This is to certify that the

dissertation entitled

Applications of Play Against Past
Strategies in Repetitions of a Game

presented by

Lei Chen

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Major professor
Dennis Gilliland

Date December 1, 1997

APPLICATIONS OF PLAY AGAINST PAST STRATEGIES IN REPETITIONS OF A GAME

By

Lei Chen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1997

ABSTRACT

APPLICATIONS OF PLAY AGAINST PAST STRATEGIES IN REPETITIONS OF A GAME

By

Lei Chen

Hannan [17] introduced and studied certain recursive "play against past" strategies in the repeated play of a game. His results include bounds on the excess of total risk to player II over an envelope risk, namely, that which results from use of the best simple strategy based on knowledge of player I's empirical distribution of past choices.

This thesis investigates the Hannan results for finite games and extends their applications to certain infinite games. This is accomplished in Chapter 2 by showing that the Hannan bounds hold for play against random perturbations of the empirical distribution of player I's past choices of randomized strategies and by identifying pure strategies in the infinite games with randomized strategies in a companion finite game. Chapter 3 uses this idea to show that the Hannan recursive strategies and resulting bounds have applications in the repeated play of certain allocation and nonadversarial multi-arm bandit problems.

Another application concerns the repeated play of an expert selection problem. Chapter 4 contains a review of some of the literature for this problem. In this component game, player II selects an expert from a set of experts for a prediction and suffers loss equal to the prediction error. In repeated play, the envelope risk is the total of the prediction errors by the best individual expert. For a set of $n = 2$ experts, Foster and Vohra [13] proposed a recursive strategy and showed that it results in average prediction error close to that of the best individual expert in the limit. Chapter 4 shows that their strategy is an example of a Hannan strategy and that the adaptation of Hannan to infinite games establishes a slightly stronger version of their asymptotic result. Moreover, this approach provides an immediate solution for the case of n experts.

To Jingjing - my dear husband

ACKNOWLEDGEMENT

I would like to express my sincerest gratitude to my advisor Professor Dennis Gilliland for his guidance and continuous encouragement in the preparation of this dissertation. He taught me the art of consulting as well as the art of teaching and research. I appreciate his patience and willingness to discuss problems at any time. I could not have finished this dissertation without his great help.

Special thanks go to Professor James Hannan for the many things I learned from him, for the conversations we had all over the campus – classroom, office, hallway, IM East, $\cdots$, and of course, for his fundamental research in game theory from which my work derives. It was amazing that he almost persuaded me to pursue an academic career even though the chance of getting it is so slim.

I could not imagine the consequence if I forgot to thank Professor James Stapleton. It is he who called me about my admission in July 1993. I am grateful for the good advice he gave me throughout my graduate study, and for his encouraging words about my English which sometimes were so good that it gave me the delusion that my English is as good as his.

I would also like to thank Professors J. Hannan, R. Lepage and H. Salehi for serving on my thesis committee, for carefully reading my thesis, and for their helpful comments.

I would like to thank all the faculty and staff of the Department of Statistics and Probability who treated me so well that I am alleged to be the only one who cried when leaving the Wells Hall.

I am very grateful to my parents - Weiyou Chen and Xinshun Zhao from whom I learned to work hard and be responsible. Their wish of me getting advanced degrees has been my inspiration over the years. I am also grateful to my parents-in-law - Shizhong Wang and Liqiong Peng who kept me out of the kitchen when I was preparing my dissertation. I owe special thanks to my parents-like host family - Fred and Charlotte Poston. Their loving and caring always made me feel warm even in the coolest days of Michigan's winter.

Words are not enough for me to express how much I owe to my best friend, walking

dictionary, personal trainer, counselor, soul mate - my husband Jing Wang. He is the one who comforted me and cheered me up when I was unhappy. He is the one who made me relax when I felt stress. He is the one who made me believe in myself when I lost confidence. Although I always complained about his frank criticisms, I know from my heart that he is the one who loves me the most, and he is the one who changed my world, and he is the one who made me become who I am. This dissertation is dedicated to him with love.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

There is current interest in a variety of problems that involve repetitions of a decision problem. Examples are on-line allocation problems, multi-armed bandit problems and expert selection problems. Hannan [17] developed recursive strategies for player II in the repeated play of a game. Our concern is the adaptation of the Hannan [17] finite game results for certain infinite games, including the aforementioned.

Prediction by combining expert advice has extensive applications. According to Clemen [8]'s review, it has been applied not only to meteorology and economics, but also to the prediction of social and technological events, football game outcomes, electrical demand, tourism, political risk and population, etc.. Clemen [8] gave an extensive survey of methods for combining forecasts. Most of these conventional methods for combining forecasts involve taking a weighted average of individual forecasts and can be viewed within a regression framework. Regression techniques, such as weighted least squares, robust-weighting techniques, ridge regression, latent root regression, have been used in combining forecast. Bayesian techniques for including prior information in a forecast combination have been studied by Clemen and Winkler [10], Clemen and Guerard [9]and Anadalingam and Chen [1]. Most of these approaches require that the probability distribution of the event being forecasted be specified.

In computer science, predicting a binary sequence by combining the predictions of a set of experts has been studied by Cesa-Bianchi et al.[7], Foster [12], Freund [14], Littlestone and Warmuth [23], Vovk [27][28] etc.. Haussler, Kivinen and Warmuth [18] also provided prediction strategies for continuous-valued outcomes using expert advice for certain classes of loss functions. Their strategies are based on the exponential weight algorithm introduced by Littlestone and Warmuth [23] and by Vovk [27]. None of their strategies requires statistical assumptions be made about the decision-maker's subjective beliefs regarding the distribution of outcomes. They proved bounds on the difference between the average loss of their strategies and the average loss of the best expert. The average loss of their strategies will approach that of the average loss of the best expert at a certain rate as the length of outcome sequence T goes to ∞.

Foster and Vohra [13] studied the problem of choosing a prediction from two experts. Assume that the prediction loss is bounded and that the decision-maker has the knowledge of the past losses incurred by the two experts. They constructed a randomized strategy for the decision-maker and indicated that the difference between the average loss of the decision-maker and the minimum average loss of the two experts goes to 0 in probability as the number of trials T goes to ∞.

By connecting the Foster and Vohra [13] randomized strategy with game theory results studied by Hannan [17], we verify that the Foster and Vohra [13] randomized strategy is a special case of Hannan's strategy for player II in a finite two-person game, that is, it plays Bayes versus a randomized perturbation of the arithmetic mean of player I's past randomized strategies. The minimum average loss of the two experts is the Bayes envelope, denoted by $\phi(\frac{1}{T}X^T)$, where X^T is the sum of player I's past randomized strategies.

We also consider the prediction problem in which there are n experts. Based on game theory results, we construct randomized strategies for the decision-maker such that the difference between the average loss of the decision-maker and the average loss of the best expert goes to 0 in probability as the number of trials T goes to ∞.

In the compound decision problem, the extended envelopes, introduced by Johns [21], is a lower envelope than the simple Bayes envelope. The extended sequence compound decision problem has been studied by Swain [25], Gilliland and Hannan [16], Ballard [3], Ballard, Gilliland and Hannan [5] and Ballard and Gilliland [4]. Vardeman [26] studied the k-extended problem in a game theoretic situation. He constructed randomized strategies with risk approaching the k-extended envelope at the rate of $O(T^{-1/2})$.

If the loss of the experts have some dependency, then we can use Vardeman [26]'s idea to construct k-extended randomized strategies for the decision-maker such that the average loss of the decision-maker approaches a lower envelope.

1.2 Summary

This dissertation connects game theory, the on-line allocation model, the multi-armed bandit problem and the on-line prediction problem using expert advice. It is organized as follows.

In Chapter 2, following Hannan [17], we present notations, useful lemmas and theorems for a finite two-person game. We also extend the results of recursive strategies to an infinite game.

In Chapter 3, we introduce two algorithms, Algorithm H and Algorithm $\tilde{H}$, based on Theorem 2.3.2. In the on-line allocation model, if the allocation agent uses Algorithm H, then the difference between the expected average loss of the allocation agent and the average loss of the best strategy converges to 0 at the rate of $O(T^{-1/2})$. In the multi-armed bandit problem if the player applies Algorithm $\tilde{H}$, then the difference between the expected average reward of the player and the expected maximum average reward of any arm in a sequence of T trials will go to 0 at the rate of $O(T^{-1/2})$.

In Chapter 4, Sections 4.2 and 4.5.1 contain a general review of the most relevant literature for the on-line prediction problem using expert advice and the k-extended idea. In Section 4.3, we present a proof of Theorem 1 of Foster and Vohra [13] using Theorem 2.3.3. In Section 4.4, Theorem 2.3.4 is used to generalize the problem studied by Foster and Vohra [13] to the case in which the choice is among n experts. Without any statistical assumption about the distribution of the sequence being predicted, if the prediction loss is bounded, then we can construct a randomized strategy for the decision-maker such that the difference between the average loss of the decision-maker and the average loss of the best expert converges to 0 in probability at the rate of $O(T^{-1/2})$. In Section 4.5, we use the Vardeman [26] technique to construct k-extended prediction strategies when the decision-maker has the knowledge of the predictions made by n experts. If the loss of the experts takes values in a finite set with cardinality q, then using the k-extended prediction strategy the average loss of the decision-maker will approach the k-extended envelope in probability at the rate of $O(q^{n(k-1)}T^{-1/2})$ uniformly in outcome sequences.

Chapter 2

Strategies for repeated play of a game

2.1 Preliminaries

2.1.1 The finite component game

Consider a finite two-person game in which players I and II have, respectively, m and n pure strategies. Their spaces of randomized strategies are denoted by $\mathbf{X}$ and $\mathbf{Y}$,

$$\mathbf{X} = \{x = (x_1, x_2, \ldots, x_m) | x_i \geq 0, \sum_1^m x_i = 1\},$$

$$\mathbf{Y} = \{y = (y_1, y_2, \ldots, y_n) | y_i \geq 0, \sum_1^n y_i = 1\},$$

and their pure strategies are represented by base vectors ϵ and δ in $\mathbf{X}$ and $\mathbf{Y}$, that is, the degenerate probability distributions.

Notations. For m-vectors, we use juxtaposition to indicate inner product, that is, $ab = \sum_1^m a_i b_i$, and we define the norms

$$|a| = \max_i |a_i| \;\; and \;\; ||a|| = \sum_1^m |a_i|.$$

Player II's inutility is denoted by a loss matrix A,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix} = (A^1, \cdots, A^n).$$

It is assumed throughout the thesis that A has no dominant column, i.e., for each j

$$\max_{\epsilon}(\epsilon A^j - \min_{r} \epsilon A^r) > 0, \tag{2.1}$$

and no duplicate and dominated columns.

Notations. Let

$$A^{qr} = A^q - A^r, \ 1 \le q < r \le n.$$

The expectation of the loss when player I or player II uses a randomized strategy is $R(x, y) = xAy$, and we will refer to it as risk. From the risk point of view, this game is identical with the s-game in which player II's pure strategies are m-vectors in the set of columns of A, $\{A^1, \ldots, A^n\}$, and player II's randomized strategies s are m-vectors in the convex hull of the columns of A,

$$s \in S = \{Ay | y \in \mathbf{Y}\}.$$

For each $x \in \mathbf{X}$, the vector inner product xs is the Bayes risk of s against x. The Bayes envelope is defined as

$$\phi(x) = \min_{s \in S} xs = \min_{j=1,\ldots,n} xA^j, \ x \in \mathbf{X}.$$

A pure strategy valued Bayes response is a function $\sigma(\cdot)$ on $\mathbf{X}$ into $\{A^1, \ldots, A^n\}$ where for each x,

$$x\sigma(x) = \phi(x).$$

Note that $\phi(\cdot)$ is uniquely defined and $\sigma(\cdot)$ is not unless, for each x, $x\sigma$ is minimized by a unique σ. It is useful to extend the domain of $\phi(\cdot)$ and any given $\sigma(\cdot)$ to the nonnegative orthant of m-space as follows. For $w \in [0, \infty)^m$, let

$$\phi(w) = \min_{s \in S} ws = \min_{j=1,\ldots,n} wA^j,$$

and $\sigma(0) = arbitrary$ and

$$\sigma(w) = \sigma(w/||w||) \text{ if } w \neq 0.$$

It is follows that $\sigma(\cdot)$ is positive homogeneous of order zero, that is, $\sigma(kw) = \sigma(w)$ for all $w \in [0, \infty)^m$ and constants $k > 0$.

Notations. In all of the following, $\sigma(\cdot)$ will denote a pure strategy valued, positive, homogeneous Bayes response defined on $[0, \infty)^m$; also, following Hannan [17, p. 102], we let

$$|B| = \sup_{x,x'} |\sigma(x) - \sigma(x')| = \max_{q,r} |A^{qr}|, \tag{2.2}$$

where the last equality follows from the fact that A has no dominated columns.

A strategy $s \in S$ may be evaluated for each $x \in X$ in terms of the additional risk above the Bayes envelope risk $\phi(x)$, that is $xs - \phi(x)$. This excess is called the regret of s at x.

2.1.2 The repeated game

This section considers repeated play of the component game and the evaluation of recursive strategies in terms of their excess total risk over ϕ evaluated at the (non-normalized) empirical distribution of player I's choices.

Suppose the component game is played repeatedly with ϵ^t and σ^t denoting the choices of pure strategies by player I and II, respectively, at time $t = 1, 2, \ldots, T$. We suppress the display of dependence on T in writing

$$\underline{\epsilon} = (\epsilon^1, \ldots, \epsilon^T) \text{ and } \underline{\sigma} = (\sigma^1, \ldots, \sigma^T).$$

The total loss across the first T plays of the game is

$$R^T(\underline{\epsilon}, \underline{\sigma}) = \sum_1^T \epsilon^t \sigma^t.$$

If player II uses the same component strategy σ at each time t, that is, $\underline{\sigma} = (\sigma, \ldots, \sigma)$, then we write $\underline{\sigma} = \sigma^T$ and note that

$$R^T(\underline{\epsilon}, \sigma^T) = \sum_1^T \epsilon^t \sigma = E^T \sigma$$

where $E^T = \sum_1^T \epsilon^t$. (Hannan [17, p. 107] calls such $\underline{\sigma}$ power strategies.) Note that

$$R^T(\underline{\epsilon}, \sigma^T) \geq \phi(E^T)$$

with equality if $\sigma = \sigma(E^T)$. We will sometimes refer to the sum E^T as the empirical distribution of player I's choices of pure strategies through time T.

The envelope risk $\phi(E^T)$ is the minimum total risk to player II resulting from the use of a power strategy across the first T plays of the game. If player II knows E^T in advance, then the power strategy with kernel $\sigma(E^T)$ achieves the envelope risk.

Hannan [17] refers to the excess risk

$$R^T(\underline{\epsilon}, \underline{\sigma}) - \phi(E^T)$$

as the modified regret of $\underline{\sigma}$ at $\underline{\epsilon}$. He develops bounds for the modified regret of recursive strategies $\underline{\sigma}$, with σ^t a function of E^{t-1} and possible randomization. These bounds are $O(T^{1/2})$ uniformly in $\underline{\epsilon}$ for finite component games.

Since we make extensive use of the Hannan [17] notations and results, it is convenient to state some of these for future reference. We let H^t, $t = 0, 1, \ldots$ be a sequence of numbers satisfying,

$$0 < H^0 \leq H^1 \leq H^2 \leq \cdots \quad \frac{H^1}{1} \geq \frac{H^2}{2} \geq \frac{H^3}{3} \geq \cdots . \tag{2.3}$$

Let μ be a probability measure on the unit m-cube $[0,1]^m$. $z \sim \mu$ provides the artificial randomization in the Hannan recursive strategies and is such that the induced distribution on

$$zA^{qr} = z(A^q - A^r)$$

satisfies a Lipschitz condition. Specifically, there exists a constant $L > 0$ such that

$$\mu\{z \mid t_1 \leq zA^{qr}/|A^{qr}| \leq t_2\} \leq L(t_2 - t_1) \text{ for all } q < r, t_1 \leq t_2. \tag{2.4}$$

Also $\theta := E_\mu||z||$. When μ is the uniform distribution on $[0,1]^m$, $\theta = m/2$ and we take $L = 1$.

