JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

GÁBOR LUGOSI

Prédiction randomisée de suites individuelles

Journal de la société française de statistique, tome 147, n° 1 (2006), p. 5-37

<http://www.numdam.org/item?id=JSFS_2006__147_1_5_0>

PRÉDICTION RANDOMISÉE DE SUITES INDIVIDUELLES

Gábor LUGOSI *

RÉSUMÉ

Cet article fait suite à la conférence que j'ai eu l'honneur de donner en hommage à Lucien Le Cam, lors des *XXXVIèmes Journées de Statistique* à Montpellier, en 2004. Il passe en revue quelques résultats récents en prédiction randomisée de suites arbitraires; l'accent est mis sur les situations où le prévisionniste n'a qu'un accès restreint au processus qu'il essaie de prédire. La plupart des résultats mentionnés dans l'article reposent sur des travaux en collaboration avec Nicolò Cesa-Bianchi, András György, Tamás Linder, György Ottucsák et Gilles Stoltz. Pour un traitement plus complet du sujet, le lecteur pourra consulter la monographie de Cesa-Bianchi et Lugosi [15].

ABSTRACT

This paper is a written version of the « Conférence Lucien Le Cam » delivered at the *XXXVIèmes Journées de Statistique* in Montpellier, 2004. It is devoted to surveying some recent results on randomized prediction of arbitrary sequences. We focus on situations in which the forecaster has limited access to the process he's trying to predict. Most results mentioned in the paper are based on joint work with Nicolò Cesa-Bianchi, András György, Tamás Linder, György Ottucsák, and Gilles Stoltz. For a more complete coverage of the subject we refer the reader to the monograph of Cesa-Bianchi and Lugosi [15].

1. Introduction

L'exposé qui suit se concentre sur différentes variantes d'un problème de prédiction séquentielle. Dans la version la plus simple de ce problème, le prévisionniste – ou pronostiqueur – observe, l'un après l'autre, les éléments d'une suite $y_1, y_2, \ldots$ de symboles. Tout au long de l'article, on supposera que $y_t \in \mathcal{Y}$, où $\mathcal{Y}$ est un ensemble de résultats possibles (fini ou infini). À chaque tour $t = 1, 2, \ldots$, avant que le t-ième symbole de la suite ne soit dévoilé, le pronostiqueur essaie de deviner sa valeur y_t en se fondant sur les $t-1$ observations précédentes.

* L'auteur tient à remercier pour leur soutien le ministère espagnol de la recherche et de la technologie et FEDER (subvention BMF2003-03324), ainsi que le réseau européen d'excellence PASCAL (subvention 506778). Il est également reconnaissant envers le BBCA pour son soutien financier dans le cadre du programme «Aprender a jugar».

ICREA and Department of Economics, Pompeu Fabra University, 08005 Barcelona, Spain
lugosi@upf.es

Dans la théorie statistique de l'estimation séquentielle, on suppose classiquement que les éléments de $y_1, y_2, \ldots$ – qu'on désignera par *résultats* – sont les réalisations d'un certain processus stochastique (stationnaire). Il s'agit alors d'estimer les caractéristiques de ce processus ; on en déduit ensuite des règles de prédiction efficaces. Dans un tel contexte, le *risque* d'une règle de prédiction peut être défini comme l'espérance d'une certaine *fonction de perte* mesurant l'écart entre la valeur prédite et la vrai résultat ; on compare différentes règles *via* le comportement de leurs risques.

On abandonne ici cette hypothèse essentielle de génération des résultats par un processus stochastique sous-jacent, et on voit la suite $y_1, y_2, \ldots$ comme le produit d'un certain mécanisme, inconnu, non spécifié (et qui pourrait être déterministe, stochastique, ou même dynamique et antagoniste). On appelle souvent cette approche prédiction de *suites individuelles*, pour l'opposer à celle qui procède par une modélisation stochastique préliminaire. Sans modèle probabiliste, on ne peut toutefois pas définir de notion de risque, et fixer les objectifs de la prédiction n'est pas chose de toute évidence.

Dans le modèle étudié dans cet exposé, le pronostiqueur choisit à chaque pas une action $i \in \{1, \ldots, N\}$ parmi N actions, et lorsque le résultat est y_t, il essuie une perte $\ell(i, y_t)$ où ℓ est une certaine fonction de perte, à valeurs dans l'intervalle $[0, 1]$. La performance de sa stratégie est comparée à celle du meilleur pronostiqueur « constant », c'est-à-dire, à celle de l'action fixe i qui, parmi toutes les autres, obtient la moindre perte cumulée lors des n tours de prédiction.

On appelle *regret* la différence entre la perte cumulée du pronostiqueur et celle d'une action constante, simplement parce qu'il mesure combien le pronostiqueur regrette, rétrospectivement, de ne pas avoir suivi cette action précise. Ainsi, il cherche à commettre (presque) aussi peu d'erreurs que la meilleure stratégie constante. Son idéal est que son regret, rapporté au nombre de pas du problème, converge vers zéro.

Comme la suite des résultats est complètement arbitraire, il est immédiat que pour toute stratégie de prédiction déterministe, il existe une suite de résultats $y_1, \ldots, y_n$ telle que le pronostiqueur ait choisi à chaque tour l'action la pire ; aucune borne sensée ne peut donc être obtenue sur le regret d'une telle stratégie. Cela peut paraître étonnant, mais dès qu'on autorise le pronostiqueur à recourir au hasard (c'est-à-dire qu'il peut lancer une pièce avant de former sa prédiction), la situation change du tout au tout ; l'introduction d'un aléa dans le choix final est un outil extrêment puissant.

Dans la partie 2, on présente de manière formelle le problème de prédiction randomisée de référence, et on décrit des pronostiqueurs simples dont le regret croît, pour toute suite possible de résultats, plus lentement que linéairement. Les parties ultérieures sont consacrées à diverses modifications du problème, dans lesquelles le pronostiqueur a un accès plus restreint aux éléments passés de la suite à prédire.

Ainsi, en partie 3, on étudie le cas où seule une faible part des résultats passés est portée à la connaissance du pronostiqueur. Etonnamment, même dans cette version « économe en termes d'observations » du jeu de prédiction, le

regret rapporté au nombre de tours de prédiction est asymptotiquement nul, sous la seule condition que le nombre de résultats observés après n pas croisse plus rapidement que $\log(n) \log\log(n)$.

La partie 4 formule dans un cadre général des problèmes de prédiction dans lesquels l'information est restreinte. La prédiction en *information imparfaite* correspond au cas où le pronostiqueur, après sa prédiction, ne reçoit qu'un signal de répercussion, au lieu de prendre connaissance de la perte qu'il a subie. Le degré de difficulté de ce problème dépend des liens entre les pertes et les répercussions. On détermine des conditions générales sous lesquelles il est possible de garantir que le regret est faible ; on établit également les vitesses optimales de convergence vers zéro du regret rapporté au nombre de pas du problème.

Enfin, la partie 5 est consacrée au problème du choix séquentiel d'un chemin dans un réseau aux arcs duquel correspondent des pertes variant arbitrairement ; et ceci, avec la difficulté supplémentaire que le pronostiqueur n'accède qu'aux pertes des arcs se situant sur le chemin pris, et pas à celles des autres arcs.

La première mention des problèmes de prédiction randomisée de suites arbitraires remonte aux années 50, et plus précisément aux travaux de Hannan [30] et Blackwell [8], qui ont formulé leurs résultats dans un cadre nommé « problèmes de décisions multiples séquentielles ». Cover [18] figure également parmi les pionniers du domaine. Le problème de la prédiction séquentielle, tel que démuni de toute hypothèse de nature probabiliste, est intimement lié à la compression de suites individuelles de données en théorie de l'information. Les recherches d'avant-garde dans ce domaine ont été menées par Ziv [63, 64] et Lempel et Ziv [43, 65] ; ils ont résolu la question de compresser une suite individuelle de données presque aussi bien que le meilleur automate fini (Feder, Merhav et Gutman [19], Merhav et Feder [48] proposent des résultats supplémentaires dans cette direction). Le paradigme de la prédiction de suites individuelles a été étudié également en théorie de l'apprentissage, où Littlestone et Warmuth [44] et Vovk [57] ont introduit le problème ; Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [12], Foster [20] et Vovk [58] présentent quelques résultats fondamentaux.

Pour un passage en revue de sujets connexes, on renvoie à Merhav et Feder [49], Foster et Vohra [22], Vovk [60]. Cesa-Bianchi et Lugosi [15] proposent un traitement détaillé de ces problèmes de prédiction et d'autres questions liées à eux. Le lecteur pourra également consulter Stoltz [55] pour avoir davantage de précisions concernant le cas de la prédiction avec accès restreint aux résultats passés.

2. Le jeu de prédiction en information complète

On commence par la version la plus simple du jeu de prédiction, celle dans laquelle le pronostiqueur a un accès total aux résultats des tours précédents.

On considère le jeu suivant, qui se déroule entre un pronostiqueur (le joueur) et son environnement. À chaque tour de jeu, le joueur choisit une action $i \in \{1, \dots, N\}$, tandis que l'environnement choisit une action (également appelée « résultat ») $y \in \mathcal{Y}$. La perte $\ell(i, y)$ du pronostiqueur est alors donnée par une fonction de perte $\ell : \{1, \dots, N\} \times \mathcal{Y} \to [0, 1]$. On suppose que le choix de l'action du joueur est randomisé, c'est-à-dire qu'au t-ième tour du jeu, le joueur choisit une probabilité $\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})$ sur son ensemble d'actions et joue l'action i avec probabilité $p_{i,t}$. Formellement, l'action I_t du joueur au tour t est définie par

$$I_t = i \quad \text{si et seulement si} \quad U_t \in \left[\sum_{j=1}^{i-1} p_{j,t}, \sum_{j=1}^{i} p_{j,t} \right)$$

de sorte que

$$\mathbb{P}\left[I_t = i \mid U_1, \dots, U_{t-1} \right] = p_{i,t} \qquad i = 1, \dots, N,$$

où $U_1, U_2, \dots$ est une suite de variables aléatoires i.i.d. selon la loi uniforme sur l'intervalle $[0, 1]$. En termes de théorie des jeux, nous dirions qu'une action est une *stratégie pure* et qu'une probabilité sur les actions est une *stratégie mixte*. Si l'environnement choisit le résultat $y_t \in \mathcal{Y}$, la perte du joueur est alors $\ell(I_t, y_t)$.

Prédiction randomisée en information complète

Paramètres : N actions, l'ensemble $\mathcal{Y}$ des résultats possibles, la fonction de perte ℓ, le nombre n de tours de jeu.

À chaque tour $t = 1, 2, \dots, n$,

(1) l'environnement choisit le résultat $y_t \in \mathcal{Y}$ à venir ;
(2) le pronostiqueur détermine sa stratégie mixte $\mathbf{p}_t$ et tire son action I_t selon la probabilité $\mathbf{p}_t$;
(3) l'environnement dévoile y_t ;
(4) le pronostiqueur subit une perte $\ell(I_t, y_t)$.

Le pronostiqueur cherche à minimiser son regret cumulé,

$$\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n} = \sum_{t=1}^{n} \ell(I_t, y_t) - \min_{i=1,\dots,N} \sum_{t=1}^{n} \ell(i, y_t)$$

à savoir, la différence entre sa perte cumulée et celle de la meilleure stratégie pure, et ce, quelle que soit la suite des résultats.

Remarque 1 (experts). — Une formulation plus générale du problème consiste à considérer différentes stratégies de prédiction, souvent appelées experts, et à définir le regret par rapport à la meilleure telle stratégie, plutôt que par rapport à la meilleure stratégie pure (c'est-à-dire, la meilleure action constante). Dans la mesure où la plupart des résultats qui suivent peuvent être étendus de manière immédiate à ce cadre plus général, ils seront présentés dans le modèle le plus simple et on soulignera simplement les endroits pour lesquels l'extension au cas des experts requiert un peu d'attention.

Avant de continuer, détaillons un point délicat qui provient de ce que le pronostiqueur utilise une stratégie randomisée. Les actions y_t de l'environnement, en tant qu'elles sont autorisées à dépendre des actions (randomisées) passées $I_1, \ldots, I_{t-1}$ du pronostiqueur, sont elles-mêmes des réalisations de variables aléatoires Y_t. On pourrait même imaginer que l'environnement utilise lui aussi une suite de variables aléatoires indépendantes pour choisir les résultats Y_t, mais comme ceci est sans effet sur les résultats mathématiques de cette partie et des suivantes, et pour simplifier les choses, on exclura en fait cette possibilité.

Plus précisément, l'adversaire est défini par une suite de fonctions $g_1, g_2, \ldots$, où $g_t : \{1, \ldots, N\}^{t-1} \to \mathcal{Y}$, et chaque résultat Y_t est donné par $Y_t = g_t(I_1, \ldots, I_{t-1})$. Ainsi, Y_t est mesurable par rapport à la tribu engendrée par les variables aléatoires $U_1, \ldots, U_{t-1}$.

