ASYMPTOTIC OPTIMALITY AND ASYMPTOTIC EQUIPARTITION PROPERTIES OF LOG-OPTIMUM INVESTMENT

by

Paul H. Algoet
and
Thomas M. Cover
Stanford University

TECHNICAL REPORT NO. 57
SEPTEMBER 1985

PREPARED UNDER THE AUSPICES OF
OF
NATIONAL SCIENCE FOUNDATION
GRANT ECS82-11568

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC OPTIMALITY AND ASYMPTOTIC EQUIPARTITION PROPERTIES OF LOG-OPTIMUM INVESTMENT

by

Paul H. Algoet
Thomas M. Cover

TECHNICAL REPORT NO. 57
September 1985

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ECS82-11568

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ASYMPTOTIC OPTIMALITY AND ASYMPTOTIC EQUIPARTITION PROPERTIES OF LOG-OPTIMUM INVESTMENT

Paul H. Algoet\textsuperscript{1} and Thomas M. Cover\textsuperscript{2}

Abstract

Let $X_t$ be a random variable that captures the random behavior of the stock market during period $t$, and $\rho(b_t, X_t)$ the capital return resulting from investment of one monetary unit according to a portfolio $b_t$ selected at the beginning of period $t$. For example if there is a finite number of stocks $j = 1, 2, \ldots, m$ then $X_t = (X_t^j)_{1 \leq j \leq m}$ will denote the vector of individual returns, $b_t = (b_t^j)_{1 \leq j \leq m}$ is a vector of nonnegative weights representing the allocation of funds across the various investment opportunities, and the return is given by the inner product or weighted sum

$$\rho(b_t, X_t) = (b_t \cdot X_t) = \sum_{1 \leq j \leq m} b_t^j X_t^j.$$ 

The capital growth over $n$ periods is given by the product $S_n = \Pi_{0 \leq t < n} \rho(b_t, X_t)$. Let $\mathcal{F}_t$ denote a $\sigma$-field representing information available at the beginning of period $t$, and $S_n^*$ the growth of our fortune over $n$ periods of investment according to portfolios $b_t^*$ achieving the maximum conditional expected log return $\sup_{b_t \in \mathcal{F}_t} \mathbb{E} \{ \log \rho(b_t, X_t) | \mathcal{F}_t \}$. This 'log-optimum' strategy $\{b_t^*\}_{0 \leq t < \infty}$ is asymptotically optimum in the sense that

$$\lim_{n \to \infty} \sup_{n} (1/n) \log(S_n/S_n^*) \leq 0 \quad \text{a.s.}$$

Suppose in addition that $\{X_t\}_{-\infty < t < \infty}$ is stationary ergodic and $\mathcal{F}_t = T^t \mathcal{F}_0$ (obtained by shifting $\mathcal{F}_t$ from period $t$ to the fixed reference period 0) increases monotonically towards a limiting $\sigma$-field $\mathcal{F}_\infty$. For example, if $\mathcal{F}_t = \sigma(X_0, \ldots, X_{t-1})$ then

\textsuperscript{1}Boston University College of Engineering, 110 Cummington Street, Boston, Mass. 02215. Work of this author was partially supported by the Joint Services Electronics Program under grant DAAG 29-84-K-0047 and the National Science Foundation under grant ECS 82-11568.

\textsuperscript{2}Stanford University, Departments of Statistics and Electrical Engineering, Durand 121, Stanford, Ca. 94305. Work by this author was partially supported by the National Science Foundation under grant ECS 82-11568.
$\mathcal{F}_t = \sigma(X_{t-1}, \ldots, X_{-1}) \wedge \mathcal{F}_\infty = \sigma(\ldots, X_{-2}, X_{-1})$. We can then prove the following asymptotic equipartition property for log-optimum investment:

$$(1/n) \log S_n^* \to W_\infty^* = \sup_{b \in F_\infty} \mathbb{E} \{ \log \rho(b, X_0) \} \quad \text{a.s.}$$

or equivalently $S_n^* = \exp[n(W_\infty^* + o(1))]$ where $o(1) \to 0$ a.s. Thus capital will grow exponentially with constant rate $W_\infty^*$ to first order in the exponent, and this rate is highest possible since

$$\limsup_{n \to \infty} (1/n) \log S_n \leq \lim_{n \to \infty} (1/n) \log S_n^* = W_\infty^* \quad \text{a.s.}$$

We use a sandwich argument to reduce the proof to applications of the strong law of large numbers for products, but martingale theory is needed to show that no gap is left in the limit.

Gambling on random outcomes is a special type of investment game. The return is defined as the Radon-Nikodym derivative of the betting measure $b$ with respect to a reference measure that defines the odds. Proportional betting is log-optimum, and the asymptotic equipartition property asserts that the likelihood ratio of the true distribution with respect to a Markovian reference measure builds up exponentially with a limiting rate equal to the relative entropy rate of the process. This is a generalization of the Shannon-McMillan-Breiman theorem.
SECTION 0.

INTRODUCTION

One question that we shall analyze in this paper is how to invest in the stock market. Suppose there is a finite collection of stocks, labelled \( j = 1, 2, \ldots, m \), and \( X_t^j \) denotes the return per unit of capital invested in stock \( j \) during period \( t \) \((0 \leq t < \infty)\). A portfolio is a vector of weights, nonnegative and summing to one; it describes the distribution of funds over the various investment opportunities. The investor must choose a portfolio \( b_t = (b_t^j)_{1 \leq j \leq m} \) at the beginning of each period \( t \). The vector of random returns \( X_t = (X_t^j)_{1 \leq j \leq m} \) is revealed at the end of the period, yielding a total return equal to the inner product or weighted sum

\[
\rho(b_t, X_t) = (b_t \cdot X_t) = \sum_{1 \leq j \leq m} b_t^j X_t^j.
\]

The investor then re-allocates his fortune during the next period according to some portfolio \( b_{t+1} \), etc. If we conveniently assume a normalized initial fortune \( S_0 = 1 \), then the capital \( S_n \) after \( n \) periods is given by the product of the factors by which it has been compounded previously:

\[
S_n = \prod_{0 \leq i < n} \rho(b_i, X_i).
\]

It makes quite a difference whether portfolios can be selected based on the history of some aggregate quantity such as the Dow-Jones average, on detailed records of the past, or perhaps with knowledge of some inside information or with the help of a clairvoyant oracle. Formally we require that \( b_t \) be \( \mathcal{F}_t \)-measurable where \( \mathcal{F}_t \) is a given \( \sigma \)-field that captures what is observable at the beginning of period \( t \). Important examples for the \( \sigma \)-field \( \mathcal{F}_t \) are the finite past \( \sigma(X_0, \ldots, X_{t-1}) \), the infinite past \( \sigma(\ldots, X_{-1}, X_0, \ldots, X_{t-1}) \), and the \( k \)-past, \( \sigma(X_{t-k}, \ldots, X_{t-1}) \) for \( k \leq t < \infty \) and \( \sigma(X_0, \ldots, X_{t-1}) \) for \( 0 \leq t < k \).

An investment strategy \( \{b_t\}_{0 \leq t < \infty} \) is permissible if \( b_t \in \mathcal{F}_t \) for all \( 0 \leq t < \infty \). We need a rule or criterion to cut down the multitude of possibilities, or, even better, to single out an optimum strategy \( \{b_t^*\}_{0 \leq t < \infty} \). Markowitz [52, 59] suggested an ad hoc approach: expected return is traded off with risk, and a quantitative measure of a portfolio's risk is given by the standard deviation of the return. This allows us to restrict our choice to portfolios located on the so-called efficient frontier. Notice that this mean-variance approach tries to optimize the first
two terms in the Taylor series expansion of the expected log return around the neutral point \( X_i^j = 1 \):

\[
E \{ \log(b \cdot X) \} = E \{ (b \cdot R) \} - (1/2) E \{ (b \cdot R)^2 \} + \text{higher order terms}
\]

where \( R = (R^j)_{1 \leq j \leq m} \) and \( R^j = X^j - 1 \).

Economic theory promotes the maximization of subjective expected utility as the supreme guiding principle. This approach is certainly appropriate if the investor's preferences are sufficiently well elucidated so that they can be summarized in a well defined utility function. But subjective utilities are difficult to assess, and many investors may prefer a less elusive and more objective criterion, if there is some rationale for its use. Some authors (e.g. Samuelson [67]) suggest that the log return should be considered as just another utility function, one among the many possibilities. We hope to convince the reader that maximizing expected log return is a strategy that deserves a more privileged status. Casting this approach in the prevailing paradigm of subjective expected utility maximization simply denies its more fundamental character.

Log-optimum investment is founded on information theoretic principles. It was introduced in the literature by Kelly [56], Breiman [60], [61], and others. Breiman considered the case where the outcomes \( \{ X_i \}_{0 \leq i < \infty} \) are independent and identically distributed, and proved that the portfolio \( b^* \) which achieves the maximum expected log return \( \sup_{b \in \mathcal{B}} E \{ \log \rho(b, X) \} \) is optimum according to a variety of criteria. Bell and Cover [80], [85] have shown that the log-optimum strategy is competitively optimum, from a game-theoretic point of view. We consider sequential investment in the case where the random outcomes \( \{ X_i \}_{0 \leq i < \infty} \) exhibit arbitrary statistical dependencies, and prove that log-optimum investment is optimum in the long run, from an asymptotic point of view.

**Theorem 1 (Asymptotic Optimality Principle).**

Let \( \{ X_i \}_{0 \leq i < \infty} \) be a sequence of outcomes with an arbitrary distribution and \( \{ \mathcal{F}_i \}_{0 \leq i < \infty} \) an arbitrary sequence of information fields. Suppose \( b^*_i \) achieves the maximum conditional expected log return

\[
w^*_i = \sup_{b \in \mathcal{B}} E \{ \log \rho(b, X_i) \} = E \{ \log \rho(b^*_i, X_i) \}.
\]

Let \( S^*_n = \prod_{0 \leq i < n} \rho(b^*_i, X_i) \) denote the capital growth over \( n \) periods of investment according to this log-optimum strategy \( \{ b^*_i \}_{0 \leq i < \infty} \), and \( S_n = \prod_{0 \leq i < n} \rho(b_i, X_i) \) the growth under any competing strategy \( \{ b_i \}_{0 \leq i < \infty} \). Then

\[
E \left( \frac{S^*_n}{S_n} \right) \leq 1
\]

(a)
and

\[
(b) \quad \limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) \leq 0 \quad \text{a.s.}
\]

Statement (b) asserts that \( S_n < \exp(n\epsilon) \cdot S_n^* \) eventually for large \( n \) and arbitrary \( \epsilon > 0 \). Thus no competing strategy can infinitely often get ahead of the log-optimum strategy by an amount which grows exponentially fast with a positive rate. Moreover if \( \{\mathcal{F}_t\}_{0 \leq t < \infty} \) is monotone nondecreasing and \( S_n/S_n^* \) is \( \mathcal{F}_n \)-measurable then \( \{S_n/S_n^*, \mathcal{F}_n\}_{0 \leq n < \infty} \) is a nonnegative supermartingale and \( S_n/S_n^* \) converges almost surely to a random variable \( Y \) such that \( \mathbb{E}\{Y\} \leq 1 \).

Part (a) follows from the Kuhn-Tucker conditions for log-optimality, and (b) follows from (a) using the Markov inequality and the Borel-Cantelli lemma. Absolutely no conditions must be imposed on the process \( \{X_t\}_{0 \leq t < \infty} \) or the information fields \( \{\mathcal{F}_t\}_{0 \leq t < \infty} \) for Theorem 1 to be valid (except the extra conditions on the information fields necessary for the submartingale result). This statement enhances results obtained by Breiman [60], [61] and Finkelstein and Whitley [81] for the case where the outcomes \( \{X_t\}_{0 \leq t < \infty} \) are independent and identically distributed.

More can be proved under additional hypotheses. First, let us assume that \( \{X_t\}_{-\infty \leq t < \infty} \) is a stationary ergodic process, that is, \( X_t(\omega) = X(T^t\omega) \) for some invertible measure-preserving and metrically transitive transformation \( T \) of the underlying probability space \((\Omega, \mathcal{F}, \mathbb{P})\). We can then define the shifted information fields

\[
\mathcal{F}_t = T^t \mathcal{F}_i = \{T^tF : F \in \mathcal{F}_i\}.
\]

\( \mathcal{F}_t \) is obtained by shifting \( \mathcal{F}_i \) from period \( t \) to period 0, where all investors face the same decision problem of selecting \( b_0 \) and where they can be compared on common grounds. Our second assumption, which brings martingale theory into play, is that \( \mathcal{F}_t \) increases monotonically towards some limiting \( \sigma \)-field \( \mathcal{F}_\infty \):

\[
\mathcal{F}_t \nearrow \mathcal{F}_\infty = \bigvee_{0 \leq t < \infty} \mathcal{F}_t = \sigma(\mathcal{F}_{0 \leq t < \infty} \mathcal{F}_i).
\]

This condition is certainly satisfied if \( \mathcal{F}_i = \sigma(X_0, \ldots, X_{i-1}) \) since the finite past \( \mathcal{F}_i = \sigma(X_{i-1}, \ldots, X_{i-1}) \) converges to the infinite past \( \mathcal{F}_\infty = \sigma(X_{i-1}, X_{i-2}, \ldots) \).

Let \( \tilde{b}_t^* \) be an \( \mathcal{F}_t \)-measurable portfolio that achieves the maximum conditional expected log return for period 0 given \( \mathcal{F}_i \):

\[
\bar{\omega}_i = \sup_{b \in \mathcal{F}_i} \mathbb{E} \left\{ \log \rho(b, X_0) | \mathcal{F}_i \right\} = \mathbb{E} \left\{ \log \rho(\tilde{b}_i^*, X_0) | \mathcal{F}_i \right\},
\]
for all $0 \leq t \leq \infty$. Furthermore let $\bar{W}_t^*$ denote the maximum (unconditional) expected log return:

$$\bar{W}_t^* = \sup_{b \in \mathcal{F}_t} \mathbb{E} \{\log \rho(b, X_0)\} = \mathbb{E} \{\log \rho(\bar{b}_t^*, X_0)\} = \mathbb{E} \{\bar{w}_t^*\}.$$  

By stationarity $\bar{W}_t^* = W_t^*$ for all $0 \leq t < \infty$, where $W_t^*$ denotes the maximum expected log return for period $0$ under decisions based on $\mathcal{F}_t$:

$$W_t^* = \sup_{b \in \mathcal{F}_t} \mathbb{E} \{\log \rho(b, X_t)\} = \mathbb{E} \{\log \rho(b_t^*, X_t)\} = \mathbb{E} \{w_t^*\}.$$  

It is clear that the maximum expected log return $W_t^*$ will not decrease when the index $t$ grows, since the maximum expectation can then be taken over a larger set of portfolios. This illustrates the principle that more information doesn’t hurt. A refinement states that the sequence $\{\bar{w}_t^*, \mathcal{F}_t\}_{0 \leq t < \infty}$ is a submartingale (strictly speaking only if $\bar{W}_\infty^* < \infty$).

Using the asymptotic optimality principle and standard results from ergodic theory and martingale theory we shall prove

**Theorem 2** (Asymptotic Equipartition Property).

Suppose $T$ is an invertible measure preserving and metrically transitive transformation of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $X_t(\omega) = X(T^t \omega)$ so that the sequence $(X_t)_{-\infty < t < \infty}$ is stationary ergodic. Assume $(\mathcal{F}_t)_{0 \leq t < \infty}$ is a sequence of information fields such that the shifted fields $\mathcal{F}_t = T^t \mathcal{F}_0$ increase monotonically towards a limiting $\sigma$-field $\mathcal{F}_\infty$. Let $S_n^* = \prod_{0 \leq t < n} \rho(b_t^*, X_t)$ denote the fortune after $n$ rounds of investment according to the log-optimum strategy $\{b_t^*\}_{0 \leq t < \infty}$. Then

$$\frac{1}{n} \log S_n^* \to \bar{W}_\infty^* = \sup_{b \in \mathcal{F}_\infty} \mathbb{E} \{\log \rho(b, X_0)\} \quad \text{a.s.}$$

or equivalently,

$$S_n^* = \exp[n(\bar{W}_\infty^* + o(1))] \quad \text{where } o(1) \to 0 \text{ a.s.}$$

where $\bar{W}_\infty^*$ is the maximum expected log return given the infinite past:

$$\bar{W}_\infty^* = \sup_{b \in \mathcal{F}_\infty} \mathbb{E} \{\log \rho(b, X_0)\}.$$  

Thus the capital $S_n^*$ will grow with a constant rate $\bar{W}_\infty^*$ almost surely. This rate is highest possible by the principle of asymptotic optimality, which can now be written in the form

$$\limsup_{n \to \infty} \frac{1}{n} \log S_n \leq \bar{W}_\infty^* = \lim_{n \to \infty} \frac{1}{n} \log S_n^* \quad \text{a.s.}$$

6
Many instances of Theorem 2 can be proved by applying an extension of Birkhoff's ergodic theorem that was used by Breiman [57, 60] to prove the asymptotic equipartition property of information theory. Although the log-optimum portfolio $\tilde{b}_t^*$ is not always unique, it can be shown that the return $\rho(\tilde{b}_t^*, X_0)$ is unambiguously defined and that

$$\rho(\tilde{b}_t^*, X_0) \rightarrow \rho(\tilde{b}_\infty^*, X_0) \quad \text{a.s.}$$

It follows that

$$\frac{1}{n} \log S_n^* = \frac{1}{n} \sum_{0 \leq t < n} \log \rho(\tilde{b}_t^*, X_0) \circ T^t \rightarrow W^*_\infty = \mathbb{E} \left\{ \log \rho(\tilde{b}_\infty^*, X_0) \right\} \quad \text{a.s.}$$

if $\{\log \rho(\tilde{b}_t^*, X_0)\}_{0 \leq t < \infty}$ is $L^1$-dominated, that is, if

$$\mathbb{E} \left\{ \sup_{0 \leq t < \infty} |\log \rho(\tilde{b}_t^*, X_0)| \right\} < \infty.$$ 

There seems to be no evidence that this condition is satisfied by an arbitrary stationary ergodic stock market process, even if $W^*_\infty$ is finite.

