CHAPTER 16

INFORMATION THEORY AND PORTFOLIO THEORY

The duality between the growth rate of wealth in the stock market and the entropy rate of the market is striking. In particular, we shall find the competitively optimal and growth rate optimal portfolio strategies. They are the same, just as the Shannon code is optimal both competitively and in the expected description rate. We also find the asymptotic growth rate of wealth for an ergodic stock market process. We end with a discussion of universal portfolios that enable one to achieve the same asymptotic growth rate as the best constant rebalanced portfolio in hindsight.

In Section 16.8 we provide a "sandwich" proof of the asymptotic equipartition property for general ergodic processes that is motivated by the notion of optimal portfolios for stationary ergodic stock markets.

16.1 THE STOCK MARKET: SOME DEFINITIONS

A stock market is represented as a vector of stocks $\mathbf{X} = (X_1, X_2, \ldots, X_m)$, $X_i \geq 0$, $i = 1, 2, \ldots, m$, where m is the number of stocks and the *price relative* X_i is the ratio of the price at the end of the day to the price at the beginning of the day. So typically, X_i is near 1. For example, $X_i = 1.03$ means that the ith stock went up 3 percent that day.

Let $\mathbf{X} \sim F(\mathbf{x})$, where $F(\mathbf{x})$ is the joint distribution of the vector of price relatives. A *portfolio* $\mathbf{b} = (b_1, b_2, \ldots, b_m)$, $b_i \geq 0$, $\sum b_i = 1$, is an allocation of wealth across the stocks. Here b_i is the fraction of one's wealth invested in stock i. If one uses a portfolio $\mathbf{b}$ and the stock vector is $\mathbf{X}$, the wealth relative (ratio of the wealth at the end of the day to the wealth at the beginning of the day) is $S = \mathbf{b}^t\mathbf{X} = \sum_{i=1}^m b_i X_i$.

We wish to maximize S in some sense. But S is a random variable, the distribution of which depends on portfolio $\mathbf{b}$, so there is controversy

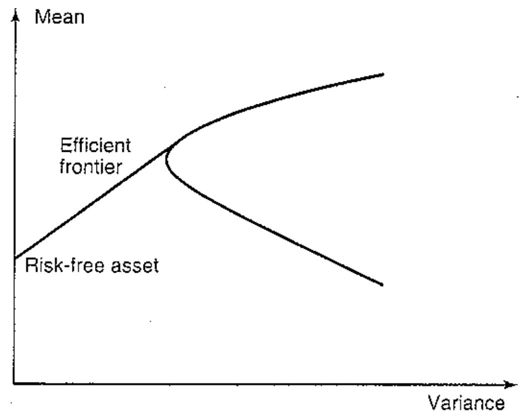

FIGURE 16.1. Sharpe–Markowitz theory: set of achievable mean–variance pairs.

over the choice of the best distribution for S. The standard theory of stock market investment is based on consideration of the first and second moments of S. The objective is to maximize the expected value of S subject to a constraint on the variance. Since it is easy to calculate these moments, the theory is simpler than the theory that deals with the entire distribution of S.

The mean–variance approach is the basis of the Sharpe–Markowitz theory of investment in the stock market and is used by business analysts and others. It is illustrated in Figure 16.1. The figure illustrates the set of achievable mean–variance pairs using various portfolios. The set of portfolios on the boundary of this region corresponds to the undominated portfolios: These are the portfolios that have the highest mean for a given variance. This boundary is called the *efficient frontier*, and if one is interested only in mean and variance, one should operate along this boundary.

Normally, the theory is simplified with the introduction of a *risk-free* asset (e.g., cash or Treasury bonds, which provide a fixed interest rate with zero variance). This stock corresponds to a point on the Y axis in the figure. By combining the risk-free asset with various stocks, one obtains all points below the tangent from the risk-free asset to the efficient frontier. This line now becomes part of the efficient frontier.

The concept of the efficient frontier also implies that there is a true price for a stock corresponding to its risk. This theory of stock prices, called the *capital asset pricing model* (CAPM), is used to decide whether the market price for a stock is too high or too low. Looking at the mean of a random variable gives information about the long-term behavior of

the sum of i.i.d. versions of the random variable. But in the stock market, one normally reinvests every day, so that the wealth at the end of n days is the product of factors, one for each day of the market. The behavior of the product is determined not by the expected value but by the expected logarithm. This leads us to define the growth rate as follows:

Definition The *growth rate* of a stock market portfolio $\mathbf{b}$ with respect to a stock distribution $F(\mathbf{x})$ is defined as

$$W(\mathbf{b}, F) = \int \log \mathbf{b}^t\mathbf{x}\, dF(\mathbf{x}) = E\left(\log \mathbf{b}^t\mathbf{X}\right). \tag{16.1}$$

If the logarithm is to base 2, the growth rate is also called the *doubling rate*.

Definition The *optimal growth rate* $W^*(F)$ is defined as

$$W^*(F) = \max_{\mathbf{b}} W(\mathbf{b}, F), \tag{16.2}$$

where the maximum is over all possible portfolios $b_i \geq 0$, $\sum_i b_i = 1$.

Definition A portfolio $\mathbf{b}^*$ that achieves the maximum of $W(\mathbf{b}, F)$ is called a *log-optimal portfolio* or *growth optimal portfolio.*

The definition of growth rate is justified by the following theorem, which shows that wealth grows as 2^{nW^*}.

Theorem 16.1.1 *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. according to $F(\mathbf{x})$. Let*

$$S_n^* = \prod_{i=1}^{n} \mathbf{b}^{*t}\mathbf{X}_i \tag{16.3}$$

be the wealth after n days using the constant rebalanced portfolio $\mathbf{b}^*$. Then

$$\frac{1}{n}\log S_n^* \to W^* \quad \text{with probability } 1. \tag{16.4}$$

Proof: By the strong law of large numbers,

$$\frac{1}{n}\log S_n^* = \frac{1}{n}\sum_{i=1}^{n} \log \mathbf{b}^{*t}\mathbf{X}_i \tag{16.5}$$

$$\to W^* \quad \text{with probability } 1. \tag{16.6}$$

Hence, $S_n^* \doteq 2^{nW^*}$. □

We now consider some of the properties of the growth rate.

Lemma 16.1.1 *$W(\mathbf{b}, F)$ is concave in $\mathbf{b}$ and linear in F. $W^*(F)$ is convex in F.*

Proof: The growth rate is

$$W(\mathbf{b}, F) = \int \log \mathbf{b}'\mathbf{x}\, dF(\mathbf{x}). \tag{16.7}$$

Since the integral is linear in F, so is $W(\mathbf{b}, F)$. Since

$$\log(\lambda \mathbf{b}_1 + (1-\lambda)\mathbf{b}_2)'\mathbf{X} \geq \lambda \log \mathbf{b}_1'\mathbf{X} + (1-\lambda) \log \mathbf{b}_2'\mathbf{X}, \tag{16.8}$$

by the concavity of the logarithm, it follows, by taking expectations, that $W(\mathbf{b}, F)$ is concave in $\mathbf{b}$. Finally, to prove the convexity of $W^*(F)$ as a function of F, let F_1 and F_2 be two distributions on the stock market and let the corresponding optimal portfolios be $\mathbf{b}^*(F_1)$ and $\mathbf{b}^*(F_2)$, respectively. Let the log-optimal portfolio corresponding to $\lambda F_1 + (1-\lambda)F_2$ be $\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2)$. Then by linearity of $W(\mathbf{b}, F)$ with respect to F, we have

$$\begin{aligned}
&W^*(\lambda F_1 + (1-\lambda)F_2) \\
&\quad = W(\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2), \lambda F_1 + (1-\lambda)F_2) &(16.9)\\
&\quad = \lambda W(\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2), F_1) \\
&\qquad + (1-\lambda) W(\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2), F_2) \\
&\quad \leq \lambda W(\mathbf{b}^*(F_1), F_1) + (1-\lambda) W^*(\mathbf{b}^*(F_2), F_2), &(16.10)
\end{aligned}$$

since $\mathbf{b}^*(F_1)$ maximizes $W(\mathbf{b}, F_1)$ and $\mathbf{b}^*(F_2)$ maximizes $W(\mathbf{b}, F_2)$. □

Lemma 16.1.2 *The set of log-optimal portfolios with respect to a given distribution is convex.*

Proof: Suppose that $\mathbf{b}_1$ and $\mathbf{b}_2$ are log-optimal (i.e., $W(\mathbf{b}_1, F) = W(\mathbf{b}_2, F) = W^*(F)$). By the concavity of $W(\mathbf{b}, F)$ in $\mathbf{b}$, we have

$$W(\lambda \mathbf{b}_1 + (1-\lambda)\mathbf{b}_2, F) \geq \lambda W(\mathbf{b}_1, F) + (1-\lambda) W(\mathbf{b}_2, F) = W^*(F). \tag{16.11}$$

Thus, $\lambda \mathbf{b}_1 + (1-\lambda)\mathbf{b}_2$ is also log-optimal. □

In the next section we use these properties to characterize the log-optimal portfolio.

16.2 KUHN–TUCKER CHARACTERIZATION OF THE LOG-OPTIMAL PORTFOLIO

Let $\mathcal{B} = \{\mathbf{b} \in \mathcal{R}^m : \mathbf{b}_i \geq 0, \sum_{i=1}^m \mathbf{b}_i = 1\}$ denote the *set of allowed portfolios*. The determination of $\mathbf{b}^*$ that achieves $W^*(F)$ is a problem of maximization of a concave function $W(\mathbf{b}, F)$ over a convex set $\mathcal{B}$. The maximum may lie on the boundary. We can use the standard Kuhn–Tucker conditions to characterize the maximum. Instead, we derive these conditions from first principles.

Theorem 16.2.1 *The log-optimal portfolio $\mathbf{b}^*$ for a stock market $\mathbf{X} \sim F$ (i.e., the portfolio that maximizes the growth rate $W(\mathbf{b}, F)$) satisfies the following necessary and sufficient conditions:*

$$E\left(\frac{X_i}{\mathbf{b}^{*\prime}\mathbf{X}}\right) \begin{cases} = 1 & \text{if } b_i^* > 0, \\ \leq 1 & \text{if } b_i^* = 0. \end{cases} \tag{16.12}$$

Proof: The growth rate $W(\mathbf{b}) = E(\ln \mathbf{b}'\mathbf{X})$ is concave in $\mathbf{b}$, where $\mathbf{b}$ ranges over the simplex of portfolios. It follows that $\mathbf{b}^*$ is log-optimum iff the directional derivative of $W(\cdot)$ in the direction from $\mathbf{b}^*$ to any alternative portfolio $\mathbf{b}$ is nonpositive. Thus, letting $\mathbf{b}_\lambda = (1-\lambda)\mathbf{b}^* + \lambda \mathbf{b}$ for $0 \leq \lambda \leq 1$, we have

$$\frac{d}{d\lambda} W(\mathbf{b}_\lambda)\Big|_{\lambda=0+} \leq 0, \qquad \mathbf{b} \in \mathcal{B}. \tag{16.13}$$

These conditions reduce to (16.12) since the one-sided derivative at $\lambda = 0+$ of $W(\mathbf{b}_\lambda)$ is

$$\begin{aligned}
&\frac{d}{d\lambda} E(\ln(\mathbf{b}_\lambda'\mathbf{X}))\Big|_{\lambda=0+} \\
&\quad = \lim_{\lambda \downarrow 0} \frac{1}{\lambda} E\left(\ln\left(\frac{(1-\lambda)\mathbf{b}^{*\prime}\mathbf{X} + \lambda \mathbf{b}'\mathbf{X}}{\mathbf{b}^{*\prime}\mathbf{X}}\right)\right) &(16.14)\\
&\quad = E\left(\lim_{\lambda \downarrow 0} \frac{1}{\lambda} \ln\left(1 + \lambda\left(\frac{\mathbf{b}'\mathbf{X}}{\mathbf{b}^{*\prime}\mathbf{X}} - 1\right)\right)\right) &(16.15)\\
&\quad = E\left(\frac{\mathbf{b}'\mathbf{X}}{\mathbf{b}^{*\prime}\mathbf{X}}\right) - 1, &(16.16)
\end{aligned}$$

where the interchange of limit and expectation can be justified using the dominated convergence theorem [39]. Thus, (16.13) reduces to

$$E\left(\frac{\mathbf{b}'\mathbf{X}}{\mathbf{b}^{*\prime}\mathbf{X}}\right) - 1 \leq 0 \tag{16.17}$$

for all $\mathbf{b} \in \mathcal{B}$. If the line segment from $\mathbf{b}$ to $\mathbf{b}^*$ can be extended beyond $\mathbf{b}^*$ in the simplex, the two-sided derivative at $\lambda = 0$ of $W(\mathbf{b}_\lambda)$ vanishes and (16.17) holds with equality. If the line segment from $\mathbf{b}$ to $\mathbf{b}^*$ cannot be extended because of the inequality constraint on $\mathbf{b}$, we have an inequality in (16.17).