Une observation-clé dans toute analyse de stratégie de prédiction randomisée est que la perte cumulée $\widehat{L}_n = \sum_{t=1}^n \ell(I_t, y_t)$ est toujours proche de la quantité dite *perte cumulée en espérance* (conditionnelle) et définie par

$$\overline{L}_n \stackrel{\text{déf.}}{=} \sum_{t=1}^n \overline{\ell}(\mathbf{p}_t, Y_t)$$

où $\overline{\ell}(\mathbf{p}_t, Y_t) = \sum_{i=1}^N p_{i,t} \ell(i, Y_t) = \mathbb{E}_t \ell(I_t, Y_t)$ est l'espérance conditionnelle de la perte $\ell(I_t, Y_t)$, étant donnés les tirages passés $U_1, \ldots, U_{t-1}$. Des inégalités de martingales permettent de quantifier aisément cette proximité. Par exemple, l'inégalité, bien connue, de Hoeffding-Azuma, qui s'applique à des sommes d'accroissements de martingales bornés (*cf.* [39],[5]), assure que pour tout $\delta \in (0,1)$, et avec probabilité au moins $1 - \delta$,

$$\widehat{L}_n \leqslant \overline{L}_n + \sqrt{\frac{n}{2} \ln \frac{1}{\delta}} \, .$$

(Comme on le verra, les inégalités de martingales jouent un rôle crucial dans l'analyse des stratégies de prédiction randomisées, et on sera amené à en considérer de plus fines.) En tout état de cause, pour toute stratégie de prédiction randomisée, $\widehat{L}_n - \overline{L}_n = O_p\left(n^{1/2}\right)$. Vu qu'il s'agit de construire des stratégies avec un regret cumulé croissant plus lentement que linéairement en n, il suffit de s'intéresser à la différence entre la perte en espérance $\overline{L}_n$ et celle de la meilleure action $\min_{i=1,\ldots,N} \sum_{t=1}^n \ell(i, y_t)$.

Étudions à présent l'algorithme de *prédiction par pondération exponentielle*, une stratégie très efficace, quoique l'une des plus simples possibles. Au tour t, il tire une action I_t selon la probabilité donnée par les

$$p_{i,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell(i, Y_s)\right)}{\sum_{k=1}^{N} \exp\left(-\eta \sum_{s=1}^{t-1} \ell(k, Y_s)\right)} \quad i = 1, \ldots, N$$

où $\eta > 0$ est un paramètre de l'algorithme. Ainsi, ce dernier donne à chacune des N actions un poids exponentiel en (l'opposé de) leurs performances passées : plus la perte accumulée par une action jusqu'au tour $t-1$ est petite, plus le poids mis sur elle est grand. La performance de cet algorithme de prédiction peut être bornée de la manière suivante.

THÉORÈME 1. — *Le regret cumulé en espérance (conditionnelle) de l'algorithme de prédiction par pondération exponentielle vérifie*

$$\overline{L}_n - \min_{i=1,\ldots,N} L_{i,n} \leqslant \frac{\ln N}{\eta} + \frac{n\eta}{8} .$$

Le choix de $\eta = \sqrt{8 \ln N / n}$ *conduit au majorant* $\sqrt{(n \ln N)/2}$.

Le théorème 1 est l'un des résultats les plus fondamentaux et les plus connus en prédiction de suites individuelles. Plusieurs versions en ont été données, par Littlestone et Warmuth [44], Vovk [57, 58], Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [12]. Nous reprenons ci-dessous une preuve élémentaire suggérée par Cesa-Bianchi [11] (voir également Cesa-Bianchi et Lugosi [13]).

Preuve. — Soit, pour $t \geqslant 1$, $L_{i,t} = \sum_{s=1}^{t} \ell(i, y_s)$ la perte cumulée de l'action i jusqu'au tour t, $w_{i,t} = e^{-\eta L_{i,t}}$ son poids exponentiel, et

$$W_t = \sum_{i=1}^{N} w_{i,t} = \sum_{i=1}^{N} e^{-\eta L_{i,t}} .$$

On pose par ailleurs $W_0 = N$. On a alors, d'une part,

$$\begin{aligned} \ln \frac{W_n}{W_0} &= \ln \left(\sum_{i=1}^{N} e^{-\eta L_{i,n}} \right) - \ln N \\ &\geqslant \ln \left(\max_{i=1,\ldots,N} e^{-\eta L_{i,n}} \right) - \ln N \\ &= -\eta \min_{i=1,\ldots,N} L_{i,n} - \ln N \end{aligned}$$

Or, d'autre part, pour tout $t = 1, \ldots, n$,

$$\begin{aligned}
\ln \frac{W_t}{W_{t-1}} &= \ln \frac{\sum_{i=1}^{N} w_{i,t-1} e^{-\eta \ell(i,y_t)}}{\sum_{j=1}^{N} w_{j,t-1}} \\
&\leqslant -\eta \frac{\sum_{i=1}^{N} w_{i,t-1} \ell(i,y_t)}{\sum_{j=1}^{N} w_{j,t-1}} + \frac{\eta^2}{8} \\
&= -\eta \overline{\ell}(\mathbf{p}_t, y_t) + \frac{\eta^2}{8}
\end{aligned}$$

où l'on a appliqué dans un premier temps une inégalité bien connue de Hoeffding [39], stipulant que pour toute variable aléatoire bornée X, à valeurs dans l'intervalle $[a,b]$, et pour tout $\eta \in \mathbb{R}$, $\ln \mathbb{E} e^{\eta X} \leqslant \eta \mathbb{E} X + \eta^2 (b-a)^2/8$; et dans un second temps, la définition de l'algorithme de prédiction par pondération exponentielle, qui entraîne que $\sum_{i=1}^{N} w_{i,t-1} \ell(i,y_t) / \sum_{j=1}^{N} w_{j,t-1} = \overline{\ell}(\mathbf{p}_t, y_t)$. En additionnant ces inégalités selon $t = 1, \ldots, n$, on obtient

$$\ln \frac{W_n}{W_0} \leqslant -\eta \sum_{t=1}^{n} \overline{\ell}(\mathbf{p}_t, y_t) + \frac{\eta^2}{8} n \; .$$

La combinaison des bornes inférieure et supérieure sur le rapport $\ln(W_n/W_0)$ donne

$$\sum_{t=1}^{n} \overline{\ell}(\mathbf{p}_t, y_t) \leqslant \min_{i=1,\ldots,N} L_{i,n} + \frac{\ln N}{\eta} + \frac{\eta}{8} n \; ,$$

ce qui conclut la preuve. □

La conjonction du théorème 1 avec l'inégalité de Hoeffding-Azuma montre que lorsque l'algorithme de prédiction par pondération exponentielle est employé avec le paramètre $\eta = \sqrt{8 \ln N / n}$, alors, avec probabilité au moins $1 - \delta$, le regret est borné par

$$\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n} \leqslant \sqrt{\frac{n \ln N}{2}} + \sqrt{\frac{n}{2} \ln \frac{1}{\delta}} \; .$$

On note que le choix proposé pour η dépend de l'*horizon* n, c'est-à-dire que la mise en œuvre de la stratégie proposée requiert de connaître à l'avance le nombre total de tours. Il est apparu que cette difficulté n'est pas insurmontable et qu'il est aisé de transformer la stratégie de prédiction présentée ci-dessus en un algorithme ne requérant pas d'information *a priori* sur le nombre de

tours et assurant que

$$\limsup_{n\to\infty} \frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n}\right) = 0 \quad \text{presque-sûrement.} \tag{1}$$

Une méthode pour ce faire consiste à employer l'algorithme de prédiction par pondération exponentielle avec un paramètre η_t variant au cours du temps selon $\eta_t = \sqrt{8\ln N/t}$ (les détails sont fournis, par exemple, par [15]). Une stratégie de prédiction satisfaisant la propriété (1) est dite *cohérente au sens de Hannan*. Cette question de cohérence, une notion-clé dans le sujet, a été résolue en premier par Hannan [30]. Son algorithme originel est fondé sur l'idée de suivre « la meilleure action bruitée », c'est-à-dire qu'on ajoute une perturbation aléatoire à la perte cumulée de chacune des actions et qu'on choisit celle de perte cumulée pertubée la plus faible. (On pourra consulter Kalai et Vempala [41], ainsi que Hutter et Poland [40], pour une analyse plus fine de ce type de stratégie de prédiction.)

Toujours et encore dans les années 50, Blackwell [9] a indiqué une construction alternative de stratégies cohérentes au sens de Hannan. Sa procédure repose sur sa généralisation efficace du théorème minimax de von Neumann au cas de paiements vectoriels, et son esprit est proche de celui de la prédiction par pondération exponentielle. En effet, cette dernière peut être étendue de la manière suivante : soit un *potentiel*

$$\Phi(\mathbf{u}) = \sum_{i=1}^{N} \phi(u_i)$$

où $\phi : \mathbb{R} \to \mathbb{R}$ est une fonction positive, croissante et deux fois dérivable. La stratégie générale de prédiction par pondération selon un potentiel définit la probabilité $\mathbf{p}_t$ comme $p_{i,1} = 1/N$ pour $t = 1$ et

$$p_{i,t} = \frac{\nabla_i \Phi(\mathbf{R}_{t-1})}{\sum_{k=1}^{N} \nabla_k \Phi(\mathbf{R}_{t-1})} = \frac{\phi'(R_{i,t-1})}{\sum_{k=1}^{N} \phi'(R_{k,t-1})}$$

pour $t > 1$ et $i = 1, \dots, N$, où $R_{i,t} = \overline{L}_t - L_{i,t}$ est le regret cumulé par rapport à l'action i après t tours de jeu et $\mathbf{R}_t = (R_{1,t}, \dots, R_{N,t})$ est le vecteur formé par ces quantités. Il est clair que le choix d'un potentiel exponentiel

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

pour la stratégie de prédiction par pondération selon un potentiel redonne l'algorithme de prédiction par pondération exponentielle introduit précédemment. La stratégie originelle de Blackwell correspond, elle, au choix du potentiel quadratique $\Phi(\mathbf{u}) = \|\mathbf{u}_+\|_2^2$. Hart et Mas-Colell [32] (voir également Cesa-Bianchi et Lugosi [14]) caractérisent toute une classe de potentiels conduisant

à des stratégies de prédiction cohérentes au sens de Hannan ; elle inclut notamment les potentiels polynomiaux $\Phi(\mathbf{u}) = \|\mathbf{u}_+\|_p^2$, pour $p \geqslant 2$. Mais c'est parce que la stratégie de prédiction par pondération exponentielle est simple, flexible et d'analyse aisée que l'on se penchera essentiellement sur elle dans cet article.

On mentionne, pour conclure cette partie, que dans cette formulation générale du problème de prédiction et de ses objectifs, il n'est pas possible de faire mieux que l'agorithme par pondération exponentielle. Cesa-Bianchi, Freund, Haussler, Helmbold, Schapire et Warmuth [12] ont été les maîtres d'œuvre de la borne inférieure suivante (on pourra également consulter Haussler, Kivinen et Warmuth [34] pour des résultats en rapport avec elle).

THÉORÈME 2. — *Soit $n, N \geqslant 1$. Il existe une fonction de perte ℓ telle que pour toute stratégie de prédiction randomisée,*

$$\sup_{y^n \in \mathcal{Y}^n} \left(\overline{L}_n - \min_{i=1,\dots,N} L_{i,n} \right) \geqslant (1 - \epsilon_{N,n}) \sqrt{\frac{n \ln N}{2}}$$

où $\lim_{N\to\infty} \lim_{n\to\infty} \epsilon_{N,n} = 0$.

L'idée principale de la preuve est de prendre des pertes $\ell(i, y_t)$ données par une suite i.i.d. de variables de Bernoulli, puis d'appliquer un argument de limite centrale.

3. Prédiction économe

On s'intéresse ici à une version du problème de prédiction séquentielle dans laquelle il est coûteux d'obtenir la valeur des résultats. Plus précisément, après avoir formé son pari au tour t, le pronostiqueur décide s'il demande à accéder au résultat Y_t (*id est*, à l'observer) ou non. Le nombre $\mu(n)$ de tels accès est cependant limité au sein d'un horizon n donné. Le jeu de prédiction économe en termes d'observations est défini formellement dans l'encadré qui suit (page 14) :

Tout comme précédemment, le pronostiqueur cherche à minimiser son regret cumulé

$$\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n} = \sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{i=1,\dots,N} \sum_{t=1}^{n} \ell(i, Y_t) \ ,$$

et ce, quelle que soit la suite des résultats.

