Universal Portfolios with Side Information

Thomas M. Cover Erik Ordentlich *

Abstract

We present a sequential investment algorithm, the μ-weighted universal portfolio with side-information, which achieves, to first order in the exponent, the same wealth as the best side-information dependent investment strategy (the best state-constant rebalanced portfolio) determined in hindsight from observed market and side-information outcomes.

This is an individual sequence result which shows that the difference between the exponential growth rates of wealth of the best state-constant rebalanced portfolio and the universal portfolio with side-information is uniformly less than $(d/(2n))\log(n+1)+(k/n)\log 2$ for every stock market and side-information sequence and for all time n. Here $d=k(m-1)$ is the number of degrees of freedom in the state-constant rebalanced portfolio with k states of side-information and m stocks. The proof of this result establishes a close connection between universal investment and universal data compression.

Keywords: Universal investment, universal data compression, portfolio theory, side information.

*This work was partially supported by NSF grant no. NCR-9205663 and JSEP contract no. DAAH04-94-G-0058. This paper was presented in part at IEEE International Symposium on Information Theory, Trondheim, Norway, 1994. Thomas M. Cover is with the Depts. of Statistics and Electrical Engineering, Stanford University. Erik Ordentlich is with the Dept. of Electrical Engineering, Stanford University. Erik Ordentlich was partially supported by an NSF Graduate Fellowship.

1 Introduction

We consider the problem of universal sequential investment in a market of m stocks with side-information. The behavior of the market is specified by an arbitrary sequence of non-negative price-relative (stock) vectors, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in R_+^m$ and an associated sequence of side-information states $y_1, y_2, \ldots y_n$ taking on values in a finite set $\mathcal{Y} = \{1, \ldots, k\}$. The j^{th} entry x_{ij} of the i^{th} price relative vector $\mathbf{x}_i$ denotes the ratio of closing to opening price of the j^{th} stock for the i^{th} trading day. The side-information sequence may depend in an arbitrary manner upon the entire stock market sequence. An investment at time i in this market is specified by a portfolio vector $\mathbf{b}_i \in R^m$ with non-negative entries summing to one. The components of $\mathbf{b}_i$ are the proportions of current wealth invested in each stock at time i.

The conventional treatment of the problem of adaptive investment is grounded in the distributional approach to investment pioneered by Kelly (Kelly gambling [1]) and many others. This approach assumes the existence of an underlying probability distribution governing the sequence of price relatives (returns on gambles in the Kelly problem). Given knowledge of this underlying distribution, it is possible to specify a sequence of investment decisions or portfolio choices which achieves a growth rate of wealth that is maximal in a probabilistic sense [2, 3, 4]. The universal investment problem in this conventional setting arises by assuming that the investor has limited knowledge of the true distribution underlying the market. It is known only that this distribution belongs to a certain set $\mathcal{P}$ of possible distributions. The goal then is to exhibit a sequential investment algorithm which achieves, with probability one, the optimal growth rate of wealth for all distributions in $\mathcal{P}$. This has been carried out by Algoet [5], with $\mathcal{P}$ being the collection of all stationary ergodic distributions on infinite sequences of price relative vectors. The algorithms described in [5] derive the next portfolio choice from estimates of the underlying conditional distribution of the next price relative given the observed past.

In this paper, we depart from the conventional approach and make *no* distributional assumptions on the sequence of price relatives and side-information states. Instead we establish a set of allowable investment actions (sequences of portfolio choices $\mathbf{b}_i$), and seek to achieve the same asymptotic growth rate of wealth as the best action in this set, not in any kind of stochastic sense, but uniformly over *all* possible sequences of price relatives and side-information states. Thus we seek an individual sequence minimax regret solution.

The set of allowable investment actions or algorithms we consider is comprised of the state-constant rebalanced portfolios. At each time $i \in \{1, \ldots, n\}$, a state-constant rebalanced portfolio investment algorithm invests in this market using one of k distinct portfolios $\mathbf{b}(1), \ldots, \mathbf{b}(k)$ depending on the current state of side-information y_i. We refer to the case of $k = 1$ as a constant rebalanced portfolio.

Our main result is summarized as follows. Let $S_n^*(\mathbf{x}^n|y^n) = e^{nW_n^*(\mathbf{x}^n|y^n)}$ be the wealth achieved in hindsight by the best state-constant rebalanced portfolio for the sequence of price relatives $\mathbf{x}^n$ and side-information states y^n. We exhibit a sequential investment algorithm, the μ-

weighted universal portfolio with side-information, which achieves a wealth $\hat{S}_n(\mathbf{x}^n|y^n) = e^{n\hat{W}_n(\mathbf{x}^n|y^n)}$ that tracks $S^*(\mathbf{x}^n|y^n)$ to first order in the exponent. Specifically it is shown that $\hat{W}_n(\mathbf{x}^n|y^n) \geq W_n^*(\mathbf{x}^n|y^n) - (d/(2n))\log(n+1) - (k/n)\log 2$, uniformly for every stock vector sequence $\mathbf{x}^n$ and side-information sequence y^n. Here $d = k(m-1)$, m is the number of stocks, and k is the cardinality of the side-information state space. The quantities $\hat{W}_n(\mathbf{x}^n|y^n)$ and $W_n^*(\mathbf{x}^n|y^n)$ are the exponential growth rates of wealth for each portfolio. Since d is exactly the number of degrees of freedom in a state-constant rebalanced portfolio algorithm we see that the cost of achieving universal performance is essentially $(1/(2n))\log n$ in the exponential growth rate of wealth per degree of freedom. This performance bound is very similar to that arising in universal data compression and source modeling. As shown later, this similarity is no accident, and is attributable to a close connection between the present universal sequential investment problem and the universal data compression of i.i.d. sources.

It will be seen, in fact, that the apparently more complicated investment problem, involving as it does investment actions and price relative vectors with more than one non-zero component, is dominated by the simple Kelly gambling or horse race problem and its well studied data compression counterpart. The horse race market turns out to be the worst case in the sense that, for this market, the bounds on the worst case performance of the universal portfolio relative to the best constant rebalanced portfolio are achieved essentially with equality. In this sense, uniformly good performance is hardest to achieve for the simplest market.

Our motivation for focusing on the collection of state-constant rebalanced portfolios arises from the distributional approach to investment discussed above. It is well known that if the price relatives are independent and identically distributed, the optimal growth rate of wealth is achieved by a constant rebalanced portfolio [2, 3, 6]. Thus the constant rebalanced portfolios possess certain optimality properties in the conventional distributional setting. The state-constant rebalanced portfolios are obtained as natural extensions of the constant rebalanced portfolios to the side-information setting. It is important to realize, however, that aside from this motivational link, the present problem and solution possess no distributional or random aspect.

Another source of motivation for considering the state-constant rebalanced portfolios is the sequential compound Bayes decision problem of Robbins, Hannan, and others [7, 8, 9]. This problem involves a sequence of repeated plays of a game against nature. The goal is to exhibit a sequential player strategy which approximates the performance of the best constant player strategy determined in hindsight for any sequence of moves by nature. Our problem fits into this framework if we identify the player's moves as portfolio choices $\mathbf{b}$, nature's moves as price relative vectors $\mathbf{x}$, and $-\log \mathbf{b}^t\mathbf{x}$ as a loss function. Allowing the player's moves to vary with side-information is a natural generalization of this basic setup.

The difficulty is that the classical approach to the compound Bayes decision problem does not readily apply to universal investment. As discussed by Merhav and Feder [10], the classical solution makes certain assumptions about the loss function and the domain of the game which are not valid in the investment problem. The principal obstacle is the unboundedness of the loss function $-\log(\mathbf{b}^t\mathbf{x})$ as $\mathbf{b}^t\mathbf{x}$ tends to zero. Our method, which is a departure from the classical approach to the general Bayes decision problem, takes advantage of the special structure of the investment problem to overcome this difficulty.

The side-information aspect of the present universal investment problem is new, but several prior works have considered the problem with no side-information. Cover and Gluss [13] show that the approachability-excludability theorem of Blackwell [11, 12] can be used to define an investment scheme with universal properties if the price relatives are restricted to a finite set. Larson [14] shows that variants of the investment scheme suggested by the compound Bayes technique have exponential growth rates arbitrarily close (but not equal) to that of the best constant rebalanced portfolio. Unlike the present setting, the price relatives are assumed to take on values in a finite set. The μ-weighted universal portfolio with side-information considered in this paper is a generalization of the universal portfolio proposed originally by Cover [15].

Cover [15] bounds the ratio of wealths achieved by the best constant rebalanced portfolio (no side-information) and the universal portfolio in terms of a sensitivity matrix depending on the texture of the stock sequence. The analysis in [15] uses Laplace's method of integration and assumes that the price relatives are bounded away from zero and bounded from above. This is in contrast to the present paper which obtains individual sequence, worst case bounds that are independent of the stock vectors $\mathbf{x}^n$ and side-information states y^n. No assumptions on the stock vectors are required.

The paper is organized as follows. Section 2 formally establishes the investment setup and notation, and defines the state-constant rebalanced portfolios. The μ-weighted universal portfolio with side-information is defined in Section 3. Section 4 contains the main theorems establishing the performance bounds on the proposed universal portfolio strategy, and Section 5 presents a simple example illustrating these results. Finally, Section 6 provides an efficient method for the exact computation of the universal portfolio.

2 Preliminaries

As stated above we will be concerned with investment opportunities in a market consisting of m stocks. A vector of price relatives $\mathbf{x} = (x_1, x_2, \ldots, x_m)^t$ (with $x_j \geq 0$, $j = 1, \ldots, m$ or $\mathbf{x} \in \mathbb{R}_+^m$) expresses the change in the prices of these stocks over one investment period. The j^{th} entry x_j of $\mathbf{x}$ is simply the ratio of the final to the initial price of the j^{th} stock for the given trading period. Thus an investment in stock j increases by a factor of x_j over that period. We refer to the vectors of price relatives as stock vectors.

An investment in the market of m stocks is specified at the beginning of each trading period by a portfolio vector $\mathbf{b} = (b_1, b_2, \ldots, b_m)^t \in \mathcal{B}$, where

$$\mathcal{B} = \{\mathbf{b} \in \mathbb{R}^m : \sum_{j=1}^{m} b_j = 1, b_j \geq 0\}. \tag{1}$$

The j^{th} entry b_j of a portfolio $\mathbf{b}$ is the proportion of wealth invested in the j^{th} stock. An investment using a portfolio $\mathbf{b}$ increases one's wealth by a factor of

$$\mathbf{b}^t\mathbf{x} = \sum_{j=1}^{m} b_j x_j, \tag{2}$$

if the market performance is specified by the stock vector $\mathbf{x}$. Further, for a sequence of n investment periods, investing according to portfolios $\mathbf{b}_i$ for periods $i = 1, \ldots, n$ increases the initial wealth by a factor of

$$\begin{aligned} S_n &= \prod_{i=1}^{n} \mathbf{b}_i^t \mathbf{x}_i \\ &= \prod_{i=1}^{n} \sum_{j=1}^{m} b_{ij} x_{ij}. \end{aligned} \tag{3}$$

if the market performance is

$$\mathbf{x}^n = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n). \tag{4}$$

A sequence of portfolio choices $\mathbf{b}_i$ constitutes an investment strategy or algorithm.