Hannan [17, Theorem 3] showed that with the component game defined earlier, if (2.3) and (2.4) hold, then for each T and $\underline{\epsilon}$,

$$-H^T\theta|B| \leq \sum_1^T \epsilon^t E_\mu \sigma(E^{t-1} + H^{t-1}z) - \phi(E^T) \tag{2.5}$$

$$\leq H^T\theta|B| + L\frac{n^2}{4}\left(\sum_1^T \frac{2}{H^t} - \frac{T}{H^T}\right)|B|, \tag{2.6}$$

where $|B|$ is defined in (2.2).

Denote the right hand side of (2.6) by U^T. For fixed μ, Hannan [17, Section 7] showed that when T is known in advance to player II, then

$$H^t = (Ln^2/4\theta)^{1/2}T^{1/2}, \; t = 1, \ldots, T,$$

minimizes $\sup_T T^{-1/2}U^T$. When T is unknown to player II and μ is the uniform distribution, then Hannan [17, Theorem 4] proved that for all T and $\underline{\epsilon}$,

$$H^t = (3n^2/2m)^{1/2}t^{1/2}, \; t = 1, 2, \ldots,$$

minimizes $\sup_T T^{-1/2}U^T$, and that

$$-\sqrt{3n^2mT/8}|B| \leq \sum_1^T \epsilon^t E_\mu \sigma(E^{t-1} + \sqrt{\frac{3n^2(t-1)}{2m}}z) - \phi(E^T) \leq \sqrt{3n^2mT/2}|B|, \tag{2.7}$$

Hannan [17, Theorem 6] produced bounds which are a uniform improvement of those in (2.7). He showed that

$$-\sqrt{1.5mT}|B| \leq \sum_1^T \epsilon^t E_\mu \sigma(E^{t-1} + \sqrt{\frac{6(t-1)}{m}}z) - \phi(E^T) \leq \sqrt{6mT}|B|. \tag{2.8}$$

2.1.3 The k-extended game

Consider the above two-person game. Player I has m pure strategies and player II has n pure strategies and the loss matrix A satisfies (2.1). Let $\mathbf{X}$ and $\mathbf{Y}$ be the set of player I and II's randomized strategies, respectively.

Denote $S = \{Ay \mid y \in \mathbf{Y}\}$. For $x \in \mathbf{X}$ and $s \in S$, xs is the risk of player II choosing strategy s against x.

Suppose that this game occurs repeatedly. At each time t, let x^t represent player I's move, and $x^1, \ldots, x^{t-1}$ are known to player II before he makes his move at time t.

Let $\mathbf{s} = (s_1, s_2, \ldots, s_T)$ be such that s_t is a function from $\mathbf{X}^{t-1}$ into S. For $\mathbf{x}^{t-1} = (x^1, \ldots, x^{t-1})$, the risk of sequence strategy $\mathbf{s}$ is

$$\sum_{t=1}^{T} x^t s_t(\mathbf{x}^{t-1}). \tag{2.9}$$

When $\mathbf{s} = (s, s, \ldots, s)$ for a fixed $s \in S$, the risk (2.9) becomes

$$\sum_{t=1}^{T} x^t s = (\sum_{t=1}^{T} x^t)s = X^T s,$$

where X^T is the sum of the randomized strategies of player I through time T. The definitions of Bayes response and Bayes envelope imply that

$$\phi(X^T) = X^T \sigma(X^T).$$

Let S^* be a set of functions from $\mathbf{X}^{k-1}$ into S. Let

$$\mathbf{x}^{*t-1} = (x^{t-k+1}, \ldots, x^{t-1}).$$

If we consider $\mathbf{s} = (s^*, s^*, \ldots, s^*)$ for a fixed $s^* \in S^*$, then the risk (2.9) for such an $\mathbf{s}$ reduces to $\sum_{t=1}^{T} x^t s^*(\mathbf{x}^{*t-1})$. We term

$$\min_{s^* \in S^*} \sum_{t=1}^{T} x^t s^*(\mathbf{x}^{*t-1})$$

a k-extended envelope.

We use Vardeman [26] 's Γ^k notation to define set $\tilde{S} \subset [0, \infty)^{m^k}$ of the form

$$\tilde{S} = \{\tilde{s} \in R^{m^k} | (x^{t-k+1} \otimes x^{t-k+2} \otimes \cdots \otimes x^t)\tilde{s} = x^t s^*(\mathbf{x}^{*t-1}), \text{for some } s^* \in S^* \},$$

where m^k-vector $x^{t-k+1} \otimes x^{t-k+2} \otimes \cdots \otimes x^t$ is given by

$$((x^{t-k+1} \otimes x^{t-k+2} \otimes \cdots \otimes x^t)_{j_1 \ldots j_k},\ j_i \in \{1, \ldots, m\},\ i = 1, \ldots, k),$$

and

$$(x^{t-k+1} \otimes x^{t-k+2} \otimes \cdots \otimes x^t)_{j_1 \ldots j_k} = \prod_{i=1}^{k} x_{j_i}^{t-k+i}.$$

Let $\tilde{\sigma}(\omega)$ denote a positive homogeneous minimizer of $\omega\tilde{s}$. Denote

$$X_T^k = \sum_{t=1}^{T} x^{t-k+1} \otimes x^{t-k+2} \otimes \cdots \otimes x^t.$$

Then

$$\begin{aligned}
&\min_{s^* \in S^*} \sum_{t=1}^{T} x^t s^*(\mathbf{x}^{*t-1}) \\
&= \min_{\tilde{s} \in \tilde{S}} \sum_{t=1}^{T} (x^{t-k+1} \otimes x^{t-k+2} \otimes \cdots \otimes x^t)\tilde{s} \\
&= \min_{\tilde{s} \in \tilde{S}} X_T^k \tilde{s} \\
&= X_T^k \tilde{\sigma}(X_T^k).
\end{aligned}$$

For $\omega \in R^{m^k}$, define $\phi^k(\omega) = \omega\tilde{\sigma}(\omega)$. Thus the k-extended envelope is

$$\phi^k(X_T^k) = X_T^k \tilde{\sigma}(X_T^k).$$

2.2 Lemmas

For later application, we determine the behavior of the Hannan recursive strategy when player I's choices are randomized strategies x^t and $\sigma(E^{t-1} + H^{t-1}z)$ is replaced by $\sigma(X^{t-1} + H^{t-1}z)$, where $X^0 = 0$ and

$$X^T = \sum_1^T x^t, \;\; T = 1, 2, \ldots.$$

For this purpose we reinterpret lemmas from Hannan [17] with ϵ^t replaced by x^t.

Remark 2.2.1 *Lemma 2.2.1 below will be an important tool in the proof of Theorems 2.3.1 and 2.3.3. Hannan [17] used it in the proof of (2.5) and (2.6).*

Lemma 2.2.1 *Let $v^t, t = 1, \ldots, T$ be any sequence of m-vectors. It follows that*

$$\sum_1^T x^t v^t = X^T v^{T+1} + \sum_1^T X^{t-1}(v^t - v^{t+1}) + \sum_1^T x^t (v^t - v^{t+1}).$$

Proof.

$$\begin{aligned}
&X^T v^{T+1} + \sum_1^T X^{t-1}(v^t - v^{t+1}) + \sum_1^T x^t(v^t - v^{t+1}) \\
&= X^T v^{T+1} + \sum_1^T X^{t-1} v^t - \sum_1^T X^{t-1} v^{t+1} + \sum_1^T x^t v^t - \sum_1^T x^t v^{t+1}
\end{aligned}$$

$$= \sum_1^T X^t v^{t+1} - \sum_1^T X^{t-1} v^{t+1} + \sum_1^T x^t v^t - \sum_1^T x^t v^{t+1}$$
$$= \sum_1^T x^t v^{t+1} + \sum_1^T x^t v^t - \sum_1^T x^t v^{t+1}$$
$$= \sum_1^T x^t v^t.$$

□

Remark 2.2.2 *Lemma 2.2.2 below will be used in the proof of Theorems 2.3.1, 2.3.2 and 2.3.3.*

Lemma 2.2.2 *If (2.3) holds, then*

$$\left\|\frac{X^{t-1}}{H^{t-1}} - \frac{X^t}{H^t}\right\| \leq \frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}$$

Proof.

$$\begin{aligned}
\left\|\frac{X^{t-1}}{H^{t-1}} - \frac{X^t}{H^t}\right\| &= \sum_{i=1}^m \left|\frac{X_i^{t-1}}{H^{t-1}} - \frac{X_i^{t-1} + x_i^t}{H^t}\right| \\
&= \sum_{i=1}^m \left|X_i^{t-1}\left(\frac{1}{H^{t-1}} - \frac{1}{H^t}\right) - \frac{x_i^t}{H^t}\right| \\
&\leq (t-1)\left(\frac{1}{H^{t-1}} - \frac{1}{H^t}\right) + \frac{1}{H^t} \\
&\leq \frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}
\end{aligned}$$

□

Remark 2.2.3 *The proofs of Lemmas 2.2.3 and 2.2.4 that follow are reinterpretations of (6.8)-(6.10) of Hannan [17]. He proved the results when player I uses pure strategies $\underline{\epsilon}$ under the assumption (without loss of generality) that for any ϵ*

$$\phi(\epsilon) = \min_{j \in \{1,\ldots,n\}} \epsilon A^j = 0. \tag{2.10}$$

Lemma 2.2.3 *Suppose that the sequence H^t is positive and nondecreasing and that σ^t is the recursive strategy*

$$\sigma^t := \sigma(X^{t-1} + H^{t-1} z), \quad t = 1, 2, \ldots. \tag{2.11}$$

It follows that

$$\sum_1^T X^t(\sigma^t - \sigma^{t+1}) \geq -H^T \|z\| |B|,$$

where $|B|$ *is defined in (2.2).*

Proof. From the definition of σ^t, we have for each t

$$(X^t + H^t z)(\sigma^t - \sigma^{t+1}) \geq 0.$$

Then

$$X^t(\sigma^t - \sigma^{t+1}) \geq -H^t z(\sigma^t - \sigma^{t+1}).$$

So

$$\begin{aligned}
&\sum_1^T X^t(\sigma^t - \sigma^{t+1}) \\
&\geq \sum_1^T H^t z\sigma^{t+1} - \sum_1^T H^t z\sigma^t \\
&= H^T z\sigma^{T+1} + \sum_1^{T-1} H^t z\sigma^{t+1} - \sum_1^T H^t z\sigma^t \\
&= H^T z\sigma^{T+1} + \sum_2^T H^{t-1} z\sigma^t - \sum_1^T H^t z\sigma^t \\
&= H^T z\sigma^{T+1} - H^0 z\sigma^1 - \sum_1^T (H^t - H^{t-1}) z\sigma^t \\
&= H^T z(\sigma^{T+1} - a^0) - H^0 z(\sigma^1 - a^0) - \sum_1^T (H^t - H^{t-1}) z(\sigma^t - a^0)
\end{aligned}$$

where $a^0 = (a_1^0, \ldots, a_m^0)$, and for each i

$$a_i^0 = \min_{j \in \{1,\ldots,n\}} a_{ij}, \tag{2.12}$$

where a_{ij} is the element of loss matrix A.

Since for any t, $\sigma^t - a^0 \geq 0$, $H^t \geq 0$ and H^t is non-decreasing, then we have

$$H^T z(\sigma^{T+1} - a^0) \geq 0.$$

$$0 \leq H^0 z(\sigma^1 - a^0) \leq H^0 \|z\| |B|.$$

$$0 \leq (H^t - H^{t-1}) z(\sigma^t - a^0) \leq (H^t - H^{t-1}) \|z\| |B|.$$

Therefore

$$\begin{aligned}\sum_1^T X^t(\sigma^t - \sigma^{t+1}) &\geq -H^0\|z\||B| - \sum_1^T (H^t - H^{t-1})\|z\||B| \\ &= -H^T\|z\||B|.\end{aligned}$$

□

Lemma 2.2.4 *Suppose that the sequence H^t is positive and nondecreasing and (2.11) holds. It follows that*

$$X^T\sigma^{T+1} - \phi(X^T) + \sum_1^T X^{t-1}(\sigma^t - \sigma^{t+1}) \leq H^T\|z\||B|,$$

where $|B|$ is defined in (2.2).

Proof. Since $(X^T + H^T z)[\sigma^{T+1} - \sigma(X^T)] \leq 0$, then

$$X^T\sigma^{T+1} - \phi(X^T) \leq -H^T z[\sigma^{T+1} - \sigma(X^T)]. \tag{2.13}$$

It follows from (2.11) that

$$(X^{t-1} + H^{t-1}z)(\sigma^t - \sigma^{t+1}) \leq 0.$$

Then

$$X^{t-1}(\sigma^t - \sigma^{t+1}) \leq -H^{t-1}z(\sigma^t - \sigma^{t+1}).$$

So

$$\begin{aligned}&\sum_1^T X^{t-1}(\sigma^t - \sigma^{t+1}) \\ &\leq \sum_1^T H^{t-1}z\sigma^{t+1} - \sum_1^T H^{t-1}z\sigma^t \\ &= \sum_1^T H^{t-1}z\sigma^{t+1} - \sum_0^{T-1} H^t z\sigma^{t+1} \\ &= H^T z\sigma^{T+1} - H^0 z\sigma^1 - \sum_1^T (H^t - H^{t-1})z\sigma^{t+1}.\end{aligned} \tag{2.14}$$

It follows from (2.13) and (2.14) that

$$\begin{aligned}
&X^T\sigma^{T+1} - \phi(X^T) + \sum_1^T X^{t-1}(\sigma^t - \sigma^{t+1}) \\
&\leq H^T z\sigma(X^T) - H^0 z\sigma^1 - \sum_1^T (H^t - H^{t-1}) z\sigma^{t+1} \\
&= H^T z[\sigma(X^T) - a^0] - H^0 z(\sigma^1 - a^0) - \sum_1^T (H^t - H^{t-1}) z(\sigma^{t+1} - a^0) \\
&\leq H^T ||z|||B|.
\end{aligned}$$

where a^0 is defined in (2.12). The second last inequality follows from the definitions of H^t, a^0 and $|B|$.

□

Remark 2.2.4 *Lemma 2.2.5 below is a slight modification of Lemma 1 of Hannan [17]. It will be used to prove Theorem 2.3.3.*

Lemma 2.2.5 *If w and w' are m-vectors and (2.4) holds, then*

$$E_\mu|\sigma(w+z) - \sigma(w'+z)| \leq L\frac{n^2}{2}|B|||w' - w||,$$

where $|B|$ is defined in (2.2).

Proof. Let $T_{jk} = \{z|\sigma(w+z) = A^j, \sigma(w'+z) = A^k\}$, and note that

$$\begin{aligned}
E_\mu|\sigma(w+z) - \sigma(w'+z)| &= \sum_{j\neq k} |A^{jk}|\mu(T_{jk}) \\
&= \sum_{j<k} |A^{jk}|[\mu(T_{jk}) + \mu(T_{kj})].
\end{aligned}$$

Note that

$$\sum_{j<k} |A^{jk}| \leq |B| \sum_{j\leq k} 1 \leq \frac{n^2}{2}|B|. \tag{2.15}$$

Hannan [17, p. 118] showed that

$$\mu(T_{jk}) + \mu(T_{kj}) \leq L||w - w'||. \tag{2.16}$$

Therefore, it follows from (2.15) and (2.16) that

$$E_\mu|\sigma(w+z) - \sigma(w'+z)| \leq L\frac{n^2}{2}|B|||w - w'||.$$

This completes the proof of the lemma.