We shall provide a more direct proof of Theorem 2 which does not require such integrability assumptions, using a sandwich argument. The general proof is reduced to special cases for which Birkhoff's ergodic theorem suffices. To illustrate the essence of the argument, let us assume that $\mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1})$ for all $0 \leq t < \infty$. We define diminished information fields $\mathcal{F}_t(k) \subseteq \mathcal{F}_t$ for all $0 \leq k < \infty$ and expanded fields $\mathcal{F}_t(\infty)$ as follows:

$$\mathcal{F}_t(k) = \begin{cases} \mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1}) & \text{if } 0 \leq t < k \\ \sigma(X_{t-k}, \ldots, X_{t-1}) & \text{if } k \leq t < \infty \end{cases}$$

and

$$\mathcal{F}_t(\infty) = \sigma(\ldots, X_{-1}, X_0, X_1, \ldots, X_{t-1}).$$

Let $b_t^*(k)$ and $b_t^*(\infty)$ denote the log-optimum decisions based on the $k$-past $\mathcal{F}_t(k)$ and the infinite past $\mathcal{F}_t(\infty)$, respectively, and let $S_n^*(k)$ and $S_n^*(\infty)$ denote the ensuing capital growths over $n$ periods:

$$S_n^*(k) = \prod_{0 \leq t < n} \rho(b_t^*(k), X_t) \quad \text{and} \quad S_n^*(\infty) = \prod_{0 \leq t < n} \rho(b_t^*(\infty), X_t).$$

Now $b_t^*(k)$ competes with $b_t^*$, which is the log-optimum portfolio selection based on $\mathcal{F}_t$. The principle of asymptotic optimality implies

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{S_n^*(k)}{S_n^*} \right) \leq 0 \quad \text{a.s.}$$

7
Similarly $b^*_t$ cannot beat $b^*_t(\infty)$ which is the log-optimum portfolio choice based on $\mathcal{F}_t(\infty)$, and hence

$$\limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S^*_n}{S^*_n(\infty)} \right) \leq 0 \text{ a.s.}$$

But the average rate $(1/n) \log S^*_n(k)$ can be written in the form

$$\frac{1}{n} \log S^*_n(k) = \frac{1}{n} \log S^*_k(k) + \frac{n-k}{n} \frac{1}{n-k} \sum_{k \leq l < n} \log \rho(b^*_k, X_k) \circ T^{l-k}$$

so that, by Birkhoff's ergodic theorem,

$$\frac{1}{n} \log S^*_n(k) \to \mathbb{E} \{ \log \rho(b^*_k, X_k) \} = W^*_k = \bar{W}^*_k \text{ a.s.}$$

Similarly

$$\frac{1}{n} \log S^*_n(\infty) = \frac{1}{n} \sum_{0 \leq l < n} \log \rho(b^*_\infty, X_0) \circ T^l \to \mathbb{E} \{ \log \rho(b^*_\infty, X_0) \} = \bar{W}^*_\infty \text{ a.s.}$$

Combining these results we obtain a chain of asymptotic inequalities

$$\lim_{n \to \infty} \frac{1}{n} \log S^*_n(k) \leq \liminf_{n \to \infty} \frac{1}{n} \log S^*_n \leq \limsup_{n \to \infty} \frac{1}{n} \log S^*_n \leq \lim_{n \to \infty} \frac{1}{n} \log S^*_n(\infty)$$

or

$$W^*_k \leq \liminf_{n \to \infty} \frac{1}{n} \log S^*_n \leq \limsup_{n \to \infty} \frac{1}{n} \log S^*_n \leq \bar{W}^*_\infty \text{ a.s.}$$

Finally, a continuity argument is needed to show that there is no gap between $\sup_{0 \leq k < \infty} \bar{W}^*_k$ and $\bar{W}^*_\infty$. Thus the average rate $(1/n) \log S^*_n$ is asymptotically sandwiched in between an upper bound $\bar{W}^*_\infty$ and lower bounds $\bar{W}^*_k$ which can be made to approach the upper bound arbitrarily closely. It must therefore converge to the upper bound $\bar{W}^*_\infty$ almost surely.

Theorem 2 continues to hold for a stationary or even asymptotically mean stationary stock market process: asymptotic mean stationarity is the most general framework in which ergodic theory applies. However the limiting rate is now a random variable, namely

$$\frac{1}{n} \log S^*_n \to \mathbb{E} \{ \log \rho(b^*_\infty, X_0) | I \} \text{ a.s.}$$

where $I$ is the $\sigma$-field of almost $T$-invariant events:

$$I = \{ F \in \mathcal{F} : F = T^{-1}F \text{ a.s.} \}$$
and $\mathbb{E}\{\cdot\}$ denotes expectation with respect to the stationary mean (ordinary expectation in the stationary case). By the ergodic decomposition theorem every stationary measure is a mixture of stationary ergodic modes, and we may interpret the limiting rate $\mathbb{E}\{\log \rho(b_{X_0}^\ast, X_0)\mid I\}$ as the maximum growth rate $\hat{W}_\infty^\ast(\epsilon)$ of the stationary ergodic mode $\epsilon = \epsilon(\omega)$ in which the process happens to be locked (in the stationary case) or towards which it is evolving (in the asymptotically mean stationary case).

How to invest in the stock market is only one of the questions studied in this paper. To cover other applications as well we need a slightly more general framework than we have been presenting so far. First, we shall not restrict ourselves to a finite set but allow an arbitrary separable metrizable space $\mathcal{A}$ of elementary investment opportunities. Portfolios are probability measures on $(\mathcal{A}, \mathcal{B}_\mathcal{A})$, not necessarily weight vectors in a finite-dimensional simplex. The random behavior of the market during period $t$ is captured by a random variable $X_t$ with values in some separable metrizable space $\mathcal{X}$. The return of capital invested in primitive option $a \in \mathcal{A}$ is given by $\rho(a, X_t)$ where $\rho : \mathcal{A} \times \mathcal{X} \to \mathbb{R}_+$ is a measurable function. If the investor chooses to distribute his money or bets according to a portfolio measure $b_t(da)$ at the beginning of period $t$ then his return at the end of that period is given by the weighted average

$$\rho(b_t, X_t) = \int_{\mathcal{A}} \rho(a, X_t) \, b_t(da).$$

We shall make one further generalization. Instead of probability measures on a space of investment alternatives $\mathcal{A}$, portfolios will be regarded as elements of some abstract separable metrizable space $\mathcal{B}$, which is defined all by itself without reference to some underlying space of more primitive objects. The task of the decision agent is to select an element $b_t$ from the space of possible actions $\mathcal{B}$ at the beginning of each period $t$, and his capital will grow by a factor $\rho(b_t, X_t)$ where $\rho : \mathcal{B} \times \mathcal{X} \to \mathbb{R}_+$ is measurable. However we do assume that $\mathcal{B}$ is a convex subset of a linear space and that $\rho(b, x)$ is concave in $b \in \mathcal{B}$ for fixed $x \in \mathcal{X}$. (In fact $\rho(b, x)$ is affine in $b$ for the concrete applications considered in this paper, but proofs carry through in general assuming only that $\rho(b, x)$ is concave in $b$.) Theorems 1 and 2 remain valid in this more general setting if the following two hypotheses hold.

(H1) A measurable selection of log-optimum portfolios exists.

(H2) The expected log return is lower semicontinuous and bounded below.
Let $\mathcal{P}$ be the space of probability measures on $(\mathcal{X}, \mathcal{B}_X)$, equipped with the weak topology. Hypothesis H1 guarantees that a portfolio $b^*(P) \in \mathcal{B}$ can be selected for every possible distribution $P$ on the space of outcomes $(\mathcal{X}, \mathcal{B}_X)$ such that $b^*(P)$ achieves the maximum expected log return

$$w^*(P) = \sup_{b \in \mathcal{B}} \mathbb{E}_P \{ \log \rho(b, X) \}$$

and $b^* : \mathcal{P} \rightarrow \mathcal{B}$ is a measurable function. Consequently, if $P_t$ denotes a regular conditional probability distribution of $X_t$ given $\mathcal{F}_t$, then $P_t$ is a sufficient summary of everything the decision agent has to know at the beginning of period $t$ in order to make an informed selection, and $b^*_t = b^*(P_t)$ is a log-optimum action based on $\mathcal{F}_t$ that achieves the maximum conditional expected log return $w^*_t = w^*(P_t)$. Hypothesis H2 asserts that the function $P \mapsto w^*(P)$ is bounded below and lower semicontinuous in $P \in \mathcal{P}$. The sandwich argument and thus the proof of Theorem 2 can be completed if H2 holds. Indeed, let $\bar{P}_t$ denote a regular conditional probability distribution of $X_0$ given $\mathcal{F}_t$, for $0 \leq t \leq \infty$. Then $\bar{b}^*_t = b^*(\bar{P}_t)$ achieves the maximum conditional expected log return $\bar{w}^*_t = w^*(\bar{P}_t)$. But $\bar{P}_t \rightarrow \bar{P}_\infty$ a.s. by Lévy’s martingale convergence theorem for conditional expectations so that, by the lower semicontinuity of $w^*$,

$$\liminf_{t \rightarrow \infty} \bar{w}^*_t = \liminf_{t \rightarrow \infty} w^*(\bar{P}_t) \geq \bar{w}^*_\infty = w^*(\bar{P}_\infty) \quad \text{a.s.}$$

Since $w^*$ is bounded below, Fatou’s lemma is applicable and yields

$$\liminf_{t \rightarrow \infty} \bar{W}^*_t = \liminf_{t \rightarrow \infty} \mathbb{E} \{ \bar{w}^*_t \} \geq \bar{W}^*_\infty = \mathbb{E} \{ \bar{w}^*_\infty \}$$

proving that there is no gap between $\bar{W}^*_\infty$ and $\sup_{0 \leq t \leq \infty} \bar{W}^*_t$. Since we already know that $\{\bar{W}^*_t\}_{0 \leq t \leq \infty}$ is monotone increasing it follows that $\bar{W}^*_t \nearrow \bar{W}^*_\infty$.

It is occasionally advantageous to decompose the return function $\rho(b, X)$ into the product of a factor $\varphi(X)$ which doesn’t depend on the portfolio $b$ and a factor $\lambda(b, u(X))$ which depends on $X$ only through a function $U = u(X)$. We call $U$ the normalized outcome, and denote its distribution by $Q$. The maximum expected log return $w^*(P)$ can then be written as the difference between an ideal reference level $r(P) = \mathbb{E}_P \{ \log \varphi(X) \}$ over which the investor has no control, and an unavoidable loss $h^*(Q) = \inf_{b \in \mathcal{B}} \mathbb{E}_Q \{ -\log \lambda(b, U) \}$ with respect to that reference level. For example, if there are finitely many stocks $j = 1, 2, \ldots, m$ and the return is given by the inner product $\rho(b, x) = (b \cdot x) = \sum_{1 \leq j \leq m} b_j x_j$ then we shall define the normalized outcome as the vector $U = X/(X^1 + \ldots + X^m)$ in the simplex which has the same direction as $X$, and the reference level as $\varphi(X) =$
The function \( h^*(Q) = \inf_{b \in \mathcal{B}} E_Q\{ -\log(b \cdot U) \} \) is then concave, continuous in \( Q \), and bounded between 0 and \( \log m \), whereas \( w^*(P) \) is not even lower semicontinuous unless we artificially restrict the outcomes to be bounded away from \((0, 0, \ldots, 0)\). The minimum loss \( h^*(Q) \) is called the generalized entropy since it reduces to Shannon's entropy function

\[
h^*(q^1, \ldots, q^m) = - \sum_{1 \leq i \leq m} q^i \log q^i
\]

if \( Q \) can be identified with a probability mass function \((q^1, \ldots, q^m)\) on the extreme points of the simplex. An analogous decomposition holds for the maximum capital growth rate of a stationary ergodic stock market process: \( \tilde{\mathcal{W}}^*_\infty \) is the difference between an ideal reference level \( R \) that describes the global trend of the market, and an unavoidable loss equal to the generalized entropy rate \( \tilde{\mathcal{H}}^*_\infty \). It makes sense to subtract out the reference level if it is infinite and if the problem of log-optimum portfolio selection is thereby reduced to minimizing a lower semicontinuous function over a compact set.

Theorem 2 is called the asymptotic equipartition property for log-optimum investment. This is no accident since it is a generalization of the fundamental theorem of information theory which is occasionally referred to by the same name and that is due to Shannon [48], McMillan [56], and Breiman [57,60]. The relevance of information theory to the portfolio selection problem was suggested by Kelly [56] who considered a special type of investment game, namely gambling on the outcome of a horse race.

We generalize the horse race problem as follows. Let \( \mathcal{A} = \mathcal{X} \) be a complete separable metrizable space of possible outcomes (a continuum of horses). The odds are defined by a fixed reference measure \( m \) on \((\mathcal{X}, \mathcal{B}_\mathcal{X})\) and the gambler distributes his money according to a betting measure \( b_t \) during round \( t \). If \( \alpha \) is a finite measurable partition of \( \mathcal{X} \) then we may consider the horse race problem with return

\[
\rho^\alpha(b_t, X_t) = \frac{b_t\{\alpha(X_t)\}}{m\{\alpha(X_t)\}}
\]

where \( \alpha(X_t) \) denotes the cell of the partition \( \alpha \) into which the outcome \( X_t \) of round \( t \) is classified. There is one strategy which is log-optimum for all possible partitions \( \alpha \), namely \( b^*_t = P_t \), where \( P_t \) is the conditional distribution of \( X_t \) given \( \mathcal{F}_t \). Thus the gambler should ignore the odds as defined by the reference measure \( m \), and place proportional bets, \( b^*_t(A) = P_t(A) \) for all measurable subsets \( A \in \alpha \). The maximum conditional expected log return for period \( t \) is then given by the Kullback-Leibler divergence \( D(P_t^\alpha||m^\alpha) \), where \( P_t^\alpha \) and \( m^\alpha \) denote the
restrictions of \( P_t \) and \( m \) to the \( \sigma \)-field \( \alpha \). But \( D(P_t^\alpha \| m^\alpha) \) increases as \( \alpha \) is refined and \( \log \rho^\alpha(b_t, X_t) \) is a submartingale indexed by the directed set of finite measurable partitions \( \alpha \). Again, more information doesn’t hurt and it is in the interest of the gambler to insist on as fine a partition \( \alpha \) of the space of outcomes as is practically feasible. Thus we shall define the return \( \rho(b_t, X_t) \) as the limit of \( \rho^\alpha(b_t, X_t) \) as the partition \( \alpha \) is more and more refined:

\[
\rho(b_t, x_t) = \lim_{\alpha \uparrow} \rho^\alpha(b_t, x_t) = \frac{b_t\{X_t \in dx_t\}}{m\{X_t \in dx_t\}}.
\]

The maximum conditional expected log return is given by the Kullback-Leibler divergence

\[
D(P_t \| m) = \sup_{\alpha} D(P_t^\alpha \| m^\alpha)
\]

and is uniquely achieved by \( b_t^* = P_t \). The maximum expected log return is equal to the relative entropy

\[
\Delta_t^* = \mathbb{E}\{D(P_t \| m)\} = \mathbb{E}\left\{ \log \left( \frac{dP_t}{dm} \right) \right\}.
\]

Choosing \( \mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1}) \) for all \( 0 \leq t < \infty \), our generalized A.E.P. for log-optimum investment reduces to the following

**Theorem 2'. (Generalized Shannon-McMillan-Breiman Theorem).**

Let \( \{X_t\}_{-\infty < t < \infty} \) be a stationary ergodic process and let \( p(X_t|X_{t-1}, \ldots, X_0) \) and \( p(X_0, \ldots, X_{n-1}) \) denote the conditional and joint densities of the process with respect to a fixed reference measure \( m \) on \( (\mathcal{X}, \mathcal{B}_\mathcal{X}) \):

\[
p(x_t|x_{t-1}, \ldots, x_0) = \frac{\mathbb{P}\{X_t \in dx_t|X_{t-1} = x_{t-1}, \ldots, X_0 = x_0\}}{m\{X_t \in dx_t\}}
\]

and

\[
p(x_0, \ldots, x_{n-1}) = \frac{\mathbb{P}\{X_0 \in dx_0, \ldots, X_{n-1} \in dx_{n-1}\}}{m\{X_0 \in dx_0\} \ldots m\{X_{n-1} \in dx_{n-1}\}} = \prod_{0 \leq t < n} p(x_t|x_{t-1}, \ldots, x_0).
\]

Then the average log-likelihood converges almost surely:

\[
\frac{1}{n} \log p(X_0, \ldots, X_{n-1}) = \frac{1}{n} \sum_{0 \leq t < n} \log p(X_t|X_{t-1}, \ldots, X_0) \to \bar{\Delta}_\infty^* \text{ a.s.}
\]

where \( \bar{\Delta}_\infty^* \) is the relative entropy rate of the process:

\[
\bar{\Delta}_\infty^* = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\{\log p(X_0, \ldots, X_{n-1})\}
\]

\[
= \lim_{n \to \infty} \mathbb{E}\{\log p(X_t|X_{t-1}, \ldots, X_0)\}
\]

\[
= \mathbb{E}\{\log p(X_0|X_{-1}, X_{-2}, \ldots)\}.
\]

12
This theorem was first proved by Barron [85]. A elementary proof that incorporates a sandwich argument but not the gambling interpretation was given by Algoet and Cover [85]. The result continues to hold if we use a stationary Markovian reference measure rather than the product of independent copies of \( m \).