Le problème de la prédiction économe a été introduit par Helmbold et Panizza [36] dans le cas particulier de la prédiction binaire, *id est*, $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $\ell(x,y) = \mathbb{I}_{[x\neq y]}$, où en outre, pour chaque suite de résultats, l'un des experts ne commet aucune erreur (voir aussi l'exemple 3). La matière présentée dans cette partie repose sur Cesa-Bianchi, Lugosi et Stoltz [16], voir également [15, chapitre 6].

Prédiction économe

Paramètres : N actions, l'ensemble $\mathcal{Y}$ des résultats possibles, la fonction de perte ℓ, le rythme d'accès aux résultats $\mu : \mathbb{N} \to \mathbb{N}$.

À chaque tour $t = 1, 2, \ldots,$

(1) l'environnement choisit le résultat $Y_t \in \mathcal{Y}$ à venir, sans le dévoiler ;

(2) le pronostiqueur détermine son action $I_t \in \{1, \ldots, N\}$;

(3) le pronostiqueur et chaque action i subissent respectivement les pertes $\ell(I_t, Y_t)$ et $\ell(i, Y_t)$, mais ces valeurs demeurent inconnues au pronostiqueur pour l'instant ;

(4) si moins de $\mu(t)$ observations ont été effectuées jusqu'à présent, le pronostiqueur peut demander à accéder au résultat Y_t, et ainsi, calculer les pertes du (3) ; sans quoi, il reste dans l'ignorance de la valeur de Y_t.

On considère dans un premier temps le cas d'un horizon fini, dans lequel le pronostiqueur cherche à contrôler son regret après un nombre n (connu et fixé à l'avance) de tours de jeu. On suppose également dans cette version simplifiée du problème de prédiction économe que, bien que seules $m = \mu(n)$ telles observations puissent être faites, aucune contrainte n'est imposée sur leur répartition au cours des n tours de jeu ; autrement dit, et pour utiliser les notations précédentes, $\mu(t) = m$ pour tout $t = 1, \ldots, n$. Dans ce cadre, on définit une stratégie de prédiction simple dont le regret en espérance (conditionnelle) est borné par $n\sqrt{2(\ln N)/m}$. Ainsi, pour $m = n$, on retrouve l'ordre de grandeur de la borne optimale de la partie 2.

On voit facilement qu'il est nécessaire qu'une stratégie recoure à des tirages aléatoires pour déterminer quand elle demande à accéder aux résultats, sans quoi son regret ne peut être contrôlé efficacement par rapport à toutes les suites de résultats possibles – et il se trouve qu'à cet effet, le lancer répété d'une pièce (biaisée) suffit.

L'idée principale est donc d'employer une famille de variables aléatoires Z_1, $Z_2, \ldots, Z_n$ i.i.d. selon une loi de Bernoulli de paramètre ε, $\mathbb{P}[Z_t = 1] = 1 - \mathbb{P}[Z_t = 0] = \varepsilon$, et d'accéder à Y_t chaque fois que $Z_t = 1$. Ici, $\varepsilon > 0$ est un paramètre à déterminer ; il sera typiquement de l'ordre de m/n, de sorte que le nombre d'observations durant les n tours de jeu soit de l'ordre de m (notons que le choix $\varepsilon = m/n$ dans le théorème 3 peut conduire le pronostiqueur à demander à accéder à plus de m observations, mais comme ceci est corrigé facilement dans le théorème 4, où une valeur légèrement plus petite est proposée, nous gardons dans un premier temps, pour la simplicité de l'analyse, la valeur $\varepsilon = m/n$).

La stratégie de prédiction économe décrite ci-dessous n'est rien d'autre qu'une stratégie de prédiction par pondération exponentielle employant au lieu des

pertes les *estimations des pertes* suivantes :

$$\widetilde{\ell}(i, Y_t) = \begin{cases} \ell(i, Y_t)/\varepsilon & \text{lorsque } Z_t = 1\,; \\ 0 & \text{sinon.} \end{cases}$$

On note que

$$\mathbb{E}_t \widetilde{\ell}(i, Y_t) = \mathbb{E}\big[\, \widetilde{\ell}(i, Y_t) \,\big|\, I_1, Z_1, \ldots, I_{t-1}, Z_{t-1} \big] = \ell(i, Y_t)$$

où $I_1, \ldots, I_{t-1}$ est la suite des actions prises aux tours de jeu précédant t. Dit autrement, $\widetilde{\ell}(i, Y_t)$ est un estimateur (conditionnellement) sans biais de la « vraie » perte $\ell(i, Y_t)$.

Stratégie de prédiction économe

Paramètres : Deux réels $\eta > 0$ et $0 \leqslant \varepsilon \leqslant 1$.

Initialisation : $\mathbf{w}_0 = (w_{1,0}, \ldots, w_{N,0}) = (1, \ldots, 1)$.

À chaque tour $t = 1, 2, \ldots,$

(1) choisir une action I_t dans $\{1, \ldots, N\}$ selon la probabilité $\mathbf{p}_t$ dont les composantes sont définies par $p_{i,t} = w_{i,t-1}/(w_{1,t-1} + \cdots + w_{N,t-1})$ pour $i = 1, \ldots, N$;

(2) tirer une variable aléatoire de Bernoulli Z_t telle que $\mathbb{P}[Z_t = 1] = \varepsilon$;

(3) lorsque $Z_t = 1$, accéder à Y_t et calculer

$$w_{i,t} = w_{i,t-1}\, e^{-\eta\, \ell(i, Y_t)/\varepsilon} \qquad \text{pour tout } i = 1, \ldots, N\ ;$$

et sinon, conserver $\mathbf{w}_t = \mathbf{w}_{t-1}$.

L'espérance du regret de cette stratégie de prédiction est bornée de la manière suivante.

THÉORÈME 3. — *Pour un horizon n fixé, la stratégie de prédiction économe utilisant les paramètres $\varepsilon = m/n$ et $\eta = (\sqrt{2m \ln N})/n$ accède en moyenne à m observations, et son regret est borné par*

$$\mathbb{E}\, \widehat{L}_n - \min_{i=1,\ldots,N} \mathbb{E}\, L_{i,n} \leqslant n \sqrt{\frac{2 \ln N}{m}}\ .$$

On pose, en vue de la preuve,

$$\widetilde{\ell}(\mathbf{p}_t, Y_t) = \sum_{i=1}^{N} p_{i,t}\, \widetilde{\ell}(i, Y_t) \qquad \text{et} \qquad \widetilde{L}_{i,n} = \sum_{t=1}^{n} \widetilde{\ell}(i, Y_t) \quad \text{pour } i = 1, \ldots, N.$$

Preuve. — La preuve est obtenue par une modification simple de celle du théorème 1 ; l'on minore et majore ici encore le rapport $\ln(W_n/W_0)$, où pour $t \geqslant 0$, $W_t = w_{1,t} + \cdots + w_{N,t}$. D'une part, et comme précédemment, on a

$$\ln \frac{W_n}{W_0} \geqslant -\eta \min_{i=1,\ldots,N} \widetilde{L}_{i,n} - \ln N \ .$$

D'autre part, pour tout $t = 1, \ldots, n$,

$$\begin{aligned}
\ln \frac{W_t}{W_{t-1}} &= \ln \sum_{i=1}^{N} p_{i,t} e^{-\eta \widetilde{\ell}(i,Y_t)} \\
&\leqslant \ln \sum_{i=1}^{N} p_{i,t} \left(1 - \eta\, \widetilde{\ell}(i,Y_t) + \frac{\eta^2}{2}\, \widetilde{\ell}^2(i,Y_t) \right) \\
&\leqslant -\eta \sum_{i=1}^{N} p_{i,t}\, \widetilde{\ell}(i,Y_t) + \frac{\eta^2}{2} \sum_{i=1}^{N} p_{i,t}\, \widetilde{\ell}^2(i,Y_t) \\
&\leqslant -\eta \widetilde{\ell}(\mathbf{p}_t, Y_t) + \frac{\eta^2}{2\epsilon}\, \widetilde{\ell}(\mathbf{p}_t, Y_t)
\end{aligned}$$

où l'on a utilisé successivement $e^{-x} \leqslant 1 - x + x^2/2$ pour $x \geqslant 0$, $\ln(1+x) \leqslant x$ pour $x \geqslant -1$ et $\widetilde{\ell}(i,Y_t) \in [0, 1/\epsilon]$.
En additionnant l'inégalité ci-dessus selon $t = 1, \ldots, n$ et en la combinant avec le minorant obtenu pour $\ln(W_n/W_0)$, il vient pour tout $i = 1, \ldots, N$,

$$\sum_{t=1}^{n} \widetilde{\ell}(\mathbf{p}_t, Y_t) - \widetilde{L}_{i,n} \leqslant \frac{\eta}{2\epsilon} \sum_{t=1}^{n} \widetilde{\ell}(\mathbf{p}_t, Y_t) + \frac{\ln N}{\eta} \ .$$

On utilise alors $\mathbb{E}\left[\sum_{t=1}^{n} \widetilde{\ell}(\mathbf{p}_t, Y_t)\right] = \mathbb{E}\, \widehat{L}_n \leqslant n$, de sorte qu'en passant aux espérances dans les deux membres, on a obtenu, pour tout $i = 1, \ldots, N$,

$$\mathbb{E}\left[\widehat{L}_n - L_{i,n}\right] \leqslant \frac{n\eta}{2\epsilon} + \frac{\ln N}{\eta} \ .$$

On conclut la preuve en remplaçant η par la valeur $\sqrt{(2\epsilon \ln N)/n}$ proposée. □

Le théorème 3 assure que l'espérance du regret, rapportée au nombre de tours, converge vers zéro dès que $m \to \infty$ lorsque $n \to \infty$. Cependant, une borne sur l'espérance du regret ne donne que peu d'informations sur le comportement réel du regret (non moyenné) $\widehat{L}_t - \min_{i=1,\ldots,N} L_{i,t}$. Le théorème précédent n'exclut pas que le regret fluctue énormément autour de sa valeur moyenne. On rappelle en effet que les résultats Y_t choisis sont donnés par des fonctions (arbitrairement complexes) des actions passées I_s du pronostiqueur et de ses demandes d'observations Z_s, $s < t$. Un choix plus fin des paramètres ε et η permet de montrer que le regret ne dépasse que peu sa valeur moyenne et

qu'il est, avec grande probabilité, borné par une quantité proportionnelle à $n\sqrt{(\ln N)/m}$.

THÉORÈME 4. — *Pour un horizon n et un niveau $\delta \in (0,1)$ fixés, on considère la stratégie de prédiction économe utilisant les paramètres*

$$\varepsilon = \max\left\{0, \frac{m - \sqrt{2m\ln(4/\delta)}}{n}\right\} \qquad \text{et} \qquad \eta = \sqrt{\frac{2\varepsilon \ln N}{n}}\,.$$

Avec probabilité plus grande que $1-\delta$, elle n'accède pas à plus de m résultats et son regret est borné selon

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

On omet la preuve détaillée du théorème 4 et on indique simplement qu'elle suit celle du théorème 3, sauf qu'au lieu de passer aux espérances lors de la combinaison des bornes inférieure et supérieure, on recourt à une inégalité de martingales, à savoir, l'inégalité de Bernstein pour les suites bornées d'accroissements de martingales (voir, par exemple, Freedman [23], Neveu [51]). Une preuve détaillée est donnée par [16] et [15].

Les stratégies de prédiction des théorèmes 3 et 4 ne sont pas en elles-mêmes cohérentes au sens de Hannan, notamment à cause de leur dépendance en la longueur n de la suite des résultats. Cependant, la cohérence peut être atteinte en mettant en œuvre les techniques désormais usuelles de réglage dynamique des paramètres. La quantité alors principalement en jeu dans cette extension est le rythme d'accès aux résultats μ (on rappelle que $\mu(n)$ est le nombre de requêtes d'accès pouvant avoir été satisfaites au tour n). Le résultat suivant montre que la cohérence est atteignable dès que $\mu(n)/(\log(n)\log\log(n)) \to \infty$. On pourra consulter [16] ou [15] pour en avoir une preuve.

COROLLAIRE 1. — *Soit $\mu : \mathbb{N} \to \mathbb{N}$ une fonction croissante à valeurs entières telle que*

$$\lim_{n\to\infty} \frac{\mu(n)}{\log(n)\log\log(n)} = \infty\,.$$

Alors il existe une stratégie de prédiction économe et cohérente au sens de Hannan qui ne demande à voir, et ce, pour tout $n \in \mathbb{N}$, qu'au plus $\mu(n)$ résultats lors des n premiers tours de prédiction.

