Elements of Information Theory
Thomas M. Cover, Joy A. Thomas
Copyright © 1991 John Wiley & Sons, Inc.
Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1

Chapter 15

Information Theory and the Stock Market

The duality between the growth rate of wealth in the stock market and the entropy rate of the market is striking. We explore this duality in this chapter. 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 expected value in data compression. We shall also find the asymptotic doubling rate for an ergodic stock market process.

15.1 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 represents 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% 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 $S = \mathbf{b}^t\mathbf{X}$.

We wish to maximize S in some sense. But S is a random variable, so there is controversy over the choice of the best distribution for S. The standard theory of stock market investment is based on the con-

sideration 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 wth 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 15.1. 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 which 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, then 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 variance 0. 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 value for a stock corresponding to its risk. This theory of stock prices is called the *Capital Assets Pricing Model* and 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 doubling rate as follows:

Definition: The *doubling rate* of a stock market portfolio $\mathbf{b}$ is defined as

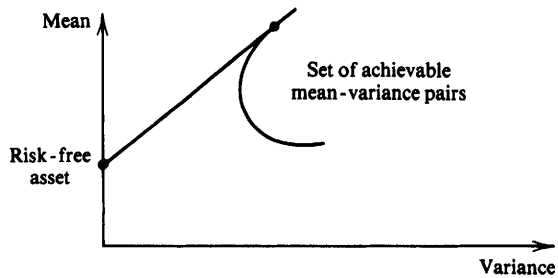

Figure 15.1. Sharpe-Markowitz theory: Set of achievable mean-variance pairs.

$$W(\mathbf{b}, F) = \int \log \mathbf{b}^t \mathbf{x}\, dF(\mathbf{x}) = E(\log \mathbf{b}^t \mathbf{X}) . \tag{15.1}$$

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

$$W^*(F) = \max_{\mathbf{b}} W(\mathbf{b}, F) , \tag{15.2}$$

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

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

The definition of doubling rate is justified by the following theorem:

Theorem 15.1.1: *Let* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *be i.i.d. according to* $F(\mathbf{x})$. *Let*

$$S_n^* = \prod_{i=1}^{n} \mathbf{b}^{*t} \mathbf{X}_i \tag{15.3}$$

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

$$\frac{1}{n} \log S_n^* \rightarrow W^* \quad \textit{with probability 1} . \tag{15.4}$$

Proof:

$$\frac{1}{n} \log S_n^* = \frac{1}{n} \sum_{i=1}^{n} \log \mathbf{b}^{*t} \mathbf{X}_i \tag{15.5}$$

$$\rightarrow W^* \quad \text{with probability 1} , \tag{15.6}$$

by the strong law of large numbers. Hence, $S_n^* \doteq 2^{nW^*}$. □

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

Lemma 15.1.1: $W(\mathbf{b}, F)$ *is concave in* $\mathbf{b}$ *and linear in* F. $W^*(F)$ *is convex in* F.

Proof: The doubling rate is

$$W(\mathbf{b}, F) = \int \log \mathbf{b}^t \mathbf{x}\, dF(\mathbf{x}) . \tag{15.7}$$

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

Since

$$\log(\lambda \mathbf{b}_1 + (1-\lambda)\mathbf{b}_2)^t \mathbf{X} \geq \lambda \log \mathbf{b}_1^t \mathbf{X} + (1-\lambda) \log \mathbf{b}_2^t \mathbf{X}, \qquad (15.8)$$

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

$$W^*(\lambda F_1 + (1-\lambda)F_2) = W(\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2), \lambda F_1 + (1-\lambda)F_2) \qquad (15.9)$$

$$= \lambda W(\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2), F_1) + (1-\lambda) \times W(\mathbf{b}^*(\lambda F_1 + (1-\lambda)F_2), F_2)$$

$$\leq \lambda W(\mathbf{b}^*(F_1), F_1) + (1-\lambda) W^*(\mathbf{b}^*(F_2), F_2), \qquad (15.10)$$

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

Lemma 15.1.2: *The set of log-optimal portfolios forms a convex set.*

Proof: Let $\mathbf{b}_1^*$ and $\mathbf{b}_2^*$ be any two portfolios in the set of log-optimal portfolios. By the previous lemma, the convex combination of $\mathbf{b}_1^*$ and $\mathbf{b}_2^*$ has a doubling rate greater than or equal to the doubling rate of $\mathbf{b}_1^*$ or $\mathbf{b}_2^*$, and hence the convex combination also achieves the maximum doubling rate. Hence the set of portfolios that achieves the maximum is convex. □

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

15.2 KUHN-TUCKER CHARACTERIZATION OF THE LOG-OPTIMAL PORTFOLIO

The determination $\mathbf{b}^*$ that achieves $W^*(F)$ is a problem of maximization of a concave function $W(\mathbf{b}, F)$ over a convex set $\mathbf{b} \in B$. The maximum may lie on the boundary. We can use the standard Kuhn-Tucker conditions to characterize the maximum. Instead, we will derive these conditions from first principles.