□

Remark 2.2.5 *Vardeman [26] proved the following lemma in the case in which player I only takes pure strategies. Lemma 2.2.6 below shows that Vardeman's lemma holds for any finite set of player I's randomized strategies. This lemma gives a decomposition of $\phi^k(X_T^k)$. By Lemma 2.2.6, we can modify the solutions of the unextended problem to produce solutions of the k-extended problem.*

Lemma 2.2.6 *Let $\mathbf{x}^{*t-1} = (x^{t-k+1}, \ldots, x^{t-1})$, $\phi^k(X_T^k)$ be the k-extended Bayes envelope. If the cardinality of $\mathbf{X}$ is finite, then*

$$\phi^k(X_T^k) = \sum_{x \in \mathbf{X}^{k-1}} \phi\left(\sum_{t:\mathbf{x}^{*t-1}=x}^{T} x^t \right).$$

Proof.

$$\sum_{t=1}^{T} x^t s^*(\mathbf{x}^{*t-1}) = \sum_{x \in \mathbf{X}^{k-1}} \left(\sum_{t:\mathbf{x}^{*t-1}=x}^{T} x^t \right) s^*(x). \tag{2.17}$$

Equation (2.17) is minimal if $s^*(x) = \sigma(\sum_{t:\mathbf{x}^{*t-1}=x}^{T} x^t)$ for each $x \in \mathbf{X}^{k-1}$.

Therefore

$$\phi^k(X_T^k) = \sum_{x \in \mathbf{X}^{k-1}} \phi\left(\sum_{t:\mathbf{x}^{*t-1}=x}^{T} x^t \right).$$

This completes the proof of the lemma.

□

2.3 Bounds for the modified regret of Hannan recursive strategies

In this section, we prove results for the recursive strategy (2.11) that are extensions of those of Hannan summarized in (2.5) - (2.8). The extensions are from $\underline{\epsilon}$ sequences to $\underline{x}$ sequences.

Remark 2.3.1 *Theorem 2.3.1 below extends the results of Hannan [17, Theorem 3] as were summarized in (2.5) and (2.6).*

Theorem 2.3.1 *Suppose (2.3) and (2.4) hold. Then for all T and $\underline{x}$,*

$$-H^T\theta|B| \leq E_\mu \sum_1^T x^t\sigma(X^{t-1} + H^{t-1}z) - \phi(X^T) \leq H^T\theta|B| + L\frac{n^2}{4}(\sum_1^T \frac{2}{H^t} - \frac{T}{H^T})|B|.$$

where $\theta = E_\mu\|z\|$, and $|B|$ is defined in (2.2).

Proof. By Lemma 2.2.1, we have that

$$\begin{aligned} \sum_1^T x^t\sigma^t - \phi(X^T) &= [X^T\sigma^{T+1} - \phi(X^T)] + \sum_1^T X^{t-1}(\sigma^t - \sigma^{t+1}) + \sum_1^T x^t(\sigma^t - \sigma^{t+1}) \\ &:= S_1 + S_2 + S_3. \end{aligned}$$

Lemma 2.2.3 implies that

$$S_2 + S_3 \geq -H^T\|z\||B|. \tag{2.18}$$

(2.18) and $X^T(\sigma^{T+1} - \phi(X^T)) \geq 0$ imply that

$$S_1 + S_2 + S_3 \geq -H^T\|z\||B|. \tag{2.19}$$

By Lemma 2.2.4, we have

$$S_1 + S_2 \leq H^T\|z\||B|. \tag{2.20}$$

So it follows from (2.19) and (2.20) that

$$-H^T\|z\||B| \leq S_1 + S_2 + S_3 \leq H^T\|z\||B| + \sum_1^T x^t(\sigma^t - \sigma^{t+1}). \tag{2.21}$$

Taking expectation with respect to μ on (2.21), we have

$$-H^T\theta|B| \leq E_\mu(S_1 + S_2 + S_3) \leq H^T\theta|B| + E_\mu\sum_1^T x^t(\sigma^t - \sigma^{t+1}). \tag{2.22}$$

where $\theta = E_\mu\|z\|$.

Applying Lemma 1 of Hannan [17, p. 131] with $w = X^{t-1}/H^{t-1}$ and $w' = X^t/H^t$, we have

$$\begin{aligned} |E_\mu(\sigma^t - \sigma^{t+1})| &\leq L\frac{n^2}{4}|B|\|\frac{X^{t-1}}{H^{t-1}} - \frac{X^t}{H^t}\| \\ &\leq L\frac{n^2}{4}|B|(\frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}). \end{aligned}$$

The last inequality follows from Lemma 2.2.2.

Thus

$$\begin{aligned} E_\mu \sum_1^T x^t(\sigma^t - \sigma^{t+1}) &\le \sum_1^T \|x^t\| |E_\mu(\sigma^t - \sigma^{t+1})| \\ &\le L\frac{n^2}{4}|B|\sum_1^T(\frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}) \\ &\le L\frac{n^2}{4}|B|(\sum_1^T \frac{2}{H^t} - \frac{T}{H^T}). \end{aligned} \tag{2.23}$$

Therefore, (2.22) and (2.23) imply that

$$\begin{aligned} -H^T\theta|B| &\le E_\mu \sum_1^T x^t\sigma(X^{t-1} + H^{t-1}z) - \phi(X^T) \\ &\le H^T\theta|B| + L\frac{n^2}{4}(\sum_1^T \frac{2}{H^t} - \frac{T}{H^T})|B|. \end{aligned}$$

□

Remark 2.3.2 *Theorem 2.3.2 below extends results of Hannan [17, Theorem 6] as were summarized in (2.8). From the proof of Theorem 2.3.2, we see that the results of Hannan [17, Theorem 4] as summarized in (2.7) hold with ϵ^t replaced by x^t.*

Theorem 2.3.2 *If μ is the uniform distribution, then, for all T and $\underline{x}$,*

$$-\sqrt{1.5mT}|B| \le E_\mu \sum_{t=1}^T x^t\sigma(X^{t-1} + \sqrt{\frac{6(t-1)}{m}}z) - \phi(X^T) \le \sqrt{6mT}|B|.$$

where $|B|$ is defined in (2.2).

Proof. Applying Lemma 2 of Hannan [17] with $w = X^{t-1}/H^{t-1}$ and $w' = X^t/H^t$, we have

$$\begin{aligned} |E_\mu(\sigma^t - \sigma^{t+1})| &\le |B| \| \frac{X^{t-1}}{H^{t-1}} - \frac{X^t}{H^t} \| \\ &\le |B|(\frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}), \end{aligned}$$

where the last inequality follows from Lemma 2.2.2. Thus

$$\begin{aligned} E_\mu \sum_1^T x^t(\sigma^t - \sigma^{t+1}) &\le \sum_1^T \|x^t\| |E_\mu(\sigma^t - \sigma^{t+1})| \\ &\le |B|\sum_1^T(\frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}) \\ &\le |B|(\sum_1^T \frac{2}{H^t} - \frac{T}{H^T}). \end{aligned} \tag{2.24}$$

Therefore, (2.22) and (2.24) imply that

$$\begin{aligned} -H^T\theta|B| &\leq E_\mu \sum_1^T x^t\sigma(X^{t-1} + H^{t-1}z) - \phi(X^T) \\ &\leq H^T\theta|B| + (\sum_1^T \frac{2}{H^t} - \frac{T}{H^T})|B|. \end{aligned}$$

Hannan [17, Section 7] showed how to minimize

$$\sup_T T^{-1/2}[H^T\theta|B| + (\sum_1^T \frac{2}{H^t} - \frac{T}{H^T})|B|]$$

by choice of the H^t satisfying (2.3). Hannan [17, (7.7)] gives

$$\sup_T T^{-1/2}[H^T\theta|B| + (\sum_1^T \frac{2}{H^t} - \frac{T}{H^T})|B|] \geq \sqrt{6m}|B|,$$

with the lower bound obtained for

$$H^t = \sqrt{\frac{6t}{m}} \quad t = 1, 2, \ldots.$$

□

Remark 2.3.3 *Theorem 2.3.3 below is a modification of Theorem 3 of Hannan [17]. In Theorem 2.3.3 the E_μ expectation is outside the absolute value rather than inside it.*

Theorem 2.3.3 *Suppose that H^t satisfies (2.3) and (2.4) holds. Then for all T and $\underline{x}$,*

$$E_\mu|\sum_1^T x^t\sigma(X^{t-1} + H^{t-1}z) - \phi(X^T)| \leq H^T\theta|B| + L\frac{n^2}{2}(\sum_1^T \frac{2}{H^t} - \frac{T}{H^T})|B|.$$

where $\theta = E_\mu\|z\|$, and $|B|$ is defined in (2.2).

Proof. By Lemma 2.2.1, we start with

$$\begin{aligned} \sum_1^T x^t\sigma^t - \phi(X^T) &= [X^T\sigma^{T+1} - \phi(X^T)] + \sum_1^T X^{t-1}(\sigma^t - \sigma^{t+1}) + \sum_1^T x^t(\sigma^t - \sigma^{t+1}) \\ &:= S_1 + S_2 + S_3. \end{aligned}$$

It follows from (2.19) and (2.20) that

$$|S_1 + S_2 + S_3| \leq H^T\|z\||B| + |\sum_1^T x^t(\sigma^t - \sigma^{t+1})|. \tag{2.25}$$

Taking expectation on both sides of (2.25), we get

$$E_\mu|S_1+S_2+S_3| \le H^T\theta|B| + E_\mu|\sum_1^T x^t(\sigma^t-\sigma^{t+1})|. \tag{2.26}$$

Applying Lemma 2.2.5 with $w = X^{t-1}/H^{t-1}$ and $w' = X^t/H^t$,

$$E_\mu|\sigma^t-\sigma^{t+1}| \le L\frac{n^2}{2}|B|\|\frac{X^{t-1}}{H^{t-1}} - \frac{X^t}{H^t}\|.$$

It follows from Lemma 2.2.2 that

$$\begin{aligned} E_\mu|\sum_1^T x^t(\sigma^t-\sigma^{t+1})| &\le \sum_1^T \|x^t\|E_\mu|\sigma^t-\sigma^{t+1}| \\ &\le L\frac{n^2}{2}|B|\sum_1^T(\frac{t-1}{H^{t-1}} - \frac{t-2}{H^t}) \\ &\le L\frac{n^2}{2}|B|(\sum_1^T\frac{2}{H^t} - \frac{T}{H^T}). \end{aligned} \tag{2.27}$$

Therefore, Theorem 2.3.3 follows from (2.26) and (2.27).

□

Remark 2.3.4 *Theorem 2.3.4 below is a modification of Theorem 4 of Hannan [17]. Similarly we can show that for a fixed μ, if T is known to player II, then*

$$H^t = (n^2/2\theta)^{1/2}T^{1/2}, \ t = 1,\dots,T,$$

minimizes $\sup_T T^{-1/2}U^T(\underline{H})$.

Theorem 2.3.4 *If μ is the uniform distribution, then, for all T and $\underline{x}$,*

$$E_\mu|\sum_1^T x^t\sigma(X^{t-1} + \sqrt{\frac{3n^2(t-1)}{m}}z) - \phi(X^T)| \le \sqrt{3n^2mT}|B|.$$

where $|B|$ is defined in (2.2).

Proof. Since μ is the uniform distribution, the result of Theorem 2.3.3 becomes

$$E_\mu|\sum_1^T x^t\sigma(X^{t-1}+H^{t-1}z) - \phi(X^T)| \le \frac{H^Tm|B|}{2} + \frac{n^2}{2}(\sum_1^T\frac{2}{H^t} - \frac{T}{H^T})|B|.$$

Let

$$U^T(\underline{H}) = \frac{H^Tm|B|}{2} + \frac{n^2}{2}(\sum_1^T\frac{2}{H^t} - \frac{T}{H^T})|B|$$

where $\underline{H} = (H^1, \ldots, H^T)$. Suppose T is unknown to player II, we want to obtain an $\underline{H}$ minimizing

$$\sup_T T^{-1/2} U^T(\underline{H})$$

Hannan [17, Section 7] showed that

$$\sup_T T^{-1/2} U^T(\underline{H}) \geq (3n^2 m)^{1/2} |B| \tag{2.28}$$

and this lower bound is obtained for $\underline{H}$ with

$$H^t = \sqrt{\frac{3n^2 t}{m}} \quad t = 1, 2, \cdots \tag{2.29}$$

Therefore, (2.28) and (2.29) imply the result of Theorem 2.3.4.

□

Remark 2.3.5 *Theorem 2.3.5 below will be used to construct k-extended predicting strategies in Chapter 4.*

Theorem 2.3.5 *Let μ be the uniform distribution. Assume that the cardinality of* **X** *is d, and $d < \infty$. Denote*

$$\sigma_t = \sigma(X_{t-1}^k | \mathbf{x}^{*t-1} + \sqrt{\frac{3n^2(T(\mathbf{x}^{*t-1}) - 1)}{m}} z),$$

where

$$X_{t-1}^k | \mathbf{x}^{*t-1} = \sum_{j:\, \mathbf{x}^{*j-1} = \mathbf{x}^{*t-1}}^{t-1} x^j,$$

and

$$T(\mathbf{x}^{*t-1}) = \sum_{j=1}^{t-1} I_{\{\mathbf{x}^{*j-1} = \mathbf{x}^{*t-1}\}},$$

Then for all T and $\underline{x}$,

$$E_\mu |\sum_1^T x^t \sigma_t - \phi^k(X_T^k)| \leq \sqrt{3n^2 m d^{k-1} T} |B|,$$

where $|B|$ is defined in (2.2).

Proof. Lemma 2.2.6 implies that

$$\begin{aligned}\sum_1^T x^t\sigma_t - \phi^k(X_T^k) &= \sum_{x\in\mathbf{X}^{k-1}} [(\sum_{t:\mathbf{x}^{*t-1}=x} x^t\sigma_t) - \phi(X_T^k|x)] \\ &:= \sum_{x\in\mathbf{X}^{k-1}} A(x).\end{aligned}$$

Denote the indices t for which $\mathbf{x}^{*t-1} = x$ by $t_1 < t_2 < \ldots < t_{T(x)}$. Then

$$A(x) = \sum_{i=1}^{T(x)} x^{t_i}\sigma(X_{t_i-1}^k|x + \sqrt{\frac{3n^2(i-1)}{m}}z) - \phi(X_T^k|x).$$

By Theorem 2.3.4, we have that

$$E_\mu|A(x)| \le \sqrt{3n^2mT(x)}|B|.$$

Applying Schwarz inequality,

$$\begin{aligned}E_\mu|\sum_1^T x^t\sigma_t - \phi^k(X_T^k)| &\le \sum_{x\in\mathbf{X}^{k-1}} \sqrt{3n^2mT(x)}|B| \\ &\le \sqrt{3n^2md^{k-1}T}|B|.\end{aligned}$$

Hence the desired result follows.

□

2.4 An extension to an infinite game

In this section we adapt the Hannan recursive strategies to produce strategies for the repeated play of an infinite component game. The infinite component game is general enough to cover the on-line allocation, multi-armed bandit and expert selection problems as will be shown in Chapters 3 and 4.

Consider an infinite component game where player I chooses a pure strategy $\hat{x}$ from

$$\hat{\mathbf{X}} = [0,1]^n$$

and player II chooses a pure strategy $\hat{y}$ from

$$\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\},$$

the set of standard base vectors in n-space. Suppose that player II's inutility is given by the loss function

$$L(\hat{x}, \hat{y}) = \hat{x}\hat{y}, \quad \hat{x} \in \hat{\mathbf{X}},\ \hat{y} \in \hat{\mathbf{Y}}, \tag{2.30}$$

where we have used the juxtaposition of vectors in n-space to indicate the ordinary inner product.

Theorem 2.4.1 *There exists a finite $2^n \times n$ game with loss matrix A and a one to one mapping f from $\hat{\mathbf{X}}$ into $\mathbf{X}$, the $(2^n - 1)$-dimensional simplex of randomized strategies for player I in the finite game, such that*

$$f(\hat{x})A^j = \hat{x}e_j = \hat{x}_j, \quad \hat{x} \in \hat{\mathbf{X}},\ j = 1, \ldots, n. \tag{2.31}$$

Proof. We see that

$$f(\hat{x})A^j = \hat{x}_j, \quad j = 1, \ldots, n$$

if the $f(\hat{x})$ mixture of the rows of A, $f(\hat{x})A$, is equal to $\hat{x}$. Thus, if the convex hull of the rows of A is equal to $\hat{\mathbf{X}} = [0,1]^n$, then (2.31) is satisfied with f, any mapping where $f(\hat{x})$ is a mixing distribution of the rows that gives mixture $\hat{x}$. The minimum generating set for $[0,1]^n$ is the set of 2^n vertices of this cube. We take A to be any matrix whose row vectors are the 2^n vertices and f as described above. This A has no dominant columns since the left hand side of (2.1) is 1. It has no duplicate or dominated columns.