Le dernier théorème de cette partie montre qu'il existe un problème de prédiction pour lequel le regret (en espérance) de toute stratégie de prédiction n'accédant pas à plus de m résultats pendant les n premiers tours de prédiction est plus grand, à une constante universelle près, que $n\sqrt{\ln N/m}$, ce qui correspond à l'ordre de grandeur des bornes supérieures établies ci-dessus. En d'autres termes, le meilleur ordre de grandeur du regret rapporté au nombre

de tours de prédiction est $\sqrt{\ln N/m}$. Chose intéressante, ce rapport ne dépend pas du nombre n de tours, seulement du nombre m d'observations auxquelles le pronostiqueur a accédé. On rappelle que dans le cas de la prédiction en information complète le meilleur rapport était de l'ordre de $\sqrt{\ln N/n}$. Une interprétation intuitive est que m joue le rôle de «taille d'échantillon» dans le problème de prédiction économe.

THÉORÈME 5. — *Pour tout entier $N \geqslant 2$, il existe une fonction de perte ℓ telle que pour tout ensemble de N actions et pour tout $n \geqslant m \geqslant 4(\sqrt{3}-1)^2 \log_2 N$, le regret en espérance de toute stratégie de prédiction contrainte à n'accéder qu'à au plus m observations dans un jeu de prédiction à n tours est minoré par*

$$\sup_{y^n \in \{0,1\}^n} \left(\mathbb{E}\, \widehat{L}_n - \min_{i=1,\dots,N} L_{i,n} \right) \geqslant cn \sqrt{\frac{\lfloor \log_2 N \rfloor}{m}}$$

où $c = \big((\sqrt{3}-1)/8\big) e^{-\sqrt{3}/2} > 0.0384$.

La preuve, comme celle du théorème 2, repose sur un choix aléatoire des pertes, bien que l'argument précis se révèle ici significativement plus délicat à mettre en œuvre. On renvoie ici encore le lecteur à [16] et [15], où deux preuves différentes sont détaillées.

4. Prédiction en information imparfaite (jeux avec signaux)

Dans nombre de problèmes cruciaux de prédiction, le pronostiqueur ignore les pertes qu'il a essuyées dans le passé et n'a reçu en leur lieu et place que des répercussions assez peu informatives sur elles. De telles situations sont souvent appelées problèmes de prédiction en *information imparfaite*.

Avant de décrire formellement le cadre mathématique de tels problèmes, on illustre le suivi partiel des résultats de la prédiction par l'exemple de l'ajustement séquentiel des prix de vente. Un vendeur dispose d'un produit qu'il propose à une suite de clients, servis l'un après l'autre. Pour chacun d'eux, il fixe un prix choisi (pour fixer les idées) dans $\{1, \dots, N\}$. Le client décide alors d'acheter ou non le produit. Aucun marchandage n'est possible et aucune autre information n'est échangée entre l'acheteur et le vendeur. Ce dernier cherche à réaliser un chiffre d'affaires presque aussi élevé que s'il avait su le prix optimal que les clients auraient été prêts à payer. De ce fait, s'il l'on note I_t le prix proposé au t-ième client et $Y_t \in \{1, \dots, N\}$ le prix maximal que ce dernier avait en tête, la perte du vendeur lors du passage du t-ième client est

$$\ell(I_t, Y_t) = \frac{(Y_t - I_t)\mathbb{I}_{\{I_t \leqslant Y_t\}} + c\mathbb{I}_{\{I_t > Y_t\}}}{N}$$

où $0 \leqslant c \leqslant N$ est défini plus loin. En réalité, lorsque client réalise l'achat, c'est-à-dire, quand $I_t \leqslant Y_t$, la perte du vendeur est un manque à gagner

correspondant à la différence entre le plus grand prix possible pour l'achat et le prix proposé ; la perte subie est identiquement égale à c/N lorsque le client n'achète pas le produit, et elle correspond par exemple aux frais de stockage. (Le facteur $1/N$ n'est là que pour assurer que les pertes sont entre 0 et 1.) Une autre mesure de la perte consiste à remplacer la constante c par Y_t. Dans les deux cas, si le vendeur connaissait à l'avance la suite des Y_t, il pourrait fixer un prix constant $i \in \{1, \ldots, N\}$ qui minimise sa perte totale. La question que l'on examine dans cette partie est de savoir s'il existe une stratégie de prédiction (randomisée) pour le vendeur telle que son regret

$$\sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{i=1,\ldots,N} \sum_{t=1}^{n} \ell(i, Y_t)$$

soit assuré d'être un $o(n)$ quelle que soit la suite $Y_1, Y_2, \ldots$ des prix maximaux des clients – c'est-à-dire qu'on cherche une stratégie de prédiction cohérente au sens de Hannan. La difficulté ici est que les seules informations à disposition du vendeur (du pronostiqueur) sont les positions respectives des I_t et des Y_t, mais pas leurs valeurs précises, ni même les valeurs des pertes $\ell(I_t, Y_t)$. (On note que dévoiler Y_t au pronostiqueur ramènerait le problème à un jeu de prédiction en information complète.)

On traite ces problèmes de répercussions restreintes sur les prédictions (ou de prédiction en *information imparfaite*) dans un cadre plus général, décrit ci-après, dont l'ajustement séquentiel des prix de vente est un cas particulier. On suppose ici que l'ensemble des résultats possibles est un ensemble fini, contenant M éléments. Sans perte de généralité, on le note $\mathcal{Y} = \{1, \ldots, M\}$. La fonction de perte ℓ étant définie sur un produit d'ensembles finis, on peut la représenter matriciellement, par $\mathbf{L} = [\ell(i,j)]_{N \times M}$; elle est supposée connue par le pronostiqueur. Lorsque ce dernier, au tour t, choisit l'action $I_t \in \{1, \ldots, N\}$ et que le résultat est $Y_t \in \mathcal{Y}$, il subit une perte $\ell(I_t, Y_t)$. Cependant, en lieu et place du résultat Y_t, il n'observe que la répercussion sur prédiction $h(I_t, Y_t)$, où h est une *fonction de répercussion* connue par le pronostiqueur, qui associe à chaque couple d'action et de résultat un élément d'un ensemble fini $\mathcal{S} = \{s_1, \ldots, s_m\}$ de *signaux*. La fonction h est représentée de même par une *matrice des répercussions* $\mathbf{H} = [h(i,j)]_{N \times M}$. Ici, comme dans les parties précédentes, on permet à l'environnement de choisir le résultat Y_t en fonction des actions passées $I_1, \ldots, I_{t-1}$ du pronostiqueur.

Le jeu de prédiction en information imparfaite peut être décrit comme dans l'encadré de la page suivante.

La notion d'information imparfaite prend sa source en théorie des jeux et a été considérée, entre autres, par Mertens, Sorin et Zamir [50], Rustichini [54], et Mannor et Shimkin [45]. Notons que certains auteurs s'intéressent à un cadre plus général dans lequel les répercussions peuvent être aléatoires ; pour la clarté de l'exposé et parce que les résultats présentés dans cette partie s'étendent au cas de signaux aléatoires (voir [17]), nous supposons les signaux déterministes. Weissman et Merhav [61], de même que Weissman, Merhav et Somekh-Baruch [62], considèrent différents problèmes de prédiction dans lesquels le pronostiqueur observe seulement une version bruitée des résultats

Prédiction en information imparfaite

Paramètres : N actions, l'ensemble fini $\mathcal{Y} = \{1, \ldots, M\}$ des résultats possibles, la fonction de perte ℓ, la fonction de répercussion h.

À chaque tour $t = 1, 2 \ldots$,

(1) l'environnement choisit le résultat $Y_t \in \mathcal{Y}$ à venir, sans le dévoiler ;
(2) le pronostiqueur détermine son action $I_t \in \{1, \ldots, N\}$;
(3) le pronostiqueur et chaque action i subissent respectivement les pertes $\ell(I_t, Y_t)$ et $\ell(i, Y_t)$, mais ces valeurs demeurent inconnues au pronostiqueur pour l'instant ;
(4) seul le signal $h(I_t, Y_t)$ est communiqué au pronostiqueur.

réels. Ils peuvent être vus comme des cas particuliers de prédiction en information imparfaite avec signaux aléatoires.

Piccolboni et Schindelhauer [52] traitent la prédiction en information imparfaite comme un problème de prédiction séquentielle. Cesa-Bianchi, Lugosi et Stoltz [17] étendent les résultats de [52] et s'intéressent aux ordres de grandeur optimaux du regret. On pourra également consulter Auer et Long [3] pour une étude de certains cas particuliers d'information imparfaite en prédiction.

Il est intéressant mais complexe de déterminer le potentiel d'un pronostiqueur à qui l'on ne fournit qu'une quantité limitée d'information sur les résultats de ses prédictions. On peut par exemple se demander sous quelles conditions il est possible d'obtenir des stratégies cohérentes au sens de Hannan. Naturellement, cela dépend des liens entre la matrice des pertes et celle des répercussions. On note qu'un algorithme de prédiction est libre de coder les valeurs $h(i, j)$ de la fonction de répercussion par des nombres réels. La seule restriction à imposer est que si, à i fixé, $h(i, j) = h(i, j')$ alors les codes réels correspondants doivent être égaux. Pour éviter les ambiguïtés éventuellement provoquées par changements d'échelle, on suppose dans la suite que $|h(i, j)| \leqslant 1$ pour tout couple (i, j). Ainsi, on fixe un encodage de $\mathbf{H} = [h(i, j)]_{N \times M}$ par une matrice de réels entre -1 et 1 et on garde à l'esprit que l'algorithme de prédiction peut toujours remplacer cette matrice par $\mathbf{H}_\phi = [\phi_i(h(i, j))]_{N \times M}$, où il n'y a pas de contraintes sur les fonctions $\phi_i : [-1, 1] \to [-1, 1]$, $i = 1, \ldots, N$, en jeu. L'ensemble $\mathcal{S}$ des signaux peut être choisi de sorte qu'il ait $m \leqslant M$ éléments, même si après l'encodage numérique, la matrice $\mathbf{H}$ peut avoir jusqu'à MN éléments distincts.

Voici quelques exemples concrets, avant que l'on ne décrive une stratégie générale de prédiction.

Exemple 1 (Jeux de bandits manchots.) — Dans de nombreux problèmes de prédiction, le pronostiqueur est capable de mesurer, après avoir agi, le montant de sa perte (ou de son gain), mais n'a pas accès à ce qu'il serait

advenu s'il avait choisi une autre action. On appelle de tels problèmes les jeux de bandits manchots. La petite histoire à l'origine de leur dénomination est la suivante; un exemple typique est en effet fourni par un joueur de casino ayant à sa disposition différentes machines à sous (également appelées bandits manchots). Il parie plusieurs fois de suite, en changeant de machine lorsqu'il le veut, et il cherche à gagner presque autant de pièces que s'il avait su à l'avance quelle machine lui rapporterait le plus. Les jeux de bandits manchots sont des cas particuliers de prédiction en information imparfaite. Il suffit de prendre $\mathbf{H} = \mathbf{L}$ pour le voir, c'est-à-dire que la seule information repercutée au pronostiqueur est le montant de sa propre perte. Ils ont été décrits à l'origine dans un cadre stochastique (voir Robbins [53] et Lai et Robbins [42]) ; plusieurs variantes du problème de départ ont été étudiées, par exemple par Berry et Fristedt [7] et Gittins [25]. Le cadre non stochastique considéré ici a été étudié en premier lieu par Baños [6], voir également Megiddo [47]. Des stratégies cohérentes au sens de Hannan ont été construites par Foster et Vohra [21], Auer, Cesa-Bianchi, Freund et Schapire [2], Hart et Mas Colell [31, 33], voir également Fudenberg et Levine [24]. Nous reviendrons sur une extension du problème des bandits manchots dans la partie 5.

Exemple 2 (Ajustement séquentiel des prix de vente). — Le problème de l'ajustement séquentiel des prix de vente décrit en introduction de cette partie correspond à $M = N$ et à une matrice des pertes égale à

$$\mathbf{L} = [\ell(i,j)]_{N\times N} \qquad \text{où} \qquad \ell(i,j) = \frac{(j-i)\mathbb{I}_{\{i\leqslant j\}} + c\mathbb{I}_{\{i>j\}}}{N} \,.$$

L'information dont le pronostiqueur (en l'occurrence, le vendeur) dispose est simplement si le prix I_t qu'il a fixé est plus grand que le prix du client Y_t, ou non, de sorte que les entrées dans la matrice des répercussions sont données par $h(i,j) = \mathbb{I}_{\{i>j\}}$ – ou, après ré-encodage idoine,

$$h(i,j) = a\mathbb{I}_{\{i\leqslant j\}} + b\mathbb{I}_{\{i>j\}} \qquad i,j = 1,\dots,N$$

où a et b sont des constantes choisies par le pronostiqueur, avec $a, b \in [-1,1]$.