Theorem 15.2.1: *The log-optimal portfolio* $\mathbf{b}^*$ *for a stock market* $\mathbf{X}$, *i.e., the portfolio that maximizes the doubling rate* $W(\mathbf{b}, F)$, *satisfies the following necessary and sufficient conditions*:

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) = 1 \quad \text{if} \quad b_i^* > 0\,, \qquad \leq 1 \quad \text{if} \quad b_i^* = 0\,. \tag{15.11}$$

Proof: The doubling rate $W(\mathbf{b}) = E(\log \mathbf{b}^t\mathbf{X})$ 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 non-positive. Thus, letting $\mathbf{b}_\lambda = (1-\lambda)\mathbf{b}^* + \lambda\mathbf{b}$ for $0 \leq \lambda \leq 1$, we have

$$\frac{d}{d\lambda} W(\mathbf{b}_\lambda)\big|_{\lambda=0+} \leq 0, \quad \mathbf{b} \in \mathcal{B}\,. \tag{15.12}$$

These conditions reduce to (15.11) since the one-sided derivative at $\lambda = 0+$ of $W(\mathbf{b}_\lambda)$ is

$$\frac{d}{d\lambda} E(\log(\mathbf{b}_\lambda^t\mathbf{X}))\bigg|_{\lambda=0+} = \lim_{\lambda \downarrow 0} \frac{1}{\lambda} E\left(\log\left(\frac{(1-\lambda)\mathbf{b}^{*t}\mathbf{X} + \lambda\mathbf{b}^t\mathbf{X}}{\mathbf{b}^{*t}\mathbf{X}}\right)\right) \tag{15.13}$$

$$= E\left(\lim_{\lambda \downarrow 0} \frac{1}{\lambda} \log\left(1 + \lambda\left(\frac{\mathbf{b}^t\mathbf{X}}{\mathbf{b}^{*t}\mathbf{X}} - 1\right)\right)\right) \tag{15.14}$$

$$= E\left(\frac{\mathbf{b}^t\mathbf{X}}{\mathbf{b}^{*t}\mathbf{X}}\right) - 1\,, \tag{15.15}$$

where the interchange of limit and expectation can be justified using the dominated convergence theorem [20]. Thus (15.12) reduces to

$$E\left(\frac{\mathbf{b}^t\mathbf{X}}{\mathbf{b}^{*t}\mathbf{X}}\right) - 1 \leq 0 \tag{15.16}$$

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, then the two-sided derivative at $\lambda = 0$ of $W(\mathbf{b}_\lambda)$ vanishes and (15.16) holds with equality. If the line segment from $\mathbf{b}$ to $\mathbf{b}^*$ cannot be extended, then we have an inequality in (15.16).

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 which characterize log-optimum $\mathbf{b}^*$ are equivalent to the following necessary and sufficient conditions:

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) = 1 \quad \text{if} \quad b_i^* > 0\,, \qquad \leq 1 \quad \text{if} \quad b_i^* = 0\,. \quad \square \tag{15.17}$$

This theorem has a few immediate consequences. One surprising result is expressed in the following theorem:

Theorem 15.2.2: *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*

$$E\,\frac{S}{S^*} \le 1\,. \tag{15.18}$$

Conversely, if $E(S/S^*) \le 1$ *for all portfolios* $\mathbf{b}$, *then* $E \log S/S^* \le 0$ *for all* $\mathbf{b}$.

Remark: This theorem can be stated more symmetrically as

$$E \ln \frac{S}{S^*} \le 0, \text{ for all S} \;\Leftrightarrow\; E\frac{S}{S^*} \le 1, \text{ for all } S\,. \tag{15.19}$$

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

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) \le 1 \tag{15.20}$$

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

$$\sum_{i=1}^{m} b_i E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) \le \sum_{i=1}^{m} b_i = 1\,, \tag{15.21}$$

which is equivalent to

$$E\,\frac{\mathbf{b}^t\mathbf{X}}{\mathbf{b}^{*t}\mathbf{X}} = E\,\frac{S}{S^*} \le 1\,. \tag{15.22}$$

The converse follows from Jensen's inequality, since

$$E \log \frac{S}{S^*} \le \log E\,\frac{S}{S^*} \le \log 1 = 0\,. \quad \square \tag{15.23}$$

Thus expected log ratio optimality is equivalent to expected ratio optimality.