The Kuhn–Tucker conditions will hold for all portfolios $\mathbf{b} \in \mathcal{B}$ if they hold for all extreme points of the simplex $\mathcal{B}$ since $E(\mathbf{b}^t\mathbf{X}/\mathbf{b}^{*t}\mathbf{X})$ is linear in $\mathbf{b}$. Furthermore, the line segment from the jth extreme point $(\mathbf{b} : b_j = 1, b_i = 0, i \neq j)$ to $\mathbf{b}^*$ can be extended beyond $\mathbf{b}^*$ in the simplex iff $b_j^* > 0$. Thus, the Kuhn–Tucker conditions that characterize the log-optimum $\mathbf{b}^*$ are equivalent to the following necessary and sufficient conditions:

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) \begin{cases} = 1 & \text{if } b_i^* > 0, \\ \le 1 & \text{if } b_i^* = 0. \quad \square \end{cases} \tag{16.18}$$

This theorem has a few immediate consequences. One useful equivalence is expressed in the following theorem.

Theorem 16.2.2 *Let $S^* = \mathbf{b}^{*t}\mathbf{X}$ be the random wealth resulting from the log-optimal portfolio $\mathbf{b}^*$. Let $S = \mathbf{b}^t\mathbf{X}$ be the wealth resulting from any other portfolio $\mathbf{b}$. Then*

$$E \ln \frac{S}{S^*} \le 0 \quad \text{for all } S \quad \Leftrightarrow \quad E\frac{S}{S^*} \le 1 \quad \text{for all } S. \tag{16.19}$$

Proof: From Theorem 16.2.1 it follows that for a log-optimal portfolio $\mathbf{b}^*$,

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) \le 1 \tag{16.20}$$

for all i. Multiplying this equation by b_i and summing over i, we have

$$\sum_{i=1}^{m} b_i E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) \le \sum_{i=1}^{m} b_i = 1, \tag{16.21}$$

which is equivalent to

$$E\frac{\mathbf{b}^t\mathbf{X}}{\mathbf{b}^{*t}\mathbf{X}} = E\frac{S}{S^*} \le 1. \tag{16.22}$$

The converse follows from Jensen's inequality, since

$$E \log \frac{S}{S^*} \le \log E\frac{S}{S^*} \le \log 1 = 0. \quad \square \tag{16.23}$$

Maximizing the expected logarithm was motivated by the asymptotic growth rate. But we have just shown that the log-optimal portfolio, in addition to maximizing the asymptotic growth rate, also "maximizes" the expected wealth relative $E(S/S^*)$ for one day. We shall say more about the short-term optimality of the log-optimal portfolio when we consider the game-theoretic optimality of this portfolio.

Another consequence of the Kuhn–Tucker characterization of the log-optimal portfolio is the fact that the expected proportion of wealth in each stock under the log-optimal portfolio is unchanged from day to day. Consider the stocks at the end of the first day. The initial allocation of wealth is $\mathbf{b}^*$. The proportion of the wealth in stock i at the end of the day is $\frac{b_i^* X_i}{\mathbf{b}^{*t}\mathbf{X}}$, and the expected value of this proportion is

$$E\frac{b_i^* X_i}{\mathbf{b}^{*t}\mathbf{X}} = b_i^* E\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}} = b_i^*. \tag{16.24}$$

Hence, the proportion of wealth in stock i expected at the end of the day is the same as the proportion invested in stock i at the beginning of the day. This is a counterpart to Kelly proportional gambling, where one invests in proportions that remain unchanged in expected value after the investment period.

16.3 ASYMPTOTIC OPTIMALITY OF THE LOG-OPTIMAL PORTFOLIO

In Section 16.2 we introduced the log-optimal portfolio and explained its motivation in terms of the long-term behavior of a sequence of investments in a repeated independent versions of the stock market. In this section we expand on this idea and prove that with probability 1, the conditionally log-optimal investor will not do any worse than any other investor who uses a causal investment strategy.

We first consider an i.i.d. stock market (i.e., $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. according to $F(\mathbf{x})$). Let

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i^t \mathbf{X}_i \tag{16.25}$$

be the wealth after n days for an investor who uses portfolio $\mathbf{b}_i$ on day i. Let

$$W^* = \max_{\mathbf{b}} W(\mathbf{b}, F) = \max_{\mathbf{b}} E \log \mathbf{b}^t\mathbf{X} \tag{16.26}$$

be the maximal growth rate, and let $\mathbf{b}^*$ be a portfolio that achieves the maximum growth rate. We only allow alternative portfolios $\mathbf{b}_i$ that depend causally on the past and are independent of the future values of the stock market.

Definition A *nonanticipating* or *causal* portfolio strategy is a sequence of mappings $b_i : \mathcal{R}^{m(i-1)} \to \mathcal{B}$, with the interpretation that portfolio $b_i(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$ is used on day i.

From the definition of W^*, it follows immediately that the log-optimal portfolio maximizes the expected log of the final wealth. This is stated in the following lemma.

Lemma 16.3.1 *Let S_n^* be the wealth after n days using the log-optimal strategy $\mathbf{b}^*$ on i.i.d. stocks, and let S_n be the wealth using a causal portfolio strategy $\mathbf{b}_i$. Then*

$$E \log S_n^* = nW^* \geq E \log S_n. \tag{16.27}$$

Proof

$$\max_{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n} E \log S_n = \max_{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n} E \sum_{i=1}^{n} \log \mathbf{b}_i^t \mathbf{X}_i \tag{16.28}$$

$$= \sum_{i=1}^{n} \max_{\mathbf{b}_i(\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_{i-1})} E \log \mathbf{b}_i^t(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})\mathbf{X}_i \tag{16.29}$$

$$= \sum_{i=1}^{n} E \log \mathbf{b}^{*t}\mathbf{X}_i \tag{16.30}$$

$$= nW^*, \tag{16.31}$$

and the maximum is achieved by a constant portfolio strategy $\mathbf{b}^*$. □

So far, we have proved two simple consequences of the definition of log-optimal portfolios: that $\mathbf{b}^*$ (satisfying (16.12)) maximizes the expected log wealth, and that the resulting wealth S_n^* is equal to 2^{nW^*} to first order in the exponent, with high probability.

Now we prove a much stronger result, which shows that S_n^* exceeds the wealth (to first order in the exponent) of any other investor for almost every sequence of outcomes from the stock market.

Theorem 16.3.1 (*Asymptotic optimality of the log-optimal portfolio*) *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a sequence of i.i.d. stock vectors drawn according to $F(\mathbf{x})$. Let $S_n^* = \prod_{i=1}^n \mathbf{b}^{*t}\mathbf{X}_i$, where $\mathbf{b}^*$ is the log-optimal portfolio, and let $S_n = \prod_{i=1}^n \mathbf{b}_i^t\mathbf{X}_i$ be the wealth resulting from any other causal portfolio. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 \qquad \textit{with probability 1.} \tag{16.32}$$

Proof: From the Kuhn–Tucker conditions and the log optimality of S_n^*, we have

$$E \frac{S_n}{S_n^*} \leq 1. \tag{16.33}$$

Hence by Markov's inequality, we have

$$\Pr\left(S_n > t_n S_n^*\right) = \Pr\left(\frac{S_n}{S_n^*} > t_n\right) < \frac{1}{t_n}. \tag{16.34}$$

Hence,

$$\Pr\left(\frac{1}{n} \log \frac{S_n}{S_n^*} > \frac{1}{n} \log t_n\right) \leq \frac{1}{t_n}. \tag{16.35}$$

Setting $t_n = n^2$ and summing over n, we have

$$\sum_{n=1}^{\infty} \Pr\left(\frac{1}{n} \log \frac{S_n}{S_n^*} > \frac{2\log n}{n}\right) \leq \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \tag{16.36}$$

Then, by the Borel–Cantelli lemma,

$$\Pr\left(\frac{1}{n} \log \frac{S_n}{S_n^*} > \frac{2\log n}{n}, \text{ infinitely often}\right) = 0. \tag{16.37}$$

This implies that for almost every sequence from the stock market, there exists an N such that for all $n > N$, $\frac{1}{n} \log \frac{S_n}{S_n^*} < \frac{2\log n}{n}$. Thus,

$$\limsup \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 \qquad \text{with probability 1.} \quad \square \tag{16.38}$$

The theorem proves that the log-optimal portfolio will perform as well as or better than any other portfolio to first order in the exponent.

16.4 SIDE INFORMATION AND THE GROWTH RATE

We showed in Chapter 6 that side information Y for the horse race X can be used to increase the growth rate by the mutual information $I(X; Y)$.

We now extend this result to the stock market. Here, $I(X; Y)$ is an upper bound on the increase in the growth rate, with equality if X is a horse race. We first consider the decrease in growth rate incurred by believing in the wrong distribution.

Theorem 16.4.1 *Let $\mathbf{X} \sim f(\mathbf{x})$. Let $\mathbf{b}_f$ be a log-optimal portfolio corresponding to $f(\mathbf{x})$, and let $\mathbf{b}_g$ be a log-optimal portfolio corresponding to some other density $g(\mathbf{x})$. Then the increase in growth rate ΔW by using $\mathbf{b}_f$ instead of $\mathbf{b}_g$ is bounded by*

$$\Delta W = W(\mathbf{b}_f, F) - W(\mathbf{b}_g, F) \le D(f||g). \tag{16.39}$$

Proof: We have

$$\Delta W = \int f(\mathbf{x}) \log \mathbf{b}_f^t \mathbf{x} - \int f(\mathbf{x}) \log \mathbf{b}_g^t \mathbf{x} \tag{16.40}$$

$$= \int f(\mathbf{x}) \log \frac{\mathbf{b}_f^t \mathbf{x}}{\mathbf{b}_g^t \mathbf{x}} \tag{16.41}$$

$$= \int f(\mathbf{x}) \log \frac{\mathbf{b}_f^t \mathbf{x}}{\mathbf{b}_g^t \mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})} \frac{f(\mathbf{x})}{g(\mathbf{x})} \tag{16.42}$$

$$= \int f(\mathbf{x}) \log \frac{\mathbf{b}_f^t \mathbf{x}}{\mathbf{b}_g^t \mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})} + D(f||g) \tag{16.43}$$

$$\stackrel{(a)}{\le} \log \int f(\mathbf{x}) \frac{\mathbf{b}_f^t \mathbf{x}}{\mathbf{b}_g^t \mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})} + D(f||g) \tag{16.44}$$

$$= \log \int g(\mathbf{x}) \frac{\mathbf{b}_f^t \mathbf{x}}{\mathbf{b}_g^t \mathbf{x}} + D(f||g) \tag{16.45}$$

$$\stackrel{(b)}{\le} \log 1 + D(f||g) \tag{16.46}$$

$$= D(f||g), \tag{16.47}$$

where (a) follows from Jensen's inequality and (b) follows from the Kuhn–Tucker conditions and the fact that $\mathbf{b}_g$ is log-optimal for g. □

Theorem 16.4.2 *The increase ΔW in growth rate due to side information Y is bounded by*

$$\Delta W \le I(\mathbf{X}; Y). \tag{16.48}$$

Proof: Let $(\mathbf{X}, Y) \sim f(\mathbf{x}, y)$, where $\mathbf{X}$ is the market vector and Y is the related side information. Given side information $Y = y$, the log-optimal investor uses the conditional log-optimal portfolio for the conditional distribution $f(\mathbf{x}|Y = y)$. Hence, conditional on $Y = y$, we have, from Theorem 16.4.1,

$$\Delta W_{Y=y} \le D(f(\mathbf{x}|Y = y)||f(\mathbf{x})) = \int_{\mathbf{x}} f(\mathbf{x}|Y = y) \log \frac{f(\mathbf{x}|Y = y)}{f(\mathbf{x})} d\mathbf{x}. \tag{16.49}$$