To clarify, note that the well known Kelly gambling problem [1] can be expressed as a special case of the general investment setup as shown in [6, chapters 6 and 15]. In Kelly gambling one must place bets on the outcome of an m-valued event, such as the winner of a race with m horses. If the event takes on the value j (i.e., horse j wins) the gambler receives $b_j o_j$, where b_j is the proportion of wealth bet on horse j and o_j is the payoff for the j^{th} horse. In the market terminology above, Kelly gambling corresponds to a market with m securities in which the price relatives are restricted to be one of m vectors where the j^{th} vector is o_j in the j^{th} component and zero in all others. The portfolio $\mathbf{b} = (b_1, b_2, \ldots, b_m)$ in this case denotes the fraction of wealth bet on each of the possible outcomes of the event.

2.1 Constant rebalanced portfolios

As stated in the introduction we restrict the set of actions or portfolio choices to the state-constant rebalanced portfolios. These are derived from the constant rebalanced portfolios, which we now describe. A constant rebalanced portfolio strategy uses the same portfolio $\mathbf{b}$ for each trading period. For a sequence of stock vectors $\mathbf{x}^n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ the constant rebalanced portfolio strategy using portfolio $\mathbf{b}$ achieves wealth $S_n(\mathbf{b}, \mathbf{x}^n)$ (which we sometimes abbreviate as $S_n(\mathbf{b})$) given by

$$S_n(\mathbf{b}, \mathbf{x}^n) = S_n(\mathbf{b}) = \prod_{i=1}^{n} \mathbf{b}^t \mathbf{x}_i. \tag{5}$$

Note that $S_n(\mathbf{b}, \mathbf{x}^n)$ depends on the sequence $\mathbf{x}^n$ only up to permutation. Also note that a constant rebalanced portfolio strategy actually involves a great deal of trading. This is because at the end of a particular trading period, say the i^{th}, the proportion of wealth invested in each stock has changed from $b_1, \ldots, b_m$ to $x_{i1}b_1/(\mathbf{b}^t\mathbf{x}_i), \ldots, x_{im}b_m/(\mathbf{b}^t\mathbf{x}_i)$ and therefore stocks must be bought and sold to restore the proportions of wealth to $b_1, \ldots, b_m$ for the next trading period.

For a sequence of stock vectors $\mathbf{x}^n$ we can determine the best constant rebalanced portfolio as the one achieving the maximum wealth. We denote this portfolio by $\mathbf{b}^*(\mathbf{x}^n)$ or just $\mathbf{b}^*$ and it is given by

$$\mathbf{b}^*(\mathbf{x}^n) = \mathbf{b}^* = \arg\max_{\mathbf{b}\in\mathcal{B}} S_n(\mathbf{b}). \tag{6}$$

We use $S_n^*(\mathbf{x}^n)$, or simply S_n^*, to denote the maximum,

$$S_n^*(\mathbf{x}^n) = S_n^* = \max_{\mathbf{b}\in\mathcal{B}} S_n(\mathbf{b}). \tag{7}$$

Thus the best constant rebalanced portfolio strategy uses portfolio $\mathbf{b}^*$ and achieves a wealth of S_n^*. We will also refer to the quantity

$$W_n^*(\mathbf{x}^n) = \frac{1}{n}\log S_n^*(\mathbf{x}^n) \tag{8}$$

as the exponential growth rate of wealth for the best constant rebalanced portfolio.

It is important to note that both S_n^* and $\mathbf{b}^*$ depend on the entire sequence $\mathbf{x}^n$ and n since they result from a maximization for that particular sequence. Thus, even if the set of investment actions is limited to the constant rebalanced portfolios, knowledge of the entire sequence $\mathbf{x}^n$ and n is required to determine the best action in this set.

2.2 Side-information

Investors use various sources of side-information to adjust and update their portfolios. We model this side-information as a finite valued variable y made available at the start of each investment period. The portfolio choice can then incorporate knowledge of y for that period. Thus, the formal domain of our market model is a sequence of pairs $\{(\mathbf{x}_i, y_i)\}$ where, as defined above, $\mathbf{x}_i$ is the stock vector for period i and $y_i \in \mathcal{Y} = \{1, 2, \ldots, k\}$ denotes the state of the side-information at time i.

The side-information can arise in numerous ways. For example, sophisticated trading strategies often develop signaling algorithms that indicate the nature of the investment opportunity about to be faced. The signal would constitute the side-information. An example of a signaling algorithm is: set $y_i = j$ if stock j has outperformed other stocks in the previous r trading days. Another signaling algorithm might use y_i to reflect whether the moving average of the last r trading days is greater or less than the average of the price relatives on the previous trading day. These examples involve side-information that is a causal function of past market performance.

In our model, however, y_i could depend on the entire sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and could, for example, identify the best stock on trading day i. This is a maximally informative side-information sequence, since it would allow the investor to invest in the best stock each day, resulting in astronomical profits. The challenge is that the significance or value of such side-information is not apparent to an investor with only sequential knowledge of the stock vector and side-information sequences. This must be "learned". However, since the sequences are arbitrary, there is nothing that can be learned in the usual sense about the continuation of the sequence.

2.3 State-constant rebalanced portfolios

The constant rebalanced portfolio discussed above is extended to the state-constant rebalanced portfolio by allowing the portfolio decisions to vary with the side-information y. A state-constant rebalanced portfolio specifies portfolios $\mathbf{b}(1), \mathbf{b}(2), \ldots, \mathbf{b}(k) \in \mathcal{B}$, and uses portfolio $\mathbf{b}(y_i)$ at time i when the side-information state takes on value $y_i \in \mathcal{Y} = \{1, 2, \ldots, k\}$. The choice of $\mathbf{b} : \mathcal{Y} \to \mathcal{B}$, results in wealth

$$S_n(\mathbf{b}(\cdot), \mathbf{x}^n | y^n) = \prod_{i=1}^{n} \mathbf{b}^t(y_i)\mathbf{x}_i \tag{9}$$

on the stock sequence $\mathbf{x}^n$ and side-information y^n. The collection of state-constant rebalanced portfolios with k states will be denoted by $\mathcal{B}^k$ (the k-wise Cartesian product of the portfolio simplex $\mathcal{B}$).

As above, for a sequence of stock vectors $\mathbf{x}^n$ and side-information states y^n we can determine the best state-constant rebalanced portfolio as the one achieving the maximum wealth. We denote this portfolio by $\mathbf{b}^*(\cdot)$, where

$$\mathbf{b}^*(\cdot) = \arg \max_{\mathbf{b}(\cdot) \in \mathcal{B}^k} S_n(\mathbf{b}(\cdot), \mathbf{x}^n | y^n), \tag{10}$$

and the maximum is over all portfolio assignments $\mathbf{b} : \mathcal{Y} \to \mathcal{B}$. Let

$$S_n^*(\mathbf{x}^n | y^n) = \max_{\mathbf{b}(\cdot) \in \mathcal{B}^k} S_n(\mathbf{b}(\cdot), \mathbf{x}^n | y^n) \tag{11}$$

denote the maximum wealth. Thus the best state-constant rebalanced portfolio strategy uses portfolio $\mathbf{b}^*(\cdot)$ and achieves a wealth of $S_n^*(\mathbf{x}^n | y^n)$. The exponential growth rate of wealth achieved by the best state-constant rebalanced portfolio at time n is

$$W_n^*(\mathbf{x}^n | y^n) = \frac{1}{n} \log S_n^*(\mathbf{x}^n | y^n). \tag{12}$$

Again, $S_n^*(\mathbf{x}^n | y^n)$ and $\mathbf{b}^*(\cdot)$ depend on the entire market sequence $\mathbf{x}^n$ and side-information sequence y^n, but only up to permutation.

The number of degrees of freedom (dimensions) in a state-constant rebalanced portfolio will be useful in characterizing the subsequent results. A state-constant rebalanced portfolio for

k states and m stocks has $k(m-1)$ degrees of freedom; $m-1$ degrees of freedom for each of the k portfolios which must be specified. The requirement that the entries sum to one gives each portfolio vector $m-1$ degrees of freedom, rather than m, where m is the number of stocks.

2.4 Sequential investment

A sequential portfolio with side-information is one which chooses $\mathbf{b}_i$, the portfolio to use at time i, based only on past (prior to time i) stock vectors and past and current (up through time i) side-information. Thus the portfolio used at time i is given by a function of the form

$$\begin{aligned} \mathbf{b}_i &= \mathbf{b}_i(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}, y_1, y_2, \ldots, y_i) \\ &= \mathbf{b}_i(\mathbf{x}^{i-1}, y^i), \end{aligned} \tag{13}$$

where $\mathbf{x}^{i-1}$ gives the past stock performance and y^i is the past and current side-information. A sequence of such investments achieves a wealth at time n of

$$S_n(\mathbf{x}^n|y^n) = \prod_{i=1}^{n} \mathbf{b}_i^t(\mathbf{x}^{i-1}, y^i)\mathbf{x}_i \tag{14}$$

and an exponential growth rate of

$$W_n(\mathbf{x}^n|y^n) = \frac{1}{n}\log S_n(\mathbf{x}^n|y^n). \tag{15}$$

We say a sequential portfolio with side-information is universal for the collection $\mathcal{B}^k$ of state-constant rebalanced portfolios if,

$$\begin{aligned} \lim_{n\to\infty} \sup_{\mathbf{x}^n, y^n} \frac{1}{n}\log\frac{S_n^*(\mathbf{x}^n|y^n)}{S_n(\mathbf{x}^n|y^n)} &= \lim_{n\to\infty} \sup_{\mathbf{x}^n, y^n} \left(W_n^*(\mathbf{x}^n|y^n) - W_n(\mathbf{x}^n|y^n)\right) \\ &= 0. \end{aligned} \tag{16}$$

In other words, a sequential investment algorithm with side-information is universal for the collection of state-constant rebalanced portfolios if it has the same asymptotic exponential growth rate of wealth as the best portfolio in $\mathcal{B}^k$ for any stock vector sequence $\mathbf{x}^n$ and any side-information sequence y^n as $n\to\infty$.

2.5 Problem statement

The problem, then, is to exhibit a sequential investment algorithm which is provably universal for the collection of state-constant rebalanced portfolios. Instead of making distributional assumptions on the stock vectors and seeking to optimize performance with respect to the true underlying distribution, we restrict the set of investment actions to the state-constant rebalanced portfolios and seek to perform as well as the best action in this set uniformly for all stock vector and side-information sequences.

3 Universal portfolios

It may seem unlikely that we could find a sequential portfolio which would be effective against arbitrary sequences, because one might ask what can be learned from the past of a sequence when the remainder of the sequence can be completely arbitrary? Isn't it possible that a sequence can be designed to fool any sequential portfolio into investing in precisely those stocks that will do the worst at each time? Of course this can be done, but the interesting fact is that, with proper care in choosing the sequential portfolio $\mathbf{b}_k : (\mathbb{R}_+^m)^{k-1} \times \mathcal{Y}^k \to \mathcal{B}$, one can still achieve $S_n^*(\mathbf{x}^n|y^n)$ to first order in the exponent, uniformly in $\mathbf{x}^n$ and y^n.

Toward this end we define the μ-weighted universal portfolio $\hat{\mathbf{b}}_i$ with side-information as follows. First, as in the preliminary section, we treat the case of no side-information or $|\mathcal{Y}| = 1$. In this case the μ-weighted universal portfolio with side-information will be referred to simply as the the μ-weighted universal portfolio.