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 wealth relative 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 $b_i^* X_i / \mathbf{b}^{*t}\mathbf{X}$, and the expected value of this proportion is

$$E \frac{b_i^* X_i}{\mathbf{b}^{*t}\mathbf{X}} = b_i^* E \frac{X_i}{\mathbf{b}^{*t}\mathbf{X}} = b_i^* 1 = b_i^* . \tag{15.24}$$

Hence the expected proportion of wealth in stock i at the end of the day is the same as the proportion invested in stock i at the beginning of the day.

15.3 ASYMPTOTIC OPTIMALITY OF THE LOG-OPTIMAL PORTFOLIO

In the previous section, 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 will 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

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i^t \mathbf{X}_i \tag{15.25}$$

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

$$W^* = \max_{\mathbf{b}} W(\mathbf{b}, F) = \max_{\mathbf{b}} E \log \mathbf{b}^t \mathbf{X} \tag{15.26}$$

be the maximal doubling rate and let $\mathbf{b}^*$ be a portfolio that achieves the maximum doubling rate.

We only allow portfolios that depend causally on the past and are independent of the future values of the stock market.

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

Lemma 15.3.1: *Let S_n^* be the wealth after n days for the investor using the log-optimal strategy on i.i.d. stocks, and let S_n be the wealth of any other investor using a causal portfolio strategy $\mathbf{b}_i$. Then*

$$E \log S_n^* = nW^* \geq E \log S_n . \tag{15.27}$$

Proof:

$$\max_{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n} E \log S_n = \max_{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n} E \sum_{i=1}^{n} \log \mathbf{b}_i^t \mathbf{X}_i \tag{15.28}$$

$$= \sum_{i=1}^{n} \max_{\mathbf{b}_i(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})} E \log \mathbf{b}_i^t(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})\mathbf{X}_i \tag{15.29}$$

$$= \sum_{i=1}^{n} E \log \mathbf{b}^{*t} \mathbf{X}_i \tag{15.30}$$

$$= nW^*, \tag{15.31}$$

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, i.e., that $\mathbf{b}^*$ (satisfying (15.11)) maximizes the expected log wealth and that the wealth S_n^* is equal to 2^{nW^*} to first order in the exponent, with high probability.

Now we will 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 15.3.1 (*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* $S_n^* = \prod \mathbf{b}^{*t} \mathbf{X}_i$, *where* $\mathbf{b}^*$ *is the log-optimal portfolio, and let* $S_n = \prod \mathbf{b}_i^t \mathbf{X}_i$ *be the wealth resulting from any other causal portfolio. Then*

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 \quad \textit{with probability 1} . \tag{15.32}$$

Proof: From the Kuhn-Tucker conditions, we have

$$E \frac{S_n}{S_n^*} \leq 1 . \tag{15.33}$$

Hence by Markov's inequality, we have

$$\Pr(S_n > t_n S_n^*) = \Pr\left(\frac{S_n}{S_n^*} > t_n \right) < \frac{1}{t_n} . \tag{15.34}$$

Hence

$$\Pr\left(\frac{1}{n} \log \frac{S_n}{S_n^*} > \frac{1}{n} \log t_n \right) \leq \frac{1}{t_n} . \tag{15.35}$$

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

$$\sum_{n=1}^{\infty} \Pr\left(\frac{1}{n} \log \frac{S_n}{S_n^*} > \frac{2 \log n}{n}\right) \leq \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} . \tag{15.36}$$

Then, by the Borel-Cantelli lemma,

$$\Pr\left(\frac{1}{n} \log \frac{S_n}{S_n^*} > \frac{2 \log n}{n}, \text{ infinitely often}\right) = 0 . \tag{15.37}$$

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

$$\limsup \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 \quad \text{with probability 1} . \quad \square \tag{15.38}$$

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

15.4 SIDE INFORMATION AND THE DOUBLING RATE

We showed in Chapter 6 that side information Y for the horse race X can be used to increase the doubling rate by $I(X; Y)$ in the case of uniform odds. We will now extend this result to the stock market. Here, $I(X; Y)$ will be a (possibly unachievable) upper bound on the increase in the doubling rate.