We conclude this introduction with a summary of the paper. Section 1 focuses on log-optimum investment for a single period, to prepare for later sections where the sequential investment problem will be considered. We examine the properties of the maximum expected log return \( w^*(P) = \sup_{b \in \mathcal{B}} E_P \{ \log \rho(b, X) \} \) as a function of \( P \), and existence of a log-optimum portfolio selection rule \( b^*(P) \) achieving \( w^*(P) \).

In section 2 we prove the asymptotic optimality principle for log-optimum investment (Theorem 1 in this introduction).

The principle that more information doesn’t hurt is the subject of section 3. If \( \{ \mathcal{F}_t \}_{0 \leq t < \infty} \) is a sequence of information fields such that \( \mathcal{F}_t \nearrow \mathcal{F}_\infty \) then the maximum conditional expected log returns \( \{ \hat{w}_t^*, \mathcal{F}_t \}_{0 \leq t < \infty} \) form a submartingale and the maximum expected log returns \( \hat{W}_t^* \) form a monotone increasing sequence. In fact, there is no gap between \( \sup_{0 \leq t < \infty} \hat{W}_t^* \) and \( \hat{W}_\infty^* \) and hence \( \hat{W}_t^* \nearrow \hat{W}_\infty^* \) in case \( w^*(P) \) is lower semicontinuous and bounded below.

The asymptotic equipartition property for log-optimum investment (Theorem 2 of this introduction) is proved in section 4. We assume that \( T \) is an invertible measure preserving and metrically transitive transformation of the underlying probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), so that the process \( X_t(\omega) = X(T^t \omega) \) \((-\infty < t < \infty)\) is stationary ergodic. We also assume that \( \{ \mathcal{F}_t \}_{0 \leq t < \infty} \) is a sequence of information fields such that the shifted fields \( \mathcal{F}_t = T^t \mathcal{F}_t \) increase monotonically towards a limiting \( \sigma \)-field \( \mathcal{F}_\infty \), so that we can invoke the martingale results of section 3. Two proofs are provided for the asymptotic equipartition property. The first goes along the lines of Breiman’s proof of the individual ergodic theorem of information theory: it uses an extension of Birkhoff’s ergodic theorem but requires that the sequence \( \{ \log \rho(b_t, X_0) \}_{0 \leq t < \infty} \) be \( L^1 \)-bounded. The second proof uses the sandwich argument, but requires that there be no gap between \( \sup_{0 \leq t < \infty} \hat{W}_t^* \) and \( \hat{W}_\infty^* \).

In section 5 we write the return function as a product \( \rho(b, x) = \varphi(x) \cdot \lambda(b, u(x)) \) and we discuss the resulting decomposition of the maximum expected log return \( w^*(P) \) into an ideal reference level \( r(P) = E_P \{ \log \varphi(X) \} \) and an unavoidable loss or generalized entropy \( h^*(Q) = \inf_{u \in \mathcal{U}} E_Q \{ -\log \lambda(b, U) \} \). Such a decomposition is advantageous if \( r(P) \) is infinite and also if \( \mathcal{B} \) is compact and \( \lambda : \mathcal{B} \times \mathcal{U} \to [0, 1] \) is
upper semicontinuous. Most interesting is the case when $h^*(Q)$ is continuous in $Q$, and this occurs for a market with a finite number of investment opportunities, as we show in section 6. The maximum expected log return $w^*(P)$ for such a market is not lower semicontinuous in $P$. Nevertheless we can prove that $\bar{W}_i^*/\bar{W}_\infty^*$ using a divide-and-conquer approach, by considering the two terms of the decomposition $\bar{W}_i^* = R - \bar{H}_i^*$ separately.

Gambling on the outcome of a horserace is a special type of investment game. The distinguishing feature is that the vector of outcomes is oriented along one of the coordinate axes, i.e., exactly one of the stocks yields a nonzero return. In section 7 we discuss how to invest on a continuum of horses and derive the generalized Shannon-McMillan-Breiman theorem (Theorem 2' of the introduction) as a special case of the asymptotic equipartition property for log-optimum investment.

Finally in section 8 we generalize Theorem 2 to stationary and even asymptotically mean stationary stock market processes, and we discuss the ergodic decomposition of the limiting growth rate $\bar{W}_\infty^*$. The definition and basic properties of lower semicontinuous functions are recalled in an appendix.

We assume throughout that the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is perfect. Perfect measures have many desirable properties. For example, any random variable defined on such a space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in a separable metric space admits a regular conditional probability distribution. Kolmogorov's extension theorem for consistent families of finite dimensional marginal distributions holds if these marginals are perfect, and the unique extension measure is then also perfect. Every probability measure on a universally measurable subset of a Polish space is perfect, and the ergodic decomposition theorem holds for any stationary measure on an infinite sequence space which is the product of copies of such a universally measurable space. See Gnedenko and Kolmogorov [49], Pachl [79], Ramachandran [79], Ryll-Nardzewski [53] and Sazonov [62] for the definition and extensive discussions of perfect measures.
SECTION 1.

LOG-OPTIMUM INVESTMENT
FOR A SINGLE PERIOD

We shall first analyze log-optimum investment for a single period, to set the stage for later sections where the sequential problem will be considered.

The stock market will be formalized by the following model. There is a random variable $X$, which captures the random behavior of the market. Let $\mathcal{X}$ be the space of possible outcomes. There is also a space of elementary investment opportunities $\mathcal{A}$, and capital invested in a stock $a \in \mathcal{A}$ at the beginning of the investment period will yield a return $\rho(a, X)$ when the random outcome $X \in \mathcal{X}$ is realized at the end. We assume that $\mathcal{A}$ and $\mathcal{X}$ are separable metrizable spaces and that the return function $\rho : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}_+$ is Borel measurable. $P$ will denote the probability distribution of $X$ on the Borel $\sigma$-field $\mathcal{B}_X$ of $\mathcal{X}$.

The problem faced by the investor is one of allocation. Investing all capital in the stock which offers the best prospects may be a bad idea if that opens up the possibility of going broke and if risk can be reduced by employing a diversified strategy. A portfolio describes how to distribute funds across the spectrum of primitive stock options. Let $\mathcal{B}$ be the space of all portfolios: $\mathcal{B}$ is the space of normalized measures on $(\mathcal{A}, \mathcal{B}_\mathcal{A})$. The investor must select an element $b \in \mathcal{B}$, knowing the distribution $P$ but not the true value of the outcome $X$. His return at the end of the period when $X$ is revealed is given by the weighted average

$$\rho(b, X) = \int_{\mathcal{A}} \rho(a, X) b(da).$$

We imagine the set of elementary investment opportunities $\mathcal{A}$ embedded as a subset in the space of diversified actions $\mathcal{B}$, identifying an atomic action $a \in \mathcal{A}$ with the Dirac measure $b = \delta_a \in \mathcal{B}$ which places unit mass on $a$. Thus the same symbol may be used to denote the return function $\rho : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}_+$ and its extension $\rho : \mathcal{B} \times \mathcal{X} \rightarrow \mathbb{R}_+$, and $\rho(a, x) = \rho(\delta_a, x)$ for $a \in \mathcal{A}$. $\mathcal{B}$ is a convex subset of a linear space, and $\rho(b, x)$ is affine in $b \in \mathcal{B}$ for fixed $x \in \mathcal{X}$.

The space $\mathcal{B}$ of probability measures on a separable metrizable space $\mathcal{A}$ is itself separable and metrizable, when equipped with the weak topology. In fact $\mathcal{B}$ is compact, complete (and thus Polish), or analytic if $\mathcal{A}$ is such. Most of the theory to be derived in this paper remains valid if we forget about the interpretation of $\mathcal{B}$ as a space of measures on some underlying space $\mathcal{A}$. Thus we will often view $\mathcal{B}$ as an abstract separable metrizable space defined all by itself, and $\rho(b, X)$ as the capital gain resulting from selection of an element $b \in \mathcal{B}$. However we
always assume that \( \mathcal{B} \) is a convex subset of a linear space, and that \( \rho(b,x) \) is Borel measurable on \( \mathcal{B} \times \mathcal{X} \) and concave in \( b \in \mathcal{B} \) for fixed \( x \in \mathcal{X} \). The reader may verify that such a representation of \( \mathcal{B} \) as the space of probability measures on a space \( \mathcal{A} \) and of \( \rho(b,x) \) as the average \( \int_{\mathcal{A}} \rho(a,x) b(da) \) of a function \( \rho(a,x) \) is possible iff the convex set \( \mathcal{B} \) is a Choquet simplex (e.g., if \( \mathcal{B} \) is compact and convex), and \( \rho(b,x) \) is affine in \( b \in \mathcal{B} \) for fixed \( x \in \mathcal{X} \). The space of investment alternatives \( \mathcal{A} \) then coincides with the set of extreme points of \( \mathcal{B} \), and every \( b \in \mathcal{B} \) can be identified with (and is equal to the barycenter of) a unique probability measure on \( \mathcal{A} \).

Our objective will be to achieve the maximum expected log return. If we choose \( b \in \mathcal{B} \) then the true log return \( w(b,X) = \log \rho(b,X) \) has expectation

\[
w(b,P) = E_P \{ w(b,X) \} = E_P \{ \log \rho(b,X) \}.
\]

We shall write \( w(b,P) = -\infty \) if this expectation is not well-defined in the usual sense, that is, if both \( E_P \{ \log^+ \rho(b,X) \} = \infty \) and \( E_P \{ \log^- \rho(b,X) \} = \infty \). This convention keeps us on the safe side when trying to achieve the maximum expected log return

\[
w^*(P) = \sup_{b \in \mathcal{B}} w(b,P) = \sup_{b \in \mathcal{B}} E_P \{ \log \rho(b,X) \}.
\]

It makes sense to call \( \rho(a,X)/\rho(b,X) \) the score function, in analogy with the theory of maximum likelihood estimation. Indeed, \( \rho(a,X)/\rho(b,X) \) defines the derivative of the log return with respect to \( b \):

\[
\log \rho(b + \delta b, X) - \log \rho(b, X) = \int_{\mathcal{A}} \frac{\rho(a,X)}{\rho(b,X)} \delta b(da).
\]

The expected score is given by

\[
\gamma(a,b;P) = E_P \left\{ \frac{\rho(a,X)}{\int_{\mathcal{A}} \rho(a,X) b(da)} \right\} = E_P \left\{ \frac{\rho(a,X)}{\rho(b,X)} \right\} = E_P \left\{ \frac{\rho(\delta a, X)}{\rho(b,X)} \right\}.
\]

Notice that \( \gamma(a,b;P) \) defines the derivative of the expected log return \( w(b,P) \) with respect to \( b \):

\[
w(b+\delta b, P) - w(b,P) = \int_{\mathcal{A}} \gamma(a,b;P) \delta b(da).
\]

We can also define \( \gamma(a,b;P) - 1 \) as the influence function or directional derivative of \( w(b,P) \) in the direction from \( b \) to \( \delta a \):

\[
\lim_{\varepsilon \to 0} \frac{w((1-\varepsilon) \cdot b + \varepsilon \cdot \delta a, P) - w(b,P)}{\varepsilon} = E_P \left\{ \frac{\rho(a,X)}{\rho(b,X)} \right\} - 1 = \gamma(a,b;P) - 1.
\]
An action \( b^* \in \mathcal{B} \) is called log-optimum if it achieves the maximum expected log return \( w^*(P) \). Let \( B^*(P) \) denote the set of all log-optimum portfolios:

\[
B^*(P) = \{ b^* \in \mathcal{B} : w(b^*, P) = w^*(P) = \sup_{b \in \mathcal{B}} w(b, P) \}.
\]

Clearly \( w^*(P) = \sup_{b \in \mathcal{B}} w(b, P) \) is a convex function of \( P \in \mathcal{P} \), since \( w(b, P) \) is affine in \( P \in \mathcal{P} \) for all fixed \( b \in \mathcal{B} \). Furthermore \( B^*(P) \) is convex since \( w(b, P) \) is concave in \( b \). The Kuhn-Tucker conditions provide equivalent criteria for log-optimality:

**Theorem 1.1** Let \( P \) be a fixed probability measure on \( (X, \mathcal{B}_X) \). Then the following are equivalent assertions about an action \( b^* \in \mathcal{B} \):

1. \( b^* \) is log-optimum, that is, \( b^* \in B^*(P) \), or \( b^* \) achieves the maximum expected log return \( w^*(P) = \sup_{b \in \mathcal{B}} w(b, P) \), or

\[
w(b, P) - w(b^*, P) = E_P \left\{ \log \left( \frac{\rho(b, X)}{\rho(b^*, X)} \right) \right\} \leq 0 \quad \text{for all } b \in \mathcal{B}.
\]

2. \( b^* \) satisfies the Kuhn-Tucker conditions

\[
E_P \left\{ \frac{\rho(b, X)}{\rho(b^*, X)} \right\} \leq 1 \quad \text{for all } b \in \mathcal{B}.
\]

3. \( b^* \) satisfies \( \gamma(a, b^*; P) \leq 1 \) for all \( a \in \mathcal{A} \), or

\[
E_P \left\{ \frac{\rho(\delta(a, X)}{\rho(b^*, X)} \right\} \leq 1 \quad \text{for all } a \in \mathcal{A}.
\]

**Proof:**

If \( E_P \left\{ \frac{\rho(b, X)}{\rho(b^*, X)} \right\} \leq 1 \) then by Jensen’s inequality,

\[
w(b, P) - w(b^*, P) = E_P \left\{ \log \left( \frac{\rho(b, X)}{\rho(b^*, X)} \right) \right\} \leq \log \left( E_P \left( \frac{\rho(b, X)}{\rho(b^*, X)} \right) \right) \leq \log 1 = 0.
\]

Thus 2 implies 1. Conversely let us assume that \( w(b, P) - w(b^*, P) \leq 0 \) for all \( b \in \mathcal{B} \), and yet there exists \( b_1 \) in \( \mathcal{B} \) such that \( E_P \{ \rho(b_1, X)/\rho(b^*, X) \} > 1 \). We will derive a contradiction using the concavity of \( \rho(b, X) \) in \( b \). Indeed, consider \( b_\lambda = \lambda b_1 + (1-\lambda) b^* \), where \( 0 < \lambda < 1 \) and \( 1-\lambda = 1 - \lambda \). Then \( \rho(b_\lambda, X) \geq \lambda \rho(b_1, X) + (1-\lambda) \rho(b^*, X) \) so that

\[
\frac{\rho(b_\lambda, X)}{\rho(b^*, X)} \geq \lambda \frac{\rho(b_1, X)}{\rho(b^*, X)} + (1 - \lambda) \frac{\rho(b^*, X)}{\rho(b^*, X)} = 1 + \lambda Z
\]

where \( Z = \frac{\rho(b_1, X)}{\rho(b^*, X)} - 1 \).
Using a Taylor series expansion and writing \( p \land q \) for the minimum of \( p \) and \( q \), we have, for any \( a > 0 \):

\[
\log(1 + \lambda Z) \geq \log(1 + \lambda (Z \land a)) \\
= \lambda (Z \land a) - \frac{1}{2} \theta \lambda^2 (Z \land a)^2 \quad \text{for some } 0 < \theta < 1 \\
\geq \lambda (Z \land a) - \frac{1}{2} \lambda^2 a^2.
\]

Since \( E_P \{ Z \} > 0 \) by hypothesis, we can choose \( a > 0 \) sufficiently large so that \( E_P \{ Z \land a \} > 0 \). Then we can choose \( \lambda < (2/a) E_P \{ Z \land a \} \) sufficiently small so that \( \lambda E_P \{ Z \land a \} - \frac{1}{2} \lambda^2 a^2 > 0 \). For these choices of \( \lambda \) and \( a \), we have

\[
E_P \left\{ \log \left( \frac{\rho(b_\lambda, X)}{\rho(b^*, X)} \right) \right\} = E_P \{ \log(1 + \lambda Z) \} \geq \lambda E_P \{ Z \land a \} - \frac{1}{2} \lambda^2 a^2 > 0.
\]

Thus \( w(b_\lambda, P) - w(b^*, P) > 0 \). The existence of such \( b_\lambda \) contradicts the initial hypothesis and it follows that \( 1 \) implies \( 2 \).