□

The infinite game $\hat{\mathbf{X}} = [0,1]^n$, $\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\}$ and $L(\hat{x}, \hat{y}) = \hat{x}\hat{y}$ has randomized strategies

$$\hat{\mathbf{X}}^* = \text{set of all probability distributions on } \hat{\mathbf{X}}$$

and

$$\hat{\mathbf{Y}}^* = \mathbf{Y} = (n-1) - \text{dimensional simplex of probability distribution on } n \text{ points.}$$

Since the loss function is linear in $\hat{x}$ and $\hat{\mathbf{X}}$ is convex, the risks from randomized strategies $\hat{x}^*$ are the same as those from the means. Thus, the extension of the infinite game through its randomized strategies remains "isomorphic" to the restricted extension of the finite game $\mathbf{X}$, $\mathbf{Y} = \hat{\mathbf{Y}}^*$, A, the restriction being $\mathbf{X}$ replaced by $f[\hat{\mathbf{X}}]$. For any $\hat{x}^* \in \hat{\mathbf{X}}^*$ and $\hat{y}^* \in \hat{\mathbf{Y}}^*$ there exists an $x \in \mathbf{X}$, namely, $x = f(\text{mean of } \hat{x}^*)$ and a $y \in \mathbf{Y}$, namely, $y = \hat{y}^*$ such that $L(\hat{x}^*, \hat{y}^*) = xAy$. Moreover, the Bayes envelope risk in the finite game at $x = f(\hat{x})$ is given by

$$\phi(x) = \min_j \hat{x}_j,$$

the Bayes envelope in the infinite game at $\hat{x}$.

Suppose the infinite component game is played repeatedly with $\hat{x}^t$ and $\hat{y}^t$ denoting the choices of strategies by player I and II, respectively, at time $t = 1, \ldots, T$. We suppress the display of dependence on T in writing

$$\underline{\hat{x}} = (\hat{x}^1, \ldots, \hat{x}^T) \text{ and } \underline{\hat{y}} = (\hat{y}^1, \ldots, \hat{y}^T).$$

Remark 2.4.1 *In the repeated play of the infinite component game, suppose that player II uses a Hannan recursive strategy*

$$\hat{y}^t = e_j \text{ if } \sigma(X^{t-1} + H^{t-1}z) = A^j,\ j = 1, 2, \ldots,$$

where σ is a positive homogeneous pure strategy valued Bayes response in the finite game described in Theorem 2.4.1, and

$$X^0 = 0,\ X^t = \sum_1^t x^s,\ x^t = f(\hat{x}^t)\ t = 1, 2, \ldots.$$

Then the modified regret in the repeated play of the infinite game at $\underline{\hat{x}}$ is

$$\sum_{t=1}^T \hat{x}^t \hat{y}^t - \min_j \sum_{t=1}^T \hat{x}^t_j = \sum_{t=1}^T x^t \sigma(X^{t-1} + H^{t-1}z) - \phi(X^T),$$

that is, it is the same as the modified regret in the finite game at $\underline{x}$. Hence the results of Theorem 2.3.1-2.3.5 cover repeated play of the infinite component game. Of course $m = 2^n$ and $|B| = 1$ in this adaptation.

Chapter 3

On-line allocation model and the multi-armed bandit problem

3.1 Introduction to on-line allocation model

3.1.1 Component allocation game

We consider the following component allocation game in which player I selects a loss vector $l \in [0,1]^n$ and player II selects a probability distribution p on n points with loss $L_p = \sum_1^n l_i p_i$. We recognize this as an example where player I's pure strategies are in the cube

$$\hat{\mathbf{X}} = [0,1]^n$$

and player II's pure strategies are in the finite set

$$\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\}.$$

By Theorem 2.4.1, there is a finite game isomorphic to the component allocation game.

3.1.2 On-line allocation model

Suppose the component allocation game is played repeatedly with l^t and p^t denoting the choices of strategies by player I and II, respectively, at time $t = 1, \ldots, T$. This repeated game was studied by Freund and Schapire [15]. They called this game the on-line allocation model.

Freund and Schapire [15] formalized the on-line allocation model as follows. The allocation agent A has n strategies to choose from. At each time $t = 1, 2, \ldots, T$, the allocator A choose a probability distribution

$$p^t = (p_1^t, p_2^t, \ldots, p_n^t)$$

over the n strategies, where p_i^t is the probability that strategy i will be chosen, for each i and t. Each strategy i suffers some loss l_i^t. Denote loss vector l^t by

$$l^t = (l_1^t, l_2^t, \ldots, l_n^t).$$

The loss suffered by A at time t is defined as

$$\sum_{i=1}^{n} p_i^t l_i^t = p^t l^t,$$

that is, the expected loss of the strategies with respect to A's chosen allocation rule.

Denote the expected total loss of A across the first T trials by

$$E[L_A^T] = \sum_{t=1}^{T} p^t l^t.$$

The goal of A is to minimize

$$E[L_A^T] - \min_{1 \le i \le n} \sum_{t=1}^{T} l_i^t.$$

So A will try to perform as well as the best strategy among these n strategies.

Assume that the loss suffered by any strategy is bounded, so without loss of generality, let $l_i^t \in [0, 1]$, for each i and t. Also there is no statistical assumption made about the loss vector l^t. Freund and Schapire [15] showed that Littlestone and Warmuth [23]'s Weighted Majority algorithm can be generalized to handle the on-line allocation problem. They constructed an algorithm, called $Hedge(\beta)$, such that at each time t, $Hedge(\beta)$ chooses the probability distribution vector

$$p^t = \frac{w^t}{\sum_{i=1}^{n} w_i^t},$$

where $w^t = (w_1^t, w_2^t, \ldots, w_n^t)$ such that for each i and parameter $\beta \in [0,1]$,

$$w_i^1 = \frac{1}{n}, \ \ w_i^{t+1} = w_i^t \beta^{l_i^t}.$$

So at each time t, after the loss vector l^{t-1} is received, the $Hedge(\beta)$ will choose the probability distribution vector p^t by updating w^t.

If β is chosen as a function of $\tilde{L}$ and n, Freund and Schapire [15, Lemma 4] showed that

$$E[L^T_{Hedge(\beta)}] \leq \min_{1\leq i\leq n} \sum_{t=1}^{T} l_i^t + \sqrt{2\tilde{L} lnn} + lnn \tag{3.1}$$

holds for all sequence of loss vectors l^1, ..., l^T which satisfies

$$\min_{1\leq i\leq n} \sum_{t=1}^{T} l_i^t \leq \tilde{L}.$$

Dividing both sides of (3.1) by T, we obtain an upper bound for the difference between the expected average loss of $Hedge(\beta)$ and the average loss suffered by the best strategy.

$$\frac{E[L^T_{Hedge(\beta)}]}{T} \leq \frac{1}{T} \min_{1\leq i\leq n} \sum_{t=1}^{T} l_i^t + \frac{\sqrt{2\tilde{L} lnn}}{T} + \frac{lnn}{T}.$$

That is, the average loss of $Hedge(\beta)$ approaches the average loss suffered by the best strategy as $T \rightarrow \infty$.

Note that, to obtain (3.1), the parameter β in $Hedge(\beta)$ depends on $\tilde{L}$. In the next section, we will give a solution of the on-line allocation model based on game theory results when $\tilde{L}$ is not available,

3.2 Application to the on-line allocation model

Our algorithm H for the on-line allocation model is described in Figure 3.1.

Lemma 3.2.1 *For any $\hat{x} \in [0,1]^n$, let $f(\hat{x}) = (f(\hat{x})_1, \ldots, f(\hat{x})_{2^n})$, where*

$$f(\hat{x})_i = \prod_{j=1}^{n} [\hat{x}_j a_{ij} + (1-\hat{x}_j)(1-a_{ij})],$$

and a_{ij} is the element of matrix A defined in Figure 3.1. Then f is a one to one mapping onto its range satisfying Theorem 2.4.1.

. Figure 3.1: Algorithm H

Algorithm H

Choose initial probability vector p^1

Repeat for $t = 2, 3, \ldots$

1. Choose allocation p^t, such that

$$p_j^t = \mu\{\sum_{s=1}^{t-1}(l_j^s - l_i^s) \le \sqrt{\frac{6(t-1)}{2^n}} z(A^i - A^j),\ \forall i\},$$

   where $z = (z_1, \ldots, z_{2^n})$, z_1, ..., z_{2^n} are i.i.d $U(0,1)$ under μ. And $A^1, \ldots, A^n$ are the columns of matrix A.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{2^n 1} & \cdots & a_{2^n n} \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_{2^n} \end{pmatrix} = (A^1, \ldots, A^n),$$

   and $A_1, \ldots, A_{2^n}$ are the distinct sequences from W.

$$W = \{w^n \mid w^n = (w_1, \ldots, w_n),\ w_i \in \{0,1\}, \forall i\}.$$

2. Receive loss vector $l^t \in [0,1]^n$.

Proof. For fixed j, let $\Lambda = \{i : a_{ij} = 1\}$, then

$$\begin{aligned} f(\hat{x})A^j &= \sum_{i=1}^{2^n} x_i^t a_{ij} \\ &= \sum_{i=1}^{2^n} \prod_{k=1}^{n} [\hat{x}_k a_{ik} + (1 - \hat{x}_k)(1 - a_{ik})] a_{ij} \\ &= \hat{x}_j \sum_{i \in \Lambda} \prod_{k \ne j} [\hat{x}_k a_{ik} + (1 - \hat{x}_k)(1 - a_{ik})] \\ &= \hat{x}_j \sum_{w^{n-1} \in \{0,1\}^{n-1}} \prod_{k \ne j} \hat{\hat{x}}_{w_k}^t \end{aligned} \tag{3.2}$$

$$= \hat{x}_j \prod_{k \neq j} [\hat{x}_k + (1 - \hat{x}_k)]$$
$$= \hat{x}_j,$$

where

$$\hat{\hat{x}}_{w_k} = \begin{cases} \hat{x}_k & \text{if } w_k = 1 \\ 1 - \hat{x}_k & \text{if } w_k = 0, \end{cases}$$

and (3.2) follows from the definition of a_{ik}.

□

Theorem 3.2.1 *The modified regret for algorithm H in the on-line allocation problem satisfies*

$$-\sqrt{2^{n-1}}\sqrt{3T} \leq E[L_H^T] - \min_{1 \leq j \leq n} \sum_{t=1}^{T} l_j^t \leq \sqrt{2^{n+1}}\sqrt{3T},$$

for all T *and all loss sequences* $\underline{l}$*, where* $E[L_H^T] = \sum_1^T p^t l^t$.

Proof. Consider the allocation game described in Section 3.1. By Theorem 2.4.1, there is a finite game isomorphic to this allocation game. Take f defined in Lemma 3.2.1 to be the one to one mapping in Theorem 2.4.1.

In the allocation game, at each time t, suppose player I chooses a strategy $\hat{x}^t$ such that

$$\hat{x}^t = (l_1^t, \ldots, l_n^t),$$

and player II's strategy is a random vector $\hat{y}^t$ taking value from a set

$$\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\}$$

such that

$$\hat{y}^t = e_j$$

if

$$\sigma(X^{t-1} + \frac{6(t-1)}{2^n} z) = A^j$$

in the corresponding finite game, where z and A^j are defined in Figure 3.1,

$$X^0 = 0, \; X^t = \sum_1^t x^s, \; x^t = f(\hat{x}^t).$$

It follows from Lemma 3.2.1 and Theorem 2.4.1 that the modified regret of the allocation game is the same as that of the finite game, i.e.,

$$\begin{aligned}&\sum_{t=1}^{T}\hat{x}^t\hat{y}^t - \min_j \sum_{t=1}^{T}\hat{x}_j^t \\ &= \sum_{t=1}^{T} x^t\sigma(X^{t-1} + \frac{6(t-1)}{2^n}z) - \phi(X^T)\end{aligned}$$

It follows from Theorem 2.3.2 that

$$-\sqrt{2^{n-1}3T} \le E_\mu \sum_{t=1}^{T} x^t\sigma(X^{t-1} + \sqrt{\frac{6(t-1)}{2^n}}z) - \phi(X^T) \le \sqrt{2^{n+1}3T}.$$

By the definition of p^t and $\hat{y}^t$, we have for any $j = 1, \ldots, n$,

$$\begin{aligned}&\mu\{\hat{y}^t = e_j\} \\ &= \mu\{\sigma(X^{t-1} + \frac{6(t-1)}{2^n}z) = A^j\} \\ &= \mu\{[\sum_{1}^{t-1} f(\hat{x}^s) + \frac{6(t-1)}{2^n}z]A^j \le [\sum_{1}^{t-1} f(\hat{x}^s) + \frac{6(t-1)}{2^n}z]A^i,\ \forall i\}\end{aligned}$$

It follows from

$$\hat{x}^t = (l_1^t, \ldots, l_n^t), \text{ and } f(\hat{x}^t)A^j = \hat{x}_j^t$$

that

$$\begin{aligned}&\mu\{\hat{y}^t = e_j\} \\ &= \mu\{\sum_{s=1}^{t-1}(l_j^s - l_i^s) \le \sqrt{\frac{6(t-1)}{2^n}}z(A^i - A^j),\ \forall i\} \\ &= p_j^t\end{aligned}$$

Thus

$$\begin{aligned} E_\mu \sum_{t=1}^{T}\hat{x}^t\hat{y}^t - \min_j \sum_{t=1}^{T}\hat{x}_j^t &= \sum_{j=1}^{n} l_j^t p_j^t - \min_j \sum_{t=1}^{T} l_j^t \\ &= E[L_H^T] - \min_j \sum_{t=1}^{T} l_j^t\end{aligned}$$

and the result follows.

□

Remark 3.2.1 *Compared with Hedge(β), Algorithm H not only gives an upper bound for* $L_H^T - \min_{1\le i\le n}\sum_{t=1}^T l_i^t$ *with the rate of* $O(T^{1/2})$, *but also gives a lower bound with the same rate. At each time t, algorithm H determines the probability distribution (allocation) without using either* $\tilde{L}$ *or* T.

The on-line allocation model described above is quite general. One of the applications is the problem of prediction using the advice of a team of experts. Suppose at each time $t = 1, \ldots, T$, the decision-maker must make a prediction and has the knowledge of the predictions made by each of the n experts. Each prediction and the outcome, which is disclosed after the decision-maker has made his prediction, determine the incurred loss.

Let the decision-maker and the n experts select their predictions from a convex set $\mathcal{D}$, and let Θ be the outcome space. Suppose loss function is a function L,

$$L \, : \, \mathcal{D}\times\Theta \to [0,1].$$

At each time t, for each i, denote the prediction of expert B_i by B_i^t, and denote the outcome by y_t. Then $L(B_i^t, y_t)$ is the loss of expert B_i at time t. If at each time t, the decision-maker chooses B_i^t to predict y_t with probability p_i^t, where p_i^t is defined in Algorithm H with $l_i^t = L(B_i^t, y_t)$. Then by Theorem 3.2.1, we have that the expected average loss of the decision-maker approaches the average loss of the best expert at the rate of $O(T^{-1/2})$.

Note that if the loss function L is convex with respect to $d \in \mathcal{D}$, then at time t, let the decision-maker predict with a nonrandomized prediction $\sum_{i=1}^n p_i^t B_i^t$. The convexity implies that

$$L(\sum_{i=1}^n p_i^t B_i^t, y_t) \le \sum_{i=1}^n p_i^t l_i^t.$$

It follows from Theorem 3.2.1 that

$$\sum_{t=1}^T L(\sum_{i=1}^n p_i^t B_i^t, y_t) \le \min_{1\le i\le n} \sum_{t=1}^T l_i^t + \sqrt{2^{n+1}}\sqrt{3T}.$$

We will continue the discussion of prediction using expert advice in Chapter 4.