Theorem 15.4.1: *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be drawn i.i.d. $\sim f(\mathbf{x})$. Let $\mathbf{b}_f^*$ be the log-optimal portfolio corresponding to $f(\mathbf{x})$ and let $\mathbf{b}_g^*$ be the log-optimal portfolio corresponding to some other density $g(\mathbf{x})$. Then the increase in doubling rate by using $\mathbf{b}_f^*$ instead of $\mathbf{b}_g^*$ is bounded by*

$$\Delta W = W(\mathbf{b}_f^*, F) - W(\mathbf{b}_g^*, F) \leq D(f \| g) \tag{15.39}$$

Proof: We have

$$\Delta W = \int f(\mathbf{x}) \log \mathbf{b}_f^{*t} \mathbf{x} - \int f(\mathbf{x}) \log \mathbf{b}_g^{*t} \mathbf{x} \tag{15.40}$$

$$= \int f(\mathbf{x}) \log \frac{\mathbf{b}_f^{*t} \mathbf{x}}{\mathbf{b}_g^{*t} \mathbf{x}} \tag{15.41}$$

$$= \int f(\mathbf{x}) \log \frac{\mathbf{b}_f^{*t} \mathbf{x}}{\mathbf{b}_g^{*t} \mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})} \frac{f(\mathbf{x})}{g(\mathbf{x})} \tag{15.42}$$

$$= \int f(\mathbf{x}) \log \frac{\mathbf{b}_f^{*t}\mathbf{x}}{\mathbf{b}_g^{*t}\mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})} + D(f\|g) \tag{15.43}$$

$$\overset{(a)}{\leq} \log \int f(\mathbf{x}) \frac{\mathbf{b}_f^{*t}\mathbf{x}}{\mathbf{b}_g^{*t}\mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})} + D(f\|g) \tag{15.44}$$

$$= \log \int g(\mathbf{x}) \frac{\mathbf{b}_f^{*t}\mathbf{x}}{\mathbf{b}_g^{*t}\mathbf{x}} + D(f\|g) \tag{15.45}$$

$$\overset{(b)}{\leq} \log 1 + D(f\|g) \tag{15.46}$$

$$= D(f\|g), \tag{15.47}$$

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 15.4.2: *The increase ΔW in doubling rate due to side information Y is bounded by*

$$\Delta W \leq I(\mathbf{X}; Y). \tag{15.48}$$

Proof: 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 15.4.1,

$$\Delta W_{Y=y} \leq D(f(\mathbf{x}|Y = y)\| f(\mathbf{x})) = \int_{\mathbf{x}} f(\mathbf{x}|Y = y) \log \frac{f(\mathbf{x}|Y = y)}{f(\mathbf{x})}\, d\mathbf{x}. \tag{15.49}$$

Averaging this over possible values of Y, we have

$$\Delta W \leq \int_y f(y) \int_{\mathbf{x}} f(\mathbf{x}|Y = y) \log \frac{f(\mathbf{x}|Y = y)}{f(\mathbf{x})}\, d\mathbf{x}\, dy \tag{15.50}$$

$$= \int_y \int_{\mathbf{x}} f(y) f(\mathbf{x}|Y = y) \log \frac{f(\mathbf{x}|Y = y)}{f(\mathbf{x})} \frac{f(y)}{f(y)}\, d\mathbf{x}\, dy \tag{15.51}$$

$$= \int_y \int_{\mathbf{x}} f(\mathbf{x}, y) \log \frac{f(\mathbf{x}, y)}{f(\mathbf{x}) f(y)}\, d\mathbf{x}\, dy \tag{15.52}$$

$$= I(\mathbf{X}; Y). \tag{15.53}$$

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

15.5 INVESTMENT IN STATIONARY MARKETS

We now extend some of the results of the previous section 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. We will 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

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i^t(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})\mathbf{X}_i \,. \tag{15.54}$$

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

$$\max_{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n} E \log S_n = \sum_{i=1}^{n} \max_{\mathbf{b}_i(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})} \log \mathbf{b}_i^t \mathbf{X}_i \tag{15.55}$$

$$= \sum_{i=1}^{n} \log \mathbf{b}_i^{*t} \mathbf{X}_i \,, \tag{15.56}$$

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, i.e., $\mathbf{b}_i^*(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1})$ is the portfolio that achieves the conditional maximum, which is denoted by

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

$$= W^*(\mathbf{X}_i | \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}) \tag{15.57}$$

Taking the expectation over the past, we write

$$W^*(\mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) = E \max_{\mathbf{b}} E[\log \mathbf{b}^{*t} \mathbf{X}_i | (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})] \tag{15.58}$$

as the conditional optimal doubling 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

$$W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = \max_{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n} E \log S_n \tag{15.59}$$

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^*:

$$W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = \sum_{i=1}^{n} W^*(\mathbf{X}_i | \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) \,. \tag{15.60}$$

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.