It is immediate that \( 2 \) implies \( 3 \), since \( \{ \delta_a : a \in A \} \) is a subset of \( B \). To show the converse it suffices to observe that

\[
E \left\{ \frac{\rho(b, X)}{\rho(b^*, X)} \right\} = \int_A \gamma(a, b^*; P) b(da).
\]

The directional derivative \( \gamma(a, b^*; P) - 1 \) of \( w(b, P) \) at \( b^* \) should be nonpositive in any direction inward into the convex set \( B \). But \( B \) is a convex set with extreme points \( A \) so that the cone of inward directions is spanned by vectors from \( b^* \) to extreme points \( \delta_a, a \in A \).

QED.

The next theorem shows that the return under the log-optimum investment strategy \( b^* \) is a well-defined quantity, even if the choice of \( b^* \) itself is ambiguous.

**Theorem 1.2** Suppose \( b_1^* \) and \( b_2^* \) are log-optimum portfolios in \( B^*(P) \). Then

\[
\rho(b_1^*, X) = \rho(b_2^*, X) \quad P\text{-a.s.}
\]

and consequently

\[
w(b_1^*, X) = w(b_2^*, X) \quad P\text{-a.s.}
\]

The return \( \rho(b^*, X) \) and the log return \( w(b^*, X) = \log \rho(b^*, X) \) under a log-optimum strategy \( b^* \in B^*(P) \) are therefore unambiguously defined.

**Proof:**

Theorem 1.1 and Jensen’s inequality imply that

\[
0 = E_P \left\{ \log \left( \frac{\rho(b_1^*, X)}{\rho(b_2^*, X)} \right) \right\} \leq \log \left( E_P \left\{ \frac{\rho(b_1^*, X)}{\rho(b_2^*, X)} \right\} \right) \leq \log 1 = 0.
\]
Since Jensen's inequality holds with equality and log is strictly concave, we may conclude that
\[
\frac{\rho(b_1^*, X)}{\rho(b_2^*, X)} = 1 \quad \text{or} \quad \rho(b_1^*, X) = \rho(b_2^*, X) \quad P\text{-a.s}
\]
QED.

A fixed point equation for log-optimum portfolios

We now generalize a result of Cover [84]. For any portfolio \( b(da) \) we define another portfolio \( b'(da) \) by the rule
\[
b'(da) = \gamma(a, b; P) b(da).
\]
Thus \( b'(da) \) is absolutely continuous with respect to \( b(da) \), with Radon-Nikodym derivative \( \gamma(a, b; P) \). Let \( D(b'\|b) \) denote the Kullback-Leibler divergence between \( b' \) and \( b \):
\[
D(b'\|b) = \int_A \log \left( \frac{dB'}{Db}(a) \right) \cdot b'(da) = \int_A \log \gamma(a, b; P) \cdot b'(da).
\]
It is well known that \( D(b'\|b) \geq 0 \), with equality iff \( b' = b \).

**Theorem 1.3** Choosing \( b' \) instead of \( b \) can only improve the expected log return:
\[
w(b', P) - w(b, P) \geq D(b'\|b) \geq 0.
\]
Thus any log-optimum portfolio \( b^*(da) \) must satisfy the fixed point equation
\[
b^*(da) = \gamma(a, b^*; P) b^*(da).
\]

**Proof:**

Observe that
\[
w(b', P) - w(b, P) = E_P \left\{ \log \left( \frac{\int_A \rho(a, X) b'(da)}{\int_A \rho(a, X) b(da)} \right) \right\}
\]
\[
= E_P \left\{ \log \left( \frac{\int_A \gamma(a, b; P) \cdot \rho(a, X) b(da)}{\int_A \rho(a, X) b(da)} \right) \right\}
\]
is the expected log of a quantity which itself is the expectation of \( \gamma(a, b; P) \) with respect to a random probability measure on \( A \). Jensen's inequality implies that
\[
\log \left( \int_A \gamma(a, b; P) \cdot \frac{\rho(a, X) b(da)}{\int_A \rho(a, X) b(da)} \right) \geq \int_A \log \gamma(a, b; P) \cdot \frac{\rho(a, X) b(da)}{\int_A \rho(a, X) b(da)}.
\]
Taking expectations \( E_P \{ \cdot \} \) yields

\[
    w(b', P) - w(b, P) \geq E_P \left\{ \int_{\mathcal{A}} \log \gamma(a, b; P) \cdot \frac{\rho(a, X) b(da)}{\int_{\mathcal{A}} \rho(a, X) b(da)} \right\} \\
    = \int_{\mathcal{A}} \log \gamma(a, b; P) \cdot E_P \left\{ \frac{\rho(a, X)}{\int_{\mathcal{A}} \rho(a, X) b(da)} \right\} \cdot b(da) \\
    = \int_{\mathcal{A}} \log \gamma(a, b; P) \cdot \gamma(a, b; P) \cdot b(da) \\
    = D(b' \| b) \geq 0.
\]

Any log-optimum portfolio \( b^* \) must satisfy the fixed point equation since otherwise the portfolio \( b^{*'}(da) = \gamma(a, b^*; P) b^*(da) \) would strictly improve the expected log return by a positive amount \( w(b^{'}, P) - w(b^*, P) \geq D(b^{'*} \| b^* ) > 0 \).

QED.

The fixed point equation expresses the fact that under log-optimum investment, the relative amount invested in various pieces of the space of elementary stocks \( \mathcal{A} \) must equal their expected relative contribution to the total return. In other words, it should be the case that

\[
b^*(A) = \int_{\mathcal{A}} \gamma(a, b^*; P) b^*(da) \\
= \int_{\mathcal{A}} E_P \left\{ \frac{\rho(a, X)}{\rho(b^*, X)} \right\} b^*(da) = E_P \left\{ \int_{\mathcal{A}} \rho(a, X) b^*(da) \right\}
\]

for all measurable \( \mathcal{A} \subseteq \mathcal{A} \). This expresses the idea of proportional betting.

Although necessary, the fixed point equation \( b^*(da) = \gamma(a, b^*; P) b^*(da) \) is usually not sufficient to guarantee log-optimality: it only expresses the requirement that \( b^* \) be log-optimum among those portfolio measures which are dominated by \( b^* \). However, it suggests an iterative procedure for computing log-optimum portfolios. Indeed, let \( b^{(0)} \) be a candidate portfolio to start with. We can then iteratively generate improvements \( b^{(1)}, b^{(2)}, \ldots, b^{(i)}, b^{(i+1)}, \ldots \) by the rule

\[
b^{(i+1)}(da) = \gamma(a, b^{(i)}; P) b^{(i)}(da)
\]

for all \( 0 \leq i < \infty \). Notice that \( w(b^{(i)}, P) \) will increase monotonically, since

\[
w(b^{(i+1)}, P) - w(b^{(i)}, P) \geq D(b^{(i+1)} \| b^{(i)}) \geq 0.
\]

If \( \mathcal{A} \) (and therefore also \( \mathcal{B} \)) is compact then the set of accumulation points of the sequence \( \{b^{(i)}\}_{0 \leq i < \infty} \) is a nonempty compact subset of \( \mathcal{B} \). We may hope that any such accumulation point is log-optimum or at least satisfies the fixed point
equation if the starting point \( b^{(0)} \) is chosen judiciously. Cover [84] demonstrates that any accumulation point of the sequence \( \{b^{(i)}\}_{0 \leq i < \infty} \) is log-optimum if \( A \) is finite and the initial portfolio measure \( b^{(0)} \) has full support. One may show that the accumulation point is unique using techniques of Csiszár and Tusnády [84].

Properties of \( w(b, P) \) and \( w^*(P) \)

We shall make progress by taking an abstract point of view and examining the maximum expected log return \( w^*(P) \) as a function of the distribution \( P \) on \( \mathcal{X} \).

Let \( \mathcal{P} \) be the space of probability measures on \( (\mathcal{X}, \mathcal{B}_X) \). We say that a sequence \( \{P_n\}_{0 \leq n < \infty} \subseteq \mathcal{P} \) converges weakly or in distribution to \( P \in \mathcal{P} \) if

\[
E_{P_n}\{f\} = \int_{\mathcal{X}} f \, dP_n \rightarrow E_P\{f\} = \int_{\mathcal{X}} f \, dP
\]

for all bounded continuous functions \( f : \mathcal{X} \rightarrow \mathcal{R} \). This notion of convergence defines the weak topology on \( \mathcal{P} \): the weakest topology such that the function \( P \mapsto E_P\{f\} \) is continuous in \( P \in \mathcal{P} \) for all bounded continuous functions \( f : \mathcal{X} \rightarrow \mathcal{R} \). The space \( \mathcal{P} \) is separable and metrizable (by the Prohorov metric) when endowed with this weak topology, and the Borel \( \sigma \)-field \( \mathcal{B}_P \) is the smallest \( \sigma \)-field on \( \mathcal{P} \) such that \( A \mapsto P(A) \) is measurable in \( P \in \mathcal{P} \) for all Borel sets \( A \) in \( \mathcal{B}_X \). We can embed \( \mathcal{X} \) as a subspace in \( \mathcal{P} \) by identifying a point \( x \in \mathcal{X} \) with the Dirac probability measure \( \delta_x \in \mathcal{P} \). Then \( x_n \rightarrow x \) in \( \mathcal{X} \) iff \( \delta_{x_n} \rightarrow \delta_x \) weakly in \( \mathcal{P} \), so that this is a homeomorphic embedding. If \( \mathcal{D} \) is a countable dense subset of \( \mathcal{X} \), then the set of finite linear combinations with rational coefficients

\[
\{ \sum_{n} a_n \delta_{x_n} : x_n \in \mathcal{D} \text{ and } a_n \text{ is rational for all } n \text{ in some finite set } (n) \}
\]

is a countable dense subset of \( \mathcal{P} \). If \( \mathcal{X} \) is compact, Polish, or analytic then \( \mathcal{P} \) is likewise.

Let \( w(b, x) = \log \rho(b, x) \) denote the log return. This function is defined and Borel measurable on \( \mathcal{B} \times \mathcal{X} \), and can be linearly extended to a function \( w(b, P) \) which is defined and Borel measurable on \( \mathcal{B} \times \mathcal{X} \):

\[
w(b, P) = E_P\{w(b, X)\} = E_P\{\log \rho(b, X)\}.
\]

The extension \( w(b, P) \) inherits many of the properties of the restriction \( w(b, x) \). For example, \( w(b, P) \) is concave in \( b \) since \( \rho(b, x) \) and also \( w(b, x) = \log \rho(b, x) \) is concave in \( b \in \mathcal{B} \).

**Theorem 1.4** If \( w(b, x) = \log \rho(b, x) \) is bounded below and lower semicontinuous (resp. bounded above and upper semicontinuous, resp. bounded and continuous), then \( w(b, P) = E_P\{w(b, X)\} = E_P\{\log \rho(b, X)\} \) is likewise.

21
Proof:
A function on a separable metrizable space is bounded below and lower semicontinuous iff it is the limit of an increasing sequence of bounded continuous functions. Now apply the monotone convergence theorem. See Bertsekas and Shreve [78], chapter 7 for more details.

QED.

Theorem 1.5 Let \( w^*(P) = \sup_{b \in B} w(b, P) \) denote the maximum expected log return.

1. If \( w : B \times \mathcal{P} \to \mathbb{R} \) is lower semicontinuous (or, more generally, if \( w(b, P) \) is lower semicontinuous in \( P \in \mathcal{P} \) for all fixed \( b \in B \)), then \( w^* : \mathcal{P} \to \mathbb{R} \) is lower semicontinuous.

2. If \( w : B \times \mathcal{P} \to \mathbb{R} \) is upper semicontinuous (resp. continuous) and \( B \) is compact then \( w^* : \mathcal{P} \to \mathbb{R} \) is upper semicontinuous (resp. continuous).

Proof:
To prove 1, it suffices to observe that the supremum of any family of lower semicontinuous functions is lower semicontinuous. To prove 2, let \( \mathcal{H} \) and \( \mathcal{H}^* \) denote the hypographs of the functions \( w : B \times \mathcal{P} \to \mathbb{R} \) and \( w^* : \mathcal{P} \to \mathbb{R} \), respectively:

\[
\mathcal{H} = \{(b, P, r) \in B \times \mathcal{P} \times \mathbb{R} : r \leq w(b, P)\}
\]

and

\[
\mathcal{H}^* = \{(P, r) \in \mathcal{P} \times \mathbb{R} : r \leq w^*(P)\}.
\]

If \( w : B \times \mathcal{P} \to \mathbb{R} \) is upper semicontinuous then \( \mathcal{H} \) is a closed subset of \( B \times \mathcal{P} \times \mathbb{R} \). It must be shown that \( w^* : \mathcal{P} \to \mathbb{R} \) is upper semicontinuous or equivalently that \( \mathcal{H}^* \) is a closed subset of \( B \times \mathcal{C} \), if \( B \) is compact. It suffices to prove the following assertion: if \( B \) and \( \mathcal{C} \) are separable metric spaces with \( B \) compact, \( \mathcal{H} \) is a closed subset of \( B \times \mathcal{C} \), and \( \mathcal{H}^* \) is the projection of \( \mathcal{H} \) on \( \mathcal{C} \), then \( \mathcal{H}^* \) is a closed subset of \( \mathcal{C} \). To prove this assertion let us consider a sequence \( \{c_n\}_n \subseteq \mathcal{H}^* \) such that \( c_n \to c \) in \( \mathcal{C} \). Then there exist \( \{b_n\}_n \subseteq B \) such that \( (b_n, c_n) \in \mathcal{H} \). By compactness of \( B \) there exists a subsequence \( \{b_{n_k}\}_k \) converging to an accumulation point \( b \in B \), and hence \( (b_{n_k}, c_{n_k}) \to (b, c) \). But \( \mathcal{H} \) is closed so that \( (b, c) \in \mathcal{H} \) and \( c \) is a point in the projection \( \mathcal{H}^* \). Thus \( \mathcal{H}^* \) is closed.

QED.
Measurable selections of log-optimum portfolios

It is an interesting question whether the maximum expected log return $w^*(P)$ can always be attained by some appropriate portfolio choice, that is, whether the set of log-optimum portfolios $B^*(P)$ is nonempty for all $P \in \mathcal{P}$. However, a related question turns out to be more relevant for our discussion, namely whether a log-optimum strategy $b^*(P)$ can be selected in such a way that $b^*(P)$ is a measurable function on $(\mathcal{P}, \mathcal{B}_\mathcal{P})$. This issue is intimately related to the continuity and measurability properties of the functions $w(b, P)$ and $w^*(P)$. For example, if a Borel measurable selection $b^* : (\mathcal{P}, \mathcal{B}_\mathcal{P}) \rightarrow (\mathcal{B}, \mathcal{B}_\mathcal{B})$ exists, then the maximum expected log return $w^* : (\mathcal{P}, \mathcal{B}_\mathcal{P}) \rightarrow (\mathcal{K}, \mathcal{B}_\mathcal{K})$ is Borel measurable since $w^*(P) = w(b^*(P), P)$ is the composition of Borel measurable functions $w : \mathcal{B} \times \mathcal{P} \rightarrow \mathcal{K}$ and $b^* : \mathcal{P} \rightarrow \mathcal{B}$. In general, we can be sure only that the map $w^*(P) = \sup_{b \in \mathcal{B}} w(b, P)$ is upper semi-analytic, even if $w(b, P)$ is Borel measurable on $\mathcal{B} \times \mathcal{P}$ and $\mathcal{B}$ and $\mathcal{P}$ are standard Borel spaces.