Averaging this over possible values of Y, we have

$$\Delta W \le \int_y f(y) \int_{\mathbf{x}} f(\mathbf{x}|Y = y) \log \frac{f(\mathbf{x}|Y = y)}{f(\mathbf{x})} d\mathbf{x}\, dy \tag{16.50}$$

$$= \int_y \int_{\mathbf{x}} f(y) f(\mathbf{x}|Y = y) \log \frac{f(\mathbf{x}|Y = y)}{f(\mathbf{x})} \frac{f(y)}{f(y)} d\mathbf{x}\, dy \tag{16.51}$$

$$= \int_y \int_{\mathbf{x}} f(\mathbf{x}, y) \log \frac{f(\mathbf{x}, y)}{f(\mathbf{x}) f(y)} d\mathbf{x}\, dy \tag{16.52}$$

$$= I(\mathbf{X}; Y). \tag{16.53}$$

Hence, the increase in growth rate is bounded above by the mutual information between the side information Y and the stock market $\mathbf{X}$. □

16.5 INVESTMENT IN STATIONARY MARKETS

We now extend some of the results of Section 16.4 from i.i.d. markets to time-dependent market processes. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$ be a vector-valued stochastic process with $\mathbf{X}_i \ge 0$. We consider investment strategies that depend on the past values of the market in a causal fashion (i.e., $\mathbf{b}_i$ may depend on $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}$). Let

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i^t(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) \mathbf{X}_i. \tag{16.54}$$

Our objective is to maximize $E \log S_n$ over all such causal portfolio strategies $\{\mathbf{b}_i(\cdot)\}$. Now

$$\max_{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n} E \log S_n = \sum_{i=1}^{n} \max_{\mathbf{b}_i(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})} E \log \mathbf{b}_i^t \mathbf{X}_i \tag{16.55}$$

$$= \sum_{i=1}^{n} E \log \mathbf{b}_i^{*t} \mathbf{X}_i, \tag{16.56}$$

where $\mathbf{b}_i^*$ is the log-optimal portfolio for the conditional distribution of $\mathbf{X}_i$ given the past values of the stock market; that is, $\mathbf{b}_i^*(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1})$ is the portfolio that achieves the conditional maximum, which is denoted by

$$\max_{\mathbf{b}} E[\log \mathbf{b}^t \mathbf{X}_i | (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1})]$$
$$= W^*(\mathbf{X}_i | \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}). \quad (16.57)$$

Taking the expectation over the past, we write

$$W^*(\mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) = E \max_{\mathbf{b}} E\left[\log \mathbf{b}^t \mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}\right] \quad (16.58)$$

as the conditional optimal growth rate, where the maximum is over all portfolio-valued functions $\mathbf{b}$ defined on $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$. Thus, the highest expected log return is achieved by using the conditional log-optimal portfolio at each stage. Let

$$W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = \max_{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n} E \log S_n, \quad (16.59)$$

where the maximum is over all causal portfolio strategies. Then since $\log S_n^* = \sum_{i=1}^m \log \mathbf{b}_i^{*t} \mathbf{X}_i$, we have the following chain rule for W^*:

$$W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = \sum_{i=1}^n W^*(\mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}). \quad (16.60)$$

This chain rule is formally the same as the chain rule for H. In some ways, W is the dual of H. In particular, conditioning reduces H but increases W. We now define the counterpart of the entropy rate for time-dependent stochastic processes.

Definition The *growth rate* W_∞^* is defined as

$$W_\infty^* = \lim_{n \to \infty} \frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} \quad (16.61)$$

if the limit exists.

Theorem 16.5.1 *For a stationary market, the growth rate exists and is equal to*

$$W_\infty^* = \lim_{n \to \infty} W^*(\mathbf{X}_n | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}). \quad (16.62)$$

Proof: By stationarity, $W^*(\mathbf{X}_n | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1})$ is nondecreasing in n. Hence, it must have a limit, possibly infinity. Since

$$\frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} = \frac{1}{n} \sum_{i=1}^n W^*(\mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}), \quad (16.63)$$

it follows by the theorem of the Cesáro mean (Theorem 4.2.3) that the left-hand side has the same limit as the limit of the terms on the right-hand side. Hence, W_∞^* exists and

$$W_\infty^* = \lim_{n \to \infty} \frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} = \lim_{n \to \infty} W^*(\mathbf{X}_n | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}). \quad \square \quad (16.64)$$

We can now extend the asymptotic optimality property to stationary markets. We have the following theorem.

Theorem 16.5.2 *Consider an arbitrary stochastic process $\{X_i\}$, $X_i \in \mathcal{R}_+^m$, conditionally log-optimal portfolios, $\mathbf{b}_i^*(X^{i-1})$ and wealth S_n^*. Let S_n be the wealth generated by any other causal portfolio strategy $\mathbf{b}_i(X^{i-1})$. Then S_n/S_n^* is a positive supermartingale with respect to the sequence of σ-fields generated by the past $X_1, X_2, \ldots, X_n$. Consequently, there exists a random variable V such that*

$$\frac{S_n}{S_n^*} \to V \quad \text{with probability 1} \quad (16.65)$$

$$EV \le 1 \quad (16.66)$$

and

$$\Pr\left\{\sup_n \frac{S_n}{S_n^*} \ge t\right\} \le \frac{1}{t}. \quad (16.67)$$

Proof: S_n/S_n^* is a positive supermartingale because

$$E\left[\frac{S_{n+1}(X^{n+1})}{S_{n+1}^*(X^{n+1})} \middle| X^n\right] = E\left[\frac{(\mathbf{b}_{n+1}^t \mathbf{X}_{n+1}) S_n(X^n)}{(\mathbf{b}_{n+1}^{*t} \mathbf{X}_{n+1}) S_n^*(X^n)} \middle| X^n\right] \quad (16.68)$$

$$= \frac{S_n(X^n)}{S_n^*(X^n)} E\left[\frac{\mathbf{b}_{n+1}^t \mathbf{X}_{n+1}}{\mathbf{b}_{n+1}^{*t} \mathbf{X}_{n+1}} \middle| X^n\right] \quad (16.69)$$

$$\le \frac{S_n(X^n)}{S_n^*(X^n)}, \quad (16.70)$$

by the Kuhn–Tucker condition on the conditionally log-optimal portfolio. Thus, by the martingale convergence theorem, S_n/S_n^* has a limit, call it V, and $EV \le E(S_0/S_0^*) = 1$. Finally, the result for $\sup(S_n/S_n^*)$ follows from Kolmogorov's inequality for positive martingales. □

We remark that (16.70) shows how strong the competitive optimality of S_n^* is. Apparently, the probability is less than 1/10 that $S_n(X^n)$ will ever be 10 times as large as $S_n^*(X^n)$. For a stationary ergodic market, we can extend the asymptotic equipartition property to prove the following theorem.

Theorem 16.5.3 *(AEP for the stock market)* *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a stationary ergodic vector-valued stochastic process. Let S_n^* be the wealth at time n for the conditionally log-optimal strategy, where*

$$S_n^* = \prod_{i=1}^{n} \mathbf{b}_i^{*t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})\mathbf{X}_i. \tag{16.71}$$

Then

$$\frac{1}{n}\log S_n^* \to W_\infty^* \quad \text{with probability } 1. \tag{16.72}$$

Proof: The proof involves a generalization of the sandwich argument [20] used to prove the AEP in Section 16.8. The details of the proof (in Algoet and Cover [21]) are omitted. □

Finally, we consider the example of the horse race once again. The horse race is a special case of the stock market in which there are m stocks corresponding to the m horses in the race. At the end of the race, the value of the stock for horse i is either 0 or o_i, the value of the odds for horse i. Thus, $\mathbf{X}$ is nonzero only in the component corresponding to the winning horse.

In this case, the log-optimal portfolio is proportional betting, known as *Kelly gambling* (i.e., $b_i^* = p_i$), and in the case of uniform fair odds (i.e., $o_i = m$, for all i),

$$W^* = \log m - H(X). \tag{16.73}$$

When we have a sequence of correlated horse races, the optimal portfolio is conditional proportional betting and the asymptotic growth rate is

$$W_\infty^* = \log m - H(\mathcal{X}), \tag{16.74}$$

where $H(\mathcal{X}) = \lim \frac{1}{n} H(X_1, X_2, \ldots, X_n)$ if the limit exists. Then Theorem 16.5.3 asserts that

$$S_n^* \doteq 2^{nW^*}, \tag{16.75}$$

in agreement with the results in chapter 6.

16.6 COMPETITIVE OPTIMALITY OF THE LOG-OPTIMAL PORTFOLIO

We now ask whether the log-optimal portfolio outperforms alternative portfolios at a given finite time n. As a direct consequence of the Kuhn–Tucker conditions, we have

$$E\frac{S_n}{S_n^*} \le 1, \tag{16.76}$$

and hence by Markov's inequality,

$$\Pr(S_n > tS_n^*) \le \frac{1}{t}. \tag{16.77}$$

This result is similar to the result derived in Chapter 5 for the competitive optimality of Shannon codes.

By considering examples, it can be seen that it is not possible to get a better bound on the probability that $S_n > S_n^*$. Consider a stock market with two stocks and two possible outcomes,

$$(X_1, X_2) = \begin{cases} \left(1, \dfrac{1}{1-\epsilon}\right) & \text{with probability } 1-\epsilon, \\ (1, 0) & \text{with probability } \epsilon. \end{cases} \tag{16.78}$$

In this market the log-optimal portfolio invests all the wealth in the first stock. [It is easy to verify that $\mathbf{b} = (1, 0)$ satisfies the Kuhn–Tucker conditions.] However, an investor who puts all his wealth in the second stock earns more money with probability $1-\epsilon$. Hence, it is not true that with high probability the log-optimal investor will do better than any other investor.

The problem with trying to prove that the log-optimal investor does best with a probability of at least $\frac{1}{2}$ is that there exist examples like the one above, where it is possible to beat the log-optimal investor by a small amount most of the time. We can get around this by allowing each investor an additional fair randomization, which has the effect of reducing the effect of small differences in the wealth.

Theorem 16.6.1 *(Competitive optimality) Let S^* be the wealth at the end of one period of investment in a stock market* $\mathbf{X}$ *with the log-optimal portfolio, and let S be the wealth induced by any other portfolio. Let U^* be a random variable independent of* $\mathbf{X}$ *uniformly distributed on $[0, 2]$, and let V be any other random variable independent of* $\mathbf{X}$ *and U^* with $V \geq 0$ and $EV = 1$. Then*

$$\Pr(VS \geq U^*S^*) \leq \frac{1}{2}. \tag{16.79}$$

Remark Here U^* and V correspond to initial "fair" randomizations of the initial wealth. This exchange of initial wealth $S_0 = 1$ for "fair" wealth U^* can be achieved in practice by placing a fair bet. The effect of the fair randomization is to randomize small differences, so that only the significant deviations of the ratio S/S^* affect the probability of winning.

Proof: We have

$$\Pr(VS \geq U^*S^*) = \Pr\left(\frac{VS}{S^*} \geq U^*\right) \tag{16.80}$$

$$= \Pr(W \geq U^*), \tag{16.81}$$

where $W = \frac{VS}{S^*}$ is a non-negative-valued random variable with mean

$$EW = E(V)E\left(\frac{S_n}{S_n^*}\right) \leq 1 \tag{16.82}$$

by the independence of V from $\mathbf{X}$ and the Kuhn–Tucker conditions. Let F be the distribution function of W. Then since U^* is uniform on $[0, 2]$,

$$\Pr(W \geq U^*) = \int_0^2 \Pr(W > w) f_{U^*}(w)\, dw \tag{16.83}$$

$$= \int_0^2 \Pr(W > w) \frac{1}{2}\, dw \tag{16.84}$$

$$= \int_0^2 \frac{1 - F(w)}{2}\, dw \tag{16.85}$$

$$\leq \int_0^\infty \frac{1 - F(w)}{2}\, dw \tag{16.86}$$

$$= \frac{1}{2} EW \tag{16.87}$$

$$\leq \frac{1}{2}, \tag{16.88}$$

using the easily proved fact (by integrating by parts) that

$$EW = \int_0^\infty (1 - F(w))\, dw \tag{16.89}$$

for a positive random variable W. Hence, we have

$$\Pr(VS \geq U^*S^*) = \Pr(W \geq U^*) \leq \frac{1}{2}. \quad \square \tag{16.90}$$

Theorem 16.6.1 provides a short-term justification for the use of the log-optimal portfolio. If the investor's only objective is to be ahead of his opponent at the end of the day in the stock market, and if fair randomization is allowed, Theorem 16.6.1 says that the investor should exchange his wealth for a uniform $[0, 2]$ wealth and then invest using the log-optimal portfolio. This is the game-theoretic solution to the problem of gambling competitively in the stock market.