Definition: The *doubling rate* W_∞^* is defined as

$$W_\infty^* = \lim_{n\to\infty} \frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} \tag{15.61}$$

if the limit exists and is undefined otherwise.

Theorem 15.5.1: *For a stationary market, the doubling rate exists and is equal to*

$$W_\infty^* = \lim_{n\to\infty} W^*(\mathbf{X}_n|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}) . \tag{15.62}$$

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

$$\frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} = \frac{1}{n}\sum_{i=1}^{n} W^*(\mathbf{X}_i|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}), \tag{15.63}$$

it follows by the theorem of the Cesáro mean that the left hand side has the same limit as the limit of the terms on the right hand side.

Hence W_∞^* exists and

$$W_\infty^* = \lim \frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} = \lim W^*(\mathbf{X}_n|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}) . \quad \square \tag{15.64}$$

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

Theorem 15.5.2: *Let S_n^* be the wealth resulting from a series of conditionally log-optimal investments in a stationary stock market $\{\mathbf{X}_i\}$. Let S_n be the wealth resulting from any other causal portfolio strategy. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 . \tag{15.65}$$

Proof: From the Kuhn-Tucker conditions for the constrained maximization, we have

$$E\,\frac{S_n}{S_n^*} \leq 1 , \tag{15.66}$$

which follows from repeated application of the conditional version of the Kuhn-Tucker conditions, at each stage conditioning on all the previous outcomes. The rest of the proof is identical to the proof for the i.i.d. stock market and will not be repeated. □

For a stationary ergodic market, we can also extend the asymptotic equipartition property to prove the following theorem.

Theorem 15.5.3 (*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*

$$S_n^* = \prod_{i=1}^{n} \mathbf{b}_i^{*t}(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})\mathbf{X}_i \,. \tag{15.67}$$

Then

$$\frac{1}{n} \log S_n^* \to W^* \quad \textit{with probability 1} \,. \tag{15.68}$$

Proof: The proof involves a modification of the sandwich argument that is used to prove the AEP in Section 15.7. The details of the proof are omitted. □

15.6 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

$$E \frac{S_n}{S_n^*} \leq 1 \,, \tag{15.69}$$

and hence by Markov's inequality,

$$\Pr(S_n > tS_n^*) \leq \frac{1}{t} \,. \tag{15.70}$$

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,

$$(X_1, X_2) = \begin{cases} (1, 1+\epsilon) & \text{with probability } 1-\epsilon\,, \\ (1, 0) & \text{with probability } \epsilon\,. \end{cases} \tag{15.71}$$

In this market, the log-optimal portfolio invests all the wealth in the first stock. (It is easy to verify $\mathbf{b} = (1, 0)$ 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, we cannot prove 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 one half 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 adding an additional fair randomization, which has the effect of reducing the effect of small differences in the wealth.

Theorem 15.6.1 (*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*

$$\Pr(VS \geq U^* S^*) \leq \frac{1}{2}\,. \tag{15.72}$$

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.

Proof: We have

$$\Pr(VS \geq U^* S^*) = \Pr\left(\frac{VS}{S^*} \geq U^*\right) \tag{15.73}$$

$$= \Pr(W \geq U^*)\,, \tag{15.74}$$

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

$$EW = E(V)E\left(\frac{S_n}{S_n^*}\right) \leq 1\,, \tag{15.75}$$

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 $[0, 2]$,

$$\Pr(W \geq U^*) = \int_0^2 \Pr(W > w) f_{U^*}(w)\, dw \tag{15.76}$$

$$= \int_0^2 \Pr(W > w)\, \frac{1}{2}\, dw \tag{15.77}$$

$$= \int_0^2 \frac{1 - F(w)}{2}\, dw \tag{15.78}$$

$$\leq \int_0^\infty \frac{1 - F(w)}{2}\, dw \tag{15.79}$$

$$= \frac{1}{2} EW \tag{15.80}$$

$$\leq \frac{1}{2}, \tag{15.81}$$

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

$$EW = \int_0^\infty (1 - F(w))\, dw \tag{15.82}$$

for a positive random variable W. Hence we have

$$\Pr(VS \geq U^*S^*) = \Pr(W \geq U^*) \leq \frac{1}{2}. \quad \square \tag{15.83}$$

This theorem 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, then the above theorem 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.

