1EXP is quite often better than EG, and even when it is outperformed by EG, the wealth then achieved by EG is just a bit smaller. The difference between

ST.	EG	B1EXP	B1POL	GBH	GBH2	B2POL	Cover's	UBH	B-CRP
2	**19.3**	19.2	11.4	13.6	13.6	10.3	17.2	13.6	20.4
3	24.8	**24.8**	13.0	16.2	15.1	10.8	21.6	13.9	28.8
5	31.6	**32.0**	16.5	23.6	19.4	11.9	29.1	15.6	47.9
8	28.2	**28.5**	16.7	25.2	19.6	13.9	27.4	15.1	59.5
10	26.2	**26.4**	17.5	24.7	19.1	15.2	25.8	14.5	67.3
12	29.0	**29.3**	18.5	27.8	20.4	15.5	28.7	14.6	87.1
15	27.6	**27.8**	18.0	27.2	20.2	15.3	27.7	14.7	98.6
18	29.3	**29.5**	19.1	29.0	21.2	16.2	29.3	15.1	121.8
20	28.1	**28.4**	18.3	28.0	20.8	16.4	28.3	15.0	120.3
25	28.9	**29.0**	19.1	28.9	21.2	17.3	29.0	15.1	153.9
2	**18.4**	18.3	9.7	15.9	15.9	8.3	17.5	15.9	19.0
3	17.4	**17.4**	8.0	15.3	14.9	6.8	16.6	14.4	21.1
5	18.6	**18.6**	5.7	17.0	15.8	4.4	18.0	14.5	28.2
8	18.9	**18.9**	5.0	18.0	15.9	3.9	18.5	13.7	36.7
10	20.3	**20.3**	5.2	19.9	17.5	3.7	20.1	15.1	43.5
12	20.9	20.9	5.3	**20.5**	17.4	4.0	20.7	14.5	51.3
15	19.7	19.6	4.6	**19.8**	17.0	3.7	19.6	14.5	55.3
18	20.7	20.6	4.8	**20.8**	17.8	3.9	20.6	14.9	66.3
20	20.3	20.2	4.2	**20.4**	17.4	3.4	20.2	14.7	71.6
25	20.5	20.3	4.5	**20.6**	17.7	3.6	20.4	15.0	83.7

TABLE 8. Arithmetic means of the wealths achieved on randomly selected sets of stocks, repeated 100 times. Daily rebalancing. A different sample was drawn for each line of this table. Top lines correspond to a no transaction cost setting, whereas the bottom lines consider the case of 1% transaction costs.

Stat.	EG	B1EXP	B1POL	GBH	GBH2	B2POL	Cover's	UBH
Min.	13.2	13.6	6.6	13.0	11.5	4.7	13.4	8.8
Ar. av.	20.9	21.5	18.1	20.5	17.3	16.0	21.1	14.6
Geo. av.	20.5	21.0	16.1	20.1	17.0	13.8	20.7	14.4
Max.	32.9	34.6	56.3	31.7	24.9	60.9	33.7	20.9
St. dev.	4.6	4.9	9.3	4.3	3.2	9.5	4.7	2.8

TABLE 9. Statistical characterization of the wealths achieved on the random sample corresponding to 12 stocks without transaction costs and monthly rebalancing. The minimum, arithmetic and geometric averages, maximum, and standard deviation of the achieved wealths are shown.

the two algorithms seems to be especially large when η is large, that is, for monthly rebalancing and/or many stocks. Table 9 reveals that B1POL and B2POL are not serious contenders because of their huge standard deviation and the extreme values. This is also illustrated by the catastrophic results of these algorithms in the presence of transaction costs and for a daily rebalancing, see Table 8. The reason is that B1POL and B2POL reallocate just too quickly, which can be good or bad. (See Figure 4.) This happens because of the property of the polynomial potential that only the nonnegative internal regrets count in the computation of the wealth allocation, and therefore when one stock dominates, almost all the weight is put on it, which is of course dangerous.

		Geom. Avg.		Arith. Avg.			Max.	
ST.	TC	EG	B1EXP	EG	B1EXP	Freq.	Δ^-	Δ^+
2	0 %	14.0	13.9	16.2	16.2	12	0.47	0.19
3	0 %	17.0	17.0	19.3	19.4	80	0.02	0.17
5	0 %	18.5	18.6	20.0	20.3	82	0.12	2.23
8	0 %	20.4	20.8	21.3	21.7	92	0.17	2.30
10	0 %	20.6	21.1	21.2	21.7	95	0.21	1.53
12	0 %	20.5	21.0	20.9	21.5	99	0.05	1.66
15	0 %	21.5	22.1	21.9	22.5	98	0.08	1.45
18	0 %	20.7	21.3	21.0	21.6	100		1.65
20	0 %	21.2	21.7	21.3	21.9	100		1.74
25	0 %	21.3	21.9	21.4	22.0	100		1.18
2	2 %	13.0	12.9	14.9	14.9	27	0.30	0.22
3	2 %	15.0	15.0	16.8	16.8	65	0.05	0.09
5	2 %	17.4	17.5	18.5	18.6	72	0.20	1.42
8	2 %	17.2	17.3	17.8	17.9	72	0.42	1.36
10	2 %	18.2	18.4	18.9	19.1	82	0.19	1.46
12	2 %	18.6	18.8	19.0	19.2	73	0.27	1.30
15	2 %	19.6	19.8	19.9	20.1	84	0.18	0.85
18	2 %	18.8	19.0	19.1	19.3	81	0.19	1.20
20	2 %	18.3	18.5	18.5	18.7	84	0.30	0.70
25	2 %	19.1	19.3	19.1	19.3	88	0.23	0.50

TABLE 10. Extensive comparison between the performances of EG and B1EXP on the samples of Table 7.

Ptf.	EG	B1EXP	B1POL	GBH	GBH2	B2POL	Cover's	UBH	U-CRP
L12	4.20	4.20	4.61	4.21	4.25	4.64	4.20	4.31	4.20
M12	4.68	4.67	6.32	4.68	4.77	6.71	4.68	4.93	4.67
H12	6.79	6.74	8.12	6.78	6.89	8.32	6.77	7.13	6.73
L24	4.32	4.30	5.66	4.31	4.40	5.84	4.31	4.55	4.30
H24	5.40	5.35	7.40	5.37	5.44	7.94	5.35	5.61	5.35
A36	4.87	4.81	6.94	4.83	4.94	7.21	4.81	5.13	4.81
L12	0.83	0.83	0.88	0.83	0.84	0.89	0.83	0.85	0.83
M12	0.88	0.88	1.11	0.88	0.90	1.14	0.88	0.93	0.88
H12	1.17	1.16	1.82	1.20	1.20	1.96	1.17	1.28	1.15
L24	0.82	0.82	1.01	0.83	0.84	1.03	0.82	0.86	0.82
H24	0.92	0.91	1.45	0.93	0.96	1.54	0.92	1.03	0.91
A36	0.85	0.85	1.25	0.85	0.88	1.28	0.85	0.94	0.85

TABLE 11. Volatilities (multiplied by 100) for portfolios chosen according to their volatilities, for monthly rebalancing (top lines) as well as for daily rebalancing (bottom lines).

Ptf.	EG	B1EXP	B1POL	GBH	GBH2	B2POL	Cover's	UBH	U-CRP
L12	10.9	**11.1**	7.6	10.8	10.1	7.7	11.0	9.4	11.2
M12	17.2	17.1	**22.9**	17.1	16.9	21.9	17.0	16.7	17.1
H12	36.3	**39.0**	12.8	34.6	25.3	10.2	37.8	17.6	39.8
L24	13.9	14.0	**19.8**	14.0	13.5	15.7	14.1	13.1	14.1
H24	26.7	27.8	**41.3**	27.1	21.8	21.7	27.6	17.2	28.0
A36	20.5	21.1	**30.9**	20.8	17.5	22.5	20.7	14.5	21.1
L12	12.3	**12.4**	6.7	12.0	11.1	6.5	12.2	10.1	12.4
M12	16.1	**16.2**	9.9	15.8	14.8	9.4	16.0	13.9	16.2
H12	78.1	**81.0**	40.8	67.9	40.2	21.9	76.0	19.5	81.9
L24	14.3	**14.4**	9.3	14.2	13.1	9.0	14.4	12.0	14.4
H24	38.2	**38.7**	25.6	38.1	26.1	21.9	38.6	16.7	38.8
A36	26.9	**27.1**	20.2	27.1	20.2	17.4	27.0	14.5	27.1

TABLE 12. Wealths achieved by the portfolios of Table 11. In each line, the wealth obtained by the best adaptive algorithm is set in bold face.

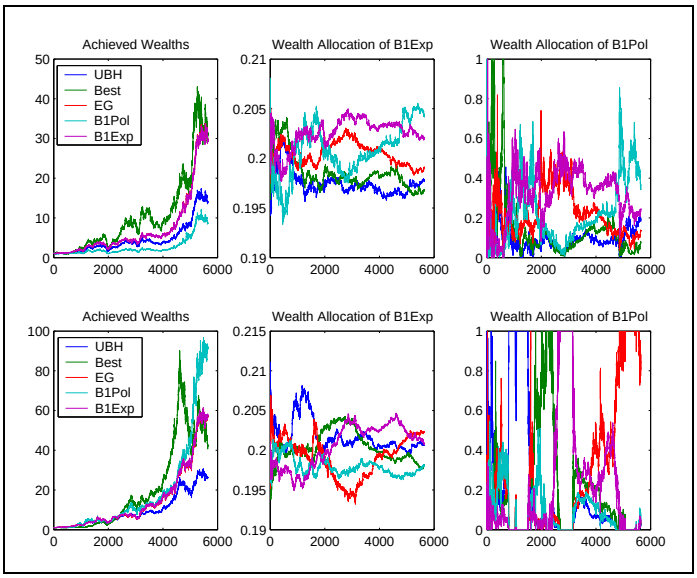

FIGURE 4. Evolution of the achieved wealths and of the wealth allocations on two typical examples. Stocks used for top graphs: Sher Will, Texaco, AHP, Espey, Helwett Packard. For bottom graphs: Gulf, JNJ, Mei Corp., Pillsbury, Schlum.

Tables 11 and 12 are given for sake of completeness as well as to allow comparison with [**HeScSiWa98**]. The algorithms are run on portfolios chosen according to the volatilities of the stocks. (See Remark 8.3 for a formal definition of the volatility.) Three groups were formed by putting the 12 lowest volatility stocks in the first group (L12), then the 12 highest in the second (H12) and the 12 remaining in the third group (M12). The group formed by L12 and M12 is called L24, the one of M12 and H12 is denoted by H24. Finally, the set of all 36 stocks is referred to as A36. Note that the B1EXP strategy has almost always the lowest volatilities. Thanks to its aggressive rebalancing, the B1POL strategy has interesting achieved wealths for monthly rebalancing. Nevertheless, the B1EXP investment scheme has globally the higher returns.

CHAPTER 8

Learning correlated equilibria in games with compact sets of strategies

In this final chapter, we study Hart and Schmeidler's extension of correlated equilibrium to games with infinite sets of strategies. General properties of the set of correlated equilibria are described. It is shown that, just like for finite games, if all players play according to an appropriate regret–minimizing strategy then the empirical frequencies of play converge to the set of correlated equilibria whenever the strategy sets are convex and compact.

Contents

Most of this chapter is based on the submitted paper [**StLu04**]. The section about convergence to linear correlated equilibria is however published here for the first time.

1. Introduction

Correlated equilibrium, introduced by Aumann [**Aum74, Aum87**] is arguably one of the most natural notions of equilibrium. Put simply, a correlated equilibrium is a joint distribution π over the set of strategies of the players that has the property that if, before taking an action, each player receives a recommendation such that the recommendations are drawn randomly according to the joint distribution of π, then, in an average sense, no player has an incentive to divert from the recommendation, provided that all other players follow theirs. The distinguishing feature of the notion is that, unlike in the definition of Nash equilibria, the recommendations do not need to be independent. Indeed, if π is a product measure, it becomes a Nash equilibrium.

A remarkable property of correlated equilibrium, pointed out by Foster and Vohra [**FoVo97**], is that if the game is repeated infinitely many times such that every player plays according to a certain regret–minimization strategy, then the empirical frequencies of play converge to the set of correlated equilibria. (See also Fudenberg and Levine [**FuLe99**], Hart and Mas–Colell [**HaMa00, HaMa01, HaMa02**].) No coordination is necessary between the players, and the players don't even need to know the others' payoff functions. Hart and Mas–Colell [**HaMa03**] show that Nash equilibrium does not share this property unless the game has quite special properties, see however Section 5.

The definition of correlated equilibrium was extended to infinite games by Hart and Schmeidler [**HaSc89**]. The purpose of this chapter is to study the correlated equilibria of a large class of infinite games. In Section 2 we recall Hart and Schmeidler's extended definition, and propose some equivalent formulations. One of them may be given by discretizing the sets of strategies, considering correlated equilibria of the discretized (finite) games, and taking appropriate limits as the discretization becomes finer (see Theorem 8.2). Some basic properties of correlated equilibria are described. In particular, under general conditions, the set of correlated equilibria is a compact convex set (see Theorem 8.1).

The main result of the chapter (Theorems 8.3 and 8.4) generalizes the above–mentioned result of Foster and Vohra to the case when the sets of strategies are compact and convex subsets of a normed space, and the payoff function of player k is continuous. It is shown that convergence of the empirical frequencies of play to the set of correlated equilibria can also be achieved in this case, by playing internal regret–minimizing strategies, where the notion of internal regret has to be generalized to the case of games with infinite strategy sets. The proof of the main theorem is given by a sequence of results, by broadening the class of departure functions in each step.

We then indicate a connection between the correlated equilibria of a finite game and those of its mixed extension. We show that in some sense, these are equivalent.

We conclude the chapter with a note about the computability of the offered procedures. We focus especially on on-line linear regression, and provide efficient internal and external regret minimizing forecasting schemes.

2. Definition of correlated equilibrium

2.1. Refined definition. The notion of correlated equilibrium was introduced by Aumann [**Aum74, Aum87**] who assumed that the sets of strategies are finite, and extended later by Hart and Schmeidler [**HaSc89**] to infinite games.

Formally, consider an N–person game in strategic (normal) form

$$\Gamma = \left(\{1, \ldots, N\}, (S^i)_{1 \leqslant i \leqslant N}, (h^i)_{1 \leqslant i \leqslant N}\right) \ ,$$

where $\{1,\dots,N\}$ is the finite set of players, player i is given a (non necessarily finite) set of strategies S^i and a payoff function $h^i : S \to \mathbb{R}$. The set of N–tuples of strategies is denoted by $S = S^1 \times S^2 \times \ldots \times S^N$. We use the notation $s = (s^{-i}, s^i)$, where

$$s^{-i} = (s^1, \dots, s^{i-1}, s^{i+1}, \dots, s^N)$$

denotes the strategies played by everyone but player i. We write $s^{-i} \in S^{-i}$, where $S^{-i} = \prod_{j\neq i} S^j$.

Some assumptions on the topology of the S^k are required. More precisely, assume that the S^k are topological spaces, equipped with their Borel σ–algebra (that is, the σ–algebra generated by the open sets). Then S is naturally equipped with a (product) topology and a (product) σ–algebra. We can now consider (Borel) probability measures over S.

Hart and Schmeidler's original definition[1] [**HaSc89**] states that a correlated equilibrium π of the game Γ is a (joint) probability distribution over S such that the extended game Γ' defined below admits the N–tuple of identity functions $S^k \to S^k$ as a Nash equilibrium. The strategy sets of Γ' are given, for player k, by the set $\mathcal{F}_k$ of all measurable maps $S^k \to S^k$, and the game is played as follows: each player k chooses his action $\psi_k \in \mathcal{F}_k$, a signal (sometimes called recommendation) $\boldsymbol{I} = (I^1, \dots, I^N) \in S$ is drawn randomly according to π, player k is told the k–th component of the signal, I^k, and he finally plays $\psi_k(I^k)$.

We remark here that the set $\mathcal{F}_k$ of allowed departures for player k may actually be taken as a (sometimes proper) subset of the set $\mathcal{L}^0(S^k)$ of all measurable departures $S^k \to S^k$, with the only restriction that it should contain the identity map. We then define a $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibrium similarly as above except that we consider departure functions ψ_k only from the class $\mathcal{F}_k$, $k = 1, \dots, N$. In the simplest cases $\mathcal{F}_k$ may be a finite set, but we also consider larger classes $\mathcal{F}_k$ given by the set of all linear functions, all continuous functions, or all measurable functions.

A more formal statement is the following one. We restrict our attention to real–valued, bounded, and measurable or, alternatively, nonnegative and measurable, payoff functions h^k.

DEFINITION 8.1. *A* $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibrium *is a (joint) distribution π over S such that for all players k and all departure functions $\psi_k \in \mathcal{F}_k$, one has*

$$\mathbb{E}_\pi\left[h^k(I^{-k}, I^k)\right] \geqslant \mathbb{E}_\pi\left[h^k(I^{-k}, \psi_k(I^k))\right] , \tag{8.1}$$

where the random vector $\boldsymbol{I} = (I^k)_{1\leqslant k\leqslant N}$, taking values in S, is distributed[2] *according to π.*

π is a $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated ε–equilibrium *if for all k and all $\psi_k \in \mathcal{F}_k$,*

$$\mathbb{E}_\pi\left[h^k(I^{-k}, I^k)\right] \geqslant \mathbb{E}_\pi\left[h^k(I^{-k}, \psi_k(I^k))\right] - \varepsilon .$$

A correlated equilibrium may be interpreted as follows. In an average sense (the average being given by the randomization associated with the signal), no player has an incentive to divert from the recommendation, provided that all other players follow theirs. The distinguishing feature of this notion is that, unlike in the definition of Nash equilibria, the random variables I^k do not need to be independent. Indeed, if π is a product measure, it becomes a Nash equilibrium. This also means that correlated equilibria always exist as soon as Nash equilibria do, which is ensured under minimal assumptions (see Remark 8.1). Their existence may however also be seen without underlying fixed point results, see Hart and Schmeidler [**HaSc89**].

[1]Note that we only consider games with finitely many players, and therefore avoid some of the difficulties arising in games with an infinite number of players.

[2]Note that one can always take for $\boldsymbol{I}$ the identity map over S, thought of as a random vector defined on the probability space (S, π).

REMARK 8.1. In the definition of correlated equilibria we consider an extension of the original game. But note that under minimal assumptions (e.g., if the sets of strategies S^k are convex compact subsets of topological vector spaces and the payoffs h^k are continuous and concave in the k–th variable) there exists a Nash equilibrium in pure strategies (see, e.g., [**Aub79**]). Each pure Nash equilibrium corresponds to a $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibrium π given by a Dirac measure over S. Clearly, π is a mixed Nash equilibrium if and only if it is a $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibrium equal to the product of its marginals.

EXAMPLE 8.1. Assume that each S^k is a convex and compact subset of a normed vector space, and that each payoff function h^k is continuous. In Section 3.6 we show that the set of $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibria coincides with the set of $(\mathcal{C}(S^k))_{1\leqslant k\leqslant N}$–correlated equilibria (where $\mathcal{C}(S^k)$ is the set of all continuous functions mapping S^k in S^k). This set is convex, compact, and contains the non–empty set of (pure and mixed) Nash equilibria.