16.7 UNIVERSAL PORTFOLIOS

The development of the log-optimal portfolio strategy in Section 16.1 relies on the assumption that we know the distribution of the stock vectors and can therefore calculate the optimal portfolio $\mathbf{b}^*$. In practice, though, we often do not know the distribution. In this section we describe a causal portfolio that performs well on individual sequences. Thus, we make no statistical assumptions about the market sequence. We assume that the stock market can be represented by a sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots \in \mathcal{R}_+^m$, where x_{ij} is the price relative for stock j on day i and $\mathbf{x}_i$ is the vector of price relatives for all stocks on day i. We begin with a finite-horizon problem, where we have n vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. We later extend the results to the infinite-horizon case.

Given this sequence of stock market outcomes, what is the best we can do? A realistic target is the growth achieved by the best constant rebalanced portfolio strategy in hindsight (i.e., the best constant rebalanced portfolio on the known sequence of stock market vectors). Note that constant rebalanced portfolios are optimal against i.i.d. stock market sequences with known distribution, so that this set of portfolios is reasonably natural.

Let us assume that we have a number of mutual funds, each of which follows a constant rebalanced portfolio strategy chosen in advance. Our objective is to perform as well as the best of these funds. In this section we show that we can do almost as well as the best constant rebalanced portfolio without advance knowledge of the distribution of the stock market vectors.

One approach is to distribute the wealth among a continuum of fund managers, each of which follows a different constantly rebalanced portfolio strategy. Since one of the managers will do exponentially better than the others, the total wealth after n days will be dominated by the largest term. We will show that we can achieve a performance of the best fund manager within a factor of $n^{\frac{m-1}{2}}$. This is the essence of the argument for the infinite-horizon universal portfolio strategy.

A second approach to this problem is as a game against a malicious opponent or nature who is allowed to choose the sequence of stock market vectors. We define a causal (nonanticipating) portfolio strategy $\hat{\mathbf{b}}_i(\mathbf{x}_{i-1}, \ldots, \mathbf{x}_1)$ that depends only on the past values of the stock market sequence. Then nature, with knowledge of the strategy $\hat{\mathbf{b}}_i(\mathbf{x}^{i-1})$, chooses a sequence of vectors $\mathbf{x}_i$ to make the strategy perform as poorly as possible relative to the best constantly rebalanced portfolio for that stock sequence. Let $\mathbf{b}^*(\mathbf{x}^n)$ be the best constantly rebalanced portfolio for a stock market sequence $\mathbf{x}^n$. Note that $\mathbf{b}^*(\mathbf{x}^n)$ depends only on the empirical distribution of the sequence, not on the order in which the vectors occur. At the end of n days, a constantly rebalanced portfolio $\mathbf{b}$ achieves wealth:

$$S_n(\mathbf{b}, \mathbf{x}^n) = \prod_{i=1}^{n} \mathbf{b}^t \mathbf{x}_i, \tag{16.91}$$

and the best constant portfolio $\mathbf{b}^*(\mathbf{x}^n)$ achieves a wealth

$$S_n^*(\mathbf{x}^n) = \max_{\mathbf{b}} \prod_{i=1}^{n} \mathbf{b}^t \mathbf{x}_i, \tag{16.92}$$

whereas the nonanticipating portfolio $\hat{\mathbf{b}}_i(\mathbf{x}^{i-1})$ strategy achieves

$$\hat{S}_n(\mathbf{x}^n) = \prod_{i=1}^{n} \hat{\mathbf{b}}_i^t(\mathbf{x}^{i-1})\mathbf{x}_i. \tag{16.93}$$

Our objective is to find a nonanticipating portfolio strategy $\hat{\mathbf{b}}(\cdot) = (\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2(\mathbf{x}_1), \ldots, \hat{\mathbf{b}}_i(\mathbf{x}^{i-1}))$ that does well in the worst case in terms of the ratio of $\hat{S}_n$ to S_n^*. We will find the optimal universal strategy and show that this strategy for each stock sequence achieves wealth $\hat{S}_n$ that is within a factor $V_n \approx n^{-\frac{m-1}{2}}$ of the wealth S_n^* achieved by the best constantly rebalanced portfolio on that sequence. This strategy depends on n, the horizon of the game. Later we describe some horizon-free results that have the same worst-case asymptotic performance as that of the finite-horizon game.

16.7.1 Finite-Horizon Universal Portfolios

We begin by analyzing a stock market of n periods, where n is known in advance, and attempt to find a portfolio strategy that does well against all possible sequences of n stock market vectors. The main result can be stated in the following theorem.

Theorem 16.7.1 *For a stock market sequence $\mathbf{x}^n = \mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathcal{R}_+^m$ of length n with m assets, let $S_n^*(\mathbf{x}^n)$ be the wealth achieved by the optimal constantly rebalanced portfolio on $\mathbf{x}^n$, and let $\hat{S}_n(\mathbf{x}^n)$ be the wealth achieved by any causal portfolio strategy $\hat{\mathbf{b}}_i(\cdot)$ on $\mathbf{x}^n$; then*

$$\max_{\hat{\mathbf{b}}_i(\cdot)} \min_{\mathbf{x}_1, \ldots, \mathbf{x}_n} \frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} = V_n, \tag{16.94}$$

where

$$V_n = \left[\sum_{n_1 + \cdots + n_m = n} \binom{n}{n_1, n_2, \ldots, n_m} 2^{-nH\left(\frac{n_1}{n}, \ldots, \frac{n_m}{n}\right)} \right]^{-1}. \tag{16.95}$$

Using Stirling's approximation, we can show that V_n is on the order of $n^{-\frac{m-1}{2}}$, and therefore the growth rate for the universal portfolio on the worst sequence differs from the growth rate of the best constantly rebalanced portfolio on that sequence by at most a polynomial factor. The logarithm of the ratio of growth of wealth of the universal portfolio $\hat{\mathbf{b}}$ to the growth of wealth of the best constant portfolio behaves like the redundancy of a universal source code. (See Shtarkov [496], where $\log V_n$ appears as the minimax individual sequence redundancy in data compression.)

We first illustrate the main results by means of an example for $n = 1$. Consider the case of two stocks and a single day. Let the stock vector for the day be $\mathbf{x} = (x_1, x_2)$. If $x_1 > x_2$, the best portfolio is one that puts all its money on stock 1, and if $x_2 > x_1$, the best portfolio puts all its money on stock 2. (If $x_1 = x_2$, all portfolios are equivalent.)

Now assume that we must choose a portfolio in advance and our opponent can choose the stock market sequence after we have chosen our portfolio to make us do as badly as possible relative to the best portfolio. Given our portfolio, the opponent can ensure that we do as badly as possible by making the stock on which we have put more weight equal to 0 and the other stock equal to 1. Our best strategy is therefore to put equal weight on both stocks, and with this, we will achieve a growth factor at least equal to half the growth factor of the best stock, and hence we will achieve at least half the gain of the best constantly rebalanced portfolio. It is not hard to calculate that $V_n = 2$ when $n = 1$ and $m = 2$ in equation (16.94).

However, this result seems misleading, since it appears to suggest that for n days, we would use a constant uniform portfolio, putting half our money on each stock every day. If our opponent then chose the stock sequence so that only the first stock was 1 (and the other was 0) every day, this uniform strategy would achieve a wealth of $1/2^n$, and we would achieve a wealth only within a factor of 2^n of the best constant portfolio, which puts all the money on the first stock for all time.

The result of the theorem shows that we can do significantly better. The main part of the argument is to reduce a sequence of stock vectors to the extreme cases where only one of the stocks is nonzero for each day. If we can ensure that we do well on such sequences, we can guarantee that we do well on any sequence of stock vectors, and achieve the bounds of the theorem.

Before we prove the theorem, we need the following lemma.

Lemma 16.7.1 *For $p_1, p_2, \ldots, p_m \ge 0$ and $q_1, q_2, \ldots, q_m \ge 0$,*

$$\frac{\sum_{i=1}^m p_i}{\sum_{i=1}^m q_i} \ge \min_i \frac{p_i}{q_i}. \tag{16.96}$$

Proof: Let I denote the index i that minimizes the right-hand side in (16.96). Assume that $p_I > 0$ (if $p_I = 0$, the lemma is trivially true). Also, if $q_I = 0$, both sides of (16.96) are infinite (all the other q_i's must also be zero), and again the inequality holds. Therefore, we can also assume that $q_I > 0$. Then

$$\frac{\sum_{i=1}^m p_i}{\sum_{i=1}^m q_i} = \frac{p_I}{q_I}\frac{1 + \sum_{i \ne I}(p_i/p_I)}{1 + \sum_{i \ne I}(q_i/q_I)} \ge \frac{p_I}{q_I} \tag{16.97}$$

because

$$\frac{p_i}{q_i} \ge \frac{p_I}{q_I} \longrightarrow \frac{p_i}{p_I} \ge \frac{q_i}{q_I} \tag{16.98}$$

for all i. □

First consider the case when $n = 1$. The wealth at the end of the first day is

$$\hat{S}_1(\mathbf{x}) = \hat{\mathbf{b}}'\mathbf{x}, \tag{16.99}$$

$$S_1(\mathbf{x}) = \mathbf{b}'\mathbf{x} \tag{16.100}$$

and

$$\frac{\hat{S}_1(\mathbf{x})}{S_1(\mathbf{x})} = \frac{\sum \hat{b}_i x_i}{\sum b_i x_i} \ge \min\left\{\frac{\hat{b}_i}{b_i}\right\}. \tag{16.101}$$

We wish to find $\max_{\hat{\mathbf{b}}} \min_{\mathbf{b},\mathbf{x}} \frac{\hat{\mathbf{b}}'\mathbf{x}}{\mathbf{b}'\mathbf{x}}$. Nature should choose $\mathbf{x} = \mathbf{e}_i$, where $\mathbf{e}_i$ is the ith basis vector with 1 in the component i that minimizes $\frac{\hat{b}_i}{b_i^*}$, and the investor should choose $\hat{\mathbf{b}}$ to maximize this minimum. This is achieved by choosing $\hat{\mathbf{b}} = (\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m})$.

The important point to realize is that

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n(\mathbf{x}^n)} = \frac{\prod_{i=1}^n \hat{\mathbf{b}}_i'\mathbf{x}_i}{\prod_{i=1}^n \mathbf{b}_i'\mathbf{x}_i} \tag{16.102}$$

can also be rewritten in the form of a ratio of terms

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n(\mathbf{x}^n)} = \frac{\hat{\mathbf{b}}'\mathbf{x}'}{\mathbf{b}'\mathbf{x}'}, \tag{16.103}$$

where $\hat{\mathbf{b}}, \mathbf{b}, \mathbf{x}' \in \mathcal{R}_+^{m^n}$. Here the m^n components of the constantly rebalanced portfolios $\mathbf{b}$ are all of the product form $b_1^{n_1} b_2^{n_2} \cdots b_m^{n_m}$. One wishes to find a universal $\hat{\mathbf{b}}$ that is uniformly close to the $\mathbf{b}$'s corresponding to constantly rebalanced portfolios.

We can now prove the main theorem (Theorem 16.7.1).