Finally, to conclude our discussion of the stock market, 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 non-zero only in the component corresponding to the winning horse.

In this case, the log-optimal portfolio is proportional betting, i.e. $b_i^* = p_i$, and in the case of uniform fair odds

$$W^* = \log m - H(X). \tag{15.84}$$

When we have a sequence of correlated horse races, then the optimal

portfolio is conditional proportional betting and the asymptotic doubling rate is

$$W^* = \log m - H(\mathscr{X}), \tag{15.85}$$

where $H(\mathscr{X}) = \lim \frac{1}{n} H(X_1, X_2, \ldots, X_n)$, if the limit exists. Then Theorem 15.5.3 asserts that

$$S_n^* \doteq 2^{nW^*}. \tag{15.86}$$

15.7 THE SHANNON-McMILLAN-BREIMAN THEOREM

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. sources. In this section, we offer a proof of the theorem for a general ergodic source. We avoid some of the technical details by using a "sandwich" argument, but this section is technically more involved than the rest of the book.

In a sense, an ergodic source is the most general dependent source for which the strong law of large numbers holds. The technical definition requires some ideas from probability theory. To be precise, an ergodic source is defined on a probability space $(\Omega, \mathscr{B}, P)$, where $\mathscr{B}$ is a sigma-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 $\mu(TA) = \mu(A)$, for all $A \in \mathscr{B}$. The transformation is called *ergodic* if every set A such that $TA = A$, a.e., satisfies $\mu(A) = 0$ 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 with a finite expected value, Birkhoff's ergodic theorem states that

$$\frac{1}{n} \sum_{i=1}^{n} X_i(\omega) \to EX = \int X \, dP \quad \text{with probability 1}. \tag{15.87}$$

Thus the law of large numbers holds for ergodic processes.

We wish to use the ergodic theorem to conclude that

$$-\frac{1}{n} \log p(X_0, X_1, \ldots, X_{n-1}) = -\frac{1}{n} \sum_{i=0}^{n-1} \log p(X_i | X_0^{i-1}) \to \lim_{n \to \infty} E[-\log p(X_n | X_0^{n-1})]. \tag{15.88}$$

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

$$H^k = E\{-\log p(X_k|X_{k-1}, X_{k-2}, \ldots, X_0)\} \tag{15.89}$$

$$= E\{-\log p(X_0|X_{-1}, X_{-2}, \ldots, X_{-k})\}, \tag{15.90}$$

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

$$H = \lim_{k\to\infty} H^k \tag{15.91}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} H^k. \tag{15.92}$$

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

$$H^\infty = E\{-\log p(X_0|X_{-1}, X_{-2}, \ldots)\}. \tag{15.93}$$

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 with the knowledge of the k past will have a growth rate of wealth $1 - H^k$, while a gambler with a knowledge of the infinite past will have a growth rate of wealth of $1 - H^\infty$. 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 $1 - H^k$ and $1 - H^\infty$. But $H^k \searrow H = H^\infty$. Thus the sandwich closes and the growth rate must be $1 - H$.

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

Theorem 15.7.1 (*AEP: The Shannon-McMillan-Breiman theorem*): *If H is the entropy rate of a finite-valued stationary ergodic process $\{X_n\}$, then*

$$-\frac{1}{n} \log p(X_0, \ldots, X_{n-1}) \to H, \quad \textit{with probability 1} \tag{15.94}$$

Proof: 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

$$p^k(X_0^{n-1}) = p(X_0^{k-1}) \prod_{i=k}^{n-1} p(X_i | X_{i-k}^{i-1}) . \tag{15.95}$$

From Lemma 15.7.3, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \leq 0 , \tag{15.96}$$

which we rewrite, taking the existence of the limit $\frac{1}{n} \log p^k(X_0^n)$ into account (Lemma 15.7.1), as

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{p(X_0^{n-1})} \leq \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{p^k(X_0^{n-1})} = H^k \tag{15.97}$$

for $k = 1, 2, \ldots$. Also, from Lemma 15.7.3, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{p(X_0^{n-1})}{p(X_0^{n-1} | X_{-\infty}^{-1})} \leq 0 , \tag{15.98}$$

which we rewrite as

$$\liminf \frac{1}{n} \log \frac{1}{p(X_0^{n-1})} \geq \lim \frac{1}{n} \log \frac{1}{p(X_0^{n-1} | X_{-\infty}^{-1})} = H^\infty \tag{15.99}$$

from the definition of H^∞ in Lemma 15.7.1.