Definition 1 (Universal portfolio) *The μ-weighted universal portfolio at time i is specified by*

$$\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_i(\mathbf{x}^{i-1}) = \frac{\int_{\mathcal{B}} \mathbf{b} S_{i-1}(\mathbf{b}, \mathbf{x}^{i-1}) d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_{i-1}(\mathbf{b}, \mathbf{x}^{i-1}) d\mu(\mathbf{b})}, \quad i = 1, 2, \ldots \tag{17}$$

with

$$\int_{\mathcal{B}} d\mu(\mathbf{b}) = 1, \tag{18}$$

and where, as defined above,

$$S_i(\mathbf{b}, \mathbf{x}^i) = S_i(\mathbf{b}) = \prod_{j=1}^{i} \mathbf{b}^t \mathbf{x}_j, \quad \textit{and} \quad S_0(\mathbf{b}, \mathbf{x}^0) = 1. \tag{19}$$

Note that if μ is symmetric, then

$$\hat{\mathbf{b}}_1 = (\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m})^t. \tag{20}$$

The overall portfolio algorithm is specified by the choice of μ, while the portfolio action at each time depends on μ and the sequence $\mathbf{x}^{i-1}$ observed so far. Later we focus on two possibilities for the measure μ; the uniform (Dirichlet$(1, \ldots, 1)$) and the Dirichlet$(1/2, \ldots, 1/2)$ distributions on the portfolio simplex $\mathcal{B}$. The Dirichlet$(1/2, 1/2, \ldots, 1/2)$ distribution has a density with respect to Lebesgue (uniform) measure on the simplex $\mathcal{B}$ given by

$$d\mu(\mathbf{b}) = \frac{\Gamma(\frac{m}{2})}{[\Gamma(\frac{1}{2})]^m} \prod_{j=1}^{m} b_j^{-\frac{1}{2}} d\mathbf{b}, \tag{21}$$

where $\Gamma(\cdot)$ denotes the Gamma function. Section 6 illustrates an efficient method for the exact computation of the universal portfolio as specified by (17) for μ equal to the Dirichlet$(1/2, \ldots, 1/2)$ distribution.

The μ-weighted universal portfolio achieves a wealth of

$$\hat{S}_n(\mathbf{x}^n) = \prod_{i=1}^{n} \hat{\mathbf{b}}_i^t(\mathbf{x}^{i-1})\mathbf{x}_i. \tag{22}$$

It is convenient that $\hat{S}_n(\mathbf{x}^n)$ can be calculated directly (rather than by calculating each $\hat{\mathbf{b}}_i$ and the resulting product $\prod_{i=1}^{n} \hat{\mathbf{b}}_i^t \mathbf{x}_i$) by using the telescoping of the product

$$\begin{aligned} \hat{S}_n(\mathbf{x}^n) &= \prod_{i=1}^{n} \frac{(\int \mathbf{b}^t S_{i-1}(\mathbf{b}) d\mu(\mathbf{b}))\mathbf{x}_i}{\int S_{i-1}(\mathbf{b}) d\mu(\mathbf{b})} \\ &= \prod_{i=1}^{n} \frac{\int S_i(\mathbf{b}) d\mu(\mathbf{b})}{\int S_{i-1}(\mathbf{b}) d\mu(\mathbf{b})} \end{aligned} \tag{23}$$

to obtain

$$\hat{S}_n(\mathbf{x}^n) = \int_{\mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n) d\mu(\mathbf{b}). \tag{24}$$

For convenience, we will sometimes refer to the μ-weighted universal portfolio as simply the universal portfolio, with the role of μ implied.

The μ-weighted universal portfolio with side-information for $|\mathcal{Y}| = k > 1$ is defined similarly, by using a fresh universal portfolio on each subsequence of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ corresponding to the times at which the side-information y_i takes on a given value in $\mathcal{Y}$.

Definition 2 (Universal portfolio with side-information) *The μ-weighted universal portfolio with side-information is specified by*

$$\hat{\mathbf{b}}_i(y) = \frac{\int_{\mathcal{B}} \mathbf{b} S_{i-1}(\mathbf{b}|y) d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_{i-1}(\mathbf{b}|y) d\mu(\mathbf{b})}, \quad i = 1, 2, \ldots \quad y \in \mathcal{Y}, \tag{25}$$

where

$$\int_{\mathcal{B}} d\mu(\mathbf{b}) = 1, \tag{26}$$

and $S_i(\mathbf{b}|y)$ is the wealth obtained by the constant rebalanced portfolio $\mathbf{b}$ along the subsequence $\{j \le i : y_j = y\}$, and is given by

$$S_i(\mathbf{b}|y) = \prod_{j \le i : y_j = y} \mathbf{b}^t \mathbf{x}_j, \quad \textit{with} \quad S_0(\mathbf{b}|y) = 1. \tag{27}$$

For stock vector and side-information sequences $\mathbf{x}^n$ and y^n the resulting wealth is

$$\hat{S}_n(\mathbf{x}^n|y^n) = \prod_{i=1}^{n} \hat{\mathbf{b}}_i^t(y_i)\mathbf{x}_i. \tag{28}$$

Again taking advantage of telescoping products along each subsequence $\{i : y_i = y\}$, the wealth can be expressed more compactly as

$$\hat{S}_n(\mathbf{x}^n|y^n) = \prod_{y=1}^{k} \int_{\mathcal{B}} S_n(\mathbf{b}|y) d\mu(\mathbf{b}). \tag{29}$$

As above, the corresponding exponential growth rate of wealth is

$$\hat{W}_n(\mathbf{x}^n|y^n) = \frac{1}{n}\log \hat{S}_n(\mathbf{x}^n|y^n). \tag{30}$$

Our goal now is to show that there exist μ for which the μ-weighted universal portfolio with side-information, $\hat{\mathbf{b}}(\cdot)$, is universal for the state-constant rebalanced portfolios $\mathcal{B}^k$ in the sense stated above; namely that,

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n,y^n} \frac{1}{n}\log \frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} = 0. \tag{31}$$

We accomplish this for μ equal to the uniform and the Dirichlet$(1/2,\ldots,1/2)$ distributions by proving the following results. For μ equal to the uniform distribution we show that

$$\begin{aligned}\sup_{\mathbf{x}^n,y^n} \frac{1}{n}\log \frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} &= \sup_{\mathbf{x}^n,y^n}\left(W_n^*(\mathbf{x}^n|y^n) - \hat{W}_n(\mathbf{x}^n|y^n)\right) \\ &\leq \frac{k(m-1)}{n}\log(n+1),\end{aligned} \tag{32}$$

and for the Dirichlet$(1/2,\ldots,1/2)$ distribution that

$$\begin{aligned}\sup_{\mathbf{x}^n,y^n} \frac{1}{n}\log \frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} &= \sup_{\mathbf{x}^n,y^n}\left(W_n^*(\mathbf{x}^n|y^n) - \hat{W}_n(\mathbf{x}^n|y^n)\right) \\ &\leq \frac{k(m-1)}{2n}\log(n+1) + \frac{k}{n}\log 2.\end{aligned} \tag{33}$$

Both bounds tend to 0 as $n \to \infty$, thereby demonstrating universality for these μ's. Furthermore, these bounds are shown to be essentially tight, indicating that the Dirichlet$(1/2,\ldots,1/2)$ weighting gives a somewhat better worst case performance than the uniform weighting. These results are proved in Section 4.

Combining expert opinion

Suppose there are m' experts, each with a portfolio selection scheme. Let $\mathbf{b}_i^r$ denote the portfolio recommendation of the r^{th} expert on day i, where $i = 1, 2, \ldots, n$, and $r = 1, 2, \ldots, m'$. Is there a sequential investment scheme which incorporates the experts' recommendations and achieves the same exponential growth rate of wealth as the best expert uniformly for all market sequences? This is easily done. Simply allocate wealth $1/m'$ per expert and invest each fraction of wealth according to each expert's sequence of portfolio selections. If $S_n^{(r)}$ is the wealth factor achieved by the r^{th} expert's strategy, the wealth accrued by this simple scheme will be $\hat{S}_n = (1/m')\sum_{r=1}^{m'} S_n^{(r)}$ which, of course, is at least $(1/m')\max_r S_n^{(r)}$ for every sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$. Since this is equal to $\max_r S_n^{(r)}$ to first order in the exponent, this simple scheme is universal for the collection of r expert strategies. The overall portfolio $\hat{\mathbf{b}}_i$ used by this investment strategy at time i is easily seen to be

$$\hat{\mathbf{b}}_i = \frac{\sum_{r=1}^{m'} S_{i-1}^{(r)}\mathbf{b}_i^r}{\sum_{r=1}^{m'} S_{i-1}^{(r)}}. \tag{34}$$

Thus the portfolio choice at time i is a performance weighted average of the portfolio choices of the m' experts. This is very similar to the universal portfolio definition (17), where the portfolio choice at each time is a performance weighted (continuous) average of the constant rebalanced portfolios. In fact the goal of universally tracking the best state-constant rebalanced portfolio can be thought of as performing as well as the best in a continuum of expert investors indexed by the portfolios in the simplex.

It is interesting to note that for the problem of tracking a finite number of investment strategies or experts, a somewhat more aggressive use of expert opinion can be obtained at the cost of an increase in the number of degrees of freedom. The market vector at time i is $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})$. Form the augmented vector $\tilde{\mathbf{x}}_i = (\mathbf{x}_i, \mathbf{b}_i^{(1)t}\mathbf{x}_i, \ldots, \mathbf{b}_i^{(m')}\mathbf{x}_i)$. This vector has $m + m'$ components. Now apply the universal portfolio to the sequence $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_n$ of augmented vectors. As claimed above, this portfolio will perform as well as the best constant rebalanced portfolio in the primary stocks x_1 through x_m and in the m' "stocks" $\mathbf{b}^{(r)t}\mathbf{x}$, $r = 1, 2, \ldots, m'$, associated with the investors. This is like investing in m stocks and m' mutual funds. Since the best constant rebalanced portfolio usually strictly out performs any of the constituent stocks, this portfolio exponentially out performs the experts as well as the constituent stocks. That is, expert opinion is valuable not only because one can ride along with the expert, but because one can rebalance among experts and the constituent stocks to take advantage of expert opinion and out perform all.

4 Main theorems : Uniform bounds

We now prove the performance bounds (equations (32) and (33)) on the μ-weighted universal portfolio with side-information for μ equal to the uniform and Dirichlet$(1/2, \ldots, 1/2)$ distributions. These results are contained in the following theorems. The first two theorems focus on the no side-information case or $k = 1$ case and the third generalizes the bounds to $k > 1$. Recall that $\hat{S}_n(\mathbf{x}^n)$ is the wealth achieved by the universal portfolio on the sequence $\mathbf{x}^n$, and $S_n^*(\mathbf{x}^n)$ is the wealth achieved by the best constant rebalanced portfolio. Also recall $\hat{S}_n(\mathbf{x}^n|y^n)$ and $S_n^*(\mathbf{x}^n|y^n)$ as the analogous quantities with side-information. Throughout, m is the number of stocks.

Theorem 1 *For μ equal to the uniform distribution,*

$$\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} \leq \binom{n+m-1}{m-1} \leq (n+1)^{m-1}, \tag{35}$$

for all n and for all stock price relative sequences $\mathbf{x}^n$.

Theorem 2 *For μ equal to the Dirichlet$(1/2, \ldots, 1/2)$,*

$$\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} \leq \frac{\Gamma(\frac{1}{2})\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\Gamma(n+\frac{1}{2})} \leq 2(n+1)^{\frac{m-1}{2}}, \tag{36}$$

for all n and for all stock price relative sequences $\mathbf{x}^n$.