Proof of Theorem 16.7.1: We will prove the theorem for $m = 2$. The proof extends in a straightforward fashion to the case $m > 2$. Denote the stocks by 1 and 2. The key idea is to express the wealth at time n,

$$S_n(\mathbf{x}^n) = \prod_{i=1}^n \mathbf{b}_i'\mathbf{x}_i, \tag{16.104}$$

which is a product of sums, into a sum of products. Each term in the sum corresponds to a sequence of stock price relatives for stock 1 or stock 2 times the proportion b_{i1} or b_{i2} that the strategy places on stock 1 or stock 2 at time i. We can therefore view the wealth S_n as a sum over all 2^n possible n-sequences of 1's and 2's of the product of the portfolio proportions times the stock price relatives:

$$S_n(\mathbf{x}^n) = \sum_{j^n \in \{1,2\}^n} \prod_{i=1}^n b_{ij_i} x_{ij_i} = \sum_{j^n \in \{1,2\}^n} \prod_{i=1}^n b_{ij_i} \prod_{i=1}^n x_{ij_i}. \tag{16.105}$$

If we let $w(j^n)$ denote the product $\prod_{i=1}^n b_{ij_i}$, the total fraction of wealth invested in the sequence j^n, and let

$$x(j^n) = \prod_{i=1}^{n} x_{ij_i} \tag{16.106}$$

be the corresponding return for this sequence, we can write

$$S_n(\mathbf{x}^n) = \sum_{j^n \in \{1,2\}^n} w(j^n)x(j^n). \tag{16.107}$$

Similar expressions apply to both the best constantly rebalanced portfolio and the universal portfolio strategy. Thus, we have

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} = \frac{\sum_{j^n \in \{1,2\}^n} \hat{w}(j^n)x(j^n)}{\sum_{j^n \in \{1,2\}^n} w^*(j^n)x(j^n)}, \tag{16.108}$$

where $\hat{w}^n$ is the amount of wealth placed on the sequence j^n by the universal nonanticipating strategy, and $w^*(j^n)$ is the amount placed by the best constant rebalanced portfolio strategy. Now applying Lemma 16.7.1, we have

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \ge \min_{j^n} \frac{\hat{w}(j^n)x(j^n)}{w^*(j^n)x(j^n)} = \min_{j^n} \frac{\hat{w}(j^n)}{w^*(j^n)}. \tag{16.109}$$

Thus, the problem of maximizing the performance ratio $\hat{S}_n/S_n^*$ is reduced to ensuring that the proportion of money bet on a sequence of stocks by the universal portfolio is uniformly close to the proportion bet by $\mathbf{b}^*$. As might be obvious by now, this formulation of S_n reduces the n-period stock market to a special case of a single-period stock market—there are 2^n stocks, one invests $w(j^n)$ in stock j^n and receives a return $x(j^n)$ for stock j^n, and the total wealth S_n is $\sum_{j^n} w(j^n)x(j^n)$.

We first calculate the weight $w^*(j^n)$ associated with the best constant rebalanced portfolio $\mathbf{b}^*$. We observe that a constantly rebalanced portfolio $\mathbf{b}$ results in

$$w(j^n) = \prod_{i=1}^{n} b_{ij_i} = b^k(1-b)^{n-k}, \tag{16.110}$$

where k is the number of times 1 appears in the sequence j^n. Thus, $w(j^n)$ depends only on k, the number of 1's in j^n. Fixing attention on j^n, we find by differentiating with respect to b that the maximum value

$$w^*(j^n) = \max_{0\le b\le 1} b^k(1-b)^{n-k} \tag{16.111}$$

$$= \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}, \tag{16.112}$$

which is achieved by

$$\mathbf{b}^* = \left(\frac{k}{n}, \frac{n-k}{n}\right). \tag{16.113}$$

Note that $\sum w^*(j^n) > 1$, reflecting the fact that the amount "bet" on j^n is chosen in hindsight, thus relieving the hindsight investor of the responsibility of allocating his investments $w^*(j^n)$ to sum to 1. The causal investor has no such luxury. How can the causal investor choose initial investments $\hat{w}(j^n)$, $\sum \hat{w}(j^n) = 1$, to protect himself from all possible j^n and hindsight-determined $w^*(j^n)$? The answer will be to choose $\hat{w}(j^n)$ proportional to $w^*(j^n)$. Then the worst-case ratio of $\hat{w}(j^n)/w^*(j^n)$ will be maximized. To proceed, we define V_n by

$$\frac{1}{V_n} = \sum_{j^n} \left(\frac{k(j^n)}{n}\right)^{k(j^n)} \left(\frac{n-k(j^n)}{n}\right)^{n-k(j^n)} \tag{16.114}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} \tag{16.115}$$

and let

$$\hat{w}(j^n) = V_n \left(\frac{k(j^n)}{n}\right)^{k(j^n)} \left(\frac{n-k(j^n)}{n}\right)^{n-k(j^n)}. \tag{16.116}$$

It is clear that $\hat{w}(j^n)$ is a legitimate distribution of wealth over the 2^n stock sequences (i.e., $\hat{w}(j^n) \ge 0$ and $\sum_{j^n} \hat{w}(j^n) = 1$). Here V_n is the normalization factor that makes $\hat{w}(j^n)$ a probability mass function. Also, from (16.109) and (16.113), for all sequences $\mathbf{x}^n$,

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \ge \min_{j^n} \frac{\hat{w}(j^n)}{w^*(j^n)} \tag{16.117}$$

$$= \min_k \frac{V_n(\frac{k}{n})^k(\frac{n-k}{n})^{n-k}}{b^{*k}(1-b^*)^{n-k}} \tag{16.118}$$

$$\ge V_n, \tag{16.119}$$

where (16.117) follows from (16.109) and (16.119) follows from (16.112). Consequently, we have

$$\max_{\hat{\mathbf{b}}} \min_{\mathbf{x}^n} \frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq V_n. \tag{16.120}$$

We have thus demonstrated a portfolio on the 2^n possible sequences of length n that achieves wealth $\hat{S}_n(\mathbf{x}^n)$ within a factor V_n of the wealth $S_n^*(\mathbf{x}^n)$ achieved by the best constant rebalanced portfolio in hindsight. To complete the proof of the theorem, we show that this is the best possible, that is, that any nonanticipating portfolio $\mathbf{b}_i(\mathbf{x}^{i-1})$ cannot do better than a factor V_n in the worst case (i.e., for the worst choice of $\mathbf{x}^n$). To prove this, we construct a set of extremal stock market sequences and show that the performance of any nonanticipating portfolio strategy is bounded by V_n for at least one of these sequences, proving the worst-case bound.

For each $j^n \in \{1, 2\}^n$, we define the corresponding extremal stock market vector $\mathbf{x}^n(j^n)$ as

$$\mathbf{x}_i(j_i) = \begin{cases} (1, 0)^t & \text{if } j_i = 1, \\ (0, 1)^t & \text{if } j_i = 2, \end{cases} \tag{16.121}$$

Let $\mathbf{e}_1 = (1, 0)^t, \mathbf{e}_2 = (0, 1)^t$ be standard basis vectors. Let

$$\mathcal{K} = \{\mathbf{x}(j^n) : j^n \in \{1, 2\}^n, \mathbf{x}_{ij_i} = \mathbf{e}_{j_i}\} \tag{16.122}$$

be the set of extremal sequences. There are 2^n such extremal sequences, and for each sequence at each time, there is only one stock that yields a nonzero return. The wealth invested in the other stock is lost. Therefore, the wealth at the end of n periods for extremal sequence $\mathbf{x}^n(j^n)$ is the product of the amounts invested in the stocks $j_1, j_2, \ldots, j_n$, [i.e., $S_n(\mathbf{x}^n(j^n)) = \prod_i b_{j_i} = w(j^n)$]. Again, we can view this as an investment on sequences of length n, and given the 0–1 nature of the return, it is easy to see for $\mathbf{x}^n \in \mathcal{K}$ that

$$\sum_{j^n} S_n(\mathbf{x}^n(j^n)) = 1. \tag{16.123}$$

For any extremal sequence $\mathbf{x}^n(j^n) \in \mathcal{K}$, the best constant rebalanced portfolio is

$$\mathbf{b}^*(\mathbf{x}^n(j^n)) = \left(\frac{n_1(j^n)}{n}, \frac{n_2(j^n)}{n}\right)^t, \tag{16.124}$$

where $n_1(j^n)$ is the number of occurrences of 1 in the sequence j^n. The corresponding wealth at the end of n periods is

$$S_n^*(\mathbf{x}^n(j^n)) = \left(\frac{n_1(j^n)}{n}\right)^{n_1(j^n)} \left(\frac{n_2(j^n)}{n}\right)^{n_2(j^n)} = \frac{\hat{w}(j^n)}{V_n}, \tag{16.125}$$

from (16.116) and it therefore follows that

$$\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n) = \frac{1}{V_n} \sum_{j^n} \hat{w}(j^n) = \frac{1}{V_n}. \tag{16.126}$$

We then have the following inequality for any portfolio sequence $\{\mathbf{b}_i\}_{i=1}^n$, with $S_n(\mathbf{x}^n)$ defined as in (16.104):

$$\min_{\mathbf{x}^n \in \mathcal{K}} \frac{S_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \leq \sum_{\tilde{\mathbf{x}}^n \in \mathcal{K}} \frac{S_n^*(\tilde{\mathbf{x}}^n)}{\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n)} \frac{S_n(\tilde{\mathbf{x}}^n)}{S_n^*(\tilde{\mathbf{x}}^n)} \tag{16.127}$$

$$= \sum_{\tilde{\mathbf{x}}^n \in \mathcal{K}} \frac{S_n(\tilde{\mathbf{x}}^n)}{\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n)} \tag{16.128}$$

$$= \frac{1}{\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n)} \tag{16.129}$$

$$= V_n, \tag{16.130}$$

where the inequality follows from the fact that the minimum is less than the average. Thus,

$$\max_{\mathbf{b}} \min_{\mathbf{x}^n \in \mathcal{K}} \frac{S_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \leq V_n. \quad \square \tag{16.131}$$

The strategy described in the theorem puts mass on all sequences of length n and is clearly dependent on n. We can recast the strategy in incremental terms (i.e., in terms of the amount bet on stock 1 and stock 2 at time 1), then, conditional on the outcome at time 1, the amount bet on each of the two stocks at time 2, and so on. Consider the weight $\hat{b}_{i,1}$ assigned by the algorithm to stock 1 at time i given the previous sequence of stock vectors $\mathbf{x}^{i-1}$. We can calculate this by summing over all sequences j^n that have a 1 in position i, giving

$$\hat{\mathbf{b}}_{i,1}(\mathbf{x}^{i-1}) = \frac{\sum_{j^{i-1} \in M^{i-1}} \hat{w}(j^{i-1}1)x(j^{i-1})}{\sum_{j^i \in M^i} \hat{w}(j^i)x(j^{i-1})}, \tag{16.132}$$

where

$$\hat{w}(j^i) = \sum_{j^n : j^i \subseteq j^n} w(j^n) \tag{16.133}$$

is the weight put on all sequences j^n that start with j^i, and

$$x(j^{i-1}) = \prod_{k=1}^{i-1} x_{kj_k} \tag{16.134}$$

is the return on those sequences as defined in (16.106).

Investigation of the asymptotics of V_n reveals [401, 496] that

$$V_n \sim \left(\sqrt{\frac{2}{n}}\right)^{m-1} \Gamma(m/2)/\sqrt{\pi} \tag{16.135}$$

for m assets. In particular, for $m = 2$ assets,

$$V_n \sim \sqrt{\frac{2}{\pi n}} \tag{16.136}$$

and

$$\frac{1}{2\sqrt{n+1}} \leq V_n \leq \frac{2}{\sqrt{n+1}} \tag{16.137}$$

for all n [400]. Consequently, for $m = 2$ stocks, the causal portfolio strategy $\hat{\mathbf{b}}_i(\mathbf{x}^{i-1})$ given in (16.132) achieves wealth $\hat{S}_n(x^n)$ such that

$$\frac{\hat{S}_n(x^n)}{S_n^*(x^n)} \geq V_n \geq \frac{1}{2\sqrt{n+1}} \tag{16.138}$$

for all market sequences x^n.

16.7.2 Horizon-Free Universal Portfolios

We describe the horizon-free strategy in terms of a weighting of different portfolio strategies. As described earlier, each constantly rebalanced portfolio $\mathbf{b}$ can be viewed as corresponding to a mutual fund that rebalances the m assets according to $\mathbf{b}$. Initially, we distribute the wealth among these funds according to a distribution $\mu(\mathbf{b})$, where $d\mu(\mathbf{b})$ is the amount of wealth invested in portfolios in the neighborhood $d\mathbf{b}$ of the constantly rebalanced portfolio $\mathbf{b}$.

Let

$$S_n(\mathbf{b}, \mathbf{x}^n) = \prod_{i=1}^{n} \mathbf{b}^t \mathbf{x}_i \tag{16.139}$$

be the wealth generated by a constant rebalanced portfolio $\mathbf{b}$ on the stock sequence $\mathbf{x}^n$. Recall that

$$S_n^*(\mathbf{x}^n) = \max_{b \in \mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n) \tag{16.140}$$

is the wealth of the best constant rebalanced portfolio in hindsight.