For the sake of completeness, we give an analog of the conditional definition which is usually proposed as a definition for correlated equilibria in the case of finite games: provided that the S^k are finite sets, a *correlated equilibrium* is a (joint) distribution π over S such that for all players k and all functions $\psi_k : S^k \to S^k$, one has

$$\sum_{s\in S} \pi(s^{-k} \,|\, s^k) \left(h^k(s^{-k}, s^k) - h^k(s^{-k}, \psi_k(s^k)) \right) \geqslant 0 \,,$$

where $\pi(\,\cdot \,|\, s^k)$ is the conditional distribution of S^{-k} given that player k is advised to play s^k. Now, recalling that we denote by $\mathcal{L}^0(S^k)$ the set of all measurable functions over S^k, we have the following conditional definition in the general case where the game may be finite or infinite. The proof is immediate.

PROPOSITION 8.1. *Under the above measurability assumptions, a distribution π over S is a $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibrium if and only if for all players k and all measurable departure maps $\psi_k : S^k \to S^k$,*

$$\mathbb{E}_\pi \left[h^k(I^{-k}, I^k) \,|\, I^k \right] \geqslant \mathbb{E}_\pi \left[h^k(I^{-k}, \psi_k(I^k)) \,|\, I^k \right] \,,$$

where the random vector $\boldsymbol{I} = (I^k)_{1\leqslant k\leqslant N}$ is distributed according to π.

Finally, a last link between the usual definition for finite games and the one for infinite games is given in Section 5.

2.2. Basic properties of $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibria. Fix the set of allowed departures $\mathcal{F}_k$, $1 \leqslant k \leqslant N$, and denote by Π the set of all $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibria. It is immediate from the definition that Π is a convex set, and that it contains the set of Nash equilibria (which is known to be non–empty under minimal assumptions, see Remark 8.1 above).

For the subsequent analysis we need to establish a topological property of Π, namely its compactness. To this end, we assume that each S^k is a Polish space, that is, S^k is a complete and separable metric space. Then the product S is also Polish, and it is well–known that Borel probability measures over S are regular. Denote by $\mathcal{C}(S, \mathbb{R})$ the set of bounded continuous real–valued functions over S. The set of Borel probability measures over S, denoted by $\mathcal{P}(S)$, (and more precisely, the set of all nonnegative and finite Borel measures over S) is equipped with the $\mathcal{C}$–weak–$*$ topology. This is the weakest topology such that, for each $f \in \mathcal{C}(S, \mathbb{R})$, the linear map $\mu \to \mu[f]$ defined for $\mu \in \mathcal{P}(S)$ is continuous, where $\mu\,[f] = \int_S f \mathrm{d}\mu$. That is, the open sets of this topology are generated by the sets

$$\{\mu \in \mathcal{P}(S) : \mu\,[f] < \alpha\} \,,$$

where f is any element of $\mathcal{C}(S, \mathbb{R})$ and α any real number.

Assume furthermore[3] that the S^k are compact. Then S is also a compact set (and $\mathcal{C}(S, \mathbb{R})$ is simply the set of all continuous real valued functions over S). Recall the following simple statement of Prohorov's theorem, see [**EtKu86**]:

PROPOSITION 8.2 (Prohorov's theorem). *If S is a compact metric space, then the space $\mathcal{P}(S)$ is compact. Its topology is equivalent to the topology of the so–called Prohorov metric. In particular, $\mathcal{P}(S)$ is sequentially compact, that is, every sequence of elements from $\mathcal{P}(S)$ contains a convergent subsequence.*

The next result summarizes some of the basic properties of the set Π. Recall that by Example 8.1, under some mild conditions, the set of correlated equilibria with respect to all measurable departures equals the set of correlated equilibria with respect to all continuous departures. Thus, the assumption in the following theorem that departure functions are continuous may be weakened in some important cases.

THEOREM 8.1. *Assume that the strategy spaces S^k are compact metric spaces. The set Π of $(\mathcal{F}_k)_{1 \leqslant k \leqslant N}$–correlated equilibria is non–empty whenever the payoff functions h^k are continuous over S. Moreover, Π is a convex set, which contains the convex hull of Nash equilibria. If, in addition, for all k, the payoff functions h^k are continuous and $\mathcal{F}_k \subset \mathcal{C}(S^k)$, where $\mathcal{C}(S^k)$ is the set of all continuous functions mapping S^k into S^k, then Π is compact.*

PROOF. The non–emptiness of Π under the assumption of continuity of the payoff functions follows either from Theorem 3 of Hart and Schmeidler [**HaSc89**] or, alternatively, from the existence of a mixed Nash equilibrium. (The latter may be shown by checking the hypotheses of a version of Nash's theorem stated in Remark 8.1, which follows easily by an application of Prohorov's and Stone–Weierstrass theorems.)

It remains to prove the compactness of Π under the given assumptions. By Prohorov's theorem, Π is included in a compact set, therefore it is enough to prove that Π is a closed set. To this end, consider the continuous real–valued function over S defined by

$$f_{k,\psi_k}(s) = h^k(s^{-k}, s^k) - h^k(s^{-k}, \psi_k(s^k)) \ ,$$

where $1 \leqslant k \leqslant N$ and $\psi_k \in \mathcal{F}_k \subset \mathcal{C}(S^k)$. Each f_{k,ψ_k} is a continuous real–valued function over S and Π is the intersection of the closed half–spaces

$$\{\mu \in \mathcal{P}(S) : \mu\,[f_{k,\psi_k}] \geqslant 0\} \ .$$

□

REMARK 8.2. Note that in general, contrary to the finite case, the set of correlated equilibria of a non-finite game, though given by an intersection of closed halfspaces, is not necessarily a convex polyhedron, as the intersection may be infinite. Nevertheless, if the $\mathcal{F}_k$ are finite sets and the S^k are subsets of finite dimensional spaces, then the above proof shows that the set of $(\mathcal{F}_k)_{1 \leqslant k \leqslant N}$–correlated equilibria is a convex polyhedron.

2.3. Discretized games. An alternative natural definition of correlated equilibrium in games with infinite strategy spaces is obtained by discretization. The idea is to "discretize" the sets of strategies and consider the set of correlated equilibria of the obtained finite game. Now appropriate "limits" may be taken as the discretization gets finer. In this section we make this definition precise and show that the obtained definition coincides, under general conditions, with the definition given above when one allows all measurable departure functions.

[3]Note that a compact metric space is always Polish.

A $(\mathcal{P},\mathcal{D})$ *discretization* of the game $\Gamma = \left(\{1,\dots,N\},(S^i)_{1\leqslant i\leqslant N},(h^i)_{1\leqslant i\leqslant N}\right)$ is given by a product partition $\mathcal{P}$, a grid $\mathcal{D}$ and induced payoffs h_d^k, $1 \leqslant k \leqslant N$. More precisely, a product partition is an N–tuple $(\mathcal{P}^1,\dots,\mathcal{P}^N)$, where each $\mathcal{P}^k$ is a finite measurable partition of the corresponding strategy set S^k, which we denote by $\mathcal{P}^k = \{V_1^k,\dots,V_{N_k}^k\}$. In every set V_i^k, $1 \leqslant k \leqslant N$, $1 \leqslant i \leqslant N_k$, we pick an arbitrary element $t_i^k \in V_i^k$. These points form a *grid* $\mathcal{D}^k = \{t_1^k,\dots,t_{N_k}^k\}$. We write $\mathcal{D} = \mathcal{D}^1 \times \ldots \times \mathcal{D}^N$. The induced payoffs h_d^k are obtained simply by restricting the original payoff functions to the grid $\mathcal{D}$.

Now, for a given discretization $(\mathcal{P},\mathcal{D})$, a distribution π over S induces a discrete distribution π_d over the grid $\mathcal{D}$ by

$$\pi_d\left(t_{i_1}^1 \times \ldots \times t_{i_N}^N\right) = \pi\left(V_{i_1}^1 \times \ldots \times V_{i_N}^N\right) .$$

The size r of a discretization $(\mathcal{P},\mathcal{D})$ is the maximal diameter of the sets V_i^k, $1 \leqslant k \leqslant N$, $1 \leqslant i \leqslant N_k$. If each S^k is compact, then every discretization has a finite size. Then we have the following characterization of correlated equilibria with respect to all measurable departures. (The fairly straightforward proof is postponed to the end of this chapter.)

THEOREM 8.2. *Assume that all strategy spaces S^k are convex and compact subsets of a normed space and that the h^k are continuous functions over S. Then a probability distribution over S is a $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibrium of the game Γ if and only if there exists a function ε with $\lim_{r\to 0}\varepsilon(r) = 0$ such that for all discretizations $(\mathcal{P},\mathcal{D})$ of size r, π induces an $\varepsilon(r)$–correlated equilibrium.*

Note that the above result is more precise than the results contained in the proofs of Theorems 2 and 3 of Hart and Schmeidler [**HaSc89**], where a correlated equilibrium of a given game with infinite strategy sets was only seen as a cluster point of the set of correlated equilibria of the discretized games.

3. Regret minimization and convergence in repeated games

One of the remarkable properties of correlated equilibrium in finite games is that if the game is played repeatedly many times such that every player plays according to a certain regret–minimization strategy then the empirical frequencies of play converge to the set of correlated equilibria. No coordination is necessary between the players, the player don't even need to know the others' payoff functions. This property was first proved by Foster and Vohra [**FoVo97**], see also Fudenberg and Levine [**FuLe99**], Hart and Mas–Colell [**HaMa00, HaMa01, HaMa02**], Lehrer [**Leh97, Leh03**].

The purpose of this section is to investigate to what extent the above–mentioned convergence result can be extended to games with possibly infinite strategy spaces.

We consider a situation in which the game Γ is played repeatedly at time instances $t = 1, 2, \ldots$. The players are assumed to know their own payoff function and the sequence of strategies played by all players up to time $t-1$. (This is known as the uncoupledness property in the literature, see Hart [**Har04**].)

The main result of the chapter, summarized in the following theorem, shows that under general conditions, if all players follow a certain regret–minimizing strategy, the empirical frequencies of play converge to the set of correlated equilibria. Thus, on the average, a correlated equilibrium is achieved without requiring any cooperation among the players.

THEOREM 8.3 (Main result). *Assume that all the strategy spaces S^k are convex and compact subsets of a normed space and all payoff functions h^k are continuous over S and concave in the k–th strategy. Then there exists a regret minimizing strategy such that, if every player follows such*

a strategy, then joint convergence of the sequence of empirical plays to the set of $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$*–correlated equilibria is achieved.*

Thus, the convergence result extends, under quite general assumptions, even if all possible measurable departure functions are allowed in the definition of correlated equilibrium. The only restrictive assumption is the concavity of the payoffs. This condition may be removed by allowing the players to use randomized strategies. The next theorem asserts that almost sure convergence of the empirical frequencies of play to the set of correlated equilibria is achieved under the only assumption that the payoff functions are continuous.

THEOREM 8.4 (Main result, randomized version). *Assume that all strategy spaces* S^k *are convex and compact subsets of a normed space and all payoff functions* h^k *are continuous over* S*. If the players are allowed to randomize, then there exists a regret minimizing strategy such that, if every player follows such a strategy, then (almost surely with respect to the auxiliary randomizations used) joint convergence of the sequence of empirical plays to the set of* $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$*–correlated equilibria is achieved almost surely.*

Of course, under the assumptions of Theorem 8.3 (or 8.4), there exist Nash equilibria in pure (or mixed) strategies, see Remark 8.1. But we note that the mentioned theorems lead to the interesting by-product that the set of correlated equilibria are non-empty (the latter property is indeed not required for their proofs). They provide a constructive and in this chapter self–contained proof of the following existence result.

COROLLARY 8.1. *Under the assumptions of Theorem 8.4, and with its notation, the set of* $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$*–correlated equilibria is non–empty.*

Note that our internal regret minimizing procedures below use the same kind of argument that those used in the direct existence proofs for correlated equilibria proposed by [**HaSc89**] – where a direct proof means a proof that does not make use of the argument that Nash equilibria are special cases of correlated equilibria.

REMARK 8.3. Even though players played independently, there is correlation in the limiting distributions, since these are given by correlated equilibria of the one-shot game. This is due to the minimal form of coordination that lies in all players' decisions to play internal regret minimizing strategies.

Theorems 8.3 and 8.4 are proved below by a series of results, some of which may be of independent interest. In particular, we define a notion of internal regret in the case of games with infinite strategy sets. Moreover, we give precise upper bounds for this internal regret in some cases, see Theorems 8.7 and 8.8, as well as Section 3.8.

REMARK 8.4. The regret minimizing strategies considered in Theorem 8.3 are deterministic in the sense that players do not need to randomize. This is made possible because of the concavity assumption on the payoffs. An example is the mixed extension of a finite game, which may be seen to satisfy the assumptions of Theorem 8.3. This means that if the game is played in the mixed extension (i.e., in each round the players output a probability distribution over the set of actions, and get as payoffs the expectations of their original payoff functions under these distributions), then joint convergence to the set of correlated equilibria (with respect to all measurable departures or just linear departures) may be achieved in the mixed extension, in a non-randomized way. It is easy to see that any of these sets of correlated equilibria of the mixed extension induces, in a natural way, the set of correlated equilibria of the underlying finite game. Thus, our algorithm

generalizes the (randomized) algorithms designed for the case of finite games. See Section 4 for more details.

3.1. Internal regret. The notion of correlated equilibrium is intimately tied to that of internal (or conditional) regret. Intuitively, internal regret is concerned with the increase of a player's payoff gained by simple modifications of the played strategy. If a simple modification results in a substantial improvement then a large internal regret is suffered.

The formal definition of internal regret (see, e.g., [**FoVo99**]) may be extended to general games in a straightforward manner as follows: let $\mathcal{F}_k$ be a class of functions $\psi_k : S^k \to S^k$. As the game Γ is repeated, at each round t, player k could play consistently $\psi_k(s_t^k)$ whenever his strategy prescribes him to play $s_t^k \in S^k$. This results in a different strategy, called the ψ_k–modified strategy. The maximal cumulative difference in the obtained payoffs for player k, for n rounds of play, equals

$$\mathcal{R}^k_{\psi_k,n} = \max_{s_1^{-k},\ldots,s_n^{-k}} \left(\sum_{t=1}^n h^k(s_t^{-k}, \psi_k(s_t^k)) - \sum_{t=1}^n h^k(s_t^{-k}, s_t^k) \right) ,$$

where the maximum is taken over all possible sequences of opponent players' actions. We call $\mathcal{R}^k_{\psi_k,n}$ the *internal regret of player k with respect to the departure ψ_k at round n*. The intuition is that if $\mathcal{R}^k_{\psi_k,n}$ is not too large, then the original strategy cannot be improved in a simple way.

We say that a strategy for player k *suffers no internal regret (or minimizes his internal regret) with respect to a class $\mathcal{F}_k$ of departures* whenever

$$\limsup_{n\to\infty} \frac{1}{n} \mathcal{R}^k_{\psi_k,n} \leqslant 0 ,$$

for all $\psi_k \in \mathcal{F}_k$. The departure functions play a similar role as in the general definition of correlated equilibria. As in the finite case, if all players minimize their internal regrets, then joint convergence of the sequence of empirical distribution of plays to the set of correlated equilibria is achieved. Denote by $s_1, \ldots, s_n$ the played strategies up to time n. We denote by π_n the empirical distribution of plays up to time n:

$$\pi_n = \frac{1}{n} \sum_{t=1}^n \delta_{s_t} ,$$

where δ_s is the Dirac mass on $s \in S$. More precisely, we have the following convergence result generalizing the corresponding statement of Foster and Vohra [**FoVo97**] for finite games.

THEOREM 8.5. *If each player k minimizes his internal regret with respect to a departure class $\mathcal{F}_k$, then, provided that the S^k are compact metric spaces, the h^k are continuous, and $\mathcal{F}_k \subset \mathcal{C}(S^k)$ for all k, the empirical distribution of plays $(\pi_n)_{n\in\mathbb{N}}$ converges to the set of $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibria.*

PROOF. The assumption on the internal regrets may be rewritten as

$$\limsup_{n\to\infty} \mathbb{E}_{\pi_n} \left[h^k(I^{-k}, \psi_k(I^k)) \right] - \mathbb{E}_{\pi_n} \left[h^k(I^{-k}, I^k) \right] \leqslant 0 \tag{8.2}$$

for all k and all $\psi_k \in \mathcal{F}_k$, where $\boldsymbol{I} = (I^k)_{1\leqslant k\leqslant N}$ is the identity map over S, defined on the probability space (S, π_n) ($\mathbb{E}_{\pi_n}$ simply denotes expectation with respect to this probability measure π_n over S). By Prohorov's theorem, the sequence $(\pi_n)_{n\in\mathbb{N}}$ lies in a compact metric space. Thus, if the whole sequence did not converge to the set of $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibria, we could extract from it a subsequence $(\pi_{\phi(n)})_{n\in\mathbb{N}}$, where ϕ is an increasing function $\mathbb{N} \to \mathbb{N}$, such that $(\pi_{\phi(n)})_{n\in\mathbb{N}}$ converges to a probability measure π which is not a $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibrium. That is,

there exists a player k, $1 \leqslant k \leqslant N$, and a departure $\psi_k \in \mathcal{F}_k$ such that

$$\mathbb{E}_\pi \left[h^k(I^{-k}, I^k) \right] < \mathbb{E}_\pi \left[h^k(I^{-k}, \psi_k(I^k)) \right] . \tag{8.3}$$

But (8.2) ensures that

$$\limsup_{n\to\infty} \mathbb{E}_{\pi_\phi(n)} \left[h^k(I^{-k}, \psi_k(I^k)) \right] - \mathbb{E}_{\pi_\phi(n)} \left[h^k(I^{-k}, I^k) \right] \leqslant 0 .$$

By continuity of the function f_{k,ψ_k} defined by

$$f_{k,\psi_k}(s) = h^k(s^{-k}, s^k) - h^k(s^{-k}, \psi_k(s^k)) , \quad s \in S$$

and by the definition of the weak–$*$ topology over $\mathcal{P}(S)$, we have

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}_{\pi_\phi(n)} \left[h^k(I^{-k}, \psi_k(I^k)) \right] - \mathbb{E}_{\pi_\phi(n)} \left[h^k(I^{-k}, I^k) \right] \\ = \mathbb{E}_\pi \left[h^k(I^{-k}, \psi_k(I^k)) \right] - \mathbb{E}_\pi \left[h^k(I^{-k}, I^k) \right] \leqslant 0 , \end{aligned}$$

which contradicts (8.3), thus proving the desired convergence. □

REMARK 8.5. The above proof shows that the uniformity over the opponents' plays required in the definition of internal regret is not needed to get the convergence result of Theorem 8.5. A kind of Hannan consistency with respect to the departures indexed by the class is enough. However, for all the algorithms introduced below, we are able to prove uniform bounds.

Theorem 8.5 shows that in order to guarantee convergence of the empirical frequencies of play to the set of correlated equilibria, it suffices that all players use a strategy that minimizes their internal regret. Note that the main issues in designing such a strategy concern the size of the set of allowed departures $\mathcal{F}_k$. For finites games, the measurable departures $S^k \to S^k$ are given by all functions $S^k \to S^k$, which are in finite number $m_k^{m_k}$, where m_k is the cardinality of S^k. If S^k is infinite (countably or continuously infinite), there is *a priori* an infinite number of departures. In particular, a simple procedure allocating a weight per each departure function, as was proposed in the finite case in Foster and Vohra [**FoVo99**] and Hart and Mas–Colell [**HaMa01**], would probably be impossible if the set of allowed departures was too large. Thus, previous learning algorithms cannot be generalized as easily as the definition could be carried over to the infinite case. Designing new learning algorithms for some general classes of infinite games will be the point of the subsequent sections.

3.2. Blackwell's condition. Regret–minimization strategies have been often derived from Blackwell's approachability theorem [**Bla56**]. Here however, we do not need the full power of Blackwell's theory, only a few simple inequalities derived in Cesa–Bianchi and Lugosi [**CeLu03**] which we briefly recall (see also Remark 3.2 and Section 4.2 of Chapter 7).