The theorems say that for any $\mathbf{x}^n$ the wealth acquired by the uniform weighted and Dirichlet$(1/2,\ldots,1/2)$ weighted universal portfolios is within a polynomial factor of the wealth $S_n^*(\mathbf{x}^n)$ acquired by the best scheme in $\mathcal{B}$ (the best constant rebalanced portfolio).

The next theorem extends the bounds of Theorems 1 and 2 to the general side-information setting. The quantity $n_r(y^n)$ denotes the number of times $y_j = r$ in the sequence y^n:

$$n_r(y^n) = \sum_{j=1}^{n} I(y_j = r), \tag{37}$$

where $I(\cdot)$ is the indicator function.

Theorem 3 *The uniform weighted universal portfolio with side-information satisfies*

$$\frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} \leq \prod_{r=1}^{k} (n_r(y^n)+1)^{(m-1)} \leq (n+1)^{k(m-1)}, \tag{38}$$

for all $n, \mathbf{x}^n \in (\mathbb{R}_+^m)^n, y^n \in \{1,2,\ldots,k\}^n$.

Similarly, the Dirichlet$(1/2,\ldots,1/2)$ weighted universal portfolio with side-information satisfies

$$\frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} \leq 2^k \prod_{r=1}^{k} (n_r(y^n)+1)^{\frac{(m-1)}{2}} \leq 2^k (n+1)^{\frac{k(m-1)}{2}}, \tag{39}$$

for all $n, \mathbf{x}^n \in (\mathbb{R}_+^m)^n, y^n \in \{1,2,\ldots,k\}^n$.

This theorem immediately implies that

$$\lim_{n\to\infty} \sup_{\mathbf{x}^n, y^n} \frac{1}{n} \log \frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} = 0, \tag{40}$$

thereby proving that for these two μ's the universal portfolio with side-information is universal for the collection of state-constant rebalanced portfolios. Equations (32) and (33) follow from (38) and (39) by taking logarithms and normalizing by n.

The proofs of the theorems rely on the following lemmas. In both lemmas we adopt the conventions that $a/0 = \infty$ if $a > 0$, and that $0/0 = 0$.

Lemma 1 *If $\alpha_1,\ldots,\alpha_n \geq 0, \quad \beta_1,\ldots,\beta_n \geq 0$, then*

$$\frac{\sum_{i=1}^{n} \alpha_i}{\sum_{i=1}^{n} \beta_i} \leq \max_j \frac{\alpha_j}{\beta_j} \tag{41}$$

Proof of Lemma 1: Let

$$J = \arg\max_j \frac{\alpha_j}{\beta_j}. \tag{42}$$

The lemma is trivially true if $\alpha_J = 0$ since both the left and right side of (41) are zero. So assume $\alpha_J > 0$. Then, if $\beta_J = 0$ the lemma is true since the right side of (41) is infinity. So assume both $\alpha_J > 0$ and $\beta_J > 0$. Then

$$\frac{\sum_{j=1}^n \alpha_j}{\sum_{j=1}^n \beta_j} = \frac{\alpha_J(1 + \sum_{j\neq J} \frac{\alpha_j}{\alpha_J})}{\beta_J(1 + \sum_{j\neq J} \frac{\beta_j}{\beta_J})} \leq \frac{\alpha_J}{\beta_J} \tag{43}$$

because

$$\frac{\alpha_j}{\beta_j} \leq \frac{\alpha_J}{\beta_J} \tag{44}$$

which implies

$$\frac{\alpha_j}{\alpha_J} \leq \frac{\beta_j}{\beta_J} \tag{45}$$

for all j. □

Lemma 2 *For the μ-weighted universal portfolio,*

$$\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} \leq \max_{j^n} \frac{\prod_{i=1}^n b_{j_i}^*}{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})} \tag{46}$$

where the maximum is over the set of sequences of indices $j^n \in \{1,\ldots,m\}^n$, and $\mathbf{b}^* = (b_1^*,\ldots,b_m^*)^t$ is the best constant rebalanced portfolio for the sequence $\mathbf{x}^n$.

Proof of Lemma 2:

First recall the definitions

$$S_n^*(\mathbf{x}^n) = \prod_{i=1}^n \mathbf{b}^{*t}\mathbf{x}_i, \tag{47}$$

and

$$\hat{S}_n(\mathbf{x}^n) = \int_{\mathcal{B}} \prod_{i=1}^n \mathbf{b}^t\mathbf{x}_i d\mu(\mathbf{b}). \tag{48}$$

Rewrite $S_n^*(\mathbf{x}^n)$ as

$$\begin{aligned} S_n^*(\mathbf{x}^n) &= \prod_{i=1}^n \mathbf{b}^{*t}\mathbf{x}_i \\ &= \prod_{i=1}^n (\sum_{j=1}^m b_j^* x_{ij}) \\ &= \sum_{j^n\in\{1,\ldots,m\}^n} \prod_{i=1}^n b_{j_i}^* x_{ij_i}, \end{aligned} \tag{49}$$

where $j^n = (j_1, j_2, \ldots, j_n) \in \{1, \ldots, m\}^n$, and we have rewritten the product of sums as a sum of products. Similarly rewrite $\hat{S}_n(\mathbf{x}^n)$ as

$$\begin{aligned}
\hat{S}_n(\mathbf{x}^n) &= \int \prod_{i=1}^{n} \mathbf{b}^t \mathbf{x}_i \, d\mu(\mathbf{b}) \\
&= \sum_{j^n \in \{1,\ldots,m\}^n} \int \prod_{i=1}^{n} b_{j_i} x_{ij_i} \, d\mu(\mathbf{b}). \qquad (50)
\end{aligned}$$

The ratio of wealths can therefore be written as,

$$\begin{aligned}
\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} &= \frac{\sum_{j^n \in \{1,\ldots m\}^n} \prod_{i=1}^n b_{j_i}^* x_{ij_i}}{\sum_{j^n \in \{1,\ldots m\}^n} \int \prod_{i=1}^n b_{j_i} x_{ij_i} \, d\mu(\mathbf{b})} \qquad (51) \\
&= \frac{\sum_{j^n : \prod_{i=1}^n x_{ij_i} > 0} \prod_{i=1}^n b_{j_i}^* x_{ij_i}}{\sum_{j^n : \prod_{i=1}^n x_{ij_i} > 0} \int \prod_{i=1}^n b_{j_i} x_{ij_i} \, d\mu(\mathbf{b})}. \qquad (52)
\end{aligned}$$

Now apply Lemma 1 with $\alpha_{(j^n)} \triangleq \prod_{i=1}^n b_{j_i}^* x_{ij_i}$ and $\beta_{(j^n)} \triangleq \int \prod_{i=1}^n b_{j_i} x_{ij_i} d\mu(\mathbf{b})$ for $j^n \in \{j^n : \prod_{i=1}^n x_{ij_i} > 0\}$ to obtain

$$\begin{aligned}
\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} &\leq \max_{j^n : \prod_{i=1}^n x_{ij_i} > 0} \frac{\prod_{i=1}^n b_{j_i}^* x_{ij_i}}{\int \prod_{i=1}^n b_{j_i} x_{ij_i} \, d\mu(\mathbf{b})} \qquad (53) \\
&= \max_{j^n : \prod_{i=1}^n x_{ij_i} > 0} \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} \, d\mu(\mathbf{b})} \qquad (54) \\
&\leq \max_{j^n} \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} \, d\mu(\mathbf{b})}, \qquad (55)
\end{aligned}$$

where (54) follows since the product of the x_{ij_i}'s factors out of numerator and denominator. □

Theorems 1 and 2 are now proved by upper bounding the ratio (46) in Lemma 2 for the two μ's in question.

Proof of Theorem 1: Lemma 2 gives

$$\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} \leq \max_{j^n} \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} \, d\mu(\mathbf{b})}. \qquad (56)$$

For $r \in \{1, \ldots, m\}$, let $n_r(j^n)$ be the number of occurrences of r in j^n. Thus the numbers $(n_1(j^n)/n, \ldots, n_m(j^n)/n)$ denote the type of the sequence j^n. Letting $\nu_r(j^n) = n_r(j^n)/n$, from [6, p. 281] [16] for any sequence j^n and any $\mathbf{b}$

$$\prod_{i=1}^{n} b_{j_i} \leq 2^{-nH(\nu_1(j^n), \ldots, \nu_m(j^n))}, \qquad (57)$$

where $H(\nu_1(j^n), \ldots, \nu_m(j^n)) = \sum_{r=1}^m -\nu_r(j^n) \log \nu_r(j^n)$. Equality is achieved by

$$\mathbf{b} = (\nu_1(j^n), \ldots, \nu_m(j^n))^t. \qquad (58)$$

Further, for μ equal to the uniform distribution, $\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})$ can be evaluated in closed form as

$$\int \prod_{i=1}^{n} b_{j_i} d\mu(\mathbf{b}) = \frac{1}{\binom{n+m-1}{m-1} T(\nu_1(j^n), \ldots, \nu_m(j^n))}, \tag{59}$$

where $T(\nu_1(j^n), \ldots, \nu_m(j^n))$ is the number of sequences of type $(\nu_1(j^n), \ldots, \nu_m(j^n))$ and $\binom{n+m-1}{m-1}$ is the number of types. It is well known [6, 16] that $T(\nu_1(j^n), \ldots, \nu_m(j^n))$ is at most $2^{nH(\nu_1(j^n),\ldots,\nu_m(j^n))}$. Therefore,

$$\int \prod_{i=1}^{n} b_{j_i} d\mu(\mathbf{b}) \geq \frac{1}{\binom{n+m-1}{m-1}} 2^{-nH(\nu_1(j^n),\ldots,\nu_m(j^n))}. \tag{60}$$

Combining (56),(57), and (60) shows that

$$\begin{aligned} \frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} &\leq \max_{j^n} \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})} \\ &\leq \max_{j^n} \frac{2^{-nH(\nu_1(j^n),\ldots,\nu_m(j^n))}}{\frac{1}{\binom{n+m-1}{m-1}} 2^{-nH(\nu_1(j^n),\ldots,\nu_m(j^n))}} \\ &= \binom{n+m-1}{m-1}, \end{aligned} \tag{61}$$

proving the first inequality in the theorem. The second inequality follows from the fact that the number of types $\binom{n+m-1}{m-1}$ is bounded from above by $(n+1)^{m-1}$. □

Proof of Theorem 2: Again Lemma 2 gives

$$\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} \leq \max_{j^n} \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})}. \tag{62}$$

As above, bound the numerator using

$$\prod_{i=1}^{n} b_{j_i}^* \leq 2^{-nH(\nu_1(j^n),\ldots,\nu_m(j^n))}. \tag{63}$$

Now for μ equal to the Dirichlet$(1/2, \ldots, 1/2)$ distribution $\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})$ can also be evaluated in closed form as

$$\int \prod_{i=1}^{n} b_{j_i} d\mu(\mathbf{b}) = \frac{\Gamma(\frac{m}{2})}{\Gamma(n+\frac{m}{2})} \prod_{r=1}^{m} \frac{\Gamma(n_r(j^n)+\frac{1}{2})}{\Gamma(\frac{1}{2})}. \tag{64}$$

Let $D(n_1(j^n), \ldots, n_m(j^n))$ denote this quantity.