Putting together (15.97) and (15.99), we have

$$H^\infty \leq \liminf -\frac{1}{n} \log p(X_0^{n-1}) \leq \limsup -\frac{1}{n} \log p(X_0^{n-1}) \leq H^k \quad \text{for all } k . \tag{15.100}$$

But, by Lemma 15.7.2, $H^k \to H^\infty = H$. Consequently,

$$\lim -\frac{1}{n} \log p(X_0^n) = H . \quad \square \tag{15.101}$$

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

Lemma 15.7.1 (*Markov approximations*): *For a stationary ergodic stochastic process* $\{X_n\}$,

$$-\frac{1}{n} \log p^k(X_0^{n-1}) \to H^k \qquad \textit{with probability 1} , \tag{15.102}$$

$$-\frac{1}{n} \log p(X_0^{n-1} | X_{-\infty}^{-1}) \to H^\infty \quad \textit{with probability 1} . \tag{15.103}$$

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

$$-\frac{1}{n}\log p^k(X_0^{n-1}) = -\frac{1}{n}\log p(X_0^{k-1}) - \frac{1}{n}\sum_{i=k}^{n-1}\log p(X_i|X_{i-k}^{i-1}) \quad (15.104)$$

$$\to 0 + H^k, \quad \text{with probability 1} \quad (15.105)$$

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

$$-\frac{1}{n}\log p(X_0^{n-1}|X_{-1}, X_{-2}, \ldots) = -\frac{1}{n}\sum_{i=k}^{n-1}\log p(X_i|X_{i-k}^{i-1}, X_{-1}, X_{-2}, \ldots) \quad (15.106)$$

$$\to H^\infty \quad \text{with probability 1}. \quad (15.107)$$

Lemma 15.7.2 (*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

$$p(x_0|X_{-k}^{-1}) \to p(x_0|X_{-\infty}^{-1}) \quad \text{with probability 1} \quad (15.108)$$

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

$$\lim_{k\to\infty} H^k = \lim_{k\to\infty} E\left\{-\sum_{x_0\in\mathcal{X}} p(x_0|X_{-k}^{-1})\log p(x_0|X_{-k}^{-1})\right\} \quad (15.109)$$

$$= E\left\{-\sum_{x_0\in\mathcal{X}} p(x_0|X_{-\infty}^{-1})\log p(x_0|X_{-\infty}^{-1})\right\} \quad (15.110)$$

$$= H^\infty. \quad (15.111)$$

Thus $H^k \searrow H = H^\infty$. □

Lemma 15.7.3 (*Sandwich*):

$$\limsup_{n\to\infty} \frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \le 0, \quad (15.112)$$

$$\limsup \frac{1}{n}\log\frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})} \le 0. \quad (15.113)$$

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

$$E\left\{\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})}\right\} = \sum_{x_0^{n-1}\in A} p(x_0^{n-1})\,\frac{p^k(x_0^{n-1})}{p(x_0^{n-1})} \tag{15.114}$$

$$= \sum_{x_0^{n-1}\in A} p^k(x_0^{n-1}) \tag{15.115}$$

$$= p^k(A) \tag{15.116}$$

$$\leq 1\,. \tag{15.117}$$

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

$$E\left\{\frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})}\right\} = E\left[E\left\{\frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})}\,\middle|\,X_{-\infty}^{-1}\right\}\right] \tag{15.118}$$

$$= E\left[\sum_{x^n\in B(X_{-\infty}^{-1})} \frac{p(x^n)}{p(x^n|X_{-\infty}^{-1})}\,p(x^n|X_{-\infty}^{-1})\right] \tag{15.119}$$

$$= E\left[\sum_{x^n\in B(X_{-\infty}^{-1})} p(x^n)\right] \tag{15.120}$$

$$\leq 1\,. \tag{15.121}$$

By Markov's inequality and (15.117), we have

$$\Pr\left\{\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \geq t_n\right\} \leq \frac{1}{t_n} \tag{15.122}$$

or

$$\Pr\left\{\frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \geq \frac{1}{n}\log t_n\right\} \leq \frac{1}{t_n}\,. \tag{15.123}$$