Exemple 3 (Dégustation de pommes). — On considère l'exemple élémentaire de prédiction binaire dans lequel $N = M = 2$, et les matrices de pertes et répercussions sont respectivement données par

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{et} \qquad \mathbf{H} = \begin{bmatrix} a & a \\ b & c \end{bmatrix} \,.$$

Ici, le pronostiqueur ne reçoit d'information sur le résultat Y_t que lorsqu'il choisit la seconde action. On appelle cet exemple le *problème de la dégustation de pommes.* (On imagine en effet qu'il s'agit de déterminer si des pommes doivent être considérées comme «bonnes à vendre» ou «pourries». Une pomme étiquetée «pourrie» peut être ouverte avant d'être jetée, histoire de voir si elle renfermait effectivement un ver; mais comme une pomme ouverte

ne peut être vendue, on ne peut jamais vérifier si une pomme déclarée bonne pour la vente contient un ver ou non.) La dégustation de pommes a été étudiée par Helmbold, Littlestone et Long [35] dans le cas particulier où l'une des deux actions a une perte cumulée nulle.

Exemple 4 (prédiction économe). — On peut également considérer une variante du problème de prédiction économe comme une situation de prédiction en information imparfaite. On prend $N = 3$, $M = 2$, et des matrices de pertes et répercussions de la forme

$$\mathbf{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{et} \qquad \mathbf{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix} .$$

Dans cet exemple, les seuls instants où le pronostiqueur reçoit une répercussion informative sont ceux où il a joué la première action, auquel cas il a subi la perte maximale 1, quel qu'ait été le résultat. Ainsi, tout comme dans le problème de prédiction économe, on ne peut choisir l'action « informative » qu'un nombre limité de fois, sans quoi il n'est pas permis d'espérer atteindre la cohérence au sens de Hannan.

On présente maintenant une stratégie générale pour la prédiction en information imparfaite ; elle est cohérente au sens de Hannan dès que la matrice des répercussions contient « suffisamment d'information » sur celle des pertes.

Cette stratégie utilise la prédiction par pondération exponentielle, et, comme dans le cas de la prédiction économe en nombre d'observations, la route du succès est ouverte par le remplacement des pertes $\ell(i, Y_t)$ par des estimations sans biais dans ladite pondération exponentielle.

Cependant, il n'est possible de construire de telles estimations que sous certaines conditions sur les matrices de pertes et répercussions. L'hypothèse centrale est qu'en un certain sens, les pertes puissent être reconstituées à partir des répercussions, ce que l'on formalise de la manière suivante : il existe une matrice $\mathbf{K} = [k(i,j)]_{N \times N}$ telle que $\mathbf{L} = \mathbf{K}\,\mathbf{H}$; à savoir,

$$\mathbf{H} \quad \text{et} \quad \begin{bmatrix} \mathbf{H} \\ \mathbf{L} \end{bmatrix}$$

ont même rang. En d'autres mots, on peut écrire, pour tout $i \in \{1, \ldots, N\}$ et tout $j \in \{1, \ldots, M\}$, que

$$\ell(i,j) = \sum_{l=1}^{N} k(i,l) h(l,j) \ .$$

Dans ce cas, on définit les estimations $\widetilde{\ell}$ des pertes par

$$\widetilde{\ell}(i, Y_t) = \frac{k(i, I_t) h(I_t, Y_t)}{p_{I_t, t}} \qquad i = 1, \ldots, N$$

et on note leurs sommes $\widetilde{L}_{i,n} = \widetilde{\ell}(i, Y_1) + \cdots + \widetilde{\ell}(i, Y_n)$. Soit $\mathbb{E}_t$ l'espérance conditionnelle sachant $I_1, \ldots, I_{t-1}$ (*id est*, l'espérance par rapport au choix aléatoire de I_t selon $\mathbf{p}_t$) ; on observe

$$\begin{aligned}\mathbb{E}_t \widetilde{\ell}(i, Y_t) &= \sum_{j=1}^{N} p_{j,t} \frac{k(i,j)h(j,Y_t)}{p_{j,t}} \\ &= \sum_{j=1}^{N} k(i,j)h(j,Y_t) = \ell(i, Y_t)\end{aligned}$$

et par conséquent, $\widetilde{\ell}(i, Y_t)$ est un estimateur (conditionnellement) sans biais de la perte $\ell(i, Y_t)$.

La stratégie de prédiction en information imparfaite est alors définie de la manière suivante, comme suggéré initialement par Piccolboni et Schindelhauer [52].

Stratégie de prédiction en information imparfaite

Paramètres : la matrice des pertes $\mathbf{L}$, la matrice des répercussions $\mathbf{H}$, une matrice $\mathbf{K}$ telle que $\mathbf{L} = \mathbf{K}\,\mathbf{H}$, deux nombres réels $0 < \eta, \gamma < 1$.

Initialisation : $\mathbf{w}_0 = (1, \ldots, 1)$.

À chaque tour $t = 1, 2, \ldots$,

(1) tirer une action $I_t \in \{1, \ldots, N\}$ selon la probabilité

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

(2) prendre connaissance de la répercussion $h_t = h(I_t, Y_t)$ et former l'estimation $\widetilde{\ell}_{i,t} = k(i, I_t)h_t / p_{I_t,t}$ pour tout $i = 1, \ldots, N$;

(3) calculer $w_{i,t} = w_{i,t-1} e^{-\eta \widetilde{\ell}(i, Y_t)}$ pour tout $i = 1, \ldots, N$.

Le résultat suivant, prouvé par Cesa-Bianchi, Lugosi et Stoltz [17], majore le regret de la stratégie de prédiction définie ci-dessus. Il exprime que le regret rapporté au nombre de tours de jeu, $\frac{1}{n}\left(\widehat{L}_n - \min_{i=1,\ldots,N} L_{i,n}\right)$, décroît vers zéro à une vitesse $n^{-1/3}$. Cette dernière est significativement plus lente que la vitesse optimale $n^{-1/2}$ obtenue dans le cas de l'information complète. On verra par la suite que cette vitesse $n^{-1/3}$ ne peut être améliorée en général : cette détérioration de la vitesse de convergence est le prix à payer pour n'avoir accès qu'à une certaine répercussion sur les résultats réels de la prédiction en lieu et place de ces derniers. Cependant, il est encore possible d'obtenir la cohérence au sens de Hannan dès que la condition $\mathbf{L} = \mathbf{K}\,\mathbf{H}$ est remplie, comme l'indiquent les commentaires qui suivent le théorème ci-dessous.

THÉORÈME 6. — *Soit un problème de prédiction en information imparfaite* $(\mathbf{L}, \mathbf{H})$ *tel que les matrices de pertes et répercussions vérifient* $\mathbf{L} = \mathbf{K}\mathbf{H}$ *pour une certaine matrice* $\mathbf{K}$ *de taille* $N \times N$, *pour laquelle on note* $k^* = \max\{1, \max_{i,j} |k(i,j)|\}$. *On fixe* $\delta \in (0,1)$. *Alors le regret de la stratégie de prédiction en information imparfaite décrite ci-dessus et utilisant les paramètres*

$$\eta = \left(\frac{\ln N}{2Nnk^*}\right)^{2/3} \qquad \text{et} \qquad \gamma = \left(\frac{(k^*N)^2 \ln N}{4n}\right)^{1/3}$$

vérifie que, pour tout

$$n \geqslant \frac{1}{k^*N}\left(\ln\frac{N+3}{\delta}\right)^{3/2}$$

et avec probabilité au moins $1-\delta$,

$$\sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{i=1,\dots,N} \sum_{t=1}^{n} \ell(i, Y_t)$$
$$\leqslant 10\,(Nnk^*)^{2/3}\,(\ln N)^{1/3}\sqrt{\ln\frac{2(N+3)}{\delta}}\,.$$

La preuve est très semblante à celle du théorème 4, et c'est pourquoi on en omet les détails, en priant le lecteur de consulter [17] ou [15] le cas échéant. Le théorème 6 assure donc l'existence d'une stratégie dont le regret est au plus de l'ordre de $n^{2/3}$ dès lors que le rang de la matrice $\mathbf{H}$ des répercussions n'est pas inférieur à celui de

$$\begin{bmatrix} \mathbf{H} \\ \mathbf{L} \end{bmatrix}\,.$$

(Et les techniques usuelles de réglage dynamique des paramètres indiquent, là encore, les modifications simples à effectuer pour rendre la stratégie cohérente au sens de Hannan.) Cette condition sur les rangs est vérifiée pour de nombreux problèmes, et pour illustrer notre propos, nous reprenons un à un les exemples proposés précédemment.

Exemple 5 (Jeux de bandits manchots, suite). — On rappelle que dans un jeu de bandits manchots, $\mathbf{H} = \mathbf{L}$, de sorte que les conditions du théorème 6 sont clairement satisfaites; on peut en effet prendre pour $\mathbf{K}$ la matrice identité, et poser $k^* = 1$. Cependant, dans le cas présent, des bornes bien plus petites que celles du théorème 6 peuvent être obtenues par une modification convenable de la stratégie générale. Auer, Cesa-Bianchi, Freund et Schapire [2] (voir aussi Auer [1]) montrent en l'occurrence qu'un regret de l'ordre de $n^{1/2}$ peut être atteint. Plus précisément, il est prouvé dans [15] que pour tout $\delta \in (0,1)$ et $n \geqslant 8N\ln(N/\delta)$, on a, avec probabilité au moins $1-\delta$,

$$\widehat{L}_n - \min_{i=1,\dots,N} L_{i,n} \leqslant \frac{11}{2}\sqrt{nN\ln(N/\delta)} + \frac{\ln N}{2}\,. \tag{2}$$

Exemple 6 (Ajustement séquentiel des prix de vente, suite). — Dans le problème d'ajustement séquentiel des prix de vente décrit en introduction de cette partie, la matrice des répercussions est donnée par $h(i,j) = a\mathbb{I}_{\{i\leqslant j\}} + b\mathbb{I}_{\{i>j\}}$, où a et b sont deux réels distincts choisis arbitrairement dans $[-1,1]$. Pour $a \neq 0$, cette matrice $\mathbf{H}$ est inversible et l'on peut par conséquent prendre $\mathbf{K} = \mathbf{L}\,\mathbf{H}^{-1}$ pour satisfaire aux conditions du théorème et obtenir, partant, une stratégie cohérente au sens de Hannan, avec un regret au plus de l'ordre de $n^{2/3}$. C'est ainsi que l'on a obtenu un moyen pour le vendeur de fixer ses prix I_t de manière dynamique de telle sorte que sa perte ne soit pas tellement plus grande que celle qu'il aurait subie s'il avait su, et à l'avance, la suite des prix Y_t des clients, et s'était décidé pour le prix constant optimal face à cette suite. Le choix numérique de $a = 1$ et $b = 0$ pour l'encodage des répercussions conduit à $k^* = 1$, ce dernier est donc indépendant de N. Ainsi, l'ordre de grandeur en n et N de la borne sur le regret de la stratégie générale de prédiction en information imparfaite est $(nN\log N)^{2/3}$.

Exemple 7 (Dégustation de pommes, suite). — Dans le problème de la dégustation de pommes décrit ci-dessus, on peut choisir l'encodage $a = b = 1$ et $c = 0$ pour les répercussions. On rend alors la matrice des répercussions inversible, et une fois encore, le théorème 6 s'applique.

Exemple 8 (Prédiction économe, suite). — On rappelle que le problème de prédiction économe vu comme problème de prédiction en information imparfaite correspond aux matrices de pertes et répercussions

$$\mathbf{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{et} \qquad \mathbf{H} = \begin{bmatrix} a & b \\ c & c \\ c & c \end{bmatrix}.$$

Comme le rang de $\mathbf{L}$ vaut ici 2, il suffit d'encoder la matrice des répercussions de telle sorte que son rang vaille 2, par exemple en choisissant $a = 1/2$, $b = 1$, et $c = 1/4$. On a alors $\mathbf{L} = \mathbf{K}\,\mathbf{H}$ pour

$$\mathbf{K} = \begin{bmatrix} 0 & 2 & 2 \\ 2 & -2 & -2 \\ -2 & 4 & 4 \end{bmatrix}.$$

Remarque 2 (Experts, ou actions multiples). — Lors de la définition du problème de prédiction en information imparfaite, on s'est restreint aux pronostiqueurs cherchant à prédire presque aussi bien que la meilleure action constante. Cela transparaît dans la définition du regret ; la perte cumulée du pronostiqueur y est en effet comparée à $\min_{i=1,\dots,N}\sum_{t=1}^{n}\ell(i,Y_t)$, à savoir, la perte cumulée de la meilleure action constante. On peut cependant définir également le regret en termes d'une classe d'experts (ou d'actions multiples). Un tel expert peut être identifié à une suite $\mathbf{i} = (i_1,\dots,i_n)$ d'actions $i_t \in \{1,\dots,N\}$. La définition du regret tenant compte d'une telle classe $\mathcal{S}$

d'experts est

$$\sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{i \in \mathcal{S}} \sum_{t=1}^{n} \ell(i_t, Y_t) \ .$$

Il est facile d'étendre les stratégies de prédiction obtenues dans le problème simple de prédiction en information imparfaite décrit dans cette partie au cas plus général des actions multiples. En particulier, lorsque $\mathcal{S}$ contient M experts, on obtient une borne de $k^{*2/3}(Nn)^{2/3}(\ln M)^{1/3}$ sur le regret.