Lemma 5, proved at the end of this section, states that

$$\frac{2^{-nH(\nu_1(j^n),\ldots,\nu_m(j^n))}}{D(n_1(j^n), \ldots, n_m(j^n))} \leq \frac{\Gamma(\frac{1}{2})\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\Gamma(n+\frac{1}{2})} \leq 2(n+1)^{\frac{m-1}{2}}, \tag{65}$$

for all j^n and all n. (The asymptotic behavior of this ratio is analyzed in [17] and other works on universal data compression, where there is less emphasis on obtaining bounds valid for all n.)

Recapping,

$$\begin{aligned}\frac{S_n^*(\mathbf{x}^n)}{\hat{S}_n(\mathbf{x}^n)} &\leq \max_{j^n} \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})} \\ &\leq \max_{j^n} \frac{2^{-nH(\nu_1(j^n),\ldots,\nu_m(j^n))}}{D(n_1(j^n),\ldots,n_m(j^n))} \\ &\leq \frac{\Gamma(\frac{1}{2})\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\Gamma(n+\frac{1}{2})} \leq 2(n+1)^{\frac{m-1}{2}}.\square \end{aligned} \tag{66}$$

The proof of Theorem 3 relies on the expression for the wealth achieved by the μ-weighted universal portfolio with side-information given originally in (29):

$$\hat{S}_n(\mathbf{x}^n|y^n) = \prod_{r=1}^k \int_{\mathcal{B}} S_n(\mathbf{b}|r)d\mu(\mathbf{b}), \tag{67}$$

where the quantity

$$\int_{\mathcal{B}} S_n(\mathbf{b}|r)d\mu(\mathbf{b}) \tag{68}$$

is the wealth acquired by an ordinary universal portfolio operating on the subsequence of $\mathbf{x}^n$ corresponding to the times i that $y_i = r$. This equation expresses the fact that the universal portfolio is extended to the universal portfolio with side-information by dividing the sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ into k subsequences indexed by the side-information states $y_i \in \{1, 2, \ldots, k\}$ and treating each subsequence as a separate universal portfolio problem.

Proof of Theorem 3: The theorem follows readily from Theorems 1 and 2 and the expression for the running wealth achieved by the μ-weighted universal portfolio with side-information as given by (67). Specifically,

$$\frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} = \prod_{r=1}^k \frac{S_n(\mathbf{b}^*|r)}{\int_{\mathcal{B}} S_n(\mathbf{b}|r)d\mu(\mathbf{b})} \tag{69}$$

$$= \prod_{r=1}^k \frac{S_n^*(\{\mathbf{x}_i : y_i = r\})}{\hat{S}_n(\{\mathbf{x}_i : y_i = r\})}, \tag{70}$$

where $S_n^*(\{\mathbf{x}_i : y_i = r\})$ and $\hat{S}_n(\{\mathbf{x}_i : y_i = r\})$ respectively denote the wealth factors achieved by the best constant rebalanced portfolio and the universal portfolio restricted to the subsequence of stock vectors $\{\mathbf{x}_i : y_i = r\}$. For μ equal to the uniform distribution, Theorem 1 shows that

$$\frac{S_n^*(\{\mathbf{x}_i : y_i = r\})}{\hat{S}_n(\{\mathbf{x}_i : y_i = r\})} \leq (n_r(y^n)+1)^{(m-1)}, \tag{71}$$

so that

$$\frac{S_n^*(\mathbf{x}^n|y^n)}{\hat{S}_n(\mathbf{x}^n|y^n)} = \prod_{r=1}^{k} \frac{S_n^*(\{\mathbf{x}_i : y_i = r\})}{\hat{S}_n(\{\mathbf{x}_i : y_i = r\})} \tag{72}$$

$$\leq \prod_{r=1}^{k} (n_r(y^n)+1)^{(m-1)}, \tag{73}$$

proving the first inequality in the theorem for this case. The second inequality follows since $n_r(y^n) \leq n$ for each r.

A similar application of Theorem 2 to each subsequence proves the theorem for μ equal to the Dirichlet$(1/2, \ldots, 1/2)$ distribution. □

Now we prove Lemma 5 used in the proof of Theorem 2. First we need to develop some properties of the Gamma function.

Lemma 3 *For $x \geq 0$,*

$$\Gamma(x+1) \leq (x+\frac{1}{2})^{\frac{1}{2}} \Gamma(x+\frac{1}{2}). \tag{74}$$

Proof: From the log convexity of $\Gamma(x)$ [18] we have,for $x \geq 0$,

$$\Gamma(x+1) = \Gamma(\frac{1}{2}(x+\frac{3}{2}) + \frac{1}{2}(x+\frac{1}{2})) \tag{75}$$

$$\leq (\Gamma(x+\frac{3}{2}))^{\frac{1}{2}} (\Gamma(x+\frac{1}{2}))^{\frac{1}{2}} \tag{76}$$

$$= (x+\frac{1}{2})^{\frac{1}{2}} \Gamma(x+\frac{1}{2}). \square \tag{77}$$

Lemma 4 *Under the constraint that the x_i are non-negative integers summing to n, the function*

$$\phi(x_1, \ldots, x_m) = \prod_{r=1}^{m} \frac{x_r^{x_r}}{\Gamma(x_r + \frac{1}{2})} \tag{78}$$

is maximized by setting $x_1 = n$, and $x_2 = x_3 = \ldots = x_m = 0$.

Proof: It suffices to prove for $r \neq s$ and $x_r \geq x_s > 0$ that

$$\frac{\phi(x_1, \ldots, x_r, \ldots, x_s, \ldots, x_m)}{\phi(x_1, \ldots, x_r+1, \ldots, x_s-1, \ldots, x_m)} = \frac{g(x_r)}{g(x_r+1)} \frac{g(x_s)}{g(x_s-1)} \tag{79}$$

$$\leq 1, \tag{80}$$

where

$$g(x) = \frac{x^x}{\Gamma(x+\frac{1}{2})}. \tag{81}$$

The symmetry of ϕ and repeated applications of (80) to pairs of non-zero x_r and x_s lead to the desired result.

The proof of (80) amounts to showing that

$$\frac{g(n)}{g(n+1)}\frac{g(m)}{g(m-1)} = \frac{f(n)}{f(m-1)} \tag{82}$$
$$\leq 1 \tag{83}$$

for all integers $n \geq m-1 \geq 0$, where

$$f(x) = \frac{g(x)}{g(x+1)} \tag{84}$$
$$= \frac{x^x}{(x+1)^{x+1}}\frac{\Gamma(x+\frac{3}{2})}{\Gamma(x+\frac{1}{2})} \tag{85}$$
$$= \frac{x^x}{(x+1)^{x+1}}(x+\frac{1}{2}). \tag{86}$$

Equation (86) follows from the identity $\Gamma(x+1) = x\Gamma(x)$.

Since $f(n)/f(m-1) = 1$ for $n = m-1$, it now suffices to show that $f(x)$ is decreasing for $x > 0$. This follows from

$$\frac{d\log f(x)}{dx} = \frac{d}{dx}(x\log x - (x+1)\log(x+1) + \log(x+\frac{1}{2})) \tag{87}$$
$$= \log x - \log(x+1) + \frac{1}{x+\frac{1}{2}} \tag{88}$$
$$= \frac{1}{x+\frac{1}{2}} - \int_x^{x+1}\frac{1}{y}dy \tag{89}$$
$$\leq \frac{1}{x+\frac{1}{2}} - \frac{1}{\int_x^{x+1} y\,dy} \tag{90}$$
$$= \frac{1}{x+\frac{1}{2}} - \frac{1}{x+\frac{1}{2}} \tag{91}$$
$$= 0, \tag{92}$$

where (90) follows from Jensen's inequality applied to the convex function $1/y$. Thus $f(x)$ is decreasing thereby implying (83). This completes the proof of the lemma. □

The proof actually demonstrates a stronger result since relation (80) implies that

$$\phi(x_1,\ldots,x_m) = \prod_{r=1}^{m}\frac{x_r^{x_r}}{\Gamma(x_r+\frac{1}{2})} \tag{93}$$

is Schur convex [19] over the integers. The theory of majorization then implies that it is maximized by assigning $x_r = n$, for some r and letting $x_s = 0$ for the other indices $s \neq r$.

(An alternative derivation of the fact that ϕ is maximized by this choice of x_r is given by Csiszár [20].)

This brings us to the main lemma justifying (65) in the proof of Theorem 2.

Lemma 5 *For all non-negative integers $n_1, \ldots, n_m$, $\sum_{r=1}^m n_r = n$,*

$$\frac{2^{-nH(\nu_1,\ldots,\nu_m)}}{\frac{\Gamma(\frac{m}{2})}{\Gamma(n+\frac{m}{2})}\prod_{r=1}^m \frac{\Gamma(n_r+\frac{1}{2})}{\Gamma(\frac{1}{2})}} \le \frac{\Gamma(\frac{1}{2})\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\Gamma(n+\frac{1}{2})} \le 2(n+1)^{\frac{m-1}{2}}, \tag{94}$$

where, as above, $\nu_r = n_r/n$.

Proof: We proceed to bound the ratio (94) as follows:

$$\begin{aligned}
\frac{2^{-nH(\nu_1,\ldots,\nu_m)}}{\frac{\Gamma(\frac{m}{2})}{\Gamma(n+\frac{m}{2})}\prod_{r=1}^m \frac{\Gamma(n_r+\frac{1}{2})}{\Gamma(\frac{1}{2})}} &= \frac{(\Gamma(\frac{1}{2}))^m}{\Gamma(\frac{m}{2})}\frac{\Gamma(n+\frac{m}{2})}{n^n}\prod_{r=1}^m \frac{n_r^{n_r}}{\Gamma(n_r+\frac{1}{2})} &(95)\\
&= \frac{(\Gamma(\frac{1}{2}))^m}{\Gamma(\frac{m}{2})}\frac{\Gamma(n+\frac{m}{2})}{n^n}\phi(n_1,\ldots,n_r), &(96)
\end{aligned}$$

where, as in Lemma 4,

$$\phi(x_1,\ldots,x_m) = \prod_{r=1}^m \frac{x_r^{x_r}}{\Gamma(x_r+\frac{1}{2})}. \tag{97}$$