By associating the set of log-optimum portfolios $B^*(P)$ with every probability measure $P$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, we have defined a multivalued function $B^*$ from $\mathcal{P}$ to $\mathcal{B}$. $B^*$ is called a closed-valued correspondence if $B^*(P)$ is a nonempty closed subset of $\mathcal{B}$ for all $P \in \mathcal{P}$. We say that $B^*$ is a closed correspondence if $Gr(B^*)$ is a closed subset of the product space $\mathcal{P} \times \mathcal{B}$, where $Gr(B^*)$ is the graph of $B^*$:

$$Gr(B^*) = \{(b^*, P) \in \mathcal{B} \times \mathcal{P} : b^* \in B^*(P)\}.$$

In other words, $B^*$ is a closed correspondence if the assertions $P_n \rightarrow P$ in $\mathcal{P}$, $b^*_n \rightarrow b^*$ in $\mathcal{B}$, and $b^*_n \in B^*(P_n)$ for all $n$ together imply that $b^* \in B^*(P)$. If $B^*$ is a closed correspondence then $B^*$ is closed-valued, but the converse need not be true.

**Theorem 1.6** Suppose $\mathcal{B}$ is compact and $w : \mathcal{B} \times \mathcal{P} \rightarrow \mathcal{K}$ is upper semicontinuous, and let $w^*(P) = \sup_{b \in \mathcal{B}} w(b, P)$. Then $w^* : \mathcal{P} \rightarrow \mathcal{K}$ is upper semicontinuous, $B^*$ is a closed-valued correspondence from $\mathcal{P}$ to $\mathcal{B}$, and there exists a Borel measurable selection $b^* : (\mathcal{P}, \mathcal{B}_\mathcal{P}) \rightarrow (\mathcal{B}, \mathcal{B}_\mathcal{B})$ such that $b^*(P) \in B^*(P)$ for all $P \in \mathcal{P}$. If in addition $w^*(P)$ is continuous in $P \in \mathcal{P}$ then $B^*$ is a closed correspondence from $\mathcal{P}$ to $\mathcal{B}$. In particular, $w^*(P)$ will be continuous on $\mathcal{P}$ if $w(b, P)$ is continuous on $\mathcal{B} \times \mathcal{P}$.

**Proof:**

Any upper semicontinuous function defined on a compact set attains its supremum on a nonempty closed and therefore compact subset of its domain. Thus $B^*$ is a closed-valued (and even compact-valued) correspondence from $\mathcal{P}$ to $\mathcal{B}$. 

23
The measurable selection theorem of Kuratowski and Ryll-Nardzewski [61] implies the existence of a Borel measurable function \( b^* : (\mathcal{P}, \mathcal{B}_\mathcal{P}) \to (\mathcal{B}, \mathcal{B}_\mathcal{B}) \) such that \( b^*(P) \in B^*(P) \) for all \( P \in \mathcal{P} \). In fact, there exists a countable number of Borel measurable selections \( b^*_n : (\mathcal{P}, \mathcal{B}_\mathcal{P}) \to (\mathcal{B}, \mathcal{B}_\mathcal{B}) \) such that \( \{b^*_n(P)\} \) is dense in \( B^*(P) \) for all \( P \in \mathcal{P} \); this strengthening is due to Castaing and Rockafellar (see Klein and Thompson [84], theorem 14.4.1, p. 167 and notes, p. 168).

If \( w : \mathcal{B} \times \mathcal{P} \to \mathcal{R} \) is continuous then so is \( w^* : \mathcal{P} \to \mathcal{R} \), by Theorem 1.4. To show that the compact-valued correspondence \( B^* \) from \( \mathcal{P} \) to \( \mathcal{B} \) is closed if \( w^* : \mathcal{P} \to \mathcal{R} \) is continuous, we argue as follows. Assume that \( P_n \to P_\infty \) in \( \mathcal{P} \), \( b^*_n \in B^*(P_n) \) for all \( 0 \leq n < \infty \), and \( b^*_n \to b^*_\infty \) in \( \mathcal{B} \). We must show that \( b^*_\infty \in B^*(P_\infty) \). Consider the sequence

\[
w^*(P_n) = w(b^*_n, P_n).
\]

On the one hand, \( P_n \to P_\infty \) in \( \mathcal{P} \) and \( w^* \) is continuous on \( \mathcal{P} \) so that

\[
w^*(P_n) \to w^*(P_\infty).
\]

On the other hand, \( w(b, P) \) is upper semicontinuous on \( \mathcal{B} \times \mathcal{P} \), so that

\[
\limsup_{n \to -\infty} w(b^*_n, P_n) \leq w(\limsup_n b^*_n, \lim_n P_n) = w(b^*_\infty, P_\infty).
\]

Thus

\[
w(b^*_\infty, P_\infty) \geq \limsup_{n \to -\infty} w(b^*_n, P_n) = \lim_{n \to -\infty} w^*(P_n) = w^*(P_\infty) = \sup_{b \in \mathcal{B}} w(b, P_\infty)
\]

proving that \( b^*_\infty \in B^*(P_\infty) \).

QED.

A continuous selection of log-optimum portfolios \( b^*(P) \in B^*(P) \) seldom exists, even if \( \mathcal{B} \) and \( \mathcal{X} \) are compact and the log return \( w(b, x) = \log \rho(b, x) \) is bounded and continuous on \( \mathcal{B} \times \mathcal{X} \). Typically there exist sequences \( \{P_n\}_{0 \leq n < \infty} \) which converge, say \( P_n \to P_\infty \) in \( \mathcal{P} \), and yet any log-optimum portfolio choices \( b^*(P_n) \in B^*(P_n) \) keep chattering forever without converging to some element in \( B^*(P_\infty) \). However:

**Theorem 1.7** Assume that \( B^* \) is a closed correspondence from \( \mathcal{P} \) to \( \mathcal{B} \). If \( P_n \to P_\infty \) in \( \mathcal{P} \) and \( b^*_n \in B^*(P_n) \) for all \( 0 \leq n < \infty \), then any accumulation point of \( \{b^*_n\}_{0 \leq n < \infty} \) is a point in \( B^*(P_\infty) \). Consequently any selection \( b^* : \mathcal{P} \to \mathcal{B} \) with \( b^*(P) \in B^*(P) \) for all \( P \in \mathcal{P} \) is continuous at every point \( P \) where \( B^*(P) \) is a singleton, say \( B^*(P) = \{b^*(P)\} \).

24
Proof:

Obvious since $Gr(B^*) = \{(b^*, P) \in B \times P : b^* \in B^*(P)\}$ is a closed subset of $B \times P$ whose projection on $P$ is all of $P$.

QED.

A less ambitious goal may be easier to attain, namely the existence of a Borel measurable selection $b^* : (P, B_P) \to (B, B_b)$ with $b^*(P) \in B^*(P)$ for all $P \in P$ and such that $w(b^*(P), X) = \log \rho(b^*(P), X)$ is continuous in $P \in P$ for all possible outcomes $X \in \mathcal{X}$. The first step towards such a result was given in Theorem 1.2, where it was shown that $w(b^*(P), X)$ is unambiguously defined independent of the choice $b^*(P) \in B^*(P)$.

Theorem 1.8 Assume that $B$ is compact, that $B^*$ is a closed correspondence from $P$ to $B$, and that $w(b, x) = \log \rho(b, x)$ is continuous on $B \times \mathcal{X}$. If $P_n \to P_\infty$ in $P$ and $b^*_n \in B^*(P_n)$ for all $0 \leq n \leq \infty$, then

$$w(b^*_n, X) \to w(b^*_\infty, X).$$

Proof:

Indeed, if this is not the case, then we can derive a contradiction, as follows. There exists a subsequence $\{P_{n_k}\}_k$ such that $w(b^*_{n_k}, X)$ does not converge to $w(b^*_\infty, X)$. We may assume that the sequence $b^*_{n_k}$ converges to some point $b^* \in B^*(P_\infty)$ (if not, then we could select a sub-subsequence). We arrive at the sought-after contradiction by observing that $w(b^*, X) = w(b^*_\infty, X)$ by Theorem 1.2 whereas by construction

$$w(b^*_\infty, X) \neq \lim_k w(b_{n_k}, X) = w(\lim_k b^*_{n_k}, X) = w(b^*, X).$$

QED.

The existence of a Borel measurable selection of log-optimum portfolios requires strong assumptions of a topological flavor, such as compactness of $B$ and upper semicontinuity of $w(b, P)$. If we relax our requirements and only insist on the existence of a universally measurable selection $b^*(P) \in B^*(P)$ which nearly attains the supremum $w^*(P)$ (to within $\epsilon$), then we will have a much better chance for success.

Theorem 1.9 Suppose $B$ and $\mathcal{X}$ are analytic spaces and $w(b, P)$ is Borel measurable. Then $w^*(P) = \sup_{b \in B} E_P\{w(b, X)\}$ is upper semi-analytic on $P$. Furthermore, let $B^*(P)$ denote the set of elements $b^* \in B$ that attain the supremum $w^*(P)$. Then $\mathcal{D} = \{P \in P : B^*(P) \neq \emptyset\}$ is a universally measurable subset of
$\mathcal{P}$, and for every $\epsilon > 0$ there exists a universally measurable function $b^* : \mathcal{P} \rightarrow \mathcal{B}$ such that $b^*(P) \in B^*(P)$ for all $P \in \mathcal{D}$ and

$$w(b^*(P), P) \geq \begin{cases} w^*(P) - \epsilon & \text{if } w^*(P) < \infty \\ 1/\epsilon & \text{if } w^*(P) = \infty \end{cases} \text{ for all } P \in \mathcal{P}.$$ 

In particular if $w^*(P)$ is attained for all $P \in \mathcal{P}$, that is, $\mathcal{D} = \mathcal{P}$, or $B^*(P) \neq \emptyset$ for all $P \in \mathcal{P}$, then there exists a universally measurable selection $b^* : \mathcal{P} \rightarrow \mathcal{B}$ such that $b^*(P) \in B^*(P)$ and hence $w(b^*(P), P) = w^*(P)$ for all $P \in \mathcal{P}$.

**Proof:**

The proof is based on the measurable selection theorem of Jankov [41] and von Neumann [49]. See Bertsekas and Shreve [78], prop. 7.50, pp. 184–86.

**QED.**

From now on we shall adopt the existence of a Borel measurable selection of log-optimum portfolios $b^* : (\mathcal{P}, \mathcal{B}_\mathcal{P}) \rightarrow (\mathcal{B}, \mathcal{B}_\mathcal{B})$ as a working hypothesis. It will be argued that this assumption is satisfied in important cases where we apply the theory. In particular, we shall make use of Theorem 1.6 whose proof is based on the measurable selection theorem of Kuratowski and Ryll-Nardzewski [61]. The maximum expected log return $w^*(P) = w(b^*(P), P)$ is then also Borel measurable. In practice $\mathcal{B}$ and $\mathcal{X}$ are analytic spaces and it suffices to assume that $w^*(P) = \sup_{b \in \mathcal{B}} E_P\{\log \rho(b, X)\}$ is achieved for all distributions $P$. The measurable selection theorem of Jankov and von Neumann then implies the existence of a universally measurable selection of log-optimum portfolios $b^* : \mathcal{P} \rightarrow \mathcal{B}$ such that $b^*(P) \in B$ for all $P \in \mathcal{P}$. Existence of a universally measurable selection is sufficient for our purposes if the underlying probability space (on which the random outcome $X$ is defined) is complete.
SECTION 2.

ASYMPTOTIC OPTIMALITY OF
SEQUENTIAL LOG-OPTIMUM
INVESTMENT

In this section we introduce the sequential investment problem and prove the
asymptotic optimality principle for log-optimum investment.

Let us first augment the basic framework of section 1 so that we obtain a
sequential investment problem. The random variable \( X : (\Omega, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{B}_X) \) is
defined on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) which is equipped with a measurable
transformation \( T : (\Omega, \mathcal{F}) \rightarrow (\Omega, \mathcal{F}) \), and we define \( X_t = X \circ T^t \) for \( 0 \leq t < \infty \).
We interpret \( X_t(\omega) = X(T^t\omega) \) as the outcome during period \( t \).

The investor must select an action \( b_t \) based on currently available information
at the beginning of each period \( t \). Let \( \mathcal{F}_t \) be the \( \sigma \)-field that represents the data
observable at the beginning of period \( t \). A most important case, one which
the reader is urged to keep in mind for intuitive motivation, is when \( \mathcal{F}_t \) is the
information contained in past outcomes:

\[
\mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1}).
\]

If in each period \( t \) we can observe only a random variable \( Y_t \) which is correlated
with \( X_t \) (e.g., a function \( \varphi(X_t) \) of the outcome \( X_t \)), then the information fields
would be given by

\[
\mathcal{F}_t = \sigma(Y_0, Y_1, \ldots, Y_{t-1}).
\]

Other choices for the information fields \( \{\mathcal{F}_t\}_{0 \leq t < \infty} \) may occasionally be useful.
We require that \( b_t \) be \( \mathcal{F}_t \)-measurable, that is, \( b_t : (\Omega, \mathcal{F}_t) \rightarrow (\mathcal{B}, \mathcal{B}_B) \), or \( b_t \in \mathcal{F}_t \)
for short. Invested capital will grow by a multiplicative factor \( \rho(b_t, X_t) \) when the
random outcome \( X_t \) is revealed at the end of period \( t \). If we normalize the initial
capital to \( S_0 = 1 \) for convenience, then the capital \( S_n \) after \( n \) periods is given by
the product of the factors with which it has been compounded previously:

\[
S_n = \prod_{0 \leq t < n} \rho(b_t, X_t).
\]

We will assume that the underlying probability space \((\Omega, \mathcal{F}, \mathbb{P})\) is perfect. A
theorem of Jirina [54] then implies that \( X_t \) admits a regular conditional prob-
bability distribution given \( \mathcal{F}_t \) for all \( 0 \leq t < \infty \). Let \( P_t \) denote such a regular
conditional probability distribution, that is, a measurable function \( P_t : (\Omega, \mathcal{F}_t) \to (\mathcal{P}, \mathcal{B}_p) \) such that conditional expectations are given by
\[
E \{ f(X_t) | \mathcal{F}_t \} = \int_X f(x_t) P_t(dx_t).
\]
If the investor selects the portfolio \( b_t^* = b^*(P_t) \) then he will attain the maximum conditional expected log return
\[
w_t^* = \sup_{b \in \mathcal{B}} E \{ \log \rho(b, X_t) | \mathcal{F}_t \} = w^*(P_t) = E \{ \log \rho(b_t^*, X_t) | \mathcal{F}_t \}.
\]
Clearly \( b_t^* = b^*(P_t) \) is \( \mathcal{F}_t \)-measurable since \( b_t^* \) is the composition of Borel measurable functions:

![Diagram](image)

Likewise \( w_t^* = w^*(P_t) = w(b^*(P_t), P_t) \) is \( \mathcal{F}_t \)-measurable.

Let \( w_t^* = \log \rho(b_t^*, X_t) \) denote the log return for period \( t \) that is actually realized under the log-optimum strategy \( b_t^* \). Notice that the conditional expectation of the true log return \( w_t^* \) is precisely the maximum conditional expected log return \( w_t^* \):
\[
w_t^* = E \{ w_t^* | \mathcal{F}_t \} = E \{ \log \rho(b_t^*, X_t) | \mathcal{F}_t \}.
\]
Finally, let \( W_t^* \) denote the maximum expected log return for period \( t \) under decisions based on the information field \( \mathcal{F}_t \):
\[
W_t^* = \sup_{b \in \mathcal{B}} E \{ \log \rho(b, X_t) \}.
\]
Thus \( W_t^* \) is the expectation of both the true log return and its conditional expectation:
\[
W_t^* = E \{ w_t^* \} = E \{ w_t^* \} = E \{ \log \rho(b_t^*, X_t) \}.
\]
It will be argued that the log-optimum strategy \( \{ b_t^* \}_{0 \leq t < \infty} \) is optimum in the long run by a variety of criteria. Let \( S_t^* \) and \( S_t \) be the capital after \( n \) rounds of investment according to the log-optimum strategy \( \{ b_t^* \}_{0 \leq t < \infty} \) and some competing policy \( \{ b_t \}_{0 \leq t < \infty} \), respectively:
\[
S_n^* = \prod_{0 \leq t < n} \rho(b_t^*, X_t) \quad \text{and} \quad S_n = \prod_{0 \leq t < n} \rho(b_t, X_t).
\]
Theorem 2.1 The log-optimum strategy is better than any competitor in the following sense:

\[ E \left\{ \frac{S_n}{S_n^*} \right\} \leq 1 \quad \text{and} \quad E \left\{ \frac{S_n^*}{S_n} \right\} \geq 1 \quad \text{for all} \quad 0 \leq n < \infty. \]

Proof:

Consider the set of tuples \( \{b_t\}_{0 \leq t < n} \) with \( b_t : (\Omega, \mathcal{F}_t) \rightarrow (\mathcal{B}, \mathcal{B}_\mathcal{B}) \) for \( 0 \leq t < n \). This set is convex, and the function

\[ \{b_t\}_{0 \leq t < n} \mapsto \prod_{0 \leq t < n} \rho(b_t, x_t) \]

is concave for all fixed \( \{x_t\}_{0 \leq t < n} \). The Kuhn-Tucker conditions for log-optimality (see Theorem 1.1) imply that

\[ E \left\{ \log \left( \frac{S_n}{S_n^*} \right) \right\} = E \left\{ \log \left( \prod_{0 \leq t < n} \frac{\rho(b_t, X_t)}{\rho(b_t^*, X_t)} \right) \right\} \leq 0 \]

for all \( \{b_t\}_{0 \leq t < n} \) iff

\[ E \left\{ \frac{S_n}{S_n^*} \right\} = E \left\{ \frac{\prod_{0 \leq t < n} \rho(b_t, X_t)}{\prod_{0 \leq t < n} \rho(b_t^*, X_t)} \right\} \leq 1 \]

for all \( \{b_t\}_{0 \leq t < n} \). But

\[ E \left\{ \log \left( \frac{S_n}{S_n^*} \right) \right\} = E \left\{ \sum_{0 \leq t < n} E \left\{ \log \left( \frac{\rho(b_t, X_t)}{\rho(b_t^*, X_t)} \right) \bigg| \mathcal{F}_t \right\} \right\} \leq 0 \]

since

\[ E \left\{ \log \left( \frac{\rho(b_t, X_t)}{\rho(b_t^*, X_t)} \right) \bigg| \mathcal{F}_t \right\} \leq 0. \]

This proves the first part. The second part follows from Jensen's inequality:

\[ \log E \left\{ \left( \frac{S_n^*}{S_n} \right) \right\} \geq E \left\{ \log \left( \frac{S_n^*}{S_n} \right) \right\} \geq 0. \]

QED.