L'exemple 8, celui de la prédiction économe, révèle que la borne du théorème 6 ne peut être améliorée significativement en général. Plus précisément, dans cet exemple (et dans d'autres où la condition $\mathbf{L} = \mathbf{K}\,\mathbf{H}$ est remplie), le regret est borné inférieurement par une quantité de l'ordre de $n^{2/3}$. On prouve cela en adaptant la preuve du théorème 5 ; voir [17] pour les détails de l'adaptation.

On peut se demander sous quelles conditions (idéalement minimales) sur les matrices de pertes et répercussions il est possible d'obtenir des stratégies cohérentes au sens de Hannan. Le théorème 6 indique que ceci est le cas dès lors qu'il existe un encodage des répercussions tel que le rang de la matrice

$$\begin{bmatrix} \mathbf{H} \\ \mathbf{L} \end{bmatrix}$$

obtenue par concaténation de celle des pertes et de celle des répercussions n'est pas plus grand que le rang de cette dernière. Cette condition suffisante n'est cependant pas nécessaire. Pour une discussion détaillée et des exemples supplémentaires, on renvoie le lecteur à [52], [17] et [15].

5. Plus court chemin dans un jeu de bandits manchots

Les résultats présentés dans la partie 2 (Théorème 1) montrent qu'il est possible de construire des stratégies de prédiction séquentielle dont la performance sur toute suite arbitraire d'observations est presque aussi bonne que celle du meilleur, sur cette même suite, parmi N experts (ou actions) ; à savoir, que la perte cumulée de la stratégie de prédiction, rapportée au nombre n de tours, est au pire égale à celle du meilleur expert plus une quantité proportionnelle à $\sqrt{\ln N/n}$, ceci valant quelle que soit la fonction de perte bornée en jeu. La dépendance logarithmique en le nombre N d'experts à disposition permet d'obtenir des bornes utiles même lorsque N est grand. Cependant, les stratégies de prédiction les plus élémentaires, telles celles par pondération selon un potentiel, ont une complexité algorithmique proportionnelle au nombre d'experts, et sont ainsi impossibles à mettre en œuvre lorsque ce dernier est très grand.

Cela dit, dans de nombreuses applications, l'ensemble des experts a une certaine structure qui peut être exploitée pour construire des algorithmes de prédiction efficaces. Une liste non exhaustive de références à l'exploitation d'une structuration sous-jacente contient Herbster et Warmuth [38], Vovk [59],

Bousquet et Warmuth [10], Helmbold et Schapire [37], Takimoto et Warmuth [56], Kalai et Vempala [41], György, Linder et Lugosi [26, 27 ,28]. On pourra également consulter, pour obtenir une vue plus globale, Cesa-Bianchi et Lugosi [15, chapitre 5]. L'abondance de la littérature sur ce point montre que ce problème a suscité un grand intérêt dans la théorie algorithmique de l'apprentissage.

On étudie dans cette partie le problème de trouver le plus court chemin entre deux points ; c'est un exemple représentatif de problème structuré, qui a retenu l'attention pour ses applications nombreuses.

On considère un réseau représenté par un ensemble de nœuds reliés par des arcs, et on suppose que l'on doit envoyer un flux de paquets d'un nœud source vers un nœud de destination. A chaque pas de temps, un paquet est envoyé selon un parcours (à choisir) reliant la source à la destination. Chaque arc du réseau a un temps de parcours propre, qui dépend de la circulation locale ; et la durée totale que met le paquet pour parcourir tout le trajet est la somme des durées associées aux arcs formant ledit trajet. Ces durées peuvent changer de manière arbitraire à chaque pas de temps, et on cherche un moyen de déterminer les chemins de telle sorte que la somme des temps de parcours des paquets sur les chemins choisis ne soit pas tellement plus grande que sur le meilleur trajet du réseau.

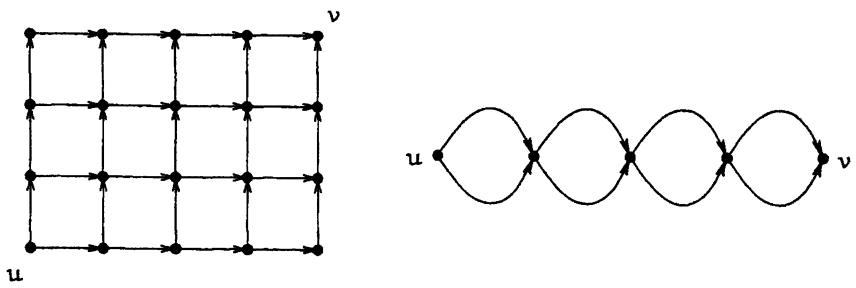

FIG 1. — Deux exemples de graphes orientés et acycliques, pour le problème du plus court chemin

Naturellement, ce problème peut être formulé comme un problème de prédiction séquentielle dans lequel chaque trajet possible correspond à une action. Cependant, comme le nombre de chemins possibles est typiquement exponentiel en le nombre d'arcs, cela réclame des algorithmes dont la mise en œuvre est algorithmiquement rationnelle. Deux solutions d'inspirations très différentes ont été proposées. L'une d'entre elles se fonde sur la stratégie du « meilleur expert bruité », voir Kalai et Vempala [41], tandis que l'autre repose sur une implémentation efficace de la stratégie de prédiction par pondération exponentielle, voir, par exemple, Takimoto et Warmuth [56]. Chacune des deux solutions a ses avantages et peut être étendue dans différentes directions.

On formalise le problème en considérant un graphe (fini), orienté et acyclique, défini par un ensemble V de sommets, et un ensemble d'arcs $E = \{e_1, \ldots, e_{|E|}\}$. Ainsi, chaque arc $e \in E$ est une paire ordonnée de sommets,

$e = (v_1, v_2)$. Soient u et v deux sommets fixés dans V. Un *chemin* de u vers v est une suite $e^{(1)}, \ldots, e^{(k)}$ d'arcs s'écrivant sous la forme $e^{(1)} = (u, v_1)$, $e^{(j)} = (v_{j-1}, v_j)$ pour tout $j = 2, \ldots, k-1$, et $e^{(k)} = (v_{k-1}, v)$. On identifie chaque chemin à un vecteur binaire $\mathbf{i} \in \{0,1\}^{|E|}$ dont la j-ième composante vaut 1 si et seulement si l'arc e_j fait partie du dit chemin. Pour simplifier les choses, on suppose que tout arc de E est dans au moins un chemin de u vers v, et que chaque sommet de V est point d'arrivée (ou de départ) d'au moins un arc de E (voir la figure 1 pour des illustrations graphiques).

À chaque tour $t = 1, \ldots, n$ du jeu de prédiction, le pronostiqueur choisit un chemin $\mathbf{I}_t$ parmi tous les chemins possibles de u vers v. Une perte $\ell_{e,t} \in [0,1]$ est alors assignée à chaque arc $e \in E$. Formellement, on peut identifier le résultat Y_t au vecteur de pertes $\ell_t \in [0,1]^{|E|}$ dont la j-ième composante est $\ell_{e_j,t}$. La perte subie par un chemin $\mathbf{i}$ au tour t est égale à la somme des pertes de chacun des arcs situés sur le chemin, c'est-à-dire que

$$\ell(\mathbf{i}, Y_t) = \mathbf{i} \cdot \ell_t \ .$$

On s'autorise un léger abus de notations en écrivant $e \in \mathbf{i}$ lorsque l'arc $e \in E$ appartient au chemin représenté par le vecteur binaire $\mathbf{i}$. On observe alors que pout tout $t = 1, \ldots, n$ et tout chemin $\mathbf{i}$,

$$\mathbf{i} \cdot \ell_t = \sum_{e \in \mathbf{i}} \ell_{e,t}$$

et que par conséquent, la perte cumulée de tout chemin $\mathbf{i}$ s'écrit comme la somme

$$\sum_{s=1}^{t} \mathbf{i} \cdot \ell_s = \sum_{e \in \mathbf{i}} L_{e,t}$$

où $L_{e,t} = \sum_{s=1}^{t} \ell_{e,s}$ est la perte accumulée par l'arc e pendant les t premiers tours de jeu. Tout comme précédemment, le pronostiqueur peut recourir à des tirages aléatoires auxiliaires, et choisir $\mathbf{I}_t$ au hasard selon une probabilité $\mathbf{p}_t$ sur l'ensemble des chemins de u vers v. On définit le regret comme la différence

$$\sum_{t=1}^{n} \ell(I_t, Y_t) - \min_{\mathbf{i}} \sum_{t=1}^{n} \ell(\mathbf{i}, Y_t)$$

dans laquelle le minimum est pris sur ledit ensemble de tous les chemins de u vers v.

Par exemple, la stratégie de prédiction par pondération exponentielle, employée sur cet ensemble, assure que le regret est borné par

$$\sum_{t=1}^{n} \ell(\mathbf{I}_t, Y_t) - \min_{\mathbf{i}} \sum_{t=1}^{n} \ell(\mathbf{i}, Y_t) \leqslant K \left(\sqrt{\frac{n \ln M}{2}} + \sqrt{\frac{n}{2} \ln \frac{1}{\delta}} \right)$$

avec probabilité au moins $1 - \delta$, où M est le nombre total de chemins de u vers v dans le graphe et K est la longueur du plus long chemin.

On étudie dans cette partie la version « bandits manchots » du problème de trouver le plus court chemin entre deux points ; ici, on suppose – et cela est effectivement plus réaliste dans les applications – que le pronostiqueur n'a accès qu'aux pertes associées aux arcs qui sont sur le chemin qu'il a choisi. Plus précisément, après avoir choisi le chemin $\mathbf{I}_t$ au tour t, la valeur de la perte $\ell_{e,t}$ est portée à sa connaissance si et seulement si $e \in \mathbf{I}_t$.

Ce problème est bien une version des jeux de bandits manchots présentés dans la partie 4. Une application directe de (2) résulte en la borne suivante : avec probabilité au moins $1 - \delta$, et avec les notations précédentes,

$$\sum_{t=1}^{n} \ell(\mathbf{I}_t, Y_t) - \min_{\mathbf{i}} \sum_{t=1}^{n} \ell(\mathbf{i}, Y_t) \leqslant \frac{11K}{2} \sqrt{nM \ln(M/\delta)} + \frac{K \ln M}{2} \,.$$

Cependant, ce majorant sur le regret est inadmissible, parce que cet emploi direct d'une stratégie de prédiction par pondération exponentielle sur les pertes des chemins entraîne une dépendance de l'ordre de $\sqrt{M \ln M}$ en le nombre M de chemins, ce qui est bien plus grand que la dépendance logarithmique, de l'ordre de $\sqrt{\ln M}$, correspondant au cas de la prédiction en information complète. Pour atteindre une borne qui ne croisse pas exponentiellement en le nombre d'arcs du graphe, il est impératif d'utiliser la structure des dépendances entre les pertes des différentes actions (chemins). L'approche d'Awerbuch et Kleinberg [4] et de McMahan et Blum [46] a été d'étendre les techniques de prédiction du type « meilleur expert bruité » au cas des jeux de bandits manchots. Toutefois, les bornes qu'ils ont obtenues ne sont pas du bon ordre de grandeur, à savoir $\sqrt{n}$, en le nombre n de tours de prédiction.

György, Linder, Lugosi et Ottuckák [29] proposent une variante soigneusement définie de la stratégie de prédiction pour les jeux de bandits manchots d'Auer, Cesa-Bianchi, Freund et Schapire [2] ; elle satisfait aux contraintes de performance requises ci-dessus.

Dans ce qui suit, on note $\mathcal{R}$ l'ensemble de tous les chemins de u vers v dans le graphe acyclique (V, E). On considère les *profits* $g_{e,t} = 1 - \ell_{e,t}$, où $e \in E$, et on note $\mathbf{g}_t$ le vecteur formé par l'ensemble des profits au tour t.

On suppose que tous les chemins $\mathbf{i} \in \mathcal{R}$ sont de la même longueur, notée $K > 0$, afin de rendre aisée la conversion du problème de minimisation des pertes en un problème de maximisation des profits à travers l'égalité $\mathbf{i} \cdot \mathbf{g}_t = K - \mathbf{i} \cdot \ell_t$.