Consider a sequential decision problem parameterized by a *decision space* $\mathcal{X}$, by an *outcome space* $\mathcal{Y}$. At each step $t = 1, 2, \ldots$, the decision maker selects an element $\widehat{x}_t$ from the decision space $\mathcal{X}$. In return, an outcome $y_t \in \mathcal{Y}$ is received, and the decision maker suffers a vector $\boldsymbol{r}_t = \boldsymbol{r}_t(\widehat{x}_t, y_t) \in \mathbb{R}^N$ of regret. The cumulative regret after t rounds of play is $\boldsymbol{R}_t = \sum_{s=1}^t \boldsymbol{r}_s$. The goal of the decision maker is to minimize $\max_{i=1,\ldots,N} R_{i,n}$, that is, the largest component of the cumulative regret vector after n rounds of play.

Similarly to Hart and Mas–Colell [**HaMa01**], we consider potential–based decision–making strategies, based on a convex and twice differentiable *potential function* $\Phi : \mathbb{R}^N \to \mathbb{R}^+$. Even though most of the theory works for a general class of potential functions, for concreteness and to get the best bounds, we restrict our attention to the special case of the exponential potential given

by

$$\Phi(\mathbf{u}) = \sum_{i=1}^{N} \exp(\eta u_i) \ ,$$

where the parameter $\eta > 0$ will be tuned by the analysis below.

We recall the following bound, proved in [**CeLu03**].

PROPOSITION 8.3. *Assume that the decision–maker plays such that in each round t of play, the regret vector $\boldsymbol{r}_t$ satisfies the so–called "Blackwell condition"*

$$\nabla\Phi(\boldsymbol{R}_{t-1}) \cdot \boldsymbol{r}_t \leqslant 0 \ . \tag{8.4}$$

If $\|\boldsymbol{r}_t\|_\infty \leqslant M$ for all t then in the case of an exponential potential and for the choice $\eta = 1/M\sqrt{2\ln N/n}$,

$$\max_{1\leqslant i\leqslant N} R_{i,n} \leqslant M\sqrt{2n\ln N} \ .$$

Observe that the value of the parameter η requires the knowledge of the number of rounds n. We remark here that similar bounds hold if, instead of the exponential potential function, the polynomial potential

$$\Phi(\mathbf{u}) = \sum_{i=1}^{N} (u_i)_+^p$$

is used (with $p = 2\ln N$), see Remark 3.2 and [**CeLu03**].

3.3. Finite classes of departure functions. As a first step assume that the set $\mathcal{F}_k$ of allowed departures for player k is finite, with cardinality m_k. For any $s \in S$ and departure $\psi_k \in \mathcal{F}_k$, denote by

$$r^k_{\psi_k}(s) = h^k(s^{-k}, \psi_k(s^k)) - h^k(s^{-k}, s^k)$$

the associated instantaneous internal regret, and by

$$\boldsymbol{r}^k(s) = \left(r^k_{\psi_k}(s)\right)_{\psi_k\in\mathcal{F}_k}$$

the regret vector formed by considering all departures. For a given sequence $s_1, \ldots, s_n \in S$ of plays, the cumulative internal regrets are given by the vector

$$\boldsymbol{R}^k(s_1^n) = \sum_{t=1}^{n} \boldsymbol{r}^k(s_t) \ ,$$

where s_1^n denotes the sequence $(s_1, \ldots, s_n)$.

Consider the following algorithm for player k. For $t = 1, 2, \ldots$, at round t, player k chooses any $s_t^k \in S^k$ such that

$$s_t^k = \sum_{\psi_k\in\mathcal{F}_k} \Delta^k_{\psi_k,t-1}\psi_k(s_t^k) \ , \tag{8.5}$$

with

$$\Delta^k_{\psi_k,t-1} = \frac{\phi\left(R^k_{\psi_k}(s_1^{t-1})\right)}{\sum_{g\in\mathcal{F}_k}\phi\left(R^k_g(s_1^{t-1})\right)} \ , \quad t \geqslant 2$$

where $\phi(x) = \exp(\eta x)$. For $t = 1$ we set $\Delta^k_{\psi_k,0} = 1/m_k$. (The parameter η will be tuned by the analysis below.)

Thus, each player is assumed to choose his action by solving the fixed–point equation (8.5). The existence of such a fixed point (under the assumptions of Theorem 8.7) follows easily by the

Schauder–Cauty fixed–point theorem [**Cau01**], which we recall below. (This is currently the most general version of Schauder's orignal theorem.)

THEOREM 8.6 (Schauder–Cauty fixed–point theorem). *Let C be a non–empty convex and compact subset of a topological Hausdorff vector space. Then each continuous map $T : C \to C$ has a fixed point.*

Note that if several fixed points of (8.5) exist, then the player is free to choose any of them.

THEOREM 8.7. *Assume that S^k is a convex and compact subset of a topological Hausdorff vector space, and that the payoff function h^k is bounded over S by $M_k \in \mathbb{R}$ and is concave in the k–th strategy. Then, whenever $\mathcal{F}_k$ is a finite subset of $\mathcal{C}(S^k)$ with cardinality m_k, the above algorithm guarantees that the cumulative internal regret satisfies*

$$\max_{\psi_k \in \mathcal{F}_k} \mathcal{R}^k_{\psi_k,n} \leqslant M_k \sqrt{2n \ln m_k} \, ,$$

if the exponential potential is used with $\eta = 1/M_k\sqrt{2 \ln m_k / n}$.

REMARK 8.6. *(Rates of convergence.)* The above theorem implies that if all $\mathcal{F}_k$ are finite, and all players play according to the above procedure, then, at round n, the empirical distribution is a $(\mathcal{F}_k)_{1 \leqslant k \leqslant N}$–correlated ε_n–equilibrium, with ε_n of the order $1/\sqrt{n}$.

REMARK 8.7. *(About the practical computation of the fixed-points.)* Note that an approximate solution of (8.5) is sufficient for our purposes. Provided that S^k is included in a normed vector space and h^k is a Lipschitz function, a simple modification of the proof of Cesa–Bianchi and Lugosi [**CeLu03**, Theorem 1] shows that the internal regret would still be $o(n)$ had we used a strategy s^k_t such that

$$\left\| s^k_t - \sum_{\psi_k \in \mathcal{F}_k} \Delta^k_{\psi_k,t-1} \psi_k(s^k_t) \right\| \leqslant \varepsilon_n \, ,$$

where ε_n decreases quickly enough to 0. In particular, when the S^k are included in finite–dimensional vector spaces, an algorithm partitioning S^k into a thin grid is able to find a suitable approximate fixed–point.

PROOF (OF THEOREM 8.7). The statement follows easily by Theorem 8.3. It suffices to prove that our choice of s^k_t satisfies the Blackwell condition

$$\nabla \Phi(\boldsymbol{R}^k(s^{t-1}_1)) \cdot \boldsymbol{r}^k(s_t) \leqslant 0$$

or equivalently

$$\sum_{\psi_k \in \mathcal{F}_k} \Delta^k_{\psi_k,t-1} h^k(s^{-k}_t, \psi_k(s^k_t)) \leqslant h^k(s^{-k}_t, s^k_t) \, ,$$

which is implied by the equality

$$h^k\left(s^{-k}_t, \sum_{\psi_k \in \mathcal{F}_k} \Delta^k_{\psi_k,t-1} \psi_k(s^k_t)\right) = h^k(s^{-k}_t, s^k_t)$$

and by the concavity of h^k in its k–th argument. This equality ensured by the choice (8.5). □

3.4. Countably infinite classes of departure functions. The next step is to extend the result of the previous section to countably infinite classes of departure functions. In this section we design an internal–regret minimizing procedure in the case when the set of allowed departures for player k is countably infinite. Denote by

$$\mathcal{F}_k = \{\psi_{k,q}, \ q \in \mathbb{N}\}$$

the set of departure functions of player k.

THEOREM 8.8. *Assume that S^k is a convex and compact subset of a topological Hausdorff vector space and that the payoff function h^k is bounded by M_k and is concave in the k–th strategy. If $\mathcal{F}_k$ is a countable subset of $\mathcal{C}(S^k)$, there exists a procedure such that for all $q \in \mathbb{N}$ and n,*

$$\mathcal{R}^k_{\psi_{k,q},n} \leqslant M_k \left(2(\ln q)^2 + 4.2n^{3/4}\right) .$$

Consequently, this procedure suffers no internal regret.

PROOF. We use a standard doubling trick (see Section 2.2 in Chapter 2) to extend the procedure of Theorem 8.7. Time is divided into blocks of increasing lengths such that the t–th block is $[\![2^{t-1}, 2^t - 1]\!]$. At the beginning of the t–th block, the algorithm for player k takes a fresh start and uses the method presented in Section 3.3, with the departures indexed by the integers between 1 and m_t and with $\eta = \eta_t$ tuned as

$$\eta_t = \frac{1}{M_k}\sqrt{2\frac{\ln m_t}{2^{t-1}}} .$$

We take, for instance, $m_t = \lfloor \exp\sqrt{2^t} \rfloor$.

Denote $\bar{n} = 2^{\lfloor \log_2 n \rfloor + 1}$. Define

$$H^k(s_1^n) = \sum_{t=1}^{n} h^k(s_t^{-k}, s_t^k) \quad \text{and} \quad H^k_{\psi_{k,q}}(s_1^n) = \sum_{t=1}^{n} h^k(s_t^{-k}, \psi_{k,q}(s_t^k)) .$$

Now, Theorem 8.7 ensures that

$$\begin{aligned} H^k(s_1^n) &= \sum_{t=1}^{\lfloor \log_2 n \rfloor} H^k(s^{2^t-1}_{2^{t-1}}) + H^k(s^n_{\bar{n}/2}) \\ &\geqslant \sum_{t=1}^{\lfloor \log_2 n \rfloor} \left(\max_{1 \leqslant q \leqslant m_t} H^k_{\psi_{k,q}}(s^{2^t-1}_{2^{t-1}}) - M_k\sqrt{2^t \ln m_t} \right) \\ &\quad + \left(\max_{1 \leqslant q \leqslant m_{\lfloor \log_2 n \rfloor + 1}} H^k_{\psi_{k,q}}(s^n_{\bar{n}/2}) - M_k\sqrt{\bar{n} \ln m_{\lfloor \log_2 n \rfloor + 1}} \right) . \end{aligned}$$

The departure function $\psi_{k,q}$ is considered from the time segment indexed by t_q, where t_q is the smallest integer such that $q \leqslant m_{t_q}$, that is, $2^{t_q - 1} < (\ln q)^2 \leqslant 2^{t_q}$. Observe that the total length of the previous time segments is $2^{t_q} - 1 \leqslant 2(\ln q)^2$. Thus, we obtain, for any $q \in \mathbb{N}$,

$$\begin{aligned} H^k(s_1^n) &\geqslant H^k_{\psi_{k,q}}(s_1^n) - M_k \left(2(\ln q)^2 + \sum_{t=1}^{\lfloor \log_2 n \rfloor + 1} \sqrt{2^t \ln(\exp\sqrt{2^t})} \right) \\ &\geqslant H^k_{\psi_{k,q}}(s_1^n) - M_k \left(2(\ln q)^2 + \sum_{t=1}^{\lfloor \log_2 n \rfloor + 1} \left(2^{3/4}\right)^t \right) \\ &\geqslant H^k_{\psi_{k,q}}(s_1^n) - M_k \left(2(\ln q)^2 + \frac{2^{3/2}}{2^{3/4} - 1} n^{3/4} \right) , \end{aligned}$$

which concludes the proof. □

REMARK 8.8. Theorem 8.8 does not provide any uniform bound for the internal regrets (where uniformity is understood with respect to the elements of the class of allowed departures $\mathcal{F}_k$), contrary to the case of finitely many departure functions of Theorem 8.7 (see Remark 8.6). In fact, in general, no non–trivial rate of convergence can be given for the convergence of the empirical distribution of plays to the set of $(\mathcal{F}_k)_{1 \leqslant k \leqslant n}$–correlated equilibria. However, in some special

cases, rates of convergence may be established, for instance for linear departures (see the end of this chapter) or for totally bounded classes of departures. In the latter case, the rates depend on the size of the classes, see Section 3.8. This means that the choice of the departure classes may be an important issue in practical situations.

3.5. Separable sets of departure functions. The extension to separable sets of departure function is now quite straightforward. Recall that compact or totally bounded spaces are special cases of separable spaces so the next result covers quite general situations.

THEOREM 8.9. *Assume that all strategy spaces S^k are convex and compact subsets of normed vector spaces. Let the payoff functions h^k be continuous over S and concave in the k–th strategy and assume that the $\mathcal{F}_k$ are separable subsets of $\mathcal{C}(S^k)$ (equipped with the supremum norm). Then there exist regret minimizing strategies such that, if every player follows such a strategy, then joint convergence of the sequence of empirical plays to the set of $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibria is achieved.*

The proof is based on the following lemma that can be shown by a simple dominated–convergence argument.

LEMMA 8.1. *Assume that the h^k are continuous, and let $(\mathcal{G}_k)$, $1 \leqslant k \leqslant N$, be classes of departure functions. Let π be a $(\mathcal{G}_k)_{1\leqslant k\leqslant N}$–correlated equilibrium. If for every k, $\mathcal{F}_k$ denotes the set of functions that may be obtained as π–almost sure limits of elements from $\mathcal{G}_k$, then π is a $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibrium.*

PROOF (OF THEOREM 8.9). For each player k, consider a countable dense subset $\mathcal{G}_k$ of $\mathcal{F}_k$ and apply the algorithm given in the proof of Theorem 8.8. Then Theorems 8.8 and 8.5 show that the empirical distribution of plays converges to the set of $(\mathcal{G}_k)_{1\leqslant k\leqslant N}$–correlated equilibria. By Lemma 8.1 the set of $(\mathcal{G}_k)_{1\leqslant k\leqslant N}$–correlated equilibria coincides with the set of $(\mathcal{F}_k)_{1\leqslant k\leqslant N}$–correlated equilibria. □

3.6. Proof of Theorem 8.3. To prove Theorem 8.3, we need two intermediate results. The first establishes separability needed to apply Theorem 8.9.

LEMMA 8.2. *If X is a convex and compact subset of a normed vector space, then the set $\mathcal{C}(X)$ of the continuous functions $X \to X$ is separable (for the supremum norm).*

The proof is an extension of Hirsch and Lacombe [**HiLa97**, Proposition 1.1]. Second we need a characterization of correlated equilibria with respect to all measurable departures. The proofs of both results are postponed to the end of this chapter.

LEMMA 8.3. *Assume that the strategy spaces S^k are convex and compact subsets of a normed vector space and that the h^k are continuous functions over S. Then the set of correlated equilibria with respect to all continuous departures $(\mathcal{C}(S^k))_{1\leqslant k\leqslant N}$ equals the set of correlated equilibria with respect to all measurable departures $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$.*

PROOF OF THE MAIN THEOREM. By the separability property stated in Lemma 8.2, Theorem 8.9 applies and gives an algorithm leading to convergence to the set of $(\mathcal{C}(S^k))_{1\leqslant k\leqslant N}$–correlated equilibria. In view of Lemma 8.3, this is equivalent to convergence to the set of $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibria, thus concluding the proof. □

3.7. Proof of Theorem 8.4. Sections 3.4, 3.5 and 3.6 only rely on the results of Section 3.3, and therefore it suffices to extend the results of Section 3.3 to the case of non–concave payoffs.

We assume that the strategy sets S^k are convex and compact subsets of normed vector spaces, and the payoff functions h^k are continuous over S^k. The players are allowed to randomize (which they do independently of each other). More, precisely, player k chooses his action s_t^k at round t according to the probability distribution $\mu_t^k \in \mathcal{P}(S^k)$, where $\mathcal{P}(S^k)$ denotes the set of probability distributions over S^k. We also assume that the departure class $\mathcal{F}_k$ is a finite subset of $\mathcal{C}(S^k)$, with cardinality m_k.

For any $\mu^k \in \mathcal{P}(S^k)$, $s^{-k} \in S^{-k}$ and any departure $\psi_k \in \mathcal{F}_k$, we denote

$$h^k(s^{-k}, \mu^k) = \int_{S^k} h^k(s^{-k}, s^k) \, \mathrm{d}\mu^k(s^k) \, ,$$

and by $(\mu^k)^{\psi_k}$ the image measure of μ^k by ψ_k, which means, in particular, that

$$h^k\left(s^{-k}, (\mu^k)^{\psi_k}\right) = \int_{S^k} h^k\left(s^{-k}, \psi_k(s^k)\right) \mathrm{d}\mu^k(s^k) \, .$$

Below we design a procedure for player k such that for all possible sequences of opponents' plays, $s_1^{-k}, s_2^{-k}, \ldots$,

$$\sum_{t=1}^{n} \left(h^k\left(s_t^{-k}, (\mu_t^k)^{\psi_k}\right) - h^k\left(s_t^{-k}, \mu_t^k\right)\right) = o(n) \, . \tag{8.6}$$

Then, thanks to the boundedness of the payoff function h^k, we may use a simple martingale convergence result such as the Hoeffding–Azuma inequality [**Azu67, Hoe63**], as well as Borel–Cantelli lemma, to show that (8.6) implies, almost surely (with respect to the auxiliary randomizations used),

$$\sum_{t=1}^{n} \left(h^k\left(s_t^{-k}, \psi_k(s_t^k)\right) - h^k\left(s_t^{-k}, s_t^k\right)\right) = o(n) \quad \text{a.s.} \, .$$

The latter is enough to apply Theorem 8.5, and prove the desired almost sure convergence.

It thus only remains to see how to design a procedure for player k guaranteeing (8.6). The techniques of Section 3.3 extend easily to this case. For any $\mu^k \in \mathcal{P}(S^k)$, $s^{-k} \in S^{-k}$ and any departure $\psi_k \in \mathcal{F}_k$, denote by

$$r_{\psi_k}^k(s^{-k}, \mu^k) = h^k\left(s^{-k}, (\mu^k)^{\psi_k}\right) - h^k(s^{-k}, \mu^k)$$

the associated instantaneous internal regret, and by

$$\boldsymbol{r}^k(s^{-k}, \mu^k) = \left(r_{\psi_k}^k(s^{-k}, \mu^k)\right)_{\psi_k \in \mathcal{F}_k}$$

the regret vector formed by considering all departures. For a given sequence $s_1^{-k}, \ldots, s_n^{-k}$ of opponents' plays, and the sequence of probability distributions $\mu_1^k, \ldots, \mu_n^k$, the cumulative internal regrets are given by the vector

$$\boldsymbol{R}^k\left((s^{-k})_1^n, (\mu^k)_1^n\right) = \sum_{t=1}^{n} \boldsymbol{r}^k(s_t^{-k}, \mu_t^k) \, ,$$

where $(s^{-k})_1^n$ denotes the sequence $(s_1^{-k}, \ldots, s_n^{-k})$, and $(\mu^k)_1^n$ is $(\mu_1^k, \ldots, \mu_n^k)$. Now, assume that player k can select his distribution μ_t^k at time t as a solution $\mu \in \mathcal{P}(S^k)$ of the equation

$$\mu = \sum_{\psi_k \in \mathcal{F}_k} \Delta_{\psi_k, t-1}^k \, \mu^{\psi_k} \, , \tag{8.7}$$

where

$$\Delta^k_{\psi_k,t-1} = \frac{\phi\left(R^k_{\psi_k}((s^{-k})_1^{t-1},(\mu^k)_1^{t-1})\right)}{\sum_{g\in\mathcal{F}_k}\phi\left(R^k_g((s^{-k})_1^{t-1},(\mu^k)_1^{t-1}))\right)}\,,\quad t\geqslant 2\,,$$

with $\phi(x)=\exp(\eta x)$, $\Delta^k_{\psi_k,0}=1/m_k$, and the parameter η is tuned as in Section 3.3. Under this assumption, we may obtain an upper bound of the order of $\sqrt{n}$ on the right–hand side of (8.6), by mimicking the argument of the proof of Theorem 8.7.