3.3 Introduction to the multi-armed bandit problem

3.3.1 Component multi-armed bandit game

We consider the following component multi-armed bandit game in which player I selects a reward vector $b \in [0,1]^n$ and player II selects a probability distribution p on n points with gain $L_p = \sum_i^n b_i p_i$. We recognize this as an example where player I's pure strategies are in the cube

$$\hat{\mathbf{X}} = [0,1]^n$$

and player II's pure strategies are in the finite set

$$\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\}.$$

By Theorem 2.4.1, there is a finite game isomorphic to the component multi-armed bandit game.

3.3.2 The multi-armed bandit problem

We are going to study the repeated play of the component multi-armed bandit game, that is, the multi-armed bandit problem. In the multi-armed bandit problem, originally proposed by Robbins [24], a gambler must decide which arm of n non-identical slot machines to play. At each trial, he plays one arm and receives a reward (maybe nonpositive). The goal of the gambler is to maximize his total reward over in a sequence of plays.

Lai and Robbins [22] studied this problem using statistical assumptions about the rewards of the slot machines. They assumed that the distribution of rewards associated with each arm is fixed and does not depend on the number of trials T. They bounded the difference between the expected total rewards of the player and the maximum of the expected total rewards of any arm with $O(logT)$.

Auer, Cesa-Bianchi, Freund and Schapire [2] presented a variant of the bandit problem in which no statistical assumptions are made about the generation of rewards.

They only assume that the rewards are bounded.

They formalized the multi-armed bandit problem as a game between a player choosing actions and an adversary with knowledge of past plays choosing the rewards associated with each action. Assume action space is $\{1, \ldots, n\}$ and all the rewards belong to the interval $[0, 1]$.

They defined the full information game and the partial information game. In the full information game, at each trial $t = 1, 2, \ldots, T$:

1. The adversary selects a vector of the current rewards
$$b^t = (b_1^t, \ldots, b_n^t),$$
where for each i, b_i^t is the reward associated with action i at trial t.

2. Without knowing b^t, the player chooses an action $i_t \in \{1, 2, \ldots, n\}$ and get the corresponding reward $b_{i_t}^t$.

3. The player observes b^t after he makes the action i_t.

The partial information game also consists of three steps. All the steps are the same as that in the full information game except step 3 is replaced by: The player only observes $b_{i_t}^t$ after he makes the action i_t.

They presented an algorithm, called *Hedge*, for the full information game. Algorithm *Hedge* is a slight variant of Algorithm *Hedge*(β) described in Section 3.1.2 The idea of *Hedge* is to choose action i at time t with probability
$$p_i^t = \frac{(1+\alpha)^{s_i^t}}{\sum_{i=1}^n (1+\alpha)^{s_i^t}},$$
where $\alpha > 0$ is a parameter and
$$s_i^{t+1} = \sum_{t=1}^{t} b_i^t.$$
Note that each reward b_i^t is defined as a random variable on the set of player's actions up to trial $t-1$. The measure of the performance of any algorithm, say A, is $E(G_A) - G_{best}$, where
$$E(G_A) = E_{i_1, \ldots, i_T}[\sum_{t=1}^{T} b_{i_t}^t],$$

and

$$G_{best} = \max_{1 \le j \le n} E_{i_1,\ldots,i_T}[\sum_{t=1}^{T} b_j^t]. \tag{3.3}$$

Auer, Cesa-Bianchi, Freund and Schapire [2, Theorem 3.2] showed that in the full information game for $\alpha > 0$,

$$E(G_{Hedge}) \ge G_{best} - \frac{\alpha}{2} G_{best} - \frac{lnn}{\alpha}.$$

For an appropriate choice of α, which is depends on G_{best}, the difference between $E(G_{Hedge})$ and G_{best} is at least $-\sqrt{2Tlnn}$ in the full information game.

Auer, Cesa-Bianchi, Freund and Schapire [2, Section 4] also gave an algorithm for the partial information game based on Hedge.

3.4 Application to the multi-armed bandit problem

In this section, we investigate an algorithm for player in the full information game under the assumption that each reward b_i^t does not depend on past play. Our algorithm is described in Figure 3.2.

Theorem 3.4.1 *If $\tilde{p}^t$ and b^t are defined in Algorithm $\tilde{H}$, then the expected gain of Algorithm $\tilde{H}$ satisfies*

$$-\sqrt{2^{n+1}}\sqrt{3T} \le E(G_{\tilde{H}}) - \max_{1 \le j \le n} \sum_{t=1}^{T} b_j^t \le \sqrt{2^{n-1}}\sqrt{3T},$$

where $E(G_{\tilde{H}}) = \sum_{j=1}^{n} \tilde{p}_j^t b_j^t$.

Proof. The proof is similar to that of Theorem 3.2.1 with loss replaced by gain. By Theorem 2.4.1, there is a finite game isomorphic to this component game. Let f be a one to one mapping in Theorem 2.4.1 associated with matrxi A defined in Figure 3.2.

At each time t, player I chooses a strategy $\hat{x}^t$ such that

$$\hat{x}^t = (b_1^t, \ldots, b_n^t),$$

Figure 3.2: Algorithm $\tilde{H}$

Algorithm $\tilde{H}$

Choose initial probability vector $\tilde{p}^1$

Repeat for $t = 2, 3, \ldots$

1. Choose action i_t according to the distribution $\tilde{p}^t$, where

$$\tilde{p}^t_j = \mu\{\sum_{s=1}^{t-1}(b^s_i - b^s_j) \leq \sqrt{\frac{6(t-1)}{2^n}} z(A^j - A^i), \forall i\},$$

   and $z = (z_1, \ldots, z_{2^n})$, $z_1, \ldots, z_{2^n}$ are i.i.d $U(0,1)$ under μ, and $A^1, \ldots, A^n$ are the columns of matrix A. A is defined in Algorithm H.

2. Receive the reward vector b^t.

---

and player II's strategy is a random vector $\hat{y}^t$ taking value from a set

$$\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\}$$

such that

$$\hat{y}^t = e_j$$

if

$$\sigma(X^{t-1} + \frac{6(t-1)}{2^n} z) = -A^j$$

in the corresponding finite game, where z and A^j are defined in Figure 3.2,

$$X^0 = 0,\ X^t = \sum_1^t x^s,\ x^t = f(\hat{x}^t).$$

It follows from Theorem 2.4.1 that

$$\begin{aligned}
&-\sum_{t=1}^T \hat{x}^t \hat{y}^t + \max_j \sum_{t=1}^T \hat{x}^t_j \\
&= \sum_{t=1}^T x^t \sigma(X^{t-1} + \frac{6(t-1)}{2^n} z) - \phi(X^T).
\end{aligned}$$

It follows from Theorem 2.3.2 that

$$-\sqrt{2^{n-1}3T} \le E_\mu \sum_{t=1}^{T} x^t \sigma(X^{t-1} + \sqrt{\frac{6(t-1)}{2^n}} z) - \phi(X^T) \le \sqrt{2^{n+1}3T}.$$

By the definition of $\tilde{p}^t$ and $\hat{y}^t$, we have for any $j = 1, \ldots, n$,

$$\begin{aligned}
&\mu\{\hat{y}^t = e_j\} \\
&= \mu\{\sigma(X^{t-1} + \frac{6(t-1)}{2^n} z) = -A^j\} \\
&= \mu\{[\sum_1^{t-1} f(\hat{x}^s) + \frac{6(t-1)}{2^n} z](-A^j) \le [\sum_1^{t-1} -f(\hat{x}^s) + \frac{6(t-1)}{2^n} z](-A^i),\ \forall i\}
\end{aligned}$$

It follows from

$$\hat{x}^t = (b_1^t, \ldots, b_n^t), \text{ and } f(\hat{x}^t)A^j = \hat{x}_j^t$$

that

$$\begin{aligned}
&\mu\{\hat{y}^t = e_j\} \\
&= \mu\{\sum_{s=1}^{t-1} (b_i^s - b_j^s) \le \sqrt{\frac{6(t-1)}{2^n}} z(A^j - A^i),\ \forall i\} \\
&= \tilde{p}_j^t
\end{aligned}$$

Thus

$$\begin{aligned}
-E_\mu \sum_{t=1}^{T} \hat{x}^t \hat{y}^t + \max_j \sum_{t=1}^{T} \hat{x}_j^t &= -\sum_{j=1}^{n} b_j^t \tilde{p}_j^t + \max_j \sum_{t=1}^{T} b_j^t \\
&= -E(G_{\tilde{H}}) + \max_j \sum_{t=1}^{T} b_j^t.
\end{aligned}$$

Therefore, Theorem 3.4.1 follows.

□

Remark 3.4.1 *Algorithm $\tilde{H}$ gives both an upper bound and a lower bound of the modified regret at the rate of $O(T^{1/2})$ without knowledge of T or G_{best}.*

Suppose the player has access to the opinion of a set of K experts. At each trial t, before choosing an action, the player is provided with a set of K probability vectors $(\xi^t(1), \ldots, \xi^t(K))$. For each j, $\xi^t(j)$ is the advice of expert j on trial t, that is

$$\xi^t(j) = (\xi_1^t(j), \ldots, \xi_n^t(j)),$$

and $\xi_i^t(j)$ is the recommended probability of choosing action i by expert j. After receiving the reward vector b^t, the expected reward for expert j is $\xi^t(j)b^t$.

At each time t, we apply Algorithm $\tilde{H}$ with new reward vector

$$(\xi^t(1)b^t, \ldots, \xi^t(K)b^t).$$

It follows from Theorem 3.4.1 that

$$-\sqrt{2^{K+1}}\sqrt{3T} \leq E(G_{\tilde{H}}) - \max_{1 \leq j \leq K} \sum_{t=1}^{T} \xi^t(j)b^t \leq \sqrt{2^{K-1}}\sqrt{3T}.$$

Therefore, using Algorithm $\tilde{H}$, the expected average reward of the player will approach to

$$\max_{1 \leq j \leq K} \sum_{t=1}^{T} \xi^t(j)b^t,$$

which is the average reward of the best expert, at the rate of $O(T^{-1/2})$.

Chapter 4

Prediction using expert advice

4.1 Introduction

In this chapter we review some of the literature dealing with prediction using expert advice. Section 4.2 is a self-contained review. However, the main thrust of this chapter is to show how the Hannan recursive strategies apply in a prediction problem considered by Foster and Vohra [13]. In particular, in Section 4.3 we prove that a strategy for $n = 2$ experts investigated by Foster and Vohra [13] is an exmaple of a Hannan strategy and we deduce a slightly stronger result than was established in their Theorem 1. Section 4.4 gives a solution for the n expert case. Finally, Section 4.5 applies the k-extended approach to the prediction problem.

4.2 Literature review

Consider the following prediction problem. Suppose no statistical assumptions made about the actual sequence

$$\underline{y} = (y_1, \ldots, y_T) \tag{4.1}$$

of outcomes that is observed. At each time $t = 1, \ldots,$ the decision-maker must predict the value of y_t. Before making the prediction, the decision-maker is given the predictions of n experts.

This decision problem can be viewed as the following game between two players, the decision-maker and nature. At each time $t = 1, 2, \ldots$,

1. Each expert B_i, $i = 1, \ldots, n$, makes a prediction $B_i^t \in \mathcal{D}$, where $\mathcal{D}$ is the prediction space.

2. The decision-maker, who has the knowledge of all B_i^s, $i = 1, \ldots, n$, $s \leq t$, and past outcomes $y_1, \ldots, y_{t-1}$, makes his prediction $\hat{y}_t \in \mathcal{D}$.

3. The nature chooses some outcome $y_t \in \Theta$, where Θ is the outcome space.

4. Each expert B_i, $i = 1, \ldots, n$, incurs loss $L(B_i^t, y_t)$ and the decision-maker incurs loss $L(\hat{y}_t, y_t)$, where $L : \mathcal{D} \times \Theta \to [0, \infty)$ is the loss function.

Suppose at each time t, the decision-maker uses a prediction algorithm A to make prediction $\hat{y}_t$, then loss of the decision-maker is equal to the loss of the algorithm A. Define the total loss of the algorithm A on a sequence of trials with respect to a sequence of outcomes $\underline{y}$ to be

$$L_A^T(\underline{y}) = \sum_{t=1}^{T} L(\hat{y}_t, y_t), \tag{4.2}$$

where $\underline{y}$ is defined in (4.1). Similarly the total loss of the expert B_i with respect to $\underline{y}$ is defined to be

$$L_{B_i}^T(\underline{y}) = \sum_{t=1}^{T} L(B_i^t, y_t).$$

The goal of the decision-maker is to find an algorithm A to minimize

$$L_A^T(\underline{y}) - \min_{1 \leq j \leq n} L_{B_j}^T(\underline{y}). \tag{4.3}$$

The goal of the nature is to maximize (4.3). So the min/max strategy for the decision-maker is the algorithm that minimizes the maximum of (4.3) over all outcome sequences.

Vovk [27] introduced a general on-line prediction algorithm when the outcomes are binary. For the case with continuous-valued predictions, Vovk proved for a large class of loss functions bounds of the form

$$L_A^T(\underline{y}) - \min_{1 \leq j \leq n} L_{B_j}^T(\underline{y}) \leq c_L \ln n,$$

where c_L is a positive constant determined by the loss function L.

In the rest of this section, we give a literature review according to the following four cases.

Case 1. $\mathcal{D} = [0,1]$, $\Theta = \{0,1\}$ and $\{B_1, \ldots, B_n\}$ is the finte set of experts.

Haussler, Kivinen and Warmuth [18] studied the *Generic* Algorithm first introduced by Vovk. At each time t, the *Generic* Algorithm predicts with any value $\hat{y}_t$ that satisfies for $y = 0$ and $y = 1$ the condition

$$L(\hat{y}_t, y) \leq -cln \sum_{i=1}^{n} \frac{w_{t,i} e^{-\eta L(B_i^t, y)}}{\sum_{i=1}^{n} w_{t,i}}, \tag{4.4}$$

where c and η are any two positive constants, for any i,

$$w_{1,i} > 0, \;\; w_{t+1,i} = w_{t,i} e^{-\eta L(B_i^t, y_t)}.$$

Haussler, Kivinen and Warmuth [18, Theorem 3.11] showed that if L is a loss function such that

$$c_L := \sup_{0<z<1} \frac{L_0'(z) L_1'(z)^2 - L_1'(z) L_0'(z)^2}{L_0'(z) L_1''(z) - L_1'(z) L_0''(z)} < \infty, \tag{4.5}$$

where $L_0(z) = L(0, z)$ and $L_1(z) = L(1, z)$, then it follows from by applying the *Generic* Algorithm with $w_{1,i} = 1$, $c = c_L$ and $\eta = 1/c_L$ that for any T,

$$\sup_{\substack{\underline{y} \in \{0,1\}^n \\ B^t \in [0,1]^n}} [L_{Generic}^T(\underline{y}) - \min_{1 \leq j \leq n} L_{B_j}^T(\underline{y})] \leq c_L lnn, \tag{4.6}$$

where $B^t = (B_1^t, \ldots, B_n^t)$.

- For logarithmic loss, that is defined by $L(\hat{y}, y) = yln\frac{y}{\hat{y}} + (1-y)ln\frac{1-y}{1-\hat{y}}$, (4.6) holds with $c_L = 1$.

- For squared loss, that is defined by $L(\hat{y}, y) = (\hat{y} - y)^2$, (4.6) holds with $c_L = 1/2$.