Lemma 4 states that ϕ is maximized by setting $x_1 = n, x_2 = 0, \ldots, x_m = 0$. This results in

$$\begin{aligned}
\frac{2^{-nH(\nu_1,\ldots,\nu_m)}}{\frac{\Gamma(\frac{m}{2})}{\Gamma(n+\frac{m}{2})}\prod_{r=1}^m \frac{\Gamma(n_r+\frac{1}{2})}{\Gamma(\frac{1}{2})}} &\le \frac{(\Gamma(\frac{1}{2}))^m}{\Gamma(\frac{m}{2})}\frac{\Gamma(n+\frac{m}{2})}{n^n}\phi(n,0,\ldots,0) &(98)\\
&= \frac{(\Gamma(\frac{1}{2}))^m}{\Gamma(\frac{m}{2})}\frac{\Gamma(n+\frac{m}{2})}{n^n}\frac{n^n}{\Gamma(n+\frac{1}{2})}\frac{1}{(\Gamma(\frac{1}{2}))^{m-1}} &(99)\\
&= \frac{\Gamma(\frac{1}{2})\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\Gamma(n+\frac{1}{2})} &(100)\\
&= \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{m}{2})}\frac{\Gamma(n+\frac{m}{2}-\lfloor\frac{m}{2}\rfloor+1)}{\Gamma(n+\frac{1}{2})}\prod_{i=1}^{\lfloor\frac{m}{2}\rfloor-1}\left(n+\frac{m}{2}-i\right) &(101)\\
&= \begin{cases} \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{m}{2})}\prod_{i=1}^{\frac{m-1}{2}}(n+\frac{m}{2}-i) & \text{for } m \text{ odd,}\\ \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{m}{2})}\frac{\Gamma(n+1)}{\Gamma(n+\frac{1}{2})}\prod_{i=1}^{\frac{m}{2}-1}(n+\frac{m}{2}-i) & \text{for } m \text{ even,}\end{cases} &(102)\\
& &(103)
\end{aligned}$$

where (101) follows from repeated applications of the identity $\Gamma(x+1) = x\Gamma(x)$. Note that (100) yields the first inequality in the lemma statement. To further simplify this expression for the case of m even, note that Lemma 3 implies $\Gamma(n+1)/\Gamma(n+1/2) \leq (n+1/2)^{\frac{1}{2}}$, yielding the final bound:

$$\frac{2^{-nH(\nu_1,\ldots,\nu_m)}}{\frac{\Gamma(\frac{m}{2})}{\Gamma(n+\frac{m}{2})}\prod_{r=1}^{m}\frac{\Gamma(n_r+\frac{1}{2})}{\Gamma(\frac{1}{2})}} \leq \begin{cases} \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{m}{2})}\prod_{i=1}^{\frac{m-1}{2}}(n+\frac{m}{2}-i) & \text{for } m \text{ odd,} \\ \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{m}{2})}(n+\frac{1}{2})^{\frac{1}{2}}\prod_{i=1}^{\frac{m}{2}-1}(n+\frac{m}{2}-i) & \text{for } m \text{ even,} \end{cases} \tag{104}$$

$$= \begin{cases} \Gamma(\frac{1}{2})\frac{\prod_{i=1}^{\frac{m-1}{2}}(n+\frac{m}{2}-i)}{\Gamma(\frac{1}{2})\prod_{i=1}^{\frac{m-1}{2}}(\frac{m}{2}-i)} & \text{for } m \text{ odd,} \\ \Gamma(\frac{1}{2})(n+\frac{1}{2})^{\frac{1}{2}}\frac{\prod_{i=1}^{\frac{m}{2}-1}(n+\frac{m}{2}-i)}{\prod_{i=1}^{\frac{m}{2}-1}(\frac{m}{2}-i)} & \text{for } m \text{ even,} \end{cases} \tag{105}$$

$$= \begin{cases} \prod_{i=1}^{\frac{m-1}{2}}\frac{(n+m/2-i)}{m/2-i} & \text{for } m \text{ odd,} \\ \Gamma(\frac{1}{2})(n+\frac{1}{2})^{\frac{1}{2}}\prod_{i=1}^{\frac{m}{2}-1}\frac{(n+m/2-i)}{m/2-i} & \text{for } m \text{ even,} \end{cases} \tag{106}$$

$$\leq \; 2(n+1)^{\frac{m-1}{2}} \text{ for all } n, \tag{107}$$

where (106) follows from the expansion of $\Gamma(m/2)$ for each case, and (107) follows from $\Gamma(1/2) = \sqrt{\pi} < 2$, and the fact that the i^{th} factor in each product in (106) is less than $(n+1)$, with the exception of one factor in the odd case which is less than $2(n+1)$, and hence the factor of 2. □

4.1 Universal data compression

The logarithm of the upper bound on $S_n^*(\mathbf{x}^n)/\hat{S}_n(\mathbf{x}^n)$ obtained in Lemma 2,

$$\max_{j^n} \log \frac{\prod_{i=1}^n b_{j_i}^*}{\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})}, \tag{108}$$

is a well studied quantity in the universal data compression of i.i.d. sources. The significance this ratio in the context of universal data compression becomes apparent if the components of a portfolio $\mathbf{b} = (b_1, \ldots, b_m)$ are viewed as probabilities on the stock indices $j \in \{1, \ldots, m\}$. The quantity $\prod_{i=1}^n b_{j_i}^*$ is then simply the probability of the index sequence $j^n \in \{1, \ldots, m\}^n$, if the indices are assumed independent and identically distributed according to $\mathbf{b}^*$. The quantity $\int \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})$ can also be viewed as the probability of j^n under a distribution which is a mixture according to μ of all the i.i.d. distributions. The logarithms of these two probabilities are, to within one bit, the codeword lengths assigned to j^n by the Shannon codes for the two distributions. We refer to the Shannon code for the $\mathbf{b}^*$ i.i.d. distribution as the $\mathbf{b}^*$ i.i.d. code, and to the Shannon code for the mixture distribution as the μ mixture code. Thus, the logarithm of the ratio in Lemma 2 is the worst case redundancy of the μ mixture code with respect to the $\mathbf{b}^*$ i.i.d. code; worst case over sequences j^n. The remarkable fact about the μ mixture code for μ equal to the uniform and Dirichlet$(1/2, \ldots, 1/2)$ distributions is that the worst case redundancies are upper bounded precisely by the logarithms of the bounds in Theorem 1 and Theorem 2 respectively. Further, these bounds are completely independent of $\mathbf{b}^*$. The proofs of Theorems 1 and 2 subsequent to Lemma 2 reargue this fact.

The importance of the μ mixture codes in universal data compression is thus apparent; for μ equal to the uniform and the Dirichlet$(1/2, \ldots, 1/2)$ distributions, the corresponding Shannon codes are pointwise universal for i.i.d. sources. This strong notion of universality means that for any i.i.d. source over m symbols and sufficiently large n, the encoding of *any* block of n source symbols according to a mixture code requires essentially no more bits per source symbol than encoding according to the Shannon code tailored to the underlying product distribution. This is stronger than requiring the expected redundancy (according to the underlying distribution) to be small. Another useful property of the mixture distributions is that they form a stochastically consistent sequence of distributions with increasing n. That is, for each μ there exists a stochastic process over source symbols with the marginal distribution on the first n symbols equal to the corresponding n^{th} order mixture distribution. This means that the above universal coding performance can be attained sequentially (as opposed to block-wise) using arithmetic coding according to the conditional distributions induced by the stochastic process [21].

It is the distribution-independent worst case redundancy result from universal data compression that enables us to obtain market independent bounds on the relative performance of the universal portfolio and the best constant rebalanced portfolio. Theorems 1 and 2 are obtained by reducing (via Lemma 2) the original investment problem to a well known problem involving the redundancies of mixture codes for i.i.d. sequences of stock indices.

4.2 Discussion

Examining the steps of the proof reveals that the bound in Lemma 2 holds with equality for Kelly gambling markets where the $\mathbf{x}_i$ are non-zero in only one component. Furthermore, it is known from universal data compression theory that the bounds on the expression in Lemma 2 given in Theorems 1 and 2 are also essentially tight. This confirms our intuition that a Kelly type market represents the least favorable investment environment, in the sense that the bounds on worst case performance are achieved. It is also apparent that the μ-weighted universal portfolio with μ equal to the Dirichlet$(1/2, \ldots, 1/2)$ distribution has a somewhat better worst case performance than μ equal to the uniform distribution.

The proof of Lemma 2 reveals another interpretation of the constant rebalanced portfolio and of the universal portfolio. Fix a time horizon n. For each sequence of stock indices $j^n \in \{1, \ldots, m\}^n$, consider the strategy which invests all wealth at time i in stock j_i. We will call these the extremal strategies indexed by j^n. The constant rebalanced portfolio strategy with portfolio $\mathbf{b} = (b_1, \ldots, b_m)^t$ now has the following interpretation. Divide the initial wealth into m^n piles indexed by $j^n = (j_1, j_2, \ldots, j_n)$, where the fraction of wealth in pile $j^n = (j_1, j_2, \ldots, j_n)$ is $\prod_{i=1}^n b_{j_i}$. Use the money in pile j^n to implement the extremal strategy corresponding to the sequence j^n. The plunging (extremal) strategy corresponding to the sequence j^n which invests an initial wealth of $\prod_{i=1}^n b_{j_i}$, accrues a wealth factor of $\prod_{i=1}^n x_{ij_i}$ and therefore achieves a final wealth of $\prod_{i=1}^n b_{j_i} x_{ij_i}$. It is easy to see that the total wealth factor at any time $l \leq n$ achieved by these n^m strategies running in parallel is exactly equal to the wealth $\sum_{j^l} \prod_{i=1}^{l} b_{j_i} x_{ij_i} = \prod_{i=1}^{l} \mathbf{b}^t \mathbf{x}_i$ of the constant rebalanced portfolio with portfolio $\mathbf{b}$. The universal portfolio can be interpreted similarly, but now the fraction of initial wealth assigned to pile j^n is $\int \prod_{i=1}^{n} b_{j_i} d\mu(\mathbf{b})$.

This interpretation gives rise to a collection of investment schemes corresponding to probability distributions $q(j_1, \ldots, j_n)$ on sequences of stock indices j^n. The probability $q(j^n)$ corresponds to the proportion of initial wealth assigned to the pile for extremal strategy j^n. The final wealth at time n for such a strategy is

$$\sum_{j^n} q(j^n) \prod_{i=1}^{n} x_{ij_i}. \tag{109}$$

A fixed time horizon is not fundamental to this interpretation. Given a stochastic process on the stock indices with conditional distributions $q_n(j_n|j_{n-1}, \ldots, j_1)$ and initial distribution $q_1(j_1)$, the following investment strategy achieves the wealth given by (109) with $q(j^n) = q_1(j_1) \prod_{i=2}^{n} q_i(j_i|j_{i-1}, \ldots, j_1)$. At time 1, let $b_{1i} = q_1(i)$ for $i \in 1, \ldots, m$. Define

$$w(j^{l-1}) = \prod_{i=1}^{l-1} b_{ij_i} x_{ij_i}. \tag{110}$$

Now for $\mathbf{b}_l = (b_{l1}, \ldots, b_{lm})$, use

$$b_{li} = \frac{\sum_{j^{l-1}} w(j^{l-1}) q_l(i|j^{l-1})}{\sum_{j^{l-1}} w(j^{l-1})}. \tag{111}$$

This is essentially a performance weighted average of the extremal strategies j^{l-1} with initial wealth proportions of $q(j^{l-1})$. Refer to such an investment scheme as a $q(\cdot)$-driven investment scheme.

What kind of investment strategies arise from various choices of $q(\cdot)$? The most naturally motivated schemes we have found in this family are the constant rebalanced portfolio, where $q(\cdot)$ is the measure induced by an i.i.d. process on the stock indices, and the universal portfolio where $q(\cdot)$ is a mixture of i.i.d. distributions. It is interesting to consider the scheme corresponding to $q(\cdot)$ given by the Lempel-Ziv induced probabilities [5, 22]. This scheme would have universal properties with respect to the collection of $q(\cdot)$-driven schemes with $q(\cdot)$ equal to any finite state (unifilar) stochastic process on the stock indices. There are, however, some motivational problems for this collection of schemes. The problem is that for a general such scheme $q(\cdot)$, the only naturally motivated stochastic process on price relatives for which this scheme is growth rate optimal in the distributional sense seems to be the Kelly market distributed according to $q(\cdot)$. This is in contrast to the constant rebalanced portfolios (corresponding to $q(\cdot)$ i.i.d. , and hence a subset of the above collection) which are distributionally growth rate optimal for i.i.d. distributions on price relatives (not just Kelly type).