But the existence of such a distribution μ^k_t follows by the Schauder–Cauty fixed–point theorem. Recall that the weak–$*$ topology put on $\mathcal{P}(S^k)$ is such that, for all $\psi\in\mathcal{C}(S^k)$, the map that assigns the element μ^ψ to $\mu\in\mathcal{P}(s^k)$ is continuous. Thus, on the right–hand side of (8.7), we have a continuous function of μ. The existence of μ^k_t follows by the application of the claimed fixed–point theorem to the convex and compact subset $\mathcal{P}(S^k)$ of the vector space of all Borel, finite, real–valued and regular measures over S^k, equipped with its weak–$*$ topology.

3.8. A note on rates of convergence. Up to this point we have only focused on asymptotic statements and have not payed attention to rates of convergence. In particular, in Sections 3.4 and 3.5, we did not consider the way the elements of the countable classes were ordered, and we set up some parameters quite arbitrarily. However, under some assumptions, precise non–asymptotic bounds may be derived for the internal regret. (See also the first section of the appendix below.)

Recall that in the case of finite classes of departure functions, the internal regret can be made of the order of $n^{1/2}$. For richer classes of departure functions this may become larger, depending on the richness of the class. In this short remark we point out this phenomenon by considering totally bounded classes of departures.

Here we assume that the strategy sets S^k are convex and compact subsets of normed vector spaces, that the payoff functions h^k are Lipschitz functions concave in the k-th strategy, and that all classes of departures $\mathcal{F}_k$ are totally bounded sets under the corresponding supremum norms.

Recall that a metric space X is said to be *totally bounded* if for all $\varepsilon>0$, there exists a finite cover of X by balls of radius ε. For a given ε, the minimal number of such balls is called the ε–covering number of X, and is denoted by $N(\varepsilon)$. Any cover of X of size $N(\varepsilon)$ will be referred to as an ε–cover of X.

Denote by $N_k(\varepsilon)$ the ε–covering number of $\mathcal{F}_k$ and let δ_k be a Lipschitz constant of h^k, and M_k an upper bound for $|h^k|$. For any $\alpha>0$, introduce[4]

$$\varepsilon_k(\alpha)=\inf\left\{\varepsilon:\alpha\delta_k^2\varepsilon^2\geqslant 4M_k^2\ln N_k(\varepsilon)\right\}\,.$$

Clearly, $\varepsilon_k(\alpha)$ is decreasing. Moreover, $\varepsilon_k(\alpha)$ tends to 0 as $\alpha\to\infty$.

To obtain a bound on the cumulative regret with respect to a totally bounded class of departure functions, we use the doubling trick similarly to Section 3.4. Time is divided again in segments such that the r–th segment ($r\geqslant 1$) corresponds to the time instances t between 2^{r-1} and 2^r-1. In the r–th segment, the procedure for player k is the one of Section 3.3, with a departure class given by the centers of the balls which form an $(\varepsilon_k(2^r)+2^{-r})$–cover of $\mathcal{F}^k$. Denoting $\varepsilon'_k(2^r)=\varepsilon_k(2^r)+2^{-r}$, this implies, using the uniform continuity of h^k that for all sequences of opponents' plays, $s_1^{-k},s_2^{-k},\ldots$, and for all departure functions $\psi_k\in\mathcal{F}_k$,

$$H^k(s_1^n)\geqslant H^k_{\psi_k}(s_1^n)-\sum_{r=1}^{\lfloor\log_2 n\rfloor+1}\left(M_k\sqrt{2^r\ln N_k\left(\varepsilon'_k(2^r)\right)}+2^{r-1}\delta_k\,\varepsilon'_k(2^r)\right)\,,$$

[4]Note that $\varepsilon_k(\alpha)$ is defined as the infimum of an interval, so that all $\varepsilon>\varepsilon_k(\alpha)$ satisfy the defining condition.

thus proving that

$$\max_{\psi_k \in \mathcal{F}_k} \mathcal{R}^k_{\psi_k,n} \leqslant \delta_k \sum_{r=1}^{\lfloor \log_2 n \rfloor + 1} 2^r \varepsilon'_k(2^r) = \delta_k \left(1 + \log_2 n + \sum_{r=1}^{\lfloor \log_2 n \rfloor + 1} 2^r \varepsilon_k(2^r) \right) . \tag{8.8}$$

Observe that the quantity on the right–hand side is always $o(n)$ by an application of Cesaro's lemma and the fact that $\varepsilon_k(2^r) \to 0$ as $r \to \infty$.

EXAMPLE 8.2. As a concrete example, consider the case when the strategy set of player k is the d–dimensional cube $S^k = [0,1]^d$ and the class $\mathcal{F}_k$ of departures is the class of all Lipschitz functions $[0,1]^d \to [0,1]^d$ with Lipschitz constant less than a given value L_k. It is equipped with the metric associated to the supremum norm. Kolmogorov and Tihomirov [**KoTi61**, Theorem XIV] show that the metric entropy $\log N_k(\varepsilon)$ of this class of functions is of the order of ε^{-d}, that is[5], $\log N_k(\varepsilon) = \Theta(\varepsilon^{-d})$. It follows that $\varepsilon_k(\alpha) = \Theta(\alpha^{-1/(d+2)})$, and (8.8) implies that

$$\max_{\psi_k \in \mathcal{F}_k} \mathcal{R}^k_{\psi_k,n} \leqslant c\, n^{\frac{d+1}{d+2}}$$

for a constant c (depending only on δ_k, M_k, and L_k).

Other examples of totally bounded classes of departure functions, with the indication of the orders of magnitude of their metric entropies $\log N_k(\varepsilon)$, may be found in Kolmogorov and Tihomirov [**KoTi61**, Sections 5–9], see also Devroye, Györfi, and Lugosi [**DeGyLu96**, Section 28.2].

4. A link with correlated equilibrium of finite games

In this final section we assume that Γ is a finite game, with strategy sets given by finite sets S^k. Assume that the players play in the mixed extension, that is, at round t, each player k chooses privately a probability distribution p^k_t over S^k, all probability distributions $p_t = (p^1_t, \dots, p^N_t)$ are made public, and player k gets the payoff $h^k(p_t)$, where we still denote by h^k the linear extension of h^k,

$$h^k(p_t) = \sum_{s \in S} \left(\prod_{j=1}^N p^j_t(s^j) \right) h^k(s) .$$

The results of the previous sections show that the players can ensure that the empirical frequencies of play in the mixed extension,

$$\mu_n = \frac{1}{n} \sum_{t=1}^n \delta_{(p^1_t, \dots, p^N_t)} ,$$

converge to some set of correlated equilibria of the mixed extension of Γ, for instance, the set $E_{\mathcal{L}^0}$ of correlated equilibria with respect to all measurable departures, or the set E_{L} of correlated equilibria with respect to all linear departures. The convergence to $E_{\mathcal{L}^0}$ may be achieved by Theorem 8.3, whereas the convergence to E_{L} is given by Theorem 8.7, since the set of all linear mappings $\mathcal{P}(S^k) \to \mathcal{P}(S^k)$ is the convex hull of a finite number of mappings.

Recall that this is done by minimizing the internal regrets, that is, by ensuring that for all players k and all $\phi_k \in \mathcal{F}_k$,

$$\sum_{t=1}^n h^k(p^{-k}_t, \phi_k(p^k_t)) - \sum_{t=1}^n h^k(p^{-k}_t, p^k_t) = o(n) , \tag{8.9}$$

[5]The notation $x_\varepsilon = \Theta(y_\varepsilon)$ means that the ratio $x_\varepsilon / y_\varepsilon$ is bounded above and below by positive numbers as ε tends to 0.

where $\mathcal{F}_k$ is either a countable dense subset of the set of all continuous functions $\mathcal{P}(S^k) \to \mathcal{P}(S^k)$, or the finite set of the mappings generating all linear functions $\mathcal{P}(S^k) \to \mathcal{P}(S^k)$.

We are actually interested in playing in the original finite game, and to do so[6], we assume that at each round $t = 1, 2, \ldots$, each player k draws finally an action $s_t^k \in S^k$ according to p_t^k. We denote by $(\widehat{\pi}_n)_{n\in\mathbb{N}^*}$ the sequence of joint empirical frequencies of play,

$$\widehat{\pi}_n = \frac{1}{n}\sum_{t=1}^{n} \delta_{(s_t^1,\ldots,s_t^N)} \ ,$$

and study its convergence properties. One may wonder whether it may converge almost surely to a set strictly smaller than the set E_Γ of correlated equilibria of the finite game Γ.

Here we point out that the results of this chapter do not imply convergence of the empirical frequencies to a set smaller than the set of correlated equilibria of the finite game. More precisely, we show that the set of correlated equilibria of the mixed extension and that of the original finite game are the same in a natural sense.

For any distribution μ over $\mathcal{P}(S^1) \times \cdots \times \mathcal{P}(S^N)$, denote by $\pi = \psi(\mu)$ the distribution over $S^1 \times \cdots \times S^N$ defined, for all $i^k \in S^k$, by

$$\pi(i^1, \ldots, i^N) = \int_{\mathcal{P}(S^1)\times\cdots\times\mathcal{P}(S^N)} p^1(i^1)\ldots p^N(i^N)\mathrm{d}\mu(p^1, \ldots, p^N) \ . \tag{8.10}$$

By this definition and by considering the linear extension of h^k, we have that $\mathbb{E}_{\psi(\mu)}h^k = \mathbb{E}_\mu h^k$ for all k and μ. (The definition of ψ indicates that we get back to the original finite game by taking averages.)

Denote $\pi_n = \psi(\mu_n)$ and note that $\|\pi_n - \widehat{\pi}_n\| \ \to \ 0$ by martingale convergence. $(\widehat{\pi}_n)_{n\in\mathbb{N}^*}$ and $(\pi_n)_{n\in\mathbb{N}^*}$ have therefore the same convergence properties. But since ψ is continuous, the $\pi_n = \psi(\mu_n)$ converge to the set $\psi(E_\mathcal{M})$, and therefore, so do the $\widehat{\pi}_n$.

Remark 8.4 alludes to the inclusion $\psi(E_{\mathcal{L}^0}) \subseteq E_\Gamma$. Below we show that, in fact, $\psi(E_{\mathcal{L}^0}) = \psi(E_\mathrm{L}) = E_\Gamma$. In this sense, the sets of correlated equilibria of the mixed extension and of the original finite game are the same. There are not fewer correlated equilibria in the mixed extension, and therefore, one cannot hope tighter convergence results by minimizing the internal regret in the mixed extension of the game.

LEMMA 8.4. $\psi(E^0_\mathcal{L}) = \psi(E_\mathrm{L}) = E_\Gamma$.

PROOF. The equality between E_Γ and $\psi(E_\mathrm{L})$ is immediate, by linearity and in view of (8.10).

We now prove that each correlated equilibrium π of Γ may be written as $\psi(\mu)$, where $\mu \in E^0_\mathcal{L}$, that is, μ is a probability distribution over $\mathcal{P}(S^1) \times \cdots \times \mathcal{P}(S^N)$ that is a correlated equilibrium with respect to all measurable departures.

For a given correlated equilibrium $\pi \in E_\Gamma$, we choose

$$\mu = \sum_{i^1,\ldots,i^N} \pi(i^1, \ldots, i^N)\, \delta_{(\delta_{i^1},\ldots,\delta_{i^N})} \ ,$$

[6]Note that by martingale convergence, (8.9) is ensured almost surely whenever for all players k and all $\phi_k \in \mathcal{F}_k$, $\sum_{t=1}^n h^k(s_t^{-k}, \phi_k(p_t^k)) - \sum_{t=1}^n h^k(s_t^{-k}, p_t^k) = o(n)$. This can be done in the finite game by using the fixed–point techniques of Section 3.4, in the sense that it can be achieved in the game where only the chosen action profiles $(s_t^1, \ldots, s_t^N)$ (and not the probability distributions $(p_t^1, \ldots, p_t^N)$) are made public.

where δ_{i^j} is the probability distribution over S^j that puts probability mass 1 on i^j. We have to prove that for all players k, for all measurable departures ϕ_k,

$$\int_{\mathcal{P}(S^1)\times\cdots\times\mathcal{P}(S^N)} h^k(p^{-k},p^k)\mathrm{d}\mu(p^1,\ldots,p^N) \geqslant \int_{\mathcal{P}(S^1)\times\cdots\times\mathcal{P}(S^N)} h^k(p^{-k},\phi_k(p^k))\mathrm{d}\mu(p^1,\ldots,p^N)\ .$$

In view of the form of μ, only the values p^k of the form δ_{i_k} where $i_k \in S^k$ matter in the above integrals. Define a linear mapping L_k from $\mathcal{P}(S^k)$ to $\mathcal{P}(S^k)$ by $L_k(\delta_{i^k}) = \phi_k(\delta_{i^k})$, for all $i^k \in S^k$. Then,

$$\int_{\mathcal{P}(S^1)\times\cdots\times\mathcal{P}(S^N)} h^k(p^{-k},\phi_k(p^k))\mathrm{d}\mu(p^1,\ldots,p^N) = \int_{\mathcal{P}(S^1)\times\cdots\times\mathcal{P}(S^N)} h^k(p^{-k},L_k(p^k))\mathrm{d}\mu(p^1,\ldots,p^N)\ .$$

This concludes the proof in view of the first equality noted above. □

REMARK 8.9. Though we do not get convergence to a smaller set of equilibria by playing in the mixed extension, it is worth noting that we may however get a deterministic convergence to correlated equilibria of the original game by doing so. The left-hand side of (8.9) is of the order of $\sqrt{n}$ when the sets of departures are given by the linear departures, so that the players achieve with π_n a $1/\sqrt{n}$–correlated equilibrium in a deterministic way after n rounds of play (instead of simply achieving a $\sqrt{\ln(1/\delta)}/\sqrt{n}$–correlated equilibrium with high probability $1-\delta$, thanks to $\widehat{\pi}_n$). A careful implementation of procedure in the mixed extension can be shown to have a computational complexity not higher than its usual randomized counterpart. It simply requires that players show the ones to the others the probability distribution they choose.

We conclude this section by pointing out that the minimization proposed by (8.9) is, using the terminology of Greenwald and Jafari [**GrJa03**], a matter of Φ–no regret, with Φ including all (extremal) linear functions as well as many other continuous maps. This solves the first half of the question posed in the conclusion of [**GrJa03**]. The second part of the question is to determine if, by performing the regret minimization (8.9), one could achieve convergence to tighter solution concepts than simply the set of all correlated equilibria. We showed strong evidence that this is not so.

5. Discussion and perspectives

5.1. Bandit strategies. In game theory, games with bandit prediction settings are often referred to as unknown games, since the players ignore the game they take part in, they do not even need to know that they are playing a game. For finite games, internal regret can be minimized in a bandit setting (in expectation, see Section 1.4 of Chapter 3, or with overwhelming probability, see Section 6 of Chapter 6). Consequently, all players of an unknown game may minimize their internal regrets simultaneously, and achieve joint convergence of the empirical frequencies of plays to the set of correlated equilibria. (See also a related procedure in [**HaMa02**].) This, however, is not easy to extend to general infinite games, even with concave payoffs.

OPEN QUESTION 8.1. Find internal-regret-minimizing strategies for a large class of infinite games, in a bandit setting.

Actually, for infinite games, we do not even know in general how to minimize the external regret with respect to all constant actions in a bandit setting. The problem is to get good estimators

for the unobserved payoffs, which correspond to all the actions but the played one. For infinite games, there are usually an uncountable number of them, and the probability distributions the players use are generally non-discrete (they may charge subsets, not only points), and we cannot simply get estimators by dividing the observed payoffs by the probability densities (when the latter exist).

5.2. Convergence to Nash equilibria. In this chapter, we get convergence to the set of correlated equilibria, and prove no convergence result to the the set of Nash equilibria. It turns out that for *finite* games, such results have been obtained recently, see Foster and Young [**FoYo03**], Germano and Lugosi [**GeLu04**], Hart and Mas-Colell [**HaMa04**]. We describe below these results, by pointing out their main limits. There is, above all, a concern about the convergence rates.

Note that we introduced strategies that only needed to keep track of the regrets to achieve convergence to correlated equilibria, so these strategies did not need to have a long memory. The story is different however when studying convergence to Nash equilibria. In this setting, Hart and Mas-Colell [**HaMa04**] are mainly concerned with bounded memory assumptions, where a strategy is said to have a memory bounded by R if the action it prescribes at each round may not rely on more than the R past played strategy profiles. They show on the negative side that for any integer R and any $\varepsilon > 0$ sufficiently small there does not exist a strategy with memory bounded by R ensuring that for all games, the sequence $(s_t)_{t\geqslant 1}$ of the played strategy profiles converges to the set of ε-Nash equilibria. On the positive side, for all $\varepsilon > 0$, they can find an integer R and a general randomized strategy, with memory bounded by R, ensuring that for all games the sequence $(\pi_n)_{n\geqslant 1}$ of the empirical distributions of plays converges to an ε-Nash equilibrium, provided that all players use the general (uncoupled) strategy.

Germano and Lugosi [**GeLu04**] do not restrict their attention to bounded memory strategies, and construct strategies using the entire past. They exhibit a general randomized strategy (and some variants of it), such that, provided that it is used by all players,

- for all games, the sequence $(\pi_n)_{n\geqslant 1}$ of the empirical distributions of plays converges almost surely to the set of ε-Nash equilibria,
- for all games, the sequence $(\mu_n)_{n\geqslant 1}$ of the played mixed profiles converges almost surely to an ε-Nash equilibrium,
- for Lebesgue-almost all games, the sequence $(\mu_n)_{n\geqslant 1}$ of the played mixed profiles converges almost surely to a Nash equilibrium.

The first drawback of these strategies is their very slow rate of convergence, see [**GeLu04**] for a discussion. The other concern is that these strategies rely on (Markov) random searches, and that they try, in some sense, all possible mixed strategy profiles. They find an approximate Nash equilibrium only by chance, and then stick to it for some time.

OPEN QUESTION 8.2. The above-mentioned results of convergence to Nash equilibria are for finite games. We believe that this chapter introduced all the mathematical techniques needed to extend them to the case of infinite games, similarly to what we did for Foster and Vohra's [**FoVo97, FoVo99**] convergence results in Section 3.

Appendix: Computable procedures for convergence to linear correlated equilibria

The general procedure described in Section 3 needs to compute, at each iteration, a fixed point, see (8.5). Whereas the existence of the latter is ensured by Schauder's theorem, there is in general no effective way to compute it. A first solution is provided by computing only an approximate fixed point, see Remark 8.7, but the resulting strategy has still a prohibitive computational complexity (exponential in n and N).

The aim of this appendix is to design algorithms, suffering no internal regret with respect to the set of all linear departures, which are first, easy to implement (with a complexity linear in n and N^2), and second, whose convergence rate to the set of linear correlated equilibria can be made precise, in the sense of Remark 8.6. We first consider the case of games with strategy sets given by simplexes, and then, in the special case of the square loss, deal with more general strategy sets.

Games with strategy sets given by simplexes. The following is a simple generalization of the results of Chapter 7 to general concave payoffs. (In Chapter 7 we actually considered convex losses, but they correspond to concave payoffs.)