We investigate the causal portfolio defined by

$$\hat{\mathbf{b}}_{i+1}(\mathbf{x}^i) = \frac{\int_{\mathcal{B}} \mathbf{b} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}. \tag{16.141}$$

We note that

$$\hat{\mathbf{b}}_{i+1}^t(\mathbf{x}^i)\mathbf{x}_{i+1} = \frac{\int_{\mathcal{B}} \mathbf{b}^t \mathbf{x}_{i+1} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})} \tag{16.142}$$

$$= \frac{\int_{\mathcal{B}} S_{i+1}(\mathbf{b}, \mathbf{x}^{i+1})\, d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}. \tag{16.143}$$

Thus, the product $\prod \hat{\mathbf{b}}_i^t \mathbf{x}_i$ telescopes and we see that the wealth $\hat{S}_n(\mathbf{x}^n)$ resulting from this portfolio is given by

$$\hat{S}_n(\mathbf{x}^n) = \prod_{i=1}^{n} \hat{\mathbf{b}}_i^t(\mathbf{x}^{i-1})\mathbf{x}_i \tag{16.144}$$

$$= \int_{b \in \mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n)\, d\mu(\mathbf{b}). \tag{16.145}$$

There is another way to interpret (16.145). The amount given to portfolio manager $\mathbf{b}$ is $d\mu(\mathbf{b})$, the resulting growth factor for the manager rebalancing to $\mathbf{b}$ is $S(\mathbf{b}, \mathbf{x}^n)$, and the total wealth of this batch of investments is

$$\hat{S}_n(\mathbf{x}^n) = \int_{\mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n)\, d\mu(\mathbf{b}). \tag{16.146}$$

Then $\hat{\mathbf{b}}_{i+1}$, defined in (16.141), is the performance-weighted total "buy order" of the individual portfolio manager $\mathbf{b}$.

So far, we have not specified what distribution $\mu(\mathbf{b})$ we use to apportion the initial wealth. We now use a distribution μ that puts mass on all possible portfolios, so that we approximate the performance of the best portfolio for the actual distribution of stock price vectors.

In the next lemma, we bound $\hat{S}_n/S_n^*$ as a function of the initial wealth distribution $\mu(\mathbf{b})$.

Lemma 16.7.2 *Let $S_n^*(\mathbf{x}^n)$ in 16.140 be the wealth achieved by the best constant rebalanced portfolio and let $\hat{S}_n(\mathbf{x}^n)$ in (16.144) be the wealth achieved by the universal mixed portfolio $\hat{\mathbf{b}}(\cdot)$, given by*

$$\hat{\mathbf{b}}_{i+1}(\mathbf{x}^i) = \frac{\int \mathbf{b} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}{\int S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}. \tag{16.147}$$

Then

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq \min_{j^n} \frac{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i}\, d\mu(\mathbf{b})}{\prod_{i=1}^n b_{j_i}^*}. \tag{16.148}$$

Proof: As before, we can write

$$S_n^*(\mathbf{x}^n) = \sum_{j^n} w^*(j^n) x(j^n), \tag{16.149}$$

where $w^*(j^n) = \prod_{i=1}^n b_{j_i}^*$ is the amount invested on the sequence j^n and $x(j^n) = \prod_{i=1}^n x_{ij_i}$ is the corresponding return. Similarly, we can write

$$\hat{S}_n(\mathbf{x}^n) = \int \prod_{i=1}^n \mathbf{b}^t \mathbf{x}_i\, d\mu(\mathbf{b}) \tag{16.150}$$

$$= \sum_{j^n} \int \prod_{i=1}^n b_{j_i} x_{ij_i}\, d\mu(\mathbf{b}) \tag{16.151}$$

$$= \sum_{j^n} \hat{w}(j^n) x(j^n), \tag{16.152}$$

where $\hat{w}(j^n) = \int \prod_{i=1}^n b_{j_i}\, d\mu(\mathbf{b})$. Now applying Lemma 16.7.1, we have

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} = \frac{\sum_{j^n} \hat{w}(j^n) x(j^n)}{\sum_{j^n} w^*(j^n) x(j^n)} \tag{16.153}$$

$$\geq \min_{j^n} \frac{\hat{w}(j^n) x(j^n)}{w^*(j^n) x(j^n)} \tag{16.154}$$

$$= \min_{j^n} \frac{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i}\, d\mu(\mathbf{b})}{\prod_{i=1}^n b_{j_i}^*}. \quad \square \tag{16.155}$$

We now apply this lemma when $\mu(\mathbf{b})$ is the Dirichlet($\frac{1}{2}$) distribution.

Theorem 16.7.2 *For the causal universal portfolio $\hat{b}_i(\,)$, $i = 1, 2, \ldots$, given in (16.141), with $m = 2$ stocks and $d\mu(\mathbf{b})$ the Dirichlet($\frac{1}{2}$, $\frac{1}{2}$) distribution, we have*

$$\frac{\hat{S}_n(x^n)}{S_n^*(x^n)} \geq \frac{1}{2\sqrt{n+1}},$$

for all n and all stock sequences x^n.

Proof: As in the discussion preceding (16.112), we can show that the weight put by the best constant portfolio b^* on the sequence j^n is

$$\prod_{i=1}^n b_{j_i}^* = \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} = 2^{-nH(k/n)}, \tag{16.156}$$

where k is the number of indices where $j_i = 1$. We can also explicitly calculate the integral in the numerator of (16.148) in Lemma 16.7.2 for the Dirichlet($\frac{1}{2}$) density, defined for m variables as

$$d\mu(\mathbf{b}) = \frac{\Gamma(\frac{m}{2})}{\left[\Gamma\left(\frac{1}{2}\right)\right]^m} \prod_{j=1}^m b_j^{-\frac{1}{2}}\, d\mathbf{b}, \tag{16.157}$$

where $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\, dt$ denotes the gamma function. For simplicity, we consider the case of two stocks, in which case

$$d\mu(b) = \frac{1}{\pi} \frac{1}{\sqrt{b(1-b)}}\, db, \qquad 0 \leq b \leq 1, \tag{16.158}$$

where b is the fraction of wealth invested in stock 1. Now consider any sequence $j^n \in \{1, 2\}^n$, and consider the amount invested in that sequence,

$$b(j^n) = \prod_{i=1}^n b_{j_i} = b^l (1-b)^{n-l}, \tag{16.159}$$

where l is the number of indices where $j_i = 1$. Then

$$\int b(j^n)\, d\mu(\mathbf{b}) = \int b^l (1-b)^{n-l} \frac{1}{\pi} \frac{1}{\sqrt{b(1-b)}}\, db \tag{16.160}$$

$$= \frac{1}{\pi} \int b^{l-\frac{1}{2}}(1-b)^{n-l-\frac{1}{2}}\, db \tag{16.161}$$

$$\triangleq \frac{1}{\pi} B\left(l+\frac{1}{2}, n-l+\frac{1}{2}\right), \tag{16.162}$$

where $B(\lambda_1, \lambda_2)$ is the beta function, defined as

$$B(\lambda_1, \lambda_2) = \int_0^1 x^{\lambda_1 - 1}(1-x)^{\lambda_2 - 1}\, dx \tag{16.163}$$

$$= \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1 + \lambda_2)} \tag{16.164}$$

and

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x}\, dx. \tag{16.165}$$

Note that for any integer n, $\Gamma(n+1) = n!$ and $\Gamma(n+\frac{1}{2}) = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n}\sqrt{\pi}$.

We can calculate $B(l+\frac{1}{2}, n-l+\frac{1}{2})$ by means of simple recursion using integration by parts. Alternatively, using (16.164), we obtain

$$B\left(l+\frac{1}{2}, n-l+\frac{1}{2}\right) = \frac{\pi}{2^{2n}} \frac{\binom{2n}{n}\binom{n}{l}}{\binom{2n}{2l}}. \tag{16.166}$$

Combining all the results with Lemma 16.7.2, we have

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq \min_{j^n} \frac{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i}\, d\mu(\mathbf{b})}{\prod_{i=1}^n b_{j_i}^*} \tag{16.167}$$

$$\geq \min_l \frac{\frac{1}{\pi} B(l+\frac{1}{2}, n-l+\frac{1}{2})}{2^{-nH(l/n)}} \tag{16.168}$$

$$\geq \frac{1}{2\sqrt{n+1}}, \tag{16.169}$$

using the results in [135, Theorem 2]. □

It follows for $m = 2$ stocks that

$$\frac{\hat{S}_n}{S_n^*} \geq \frac{1}{\sqrt{2\pi}} V_n \tag{16.170}$$

for all n and all market sequences $x_1, x_2, \ldots, x_n$. Thus, good minimax performance for all n costs at most an extra factor $\sqrt{2\pi}$ over the fixed horizon minimax portfolio. The cost of universality is V_n, which is asymptotically negligible in the growth rate in the sense that

$$\frac{1}{n}\ln \hat{S}_n(\mathbf{x}^n) - \frac{1}{n}\ln S_n^*(\mathbf{x}^n) \geq \frac{1}{n}\ln \frac{V_n}{\sqrt{2\pi}} \to 0. \tag{16.171}$$

Thus, the universal causal portfolio achieves the same asymptotic growth rate of wealth as the best hindsight portfolio.

Let's now consider how this portfolio algorithm performs on two real stocks. We consider a 14-year period (ending in 2004) and two stocks, Hewlett-Packard and Altria (formerly, Phillip Morris), which are both components of the Dow Jones Index. Over these 14 years, HP went up by a factor of 11.8, while Altria went up by a factor of 11.5. The performance of the different constantly rebalanced portfolios that contain HP and Altria are shown in Figure 16.2. The best constantly rebalanced portfolio (which can be computed only in hindsight) achieves a growth of a factor of 18.7 using a mixture of about 51% HP and 49% Altria. The universal portfolio strategy described in this section achieves a growth factor of 15.7 without foreknowledge.

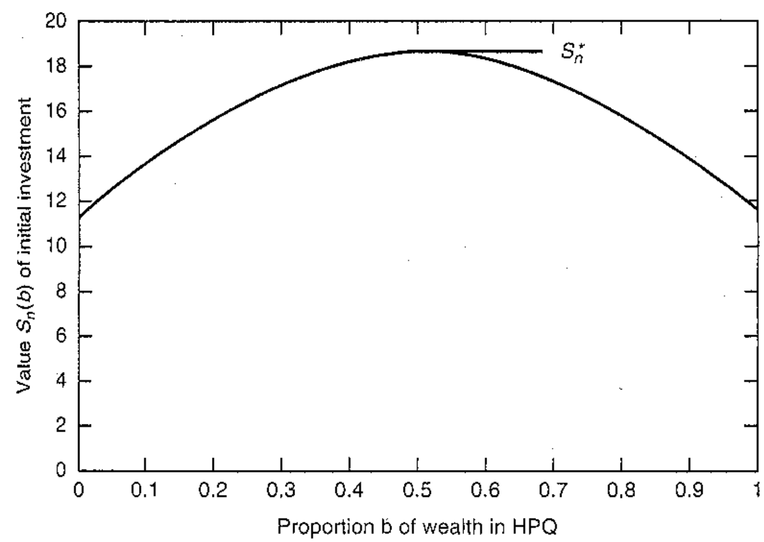

FIGURE 16.2. Performance of different constant rebalanced portfolios $\mathbf{b}$ for HP and Altria.

16.8 SHANNON–MCMILLAN–BREIMAN THEOREM (GENERAL AEP)

The AEP for ergodic processes has come to be known as the *Shannon–McMillan–Breiman theorem*. In Chapter 3 we proved the AEP for i.i.d. processes. In this section we offer a proof of the theorem for a general ergodic process. We prove the convergence of $\frac{1}{n}\log p(X^n)$ by sandwiching it between two ergodic sequences.

In a sense, an ergodic process is the most general dependent process for which the strong law of large numbers holds. For finite alphabet processes, ergodicity is equivalent to the convergence of the kth-order empirical distributions to their marginals for all k.

The technical definition requires some ideas from probability theory. To be precise, an ergodic source is defined on a probability space $(\Omega, \mathcal{B}, P)$, where $\mathcal{B}$ is a σ-algebra of subsets of Ω and P is a probability measure. A random variable X is defined as a function $X(\omega)$, $\omega \in \Omega$, on the probability space. We also have a transformation $T : \Omega \to \Omega$, which plays the role of a time shift. We will say that the transformation is *stationary* if $P(TA) = P(A)$ for all $A \in \mathcal{B}$. The transformation is called *ergodic* if every set A such that $TA = A$, a.e., satisfies $P(A) = 0$ or 1. If T is stationary and ergodic, we say that the process defined by $X_n(\omega) = X(T^n\omega)$ is stationary and ergodic. For a stationary ergodic source, Birkhoff's ergodic theorem states that

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \to EX = \int X\,dP \quad \text{with probability 1.} \tag{16.172}$$

Thus, the law of large numbers holds for ergodic processes.