An important aspect of Theorem 3 is the absence of any assumptions about the dependence between stock vectors and side-information states. This dependence can be arbitrary and, in fact, the side-information can be a function of the entire market sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. An interesing example is if y_i indicates which stock will be the best performer on day i. Of course, in this case, with knowledge of this dependence the investor can take full advantage of the sequence y^n to invest all wealth in the best stock each trading day. This usually results in astronomically high performance, even for the actual market. The problem is that the investor doesn't know ahead of time how valuable this side-information sequence is. Nonetheless, the μ-weighted universal portfolio with side-information is able to learn the association cautiously yet rapidly, and then perform as well, to first order in the exponent, as if this dependence of y and $\mathbf{x}$ were known ahead of time. The cost in the exponent, is $(1/(2n))\log n$ per degree of freedom (for the Dirichlet$(1/2, \ldots, 1/2)$ weighting). Since there are $k = m$ states of side-information (necessary for indicating the best stock on each trading day), and there are $m-1$ degrees of freedom in the portfolio $\mathbf{b} \in \mathcal{B}$, the loss in the exponential growth rate is $(m(m-1)/(2n))\log n$. This is asymptotically negligible as $n \to \infty$.

5 Example

We present a simple example illustrating the above results. Let $a > 1$, and let

$$\mathbf{x}_1, \mathbf{x}_2, \ldots = (1,\ a)^t,\ (1,\ \frac{1}{a})^t,\ (1,\ a)^t,\ (1,\ \frac{1}{a})^t, \ldots$$

be the sequence of stock market vectors. Note that the first component x_{i1} of the stock vector $\mathbf{x}_i$ at time i is constantly equal to 1 for $i = 1, 2, \ldots, n$. This first component represents a risk free asset (or cash). An investment in the first stock returns the investment — a dollar in is a dollar out. On the other hand, the second stock x_{i2} is highly volatile, jumping up and down by a factor of a or $1/a$ each investment day. Both stocks are going nowhere. A buy-and-hold strategy in stock 1 results in $S_n = \prod x_{i1} = 1$; a buy-and-hold of stock 2 results in $S_n = \prod x_{i2} = 1$, when n is even. Also, this sequence has been maliciously chosen to perform contrary to naive expectation. For example, whenever stock 2 has outperformed stock 1 in the past, it plunges by a factor of $1/a$.

Now consider the behavior of a constant rebalanced portfolio $\mathbf{b} = (b, 1-b)$ on this sequence. Then, for n even,

$$S_n(\mathbf{b}) = (b + (1-b)a)^{\frac{n}{2}}(b + \frac{(1-b)}{a})^{\frac{n}{2}}. \tag{112}$$

Setting the derivative to 0, we find the maximum wealth is achieved by rebalancing each time to

$$\mathbf{b}^* = (\frac{1}{2}, \frac{1}{2}), \tag{113}$$

resulting in wealth

$$S_n^*(\mathbf{x}^n) = \left(\sqrt{(1+a)(1+1/a)(1/4)}\right)^n, \tag{114}$$

for $n = 2, 4, 6, \ldots$ (n even). Since $(1+a)(1+1/a)(1/4) > 1$, for $a \neq 1$, the wealth grows exponentially to infinity.

Now consider side-information with $|\mathcal{Y}| = 2$ states :

$$\begin{aligned} y_i &= \begin{cases} 1, & \prod_{l=1}^{i-1} x_{l2} \leq \prod_{l=1}^{i-1} x_{l1} \\ 2, & \prod_{l=1}^{i-1} x_{l2} > \prod_{l=1}^{i-1} x_{l1} \end{cases} \\ y_1 &= 1. \end{aligned}$$

Thus y_i indicates whether the running price of stock 2 exceeds stock 1 (cash) at time i. The sequences look like this:

$$\begin{array}{lcccccc} \mathbf{x}_i & : & (1,\ a)^t, & (1,\ \frac{1}{a})^t, & (1,\ a)^t, & (1,\ \frac{1}{a})^t, & \ldots \\ y_i & : & 1, & 2, & 1, & 2, & \ldots \end{array}$$

Note that this simple calculation based on the past yields side-information that gives perfect investment information. An investor knowing $(\mathbf{x}^n, y^n)$ would make perfect investment decisions, and hence the best state-constant rebalanced portfolio is

$$\begin{array}{lcccccc} \mathbf{b}_i(y_i) & : & (0,\ 1)^t, & (1,\ 0)^t, & (0,\ 1)^t, & (1,\ 0)^t & \ldots \\ S_n^*(\mathbf{x}^k|y^k) & : & a, & a, & a^2, & a^2, & \ldots \end{array}$$

By investing in the best stock each time, the wealth gained by the best state-constant rebalanced portfolio is

$$S_n^*(\mathbf{x}^n|y^n) = a^{\frac{n}{2}}, \tag{115}$$

for n even. Of course this is much larger than (114).

We now investigate the performance of the universal portfolio on the same sequence. As given by (29), for n even, the universal portfolio with side-information achieves a wealth of

$$\hat{S}_n(\mathbf{x}^n|y^n) = \left(\int_{\mathcal{B}} \prod_{i:y_i=1} \mathbf{b}^t \mathbf{x}_i d\mu(\mathbf{b}) \right) \left(\int_{\mathcal{B}} \prod_{i:y_i=2} \mathbf{b}^t \mathbf{x}_i d\mu(\mathbf{b}) \right), \tag{116}$$

which, using results from Section 6 for the Dirichlet$(1/2, \ldots, 1/2)$ distribution, we can express by

$$\begin{aligned} &\hat{S}_n(\mathbf{x}^n|y^n) \\ &= \left(\int_{\mathcal{B}} \prod_{i:y_i=1} \mathbf{b}^t \mathbf{x}_i d\mu(\mathbf{b}) \right) \left(\int_{\mathcal{B}} \prod_{i:y_i=2} \mathbf{b}^t \mathbf{x}_i d\mu(\mathbf{b}) \right) \\ &= \left(\int_0^1 (b + a(1-b))^{\frac{n}{2}} d\mu(b) \right) \left(\int_0^1 (b + \frac{1}{a}(1-b))^{\frac{n}{2}} d\mu(b) \right) \\ &= \left(\frac{1}{2^n} \sum_{l=0}^{\frac{n}{2}} \binom{2l}{l} \binom{2(n/2-l)}{n/2-l} a^{\frac{n}{2}-l} \right) \end{aligned}$$

$$
\cdot \left(\frac{1}{2^n} \sum_{l=0}^{\frac{n}{2}} \binom{2l}{l} \binom{2(n/2-l)}{n/2-l} a^{-(\frac{n}{2}-l)} \right)
$$

$$
\geq \left(\frac{1}{2^n} \binom{n}{n/2} a^{\frac{n}{2}} \right) \left(\frac{1}{2^n} \binom{n}{n/2} \right). \tag{117}
$$

Using Stirling's approximation, we note, for purposes of comparison, that

$$
\begin{aligned}
\left(\frac{1}{2^n} \binom{n}{n/2} a^{\frac{n}{2}} \right) \left(\frac{1}{2^n} \binom{n}{n/2} \right) &\sim \left(\frac{1}{\sqrt{\pi n/2}} a^{\frac{n}{2}} \right) \left(\frac{1}{\sqrt{\pi n/2}} \right) \\
&= \frac{2}{\pi n} a^{\frac{n}{2}}.
\end{aligned} \tag{118}
$$

A similar analysis is easily carried out for n odd. We see that (118) and (115) agree to first order in the exponent. In fact, Theorem 3 states that for any n the overall wealth factor for the universal portfolio with side-information is no smaller than

$$
\begin{aligned}
\hat{S}_n(\mathbf{x}^n|y^n) &\geq \frac{1}{4(\frac{n}{2}+1)} a^{\frac{n}{2}} \\
&= \frac{1}{4(\frac{n}{2}+1)} S_n^*(\mathbf{x}^n|y^n).
\end{aligned} \tag{119}
$$

Thus the target wealth of the best state-constant rebalanced portfolio is attained to within a factor of $4(n/2+1)$. A factor of $2\sqrt{(n/2+1)}$ arises from each of the subsequences $\{i : y_i = 1\}$ and $\{i : y_i = 2\}$.

6 Computing the universal portfolio

This section presents a simple method for the exact computation of the μ-weighted universal portfolio for μ equal to the Dirichlet$(1/2, \ldots, 1/2)$ distribution. The idea is to compute the universal portfolio recursively in a manner similar to the recursive calculation of the binomial coefficients. We illustrate this for the case of $m = 2$ stocks and $|\mathcal{Y}| = 1$ (no side-information). The method generalizes easily to more stocks and side-information.

As in the proof of Lemma 2, we begin by rewriting the wealth accrued by a constant rebalanced portfolio,

$$
\begin{aligned}
S_n(\mathbf{x}^n, \mathbf{b}) &= \prod_{i=1}^{n} \mathbf{b}^t \mathbf{x}_i \\
&= \prod_{i=1}^{n} (b_1 x_{i1} + b_2 x_{i2}) \\
&= \sum_{j^n \in \{1,2\}^n} \prod_{k=1}^{n} b_{j_i} x_{i j_i}
\end{aligned}
$$

$$= \sum_{l=0}^{n} b_1^l b_2^{(n-l)} \Big(\sum_{j^n \in T_n(l)} \prod_{i=1}^{n} x_{ij_i} \Big), \tag{120}$$

where $T_n(l)$ is the set of all sequences $j^n \in \{1,2\}^n$ with l 1's and $n-l$ 2's. Letting

$$X_n(l) \;=\; \sum_{j^n \in T_n(l)} \prod_{i=1}^{n} x_{ij_i}, \tag{121}$$

we have

$$S_n(\mathbf{x}^n, \mathbf{b}) = \sum_{l=0}^{n} b_1^l b_2^{(n-l)} X_n(l). \tag{122}$$

Integrating this expression yields the wealth accrued by the universal portfolio at time n,

$$\begin{aligned}
\hat{S}_n(\mathbf{x}^n) &= \int \sum_{l=0}^{n} b_1^l b_2^{(n-l)} X_n(l)\, d\mu(\mathbf{b}) \\
&= \sum_{l=0}^{n} X_n(l) \int b_1^l b_2^{(n-l)}\, d\mu(\mathbf{b}) \\
&= \sum_{l=0}^{n} X_n(l) C_n(l),
\end{aligned} \tag{123}$$

where we have defined

$$C_n(l) = \int b_1^l b_2^{(n-l)}\, d\mu(\mathbf{b}). \tag{124}$$

This integral, with μ equal to the Dirichlet$(1/2, 1/2)$ density, can be evaluated in closed form to obtain

$$\begin{aligned}
C_n(l) &= \frac{(l-\frac{1}{2})\cdot(l-1-\frac{1}{2})\cdots(\frac{1}{2})\cdot(n-l-\frac{1}{2})\cdots(\frac{1}{2})}{n\cdot(n-1)\cdots 2\cdot 1} \\
&= \frac{l!(n-l)!\binom{2l}{l}\binom{2(n-l)}{n-l}}{n! 2^{2n}}.
\end{aligned} \tag{125}$$

We note for reference that, by Stirling's formula,

$$C_n(l) \sim \frac{1}{\binom{n}{l}\pi\sqrt{l(n-l)}}. \tag{126}$$

We can obtain an expression analogous to (123) for $\hat{\mathbf{b}}_n$, the specified portfolio at time n. First recall that

$$\begin{aligned}
\hat{\mathbf{b}}_n &= \frac{1}{\hat{S}_{n-1}(\mathbf{x}^{n-1})} \int \prod_{i=1}^{n-1} \mathbf{b}^t \mathbf{x}_i \mathbf{b}\, d\mu(\mathbf{b}) \\
&= \frac{1}{\hat{S}_{n-1}(\mathbf{x}^{n-1})} \begin{bmatrix} \int \prod_{i=1}^{n-1} \mathbf{b}^t \mathbf{x}_i b_1\, d\mu(\mathbf{b}) \\ \int \prod_{i=1}^{n-1} \mathbf{b}^t \mathbf{x}_i b_2\, d\mu(\mathbf{b}) \end{bmatrix}.
\end{aligned} \tag{127}$$