We now prove a crucial statement, namely the asymptotic optimality of the log-optimum investment strategy. This result is not only very important for its own sake, but it will also be the key to our proof of the asymptotic equipartition property for a stationary ergodic stock market in section 4.

Theorem 2.2 (Asymptotic Optimality Principle)

\[ \limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) \leq 0 \quad \text{a.s.} \]
PROOF:

We prove the following lemma: if \( \{Y_n\}_{n \leq n < \infty} \) is a sequence of nonnegative random variables with expectations \( \mathbb{E}\{Y_n\} \leq 1 \), then

\[
\limsup_{n \to \infty} \frac{1}{n} \log Y_n \leq 0.
\]

Indeed, the Markov inequality implies that

\[
P\{Y_n \geq r\} \leq \frac{1}{r} \mathbb{E}\{Y_n\} \leq \frac{1}{r} \quad \text{if} \quad r \geq 1.
\]

Using \( \epsilon = (1/n) \log r > 0 \), we can write

\[
P\left\{ \frac{1}{n} \log Y_n \geq \epsilon \right\} \leq \exp(-n\epsilon)
\]

and hence

\[
\sum_{0 \leq n < \infty} P\left\{ \frac{1}{n} \log Y_n \geq \epsilon \right\} \leq \sum_{0 \leq n < \infty} \exp(-n\epsilon) < \infty.
\]

The Borel-Cantelli lemma implies

\[
P\left\{ \frac{1}{n} \log Y_n \geq \epsilon \quad \text{infinitely often} \right\} = 0.
\]

Since \( \epsilon > 0 \) was arbitrary we may conclude that

\[
\limsup_{n \to \infty} \frac{1}{n} \log Y_n \leq 0 \quad \text{a.s.}
\]

To prove the theorem we apply the lemma to \( Y_n = S_n/S_n^* \). We know that \( \mathbb{E}\{S_n/S_n^*\} \leq 1 \), from Theorem 2.1.

QED.

Remark 1. Theorem 2.2 may be interpreted as a game-theoretic result. Consider two players \( i = 1, 2 \). The permissible actions for player \( i \) are sequential strategies \( \{b^{(i)}_t\}_{0 \leq t < \infty} \) with \( b^{(i)}_t : (\Omega, \mathcal{F}) \to (B, B_\beta) \) for all \( 0 \leq t < \infty \). Let \( S^{(i)}_n = \prod_{0 \leq t < n} \rho(b^{(i)}_t, X_t) \) be the capital of player \( i \) after \( n \) rounds. Player \( i \) will win if his capital \( S^{(i)}_n \) grows with a higher exponential rate in the long run than the other player’s capital. Specifically, the theorem asserts that the game with payoff

\[
\limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S^{(1)}_n}{S^{(2)}_n} \right)
\]

to player 1 has a value (namely the value 0, like any zero-sum game), and that the log-optimum strategy \( \{b^{(i)}_t\}_{0 \leq t < \infty} \) is an optimum strategy for both players.
Thus log-optimum investment is competitively optimum in the long run. Notice however that the optimum strategy $\{b_t^*\}_{0 \leq t < \infty}$ is in general not unique, since any strategy $\{b_t'\}_{0 \leq t < \infty}$ obtained from $\{b_t^*\}_{0 \leq t < \infty}$ by modifying $b_t^*$ for an arbitrarily large but finite number of periods $t$ gives rise to the same payoff, no matter what the opponent does.

**Remark 2.** Suppose $\{\mathcal{F}_t\}_{0 \leq t < \infty}$ is a sequence of information fields which are less refined than the corresponding fields in the sequence $\{\mathcal{F}_t\}_{0 \leq t < \infty}$, i.e., $\mathcal{F}_t' \subseteq \mathcal{F}_t$ for all $0 \leq t < \infty$. The log-optimum $\mathcal{F}_t'$-measurable portfolio, let's call it $b_t^*$, is also $\mathcal{F}_t$-measurable and competes with the log-optimum portfolio $b_t^*$ which is the best choice based on the richer field $\mathcal{F}_t$. The corresponding capital return $S_{t \uparrow}^* = \prod_{0 \leq t < n} \rho(b_{t \uparrow}^*, X_t)$ therefore satisfies

$$\limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_{t \uparrow}'}{S_{t \uparrow}^*} \right) \leq 0 \quad \text{a.s.}$$

**Remark 3.** If $\lim_{n \to \infty} (1/n) \log S_n^*$ exists, then Theorem 2.2 is equivalent to

$$\limsup_{n \to \infty} \frac{1}{n} \log S_n \leq \lim_{n \to \infty} \frac{1}{n} \log S_n^* \quad \text{a.s.}$$

Similarly, if $(1/n) \log S_n$ converges, then we can write the statement in the form

$$\lim_{n \to \infty} \frac{1}{n} \log S_n \leq \liminf_{n \to \infty} \frac{1}{n} \log S_n^* \quad \text{a.s.}$$

Thus log-optimum investment is truly optimum in the long run since it leads to the highest possible rate of exponential growth.

Although we have shown that $E \{S_n/S_n^* \} \leq 1$ for all $0 \leq n < \infty$, it is still an open question whether $\{E \{S_n/S_n^* \}\}_{0 \leq n < \infty}$ is a decreasing sequence. In the next theorem we resolve this question and derive the even stronger conclusion that $S_n/S_n^*$ is a nonnegative supermartingale, under an additional hypothesis. A similar theorem was proved by Breiman [61], Breiman [62], and Finkelstein and Whitley [81].

**Theorem 2.3** Suppose the sequence of information fields $\{\mathcal{F}_t\}_{0 \leq t < \infty}$ is monotone increasing, and

$$\sigma(X_0, X_1, \ldots, X_{t-1}) \subseteq \mathcal{F}_t \quad \text{for all } 0 \leq t < \infty.$$

(In particular, the sequence $\mathcal{F}_t = \sigma(X_0, \ldots, X_{t-1})$ satisfies this condition.) Let furthermore $\{\mathcal{G}_t\}_{0 \leq t < \infty}$ be any monotone increasing sequence of $\sigma$-fields satisfying

$$\mathcal{F}_{t-1} \vee \sigma(X_0, X_1, \ldots, X_{t-1}) \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t \quad \text{for all } 0 \leq t < \infty.$$
(For example, $\mathcal{G}_t = \mathcal{F}_{t-1} \vee \sigma(X_{t-1})$ or $\mathcal{G}_t = \mathcal{F}_t$ for all $0 \leq t < \infty$ are satisfactory choices.) Then

$$\left\{ \frac{S_n}{S_n^*}, \mathcal{G}_n \right\}_{0 \leq n < \infty}$$

is a nonnegative supermartingale.

**Proof:**

Clearly the ratio $\frac{S_n}{S_n^*} = \prod_{0 \leq t < n} \frac{\rho(b_t, X_t)}{\rho(b_t^*, X_t)}$ is $\mathcal{F}_{t-1} \vee \sigma(X_{t-1})$-measurable, and hence $\mathcal{G}_n$-measurable for all $n \geq 1$ since $\mathcal{G}_n$ is a larger $\sigma$-field. Therefore

$$E \left\{ \frac{S_{n+1}}{S_{n+1}^*} \mid \mathcal{G}_n \right\} = E \left\{ \frac{S_n}{S_n^*} \cdot \frac{\rho(b_n, X_n)}{\rho(b_n^*, X_n)} \mid \mathcal{G}_n \right\}$$

$$= \frac{S_n}{S_n^*} \cdot E \left\{ \frac{\rho(b_n, X_n)}{\rho(b_n^*, X_n)} \mid \mathcal{G}_n \right\} \text{ since } \frac{S_n}{S_n^*} \text{ is } \mathcal{G}_n\text{-measurable}$$

$$= \frac{S_n}{S_n^*} \cdot E \left\{ E \left\{ \frac{\rho(b_n, X_n)}{\rho(b_n^*, X_n)} \mid \mathcal{F}_n \right\} \mid \mathcal{G}_n \right\} \text{ since } \mathcal{G}_n \subseteq \mathcal{F}_n$$

$$\leq \frac{S_n}{S_n^*} \text{ since } E \left\{ \frac{\rho(b_n, X_n)}{\rho(b_n^*, X_n)} \mid \mathcal{F}_n \right\} \leq 1.$$

The last inequality follows from a conditional version of the Kuhn-Tucker conditions for log-optimality. We conclude that $\{S_n/S_n^*, \mathcal{G}_n\}_{0 \leq n < \infty}$ is a nonnegative supermartingale.

QED.

**Remark 1.** Since $\mathcal{F}_{t-1} \vee \sigma(X_{t-1}) \subseteq \mathcal{F}_t$ for all $0 < t < \infty$, $E \{S_n/S_n^*\}$ can be written as the expectation of a nested product of conditional expectations:

$$E \left\{ \frac{S_n}{S_n^*} \right\} = E \left\{ \prod_{0 \leq t < n} \frac{\rho(b_t, X_t)}{\rho(b_t^*, X_t)} \right\} =$$

$$E \left\{ E \left\{ \frac{\rho(b_0, X_0)}{\rho(b_0^*, X_0)} \cdot E \left\{ \frac{\rho(b_1, X_1)}{\rho(b_1^*, X_1)} \cdot \ldots \cdot E \left\{ \frac{\rho(b_{n-1}, X_{n-1})}{\rho(b_{n-1}^*, X_{n-1})} \mid \mathcal{F}_{n-1} \right\} \mid \mathcal{F}_1 \right\} \mid \mathcal{F}_0 \right\} \right\}.$$

**Remark 2.** Any nonnegative supermartingale converges a.s. and the expectations decrease monotonically to a limit which is no smaller than the expectation of the limit, by Fatou's lemma. Thus there exists a random variable $Y$ with $Y \geq 0$, $E \{Y\} \leq 1$ and

$$1 = \frac{S_0}{S_0^*} = E \left\{ \frac{S_0}{S_0^*} \right\} \geq \ldots \geq E \left\{ \frac{S_n}{S_n^*} \right\} \geq \ldots \geq \lim_{n \to \infty} E \left\{ \frac{S_n}{S_n^*} \right\} \geq E \{Y\} \geq 0.$$

**Remark 3.** If the assumptions of the last theorem are satisfied, then there is a simpler proof of the asymptotic optimality principle. Indeed, let $Y =$
\[ \lim_{n \to \infty} S_n/S_n^* \] Then \( Y \) is well-defined and \( 0 \leq Y < \infty \) a.s. since \( \mathbb{E}\{Y\} \leq 1 \).

On \( \{Y \neq 0\} \), we clearly have

\[ \lim_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) = \lim_{n \to \infty} \frac{1}{n} \log Y = 0. \]

If \( Y = 0 \), then \( S_n/S_n^* \leq \epsilon \) eventually for large \( n \) and any fixed but arbitrarily chosen \( \epsilon > 0 \), so that

\[ \limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) \leq \limsup_{n \to \infty} \frac{1}{n} \log \epsilon = 0. \]

Thus we may conclude in any case that

\[ \limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) \leq 0 \quad \text{a.s.} \]

**Remark 4.** If \( \varphi : \mathcal{R}_+ \to \mathcal{R}_+ \) is concave and increasing then \( \{\varphi(S_n/S_n^*), \mathcal{G}_n\}_{0 \leq n < \infty} \) is also a nonnegative supermartingale, by Jensen's inequality. It converges to \( \varphi(Y) \) if \( \varphi \) is continuous. Similarly, if \( \varphi \) is convex and monotone decreasing then \( \{\varphi(S_n/S_n^*), \mathcal{G}_n\}_{0 \leq n < \infty} \) is a nonnegative submartingale (strictly speaking only if the expectations \( \mathbb{E}\{\varphi(S_n/S_n^*)\} \) are finite). In particular, \( \varphi(r) = 1/r \) is such a function and \( \mathbb{E}\{S_n^*/S_n\} \) is monotone increasing:

\[ 1 = \frac{S_0^*}{S_0} = \mathbb{E}\{S_0^*\} \leq \mathbb{E}\{S_1^*/S_0\} \leq \ldots \leq \mathbb{E}\{S_n^*/S_n\} \leq \ldots \]
SECTION 3.

A SUBMARTINGALE PROPERTY:
MORE INFORMATION DOESN’T HURT

In the present section we again consider log-optimum investment for a single period. This time we examine how well investors can perform in terms of expected log return when given increasing amounts of data on which to base their decisions. Obviously the maximum (unconditional) expected log return cannot decrease because there will be more measurable portfolios to choose from so that the maximum can be taken over a larger set. But the sequence of maximum conditional expected log returns is in general not monotone increasing; it is only a submartingale. Non-monotonicity means that ignoring information may occasionally be beneficial. Thus ignorance can be bliss: information may be misleading and suggest the wrong course of action. Of course, this can happen only infrequently, because more information does not hurt on the average. Investors are in general willing to pay in advance for the privilege of inspecting data that decreases the uncertainty of the outcome, fully aware that they may regret this in the future when they see more clarifying evidence or when they see the real truth. More information has no value only if it and the outcome are conditionally independent given the current state of knowledge.

The outcome is a random variable $X : (\Omega, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{B}_\mathcal{X})$ defined on a perfect probability space $(\Omega, \mathcal{F}, P)$. Thus the distribution $P$ of $X$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ is the image measure of $P$ via $X$, or $P(B) = P\{X \in B\} = P(X^{-1}(B))$ for all Borel sets $B \in \mathcal{B}_\mathcal{X}$. $X$ admits a regular conditional probability distribution (r.c.p.d.) given any sub-$\sigma$-field $\mathcal{G}$ of $\mathcal{F}$ (see Jiřina [54] or Sokal [81]). In particular, if $\mathcal{G} = \mathcal{F}$ then we have full information or perfect foresight and this r.c.p.d. is the Dirac probability measure $\delta_X$ concentrated at the random outcome $X \in \mathcal{X}$. On the other hand, the distribution $P$ of $X$ is the r.c.p.d. of $X$ given the trivial $\sigma$-field $\{\emptyset, \Omega\}$ (case of no information).

Let us consider a monotone nondecreasing sequence of information fields $\{\mathcal{F}_n\}_{0 \leq n < \infty}$ with limiting $\sigma$-field $\mathcal{F}_\infty$:

$$\mathcal{F}_n \uparrow \mathcal{F}_\infty = \bigvee_{0 \leq n < \infty} \mathcal{F}_n = \sigma(\bigcup_{0 \leq n < \infty} \mathcal{F}_n).$$

Let $\tilde{\delta}_n$ be a log-optimum portfolio based on $\mathcal{F}_n$, and $\bar{\omega}_n$ and $\bar{W}_n$ the maximum conditional and unconditional expected log returns under decisions based on $\mathcal{F}_n$:

$$\bar{\omega}_n = \sup_{\tilde{\delta}_n} \mathbb{E}\left\{\log \rho(b, X) \mid \mathcal{F}_n\right\} = \mathbb{E}\left\{\log \rho(\tilde{\delta}_n, X) \mid \mathcal{F}_n\right\}$$

$$\bar{W}_n = \sup_{\{} \mathbb{E}\left\{\log \rho(b, X) \mid \mathcal{F}_n\right\}.$$
\[ W_0^* = \sup_{b \in \mathcal{F}_n} \mathbb{E}\{\log \rho(b, X)\} = \mathbb{E}\{\bar{w}_n^*\} = \mathbb{E}\{\log \rho(\bar{b}_n, X)\}. \]

It is clear that \( \bar{b}_n = b^* (P_n) \) and \( \bar{w}_n^* = w^* (\bar{b}_n, P_n) \) where \( P_n \) is a r.c.p.d. of \( X \) given \( \mathcal{F}_n \). The following theorem illustrates the intuitive principle that more information seldom hurts, and never on the average.