  Foster [12] also proposed a prediction method with the loss function defined as squared loss. At each time t, the decision-maker predicts with

$$\hat{y}_t = \sum_{i=1}^{n} w_{t,i} B_i^t,$$

where $(w_{t,1}, \ldots, w_{t,n})$ minimizes

$$\sum_{s=1}^{t-1}(\sum_{i=1}^{n} w_i B_i^s - y_s)^2 + \sum_{i=1}^{n} w_i^2$$

over all probability vectors w. It follows from Foster [12, Theorem 1] that for any outcome sequence $\underline{y}$ and any probability vector w,

$$\sum_{t=1}^{T}(\hat{y}_t - y_t)^2 - \sum_{t=1}^{T}(\sum_{i=1}^{n} w_i B_i^t - y_t)^2 \leq 2 + nlnn(T+1).$$

Therefore using this strategy, the decision-maker can perform as well as any convex combination of the n experts, namely, the difference between the average loss of the decision-maker and the average loss of any convex combination of the n experts will converge to 0 as T goes to ∞.

- For absolute loss, that is defined by $L(\hat{y}, y) = |\hat{y} - y|$, $c_L = \infty$. Then the *Generic* Algorithm can not be applies directly. Cesa-Bianchi et al [7] studies some variants of the *Generic* Algorithm for the absolute loss.

  Cesa-Bianchi et al [7] constructed a min/max strategy, called Algorithm MM, for the decision-maker when the loss function is absolute loss. At each time $t = 1, \ldots, T$, Algorithm MM predicts with

$$\hat{y}_t = \frac{v(M^r + Z^r, r-1) - v(M^r + \mathbf{1} - r, r-1) + 1}{2},$$

  where

$$r = T + 1 - t, \; Z^r = (B_1^t, \ldots, B_n^t)$$

$$M^r = (M_1^r, \ldots, M_n^r), \; M_j^T = \mathbf{0} \; M_j^t = \sum_{s=1}^{t-1} |B_j^s - y_s|, \; \forall j,$$

  and v is defined inductively as

$$v(M, 0) = \min_{1 \leq j \leq n} M_j,$$

$$v(M, r) = \min_{z \in [0,1]} \frac{v(M + z, r-1) + v(M + \mathbf{1} - z, r-1)}{2}.$$

  Cesa-Bianchi et al [7, Theorem 2] showed that for any set of n experts and for any outcome sequence $\underline{y}$,

$$L_{MM}^T(\underline{y}) - \min_{1 \leq j \leq n} L_{B_j}^T(\underline{y}) \leq \frac{T}{2} - v(0, T),$$

and Algorithm MM achieves the value of the game, $\frac{T}{2} - v(0,T)$. So Algorithm MM is a min/max strategy.

The disadvantage of the Algorithm MM is that T must be known in the beginning of the game. And the algorithm is computational expensive since the calculation of $v(M,r)$ involves minimizing a recursively defined function over all choices of $z \in [0,1]^n$. Therefore simple algorithms, Algorithm P, Algorithm P' and Algorithm P^*, were introduced so that they can be implemented efficiently.

Algorithm P works as following. At each time t, Algorithm P makes a prediction

$$\hat{y}_t = F_\beta(r_t),$$

where $\beta \in [0,1)$,

$$r_t = \frac{\sum_{i=1}^n w_{t,i} B_i^t}{\sum_{i=1}^n w_{t,i}},$$

$$w_{t,i} = 1, \ \ w_{t+1,i} = w_{t,i} U_\beta(|B_i^t - y_t|), \ \ \forall i,$$

and $F_\beta(r)$ and $U_\beta(q)$ be any functions such that

$$1 + \frac{ln((1-r)\beta + r)}{2ln(\frac{2}{1+\beta})} \leq F_\beta(r) \leq \frac{-ln(1-r+r\beta)}{2ln(\frac{2}{1+\beta})},$$

for $0 \leq r \leq 1$, and

$$\beta^q \leq U_\beta(q) \leq 1 - (1-\beta)q,$$

for $0 \leq q \leq 1$.

The performance of Algorithm P depends on the parameter β. Cesa-Bianchi et al [7] showed how to choose β in according to the type of knowledge available to the decision-maker.

- ⋆ If the decision-maker knows an upper bound on the total loss of the best expert, Cesa-Bianchi et al [7, Theorem 15] showed that for any $K \geq 0$, taking $\beta = g(\sqrt{\frac{lnn}{K}})$, where

$$g(z) = \frac{1}{1 + 2z + \frac{z^2}{lnz}}, \tag{4.7}$$

for any set of n experts and any outcome sequence $\underline{y}$ such that

$$\min_{1\leq j\leq n} L^T_{B_j}(\underline{y}) \leq K,$$

we have

$$L^T_P(\underline{y}) - \min_{1\leq j\leq n} L^T_{B_j}(\underline{y}) \leq \sqrt{Klnn} + \frac{log_2 n}{2}.$$

⋆ If T is known to the decision-maker in advance, then use a slight variant of Algorithm P by adding a new expert B_{n+1}, where at time t, $B^t_{n+1} = 1 - B^t_1$. Denote the algorithm that uses the expanded $n+1$ experts by Algorithm P'. Cesa-Bianchi et al [7, Theorem 16] showed that taking $\beta = g(\sqrt{\frac{2ln(n+1)}{T}})$, where g is defined in (4.7), for any set of n experts and any outcome sequence $\underline{y}$ of length T, we have

$$L^T_{P'}(\underline{y}) - \min_{1\leq j\leq n} L^T_{B_j}(\underline{y}) \leq \sqrt{\frac{Tln(n+1)}{2}} + \frac{log_2(n+1)}{2}.$$

⋆ If there is no prior knowledge about the upper bound on the total loss of the best expert or the length T of the sequence, the following procedure, called Algorithm P^*, can be used. For $z = 0$ to ∞, Algorithm P^* repeatedly runs Algorithm $P(g(\sqrt{\frac{lnn}{k_z}}))$ until the total loss exceeds b_z, where g is defined in (4.7),

$$k_z = 4(\frac{1+\sqrt{5}}{2})^{2z} lnn,$$

and

$$b_z = k_z + \sqrt{k_z lnn} + \frac{log_2 n}{2}.$$

Cesa-Bianchi et al [7, Corollary 22] showed that if $n \geq 7$, then for any outcome sequence $\underline{y}$,

$$L^T_{P^*}(\underline{y}) - \min_{1\leq j\leq n} L^T_{B_j}(\underline{y}) \leq 4\sqrt{\min_{1\leq j\leq n} L^T_{B_j}(\underline{y}) lnn} + 2.8 lnn.$$

Case 2. $\mathcal{D} = [0,1]$, $\Theta = [0,1]$ and $\{B_1, \ldots, B_n\}$ is the finite set of experts.

Haussler, Kivinen and Warmuth [18, Theorem 4.2] showed that for loss function such that

$$\frac{\partial^2 g(y,a,b)}{\partial y^2} + (\frac{\partial g(y,a,b)}{\partial y})^2 \geq 0 \quad (4.8)$$

holds for all $y, a, b \in [0,1]$, where

$$g(y,a,b) = \frac{L(y,a)}{c_L} - \frac{L(y,b)}{c_L},$$

c_L is defined in (4.5), and $c_L < \infty$, the *Generic* Algorithm satisfies

$$\sup_{\substack{\underline{y} \in \{0,1\}^n \\ B^t \in [0,1]^n}} [L^T_{Generic}(\underline{y}) - \min_{1\leq j\leq n} L^T_{B_j}(\underline{y})] \leq c_L ln n, \tag{4.9}$$

where $B_t = (B^t_1, \ldots, B^t_n)$.

- For logarithmic loss, that is defined by $L(\hat{y}, y) = y ln\frac{y}{\hat{y}} + (1-y) ln\frac{1-y}{1-\hat{y}}$, (4.8) holds and $c_L = 1$. Then (4.9) follows.

- squared loss, that is defined by $L(\hat{y}, y) = (\hat{y} - y)^2$, (4.8) holds and $c_L = 1/2$. Then (4.9) follows.

- For absolute loss, that is defined by $L(\hat{y}, y) = |\hat{y} - y|$, we have $c_L = \infty$. Then we can not apply the *Generic* Algorithm. Haussler, Kivinen and Warmuth [18] constructed an new algorithm, the *Vee* Algorithm, that predicts with any value $\hat{y}_t$ that satisfies the condition

$$\begin{aligned} &\max_{y\in Y}\{y + [ln(\sum_{i=1}^{n} \frac{w_{t,i}e^{-\eta|B^t_i - y|}}{\sum_{i=1}^n w_{t,i}})]/[2ln\frac{2}{1+e^{-\eta}}]\} \\ &\leq \hat{y}_t \\ &\leq \min_{y\in Y}\{y - [ln(\sum_{i=1}^{n} \frac{w_{t,i}e^{-\eta|B^t_i - y|}}{\sum_{i=1}^n w_{t,i}})]/[2ln\frac{2}{1+e^{-\eta}}]\}, \end{aligned}$$

where for any i,

$$w_{1,i} > 0, \quad w_{t+1,i} = w_{t,i}e^{-\eta L(B^t_i, y_t)},$$

$$Y = \{0, 1, B^t_1, \ldots, B^t_n\}.$$

Haussler, Kivinen and Warmuth [18, Theorem 4.7] showed that for any outcome sequence $\underline{y}$ and any i, the *Vee* algorithm satisfies

$$L^T_{Vee}(\underline{y}) \leq [-ln\frac{w_{t,i}}{\sum_{i=1}^n w_{t,i}} + \eta \min_{1\leq j\leq n} L^T_{B_l}(\underline{y})]/[2ln\frac{2}{1+e^{-\eta}}].$$

Case 3. $\mathcal{D} = [0,1]$, $\Theta = \{0,1\}$, and there are unaccountably infinite set of experts.

Freund [14] generalized the Weighted Majority algorithm of Littlestone and Warmuth [23] to the case in which there are unaccountably infinite set of experts. The algorithm he used is called the exponential weights (EW) algorithm. The EW Algorithm gives a prediction $\hat{y}_t$ that is any value in $[0,1]$, such that for $y \in \{0,1\}$,

$$L(\hat{y}_t, y_t) \le -c[\int_0^1 e^{-\frac{1}{c}L(p,y)} d\mu_t(p)]/[\int_0^1 d\mu_t(p)], \tag{4.10}$$

where for each t, c is a positive parameter, and the measure $\mu_t(A)$ is defined as

$$\mu_{t+1}(A) = \int_A e^{-\frac{1}{c}L(p,y_t)} d\mu_t(p),$$

and μ_1 is a probability measure on [0, 1].

Assumption 4.2.1 *Suppose the loss function satisfies the following properties*

- $\exists\, c$ *and* $\eta = 1/c$, *such that (4.4) holds.*
- $\forall\, y$, $L(p,y)$ *has a continuous second derivative with respect to* p.
- $\exists\, \hat{p} : [0,1] \to [0,1]$, $\hat{p}(\hat{\theta})$ *is the unique minimizer of*

  $$\sum_{t=1}^{T} L(p, y_t)$$

  over all outcome sequence $\underline{y}$ *whose empirical distribution is* $\hat{\theta}$.

Choose the initial probability measure to be

$$\mu_1(A) = \int_A w(x)dx,$$

where

$$w(x) = -\frac{1}{Z}[\frac{\partial^2}{\partial p^2} g(x,p)]_{p=\hat{p}(x)},$$

$$Z = \int_0^1 [\frac{\partial^2}{\partial p^2} g(x,p)]_{p=\hat{p}(x)} dx, \tag{4.11}$$

and

$$g(x,p) = \frac{1}{c}[x(L(p,1) - L(\hat{p},1)) + (1-x)(L(p,0) - L(\hat{p},0))].$$

Freund [14, Theorem 1] showed that for all outcome sequence $\underline{y}$ whose empirical distribution is $\hat{\theta}$,

$$\max_{\underline{y}:\frac{1}{T}\sum_1^T y_t=\hat{\theta}} [L_{EW}^T(\underline{y}) - \min_{p\in[0,1]}\sum_{t=1}^T L(p,y_t)] \le \frac{c}{2}ln\frac{T}{2\pi} - \frac{c}{2}lnZ + O(\frac{1}{T}),$$

where Z is defined in (4.11).

- For log-loss, that is defined by $L(d,\theta) = -log(1-|d-\theta|)$, Assumption 4.2.1 holds with $c = 1$. Freund [14, Theorem 3] showed that

$$\max_{\underline{y}}[L_{EW}^T(\underline{y}) - \min_{p\in[0,1]}\sum_{t=1}^T L(p,y_t)] \le \frac{1}{2}ln(T+1) + 1.$$

- For square loss, that is defined by $L(d,\theta) = (d-\theta)^2$, Assumption 4.2.1 holds with $c = 1/2$. Freund [14, Theorem 4] showed that

$$\max_{\underline{y}}[L_{EW}^T(\underline{y}) - \min_{p\in[0,1]}\sum_{t=1}^T L(p,y_t)] \le \frac{1}{4}lnT + \frac{1}{2}ln\frac{2}{erf(\sqrt{2})} - \frac{1}{4}ln\frac{\pi}{2},$$

  where $erf(p) = \frac{2}{\pi}\int_0^p e^{-x^2}dx$.

- For absolute loss, that is defined by $L(d,\theta) = |d-\theta|$, we have $c = \infty$. Since

$$\min_{p\in[0,1]}\sum_{t=1}^T |p-y_t| = \min_{p\in\{0,1\}}\sum_{t=1}^T |p-y_t|,$$

  this prediction problem can be treated as a prediction problem using two expert advice.

Case 4. $L(d,\theta) \le K$, for all $d \in \mathcal{D}$ and $\theta \in \Theta$.

Foster and Vohra [13] studied the problem of choosing between two expert forecasts. Suppose that B_1 and B_2 are two experts with bounded forecasting errors or called loss. At each time $t = 1, \ldots, T$, the decision-maker has the knowledge of the past loss incurred by Experts B_1 and B_2. Foster and Vohra [13] constructed a randomized strategy C from B_1 and B_2 based on their past average loss. They indicated that the difference between the average loss of C and the minimum average loss of B_1 and B_2 will converge to 0 in probability at the rate of $O(T^{-1/2})$. We will give a detail description of the strategy C and a proof in Section 4.2 using game theory results.

One of the applications of the on-line allocation problem studied by Freund and Schapire [15] is predicting using expert advice when the loss function is bounded. By the *Hedge*(β) Algorithm, at each time t, the decision-maker predicts with $\hat{y}_t = B_i^t$ with probability p_i^t, where B_i^t is the prediction of Expert B_i at time t,

$$p_i^t = \frac{w^t}{\sum_{i=1}^n w_i^t},$$

where the parameter $\beta \in [0,1]$, for any i,

$$w_i^1 = 1/n, \;\; w_i^{t+1} = w_i^t \beta^{L(B_i^t, y_t)}.$$

Freund and Schapire [15, Lemma 4] showed for all outcome sequence of $\underline{y}$ such that

$$\min_{1 \leq j \leq n} L_{B_j}^T(\underline{y}) \leq \tilde{L},$$

choosing β as a function of $\tilde{L}$,

$$E[L_{Hedge(\beta)}^T(\underline{y})] - \min_{1 \leq j \leq n} L_{B_j}^T(\underline{y}) \leq \sqrt{2\tilde{L}lnn} + lnn,$$

where

$$E[L_{Hedge(\beta)}^T(\underline{y})] = \sum_{t=1}^T \sum_{i=1}^n p_i^t L(B_i^t, y_t),$$

is the expected total loss of the *Hedge*(β) Algorithm over $\underline{y}$.

In Chapter 3, we use game theory results to construct an algorithm H for the decision-maker when the loss function is bounded. Without loss of generality, for any $d \in \mathcal{D}$, and $\theta \in \Theta$, let

$$0 \leq L(d, \theta) \leq 1.$$

At each time t, the decision-maker predicts $\hat{y}_t = B_i^t$ with probability p_i^t, where p_i^t is defined in Algorithm H with $l_i^t = L(B_i^t, y_t)$. Theorem 3.2.1 shows that for all outcome sequence of $\underline{y}$,

$$-\sqrt{2^{n-1}}\sqrt{3T} \leq E[L_H^T(\underline{y})] - \min_{1 \leq i \leq n} \sum_{t=1}^T L_{B_i}^T(\underline{y}) \leq \sqrt{2^{n+1}}\sqrt{3T},$$

where

$$E[L_H^T(\underline{y})] = \sum_{t=1}^T \sum_{i=1}^n p_i^t L(B_i^t, y_t),$$

is the expected total loss of Algorithm H over $\underline{y}$.