We wish to use the ergodic theorem to conclude that

$$\begin{aligned} -\frac{1}{n}\log p(X_0, X_1, \ldots, X_{n-1}) &= -\frac{1}{n}\sum_{i=0}^{n-1}\log p(X_i|X_0^{i-1}) \\ &\to \lim_{n\to\infty} E[-\log p(X_n|X_0^{n-1})]. \end{aligned} \tag{16.173}$$

But the stochastic sequence $p(X_i|X_0^{i-1})$ is not ergodic. However, the closely related quantities $p(X_i|X_{i-k}^{i-1})$ and $p(X_i|X_{-\infty}^{i-1})$ are ergodic and have expectations easily identified as entropy rates. We plan to sandwich $p(X_i|X_0^{i-1})$ between these two more tractable processes.

We define the kth-order entropy H^k as

$$H^k = E\{-\log p(X_k|X_{k-1}, X_{k-2}, \ldots, X_0)\} \tag{16.174}$$

$$= E\{-\log p(X_0|X_{-1}, X_{-2}, \ldots, X_{-k})\}, \tag{16.175}$$

where the last equation follows from stationarity. Recall that the entropy rate is given by

$$H = \lim_{k\to\infty} H^k \tag{16.176}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} H^k. \tag{16.177}$$

Of course, $H^k \searrow H$ by stationarity and the fact that conditioning does not increase entropy. It will be crucial that $H^k \searrow H = H^\infty$, where

$$H^\infty = E\{-\log p(X_0|X_{-1}, X_{-2}, \ldots)\}. \tag{16.178}$$

The proof that $H^\infty = H$ involves exchanging expectation and limit.

The main idea in the proof goes back to the idea of (conditional) proportional gambling. A gambler receiving uniform odds with the knowledge of the k past will have a growth rate of wealth $\log|\mathcal{X}| - H^k$, while a gambler with a knowledge of the infinite past will have a growth rate of wealth of $\log|\mathcal{X}| - H^\infty$. We don't know the wealth growth rate of a gambler with growing knowledge of the past X_0^n, but it is certainly sandwiched between $\log|\mathcal{X}| - H^k$ and $\log|\mathcal{X}| - H^\infty$. But $H^k \searrow H = H^\infty$. Thus, the sandwich closes and the growth rate must be $\log|\mathcal{X}| - H$.

We will prove the theorem based on lemmas that will follow the proof.

Theorem 16.8.1 (*AEP: Shannon–McMillan–Breiman Theorem*) *If H is the entropy rate of a finite-valued stationary ergodic process $\{X_n\}$, then*

$$-\frac{1}{n}\log p(X_0, \ldots, X_{n-1}) \to H \quad \textit{with probability 1.} \tag{16.179}$$

Proof: We prove this for finite alphabet $\mathcal{X}$; this proof and the proof for countable alphabets and densities is given in Algoet and Cover [20]. We argue that the sequence of random variables $-\frac{1}{n}\log p(X_0^{n-1})$ is asymptotically sandwiched between the upper bound H^k and the lower bound H^∞ for all $k \geq 0$. The AEP will follow since $H^k \to H^\infty$ and $H^\infty = H$. The kth-order Markov approximation to the probability is defined for $n \geq k$ as

$$p^k(X_0^{n-1}) = p(X_0^{k-1})\prod_{i=k}^{n-1} p(X_i|X_{i-k}^{i-1}). \tag{16.180}$$

From Lemma 16.8.3 we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \le 0, \tag{16.181}$$

which we rewrite, taking the existence of the limit $\frac{1}{n}\log p^k(X_0^n)$ into account (Lemma 16.8.1), as

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{p(X_0^{n-1})} \le \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{p^k(X_0^{n-1})} = H^k \tag{16.182}$$

for $k = 1, 2, \ldots$. Also, from Lemma 16.8.3, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})} \le 0, \tag{16.183}$$

which we rewrite as

$$\liminf \frac{1}{n} \log \frac{1}{p(X_0^{n-1})} \ge \lim \frac{1}{n} \log \frac{1}{p(X_0^{n-1}|X_{-\infty}^{-1})} = H^\infty \tag{16.184}$$

from the definition of H^∞ in Lemma 16.8.1.

Putting together (16.182) and (16.184), we have

$$\begin{aligned} H^\infty &\le \liminf -\frac{1}{n}\log p(X_0^{n-1}) \le \limsup -\frac{1}{n}\log p(X_0^{n-1}) \\ &\le H^k \quad \text{for all } k. \end{aligned} \tag{16.185}$$

But by Lemma 16.8.2, $H^k \to H^\infty = H$. Consequently,

$$\lim -\frac{1}{n}\log p(X_0^n) = H. \quad \square \tag{16.186}$$

We now prove the lemmas that were used in the main proof. The first lemma uses the ergodic theorem.

Lemma 16.8.1 (*Markov approximations*) *For a stationary ergodic stochastic process* $\{X_n\}$,

$$-\frac{1}{n}\log p^k(X_0^{n-1}) \to H^k \quad \textit{with probability 1}, \tag{16.187}$$

$$-\frac{1}{n}\log p(X_0^{n-1}|X_{-\infty}^{-1}) \to H^\infty \quad \textit{with probability 1}. \tag{16.188}$$

Proof: Functions $Y_n = f(X_{-\infty}^n)$ of ergodic processes $\{X_i\}$ are ergodic processes. Thus, $\log p(X_n|X_{n-k}^{n-1})$ and $\log p(X_n|X_{n-1}, X_{n-2}, \ldots,)$ are also ergodic processes, and

$$-\frac{1}{n}\log p^k(X_0^{n-1}) = -\frac{1}{n}\log p(X_0^{k-1}) - \frac{1}{n}\sum_{i=k}^{n-1}\log p(X_i|X_{i-k}^{i-1}) \tag{16.189}$$

$$\to 0 + H^k \quad \text{with probability 1}, \tag{16.190}$$

by the ergodic theorem. Similarly, by the ergodic theorem,

$$-\frac{1}{n}\log p(X_0^{n-1}|X_{-1}, X_{-2}, \ldots) = -\frac{1}{n}\sum_{i=0}^{n-1}\log p(X_i|X_{i-1}, X_{i-2}, \ldots) \tag{16.191}$$

$$\to H^\infty \quad \text{with probability 1.} \quad \square \tag{16.192}$$

Lemma 16.8.2 (*No gap*) $H^k \searrow H^\infty$ *and* $H = H^\infty$.

Proof: We know that for stationary processes, $H^k \searrow H$, so it remains to show that $H^k \searrow H^\infty$, thus yielding $H = H^\infty$. Levy's martingale convergence theorem for conditional probabilities asserts that

$$p(x_0|X_{-k}^{-1}) \to p(x_0|X_{-\infty}^{-1}) \quad \text{with probability 1} \tag{16.193}$$

for all $x_0 \in \mathcal{X}$. Since $\mathcal{X}$ is finite and $p\log p$ is bounded and continuous in p for all $0 \le p \le 1$, the bounded convergence theorem allows interchange of expectation and limit, yielding

$$\lim_{k\to\infty} H^k = \lim_{k\to\infty} E\left\{-\sum_{x_0\in\mathcal{X}} p(x_0|X_{-k}^{-1})\log p(x_0|X_{-k}^{-1})\right\} \tag{16.194}$$

$$= E\left\{-\sum_{x_0\in\mathcal{X}} p(x_0|X_{-\infty}^{-1})\log p(x_0|X_{-\infty}^{-1})\right\} \tag{16.195}$$

$$= H^\infty. \tag{16.196}$$

Thus, $H^k \searrow H = H^\infty$. $\square$

Lemma 16.8.3 (*Sandwich*)

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \le 0, \tag{16.197}$$

$$\limsup \frac{1}{n} \log \frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})} \le 0. \tag{16.198}$$

Proof: Let A be the support set of $p(X_0^{n-1})$. Then

$$E\left\{\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})}\right\} = \sum_{x_0^{n-1}\in A} p(x_0^{n-1}) \frac{p^k(x_0^{n-1})}{p(x_0^{n-1})} \tag{16.199}$$

$$= \sum_{x_0^{n-1}\in A} p^k(x_0^{n-1}) \tag{16.200}$$

$$= p^k(A) \tag{16.201}$$

$$\le 1. \tag{16.202}$$

Similarly, let $B(X_{-\infty}^{-1})$ denote the support set of $p(\cdot|X_{-\infty}^{-1})$. Then we have

$$E\left\{\frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})}\right\} = E\left[E\left\{\frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})}\middle| X_{-\infty}^{-1}\right\}\right] \tag{16.203}$$

$$= E\left[\sum_{x^n\in B(X_{-\infty}^{-1})} \frac{p(x^n)}{p(x^n|X_{-\infty}^{-1})} p(x^n|X_{-\infty}^{-1})\right] \tag{16.204}$$

$$= E\left[\sum_{x^n\in B(X_{-\infty}^{-1})} p(x^n)\right] \tag{16.205}$$

$$\le 1. \tag{16.206}$$

By Markov's inequality and (16.202), we have

$$\Pr\left\{\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \ge t_n\right\} \le \frac{1}{t_n} \tag{16.207}$$

or

$$\Pr\left\{\frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \ge \frac{1}{n}\log t_n\right\} \le \frac{1}{t_n}. \tag{16.208}$$

Letting $t_n = n^2$ and noting that $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$, we see by the Borel–Cantelli lemma that the event

$$\left\{\frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \ge \frac{1}{n}\log t_n\right\} \tag{16.209}$$

occurs only finitely often with probability 1. Thus,

$$\limsup \frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \le 0 \qquad \text{with probability 1.} \tag{16.210}$$

Applying the same arguments using Markov's inequality to (16.206), we obtain

$$\limsup \frac{1}{n}\log\frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})} \le 0 \qquad \text{with probability 1,} \tag{16.211}$$

proving the lemma. □

The arguments used in the proof can be extended to prove the AEP for the stock market (Theorem 16.5.3).

SUMMARY

Growth rate. The *growth rate* of a stock market portfolio $\mathbf{b}$ with respect to a distribution $F(\mathbf{x})$ is defined as

$$W(\mathbf{b}, F) = \int \log \mathbf{b}^t\mathbf{x}\, dF(\mathbf{x}) = E\left(\log \mathbf{b}^t\mathbf{x}\right). \tag{16.212}$$

Log-optimal portfolio. The *optimal growth rate* with respect to a distribution $F(x)$ is

$$W^*(F) = \max_{\mathbf{b}} W(\mathbf{b}, F). \tag{16.213}$$

The portfolio $\mathbf{b}^*$ that achieves the maximum of $W(\mathbf{b}, F)$ is called the *log-optimal portfolio*.

Concavity. $W(\mathbf{b}, F)$ is concave in $\mathbf{b}$ and linear in F. $W^*(F)$ is convex in F.