**Theorem 3.1** Assume that \( \mathcal{F}_n \uparrow \mathcal{F}_\infty \). Then

1. The maximum expected log returns are monotonically nondecreasing:
   \[ W_0^* \leq W_1^* \leq \ldots \leq W_n^* \leq \ldots \leq W_\infty^*. \]

2. \( \{ar{w}_n^*, \mathcal{F}_n\}_{0 \leq n < \infty} \) is a submartingale, at least if \( \bar{W}_n^* < \infty \) for all \( 0 \leq n < \infty \). In any case,
   \[ \limsup_{n \to \infty} \bar{w}_n^* \leq \bar{w}_\infty^* \text{ a.s.} \]

3. If a stronger condition holds, namely \( \sup_{0 \leq n < \infty} \mathbb{E}\{\bar{w}_n^{**}\} < \infty \), then the submartingale \( \{ar{w}_n^*, \mathcal{F}_n\}_{0 \leq n < \infty} \) converges a.s. towards a limit satisfying
   \[ \lim_{n \to \infty} \bar{w}_n^* \leq \bar{w}_\infty^* \text{ a.s.} \]

This stronger condition will hold if \( \bar{W}_\infty^* < \infty \), and \( \{ar{w}_n^*, \mathcal{F}_n\}_{0 \leq n \leq \infty} \) is a submartingale in that case. There is not only convergence a.s. but also in \( L^1 \) if \( \{\bar{w}_n^*\}_{0 \leq n < \infty} \) is uniformly integrable. Uniform integrability holds in particular if \( \{\bar{w}_n^*\}_{0 \leq n < \infty} \) is \( L^1 \)-dominated, that is, if
\[ \mathbb{E}\{\sup_{0 \leq n < \infty} |\bar{w}_n^*|\} < \infty. \]

4. If \( \bar{W}_n^* \uparrow \bar{W}_\infty^* < \infty \) then \( \bar{w}_n^* \to \bar{w}_\infty^* \) a.s.

**Proof:**

Clearly \( W_n^* = \sup_{b \in \mathcal{F}_n} \mathbb{E}\{\log \rho(b, X)\} \) increases as \( n \) increases, since the supremum can then be taken over a larger set. This proves assertion 1. To prove 2, first observe that \( \bar{b}_n = b^* (P_n) \) and \( \bar{w}_n^* = w^* (\bar{b}_n, P_n) \) are \( \mathcal{F}_n \)-measurable. Hence to prove that \( \{ar{w}_n^*, \mathcal{F}_n\}_{0 \leq n \leq \infty} \) is a submartingale (at least in case \( \bar{W}_\infty^* < \infty \)), it suffices to show that \( \bar{w}_m \leq \mathbb{E}\{\bar{w}_n^*|\mathcal{F}_m\} \) for \( 0 \leq m < n \leq \infty \). But
\[ \sup_{b \in \mathcal{F}_m} \mathbb{E}\{\log \rho(b, X)|\mathcal{F}_m\} \leq \bar{w}_n^* = \sup_{b \in \mathcal{F}_n} \mathbb{E}\{\log \rho(b, X)|\mathcal{F}_n\} \]

35
since any $\mathcal{F}_m$-measurable portfolio is $\mathcal{F}_n$-measurable (i.e., $b \in \mathcal{F}_m$ implies that $b \in \mathcal{F}_n$). Taking conditional expectations given $\mathcal{F}_m$, we obtain

\[
\bar{w}_n^* = \sup_{b \in \mathcal{F}_m} \mathbb{E} \{ \log \rho(b, X) | \mathcal{F}_m \}
\]

\[
= \sup_{b \in \mathcal{F}_m} \mathbb{E} \left\{ \mathbb{E} \left\{ \log \rho(b, X) | \mathcal{F}_n \right\} | \mathcal{F}_m \right\}
\]

\[
\leq \sup_{b \in \mathcal{F}_n} \mathbb{E} \left\{ \mathbb{E} \left\{ \log \rho(b, X) | \mathcal{F}_n \right\} | \mathcal{F}_m \right\}
\]

\[
\leq \mathbb{E} \left\{ \sup_{b \in \mathcal{F}_n} \mathbb{E} \left\{ \log \rho(b, X) | \mathcal{F}_n \right\} | \mathcal{F}_m \right\} = \mathbb{E} \left\{ \bar{w}_n^* | \mathcal{F}_m \right\}.
\]

Since

\[
\bar{w}_n^* \leq \mathbb{E} \left\{ \bar{w}_\infty^* | \mathcal{F}_n \right\} \quad \text{and} \quad \mathbb{E} \left\{ \bar{w}_\infty^* | \mathcal{F}_n \right\} \to \mathbb{E} \left\{ \bar{w}_\infty^* | \mathcal{F}_\infty \right\} = \bar{w}_\infty^* \quad \text{a.s.}
\]

by Lévy's theorem on convergence of conditional expectations we may conclude that

\[
\limsup_{n \to \infty} \bar{w}_n^* \leq \bar{w}_\infty^*.
\]

The claims under 3 are general results about submartingales (see Neveu [70], pp. 131, 133). Finally, to show 4, observe that

\[
\bar{w}_\infty^* - \limsup_{n \to \infty} \bar{w}_n^* = \liminf_{n \to \infty} \left[ \mathbb{E} \left\{ \bar{w}_\infty^* | \mathcal{F}_n \right\} - \bar{w}_n^* \right] \geq 0 \quad \text{a.s.}
\]

The quantity on the left hand side cannot be strictly positive with positive probability. Indeed, Fatou's lemma for sequences of nonnegative random variables gives

\[
0 = \bar{W}_\infty^* - \lim_{n \to \infty} \bar{W}_n^* = \liminf_{n \to \infty} \mathbb{E} \left\{ \bar{w}_\infty^* | \mathcal{F}_n \right\} - \bar{w}_n^*
\]

\[
\geq \mathbb{E} \left\{ \bar{w}_\infty^* - \limsup_{n \to \infty} \bar{w}_n^* \right\} \geq 0.
\]

QED.

The second issue discussed in this section is a continuity question. Given an increasing sequence of information fields, do the maximum expected log returns increase monotonically to the maximum expected log return given the limiting $\sigma$-field, or is there perhaps a gap between the limit of the expectations and the expectation of the limit? The previous theorem shows only that

\[
\bar{W}_n^* \not\to \bar{W}_\infty^* \quad \text{and} \quad \limsup_{n \to \infty} \bar{w}_n^* \leq \bar{w}_\infty^* \quad \text{a.s.}
\]

But in later sections we will have to assume that $\bar{W}_n^* \to \bar{W}_\infty^*$ and hence (at least if $\bar{W}_\infty^* < \infty$) $\bar{w}_n^* \to \bar{w}_\infty^*$ a.s. In the next theorem we exhibit a condition which guarantees the absence of a gap between $\lim_n \bar{W}_n^*$ and $\bar{W}_\infty^*$: it suffices that the
maximum expected log return $w^*(P)$ is nonnegative and lower semicontinuous as a function of the distribution $P$ of the outcome.

**Theorem 3.2** Suppose $\mathcal{F}_n \to \mathcal{F}_\infty$ and $\tilde{P}_n$ is a regular conditional probability distribution of $X$ given $\mathcal{F}_n$, for $0 \leq n \leq \infty$. Then $\tilde{P}_n \to \tilde{P}_\infty$ weakly a.s. If the maximum expected log return $w^*(P) = \sup_{b \in B} \mathbb{E}\{\log \rho(b, X)\}$ is lower semicontinuous then $\tilde{w}_n^* \to \tilde{w}_\infty^*$ a.s. If moreover $w^*: \mathcal{P} \to \mathcal{R}$ is bounded below, then also $\tilde{W}_n^* \to \tilde{W}_\infty^*$, and there is convergence in $L^1$ if $\tilde{W}_\infty^* < \infty$. In particular, if $w^*(P)$ is bounded and continuous in $P \in \mathcal{P}$, then $\tilde{w}_n^* \to \tilde{w}_\infty^*$ a.s. and in $L^1$, and $\tilde{W}_n^* \to \tilde{W}_\infty^*$.

**Proof:**

If $f: \mathcal{X} \to \mathcal{R}$ is integrable then $\mathbb{E}\{f|\mathcal{F}_n\} \to \mathbb{E}\{f|\mathcal{F}_\infty\}$ a.s. by Lévy's theorem for convergence of conditional expectations. But

$$\mathbb{E}\{f|\mathcal{F}_n\} = \int_{\mathcal{X}} f \cdot d\tilde{P}_n \quad \text{for } 0 \leq n \leq \infty$$

so that

$$\int_{\mathcal{X}} f \cdot d\tilde{P}_n \to \int_{\mathcal{X}} f \cdot d\tilde{P}_\infty \quad \text{a.s.}$$

for all bounded continuous $f: \mathcal{X} \to \mathcal{R}$. This proves that $\tilde{P}_n \to \tilde{P}_\infty$ weakly a.s., and it follows that $\tilde{w}_n^* \to \tilde{w}_\infty^*$ a.s. if $w^*: \mathcal{P} \to \mathcal{R}$ is continuous, since $\tilde{w}_n^* = w^*(\tilde{P}_n)$.

If $w^*: \mathcal{P} \to \mathcal{R}$ is lower semicontinuous then

$$\liminf_{n \to \infty} \tilde{w}_n^* = \liminf_{n \to \infty} w^*(\tilde{P}_n) \geq w^*(\tilde{P}_\infty) = \tilde{w}_\infty^*.$$

On the other hand, part 2 of Theorem 3.1 implies that

$$\limsup_{n \to \infty} \tilde{w}_n^* \leq \tilde{w}_\infty^* \quad \text{a.s.}$$

Thus there will actually be equality:

$$\lim_{n \to \infty} \tilde{w}_n^* = \tilde{w}_\infty^* \quad \text{a.s.}$$

If $\{\bar{w}_n\}_{0 \leq n < \infty}$ is also bounded below by an integrable function then Fatou's lemma gives

$$\tilde{W}_\infty^* = \mathbb{E}\{\bar{w}_\infty^*\} = \mathbb{E}\left\{\lim_{n} \bar{w}_n^*\right\} \leq \lim_{n \to \infty} \mathbb{E}\{\bar{w}_n^*\} = \lim_{n \to \infty} \tilde{W}_n^* \leq \tilde{W}_\infty^*$$

so that $\tilde{W}_n^* \to \tilde{W}_\infty^*$.

QED.
SECTION 4.

ASYMPTOTIC EQUIPARTITION FOR A STATIONARY ERGODIC MARKET

The name asymptotic equipartition property (or A.E.P.) has its roots in statistical mechanics and is used in the information theory literature as an alternative for what is also known as the Shannon-McMillan-Breiman theorem. In the present section we shall prove the following generalized A.E.P. for log-optimum investment in a stationary ergodic stock market:

\[ \frac{1}{n} \log S_n^* \rightarrow W_\infty^* \quad \text{a.s. (and in } L^1 \text{ if } W_\infty^* \text{ is finite)} \]

or equivalently

\[ S_n^* = \exp[n(W_\infty^* + o(1))] \quad \text{where } o(1) \rightarrow 0 \text{ a.s. (and in } L^1 \text{ if } W_\infty^* \text{ is finite).} \]

Here \( S_n^* = \prod_{0 \leq t < n} \rho(b_t^*, X_t) \) is the fortune after \( n \) periods of log-optimum investment, and \( S_n = \prod_{0 \leq t < n} \rho(b_t, X_t) \) is the analogous quantity under a competing strategy \( \{b_t\}_{0 \leq t < \infty} \). Thus capital will grow exponentially almost surely in the long run with constant asymptotic rate \( W_\infty^* \). This rate must be highest possible by the principle of asymptotic optimality of section 2, which can now be written in the form

\[ \limsup_{n \to \infty} \frac{1}{n} \log S_n \leq \lim_{n \to \infty} \frac{1}{n} \log S_n^* = W_\infty^* \quad \text{a.s.} \]

Breiman [61] considered independent and identically distributed outcomes \( \{X_t\}_{0 \leq t < \infty} \). If \( b^* \in B \) attains the maximum expected log return

\[ W^* = \sup_{b \in B} \mathbb{E} \{ \log \rho(b, X) \} = \mathbb{E} \{ \log \rho(b^*, X) \} \]

then by the strong law of large numbers

\[ \frac{1}{n} \log S_n^* = \frac{1}{n} \sum_{0 \leq t < n} \log \rho(b^*, X_t) \rightarrow W^* \quad \text{a.s.} \]

We consider the more general case where \( T \) is an invertible measure preserving and metrically transitive transformation of the underlying probability space \((\Omega, \mathcal{F}, P)\). Thus \( \{X_t\}_{-\infty < t < \infty} \) is a stationary ergodic process, where \( X_t(\omega) = X(T^t\omega) \).

Another key ingredient that goes into the proof of the A.E.P., besides ergodicity, is the martingale convergence theorem. To bring martingale theory into
play, we need to impose a monotonicity condition on information fields. An increasing sequence of \( \sigma \)-fields \( \{ \mathcal{F}_t \}_{0 \leq t < \infty} \) models the situation where the investor keeps accumulating knowledge without ever having to perform data reductions or erasing his memory as time progresses. Somewhat surprisingly this is not the appropriate condition we need: the reason is an inter-temporal issue. Investors at different periods \( t \) face different decision problems, namely gambling on various outcomes \( X_t \) based on varying information fields \( \mathcal{F}_t \). But their capabilities should only be compared when all are placed in the same context. Thus we shall require monotonicity only after all investors have been shifted to the same reference point in time, say the beginning of period 0, where all are confronted with the same problem, namely that of selecting \( b_0^* \). The information field \( \mathcal{F}_t \) accessible at the beginning of period \( t \) is thereby transformed into the \( \sigma \)-field

\[
\mathcal{F}_t = T^t \mathcal{F} = \{ T^t F : F \in \mathcal{F}_t \}.
\]

The proper condition is that the sequence \( \{ \mathcal{F}_t \}_{0 \leq t < \infty} \) of shifted information fields be monotone increasing, say towards a limiting \( \sigma \)-field \( \mathcal{F}_\infty \):

\[
\mathcal{F}_t \nearrow \mathcal{F}_\infty = \bigvee_{0 \leq t < \infty} \mathcal{F}_t = \sigma(\bigcup_{0 \leq t < \infty} \mathcal{F}_t).
\]

We can therefore apply the notation and the martingale convergence results of section 3. Notice that the quantities as they appear to the agent at period \( t \) after he is translated back to the origin at time \( t = 0 \) (where comparisons can be made on common grounds) are distinguished from the corresponding unshifted quantities by writing an overbar. Thus

\[
P_t = \bar{P}_t \circ T^t, \quad b_t^* = \bar{b}_t^* \circ T^t, \quad w_t^* = \bar{w}_t^* \circ T^t, \quad \text{and} \quad w_t^* = \bar{w}_t^* \circ T^t
\]

for all \( 0 \leq t < \infty \). Furthermore \( \bar{W}_t^* = W_t^* \), by stationarity.

**Chain rules for the maximum expected log return**

Maximum expected log returns satisfy chain rules resembling those for entropy and Kullback-Leibler divergence in information theory. Let us restrict ourselves to the case where \( \mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1}) \) and hence \( \mathcal{F}_t = \sigma(X_{t-1}, X_{t-2}, X_{t-1}) \) for all \( 0 \leq t < \infty \). Then \( W_t^* = W^*(X_t|X_0, \ldots, X_{t-1}) \) denotes the maximum expected log return for period \( t \) based on the past outcomes \( X_0, \ldots, X_{t-1} \). We may write

\[
W_t^* = W^*(X_t|X_0, X_1, \ldots, X_{t-1}) = \sup_{b \in \mathcal{B}(X_0, X_1, \ldots, X_{t-1})} \mathbb{E}\{ \log \rho(b, X_t) \};
\]
\[ W_t^* = W^*(X_0|X_{-t}, \ldots, X_{-2}, X_{-1}) = \sup_{b=b(X_{-t}, \ldots, X_{-2}, X_{-1})} \mathbb{E}\{\log \rho(b, X_0)\}; \]

\[ W_\infty^* = W^*(X_0|\ldots, X_{-2}, X_{-1}) = \sup_{b=b(\ldots, X_{-2}, X_{-1})} \mathbb{E}\{\log \rho(b, X_0)\}. \]

Notice that \( W^*(X_t|X_0, X_1, \ldots, X_{t-1}) = W^*(X_0|X_{-t}, \ldots, X_{-2}, X_{-1}), \) by stationarity. We also define

\[ W^*(X_0, X_1, \ldots, X_{n-1}) = \mathbb{E}\{\log S_n^*\} \]

so that the following chain rule holds:

\[ W^*(X_0, X_1, \ldots, X_{n-1}) = \sum_{0 \leq t < n} W^*(X_t|X_0, X_1, \ldots, X_{t-1}). \]

The order of the expansion seems to matter, unlike the case for the entropy functional \( H^*, \) which satisfies

\[ H^*(X_0, X_1, \ldots, X_{n-1}) = \sum_{0 \leq i < n} H^*(X_{i_0}, X_{i_1}, \ldots, X_{i_{n-1}}) \]

for any permutation \( i_0, i_1, \ldots, i_{n-1} \) of \( 0, 1, \ldots, n - 1. \) If there is no gap between \( \lim_{t \to \infty} W_t^* \) and \( W_\infty^* \), then there are many equivalent definitions of \( W_\infty^* \), such as

\[ W_\infty^* = W^*(X_0, \ldots, X_{-2}, X_{-1}) = \lim_{t \to \infty} \uparrow W^*(X_0|X_{-t}, \ldots, X_{-2}, X_{-1}) = \lim_{n \to \infty} \uparrow W^*(X_t|X_0, X_1, \ldots, X_{t-1}) = \lim_{n \to \infty} \uparrow (1/n) W^*(X_0, X_1, \ldots, X_{n-1}). \]

**Breiman’s method for the dominated case**

Breiman [57] used the following generalization of Birkhoff’s ergodic theorem as a lemma to prove the classical A.E.P. of information theory.