4.3 A proof of Theorem 1 of Foster and Vohra

Suppose B_1 and B_2 are two experts. At each time t, let B_1^t and B_2^t be their predictions of outcome y_t, respectively. Suppose for $i = 1, 2$,

$$B_i^t \in \mathcal{D}, \; y_t \in \Theta,$$

and the loss function L is bounded. Without loss of generality, for any $d \in \mathcal{D}$, and $\theta \in \Theta$, let

$$0 \leq L(d, \theta) \leq 1.$$

For $i = 1, 2$, denote

$$b_i^t = L(B_i^t, y_t).$$

In their paper, Foster and Vohra [13] constructed a randomized strategy $\tilde{C}$ for the decision-maker. At each time t, the decision-maker predicts with

$$\hat{y}_t = \begin{cases} B_1^t & \text{with probability } \min(\max[0, \frac{D_{t-1}+(t-1)^s}{2(t-1)^s}], 1) \\ B_2^t & \text{otherwise,} \end{cases}$$

where s is a constant that satisfies $0.5 \leq s < 1$,

$$D_t := \sum_{k=1}^{t} b_2^k - \sum_{k=1}^{t} b_1^k. \tag{4.12}$$

Now we will write the randomized strategy $\tilde{C}$ in the following equivalent way. At each time t, the decision-maker predicts with

$$\hat{y}_t = \begin{cases} B_1^t & \text{if } D_{t-1} > 2(t-1)^s(z - 1/2) \\ B_2^t & \text{if } D_{t-1} \leq 2(t-1)^s(z - 1/2), \end{cases}$$

where z is $U(0, 1)$ random variable under μ.

Foster and Vohra indicated that if the decision-maker uses the randomized strategy $\tilde{C}$, then the difference between the average loss of the decision-maker and the minimum average loss of the two experts will be bounded by a nonnegative random variable ϵ_T, and ϵ_T goes to 0 in probability as T goes to ∞. On its face, this randomized strategy $\tilde{C}$ is similar to the Hannan [17]'s recursive strategies for player II in a two-person finite game, that is player II plays Bayes versus a randomized perturbation of the player I 's empirical distribution. We will prove Theorem 1 of Foster and Vohra [13] using the game theory results of Chapter 2.

Theorem 1 of Foster and Vohra [13] *If B_1 and B_2 are two experts with bounded loss b_1^t, b_2^t for all t, and the strategy $\tilde{C}$ is defined as above, then*

$$\frac{1}{T}\sum_1^T \tilde{c}_t \leq \frac{1}{T}\min\left(\sum_1^T b_1^t, \sum_1^T b_2^t\right) + \epsilon_T,$$

and

$$0 \leq \epsilon_T \xrightarrow{P} 0,$$

where $\tilde{c}_t$ is the loss of strategy $\tilde{C}$ at time t.

Proof. Without loss of generality, for all t, let

$$0 \leq b_1^t \leq 1 \text{ and } 0 \leq b_2^t \leq 1.$$

Consider a finite two-person game in which player I has 4 pure strategies, player II has 2 pure strategies, and the loss matrix A is defined as following:

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Let $z = (z_1, z_2, z_3, z_4)$. Under a probability measure μ, z_1, z_2, z_3 and z_4 are independent. z_i is $U(0,1)$ for $i = 1, 2$ and 4. z_3 is degenerate at $1/2$.

Since $A^{qr} = A^q - A^r = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$, for $1 \leq q < r \leq 2$, (2.4) becomes

$$\mu\{z \mid t_1 \leq z_2 - z_3 \leq t_2\} \leq t_2 - t_1.$$

So (2.4) holds with $L = 1$. $\theta = E_\mu \|z\| = m/2 = 2$.

Consider that player I chooses randomized strategies $\underline{x}$ from the class $\tilde{\mathbf{X}}$ that consists of all x^t such that for each t,

$$x^t = \{(1-b_1^t)(1-b_2^t), b_1^t(1-b_2^t), (1-b_1^t)b_2^t, b_1^t b_2^t\}.$$

For any t and $i = 1, 2$,

$$x^t A^i = b_i^t.$$

Then the Bayes envelope $\phi(X^T)$ becomes

$$\phi(X^T) = \min_{\sigma \in \{A^1, A^2\}} X^T \sigma = \min_{i=1,2} \sum_{t=1}^{T} b_i^t.$$

If $H^t = 2t^s$ and z is defined as above, we have

$$(X^{t-1} + H^{t-1}z)A^1 = \sum_{k=1}^{t-1} b_1^k + 2(t-1)^s(z_2 + z_4),$$

and

$$(X^{t-1} + H^{t-1}z)A^2 = \sum_{k=1}^{t-1} b_2^k + 2(t-1)^s(z_3 + z_4).$$

Then the Bayes response $\sigma(X^{t-1} + H^{t-1}z)$ becomes

$$\sigma(X^{t-1} + H^{t-1}z) = \begin{cases} A^1 & \text{if } D_{t-1} > 2(t-1)^s(z_2 - 1/2) \\ A^2 & \text{if } D_{t-1} \leq 2(t-1)^s(z_2 - 1/2), \end{cases}$$

where D_{t-1} is defined in (4.12).

Since

$$x^t \sigma(X^{t-1} + H^{t-1}z) = \begin{cases} b_1^t & \text{if } D_{t-1} > 2(t-1)^s(z_2 - 1/2) \\ b_2^t & \text{if } D_{t-1} \leq 2(t-1)^s(z_2 - 1/2), \end{cases}$$

the definition of the strategy $\tilde{C}$ implies that

$$\tilde{c}_t = x^t \sigma(X^{t-1} + H^{t-1}z).$$

Then applying Theorem 2.3.3 with $m = 4$, $n = 2$, $|B| = 1$, $L = 1$, $\theta = 2$, $H^t = 2t^s$ and for any player I's strategy x^t in $\tilde{\mathbf{X}}$, we have

$$E_\mu |\sum_1^T \tilde{c}_t - \min(\sum_1^T b_1^t, \sum_1^T b_2^t)|$$

$$= E_\mu|\sum_1^T x^t\sigma(X^{t-1} + H^{t-1}z) - \phi(X^T)|$$

$$\leq 2T^s 2|B| + \frac{2^2}{2}(\sum_1^T \frac{2}{2t^s} - T^{1-s})|B|$$

$$\leq 4T^s + \frac{2s}{1-s}T^{1-s}.$$

By Markov inequality, we have that

$$\frac{1}{T}|\sum_1^T \tilde{c}_t - \min(\sum_1^T b_1^t, \sum_1^T b_2^t)| \xrightarrow{P} 0.$$

Hence the desired result follows.

□

Remark 4.3.1 $\frac{1}{T}|\sum_1^T \tilde{c}_t - \min(\sum_1^T b_1^t, \sum_1^T b_2^t)| \xrightarrow{P} 0$ *is slightly stronger than the result of Foster and Vohra [13, Theorem 1]. The optimal choice of* s *is 1/2, and the optimal convergence rate is* $O(T^{-1/2})$.

4.4 A generalization to more than two experts

In the last section, we have proved that given the predictions of two experts, we can construct a randomized strategy for the decision-maker such that the decision-maker performs as well as the best experts in the sense that the difference between the average loss of the decision-maker and the average loss of the best expert converges to 0 in probability. In this section we consider how to construct a randomized strategy for the decision-maker when there are more than two experts.

Suppose at time t, $B_1^t, \ldots, B_n^t$ are the predictions of n experts, respectively. For any t and i, let outcome $y_t \in \Theta$, $B_i^t \in \mathcal{D}$. Suppose the loss function L is bounded. Denote

$$b_i^t = L(B_i^t, y_t), \ \forall\, i, t.$$

Without loss of generality, we assume that $0 \leq b_i^t \leq 1$ for all i and t.

Suppose at each time t, for all i, the decision-maker has the knowledge of $b_i^1, \ldots, b_i^{t-1}$. We define a randomized strategy C for the decision-maker, which is described in Figure 4.1.

Figure 4.1: Strategy C

Strategy C

Choose initial prediction from the set of $\{B_i^1,\ i = 1, \ldots, n\}$

Repeat for $t = 2, 3, \ldots$

predict with B_j^t if for any i,

$$\sum_{t=1}^{t-1}(b_j^t - b_i^t) \leq \sqrt{\frac{3n^2(t-1)}{2^n}}z(A^i - A^j),$$

where $z = (z_1, \ldots, z_{2^n})$, and $z_1, \ldots, z_{2^n}$ are i.i.d. $U(0,1)$ random variables under μ, $A^1, \ldots, A^n$ are the columns of matrix A.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{2^n1} & \cdots & a_{2^nn} \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_{2^n} \end{pmatrix}.$$

And $A_1, \ldots, A_{2^n}$ are the distinct sequences from W.

$$W = \{w^n \mid w^n = (w_1, \ldots, w_n),\ w_i \in \{0,1\}, \forall i\}.$$

In the next theorem we will show that using strategy C, the decision-maker will perform as well as the best experts among the n experts.

Theorem 4.4.1 *Let c_t be the loss of the strategy C at time t, b_i^t be the loss of the expert B_i at time t with*

$$0 \leq b_i^t \leq 1\ \forall\ i, t.$$

Then as T goes to ∞,

$$\frac{1}{T}\left|\sum_{1}^{T} c_t - \min_{1\leq i\leq n}\sum_{t=1}^{T} b_i^t\right| \xrightarrow{P} 0,$$

and the convergence rate is $O(T^{-1/2})$.

Proof. We consider the two-person game infinite game described in Section 2.4. By Theorem 2.4.1, there is a finite game isomorphic to this infinite game. Let f be a one to one mapping in Theorem 2.4.1 associated with the loss matrix A defined in Figure 4.1..

In the infinite game, at each time t, suppose player I chooses a strategy $\hat{x}^t$ such that

$$\hat{x}^t = (b_1^t, \ldots, b_n^t),$$

and player II's strategy is a random vector $\hat{y}^t$ taking value from a set

$$\hat{\mathbf{Y}} = \{e_1, \ldots, e_n\}$$

such that

$$\hat{y}^t = e_j$$

if

$$\sigma(X^{t-1} + H^{t-1}z) = A^j$$

in the corresponding finite game, where z and A^j are defined in Figure 4.1,

$$X^0 = 0,\ \ X^t = \sum_1^t x^s,\ \ x^t = f(\hat{x}^t),$$

and

$$H^t = \sqrt{\frac{3n^2t}{2^n}} \quad t = 1, 2, \ldots$$

It follows from Theorem 2.4.1 that the modified regret of the allocation game is the same as that of the finite game, i.e.,

$$\sum_{t=1}^T \hat{x}^t \hat{y}^t - \min_j \sum_{t=1}^T \hat{x}_j^t$$
$$= \sum_{t=1}^T x^t \sigma(X^{t-1} + H^{t-1}z) - \phi(X^T)$$

By the definition of $\hat{y}^t$, we have

$$\sum_1^T c_t - \min_{1 \le j \le n} \sum_1^T b_j^t$$
$$= \sum_{t=1}^T x^t \sigma(X^{t-1} + H^{t-1}z) - \phi(X^T)$$

It follows from Theorem 2.3.4 that

$$E_\mu \left| \sum_{t=1}^T x^t \sigma(X^{t-1} + H^{t-1}z) - \phi(X^T) \right| \le \sqrt{3n^2 2^n T}.$$

Thus

$$E_\mu|\sum_1^T c_t - \min_{1\le j\le n}\sum_1^T b_j^t| \le \sqrt{3n^2 2^n T}.$$

By Markov inequality, we have that

$$\frac{1}{T}|\sum_1^T c_t - \min_{1\le j\le n}\sum_1^T b_j^t| \xrightarrow{P} 0.$$

Hence the desired result follows.

□

4.5 The k-extended prediction strategies

4.5.1 Introduction

In the on-line prediction model described in Section 4.2, at each time t, the decision-maker predicts the outcome $y_t \in \Theta$. The decision-maker has the knowledge of the predictions made by each of n experts, and makes a prediction based on the past and current expert predictions and the past outcome sequence. The goal of the decision-maker is to find a prediction strategy such that the total loss is as small as possible. Since no statistical assumptions are made about the distribution of the outcome sequence, a reasonable goal for the decision-maker is to perform as well as the best expert. In Chapter 3 and the first four sections of Chapter 4, different strategies have been introduced to give upper bounds on the difference between the total loss of the decision-maker and the total loss incurred by the best expert such that the average loss of the decision-maker approaches the average loss incurred by the best expert as T goes to ∞. We observe that the average loss incurred by the best expert is equal to the Bayes envelope in a finite two-person game. It is of interest to find a strategy for the decision-maker such that the average loss of the decision-maker approaches a lower envelope as T goes to ∞.

Herbster and Warmuth [20] considered the prediction model where $\mathcal{D} = [0,1]$, $\Theta = [0,1]$ and loss function is L. They studied the case in which the outcome sequence is divided into at most $k+1$ arbitrary segments. Each segment has a best expert. The sequences of segments and its associated sequence of best experts is called a partitioning. Now the goal of the decision-maker is to perform well relative to the best partition.

Let $\underline{y} = (y_1, \ldots, y_T)$ be any outcome sequence. For any $(t_1, \ldots, t_k)$ such that $1 < t_i \leq T$ and $t_i \leq t_{i+1}$,

$$[y_0, y_{t_1}), [y_{t_1}, y_{t_2}), \ldots, [y_{t_k}, y_{T+1})$$

is called a k-partition of $\underline{y}$, denoted by $P_{k,e}(\underline{y})$, where

$$e = (e_0, \ldots, e_k),$$

such that $1 \leq e_i \leq n$ and $e_i \neq e_{i+1}$. Expert B_{e_i} is the best expert associated with the ith segment $[y_{t_i}, y_{t_{i+1}})$. Define the total loss of $P_{k,e}(\underline{y})$ to be

$$L^T_{P_{k,e}}(\underline{y}) = \sum_{i=1}^{k} L^{t_{i+1}-t_i}_{B_{e_i}}([y_{t_i}, y_{t_{i+1}})),$$

where $L^{t_{i+1}-t_i}_{B_{e_i}}([y_{t_i}, y_{t_{i+1}}))$ is defined in (4.2). Since

$$L^T_{P_{k,e}}(\underline{y}) \leq \min_{1 \leq j \leq N} L^T_{B_j}(\underline{y}),$$

$L_{P_{k,e}}(\underline{y})$ can be considered as a lower envelope.

Herbster and Warmuth [20] modified Vovk's *Generic* Algorithm, which is described in Section 4.2, by adding an additional update to obtain two algorithms: the $Fixed-share$ Algorithm and the $Variable-share$ Algorithm. Each algorithm has a parameter $\alpha \in [0, 1]$.

Herbster and Warmuth [20, Theorem 4.4] showed that for any positive integers $\hat{k}$ and $\hat{T}$, by setting $\alpha = \frac{\hat{k}}{\hat{T}}$, for any outcome sequence $\underline{y}$ with $T \leq \hat{T}$, and any $P_{k,e}(\underline{y})$ with $k \leq \hat{k}$, the $Fixed-share$ Algorithm satisfies

$$L^T_{Fixed}(\underline{y}) \leq c ln n + c\eta L^T_{P_{k,e}}(\underline{y}) + ck(ln\frac{\hat{T}}{\hat{k}} + ln(n-1)) + 2c\hat{k},$$

where c and η are determined by the loss function, and $L^T_{Fixed}(\underline{y})$ is defined in (4.2).

Herbster and Warmuth [20, Theorem 5.8] showed that for any positive integers $\hat{k}$ and $\hat{L}$, by setting $\alpha = \frac{\hat{k}}{2\hat{k}+\hat{L}}$, for any $P_{k,e}(\underline{y})$ with $k \leq \hat{k}$, $L^T_{P_{k,e}}(\underline{y}) \leq \hat{L}$, the $Variable-$

share Algorithm satisfies

$$L^T_{Variable}(\underline{y}) \leq clnn + c\eta L^T_{P_{k,e}}(\underline{y}) + O(k(ln\frac{\hat{L}}{\hat{k}}) + O(\hat{k}),$$

where c and η are constants determined by the loss function, and $L^T_{Variable}(\underline{y})$ is defined in (4.2).