Un des traits essentiels de l'algorithme ci-dessous est que les profits sont estimés arc par arc et non pas chemin par chemin. Cette modification est la clé de l'amélioration de la borne sur le regret où le nombre $|E|$ d'arcs remplace le nombre M de chemins. En outre, les techniques de programmation dynamique, telles celles employées par Takimoto and Warmuth [56], conduisent à une implémentation efficace de l'algorithme. Un autre point important est qu'il faut s'assurer que chaque arc est visité suffisamment souvent. A cet effet, on introduit un ensemble $\mathcal{C}$ de *chemins de recouvrement* dont la caractéristique est que pour tout arc $e \in E$, il existe un chemin $\mathbf{i} \in \mathcal{C}$ tel que $e \in \mathcal{C}$. On remarque que l'on peut toujours trouver un tel ensemble de recouvrement, de taille $|\mathcal{C}| \leqslant |E|$.

Stratégie pour trouver le plus court chemin dans un jeu de bandits manchots

Paramètres : trois nombres réels $\beta > 0$, $0 < \eta, \gamma < 1$.

Initialisation : $w_{e,0} = 1$ pour tout $e \in E$; $\mathbf{w}_{\mathbf{i},0} = 1$ pour tout $\mathbf{i} \in \mathcal{R}$; $\overline{W}_0 = |\mathcal{R}|$.

À chaque tour $t = 1, 2, \ldots$,

(1) tirer un chemin I_t selon la probabilité $\mathbf{p}_t$ définie par

$$p_{\mathbf{i},t} = \begin{cases} (1-\gamma)\dfrac{\mathbf{w}_{\mathbf{i},t-1}}{\overline{W}_{t-1}} + \dfrac{\gamma}{|\mathcal{C}|} & \text{si } \mathbf{i} \in \mathcal{C}, \\ (1-\gamma)\dfrac{\mathbf{w}_{\mathbf{i},t-1}}{\overline{W}_{t-1}} & \text{si } \mathbf{i} \notin \mathcal{C}; \end{cases}$$

(2) calculer la probabilité de choisir chacun des arcs e,

$$q_{e,t} = \sum_{\mathbf{i}: e \in \mathbf{i}} p_{\mathbf{i},t} = (1-\gamma)\frac{\sum_{\mathbf{i}: e \in \mathbf{i}} \mathbf{w}_{\mathbf{i},t-1}}{\overline{W}_{t-1}} + \gamma \frac{|\{\mathbf{i} \in \mathcal{C} : e \in \mathbf{i}\}|}{|\mathcal{C}|} ;$$

(3) estimer les profits

$$g'_{e,t} = \begin{cases} \dfrac{g_{e,t} + \beta}{q_{e,t}} & \text{si } e \in I_t, \\ \dfrac{\beta}{q_{e,t}} & \text{sinon}; \end{cases}$$

(4) mettre à jour les poids des arcs et ceux des chemins,

$$w_{e,t} = w_{e,t-1} e^{\eta g'_{e,t}}$$
$$\mathbf{w}_{\mathbf{i},t} = \prod_{e \in \mathbf{i}} w_{e,t} = \mathbf{w}_{\mathbf{i},t-1} e^{\eta g'_{\mathbf{i},t}}$$

où $g'_{\mathbf{i},t} = \sum_{e \in \mathbf{i}} g'_{e,t}$, et déterminer la somme des poids associés aux chemins,

$$\overline{W}_t = \sum_{\mathbf{i} \in \mathcal{R}} \mathbf{w}_{\mathbf{i},t} .$$

Au cours de la preuve, on fera usage des notations

$$G_{\mathbf{i},n} = \sum_{t=1}^{n} \mathbf{i} \cdot \mathbf{g}_t \qquad \text{et} \qquad G'_{\mathbf{i},n} = \sum_{t=1}^{n} g'_{\mathbf{i},t}$$

pour tout $\mathbf{i} \in \mathcal{R}$,

$$\widehat{G}_n = \sum_{t=1}^{n} g_{\mathbf{I}_t,t} ,$$

ainsi que

$$G_{e,n} = \sum_{t=1}^{n} g_{e,t} \qquad \text{et} \qquad G'_{e,n} = \sum_{t=1}^{n} g'_{e,t}$$

pour tout $e \in E$.

THÉORÈME 7. — *Pour tout $\delta \in (0,1)$ et tout choix des paramètres β, γ, η tel que $0 \leqslant \gamma < 1/2$, $0 \leqslant \beta \leqslant 1$, $\eta > 0$, et $2\eta K|\mathcal{C}| \leqslant \gamma$, le regret de la stratégie de prédiction définie ci-dessus est borné, avec probabilité au moins $1-\delta$, par*

$$\widehat{L}_n - \min_{\mathbf{i}\in\mathcal{R}} L_{\mathbf{i},n} \leqslant Kn\gamma + 2\eta n K^2|\mathcal{C}| + \frac{K}{\beta}\ln\frac{|E|}{\delta} + \frac{\ln N}{\eta} + n|E|\beta \ .$$

En particulier, le choix de $\beta = \sqrt{\dfrac{K}{n|E|}\ln\dfrac{|E|}{\delta}}$, $\gamma = 2\eta K|\mathcal{C}|$ et $\eta = \sqrt{\dfrac{\ln N}{4nK^2|\mathcal{C}|}}$ implique que pour tout $n \geqslant \max\left\{\dfrac{K}{|E|}\ln\dfrac{|E|}{\delta}, 4|\mathcal{C}|\ln N\right\}$,

$$\widehat{L}_n - \min_{\mathbf{i}\in\mathcal{R}} L_{\mathbf{i},n} \leqslant 2\sqrt{nK}\left(\sqrt{4K|\mathcal{C}|\ln N} + \sqrt{|E|\ln\frac{|E|}{\delta}}\right) \ .$$

On ne donne ici que les grandes lignes de la preuve et on prie le lecteur intéressé par les détails de consulter [29]. On commence par le lemme suivant.

Lemme 1. — Pour tous $\delta \in (0,1)$, $0 \leqslant \beta \leqslant 1$ et $e \in E$, on a

$$\mathbb{P}\left[G_{e,n} > G'_{e,n} + \frac{1}{\beta}\ln\frac{|E|}{\delta}\right] \leqslant \frac{\delta}{|E|} \ .$$

En outre, pour tout $\mathbf{i} \in \mathcal{R}$,

$$\mathbb{P}\left[G_{\mathbf{i},n} > G'_{\mathbf{i},n} + \frac{K}{\beta}\ln\frac{|E|}{\delta}\right] \leqslant \delta \ .$$

Éléments de preuve. — La deuxième assertion découle de la première en considérant des réunions d'ensembles de probabilité au plus $\delta/|E|$. Pour démontrer la première inégalité, on fixe $e \in E$, et on remarque que l'inégalité de Markov entraîne que, pour tout $u > 0$,

$$\mathbb{P}[G_{e,n} > G'_{e,n} + u] \leqslant e^{-\beta u}\mathbb{E}e^{\beta(G_{e,n}-G'_{e,n})}. \tag{3}$$

Avec le choix de $u = \ln(|E|/\delta)/\beta$, il s'agit de montrer que $\mathbb{E}e^{\beta(G_{e,n}-G'_{e,n})} \leqslant 1$ pour tout n.

À cet effet, on considère $Z_t = e^{\beta(G_{e,n}-G'_{e,n})}$. La réécriture

$$Z_t = Z_{t-1} e^{\beta\left(g_{e,t} - \frac{\mathbb{I}_{\{\{e\in\mathbf{I}_t\}\}} g_{e,t}+\beta}{q_{e,t}}\right)}$$

permet de voir facilement que $\mathbb{E}_t Z_t \leqslant Z_{t-1}$ pour $t \geqslant 2$, où $\mathbb{E}_t$ désigne l'espérance conditionnelle sachant $\mathbf{I}_{t-1}, \ldots, \mathbf{I}_1$; en passant aux espérances, on a alors prouvé $\mathbb{E}Z_t \leqslant \mathbb{E}Z_{t-1}$. Un argument analogue montre $\mathbb{E}Z_1 \leqslant 1$, ce qui implique finalement $\mathbb{E}Z_n \leqslant 1$, comme demandé. □

Eléments de preuve pour le théorème 7. — La preuve commence, comme d'habitude, par la minoration et la majoration du log-rapport $\ln \overline{W}_n/\overline{W}_0$. On obtient la borne inférieure en écrivant simplement

$$\ln \frac{\overline{W}_n}{\overline{W}_0} = \ln \sum_{\mathbf{i}\in\mathcal{R}} e^{\eta G'_{\mathbf{i},n}} - \ln N \geqslant \eta \max_{\mathbf{i}\in\mathcal{R}} G'_{\mathbf{i},n} - \ln N \tag{4}$$

où l'on a utilisé que, par définition, $\mathbf{w}_{\mathbf{i},n} = e^{\eta G'_{\mathbf{i},n}}$.

Pour la borne supérieure, on note tout d'abord que la condition $\eta \leqslant \frac{\gamma}{2K|\mathcal{C}|}$ implique que $\eta g'_{\mathbf{i},t} \leqslant 1$ pour tous $\mathbf{i}$ et t, puisque

$$\eta g'_{\mathbf{i},t} = \eta \sum_{e\in\mathbf{i}} g'_{e,t} \leqslant \eta \sum_{e\in\mathbf{i}} \frac{1+\beta}{q_{e,t}} \leqslant \frac{\eta K(1+\beta)|\mathcal{C}|}{\gamma} \leqslant 1$$

(où la deuxième inégalité découle de ce que $q_{e,t} \geqslant \gamma/|\mathcal{C}|$ pour tout e). Par conséquent, l'inégalité $e^x < 1+x+x^2$ pour tout $x \leqslant 1$ montre, après quelques calculs sans difficulté, que pour tout $t \geqslant 1$,

$$\ln \frac{\overline{W}_t}{\overline{W}_{t-1}} \leqslant \frac{\eta}{1-\gamma} \sum_{\mathbf{i}\in\mathcal{R}} p_{\mathbf{i},t} g'_{\mathbf{i},t} + \frac{\eta^2}{1-\gamma} \sum_{\mathbf{i}\in\mathcal{R}} p_{\mathbf{i},t} g'^2_{\mathbf{i},t} \,.$$

Pour majorer les sommes ci-dessus, on observe premièrement que

$$\begin{aligned}\sum_{\mathbf{i}\in\mathcal{R}} p_{\mathbf{i},t} g'_{\mathbf{i},t} &= \sum_{\mathbf{i}\in\mathcal{R}} p_{\mathbf{i},t} \sum_{e\in\mathbf{i}} g'_{e,t} = \sum_{e\in E} g'_{e,t} \sum_{\mathbf{i}\in\mathcal{R}:e\in\mathbf{i}} p_{\mathbf{i},t} \\ &= \sum_{e\in E} g'_{e,t} q_{e,t} = g_{\mathbf{I}_t,t} + |E|\beta \,.\end{aligned}$$

Ensuite, il n'est pas plus difficile de voir que

$$\sum_{\mathbf{i}\in\mathcal{R}} p_{\mathbf{i},t} g'^2_{\mathbf{i},t} \leqslant K(1+\beta) \sum_{e\in E} g'_{e,t} \,.$$

Par conséquent,

$$\ln \frac{\overline{W}_t}{\overline{W}_{t-1}} \leqslant \frac{\eta}{1-\gamma} \left(g_{\mathbf{I}_t,t} + |E|\beta\right) + \frac{\eta^2(1+\beta)K}{1-\gamma} \sum_{e\in E} g'_{e,t}.$$

En additionnant ces inégalités selon $t = 1, \ldots, n$, il vient

$$\ln\frac{\overline{W}_n}{\overline{W}_0} \leqslant \frac{\eta}{1-\gamma}\left(\widehat{G}_n + n|E|\beta\right) + \frac{\eta^2 K(1+\beta)}{1-\gamma}\sum_{e\in E} G'_{e,n}$$
$$\leqslant \frac{\eta}{1-\gamma}\left(\widehat{G}_n + n|E|\beta\right) + \frac{\eta^2 K(1+\beta)}{1-\gamma}|\mathcal{C}|\max_{\mathbf{i}\in\mathcal{R}} G'_{\mathbf{i},n}$$

où la deuxième inégalité provient du fait que $\sum_{e\in E} G'_{e,n} \leqslant \sum_{\mathbf{i}\in\mathcal{C}} G'_{\mathbf{i},n}$.
La combinaison de cette borne supérieure avec la borne inférieure (4) entraîne

$$\widehat{G}_n \geqslant (1-\gamma-\eta K(1+\beta)|\mathcal{C}|)\max_{\mathbf{i}\in\mathcal{R}} G'_{\mathbf{i},n} - \frac{1-\gamma}{\eta}\ln N - n|E|\beta.$$