We first describe the class of games we consider. We assume, in a first time, that the strategy sets are all given by simplexes (of possibly different orders) $S^k = \mathcal{X}_{d_k}$, where $d_k \in \mathbb{N}^*$ and for every $d \in \mathbb{N}^*$, $\mathcal{X}_d$ denotes the simplex of $\mathbb{R}^d$. For each player k, the set of allowed departures $\mathcal{F}_k$ is formed by the linear maps $\mathcal{X}_{d_k} \to \mathcal{X}_{d_k}$. (It may be seen easily that all $\mathcal{F}_k$ are non-empty. Actually, for $d \in \mathbb{N}^*$, each linear map $\mathcal{X}_d \to \mathcal{X}_d$ may be represented by a row-stochastic matrix, see Section 5.2 in Chapter 7.) We refer to $(\mathcal{F}_k)_{1 \leqslant k \leqslant N}$-correlated equilibria as the *linear correlated equilibria* of the game.

The strategy of player k proceeds from a simple prior linearization of the instantaneous internal regrets. Assume that for all fixed opponent plays s^{-k}, the payoff function h^k is differentiable as a function $t^k \in \mathcal{X}_{d_k} \mapsto h^k(s^{-k}, t^k)$ of the k-th variable, with a gradient at $s^k \in \mathcal{X}_{d_k}$ denoted by $g^k(s) = \nabla_k h^k(s^{-k}, s^k)$. At round t, once all the players have output their strategies $s_t \in S$, consider the instantaneous regret vector $\boldsymbol{r}^k(s_t)$ given by, for $1 \leqslant i, j \leqslant d_k$, $i \neq j$:

$$r^k_{(i,j)}(s_t) = s^k_{i,t} \left(g^k_j(s_t) - g^k_i(s_t) \right) ,$$

where $g^k_i(s_t)$ (respectively $s^k_{i,t}$) denotes the i-th component of $g^k(s_t)$ (respectively s^k_t).

For a given history $s_1, \ldots, s_n$, the cumulative internal regrets are given by the vector

$$\boldsymbol{R}^k(s_1^n) = \sum_{t=1}^{n} \boldsymbol{r}^k(s_t) ,$$

where s_1^n simply denotes the sequence of actions $(s_1, \ldots, s_n)$. (The proof below shows indeed that these regrets are linear upper bounds on the original regrets.)

Denote by $L^k_{(i,j)}$ the linear function $\mathcal{X}_{d_k} \to \mathcal{X}_{d_k}$ that maps an element $s^k \in \mathcal{X}_{d_k}$ to $u^k \in \mathcal{X}_{d_k}$ given by $u^k_i = 0$, $u^k_j = s^k_i + s^k_j$, and $u^k_m = s^k_m$ if $m \neq i$, $m \neq j$.

Now, for $t = 1, 2, \ldots$, if the sequence of played profiles is given by s_1^{t-1}, then player k chooses, at round t, an element $s^k_t \in \mathcal{X}_{d_k}$ such that the fixed point equality

$$s^k_t = \sum_{i \neq j} \Delta^k_{(i,j),t-1} L^k_{(i,j)}(s^k_t) \tag{8.11}$$

holds, where the summation is over all pairs (i, j), $i \neq j$, and where for $t \geqslant 2$,

$$\Delta^k_{(i,j),t-1} = \frac{\phi\left(R^k_{(i,j)}(s_1^{t-1})\right)}{\sum_{l \neq m} \phi\left(R^k_{(l,m)}(s_1^{t-1})\right)} ,$$

with $\phi(x) = (x)_+^{p-1}$, and for $t = 1$, the $\Delta^k_{(i,j),0}$ are taken equal to $1/(d_k(d_k-1))$. (The parameter p will be tuned by the analysis below.) The existence and an efficient method to compute such a fixed point s^k_t are detailed by Lemma 3.1.

THEOREM 8.10. *Assume that S^k is $\mathcal{X}_{d_k}$, the simplex of order $d_k \in \mathbb{N}^*$. Provided that for all fixed opponents' plays $s^{-k} \in S^{-k}$, the payoff function $s^k \in S^k \mapsto h^k(s^{-k}, s^k)$ is concave and differentiable, with a gradient uniformly bounded in norm by M_k, the above algorithm suffers no internal regret, with the uniform bound*

$$\max_{L_k \in \mathcal{L}_k} \mathcal{R}^k_{L_k,n} \leqslant d_k M_k \sqrt{(4 \ln d_k - 1)\, en}\ ,$$

where $\mathcal{L}_k$ denotes the class of all linear maps $\mathcal{X}_{d_k} \to \mathcal{X}_{d_k}$ and p is chosen as $p = 4 \ln d_k$.

COROLLARY 8.2. *If the assumptions of Theorem 8.10 are satisfied for all players k, and if all of them minimize their linear internal regrets with the general procedure described above, then at round n, the empirical distribution of played profiles π_n defines an ε_n-linear correlated equilibrium, where*

$$\varepsilon_n = \left(\max_{1 \leqslant k \leqslant N} d_k M_k \sqrt{4 \ln d_k - 1} \right) \frac{\sqrt{e}}{n}\ .$$

PROOF. The proof is a simple generalization of the proof techniques used in Section 4.1 of Chapter 7. By the assumption of concavity and differentiability, we have, for all $s \in S$ and all $L_k \in \mathcal{L}_k$, the linear upper bounding

$$h^k\left(s^{-k}, L_k(s^k)\right) - h^k(s^{-k}, s^k) \leqslant \nabla_k h^k(s^{-k}, s^k) \cdot \left(L_k(s^k) - s^k\right)\ , \tag{8.12}$$

where $\cdot$ is the standard inner product in $\mathbb{R}^{d_k}$.

By the representation of linear maps from simplexes into themselves by row-stochastic matrices, it may be seen that for all $\mathbf{u} \in \mathbb{R}^{d_k}$, $s^k \in S^k$ and $L_k \in \mathcal{L}_k$,

$$\mathbf{u} \cdot \left(L_k(s^k) - s^k\right) \leqslant d_k \max_{i \neq j} s^k_i\, (u_j - u_i)\ ,$$

thus leading (with the above notation for the gradient) to

$$h^k\left(s^{-k}, L_k(s^k)\right) - h^k(s^{-k}, s^k) \leqslant d_k \max_{i \neq j} s^k_i \left(g^k_j(s) - g^k_i(s)\right)\ .$$

We thus have shown that the algorithm for player k satisfies

$$\max_{L_k \in \mathcal{L}_k} \mathcal{R}^k_{L_k,n} \leqslant d_k \max_{s_1^{-k}, \ldots, s_n^{-k}} \max_{i \neq j} R^k_{(i,j)}(s_1^n)\ .$$

The argument is concluded by noting that the definition of the algorithm, Theorem 3.1, as well as the boundedness assumption on the gradient function, show that the maximum on the right-hand side is upper bounded by

$$M_k \sqrt{(p-1) d_k^{4/p} n} = M_k \sqrt{(4 \ln d_k - 1)\, en}$$

for the proposed choice $p = 4 \ln d_k$. □

REMARK 8.10. Label-efficient settings (see Chapter 5) may also be considered in games where all players have concave payoff functions. Due to the linearization of the regrets, it is easy to extend the procedures of Chapter 5, and to still get convergence to correlated equilibria, thanks to Corollary 5.1 and Remark 5.3.

EXAMPLE 8.3. *(Penalizing the volatility in on-line investment.)* Consider the setting and the notation of Chapter 7. The *volatility* V of an investment strategy is defined as follows. Assume that

on the sequence of market vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the investment strategy was given by the sequence of portfolios $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_n$ over the N assets. Then, the volatility V equals the variance of the log-wealth ratios,

$$V = \left(\frac{1}{n} \sum_{t=1}^{n} \left(\ln \left(\boldsymbol{P}_t \cdot \boldsymbol{x}_t \right) \right)^2 \right) - \left(\frac{1}{n} \sum_{t=1}^{n} \ln \left(\boldsymbol{P}_t \cdot \boldsymbol{x}_t \right) \right)^2 .$$

V is invariant under rescalings of the market vectors, and so is the worst-case logarithmic wealth ratio (see the beginning of Section 6.1 in Chapter 7). Under a boundedness assumption over the market, we may renormalize the market vectors (only for the definition of the investment algorithm), such that they all lie in, say, $[1/e, e]^N$. We upper bound V by the first sum appearing in its definition, and consider the penalized log-returns given by $h(\boldsymbol{P}_1, \boldsymbol{x}_1) + \ldots + h(\boldsymbol{P}_n, \boldsymbol{x}_n)$, where

$$h(\boldsymbol{P}, \boldsymbol{x}) = \ln \left(\boldsymbol{P} \cdot \boldsymbol{x} \right) - \alpha \left(\ln \left(\boldsymbol{P} \cdot \boldsymbol{x} \right) \right)^2 .$$

For any (fixed) renormalized market vector $\boldsymbol{x} \in [1/e, e]^N$, the function $h(\cdot, \boldsymbol{x})$ is concave, as a sum of two concave functions, and has bounded gradient. Consequently, the above procedures apply, and we may minimize internal and external regrets with respect to this new loss (or payoff) function, similarly to what we did in Chapter 7. Doing so, we are not only interested in the returns, but we want to trade off between high returns and low volatility. This should result in stable investment strategies, with large returns.

REMARK 8.11. The results of Section 5 of Chapter 7 may be extended as well, to the so-called exp-concave payoff functions, see Kivinen and Warmuth [**KiWa99**]: a payoff function h^k is exp-concave (in its k-th variable) if there exists a constant $\eta > 0$ such that $\exp(\eta h^k)$ is concave (in its k-th variable). For such payoff functions, the (fast) rates of convergence to correlated equilibria are of the order of $(\ln N)/n$. However, the resulting procedures suffer from the same computational drawbacks as in Remark 7.8.

More general parametric strategy sets, and application to on-line linear regression. We indicate now how convergence to linear correlated equilibria may be obtained in games with more general strategy sets than simplexes. These strategy sets are parametric, that is, they are included in finite-dimensional vector spaces.

Whenever the parametric strategy set of player k contains a simplex and his payoff function is concave, he may restrict this set to the smaller strategy set given by the included simplex. Then, the results of the previous section indicate him a possible no-linear-internal-regret strategy. This strategy minimizes in particular the external regret with respect to the class of constant strategies indexed by this simplex.

This restriction is not necessary in some special cases, like in on-line regression, and more satisfactory procedures can be designed. This is the point of the present section.

EXAMPLE 8.4. *(On-line linear or polynomial regression.)* On-line linear regression is a prediction game that corresponds to a repeated zero-sum game between a forecaster and an environment. It is played as follows (see, among many others, [**KiWa97, Ces99**]). At round t, the environment chooses an input variable $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{d,t}) \in \mathbb{R}^d$, and the forecaster is asked to form a prediction of the form $\hat{y}_t = \boldsymbol{u}_t \cdot \boldsymbol{x}_t$, where $\boldsymbol{u}_t = (u_{1,t}, \ldots, u_{d,t}) \in \mathbb{R}^d$. The environment then reveals the true outcome $y_t \in \mathbb{R}$. The loss of the forecaster is measured by $\ell(\boldsymbol{u}, (\boldsymbol{x}, y)) = (\boldsymbol{u} \cdot \boldsymbol{x} - y)^2$, or, alternatively, its payoff is given by $h(\boldsymbol{u}, (\boldsymbol{x}, y)) = -(\boldsymbol{u} \cdot \boldsymbol{x} - y)^2$.

We note that on-line m–polynomial regression, $m \in \mathbb{N}^*$, is a straightforward extension of on-line linear regression. It corresponds to predictions of the form

$$\hat{y}_t = \sum_{p=0}^{m} \sum_{i=1}^{d} u_{i,(p,t)} \, x_{i,t}^p \ ,$$

where the forecaster outputs $m+1$ weight vectors $u_{(p,t)} \in \mathbb{R}^d$, $p = 0, \ldots, m$. This more general problem can be encompassed in the previous one, by considering as an input variable at round t the $(m+1)d$–dimensional vector $(\mathbf{1}, \boldsymbol{x}_t, \boldsymbol{x}_t^2, \ldots, \boldsymbol{x}_t^p)$, where we used obvious notation. This is why we concentrate below on linear regression.

In on-line linear regression, the aim of the forecaster is to minimize his external regret,

$$\sum_{t=1}^{n} \left(\boldsymbol{u}_t \cdot \boldsymbol{x}_t - y_t \right)^2 - \min_{\boldsymbol{v} \in \mathcal{V}} \sum_{t=1}^{n} \left(\boldsymbol{v} \cdot \boldsymbol{x}_t - y_t \right)^2 \ ,$$

where $\mathcal{V}$ is an (often strict) subset of $\mathbb{R}^d$, and his linear internal regret,

$$\sum_{t=1}^{n} \left(\boldsymbol{u}_t \cdot \boldsymbol{x}_t - y_t \right)^2 - \min_{L \in \mathcal{L}} \sum_{t=1}^{n} \left(L\left(\boldsymbol{u}_t\right) \cdot \boldsymbol{x}_t - y_t \right)^2 \ ,$$

where $\mathcal{L}$ denotes the set of all linear functions $\mathbb{R}^d \to \mathbb{R}^d$, or a large class of such linear functions. (Note that in this section, we state the results in terms of losses, rather than payoffs.) The results of the previous section (including those of Remark 8.11, with $\eta = 1/2$) apply, and yield a prediction scheme for the forecaster such that both the external regret with respect to $\mathcal{V} = \mathcal{X}_d$ and the linear internal regret are minimized. This scheme only outputs weight vectors from the simplex.

However, in on-line linear regression, the notion of external regret is usually meaningful only if it is defined with respect to a larger class that also contains vectors with non-positive components. The classes of interest are typically of the form

$$\mathcal{V}_U = \left\{ \boldsymbol{u} \in \mathbb{R}^d : \|u\| \leqslant U \right\} \ ,$$

where $\|\cdot\|$ is some norm on $\mathbb{R}^d$, for instance, the ℓ^1, Euclidian, or supremum norms. We concentrate on the ℓ^1 norm, defined by

$$\|\boldsymbol{u}\| = \sum_{i=1}^{d} |u_i| \ ,$$

and study the forecaster introduced in Figure 1.

THEOREM 8.11. *We assume that the input values and the outcomes all satisfy $|y_t| \leqslant M$, $|x_{i,t}| \leqslant M$, for some $M > 0$ and for all i and t. The forecaster of Figure 1 outputs weight vectors in $\mathcal{V}_U$, and suffers both no linear internal regret and no external regret with respect to $\mathcal{V}_U$, with the upper bounds*

$$\sum_{t=1}^{n} \left(\boldsymbol{u}_t \cdot \boldsymbol{x}_t - y_t \right)^2 - \min_{\boldsymbol{v} \in \mathcal{V}_U} \sum_{t=1}^{n} \left(\boldsymbol{v} \cdot \boldsymbol{x}_t - y_t \right)^2 \quad \leqslant \quad 4\, dUM^2 \sqrt{(4 \ln(2d) - 1)en} \ ,$$

$$\sum_{t=1}^{n} \left(\boldsymbol{u}_t \cdot \boldsymbol{x}_t - y_t \right)^2 - \min_{L \in \mathcal{L}} \sum_{t=1}^{n} \left(L\left(\boldsymbol{u}_t\right) \cdot \boldsymbol{x}_t - y_t \right)^2 \quad \leqslant \quad 4\, dUM^2 \sqrt{(4 \ln(2d) - 1)en} \ ,$$

where $\mathcal{L}$ denotes the set of all linear mappings $\mathbb{R}^d \to \mathbb{R}^d$ satisfying $L(\mathcal{X}_d) \subseteq \mathcal{X}_d$.

Parameters: $U > 0$, a bound on the weight vectors.
For each round $t = 1, 2, \ldots$,

(1) get the input value $\boldsymbol{x}_t$, and denote
$$\boldsymbol{x}'_t = (Ux_{1,t}, \ldots, Ux_{d,t}, -Ux_{1,t}, \ldots, -Ux_{d,t}) \in \mathbb{R}^{2d} ;$$
(2) at round t, denote, for $s \leqslant t-1$ and $i = 1, \ldots, N$,
$$\ell_{i,s} = 2(\boldsymbol{u}'_s \cdot \boldsymbol{x}'_s - y_s)\, x'_{i,s} ,$$
and get the probability distribution $\boldsymbol{u}'_t \in \mathcal{X}_{2d}$ chosen by the forecaster of Section 1.2 of Chapter 3, with a polynomial potential of order $p = 4\ln(2d)$;
(3) predict with $\boldsymbol{u}_t$ given by $u_{i,t} = U(u'_{i,t} - u'_{i+d,t}),\ i = 1, \ldots, d$.

FIGURE 1. A forecasting scheme for on-line linear regression.

We used polynomial reweightings in Figure 1 simply to avoid tuning issues. Of course, in view of the standard adaptive techniques, like Lemma 4.3, exponential potentials may be used as well.

The forecaster of Figure 1 relies on a trick introduced by Kivinen and Warmuth [**KiWa97**]. They faced the same situation as here, that is, they first introduced a forecasting scheme suffering no external regret with respect to all weight vectors of the simplex, and then extended it thanks to this trick to a forecasting scheme suffering no external regret with respect to $\mathcal{V}_U$. (The bound U has to be known beforehand by the forecaster.)

This trick consists in noting that, with the notation of Figure 1,

$$\boldsymbol{u}_t \cdot \boldsymbol{x}_t = \boldsymbol{u}'_t \cdot \boldsymbol{x}'_t , \tag{8.13}$$

where $\boldsymbol{u}_t \in \mathcal{V}_U$, whereas $\boldsymbol{u}'_t$ is a probability distribution. Thus, it is enough to compute a weight vector lying in the simplex, which we know how to do. For the sake of readability, we indicated the values of the gradient of the losses (these are the $\ell_{i,s}$) and referred for the sub-algorithm to Chapter 3. That is, we exactly apply the procedure of the previous section and simply write explicitly the gradients here.

PROOF. In view of step (2) in the definition of the forecaster, by Theorems 3.1 or 8.10, and by the linear upper bounding (8.12) (see also (7.2)),

$$\sum_{t=1}^{n} \left(\boldsymbol{u}'_t \cdot \boldsymbol{x}'_t - y_t\right)^2 - \min_{L' \in \mathcal{L}'} \sum_{t=1}^{n} \left(L'\left(\boldsymbol{u}'_t\right) \cdot \boldsymbol{x}'_t - y_t\right)^2 \leqslant 4\, dUM^2 \sqrt{(4\ln(2d) - 1)en} , \tag{8.14}$$

where $\mathcal{L}'$ is the set of all linear functions $\mathcal{X}_{2d} \to \mathcal{X}_{2d}$, since the gradients $2(\boldsymbol{u}'_s \cdot \boldsymbol{x}'_s - y_s)\, x'_{i,s}$ are bounded by $4UM^2$. Now, to any linear mapping $L : \mathbb{R}^d \to \mathbb{R}^d$ preserving the simplex, we may associate a linear function $L' : \mathcal{X}_{2d} \to \mathcal{X}_{2d}$ such that, with the notation of Figure 1, $L(\boldsymbol{u}_t) \cdot \boldsymbol{x}_t = L'(\boldsymbol{u}'_t) \cdot \boldsymbol{x}'_t$ for all t. This function is defined, with obvious notation, by $L'(\boldsymbol{u}_1, \boldsymbol{u}_2) = (L(\boldsymbol{u}_1), L(\boldsymbol{u}_2))$. The proof is therefore concluded for the internal regret.