Now proceeding exactly as in the simplification of $\hat{S}_n(\mathbf{x}^n)$ above we get

$$
\begin{aligned}
\hat{\mathbf{b}}_n &= \frac{1}{\hat{S}_{n-1}(\mathbf{x}^{n-1})} \begin{bmatrix} \sum_{l=0}^{n-1} X_{n-1}(l) \int b_1^{(l+1)} b_2^{(n-1-l)}\, d\mu(\mathbf{b}) \\ \sum_{l=0}^{n-1} X_{n-1}(l) \int b_1^{l} b_2^{(n-l)}\, d\mu(\mathbf{b}) \end{bmatrix} \\
&= \frac{1}{\hat{S}_{n-1}(\mathbf{x}^{n-1})} \begin{bmatrix} \sum_{l=0}^{n-1} C_n(l+1) X_{n-1}(l) \\ \sum_{l=0}^{n-1} C_n(l) X_{n-1}(l) \end{bmatrix}. \qquad (128)
\end{aligned}
$$

The next step is to observe that the quantities $C_n(l)$ and $X_n(l)$ can be computed recursively. Taking $C_0(0) = 1$ we have two recursions for $C_n(l)$,

$$
\begin{aligned}
C_n(l) &= \frac{(l - \frac{1}{2})}{n} C_{n-1}(l-1) \\
C_n(l) &= \frac{(n - l - \frac{1}{2})}{n} C_{n-1}(l). \qquad (129)
\end{aligned}
$$

For $X_n(l)$ we have the recursion

$$
X_n(l) = x_{n1} X_{n-1}(l-1) + x_{n2} X_{n-1}(l), \qquad (130)
$$

valid for $1 \le l \le n-1$, with the obvious endpoint conditions

$$
\begin{aligned}
X_n(0) &= x_{n2} X_{n-1}(0) \\
X_n(n) &= x_{n1} X_{n-1}(n-1). \qquad (131)
\end{aligned}
$$

We now have all the ingredients to compute the universal portfolio expressions. Recursively compute and store the quantities $X_n(l)$ and $C_n(l)$ and then insert them into the above expressions for $\hat{S}_n(\mathbf{x}^n)$ and $\hat{\mathbf{b}}_n$.

We can further simplify the computation by noticing that $\hat{S}_n(\mathbf{x}^n)$ and $\hat{\mathbf{b}}_n$ depend on $X_n(l)$ and $C_n(l)$ only through their product $Q_n(l)$,

$$
Q_n(l) = X_n(l) C_n(l). \qquad (132)
$$

In particular, we can write $\hat{S}_n(\mathbf{x}^n)$ as

$$
\hat{S}_n(\mathbf{x}^n) = \sum_{l=0}^{n} Q_n(l). \qquad (133)
$$

We can also use the $Q_n(l)$ to express the quantities $C_n(l+1)X_{n-1}(l)$ and $C_n(l)X_{n-1}(l)$ appearing in the expression for $\hat{\mathbf{b}}_n$. This is given by

$$
\begin{aligned}
C_n(l+1) X_{n-1}(l) &= \frac{(l + 1 - \frac{1}{2})}{n} Q_{n-1}(l), \\
C_n(l) X_{n-1}(l) &= \frac{(n - l - \frac{1}{2})}{n} Q_{n-1}(l). \qquad (134)
\end{aligned}
$$

Therefore, $\hat{\mathbf{b}}_n$ is given by

$$\hat{\mathbf{b}}_n = \frac{1}{\sum_{l=0}^{n-1} Q_{n-1}(l)} \begin{bmatrix} \sum_{l=0}^{n-1} \frac{(l+1-\frac{1}{2})}{n} Q_{n-1}(l) \\ \sum_{l=0}^{n-1} \frac{(n-l-\frac{1}{2})}{n} Q_{n-1}(l) \end{bmatrix}. \tag{135}$$

Thus it makes sense to compute only the $Q_n(l)$. This is advantageous from a numerical standpoint as well since $C_n(l)$ and $X_n(l)$ are respectively exponentially decreasing and increasing in n. Computation of the $Q_n(l)$ can also be done recursively by combining the above recursions for $X_n(l)$ and $C_n(l)$. The $Q_n(l)$ recursions are given by

$$Q_n(l) = x_{n1} \frac{(l-\frac{1}{2})}{n} Q_{n-1}(l-1) + x_{n2} \frac{(n-l-\frac{1}{2})}{n} Q_{n-1}(l), \tag{136}$$

again valid for $1 \le l \le n-1$, and the endpoint recursions,

$$\begin{aligned} Q_n(0) &= x_{n2} \frac{(n-\frac{1}{2})}{n} Q_{n-1}(0) \\ Q_n(n) &= x_{n1} \frac{(n-\frac{1}{2})}{n} Q_{n-1}(n-1). \end{aligned} \tag{137}$$

The initial condition is $Q_0(0) = 1$.

To summarize, we have written the two universal portfolio expressions, $\hat{S}_n(\mathbf{x}^n)$ and $\hat{\mathbf{b}}_n$, in terms of the $Q_n(l)$. The accrued wealth is given by

$$\hat{S}_n(\mathbf{x}^n) = \sum_{l=0}^{n} Q_n(l). \tag{138}$$

and the portfolio at time n by

$$\hat{\mathbf{b}}_n = \frac{1}{\hat{S}_{n-1}(\mathbf{x}^{n-1})} \begin{bmatrix} \sum_{l=0}^{n-1} \frac{(l+1-\frac{1}{2})}{n} Q_{n-1}(l) \\ \sum_{l=0}^{n-1} \frac{(n-l-\frac{1}{2})}{n} Q_{n-1}(l) \end{bmatrix}. \tag{139}$$

The exact computation of the two-stock universal portfolio at time n thus simplifies to recursively computing and storing the $n+1$ quantities $Q_n(l)$ and then substituting them into the above two expressions. This method generalizes easily to m stocks, with the storage requirements for the recursive computations growing like n^{m-1}.

7 Conclusion

There are some practical considerations concerning the universal portfolio with side-information which we have not addressed. One such consideration is the question of trading costs. Optimal investing in the presence of commissions is poorly understood throughout finance and portfolio theory. For example, it was pointed out that constant rebalanced portfolios are growth rate optimal for i.i.d. markets. This is only valid in the absence of trading costs and there is no general solution to the growth rate optimal investment problem with commissions. As for universal investment, it is not even clear what that means in the presence of trading costs. Perhaps the exponential growth rate of wealth of the best constant rebalanced portfolio computed in hindsight is, in fact, not achievable sequentially when trading costs must be taken into account.

Another practical consideration is the nature of side-information. In this paper we assumed that an abstract source of side-information is available for portfolio decision making. A challenging yet useful effort would be to isolate practical sources of side-information appropriate for real world markets. An important point to keep in mind in such an undertaking is the dimensionality tradeoff evident in Theorem 3. Increasing the number of states k of side-information (and hence the dimension $d = k(m-1)$) increases $S_n^*(\mathbf{x}^n|y^n)$, the wealth achieved by the best state-constant rebalanced portfolio. This is good. However, the factor by which the universal portfolio underperforms this improved target wealth increases with k roughly like $(\sqrt{n})^{k(m-1)}$. This is bad and hence the dimensionality tradeoff. This tradeoff is also an issue in data compression and prediction.

The main points of this paper are the extension of the universal portfolio to incorporate side-information and the derivation of worst case performance bounds for the universal portfolio with side-information. Theorem 3 shows that the wealth achieved by the μ-weighted universal portfolio with side-information is within a polynomial factor of the wealth achieved by the best state-constant rebalanced portfolio computed in hindsight up through any time n and any sequence of price relatives $\mathbf{x}^n$ and side-information y^n. This is an individual sequence result in the sense that the polynomial factor holds for every sequence and not just on the average or in probability.

The proof of this result establishes a connection between universal investment and universal data compression, which accounts for the $-(d/(2n))\log n$ lower bound on the universal portfolio's exponential growth rate of wealth relative to the best achievable. The strongest bounds on the redundancies of universal source codes are also of this form with d related to the source alphabet size. In our case $d = (m-1)k$ is the number of degrees of freedom in a state-constant rebalanced portfolio with k states and m stocks.

Acknowledgements

We would like to acknowledge the anonymous referees and Meir Feder for their helpful comments and suggestions.

References

[1] J. L. Kelly, Jr. A new interpretation of information rate. *Bell Syst. Tech. J.*, 35:917–926, 1956.

[2] P. H. Algoet and T. M. Cover. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *Ann. Probab.*, 16(2):876–898, 1988.

[3] R. Bell and T. M. Cover. Competitive optimality of logarithmic investment. *Math. Oper. Res.*, 5:161–166, 1980.

[4] R. Bell and T. M. Cover. Game-theoretic optimal portfolios. *Management Sci.*, 34:724–733, 1988.

[5] P. H. Algoet. Universal schemes for prediction, gambling, and portfolio selection. *Ann. Probab.*, pages 901–941, April 1992.

[6] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley, New York, 1991.

[7] H. Robbins. Asymptotically subminimax solutions of compound statistical decision problems. In *Proc. 2nd Berkeley Symp. Math. Statistics. Probab.*, pages 131–148, 1951.

[8] J. F. Hannan. *Contributions to the Theory of Games, vol. III, Annals of Mathematics Studies*, pages 97–139. Princeton Univ. Press, 1957.

[9] D. C. Gilliland. Asymptotic risk stability resulting from play against the past in a sequence of decision problems. *IEEE Trans. Info. Theory*, 18(5):614–617, 1972.

[10] N. Merhav and M. Feder. Universal schemes for sequential decision from individual data sequences. *IEEE Trans. Info. Theory*, 39(4):1280–1292, July 1993.

[11] D. Blackwell. Controlled random walks. In *Proceedings of International Congress of Mathematics, vol. III*, pages 336–338, Amsterdam, North Holland, 1956.

[12] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[13] T. M. Cover and D. Gluss. Empirical Bayes stock market portfolios. *Adv. Appl. Math.*, 7:170–181, 1986.

[14] D. C. Larson. *Growth Optimal Trading Strategies.* PhD thesis, Stanford University, Stanford, California, 1986.

[15] T. M. Cover. Universal portfolios. *Math. Finance*, 1(1):1–29, January 1991.

[16] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.

[17] R. E. Krichevsky and V. K. Trofimov. The performance of universal encoding. *IEEE Trans. Info. Theory*, 27(2):199–207, March 1981.

[18] W. Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, Third edition, 1976.

[19] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*, volume 143 of *Mathematics in Science and Engineering.* Academic Press, London, 1979.

[20] I. Csiszár. Notes from lecture on universal data compression delivered at Stanford University. Unpublished, March 1993.

[21] J. Rissanen and G. G. Langdon, Jr. Universal modeling and coding. *IEEE Trans. Info. Theory*, 27(1):12–23, January 1981.

[22] J. Rissanen. A universal data compression system. *IEEE Trans. Info. Theory*, 29:656–664, 1983.