For certain loss functions, such as square loss, c and η can be chosen such that $c\eta = 1$. Then using either $Fixed - share$ Algorithm or $Variable - share$ Algorithm, we have bounds for the difference between the average loss of the decision-maker and $L^T_{P_{k,e}}(\underline{y})$. As an example, Herbster and Warmuth [20, Section 6] considered a sequence of 800 trials with four distinct segments. They compared Vovk's *Generic* Algorithm to the Variable-share Algorithm. The simulation results showed that the Variable-share Algorithm performs better than Vovk's *Generic* Algorithm for this sequence.

Cover and Shenhar [11] introduced a prediction strategy whose average loss approaches to the k-th order Bayes envelope in the situation of sequential prediction of binary sequences with apparent Markov structure. In their paper, $\mathcal{D} = [0,1]$, $\Theta = \{0,1\}$, and they use a score function instead of a loss function. Without considering the Markov structure, Cover and Shenhar [Section 4][11] gave a random predictor that predicts $\hat{y}_t = 1$ with probability p_t at each time t, where p_t is constructed based on Blackwell [6]'s procedure.

For any outcome sequence $\underline{y}$, denote the expected average score by

$$E(S_T) = \frac{1}{T}\sum_{t=1}^{T}[p_t y_t + (1-p_t)(1-y_t)],$$

and the Bayes envelope by

$$\phi(S_T) = \max(\frac{1}{T}\sum_{t=1}^{T} y_t, 1 - \frac{1}{T}\sum_{t=1}^{T} y_t).$$

It follows from Hannan [17, p.139] that for any $\underline{y}$,

$$\phi(S_T) - E(S_T) \leq \frac{3}{\sqrt{T}}.$$

Considering the k-th Markov structure, Cover and Shenhar [11, Section 5] gave a k-th order Markov predictor. At time T, $T = k$, $k+1$, ..., let

$$z = (y_{T-k+1}, \ldots, y_T).$$

Denote $T'(z) = T(z,1) + T(z,0)$, where $T(z,1)$ and $T(z,0)$ are the number of times the sequence $z,1$ and $z,0$ were observed in $\underline{y}$. For each z, the k-th order Markov predictor uses the Blackwell procedure to make predictions. It follows that

$$\phi^k(S_T) - E(S_T) \le \frac{2^k 3}{\sqrt{T}} + \frac{k}{T},$$

where

$$\phi^k(S_T) = \frac{1}{T-k} \sum_{z\in[0,1]^k} T'(z)\phi(S_{T'(z)})$$

is the k-th order Bayes envelope.

In the compound decision problem, a lower envelope than the simple Bayes envelope is the extended envelope introduced by Johns [21]. The idea of the extended version is to take advantage of higher order empirical dependencies in the parameter sequence. Johns [21] proposed extended compound rules whose risks achieve these envelopes in the limit. Gilliland and Hannan [16] generalized and strengthened some results previously reported by Swain [25] and Johns [21]. Theorem 3 of Gilliland and Hannan [16] implies that the extended rules should compare favorably with unextended rules relative to the parameter sequence generated by a strictly stationary process.

Vardeman [26] treated a sequence version of the finite state compound decision problem. He also studied the k-extended problem in a game theoretic situation. After proving a simple game theoretic decomposition of k-extended envelope, he constructed randomized strategies with risk approximating the k-extended envelope at the rate of $O(T^{1/2})$.

Suppose that the loss sequence of the experts has some dependencies. For instance, in the case of two experts, assume that the loss sequence $\{(b_1^1, b_2^1)$, (b_1^2, b_2^2), (b_1^3, b_2^3), $(b_1^4, b_2^4), \ldots\}$ is $\{(0,1), (1,0), (0,1), (1,0), \ldots\}$. Taking advantage of the dependencies, we can use the idea of the k-extended problem in decision theory to construct a strategy, named k-extended strategy, such that the average prediction error of the k-extended strategy approximates the k-extended envelope, which is a lower envelope than the simple Bayes envelope.

4.5.2 The k-extended prediction strategies

Consider the prediction problem discussed in Section 4.4. Suppose at time t, $B_1^t, \ldots,$ B_n^t are the predictions of n experts, respectively. For any t and i, let outcome $y_t \in \Theta$, $B_i^t \in \mathcal{D}$. Suppose the loss function L is bounded. Denote

$$b_i^t = L(B_i^t, y_t), \ \forall\, i, t.$$

Without loss of generality, we assume that $0 \leq b_i^t \leq 1$ for all i and t. Let

$$Q = \{b_i^t \ \forall i, t\}$$

and suppose that Q is a finite set with cardinality q.

Assume at each time t, for all i, the decision-maker has the knowledge of $b_i^1, \ldots,$ b_i^{t-1}. Using Vardeman's technique, we define a k-extended randomized strategy C^k for the decision-maker, which is described in Figure 4.2. So at each time t, strategy C^k only uses the past stages, at which the k previous predicting errors are the same as $\mathbf{b}^{*t-1}$, to determine the prediction.

Theorem 4.5.1 *Let c_t^k be the loss of strategy C^k at each time t, b_i^t be the loss of the expert B_i at time t with*

$$0 \leq b_i^t \leq 1 \ \forall \ i, t.$$

Suppose $Q = \{b_i^t \ \forall i, t\}$ is a finite set with cardinality q. Then as $T \to \infty$,

$$\frac{1}{T}|\sum_{t=1}^{T} c_t^k - \sum_{y \in Q^{n(k-1)}} \min_{1 \leq i \leq n} (\sum_{t:\mathbf{b}^{*t-1}=y}^{T} b_i^t)| \xrightarrow{P} 0,$$

and the convergence rate is $O(T^{-1/2})$.

Proof. Denote

$$\tilde{\mathbf{X}} = | \{x^t = \{x_1^t, \ldots, x_{2^n}^t\}, \ x_i^t = \prod_{l=1}^{n} [b_l^t a_{il} + (1 - b_l^t)(1 - a_{il})], \ \forall i\}$$

Figure 4.2: Strategy C^k

Strategy C^k

Choose initial predictions $\hat{y}_t,\ \ t < k$, from the set of $\{B_i^1,\ \forall i\}$

Repeat for $t = k, k+1, \ldots$

predict with

$$\hat{y}_t = B_j^t \text{ if } \sum_{s:\mathbf{b}^{*s-1}=\mathbf{b}^{*t-1}} (b_j^s - b_i^s) \leq \sqrt{\frac{3n^2(T(\mathbf{b}^{*t-1})-1)}{2^n}} z(A^i - A^j) \forall i,$$

where

$$\mathbf{b}^{*t-1} = \{b^{t-k+1}, b^{t-k+2}, \ldots, b^{t-1}\},$$
$$b^{t-1} = \{b_1^{t-1}, \ldots, b_n^{t-1}\},$$
$$T(\mathbf{b}^{*t-1}) = \sum_{j=1}^{t-1} I_{\{\mathbf{b}^{*j-1}=\mathbf{b}^{*t-1}\}}.$$

and $z = (z_1, \ldots, z_{2^n})$, and z_1, …, z_{2^n} are i.i.d. $U(0,1)$ random variables, A^1, …, A^n are the columns of matrix A.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{2^n1} & \cdots & a_{2^nn} \end{pmatrix} = \begin{pmatrix} A_1 \\ \vdots \\ A_{2^n} \end{pmatrix}.$$

And A_1, …, A_{2^n} are the distinct sequences from W.

$$W = \{w^n \mid w^n = (w_1, \ldots, w_n),\ \ w_i \in \{0,1\}, \forall i\}.$$

Since the loss of experts take values in a finite set Q, which has cardinality q, the cardinality of $\tilde{\mathbf{X}}$ is q^n. There exits a one to one mapping between Q^n and $\tilde{\mathbf{X}}$.

Consider a finite two-person game in which player I has 2^n pure strategies and player II has n pure strategies. The loss matrix is the matrix A in Figure 4.2. Suppose player I only takes strategies from $\tilde{\mathbf{X}}$. Denote

$$\sigma_t = \sigma(X^k_{t-1}|\mathbf{x}^{*t-1} + \sqrt{\frac{3n^2(T(\mathbf{x}^{*t-1}) - 1)}{2^n}}z),$$

where

$$X^k_{t-1}|\mathbf{x}^{*t-1} = \sum_{j:\,\mathbf{x}^{*j-1}=\mathbf{x}^{*t-1}}^{t-1} x^j,$$

and

$$T(\mathbf{x}^{*t-1}) = \sum_{j=1}^{t-1} I_{\{\mathbf{x}^{*j-1}=\mathbf{x}^{*t-1}\}}.$$

Then the definition of the strategy C^k implies that

$$\begin{aligned}
\sum_1^T x^t\sigma_t &= \sum_{x\in\tilde{\mathbf{X}}^{k-1}} (\sum_{t:\mathbf{x}^{*t-1}=x} x^t\sigma_t) \\
&= \sum_{y\in Q^{n(k-1)}} (\sum_{t:\mathbf{b}^{*t-1}=y} c^k_t) \\
&= \sum_{t=1}^T c^k_t.
\end{aligned}$$

It follows from Lemma 2.2.6 that

$$\begin{aligned}
\phi^k(X^k_T) &= \sum_{x\in\tilde{\mathbf{X}}^{k-1}} \phi(X^k_T|x) \\
&= \sum_{y\in Q^{n(k-1)}} (\min_i(\sum_{t:\mathbf{b}^{*t-1}=y}^T b^t_i)).
\end{aligned}$$

Thus, applying Theorem 2.3.5 with $|B| = 1$, $m = 2^n$, and x^t in $\tilde{\mathbf{X}}$, we have

$$\begin{aligned}
& E_\mu|\sum_{t=1}^T c^k_t - \sum_{y\in Q^{n(k-1)}} (\min_i(\sum_{t:\mathbf{b}^{*t-1}=y}^T b_{i,t}))| \\
&= E_\mu|\sum_1^T x^t\sigma_t - \phi^k(X^k_T)| \\
&\leq \sqrt{3n^2 2^n q^{n(k-1)} T}.
\end{aligned}$$

Therefore, by Markov inequality, Theorem 4.4.1 follows.

□

Remark 4.5.1

$$\sum_{y \in Q^{n(k-1)}} (\min_i(\sum_{t:\mathbf{b}^{*t-1}=y}^{T} b_{i,t})) \leq \min_{1\leq j\leq n} \sum_{t=1}^{T} b_i^t.$$

Example 4.5.1 *Suppose there are two experts, and their loss sequence is as following:*

$$\{(b_1^1, b_2^1), (b_1^2, b_2^2), (b_1^3, b_2^3), (b_1^4, b_2^4), \ldots\} = \{(0,1), (1,0), (0,1), (1,0), \ldots\}$$

Then the simple envelope is

$$\phi(X^T) = \begin{cases} \frac{T-1}{2} & \text{if } T > 1 \text{ and } T \text{ is odd} \\ \frac{T}{2} & \text{if } T \text{ is even} \end{cases}.$$

For the simple case of $k = 2$, the k-extended envelope is

$$\phi^2(X_T^2) = 0.$$

Therefore, we see that the 2-extended envelope is indeed a lower envelope.

Let $T = 5000$. To compare the average loss of strategy C^2 with the average loss of the strategy C defined in Section 4.3, we used S-PLUS to determine:

$$\frac{1}{4998} \sum_{t=3}^{5000} c_t = 0.5196078,$$

$$\frac{1}{4998} \sum_{t=3}^{5000} c_t^2 = 0.$$

Thus, the 2-extended prediction strategy performs better for this sequence.

Bibliography

[1] G. Anandalingam and L. Chen. Linear combination of forecasts: a general Bayesain Model. *Journal of Forecasting*, V8, 199-214, 1989.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund and R. E. Schapire. Gambling in a rigged casino: The adversarial muti-armed bandit problem, In *Proceedings, 36th Annual Symposium on Foundations of Computer Science*, November, 1995.

[3] R. J. Ballard. Extended rules for the sequence compound decision problem with $m \times n$ component, Doctoral Thesis at Michigan State University, 1974.

[4] R. J. Ballard and D. Gilliland. On the risk performance of extended sequence compound rules for classification between $N(0,1)$ and $N(1,1)$, *J. Statist. Comput. Simul.*, V6, 265-280, 1978.

[5] R. J. Ballard, D. Gilliland and J. Hannan. $O(N^{-1/2})$ convergence to k-extended Bayes risk in the sequence compound decision problem with $m \times n$ component, RM-333, Statistics and Probability, MSU, 1974.

[6] D. Blackwell. Controlled random walks. *J. Proc. Int. Congr. Math.* 3, 336-338, 1956.

[7] N. Cesa-Bianch, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire and M. K. Warmuth. How to use expert advice. Unpublished manuscript, 1995.

[8] R. T. Clemen. Combining forecasts: a review and annotated bibliography. *Intl. J. Forecast* 5, 559-583, 1989.

[9] R. T. Clemen and J. B. Guerard, Jr. Econometric GNP forecasts: incremental information relative to nave extrapolation. *Intl. J. Forecast* 5, 417-426, 1989

[10] R. T. Clemen, and R. L. Winkler. Combining economic forecasts. *J. Bus. And Econ. Statist.* 4, 39-46, 1986.

[11] T. M. Cover and A. Shenhar. Compound Bayes predictors for sequences with apparent Markov Structure, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-7, No.6, 1977.

[12] D. P. Foster. Prediction in the worst case. *The annals of Statistics*, V19, No. 2, 1084-1090, 1991.

[13] D. P. Foster and R. Vohra. A randomization rule for selecting forecasts. *Operations Research*, 41(4):704-709, July-August 1993.

[14] Y. Freund. Predicting a binary sequence almost as well as the optimal biased coin. In *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, 1996.

[15] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting . In *Computational Learning Theory: Second European Conference, EuroCOLT '95*, 23-37, Springer-Verlag, 1995.

[16] D. Gilliland and J. Hannan. On the extended compound decision problem. *Ann Math. Statist.* 40: 1536-1541, 1969.

[17] J. Hannan. Approximation to Bayes risk in repeated play. In *Contribution to the theory of games*, V3, 97-139, Princeton University Press, 1957.

[18] D. Haussler, J. Kivinen, and M. K. Warmuth. Tight worst-case loss bounds for predicting with expert advice. In *Computational Learning Theory: Second European Conference, EuroCOLT '95*, 69-83, Springer-Verlag, 1995.

[19] D. P. Helmbold, R. E. Schapire, Y. Singer and M. K. Warmuth. On-line portfolio selection using multiplicative updates. Unpublished manuscript, 1996.

[20] M. Herbster and M. K. Warmuth. Tracking the best expert. In *Proceedings of the Twelfth International Conference on Machine Learning*, 286-294, 1995.

[21] M. V. Jr., Johns. Two-action compound decision problems. *Proc. Fifth Berkeley Symp. Math. Statist. Prob*, 1 pages 463-478, University of California Press, 1967.

[22] T. L. Lai and H. Robbins. Asymptotically Efficient Adaptive Allocation Rules. *Advances in Applied Mathematics*, 6:4-22, 1985.

[23] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108:212-261, 1994.

[24] H. Robbins. Some Aspects of the Sequential Design of Experiments. *Bulletin American Mathematical Society*, 55:527-535, 1952.

[25] D. D. Swain. Bounds and rates of convergence for the extended compound estimation problem in the sequence case, *Tech. Report*, No. 81, Department of Statistics, Stanford University, 1965.

[26] S. B. Vardeman. Approximation to minimum k-extended Bayes risk in sequences of finite state decision problems and games, *Bulletin of the Institute of Mathematics Academia Sinica*, V10, No. 1, 1982.

[27] V. G. Vovk. Aggregating strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory*, 371-383, 1990.

[28] V. G. Vovk. A game of prediction with expert advice. Unpublished manuscript, 1996.