The bound for the external regret is almost a special case of the one for the internal regret. Let $\boldsymbol{v}$ be any element of $\mathcal{V}_U$. Provided that we can find $\boldsymbol{v}' \in \mathcal{X}_{2d}$ such that $\boldsymbol{v} \cdot \boldsymbol{x}_t = \boldsymbol{v}' \cdot \boldsymbol{x}'_t$ for all t, the first bound of the theorem follows from (8.14), used with the constant, thus linear, function $L' \equiv \boldsymbol{v}'$. But [**KiWa97**] explains precisely how to do this. We denote respectively by v^+ and v^-

the nonnegative and non-positive parts of any real number v, and simply let

$$\boldsymbol{v}' = \frac{1}{U}\left((v_1^+, \ldots, v_d^+, v_1^+, \ldots, v_d^+) + \frac{U - \|\boldsymbol{v}\|}{2d}(1, \ldots, 1)\right) .$$

□

REMARK 8.12. We note here that, up to the multiplicative factor d, we recovered the same orders of magnitude in all parameters for the bound on external regret as [**KiWa97**], see also [**Ces99**]. (But the bound on internal regret is new.) In both papers, the bounds are derived by using a more general method, based on Bregman divergences, which in this case reduce to Kullback-Leibler divergences. Interestingly enough, this was the same for the first derivation of the EG strategy in [**HeScSiWa98**]. In Chapter 7 we showed how a simple linear upper bound and the usual techniques used prediction with expert advice yield instead a simple analysis of the EG strategy, and even lead to no internal regret algorithms of the same flavour. This is exactly what we did for on-line linear regression with the square loss in this section.

Appendix: Technical proofs

Proof of Theorem 8.2. For $s = (s^1, \ldots, s^N) \in S$, we write $\|s\|_\infty = \max_{i=1,\ldots,N} \|s^i\|$ where $\|s^i\|$ is the norm of s^i in S^i.

First, we prove the direct implication. Fix π, a $(\mathcal{L}^0(S^k))_{1\leqslant k\leqslant N}$–correlated equilibrium of the continuous game Γ. Choose any number $\varepsilon > 0$. It suffices to show that there exists a r_0 such that for every discretization of size $r < r_0$, π induces a 2ε–correlated equilibrium.

Each h^k is uniformly continuous, so we may choose r_0 such that for all $k \leqslant N$, $s, t \in S$, $\|s - t\|_\infty \leqslant r_0$ implies $\left| h^k(s) - h^k(t) \right| \leqslant \varepsilon$. Fix a discretization $(\mathcal{P}, \mathcal{D})$ of size r less than r_0.

Fix a player k and a departure $g_k : \mathcal{D}^k \to \mathcal{D}^k$. We need to prove that

$$\sum_{t\in\mathcal{D}} \pi_d(t) h^k(t) + 2\varepsilon \geqslant \sum_{t\in\mathcal{D}} \pi_d(t) h^k(t^{-k}, g_k(t^k)) . \tag{8.15}$$

Define $\psi_k : S \to S$ by $\psi_k(s^k) = g_k(t_j^k)$ for all $s^k \in V_j^k$. ψ_k is a measurable function. Now, for all $s \in V_{i_1}^1 \times \ldots \times V_{i_N}^N$,

$$\left\| (s^{-k}, \psi_k(s^k)) - (t_{i_1}^1, \ldots, g_k(t_{i_k}^k), \ldots, t_{i_N}^N) \right\|_\infty \leqslant r .$$

Therefore, due to the uniform continuity of the h^k, we have for all k,

$$\left| \int_S h^k(s^{-k}, \psi_k(s^k)) \mathrm{d}\pi(s) - \sum_{t\in\mathcal{D}} \pi_d(t) h^k(t^{-k}, g_k(t^k)) \right| \leqslant \varepsilon .$$

It is even easier to see that

$$\left| \int_S h^k(s^{-k}, s^k) \mathrm{d}\pi(s) - \sum_{t\in\mathcal{D}} \pi_d(t) h^k(t^{-k}, t^k) \right| \leqslant \varepsilon .$$

Now, as π is a correlated equilibrium of the original game,

$$\int_S h^k(s^{-k}, s^k) \mathrm{d}\pi(s) \geqslant \int_S h^k(s^{-k}, \psi_k(s^k)) \mathrm{d}\pi(s) .$$

Combining the last three inequalities leads to (8.15), thus concluding the direct part.

The converse implication is proved in a similar way. First, note that thanks to Lemma 8.3, we can restrict our attention to continuous departures. Assume that there exists a function ε with $\lim_{r\to 0}\varepsilon(r) = 0$ such that for all discretizations $(\mathcal{P}, \mathcal{D})$ of size r, π induces an $\varepsilon(r)$–correlated equilibrium.

Fix an arbitrary $\eta > 0$. We show that for all k and all continuous functions $\psi_k : S^k \to S^k$,

$$\int_S h^k(s^{-k}, s^k)\mathrm{d}\pi(s) + \eta \geqslant \int_S h^k(s^{-k}, \psi_k(s^k))\mathrm{d}\pi(s)\ . \tag{8.16}$$

(The conclusion will follow by letting η decrease to 0.)

Fix a player k and a continuous departure ψ_k. As h^k is uniformly continuous, we can choose $\delta > 0$ such that for all $s, t \in S$, $\|s - t\|_\infty \leqslant \delta$ implies $\left| h^k(s) - h^k(t) \right| \leqslant \eta/3$. Now, ψ_k is also uniformly continuous, so that there exists $\delta' > 0$ such that for all $s, t \in S$, $\|s - t\|_\infty \leqslant \delta'$ implies $|\psi_k(s) - \psi_k(t)| \leqslant \delta/2$. Finally, take $r_0 > 0$ sufficiently small so that for all $r \leqslant r_0$, $\varepsilon(r) \leqslant \eta/3$. We consider $r = \min(r_0, \delta, \delta')$.

There exists a finite cover of each S^j by open balls of radius r, denoted by $B(x_i^j, r)$, $1 \leqslant j \leqslant N$, $1 \leqslant i \leqslant N_j$. Each open cover is converted into a measurable partition in the following way. For $1 \leqslant j \leqslant N$, $1 \leqslant i \leqslant N_j$,

$$V_i^j = B(x_i^j, r) \backslash \left(\cup_{m=1}^{i-1} B(x_m^j, r) \right)\ .$$

We take the grid given by the centers, that is, with the above notation, $t_i^j = x_i^j$, $1 \leqslant j \leqslant N$, $1 \leqslant i \leqslant N_j$. We thus have obtained a discretization of size less than r, and denote by π_d the probability measure induced by π.

We define $g_k : \mathcal{D}^k \to \mathcal{D}^k$ as follows. For $1 \leqslant j \leqslant N_k$, $g_k(x_j^k) = x_m^k$ where $1 \leqslant m \leqslant N_k$ is the index such that $\psi_k(x_j^k) \in V_m^k$. Note that in particular, $\left\| g_k(x_j^k) - \psi_k(x_j^k) \right\| \leqslant r \leqslant \delta/2$.

But if $s^k \in V_{i_k}^k$, $\left\| s^k - x_{i_k}^k \right\| \leqslant r \leqslant \delta'$, so that $\left\| \psi_k(s^k) - \psi_k(x_{i_k}^k) \right\| \leqslant \delta/2$. Finally, $\left\| \psi_k(s^k) - g_k(x_{i_k}^k) \right\| \leqslant \delta$. Thus, if $s \in V_{i_1}^1 \times \ldots \times V_{i_N}^N$,

$$\left\| (s^{-k}, \psi_k(s^k)) - (x_{i_1}^1, \ldots, g_k(x_{i_k}^k), \ldots, x_{i_N}^N) \right\|_\infty \leqslant \delta\ .$$

Therefore, by uniform continuity of h^k,

$$\left| \int_S h^k(s^{-k}, \psi_k(s^k))\mathrm{d}\pi(s) - \sum_{x\in\mathcal{D}} \pi_d(x) h^k(x^{-k}, g_k(x^k)) \right| \leqslant \frac{\eta}{3}\ .$$

Again, it is even easier to see that

$$\left| \int_S h^k(s^{-k}, s^k)\mathrm{d}\pi(s) - \sum_{x\in\mathcal{D}} \pi_d(x) h^k(x^{-k}, x^k) \right| \leqslant \frac{\eta}{3}\ .$$

Since π is an $\varepsilon(r)$–correlated equilibrium (and since $\varepsilon(r) \leqslant \eta/3$), it is true that

$$\sum_{x\in\mathcal{D}} \pi_d(x) h^k(x) + \frac{\eta}{3} \geqslant \sum_{x\in\mathcal{D}} \pi_d(x) h^k(x^{-k}, g_k(x^k))\ .$$

Combining these last three inequalities, we get (8.16), concluding the proof.

Proof of Lemma 8.2. Hirsch and Lacombe [**HiLa97**] consider the set of continuous functions defined on a compact metric set X into $\mathbb{R}$ and show that this set is separable (Proposition 1.1). But it turns out that this proof easily extends to the case of Lemma 8.2, giving, in addition, an example of a dense countable subset of $\mathcal{C}(X)$. We simply need the following well–known lemma, see for instance Rudin [**Rud74**].

LEMMA 8.5 (Partition of unity). *If X is a locally compact Hausdorff space, then, given a finite number of open sets $V_1, \dots, V_N$ and a compact $K \subset \cup_{i=1,\dots,N} V_i$, there exist N nonnegative continuous functions $h_1, \dots, h_N$ summing to 1 over K, such that h_i vanishes outside V_i.*

PROOF (OF LEMMA 8.2). As X is a compact set, for a given $n \in \mathbb{N}^*$, there exist finitely many x_n^j, $j = 1, \dots, N_n$, such that the collection of open balls of common radius $1/n$ and centered in these x_n^j forms a finite cover of X,

$$X = \cup_{j=1}^{N_n} B(x_n^j, 1/n) \ .$$

We denote the set formed by these x_n^j by X_n. By Lemma 8.5 (with $K = X$), denote by ϕ_j^n, $j = 1, \dots, N_n$, a partition of unity constructed over this open cover of X. We denote by A_n the set formed by

$$A_n = \left\{ \sum_{j=1}^{N_n} y_n^j \phi_j^n \ , \ (y_n^j)_{j=1,\dots,N_n} \in (X_n)^{N_n} \right\} \ .$$

A_n is a finite set. By convexity of X, each element of A_n maps X into X. By continuity of the ϕ_j^n, A_n is finally seen as a subset of $\mathcal{C}(X)$.

We consider the countable subset A formed by the union of the A_n, $A = \cup_{n \in \mathbb{N}^*} A_n$, and claim that A is dense in $\mathcal{C}(X)$. To see this, fix a continuous function $f \in \mathcal{C}(X)$. As f maps the compact metric space X into itself, f is uniformly continuous over X. Fix $\varepsilon > 0$ and choose $\delta > 0$ small enough to ensure that $\|x - y\| < \delta$ implies $\|f(x) - f(y)\| < \varepsilon$, where $\|\cdot\|$ denotes the norm of the underlying normed space that contains X. Now, fix a sufficiently large integer n such that $1/n < \min(\delta, \varepsilon)$. For every $j = 1, \dots, N_n$, choose y_n^j such that $\left\| y_n^j - f(x_n^j) \right\| \leqslant \varepsilon$. Introduce the functions

$$g = \sum_{j=1}^{N_n} f(x_n^j) \phi_j^n \ , \quad h = \sum_{j=1}^{N_n} y_n^j \phi_j^n \ .$$

It is clear that $h \in A$, and we prove that $\|f - h\|_\infty \leqslant 2\varepsilon$.

For a given $x \in X$,

$$\begin{aligned} \|f(x) - g(x)\| &= \left\| \sum_{j=1}^{N_n} \left(f(x) - f(x_n^j) \right) \phi_n^j(x) \right\| \\ &\leqslant \sum_{j=1}^{N_n} \left\| f(x) - f(x_n^j) \right\| \phi_n^j(x) \ . \end{aligned}$$

Now, $\left\| f(x) - f(x_n^j) \right\| \phi_n^j(x) \leqslant \varepsilon \phi_n^j(x)$, simply because ϕ_n^j vanishes outside $B(x_n^j, 1/n)$ (which is included in $B(x_n^j, \delta)$), whereas, thanks to uniform continuity, the norm of the difference $f(x) - f(x_n^j)$ is less than ε over this ball. Finally, recalling that the ϕ_j^n sum to 1, we get $\|f - g\|_\infty \leqslant \varepsilon$.

A similar argument, using the fact that for every j, $\left\| y_n^j - f(x_n^j) \right\| \leqslant \varepsilon$, indicates that $\|g - h\|_\infty \leqslant \varepsilon$, thus concluding the proof. □

Proof of Lemma 8.3. The proof is a combination of Lemma 8.1 and Corollary 8.3, which is derived from the following version of Lusin's theorem tailored for our needs.

PROPOSITION 8.4. *If X is a convex and compact subset of a normed space, equipped with a probability measure μ (defined over the Borel σ–algebra), then for every measurable function $f : X \to X$ and for every $\delta, \varepsilon > 0$, there exists a continuous function $g : X \to X$ such that*

$$\mu \{ \|f - g\| \geqslant \delta \} \leqslant \varepsilon \ .$$

PROOF. We use the notation (and the techniques) of the proof of Lemma 8.2. First note that μ is regular, since it is a finite measure over the Borel σ–algebra of a Polish space (compact metric spaces are Polish).

Fix n large enough such that $1/n < \delta$. Consider the N_n measurable sets

$$M_j^n = f^{-1}\left(B(x_j^n, 1/n)\right) \ .$$

By regularity of μ, one can find compact sets K_j^n and open sets V_j^n such that, for all j,

$$K_n^j \subset M_j^n \subset V_j^n \ , \quad \mu(V_j^n \backslash K_j^n) \leqslant \frac{\varepsilon}{N_n} \ .$$

By construction, the M_j^n form a cover of X. Therefore, the V_j^n form an open cover of X. By Lemma 8.5 (with $K = X$), fix a partition of unity based on this open over, which we denote by $\xi_1^n, \ldots, \xi_{N_n}^n$. Consider the continuous function g given by

$$g = \sum_{j=1}^{N_n} x_j^n \xi_j^n \ .$$

By convexity of X, g maps X into X. Now, as above, for all $x \in X$,

$$\|f(x) - g(x)\| \leqslant \sum_{j=1}^{N_n} \left\| f(x) - x_n^j \right\| \xi_n^j(x) \ .$$

By construction, $\left\| f(x) - x_n^j \right\| \xi_n^j(x) \leqslant \xi_n^j(x)/n$ provided that $x \in M_j^n \cup (V_j^n)^c$. Therefore, $\|f(x) - g(x)\| \leqslant 1/n < \delta$, except, possibly, on the measurable subset Δ defined by

$$\Delta = \cup_{j=1}^{N_n} V_j^n \backslash M_j^n \ ,$$

whose μ–measure is seen to be less than ε by subadditivity of the measure. □

Now, setting $\delta_n = \varepsilon_n = 1/2^n$, and using Borel–Cantelli lemma, one easily gets the following corollary.

COROLLARY 8.3. *If X is a convex and compact subset of a normed space, equipped with a probability measure μ (over the Borel σ–algebra), then every measurable function $f : X \to X$ may be obtained as a μ–almost sure limit of continuous functions $(g_n)_{n\in\mathbb{N}^*}$ mapping X into X.*

Part 4

Additional material and bibliography

APPENDIX A

Statistical background

Contents

1. Hoeffding-Azuma maximal inequality

We recall in this section the maximal version of the Hoeffding-Azuma inequality (see Hoeffding [**Hoe63**], Azuma [**Azu67**], see also McDiarmid [**McD89**]).

A sequence of random variables $X_1, X_2, \ldots$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$ if for every $t \geqslant 1$, X_t is $\mathcal{F}_{t-1}$-measurable and a.s.,

$$\mathbb{E}\left[X_t \,|\mathcal{F}_{t-1}\right] = 0 \,.$$

The following lemma is the key step in the proof of the Hoeffding-Azuma maximal inequality.

LEMMA A.1. *Let $(X_1, \ldots, X_n)$ be a martingale difference sequence with respect to the filtration $\mathcal{F} = (\mathcal{F}_t)_{0 \leqslant t \leqslant n}$, such that for all $t = 1, \ldots, n$, there exists a $\mathcal{F}_{t-1}$-measurable random variable V_t and a nonnegative constant c_t with $V_t \leqslant X_t \leqslant V_t + c_t$ a.s. Then, denoting by $M_n = X_1 + \ldots + X_n$ the associated martingale, for any $x > 0$,*

$$\log \mathbb{E}\left[e^{sM_n}\right] \leqslant \frac{s^2}{8} \sum_{t=1}^{n} c_t^2 \,.$$

Now, Doob's maximal inequality and simple algebra imply the following.

LEMMA A.2 (Hoeffding-Azuma maximal inequality). *Let $(X_1, \ldots, X_n)$ be a martingale difference sequence with respect to the filtration $\mathcal{F} = (\mathcal{F}_t)_{0 \leqslant t \leqslant n}$, such that for all $t = 1, \ldots, n$, there exists a $\mathcal{F}_{t-1}$-measurable random variable V_t and a nonnegative constant c_t with $V_t \leqslant X_t \leqslant V_t + c_t$ a.s. Denote by $(M_1, \ldots, M_n)$ the associated martingale, where $M_t = \sum_{s=1}^{t} X_s$ for all t. Then, for any $x > 0$,*

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > x\right] \leqslant \exp\left(-\frac{2\,x^2}{\sum_{t=1}^{n} c_t^2}\right) \,,$$

or, equivalently,

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > \sqrt{\frac{x}{2}\sum_{t=1}^{n} c_t^2}\right] \leqslant e^{-x} .$$

2. Bernstein's maximal inequality for martingales

The Hoeffding-Azuma inequality is in a sense not sharp enough for most of our purposes, for it does not involve the variances of the martingale differences, which turn out to be in practice far smaller than the simple sum of the squared conditional ranges. Bernstein's inequality fixes this, and offers a bound where the variances replace the squared ranges. The crux of the proof is the following classical inequality for random variables bounded from above, see e.g. [**Fre75**, Section 3]. The previous chapters showed that it was also of independent interest.

LEMMA A.3. *Let Z be any random variable, bounded from above by 1 and with nonpositive expectation,*

$$Z \leqslant 1 \ \textit{a.s.} \quad \textit{and} \quad \mathbb{E}[Z] \leqslant 0 .$$

Then, for all $\lambda \geqslant 0$,

$$\ln \mathbb{E}\left[e^{\lambda Z}\right] \leqslant \left(e^{\lambda} - 1 - \lambda\right) \operatorname{Var} Z .$$

Moreover, the factor $e^{\lambda} - 1 - \lambda$ in the inequality above is optimal.

We state now a version of Bernstein's inequality suited for maxima of martingale difference sequences (see, e.g. [**Fre75**] or [**DaDu83**]), and prove a corollary tailored to the needs of Section 4 of Chapter 5.