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

$$\left\{\frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \geq \frac{1}{n}\log t_n\right\} \tag{15.124}$$

occurs only finitely often with probability 1. Thus

$$\limsup\,\frac{1}{n}\log\frac{p^k(X_0^{n-1})}{p(X_0^{n-1})} \leq 0 \text{ with probability } 1\,. \tag{15.125}$$

Applying the same arguments using Markov's inequality to (15.121), we obtain

$$\limsup \frac{1}{n} \log \frac{p(X_0^{n-1})}{p(X_0^{n-1}|X_{-\infty}^{-1})} \leq 0 \quad \text{with probability 1}, \tag{15.126}$$

proving the lemma. □

The arguments used in the proof can be extended to prove the AEP for the stock market (Theorem 15.5.3).

SUMMARY OF CHAPTER 15

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

$$W(\mathbf{b}, F) = \int \log \mathbf{b}^t \mathbf{x} \, dF(\mathbf{x}) = E(\log \mathbf{b}^t \mathbf{X}). \tag{15.127}$$

Log-optimal portfolio: The *optimal doubling rate* is

$$W^*(F) = \max_{\mathbf{b}} W(\mathbf{b}, F) \tag{15.128}$$

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

$$E\left(\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}}\right) = 1 \quad \text{if } b_i^* > 0, \qquad \leq 1 \quad \text{if } b_i^* = 0. \tag{15.129}$$

Expected ratio optimality: Letting $S_n^* = \prod_{i=1}^n \mathbf{b}^{*t}\mathbf{X}_i$, $S_n = \prod_{i=1}^n \mathbf{b}_i^t \mathbf{X}_i$, we have

$$E \frac{S_n}{S_n^*} \leq 1. \tag{15.130}$$

Growth rate (AEP):

$$\frac{1}{n} \log S_n^* \to W^*(F) \quad \text{with probability 1}. \tag{15.131}$$

Asymptotic optimality:

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{S_n}{S_n^*} \leq 0 \quad \text{with probability 1}. \tag{15.132}$$

Wrong information: Believing g when f is true loses

$$\Delta W = W(\mathbf{b}_f^*, F) - W(\mathbf{b}_g^*, F) \leq D(f\|g)\,. \tag{15.133}$$

Side information Y:

$$\Delta W \leq I(\mathbf{X}; Y)\,. \tag{15.134}$$

Chain rule:

$$W^*(\mathbf{X}_i|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1}) = \max_{\mathbf{b}_i(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1})} E \log \mathbf{b}_i^t \mathbf{X}_i \tag{15.135}$$

$$W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) = \sum_{i=1}^{n} W^*(\mathbf{X}_i|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{i-1})\,. \tag{15.136}$$

Doubling rate for a stationary market:

$$W_\infty^* = \lim \frac{W^*(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)}{n} \tag{15.137}$$

Competitive optimality of log-optimal portfolios:

$$\Pr(VS \geq U^*S^*) \leq \frac{1}{2}\,. \tag{15.138}$$

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

$$-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(\mathcal{X}) \text{ with probability } 1\,. \tag{15.139}$$

PROBLEMS FOR CHAPTER 15

1. *Doubling rate.* Let

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

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 doubling rate.

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

(b) Find the doubling rate $\mathbf{W}^*$.

(c) Find the asymptotic behavior of

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

for all $\mathbf{b}$.

2. *Side information.* Suppose, in the previous problem, 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 doubling rate W^{**} and verify that $\Delta W = W^{**} - W^{*}$ satisfies

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

3. *Stock market.* Consider a stock market vector

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

Suppose $X_1 = 2$ with probability 1.

(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 that the doubling rate satisfies $W^* \geq 1$.

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 [250]. Log optimal portfolios were introduced by Kelly [150] and Latané [172] and generalized by Breiman [45]. See Samuelson [225, 226] for a criticism of log-optimal investment. An adaptive portfolio counterpart to universal data compression is given in Cover [66].

The proof of the competitive optimality of the log-optimal portfolio is due to Bell and Cover [20, 21]. The AEP for the stock market and the asymptotic optimality of log-optimal investment are given in Algoet and Cover [9]. The AEP for ergodic processes was proved in full generality by Barron [18] and Orey [202]. The sandwich proof for the AEP is based on Algoet and Cover [8].