Optimality conditions. The portfolio $\mathbf{b}^*$ is log-optimal if and only if

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) = 1 \quad \text{if } b_i^* > 0, \qquad \le 1 \quad \text{if } b_i^* = 0. \tag{16.214}$$

Expected ratio optimality. If $S_n^* = \prod_{i=1}^n \mathbf{b}^{*t}\mathbf{X}_i$, $S_n = \prod_{i=1}^n \mathbf{b}_i^t\mathbf{X}_i$, then

$$E\frac{S_n}{S_n^*} \le 1 \quad \text{if and only if} \quad E \ln \frac{S_n}{S_n^*} \le 0. \tag{16.215}$$

Growth rate (AEP)

$$\frac{1}{n}\log S_n^* \to W^*(F) \qquad \text{with probability 1.} \tag{16.216}$$

Asymptotic optimality

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \le 0 \qquad \text{with probability 1.} \tag{16.217}$$

Wrong information. Believing g when f is true loses

$$\Delta W = W(\mathbf{b}_f^*, F) - W(\mathbf{b}_g^*, F) \le D(f||g). \tag{16.218}$$

Side information Y

$$\Delta W \le I(\mathbf{X}; Y). \tag{16.219}$$

Chain rule

$$W^*(\mathbf{X}_i|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) = \max_{\mathbf{b}_i(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1})} E \log \mathbf{b}_i^t\mathbf{X}_i \tag{16.220}$$

$$W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = \sum_{i=1}^n W^*(\mathbf{X}_i|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}). \tag{16.221}$$

Growth rate for a stationary market.

$$W_\infty^* = \lim \frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} \tag{16.222}$$

$$\frac{1}{n}\log S_n^* \to W_\infty^*. \tag{16.223}$$

Competitive optimality of log-optimal portfolios.

$$\Pr(VS \ge U^*S^*) \le \frac{1}{2}. \tag{16.224}$$

Universal portfolio.

$$\max_{\hat{\mathbf{b}}_i(\cdot)} \min_{\mathbf{x}^n, \mathbf{b}} \frac{\prod_{i=1}^n \hat{\mathbf{b}}_i^t(\mathbf{x}^{i-1})\mathbf{x}_i}{\prod_{i=1}^n \mathbf{b}^t\mathbf{x}_i} = V_n, \tag{16.225}$$

where

$$V_n = \left[\sum_{n_1+\cdots+n_m=n} \binom{n}{n_1, n_2, \ldots, n_m} 2^{-nH(n_1/n, \ldots, n_m/n)}\right]^{-1}. \tag{16.226}$$

For $m = 2$,

$$V_n \sim \sqrt{2/\pi n} \tag{16.227}$$

The causal universal portfolio

$$\hat{\mathbf{b}}_{i+1}(\mathbf{x}^i) = \frac{\int \mathbf{b} S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})}{\int S_i(\mathbf{b}, \mathbf{x}^i)\, d\mu(\mathbf{b})} \tag{16.228}$$

achieves

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \ge \frac{1}{2\sqrt{n+1}} \tag{16.229}$$

for all n and all $\mathbf{x}^n$.

AEP. If $\{X_i\}$ is stationary ergodic, then

$$-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H(\mathcal{X}) \qquad \text{with probability 1.} \tag{16.230}$$

PROBLEMS

16.1 *Growth rate.* Let

$$\mathbf{X} = \begin{cases} (1, a) & \text{with probability } \frac{1}{2} \\ (1, 1/a) & \text{with probability } \frac{1}{2} \end{cases}$$

where $a > 1$. This vector $\mathbf{X}$ represents a stock market vector of cash vs. a hot stock. Let

$$W(\mathbf{b}, F) = E \log \mathbf{b}^t\mathbf{X}$$

and

$$W^* = \max_{\mathbf{b}} W(\mathbf{b}, F)$$

be the growth rate.

(a) Find the log optimal portfolio $\mathbf{b}^*$.

(b) Find the growth rate W^*.

(c) Find the asymptotic behavior of

$$S_n = \prod_{i=1}^{n} \mathbf{b}^t\mathbf{X}_i$$

for all $\mathbf{b}$.

16.2 *Side information.* Suppose, in Problem 16.1, that

$$\mathbf{Y} = \begin{cases} 1 & \text{if } (X_1, X_2) \geq (1, 1), \\ 0 & \text{if } (X_1, X_2) \leq (1, 1). \end{cases}$$

Let the portfolio $\mathbf{b}$ depend on $\mathbf{Y}$. Find the new growth rate W^{**} and verify that $\Delta W = W^{**} - W^*$ satisfies

$$\Delta W \leq I(X; Y).$$

16.3 *Stock dominance.* Consider a stock market vector

$$\mathbf{X} = (X_1, X_2).$$

Suppose that $X_1 = 2$ with probability 1. Thus an investment in the first stock is doubled at the end of the day.

(a) Find necessary and sufficient conditions on the distribution of stock X_2 such that the log-optimal portfolio $\mathbf{b}^*$ invests all the wealth in stock X_2 [i.e., $\mathbf{b}^* = (0, 1)$].

(b) Argue for any distribution on X_2 that the growth rate satisfies $W^* \geq 1$.

16.4 *Including experts and mutual funds.* Let $\mathbf{X} \sim F(\mathbf{x})$, $\mathbf{x} \in \mathcal{R}_+^m$, be the vector of price relatives for a stock market. Suppose that an "expert" suggests a portfolio $\mathbf{b}$. This would result in a wealth factor $\mathbf{b}^t\mathbf{X}$. We add this to the stock alternatives to form $\tilde{\mathbf{X}} = (X_1, X_2, \ldots, X_m, \mathbf{b}^t\mathbf{X})$. Show that the new growth rate,

$$\tilde{W}^* = \max_{b_1,\ldots,b_m,b_{m+1}} \int \ln(\mathbf{b}^t\tilde{\mathbf{x}})\, dF(\tilde{\mathbf{x}}), \tag{16.231}$$

is equal to the old growth rate,

$$W^* = \max_{b_1,\ldots,b_m} \int \ln(\mathbf{b}^t\mathbf{x})\, dF(\mathbf{x}). \tag{16.232}$$

16.5 *Growth rate for symmetric distribution.* Consider a stock vector $\mathbf{X} \sim F(\mathbf{x})$, $\mathbf{X} \in \mathcal{R}^m$, $\mathbf{X} \geq 0$, where the component stocks are exchangeable. Thus, $F(x_1, x_2, \ldots, x_m) = F(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(m)})$ for all permutations σ.

(a) Find the portfolio $\mathbf{b}^*$ optimizing the growth rate and establish its optimality. Now assume that $\mathbf{X}$ has been normalized so that $\frac{1}{m}\sum_{i=1}^m X_i = 1$, and F is symmetric as before.

(b) Again assuming $\mathbf{X}$ to be normalized, show that all symmetric distributions F have the same growth rate against $\mathbf{b}^*$.

(c) Find this growth rate.

16.6 *Convexity.* We are interested in the set of stock market densities that yield the same optimal porfolio. Let $P_{\mathbf{b}_0}$ be the set of all probability densities on $\mathcal{R}_+^m$ for which $\mathbf{b}_0$ is optimal. Thus, $P_{\mathbf{b}_0} = \{p(x) : \int \ln(\mathbf{b}^t x) p(x)\, dx$ is maximized by $\mathbf{b} = \mathbf{b}_0\}$. Show that $P_{\mathbf{b}_0}$ is a convex set. It may be helpful to use Theorem 16.2.2.

16.7 *Short selling.* Let

$$X = \begin{cases} (1, 2), & p, \\ (1, \frac{1}{2}), & 1 - p. \end{cases}$$

Let $B = \{(b_1, b_2) : b_1 + b_2 = 1\}$. Thus, this set of portfolios B does not include the constraint $b_i \geq 0$. (This allows short selling.)

(a) Find the log optimal portfolio $\mathbf{b}^*(p)$.

(b) Relate the growth rate $W^*(p)$ to the entropy rate $H(p)$.

16.8 *Normalizing* $\mathbf{x}$. Suppose that we define the log-optimal portfolio $\mathbf{b}^*$ to be the portfolio maximizing the relative growth rate

$$\int \ln \frac{\mathbf{b}^t\mathbf{x}}{\frac{1}{m}\sum_{i=1}^m x_i} dF(x_1, \ldots, x_m).$$

The virtue of the normalization $\frac{1}{m}\sum X_i$, which can be viewed as the wealth associated with a uniform portfolio, is that the relative growth rate is finite even when the growth rate $\int \ln b^t x dF(x)$ is not. This matters, for example, if X has a St. Petersburg-like distribution. Thus, the log-optimal portfolio $\mathbf{b}^*$ is defined for all distributions F, even those with infinite growth rates $W^*(F)$.

(a) Show that if $\mathbf{b}$ maximizes $\int \ln(\mathbf{b}^t\mathbf{x})\, dF(x)$, it also maximizes $\int \ln \frac{\mathbf{b}^t\mathbf{x}}{u^t x}\, dF(x)$, where $u = (\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m})$.

(b) Find the log optimal portfolio $\mathbf{b}^*$ for

$$\mathbf{X} = \begin{cases} (2^{2^k+1}, 2^{2^k}), & 2^{-(k+1)}, \\ (2^{2^k}, 2^{2^k+1}), & 2^{-(k+1)}, \end{cases}$$

where $k = 1, 2, \ldots$.

(c) Find EX and W^*.

(d) Argue that $\mathbf{b}^*$ is competitively better than any portfolio $\mathbf{b}$ in the sense that $\Pr\{\mathbf{b}^t\mathbf{X} > c\mathbf{b}^{*t}\mathbf{X}\} \le \frac{1}{c}$.

16.9 *Universal portfolio*. We examine the first $n = 2$ steps of the implementation of the universal portfolio in (16.7.2) for $\mu(b)$ uniform for $m = 2$ stocks. Let the stock vectors for days 1 and 2 be $\mathbf{x}_1 = (1, \frac{1}{2})$, and $\mathbf{x}_2 = (1, 2)$. Let $\mathbf{b} = (b, 1-b)$ denote a portfolio.

(a) Graph $S_2(\mathbf{b}) = \prod_{i=1}^2 \mathbf{b}^t\mathbf{x}_i, \quad 0 \le b \le 1$.

(b) Calculate $S_2^* = \max_{\mathbf{b}} S_2(\mathbf{b})$.

(c) Argue that $\log S_2(\mathbf{b})$ is concave in $\mathbf{b}$.

(d) Calculate the (universal) wealth $\hat{S}_2 = \int_0^1 S_2(\mathbf{b}) d\mathbf{b}$.

(e) Calculate the universal portfolio at times $n = 1$ and $n = 2$:

$$\hat{\mathbf{b}}_1 = \int_0^1 \mathbf{b}\, d\mathbf{b}$$

$$\hat{\mathbf{b}}_2(\mathbf{x}_1) = \frac{\int_0^1 \mathbf{b} S_1(\mathbf{b})\, d\mathbf{b}}{\int_0^1 S_1(\mathbf{b})\, d\mathbf{b}}.$$

(f) Which of $S_2(\mathbf{b})$, S_2^*, $\hat{S}_2$, $\hat{\mathbf{b}}_2$ are unchanged if we permute the order of appearance of the stock vector outcomes [i.e., if the sequence is now $(1, 2), (1, \frac{1}{2})$]?

16.10 *Growth optimal*. Let $X_1, X_2 \ge 0$, be price relatives of two independent stocks. Suppose that $EX_1 > EX_2$. Do you always want some of X_1 in a growth rate optimal portfolio $S(\mathbf{b}) = bX_1 + \bar{b}X_2$? Prove or provide a counterexample.

16.11 *Cost of universality*. In the discussion of finite-horizon universal portfolios, it was shown that the loss factor due to universality is

$$\frac{1}{V_n} = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}. \tag{16.233}$$

Evaluate V_n for $n = 1, 2, 3$.

16.12 *Convex families*. This problem generalizes Theorem 16.2.2. We say that $\mathcal{S}$ is a convex family of random variables if $S_1, S_2 \in \mathcal{S}$ implies that $\lambda S_1 + (1-\lambda)S_2 \in \mathcal{S}$. Let $\mathcal{S}$ be a closed convex family of random variables. Show that there is a random variable $S^* \in \mathcal{S}$ such that

$$E \ln\left(\frac{S}{S^*}\right) \le 0 \tag{16.234}$$

for all $S \in \mathcal{S}$ if and only if

$$E\left(\frac{S}{S^*}\right) \le 1 \tag{16.235}$$

for all $S \in \mathcal{S}$.

HISTORICAL NOTES

There is an extensive literature on the mean–variance approach to investment in the stock market. A good introduction is the book by Sharpe [491]. Log-optimal portfolios were introduced by Kelly [308] and Latané [346], and generalized by Breiman [75]. The bound on the increase in the growth rate in terms of the mutual information is due to Barron and Cover [31]. See Samuelson [453, 454] for a criticism of log-optimal investment.

The proof of the competitive optimality of the log-optimal portfolio is due to Bell and Cover [39, 40]. Breiman [75] investigated asymptotic optimality for random market processes.

The AEP was introduced by Shannon. The AEP for the stock market and the asymptotic optimality of log-optimal investment are given in Algoet and Cover [21]. The relatively simple sandwich proof for the AEP is due to Algoet and Cover [20]. The AEP for real-valued ergodic processes was proved in full generality by Barron [34] and Orey [402].

The universal portfolio was defined in Cover [110] and the proof of universality was given in Cover [110] and more exactly in Cover and Ordentlich [135]. The fixed-horizon exact calculation of the cost of universality V_n is given in Ordentlich and Cover [401]. The quantity V_n also appears in data compression in the work of Shtarkov [496].