LEMMA A.4 (Bernstein's maximal inequality for martingales). *Let $(X_1, \ldots, X_n)$ be a martingale difference sequence with respect to the filtration $\mathcal{F} = (\mathcal{F}_t)_{0 \leqslant t \leqslant n}$ and with increments bounded by a constant $K > 0$: for all t, $|X_t| \leqslant K$ a.s. Consider the associated martingale $(M_1, \ldots, M_n)$, where $M_t = \sum_{s=1}^{t} X_s$ for all t. Denote the sum of the conditional variances by*

$$V_n = \sum_{t=1}^{n} \mathbb{E}\left[X_t^2 \,|\, \mathcal{F}_{t-1}\right] .$$

Then, for all $\lambda > 0$,

$$\left(\exp\left(\lambda M_n - \phi_K(\lambda) V_n\right)\right)_{n \geqslant 0}$$

is a supermartingale (with respect to the same filtration $\mathcal{F}$), where

$$\phi_K(\lambda) = \frac{1}{K^2}\left(e^{\lambda K} - 1 - \lambda K\right) .$$

In particular, for all constants $x, v > 0$,

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > x \text{ and } V_n \leqslant v\right] \leqslant \exp\left(-\frac{x^2}{2\left(v + Kx/3\right)}\right)$$

and therefore,

$$\mathbb{P}\left[\max_{t=1,\ldots,n} M_t > \sqrt{2vx} + (\sqrt{2}/3)Kx \ \text{ and } \ V_n \leqslant v\right] \leqslant e^{-x} .$$

Simple calculations yield the following corollary, in which the bounds involve directly the sum of the conditional variances rather than a constant upper bound v on it. We do not know whether this issue had already been considered, and in particular, whether the extra $\sqrt{\ln n}$ which appears below is necessary.

COROLLARY A.1. *Under the assumptions of Lemma A.4, for all $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\max_{t=1,\dots,n} M_t \leqslant \sqrt{2(V_n + K^2)\ln(n/\delta)} + (\sqrt{2}/3)K\ln(n/\delta)\ .$$

PROOF. Denote

$$M = \max_{t=1,\dots,n} M_t\ .$$

We apply the previous lemma n times and use a union-of-events bound. For $t = 1,\dots,n$,

$$\begin{aligned}
&\mathbb{P}\left[M > \sqrt{2(V_n + K^2)\ln(n/\delta)} + (\sqrt{2}/3)K\ln(n/\delta) \text{ and } V_n \in K^2\,[t-1,\,t]\right] \\
&\quad\leqslant\quad \mathbb{P}\left[M > \sqrt{2K^2 t\ln(n/\delta)} + (\sqrt{2}/3)K\ln(n/\delta) \text{ and } V_n \leqslant K^2 t\right] \\
&\quad\leqslant\quad \delta/n\ ,
\end{aligned}$$

where we used Lemma A.4 in the last step. By boundedness of the X_t, V_n lies between 0 and $K^2\,n$, and therefore a union-of-events bound over $t = 1,\dots,n$ concludes the proof. □

3. Some elements of information theory

We essentially deal with the Kullback-Leibler divergence (or relative entropy) in this section. A good reference is the monography by Cover and Thomas [**CoTh91**, Chapter 2]. The *Kullback-Leibler divergence* between two probability distributions $\mathbb{P}$ and $\mathbb{Q}$, with common dominating measure μ, and with densities $\mathrm{d}\mathbb{P} = p\ \mathrm{d}\mu$ and $\mathrm{d}\mathbb{Q} = q\,\mathrm{d}\mu$, equals

$$\mathcal{K}(\mathbb{P},\mathbb{Q}) = \int p\ln\frac{p}{q}\,\mathrm{d}\mu\ .$$

The definition does not depend on the choice of μ, so that the only case when $\mathcal{K}(\mathbb{P},\mathbb{Q})$ is not defined yet is when $\mathbb{P}$ is not absolutely continuous with respect to $\mathbb{Q}$. In this case, we simply let $\mathcal{K}(\mathbb{P},\mathbb{Q}) = +\infty$.

To illustrate the definition, we show simple upper bounds on the Kullback-Leibler divergences between two Bernoulli distributions $\mathbb{B}(p)$ and $\mathbb{B}(q)$, the one with parameter $p = 1/2$, and the other with parameter $q = 1/2 - \varepsilon$. (The dominating measure μ is for instance the sum of the Dirac measures at 0 and 1.)

LEMMA A.5. *For all $0 \leqslant \varepsilon \leqslant 1/\sqrt{6}$,*

$$\mathcal{K}\left(\mathbb{B}_{1/2},\mathbb{B}_{1/2-\varepsilon}\right) \leqslant 6\,\varepsilon^2\ .$$

For all $0 \leqslant \varepsilon \leqslant 1/10$,

$$\mathcal{K}\left(\mathbb{B}_{1/2-\varepsilon},\mathbb{B}_{1/2}\right) + \mathcal{K}\left(\mathbb{B}_{1/2},\mathbb{B}_{1/2-\varepsilon}\right) \leqslant 5\,\varepsilon^2\ .$$

PROOF. We simply use the definition of $\mathcal{K}$,

$$\begin{aligned}
\mathcal{K}\left(\mathbb{B}_{1/2},\mathbb{B}_{1/2-\varepsilon}\right) = \frac{1}{2}\left(\ln\frac{1}{1-2\varepsilon} + \ln\frac{1}{1+2\varepsilon}\right) \quad &= \quad \frac{1}{2}\left(\ln\frac{1}{1-4\varepsilon^2}\right) = \frac{1}{2}\ln\left(1 + \frac{4\varepsilon^2}{1-4\varepsilon^2}\right) \\
&\leqslant \quad \frac{2\varepsilon^2}{1-4\varepsilon^2}\ ,
\end{aligned}$$

where we used at the last step $\ln(1+u) \leqslant u$. The proof of the first inequality is now concluded by using $0 \leqslant \varepsilon \leqslant 1/\sqrt{6}$.

As for the second one, the same techniques yield

$$\begin{aligned} & \mathcal{K}\left(\mathbb{B}_{1/2}, \mathbb{B}_{1/2-\varepsilon}\right) + \mathcal{K}\left(\mathbb{B}_{1/2-\varepsilon}, \mathbb{B}_{1/2}\right) \\ = \quad & \left(-\frac{1}{2} + \left(\frac{1}{2} - \varepsilon\right)\right) \ln(1-2\varepsilon) + \left(-\frac{1}{2} + \left(\frac{1}{2} + \varepsilon\right)\right) \ln(1+2\varepsilon) \\ = \quad & \varepsilon \ln\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right) = \varepsilon \ln\left(1 + \frac{4\varepsilon}{1-2\varepsilon}\right) \leqslant \frac{4\varepsilon^2}{1-2\varepsilon} \,, \end{aligned}$$

and the proof is concluded by lower bounding the denominator thanks to $0 \leqslant \varepsilon \leqslant 1/10$. □

We now motivate the computation of the Kullback-Leibler divergence between two probability distributions by Pinsker's equality (see, e.g., [**Tsy04**]).

LEMMA A.6 (Pinsker's inequality). *For all measurable sets A,*

$$\mathbb{P}[A] - \mathbb{Q}[A] \leqslant \sqrt{\frac{1}{2}\mathcal{K}(\mathbb{P}, \mathbb{Q})} \,.$$

(The supremum over all measurable sets A in the left-hand side is called the variational distance between $\mathbb{P}$ and $\mathbb{Q}$.)

We end this section by indicating two useful ways of computing or upper bounding Kullback-Leibler divergences.

LEMMA A.7 (Convexity of $\mathcal{K}$). *The map $(\mathbb{P}, \mathbb{Q}) \mapsto \mathcal{K}(\mathbb{P}, \mathbb{Q})$ is a convex function (in the pair). Consequently, for any given random variable X, denoting by $\mathbb{P}^X$ and $\mathbb{Q}^X$ the laws of X under the distributions $\mathbb{P}$ and $\mathbb{Q}$, we have*

$$\mathcal{K}\left(\mathbb{P}^X, \mathbb{Q}^X\right) \leqslant \mathcal{K}(\mathbb{P}, \mathbb{Q}) \,.$$

The definition of the Kullback-Leibler divergence shows that when $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ and $\mathbb{Q} = \mathbb{Q}_1 \otimes \mathbb{Q}_2$ are given by product measures, then

$$\mathcal{K}\left(\mathbb{P}_1 \otimes \mathbb{P}_2,\, \mathbb{Q}_1 \otimes \mathbb{Q}_2\right) = \mathcal{K}\left(\mathbb{P}_1,\, \mathbb{Q}_1\right) + \mathcal{K}\left(\mathbb{P}_2,\, \mathbb{Q}_2\right) \,. \tag{A.1}$$

We extend this to general probability distributions over (discrete) product spaces as follows (by stating a "chain rule"). We consider probability distributions $\mathbb{P}$ and $\mathbb{Q}$ over a discrete product image set $A \times B$, and denote respectively by $p(a,b)$ and $q(a,b)$ their densities with respect to the counting measure over $A \times B$. We also denote by $\mathbb{P}_B$ the second marginal of $\mathbb{P}$, and by $p(\cdot \,|\, b)$ the conditional density of $\mathbb{P}$ given $\{b\}$ with respect to the counting measure over A, and use analogous notation for $\mathbb{Q}$. We define the *conditional Kullback-Leibler divergence* between $\mathbb{P}$ and $\mathbb{Q}$ as

$$\mathcal{K}\left(\mathbb{P}_{A\,|\,B},\, \mathbb{Q}_{A\,|\,B}\right) = \sum_{b \in B} p(b) \sum_{a \in A} p(a \,|\, b) \ln \frac{p(a \,|\, b)}{q(a \,|\, b)} \,.$$

LEMMA A.8 ("Chain rule"). *With the notation above,*

$$\mathcal{K}(\mathbb{P}, \mathbb{Q}) = \mathcal{K}(\mathbb{P}_B,\, \mathbb{Q}_B) + \mathcal{K}\left(\mathbb{P}_{A\,|\,B},\, \mathbb{Q}_{A\,|\,B}\right) \,.$$

4. On Fano's lemma

Fano's lemma often yields sharper lower bounds than Pinsker's inequality, with an additional $\ln N$ factor. We illustrated this general fact in Section 5 of Chapter 5 and in Remark 6.2. Let us briefly sketch a (high-level) picture of why this is so. To this end, we first state a possible version of Fano's lemma. This version is a corollary of an information theoretic result (see, e.g., Cover and Thomas [**CoTh91**, Chapter 2]).

LEMMA A.9 (Fano's lemma). *Consider a probability space Ω, equipped with N probability measures $\mathbb{P}_1, \ldots, \mathbb{P}_N$. For all partitions $A_1, \ldots, A_N$ of Ω,*

$$\frac{1}{N}\sum_{j=1}^{N} \mathbb{P}_j[A_j] \leqslant \frac{K + \ln 2}{\ln (N-1)}$$

where

$$K = \frac{1}{N}\sum_{j=1}^{N} \mathcal{K}\left(\mathbb{P}_j, \bar{\mathbb{P}}\right) \ , \quad \textit{with } \bar{\mathbb{P}} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{P}_j \ .$$

It turns out that in our lower bound proofs, the problem is quite symmetric, by construction. The sets A_j are of the form $\{I = j\}$, where I is the action taken by the forecaster, and the probability distribution $\mathbb{P}_j$ only favors action j. The corresponding average distribution $\bar{\mathbb{P}}$ is often the uniform distribution over the outcomes, so that $\bar{\mathbb{P}}[A_j] = 1/N$ for all j. Due to the symmetry of the problem and of the "good" forecasters, (these are usually invariant under a relabelling of the experts), all quantities $\mathcal{K}(\mathbb{P}_j, \bar{\mathbb{P}})$ are equal, with common value denoted by K, and the same holds even for the $\mathbb{P}_j[A_j]$. Thus, we have the following bounds, respectively by Pinsker's inequality (see Lemma A.6) and Fano's lemma,

$$\begin{aligned} \min_{j=1,\ldots,N} \mathbb{P}_j[A_j] &\leqslant \frac{1}{N} + \sqrt{\frac{1}{2}K} \qquad \text{(Pinsker's inequality)} \ , \\ \min_{j=1,\ldots,N} \mathbb{P}_j[A_j] &\leqslant \frac{K + \ln 2}{\ln (N-1)} \qquad \text{(Fano's lemma)} \ . \end{aligned}$$

As K is often small (of the order of ε^2, where ε is a small parameter), the second bound is an important improvement, at least from an asymptotical viewpoint, due to the extra $\ln (N-1)$ in the denominator. (Compare the proofs of Theorems 3.3, 5.5, and 6.3.)

Our goal is now to have Fano-like bounds which are interesting for moderate values of N. The bound proposed by Lemma A.9 is useless for $N \leqslant 3$, and more generally speaking, the extra $\ln 2$ factor is very unconvenient for non-asymptotic purposes. The solution is offered by a recent paper of Birgé [**Bir05**], and is presented in the following lemma. (The second half of the Lemma is actually stated in Massart [**Mas05**] and is an easy consequence of the proof given in [**Bir05**]. The interest is to get a clean $\ln N$ factor, instead of simply a $\ln (N-1)$.)

LEMMA A.10 (Birgé's version of Fano's lemma). *Consider a probability space Ω, equipped with N probability measures $\mathbb{P}_1, \ldots, \mathbb{P}_N$. For all partitions $A_1, \ldots, A_N$ of Ω,*

$$\min_{j=1,\ldots,N} \mathbb{P}_j[A_j] \leqslant \max\left\{\frac{e}{1+e}, \frac{\bar{K}}{\ln (N-1)}\right\} \ ,$$

where

$$\bar{K} = \frac{1}{N-1}\sum_{j=2}^{N} \mathcal{K}(\mathbb{P}_j, \mathbb{P}_1) \ .$$

Another valid upper bound is

$$\min_{j=1,\ldots,N} \mathbb{P}_j[A_j] \leqslant \max\left\{\frac{2e}{1+2e}, \frac{\bar{K}}{\ln N}\right\} \ ,$$

Now, the crucial point in the proof of Theorem 5.5 is an extension of Birgé's version of Fano's lemma to a convex combination of probability masses. This extension, stated in Lemma A.13 below is proved thanks to a straightforward modification of the proofs techniques used in Birgé [**Bir05**] (see also Massart [**Mas05**]). Below, we state and use their main lemmas. Recall first a consequence of the variational formula for entropy.

LEMMA A.11. *For arbitrary probability distributions $\mathbb{P}, \mathbb{Q}$ and for each $\lambda > 0$,*

$$\lambda\, \mathbb{P}[A] - \psi_{\mathbb{Q}[A]}(\lambda) \leqslant \mathcal{K}(\mathbb{P}, \mathbb{Q})$$

where $\psi_p(\lambda) = \ln\big(p\,(e^\lambda - 1) + 1\big)$.

We now need to know the behavior of the Cramer transform of this function ψ_p. This is indicated in the following lemma (see [**Mas05**, Section 2.3.4]).

LEMMA A.12. *For all $p > 0$, the Cramer transform ψ_p of ψ_p satisfies, at $p \leqslant a \leqslant 1$,*

$$\psi_p^*(a) = \sup_{\lambda \geqslant 0} \left(\lambda a - \psi_p(\lambda)\right) = a \ln\left(\frac{a}{p}\right) + (1-a)\ln\left(\frac{1-a}{1-p}\right) \geqslant a \ln\left(\frac{a}{e\,p}\right)\ .$$

Next we are ready to extend Lemma A.10 to the case of convex combinations of probability distributions. This extension does not follow from Birgé's lemma, but may be obtained by a simple modification of its proof, for the latter already deals with convex combinations.

LEMMA A.13 (Fano's lemma for convex combinations). *Let*

$$\{A_{s,j} \,:\, s = 1, \ldots, S,\ j = 1, \ldots, N\}$$

be a family of subsets of a set Ω such that $A_{s,1}, \ldots, A_{s,N}$ form a partition of Ω for each fixed s. Let $\alpha_1, \ldots, \alpha_s$ be such that $\alpha_s \geqslant 0$ for $s = 1, \ldots, S$ and $\alpha_1 + \ldots + \alpha_S = 1$. Then, for all sets $\{\mathbb{P}_{s,1}, \ldots, \mathbb{P}_{s,N}\}$, $s = 1, \ldots, S$, of probability distributions on Ω,

$$\min_{j=1,\ldots,N} \sum_{s=1}^{S} \alpha_s\, \mathbb{P}_{s,j}[A_{s,j}] \leqslant \max\left\{\frac{e}{1+e}, \frac{\bar{K}}{\ln\,(N-1)}\right\}\ ,$$

where

$$\bar{K} = \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1}\, \mathcal{K}(\mathbb{P}_{s,j}, \mathbb{P}_{s,1})\ .$$

PROOF. Using Lemma A.11, we have that

$$\begin{aligned}
&\sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \lambda\, \mathbb{P}_{s,j}[A_{s,j}] - \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \psi_{\mathbb{P}_{s,1}[A_{s,j}]}(\lambda) \\
&\qquad \leqslant \quad \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \mathcal{K}(\mathbb{P}_{s,j}, \mathbb{P}_{s,1}) = \bar{K}\ .
\end{aligned}$$

Now, for each fixed $\lambda > 0$, the function that maps p to $-\psi_p(\lambda)$ is convex. Hence, letting

$$p_1 = \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \mathbb{P}_{s,1}[A_{s,j}] = \frac{1}{N-1}\left(1 - \sum_{s=1}^{S} \alpha_s \mathbb{P}_{s,1}[A_{s,1}]\right)\ ,$$

by Jensen's inequality we get

$$\begin{aligned}
&\sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \lambda\, \mathbb{P}_{s,j}[A_{s,j}] - \psi_{p_1}(\lambda) \\
&\qquad \leqslant \quad \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \lambda\, \mathbb{P}_{s,j}[A_{s,j}] - \sum_{s=1}^{S}\sum_{j=2}^{N} \frac{\alpha_s}{N-1} \psi_{\mathbb{P}_{s,1}[A_{s,j}]}(\lambda)\ .
\end{aligned}$$

Recalling that the right-hand side of the above inequality above is less than $\bar{K}$, and introducing the quantities

$$a_j = \sum_{s=1}^{S} \alpha_s \mathbb{P}_{s,j}[A_{s,j}] \qquad \text{for } j = 1, \ldots, N,$$

we conclude

$$\lambda \min_{j=1,\ldots,N} a_j - \psi_{\frac{1-a_1}{N-1}}(\lambda) \leqslant \lambda \frac{1}{N-1} \sum_{j=2}^{N} a_j - \psi_{\frac{1-a_1}{N-1}}(\lambda) \leqslant \bar{K} \, .$$

Denote by a the minimum of the a_j's and let $p^* = (1-a)/(N-1) \geqslant p_1$. We only have to deal with the case when $a \geqslant e/(1+e)$. As for all $\lambda > 0$, the function that maps p to $-\psi_p$ is decreasing, we have

$$\bar{K} \geqslant \sup_{\lambda > 0} \left(\lambda \, a - \psi_{p^*}(\lambda) \right) \geqslant a \, \ln \frac{a}{e \, p^*} \geqslant a \, \ln \frac{a \, (N-1)}{(1-a) \, e} \geqslant a \, \ln (N-1) \, ,$$

whenever $p^* \leqslant a \leqslant 1$ for the second inequality to hold (thanks to Lemma A.12), and by using $a \geqslant e/(1+e)$ for the last one. As $p^* \leqslant 1/(N-1) \leqslant e/(1+e)$ whenever $N \geqslant 3$, the case $a < p^*$ may only happen when $N = 2$, but then the result is trivial. □

REMARK A.1. We simply remark here that in some situations, such as the symmetric toy situation described at the beginning of this section, *two* probability distributions are enough to get an extra $\ln N$ factor. Assume we introduced two probability distributions $\mathbb{P}_1$ and $\bar{\mathbb{P}}$ and an event A_1 such that $\mathbb{P}_1[A_1] \geqslant \bar{\mathbb{P}}[A_1] = 1/N$. (This was the case in the toy situation.) Then, thanks to Lemmas A.11 and A.12, we may write

$$\mathcal{K}\left(\mathbb{P}_1, \bar{\mathbb{P}}\right) \geqslant \sup_{\lambda \geqslant 0} \left(\lambda \mathbb{P}_1[A_1] - \psi_{\bar{\mathbb{P}}[A_1]}(\lambda) \right) \geqslant \mathbb{P}_1[A_1] \ln \left(\frac{\mathbb{P}_1[A_1]}{e \, \bar{\mathbb{P}}[A_1]} \right) \geqslant \mathbb{P}_1[A_1] \ln \frac{N}{4} \, ,$$

provided (for the last step) that $\mathbb{P}_1[A_1] \geqslant e/4$. Thus, we have that

$$\mathbb{P}_1[A_1] \leqslant \max \left\{ \frac{e}{4}, \frac{\mathcal{K}\left(\mathbb{P}_1, \bar{\mathbb{P}}\right)}{\ln (N/4)} \right\} \, .$$

5. A lemma for solving for the regrets

At various points in the previous chapters (see, e.g., the proof of Corollary or Section 4 of Chapter 5), we had to solve an inequality for the regrets or for the cumulative (expected) losses. The lemma below offers a straightforward upper bound over the solution, which is all we need.