Or, $1-\gamma-\eta K(1+\beta)|\mathcal{C}| \geqslant 0$ vu les conditions imposées sur les paramètres, de sorte que le lemme 1 implique qu'avec probabilité au moins $1-\delta$,

$$\widehat{G}_n \geqslant (1-\gamma-\eta K(1+\beta)|\mathcal{C}|)\left(\max_{\mathbf{i}\in\mathcal{R}} G_{\mathbf{i},n} - \frac{K}{\beta}\ln\frac{|E|}{\delta}\right) - \frac{1-\gamma}{\eta}\ln N - n|E|\beta.$$

En se rappelant que $\widehat{G}_n = Kn - \widehat{L}_n$ et $G_{\mathbf{i},n} = Kn - L_{\mathbf{i},n}$, on arrive à

$$\widehat{L}_n \leqslant Kn\,(\gamma + \eta(1+\beta)K|\mathcal{C}|) + (1-\gamma-\eta(1+\beta)K|\mathcal{C}|)\min_{\mathbf{i}\in\mathcal{R}} L_{\mathbf{i},n}$$
$$+ (1-\gamma-\eta(1+\beta)K|\mathcal{C}|)\frac{K}{\beta}\ln\frac{|E|}{\delta} + \frac{1-\gamma}{\eta}\ln N + n|E|\beta.$$

Cette dernière inégalité entraîne

$$\widehat{L}_n - \min_{\mathbf{i}\in\mathcal{R}} L_{\mathbf{i},n} \leqslant Kn\gamma + \eta(1+\beta)nK^2|\mathcal{C}| + \frac{K}{\beta}\ln\frac{|E|}{\delta} + \frac{1-\gamma}{\eta}\ln N + n|E|\beta$$
$$\leqslant Kn\gamma + 2\eta nK^2|\mathcal{C}| + \frac{K}{\beta}\ln\frac{|E|}{\delta} + \frac{\ln N}{\eta} + n|E|\beta$$

avec probabilité au moins $1-\delta$.
Le choix de

$$\beta = \sqrt{\frac{K}{n|E|}\ln\frac{|E|}{\delta}} \qquad \text{et} \qquad \gamma = 2\eta K|\mathcal{C}|$$

a pour résultat, toujours avec probabilité au moins $1-\delta$, que

$$\widehat{L}_n - \min_{\mathbf{i}\in\mathcal{R}} L_{\mathbf{i},n} \leqslant 4\eta nK^2|\mathcal{C}| + \frac{\ln N}{\eta} + 2\sqrt{nK|E|\ln\frac{|E|}{\delta}}\,,$$

dès lors que $n \geqslant (K/|E|)\ln(|E|/\delta)$ (condition qui assure que $\beta \leqslant 1$). Enfin, il suffit de prendre

$$\eta = \sqrt{\frac{\ln N}{4nK^2|\mathcal{C}|}}$$

pour obtenir la deuxième assertion du théorème (et $n \geqslant 4\ln N|\mathcal{C}|$ est requis pour garantir que $\gamma \leqslant 1/2$).

Remerciements : La plupart des résultats présentés dans cet article reposent sur des collaborations avec mes coauteurs Nicolò Cesa-Bianchi, András György, Tamás Linder, György Ottucsák, and Gilles Stoltz. Je suis reconnaissant à Gilles Stoltz d'avoir traduit en français cet article, et des remarques pertinentes qu'il a formulées au passage.

Références

[1] AUER P. (2002). Using confidence bounds for exploitation-exploration tradeoffs. *Journal of Machine Learning Research*, 3 :397-422.

[2] AUER P., CESA-BIANCHI N., FREUND Y. et SCHAPIRE R. (2002). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32 :48-77.

[3] AUER P. et LONG P.M. (1999). Structural results for online learning models with and without queries. *Machine Learning*, 36 :147-181.

[4] AWERBUCH B. et KLEINBERG R. D. (2004). Adaptive routing with end-to-end feedback : distributed learning and geometric approaches. In *Proceedings of the 36th ACM Symposium on the Theory of Computing*, pages 45-53. ACM Press.

[5] AZUMA K. (1967). Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 68 :357-367.

[6] BAÑOS A. (1968). On pseudo-games. *Annals of Mathematical Statistics*, 39 :1932-1945.

[7] BERRY D.A. et FRISTEDT B. (1985). *Bandit Problems*. Chapman and Hall.

[8] BLACKWELL D. (1956). An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6 :1-8.

[9] BLACKWELL D. (1956). Controlled random walks. In *Proceedings of the International Congress of Mathematicians, 1954*, volume III, pages 336-338. North-Holland.

[10] BOUSQUET O. et WARMUTH M. (2002). Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3 :363-396.

[11] CESA-BIANCHI N. (1999). Analysis of two gradient-based algorithms for on-line regression. *Journal of Computer and System Sciences*, 59(3) :392-411.

[12] CESA-BIANCHI N., FREUND Y., HAUSSLER D., HELMBOLD D.P., SCHAPIRE R. et WARMUTH M. (1997). How to use expert advice. *Journal of the ACM*, 44(3) :427-485.

[13] CESA-BIANCHI N. et LUGOSI G. (1999). On prediction of individual sequences. *Annals of Statistics*, 27 :1865-1895.

[14] CESA-BIANCHI N. et LUGOSI G. (2003). Potential-based algorithms in on-line prediction and game theory. *Machine Learning*, 51 :239-261.

[15] CESA-BIANCHI N. et LUGOSI G. (2006). *Prediction, Learning, and Games*. Cambridge University Press, New York.

[16] CESA-BIANCHI N., LUGOSI G. et STOLTZ G. (2004). Minimizing regret with label efficient prediction. *IEEE Transactions on Information Theory*, 51 :2152-2162.

[17] CESA-BIANCHI N., LUGOSI G. et STOLTZ G. (2004). Regret minimization under partial monitoring. Prépublication DMA-04-18, Ecole normale supérieure, Paris.

[18] COVER T. (1965). 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. Maison d'édition de l'Académie des sciences de Tchécoslovaquie, Prague.

[19] FEDER M., MERHAV N. et GUTMAN M. (1992). Universal prediction of individual sequences. *IEEE Transactions on Information Theory*, 38 :1258-127.

[20] FOSTER D. (1991). Prediction in the worst-case. *Annals of Statistics*, 19 :1084-1090.

[21] FOSTER D. et VOHRA R. (1998). Asymptotic calibration. *Biometrika*, 85 :379-390.

[22] FOSTER D. et VOHRA R. (1999). Regret in the on-line decision problem. *Games and Economic Behavior*, 29 :7-36.

[23] FREEDMAN D.A. (1975). On tail probabilities for martingales. *Annals of Probability*, 3 :100-118.

[24] FUDENBERG D. et LEVINE D.K. (1998). *The Theory of Learning in Games*. MIT Press.

[25] GITTINS J.C. (1989). *Multi-Armed Bandit Allocation Indices*. Wiley-Interscience series in Systems and Optimization. Wiley.

[26] GYÖRGY A., LINDER T. et LUGOSI G. (2004). Efficient algorithms and minimax bounds for zero-delay lossy source coding. *IEEE Transactions on Signal Processing*, 52 :2337-2347.

[27] GYÖRGY A., LINDER T. et LUGOSI G. (2004). A "follow the perturbed leader"-type algorithm for zero-delay quantization of individual sequences. In *Data Compression Conference*, pages 342-351.

[28] GYÖRGY A., LINDER T. et LUGOSI G. (2005). Tracking the best of many experts. In *Proceedings of the 18th Annual Conference on Learning Theory*, pages 204-216.

[29] GYÖRGY A., LINDER T., LUGOSI G. et OTTUCSÁK Gy. (2005). The on-line shortest path bandit problem. Manuscrit.

[30] HANNAN G. (1957). Approximation to Bayes risk in repeated play. *Contributions to the theory of games*, 3 :97-139.

[31] HART S. et MAS-COLELL A. (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68 :1127-1150.

[32] HART S. et MAS-COLELL A. (2001) A general class of adaptive strategies. *Journal of Economic Theory*, 98 :26-54.

[33] HART S. et MAS-COLELL A. (2002). A reinforcement procedure leading to correlated equilibrium. In *Economic Essays : A Festschrift for Werner Hildenbrand*, pages 181-200. Springer.

[34] HAUSSLER D., KIVINEN J. et WARMUTH M. (1998). Sequential prediction of individual sequences under general loss functions. *IEEE Transactions on Information Theory*, 44 :1906-1925, 1998.

[35] HELMBOLD D.P., LITTLESTONE N. et LONG P.M. (2000). Apple tasting. *Information and Computation*, 161(2) :85-139.

[36] HELMBOLD D.P. et PANIZZA S. (1997). Some label efficient learning results. In *Proceedings of the 10th Annual Conference on Computational Learning Theory*, pages 218-230. ACM Press.

[37] HELMBOLD D.P. et SCHAPIRE R. (1997). Predicting nearly as well as the best pruning of a decision tree. *Machine Learning*, 27(1) :51-68.

[38] HELMBOLD D.P. et WARMUTH M. (1998). Tracking the best expert. *Machine Learning*, 32(2) :151-178.

[39] HOEFFDING W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58 :13-30.

[40] HUTTER M. et POLAND J. (2004). Prediction with expert advice by following the perturbed and penalized leader. Prépublication IDSIA-20-04, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Suisse.

[41] KALAI A. et VEMPALA S. (2003). Efficient algorithms for the online decision problem. In *Proceedings of the 16th Annual Conference on Learning Theory*, pages 26-40. Springer.

[42] LAI T.L. et ROBBINS H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6 :4-22.

[43] LEMPEL A. et ZIV J. (1976). On the complexity of an individual sequence. *IEEE Transactions on Information Theory*, 22 :75-81.

[44] LITTLESTONE N. et WARMUTH M. (1994). The weighted majority algorithm. *Information and Computation*, 108 :212-261.

[45] MANNOR S. et SHIMKIN N. (2003). On-line learning with imperfect monitoring. In *Proceedings of the 16th Annual Conference on Learning Theory*, pages 552-567. Springer.

[46] MCMAHAN H. B. et BLUM A. (2004). Online geometric optimization in the bandit setting against an adaptive adversary. In *Proceedings of the 17th Annual Conference on Learning Theory*, pages 109-123. Springer, 2004.

[47] MEGIDDO N. (1980). On repeated games with incomplete information played by non-Bayesian players. *International Journal of Game Theory*, 9 :157-167.

[48] MERHAV N. et FEDER M. (1993). Universal schemes for sequential decision from individual data sequences. *IEEE Transactions on Information Theory*, 39 :1280-1292.

[49] MERHAV N. et FEDER M. (1998). Universal prediction. *IEEE Transactions on Information Theory*, 44 :2124-2147.

[50] MERTENS J.-F., SORIN S. et ZAMIR S. (1994). Repeated games. core Discussion paper, numéros 9420, 9421, 9422, Louvain-la-Neuve, Belgique.

[51] NEVEU J. (1975). *Discrete Parameter Martingales*. North-Holland.

[52] PICCOLBONI A. et SCHINDELHAUER C. (2001). Discrete prediction games with arbitrary feedback and loss. In *Proceedings of the 14th Annual Conference on Computational Learning Theory*, pages 208-223.

[53] ROBBINS H. (1952). Some aspects of the sequential design of experiments. *Bullettin of the American Mathematical Society*, 55 :527-535.

[54] RUSTICHINI A. (1999). Minimizing regret : The general case. *Games and Economic Behavior*, 29 :224-243.

[55] STOLTZ G. (2005). *Information incomplète et regret interne en prédiction de suites individuelles*. Thèse de doctorat, Université Paris-Sud, Orsay.

[56] TAKIMOTO E. et WARMUTH M. (2004). Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4 :773-818.

[57] VOVK V. (1990). Aggregating strategies. In *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, pages 372-383.

[58] VOVK V. (1998). A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56 :153-173.

[59] VOVK V. (1999). Derandomizing stochastic prediction strategies. *Machine Learning*, 35 :247-282.

[60] VOVK V. (2001). Competitive on-line statistics. *International Statistical Review*, 69 :213-248.

[61] WEISSMAN T. et MERHAV N. (2001). Universal prediction of binary individual sequences in the presence of noise. *IEEE Transactions on Information Theory*, 47 :2151-2173.

[62] WEISSMAN T., MERHAV N. et SOMEKH-BARUCH (2001). Twofold universal prediction schemes for achieving the finite state predictability of a noisy individual binary sequence. *IEEE Transactions on Information Theory*, 47 :1849-1866.

[63] ZIV J. (1978). Coding theorems for individual sequences. *IEEE Transactions on Information Theory*, 24 :405-412.

[64] ZIV J. (1980). Distortion-rate theory for individual sequences. *IEEE Transactions on Information Theory*, 26 :137-143.

[65] ZIV J. et LEMPEL A. (1977). A universal algorithm for sequential data-compression. *IEEE Transactions on Information Theory*, 23 :337-343.