LEMMA A.14. *If $x_t, y_t \geqslant 0$, and $b \geqslant 0$, are such that for all $t = 1, \ldots, n$,*

$$x_t \leqslant y_t + b\sqrt{x_n} \, , \tag{A.2}$$

then

$$\forall t = 1, \ldots, n, \qquad x_t \leqslant y_t + b\sqrt{y_n} + b^2 \, .$$

PROOF. We obtain a bound over $\sqrt{x_n}$ and substitute it into (A.2) to conclude. The inequality

$$x_n \leqslant y_n + b\sqrt{x_n}$$

rewrites as

$$\left(\sqrt{x_n} - \frac{b}{2} \right)^2 \leqslant y_n + \frac{b^2}{4} \, ,$$

that is, either $\sqrt{x_n} \leqslant b/2$ or

$$\sqrt{x_n} - \frac{b}{2} = \left| \sqrt{x_n} - \frac{b}{2} \right| \leqslant \sqrt{y_n + \frac{b^2}{4}} \leqslant \sqrt{y_n} + \frac{b}{2} \, .$$

In both cases,

$$\sqrt{x_n} \leqslant b + \sqrt{y_n}$$

concluding the proof. □

Bibliography

[**AlAu04**] C. Allenberg-Neeman and P. Auer. Personal communication.

[**AlNe04**] C. Allenberg-Neeman and B. Neeman. Full information game with gains and losses. In *Proceedings of the 15th International Conference on Algorithmic Learning Theory*, pages 264-278. Springer, 2004.

[**Aub79**] J. Aubin. *Mathematical Methods of Game and Economic Theory*. North Holland Publishing, Amsterdam, 1979.

[**Aue02**] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422, 2002.

[**AuCeFrSc02**] P. Auer, N. Cesa-Bianchi, Y. Freund, and R.E. Schapire. The nonstochastic multi-armed bandit problem. *SIAM Journal on Computing*, 32:48–77, 2002.

[**AuCeGe02**] P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64:48–75, 2002.

[**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.

[**Azu67**] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 68:357–367, 1967.

[**Ban68**] A. Baños. On pseudo-games. *Annals of Mathematical Statistics*, 39:1932–1945, 1968.

[**Ber90**] M. Berger. *Géométrie*. Nathan, Paris, 1990.

[**Bir05**] L. Birgé. A new lower bound for multiple hypothesis testing. *IEEE Transactions on Information Theory*, 2005. To appear.

[**Bla56**] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[**BlKa99**] A. Blum and A. Kalai. Universal portfolios with and without transaction costs. *Machine Learning*, 35:193–205, 1999.

[**BlMa05**] A. Blum and Y. Mansour. From external to internal regret. In P. Auer and R. Meir, editors, *Proceedings of the 18th Annual Conference on Learning Theory*. Springer, 2005. In press.

[**BoElGo00**] A. Borodin, R. El-Yaniv, and V. Gogan. On the competitive theory and practice of portfolio selection (extended abstract). In *Proceedings of the 4th Latin American Symposium on Theoretical Informatics*, pages 173–196, Punta del Este, Uruguay, 2000.

[**BoBoLu05**] S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: a survey of recent advances. *ESAIM: Probability and Statistics*, 2005. To appear.

[**Cat01**] O. Catoni. Optimisation stochastique et reconnaissance des formes. Lectures presented at "2001, a Statistics Odyssey", Institut Henri-Poincaré. Paris, 2001.

[**Cau01**] R. Cauty. Solution du problème de point fixe de Schauder. *Fundamenta Mathematicæ*, 170:231–246, 2001.

[**Ces99**] N. Cesa-Bianchi. Analysis of two gradient-based algorithms for on-line regression. *Journal of Computer and System Sciences*, 59(3):392–411, 1999.

[**CeFrHaHeScWa97**] N. Cesa-Bianchi, Y. Freund, D.P. Helmbold, D. Haussler, R. Schapire, and M.K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

[**CeLu99**] N. Cesa-Bianchi and G. Lugosi. On prediction of individual sequences. *Annals of Statistics*, 27:1865–1895, 1999.

[**CeLu00**] N. Cesa-Bianchi and G. Lugosi. Minimax values and entropy bounds for portfolio selection problems. In *Proceedings of the First World Congress of the Game Theory Society*, 2000.

[**CeLu03**] N. Cesa-Bianchi and G. Lugosi. Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51:239–261, 2003.

[**CeLu05**] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, to appear.

[**CeLuSt04a**] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. In J. Shawe-Taylor and Y. Singer, editors, *Proceedings of the 17th Annual Conference on Learning Theory*, pages 77–92. Springer, 2004.

[**CeLuSt04b**] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Regret minimization under partial monitoring. Technical report, Ecole Normale Supérieure, 2004.

[**CeLuSt05**] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. *IEEE: Transactions on Information Theory*, 2005. To appear.

[**CeMaSt05**] N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds in prediction with expert advice. In P. Auer and R. Meir, editors, *Proceedings of the 18th Annual Conference on Learning Theory*. Springer, 2005. In press.

[**Cov65**] T.M. Cover. Behavior of sequential predictors of binary sequences. In *Proceedings of the 4th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 263–272. 1965.

[**Cov84**] T.M. Cover. An algorithm for maximizing expected log investment return. *IEEE Transactions on Information Theory*, 30:369–373, 1984.

[**Cov91**] T.M. Cover. Universal portfolios. *Mathematical Finance*, 1:1–29, 1991.

[**CoOr96**] T.M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42:348–363, 1996.

[**ChTe88**] Y.S. Chow and H. Teicher. *Probability Theory*. Springer, 1988.

[**Chu94**] T.H. Chung. *Minimax Learning in Iterated Games via Distributional Majorization*. PhD thesis, Stanford University, 1994.

[**CoTh91**] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.

[**DaDu83**] D. Dacunha-Castelle and M. Duflo. *Probabilités et statistiques : Problèmes à temps mobile*. Masson, Paris, 1983.

[**DeGyLu96**] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition.* Springer-Verlag, New-York. 1996.

[**EtKu86**] S. Ethier and T. Kurtz. *Markov processes.* Wiley, New York, 1986.

[**FaMe03**] D.P. de Farias and 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*, 2003.

[**FeMeGu92**] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Transactions on Information Theory*, 38:1258–1270, 1992.

[**FoVo97**] D. Foster and R. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.

[**FoVo98**] D. Foster and R. Vohra. Asymptotic calibration. *Biometrika*, 85:379–390, 1998.

[**FoVo99**] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–36, 1999.

[**FoYo03**] D. Foster and P. Young. Regret testing: a simple payoff-based procedure for learning Nash equilibrium. Manuscript, 2003.

[**Fre75**] D. A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3:100–118, 1975.

[**FrSc97**] Y. Freund and 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.

[**FuLe95**] D. Fudenberg and D. Levine. Universal consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19:1065–1089, 1995.

[**FuLe98**] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, 1998.

[**FuLe99**] D. Fudenberg and D. Levine. Universal conditional consistency. *Games and Economic Behavior*, 29:104–130, 1999.

[**GeLu04**] F. Germano and G. Lugosi. Global Nash convergence of Foster and Young's regret testing. Manuscript, 2004.

[**GrJa03**] A. Greenwald and A. Jafari. A general class of no–regret learning algorithms and game–theoretic equilibria. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop*, 2–12. Springer, 2003.

[**Han57**] J. Hannan. Approximation to Bayes risk in repeated play. *Contributions to the theory of games*, 3:97–139, 1957.

[**Har04**] S. Hart. Adaptive heuristics. Technical report, The Hebrew University of Jerusalem, Center for Rationality DP-372, 2004.

[**HaMa00**] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[**HaMa01**] S. Hart and A. Mas-Colell. A general class of adaptive strategies. *Journal of Economic Theory*, 98:26–54, 2001.

[**HaMa02**] S. Hart and A. Mas-Colell. A reinforcement procedure leading to correlated equilibrium. In G. Debreu, W. Neuefeind, and W. Trockel, editors, *Economic Essays: A Festschrift for Werner Hildenbrand*, pages 181–200. Springer, New York, 2002.

[**HaMa03**] S. Hart and A. Mas–Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93:1830–1836, 2003.

[**HaMa04**] S. Hart and A. Mas–Colell. Stochastic uncoupled dynamics and Nash equilibrium. Technical report, The Hebrew University of Jerusalem, Center for Rationality DP-371, 2004.

[**HaSc89**] S. Hart and D. Schmeidler. Existence of correlated equilibria. *Mathematics of Operations Research*, 14:18–25, 1989.

[**HaKiWa98**] D. Haussler, J. Kivinen, and M. Warmuth. Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 44:1906–1925, 1998.

[**HePa97**] D. P. Helmbold and S. Panizza. Some label efficient learning results. In *Proceedings of the 10th Annual Conference on Computational Learning Theory*, pages 218–230. ACM Press, 1997.

[**HeScSiWa98**] D. P. Helmbold, R. E. Schapire, Y. Singer, and M. K. Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8:325–344, 1998.

[**HiLa97**] F. Hirsch and G. Lacombe. *Eléments d'Analyse Fonctionnelle*. Masson, Paris, 1997.

[**HeLiLo00**] D.P. Helmbold, N. Littlestone, and P.M. Long. Apple tasting. *Information and Computation*, 161(2):85–139, 2000.

[**Hoe63**] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[**HuPo04**] M. Hutter and J. Poland. Prediction with expert advice by following the perturbed leader for general weights. In *Proceedings of the 15th International Conference on Algorithmic Learning Theory*, pages 279-293. Springer, 2004.

[**KaVe03a**] A. Kalai and S. Vempala. Efficient algorithms for universal portfolios. *Journal of Machine Learning Research*, 3(3):423–440, 2003.

[**KaVe03b**] A. Kalai and S. Vempala. Efficient algorithms for the on-line decision problem. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop*, pages 26–40. Springer, 2003.

[**KiWa97**] J. Kivinen and M.K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

[**KiWa99**] J. Kivinen and M.K. Warmuth. Averaging expert predictions. In H.U. Simon and P. Fischer, editors, *Proceedings of the 4th European Conference on Computational Learning Theory*, pages 153–167. Springer, 1999.

[**KlLe03**] R. Kleinberg and T. Leighton. The value of knowing a demand curve: Bounds on regret for on-line posted-price auctions. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 594–605. IEEE Press, 2003.

[**KoTi61**] A.N. Kolmogorov and V.M. Tikhomirov. ε–entropy and ε–capacity of sets in function spaces. *American Mathematical Society Translations*, 17:277–364, 1961.

[**Leh97**] E. Lehrer. Approachability in infinite dimensional spaces. *International Journal of Game Theory*, 31:255–270, 1997.

[**Leh03**] E. Lehrer. A wide range no–regret theorem. *Games and Economic Behavior*, 42:101–115, 2003.

[**LeZi76**] A. Lempel and J. Ziv. On the complexity of an individual sequence. *IEEE: Transactions on Information Theory*, 22:75–81, 1976.

[**Lit88**] N. Littlestone. Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm. *Machine Learning*, 2(4):285–318, 1988.

[**Lit89**] N. Littlestone. *Mistake Bounds and Logarithmic Linear-threshold Learning Algorithms*. PhD thesis, University of California at Santa Cruz, 1989.

[**LiWa94**] N. Littlestone and M.K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212–261, 1994.

[**Lug01**] G. Lugosi. Lectures on prediction of individual sequences. Presented at "2001, a Statistics Odyssey", Institut Henri-Poincaré. Paris, 2001.

[**MaTs99**] E.J. Mammen and A. Tsybakov. Smooth discrimination analysis. *Annals of Statistics*, 27:1808–1829, 1999.

[**MaSh03**] S. Mannor and N. Shimkin. On-line learning with imperfect monitoring. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop*, pages 552–567. Springer, 2003.

[**Mas05**] P. Massart. *Concentration inequalities and model selection*. Saint-Flour summer school lecture notes, Springer, New-York, 2005. To appear.

[**MaNe03**] P. Massart and E. Nédélec. Risk bounds for statistical learning. Technical report, Université Paris-Sud, 2003.

[**McD89**] C. McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, Proceedings of the 12th British Combinatorial Conference*, pages 148–188. Cambridge University Press, 1989.

[**Meg80**] N. Megiddo. On repeated games with incomplete information played by non-Bayesian players. *International Journal of Game Theory*, 9:157–167, 1980.

[**MeFe98**] N. Merhav and M. Feder. Universal prediction. *IEEE Transactions on Information Theory*, 44:2124–2147, 1998.

[**MeSoZa94**] J.-F. Mertens, S. Sorin, and S. Zamir. Repeated games. CORE Discussion paper, no. 9420,9421,9422, Louvain-la-Neuve, 1994.

[**OrCo98**] E. Ordentlich and T.M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23:960–982, 1998.

[**PiSc01**] A. Piccolboni and C. Schindelhauer. Discrete prediction games with arbitrary feedback and loss. In *Proceedings of the 14th Annual Conference on Computational Learning Theory*, pages 208–223, 2001.

[**Rud74**] W. Rudin. *Real and Complex Analysis.* Mcgraw–Hill, New–York. 1974.

[**Rus99**] A. Rustichini. Minimizing regret: The general case. *Games and Economic Behavior*, 29:224–243, 1999.

[**SaMaWe88**] A. de Santis, G. Markowski, and M.N. Wegman. Learning probabilistic prediction functions. In *Proceedings of the 1st Annual Workshop on Computational Learning Theory*, pages 312–328, 1988.

[**Sin97**] Y. Singer. Switching portfolios. *International Journal of Neural Systems*, 8:445–455, 1997.

[**Sor02**] S. Sorin. *A First Course on Zero-Sum Repeated Games*. Série "Mathématiques et Applications" de la SMAI, Springer, 2002.

[**StLu03**] G. Stoltz and G. Lugosi. Internal regret in on-line portfolio selection. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop*, pages 403–417. Springer, 2003.

[**StLu04**] G. Stoltz and G. Lugosi. Learning correlated equilibria in games with compact sets of strategies. Technical report, Université Paris-Sud, 2004.

[**StLu05**] G. Stoltz and G. Lugosi. Internal regret in on-line portfolio selection. *Machine Learning*, 2005. To appear.

[**StSc05**] I. Steinwart and C. Scovel. Fast rates for Support Vectors Machines. In P. Auer and R. Meir, editors, *Proceedings of the 18th Annual Conference on Learning Theory*. Springer, 2005. In press.

[**Tsy02**] A. Tsybakov. Optimal aggregation of classifiers in statistical learning. *Annals of Statistics*, 32(1):135–166, 2004.

[**Tsy04**] A. Tsybakov. *Introduction à l'estimation non-paramétrique.* Springer, New-York, 2004.

[**Vov90**] V.G. Vovk. Aggregating strategies. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 372–383, 1990.

[**Vov98**] V.G. Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–73, 1998.

[**Vov01**] V.G. Vovk. Competitive on-line statistics. *International Statistical Review*, 69:213–248, 2001.

[**WeMe01**] T. Weissman and N. Merhav. Universal prediction of binary individual sequences in the presence of noise. *IEEE Transactions on Information Theory*, 47:2151–2173, 2001.

[**WeMeSo01**] T. Weissman, N. Merhav, and A. Somekh-Baruch. Twofold universal prediction schemes for achieving the finite state predictability of a noisy individual binary sequence. *IEEE Transactions on Information Theory*, 47:1849–1866, 2001.

[**WiShTj95**] F. Willems, Y. Shtarkov and T. Tjalkens. The context-tree weighting method: basic properties. *IEEE Transactions on Information Theory*, 41:653-664, 1995.

[**WiShTj96**] F. Willems, Y. Shtarkov and T. Tjalkens. Context weighting for general finite-context sources. *IEEE Transactions on Information Theory*, 42:1514-1520, 1996.

[**YaElSe04**] R. Yaroshinsky, R. El-Yaniv, and S. Seiden. How to better use expert advice. *Machine Learning*, 55(3):271–309, 2004.

[**Ziv78**] J. Ziv. Coding theorems for individual sequences. *IEEE: Transactions on Information Theory*, 24:405–412, 1978.

[**ZiLe77**] J. Ziv and A. Lempel. A universal algorithm for sequential data-compression. *IEEE: Transactions on Information Theory*, 23:337–343, 1977.

N° d'impression : 2670
Deuxième trimestre 2005

Information incomplète et regret interne en prédiction de suites individuelles

Résumé : Le domaine de recherche dans lequel s'inscrit ce travail de thèse est la théorie de la prédiction des suites individuelles. Cette dernière considère les problèmes d'apprentissage séquentiel pour lesquels on ne peut ou ne veut pas modéliser le problème de manière stochastique, et fournit des stratégies de prédiction très robustes. Elle englobe aussi bien des problèmes issus de la communauté du *machine learning* que de celle de la théorie des jeux répétés, et ces derniers sont traités avec des méthodes statistiques, incluant par exemple les techniques de concentration de la mesure ou de l'estimation adaptative. Les résultats obtenus aboutissent, entre autres, à des stratégies de minimisation des regrets externe et interne dans les jeux à information incomplète, notamment les jeux répétés avec signaux. Ces stratégies s'appliquent au problème d'ajustement séquentiel des prix de vente, ou d'allocation séquentielle de bande passante. Le regret interne est ensuite plus spécifiquement étudié, d'abord dans le cadre de l'investissement séquentiel dans le marché boursier, pour lequel des simulations sur des données historiques sont proposées, puis pour l'apprentissage des équilibres corrélés des jeux infinis à ensembles de stratégies convexes et compacts.

Mots-clés : Suites individuelles, prédiction séquentielle, prédiction avec avis d'experts, regret externe, regret interne, jeux répétés avec signaux, sélection séquentielle de portefeuilles, equilibres corrélés des jeux infinis.

Incomplete Information and Internal Regret in Prediction of Individual Sequences

Abstract: This thesis takes place within the theory of prediction of individual sequences. The latter avoids any modelling of the data and aims at providing some techniques of robust prediction and discuss their possibilities, limitations, and difficulties. It considers issues arising from the machine learning as well as from the game-theory communities, and these are dealt with thanks to statistical techniques, including martingale concentration inequalities and minimax lower bound techniques. The obtained results consist, among others, in external and internal regret minimizing strategies for label efficient prediction or in games with partial monitoring. Such strategies are valuable for the on-line pricing problem or for on-line bandwidth allocation. We then focus on internal regret for general convex losses. We consider first the case of on-line portfolio selection, for which simulations on real data are provided, and generalize later the results to show how players can learn correlated equilibria in games with compact sets of strategies.

Keywords: On-line learning, individual sequences, sequential prediction, prediction with expert advice, external regret, internal regret, repeated games, prediction with partial monitoring, on-line portfolio selection, correlated equilibrium of infinite games, game-theoretic learning.

AMS Classification: 68Q17, 68Q32, 91A10, 91A20, 91A26, 62L12, 62P05.