**Theorem 4.1** Suppose \( T \) is a measure preserving and metrically transitive transformation of the underlying probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Let \( \{g_t\}_{0 \leq t < \infty} \) be an \( L^1 \)-dominated sequence that converges a.s. to some limiting random variable \( g_\infty \):

\[ \mathbb{E}\{\sup_{0 \leq t < \infty} |g_t|\} < \infty \quad \text{and} \quad g_t \to g_\infty \quad \text{a.s.} \]

Then

\[ \frac{1}{n} \sum_{0 \leq t < n} g_t \circ T^t \to \mathbb{E}\{g_\infty\} \quad \text{a.s. and in } L^1. \]

**Proof:**
We prove almost sure convergence if \( g_t \to g_\infty \) a.s. and \( \mathbb{E} \{ G_k \} < \infty \) for some \( 0 \leq k < \infty \), where
\[
G_k = \sup_{k \leq t < \infty} |g_t - g_\infty|.
\]
The same line of reasoning will also prove \( L^1 \)-convergence if in addition \( g_\infty \) and \( \sum_{0 \leq t < k} g_t \) are integrable.

It will suffice to show that \( \frac{1}{n} \sum_{0 \leq t < n} g_t \circ T^t = \frac{1}{n} \sum_{0 \leq t < n} g_\infty \circ T^t \) converges to zero, since the ergodic theorem asserts that
\[
\frac{1}{n} \sum_{0 \leq t < n} g_\infty \circ T^t \to \mathbb{E} \{ g_\infty \} \quad \text{a.s. (and in } L^1 \text{ since } \mathbb{E} \{ g_\infty \} \text{ is finite).}
\]
But for \( k \leq N \leq n \) we can write
\[
\left| \frac{1}{n} \sum_{0 \leq t < n} g_t \circ T^t - \frac{1}{n} \sum_{0 \leq t < n} g_\infty \circ T^t \right| 
\leq \frac{1}{n} \sum_{0 \leq t < k} (g_t \circ T^t - g_\infty \circ T^t) + \frac{1}{n} \sum_{k \leq t < N} G_k \circ T^t + \frac{1}{n} \sum_{N \leq t < n} G_N \circ T^t.
\]
The first two terms on the right hand side of the above inequality vanish in the limit and the third term converges to \( \mathbb{E} \{ G_N \} \). But \( G_N \searrow 0 \) as \( N \to \infty \) and \( \mathbb{E} \{ G_k \} < \infty \) by hypothesis so that, by the monotone convergence theorem, \( \mathbb{E} \{ G_N \} \searrow 0 \) as \( N \to \infty \).

QED.

Whether \( \mathbb{E} \{ \sup_{0 \leq t < \infty} |g_t| \} < \infty \) is in general hard to check. However there are some instances where this integrability condition follows from assumptions that are easier to verify. The next theorem provides such an example.

**Theorem 4.2** Let \( \{ \mathcal{G}_t \}_{0 \leq t \leq \infty} \) is a sequence of \( \sigma \)-fields such that \( \mathcal{G}_t \wedge \mathcal{G}_\infty \), and let \( g \) be a random variable with a well-defined expectation. Let \( g_t = \mathbb{E} \{ g | \mathcal{G}_t \} \) for \( 0 \leq t \leq \infty \) and suppose \( g_\infty \in L \log L \), that is, \( \mathbb{E} \{ |g_\infty| \log^+ |g_\infty| \} < \infty \). If \( T \) is measure preserving and metrically transitive then
\[
\frac{1}{n} \sum_{0 \leq t < n} g_t \circ T^t \to \mathbb{E} \{ g_\infty \} \quad \text{a.s. and in } L^1.
\]

**Proof:**

Note that \( \{ |g_t|, \mathcal{G}_t \}_{0 \leq t \leq \infty} \) is a nonnegative submartingale since \( \{ g_t, \mathcal{G}_t \}_{0 \leq t \leq \infty} \) is a martingale. By Wiener's dominated ergodic theorem (see Chung [74], theorem 4, p. 356 and exercise 7, p. 355), we may write
\[
\mathbb{E} \{ \sup_{0 \leq t < \infty} |g_t| \} \leq \frac{e}{e - 1} \left[ 1 + \sup_{0 \leq t < \infty} \mathbb{E} \{ |g_t| \log^+ |g_t| \} \right]
\]
where \( \log^+ x = \log x \) if \( x \geq 1 \) and 0 if \( 0 \leq x \leq 1 \). If \( \phi(x) = x \log^+ x \) then \( \phi: \mathcal{R}_+ \rightarrow \mathcal{R}_+ \) is convex and monotone increasing so that \( \{ \phi(|g_t|), \mathcal{G}_t \}_{0 \leq t \leq \infty} \) is also a submartingale (at least if the expectations are finite), and consequently

\[
\sup_{0 \leq t < \infty} \mathbb{E}\{|g_t| \log^+ |g_t|\} \leq \mathbb{E}\{|g_\infty| \log^+ |g_\infty|\}.
\]

Thus \( \mathbb{E}\{\sup_{0 \leq t < \infty} |g_t|\} < \infty \) if \( \mathbb{E}\{|g_\infty| \log^+ |g_\infty|\} < \infty \), that is, if \( |g_\infty| \in L \log L \). Since the integrability condition necessary for the application of Breiman's extension of the ergodic theorem holds, the theorem follows.

\textit{QED.}

We are now in a position to prove the generalized A.E.P. for log-optimum investment in the dominated case.

\textbf{Theorem 4.3} Let \( X_t(\omega) = X(T^t \omega) \) \((-\infty < t < \infty)\) where \( T \) is an invertible measure preserving and metrically transitive transformation, and let \( \{ \mathcal{F}_t \}_{0 \leq t < \infty} \) be a sequence of information fields such that the shifted fields \( \mathcal{F}_t = T^t \mathcal{F}_0 \) increase monotonically towards a limiting \( \sigma \)-field \( \mathcal{F}_\infty \).

Let \( b_1^* = b^*(P_t) \) be a log-optimum portfolio choice for period 0 based on \( \mathcal{F}_0 \) and suppose the log returns \( \bar{w}_t^* = \log \rho(b_t^*, X_0) \) \((0 \leq t < \infty)\) are \( L^1 \)-dominated and converge to a limit \( \bar{w}^* \) a.s. and in \( L^1 \):

\[
\mathbb{E}\{\sup_{0 \leq t < \infty} |\bar{w}_t^*|\} < \infty \quad \text{and} \quad \bar{w}_t^* \rightarrow \bar{w}^* \quad \text{a.s. and in } L^1.
\]  

Then the maximum conditional expected log returns \( \bar{w}_t^* = \mathbb{E}\{\bar{w}_t^* | \mathcal{F}_t\} \) form a uniformly integrable submartingale with respect to the \( \sigma \)-fields \( \mathcal{F}_t \), with limit \( \bar{w}^* = \mathbb{E}\{\bar{w}^* | \mathcal{F}_\infty\} \). Consequently

\[
\bar{w}_t^* \rightarrow \bar{w}^* = \mathbb{E}\{\bar{w}^* | \mathcal{F}_\infty\} \quad \text{a.s. and in } L^1.
\]  

The maximum expected log returns \( \bar{W}_t^* = \mathbb{E}\{\bar{w}_t^*\} = \mathbb{E}\{\bar{w}_t^*\} \) also converge:

\[
\bar{W}_t^* \rightarrow \bar{W}^* = \mathbb{E}\{\bar{w}^*\} = \mathbb{E}\{\bar{w}_t^*\}.
\]  

Let \( b_t^* = b^*(P_t) \) be a log-optimum portfolio selection based on \( \mathcal{F}_t \). The Cesàro averages of the true log returns \( w_t^* = \log \rho(b_t^*, X_t) \), the maximum conditional expected log returns \( \bar{w}_t^* = \mathbb{E}\{w_t^* | \mathcal{F}_t\} \), and the maximum expected log returns \( \bar{W}_t^* = \mathbb{E}\{w_t^*\} = \mathbb{E}\{w_t^*\} \) all converge to the same limit, namely \( \bar{W}^* \). In other words,

\[
\frac{1}{n} \sum_{0 \leq t < n} \bar{W}_t^* \rightarrow \bar{W}^*,
\]  

42
\[
\frac{1}{n} \sum_{0 \leq i < n} w_i^* \to \bar{W}^* \quad \text{a.s. and in } L^1, \quad \text{and} \quad (5)
\]

\[
\frac{1}{n} \log S_n^* = \frac{1}{n} \sum_{0 \leq i < n} w_i^* \to \bar{W}^* \quad \text{a.s. and in } L^1. \quad (6)
\]

Convergence of the conditional expected log returns, assertion (5), could be proved only if \(\{\bar{w}_i^*\}_{0 \leq i < \infty}\) is \(L^1\)-dominated. In particular, if \(\bar{w}_i^* \geq 0\) for all \(0 \leq t < \infty\) then it suffices that \(\bar{w}^* \in L \log L\), i.e., \(\mathbb{E}\{\bar{w}^* \log^+ \bar{w}^*\} < \infty\).

**Proof:**

To prove (2), observe that \(\{\bar{w}_i^*\}_{0 \leq i < \infty}\) is a submartingale, by Theorem 3.1. Chow and Teicher [78], cor. 1, p. 233 show that \(\bar{w}_i^* \to \bar{w}^*\) a.s. Part (3) follows from the \(L^1\)-convergence in (1) and the monotonicity of \(\{\mathcal{F}_t\}_{0 \leq t < \infty}\). To prove (4), observe that \(W_i^* = \bar{W}_i^*\) by stationarity and then use (3). The asymptotic equipartition property (6) follows from (1) by an application of Breiman's extension of the ergodic theorem since the required integrability condition is satisfied by hypothesis. Similarly (5) follows from (2) if the sequence \(\{\bar{w}_i^*\}_{0 \leq i < \infty}\) is \(L^1\)-dominated. For example if \(\bar{w}_i^* \geq 0\) for all \(0 \leq t < \infty\) and \(\bar{w}^* \in L \log L\), then \(\{\bar{w}_i^*, \mathcal{F}_t\}_{0 \leq t < \infty}\) is a nonnegative submartingale which is uniformly integrable and converges to \(\bar{w}^*\). Using the same proof as that of Theorem 4.2 but writing \(\bar{w}_i^*\) instead of \(|g_t|\) and \(\bar{w}^*\) instead of \(|g_\infty|\) we can show that

\[
\mathbb{E}\left\{\sup_{0 \leq t < \infty} |\bar{w}_i^*| \right\} \leq \frac{e}{\epsilon - 1} \left[ 1 + \mathbb{E}\{\bar{w}^* \log^+ \bar{w}^*\} \right].
\]

**QED.**

It will require some effort to check that hypothesis (1) is satisfied in each case where we want to apply the above theorem. Of course \(\{\bar{w}_i^*\}_{0 \leq i < \infty}\) is bounded and thus \(L^1\)-dominated if \(\rho(b, x)\) is bounded and bounded away from 0 on \(\mathcal{B} \times \mathcal{X}\). Proving that the conditional expected log returns \(\{\bar{w}_i^*\}_{0 \leq i < \infty}\) converge may be relatively easy. Indeed, if \(\sup_{0 \leq t < \infty} \mathbb{E}\{\bar{w}_i^*\}\) is finite then the submartingale \(\{\bar{w}_i^*, \mathcal{F}_t\}_{0 \leq t < \infty}\) converges a.s. and in \(L^1\) to a limit \(\bar{w}^*\) satisfying

\[
\bar{w}^* = \lim_{i \to \infty} \bar{w}_i^* \leq \bar{w}_\infty^* = \mathbb{E}\left\{\log \rho(b_\infty^*, X_0) | \mathcal{F}_\infty \right\} = \sup_{b \in \mathcal{B}_\infty} \mathbb{E}\left\{\log \rho(b, X_0) | \mathcal{F}_\infty \right\}.
\]

Moreover there is equality \(\bar{w}^* = \bar{w}_\infty^*\) if \(P \mapsto \omega^*(P)\) is lower semicontinuous in \(P \in \mathcal{P}\) (see Theorem 3.1). Proving that the sequence of true log returns \(\{\bar{w}_i^*\}_{0 \leq i < \infty}\) converges may be much more difficult because this sequence has in general no martingale properties. Recall that we even needed a careful argument using the strict convexity of the log function to show that \(\bar{w}_i^* = \log \rho(b_i^*, X_0)\) is unambiguously defined independently of the choice of log-optimum portfolio.
\( \bar{b}_t \in B^*(\bar{P}_t) \) (see Theorem 1.2). There is one case where we can be sure that \( \bar{w}_t = \log \rho(\bar{b}_t, X_0) \) converges to \( \bar{w}_\infty = \log \rho(\bar{b}_\infty, X_0) \), namely when \( B \) is compact, \( B^* \) is a closed correspondence from \( \mathcal{P} \) to \( B \), and \( \rho(b, z) \) is continuous in \( b \in B \) for fixed \( z \in X \) (see Theorem 1.8).

**The sandwich argument**

We now provide proof of the generalized A.E.P. for log-optimum investment that does not require any integrability assumptions. The only hypothesis we need is that there be no gap between \( \lim_{t \to \infty} \bar{W}_t^* \) and \( \bar{W}_\infty^* \).

**Theorem 4.4 (Asymptotic Equipartition Property)**

Let \( \{X_t\}_{-\infty < t < \infty} \) be a stationary ergodic process and \( \{\bar{F}_t\}_{0 \leq t < \infty} \) a sequence of information fields such that the shifted fields \( \bar{F}_i = T^i \bar{F}_0 \) increase monotonically towards a limiting \( \sigma \)-field \( \bar{F}_\infty \). If \( \bar{W}_t^* / \bar{W}_\infty^* \) (in particular, if \( P \to w^*(P) \) is lower semicontinuous in \( P \in \mathcal{P} \)) then

\[
\lim_{n \to \infty} \frac{1}{n} \log S^*_n = \lim_{n \to \infty} \frac{1}{n} \sum_{0 \leq t < n} \log \rho(b^*_t, X_t) = \bar{W}_\infty^* \quad \text{a.s.}
\]

or equivalently

\[
S^*_n = \exp[n(\bar{W}_\infty^* + o(1))] \quad \text{where } o(1) \to 0 \text{ a.s.}
\]

If \( \bar{W}_\infty^* \) is finite then there is also convergence in \( L^1 \).

**Proof:**

As a preliminary let us first consider the special case where \( \{w^*_t\}_{k \leq t < \infty} \) is stationary and ergodic, for some finite \( 0 \leq k < \infty \). Birkhoff’s ergodic theorem then suffices to prove the claim. Indeed, using \( w^*_t = \log \rho(b^*_t, X_t) \) we may write

\[
\frac{1}{n} \log S^*_n = \frac{1}{n} \sum_{0 \leq t < n} w^*_t = \frac{1}{n} \sum_{0 \leq t < k} w^*_t + \frac{n - k}{n} \sum_{k \leq t < \infty} w^*_t \circ T^{t-k}.
\]

It suffices to observe that

\[
\frac{1}{n} \sum_{0 \leq t < k} w^*_t \to 0 \quad \text{a.s. (and in } L^1 \text{ if } \bar{W}_t^* \text{ is finite for all } 0 \leq t < k),
\]

and

\[
\frac{1}{n - k} \sum_{k \leq t < \infty} w^*_t \circ T^{t-k} \to \bar{W}_k^* \quad \text{a.s. (and in } L^1 \text{ if } \bar{W}_k^* \text{ is finite).}
\]

This simple case will occur if \( X_0 \) and \( \bar{F}_t \) are conditionally independent given \( \bar{F}_k \) for all \( k \leq t < \infty \). (If \( \bar{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1}) \) then this condition means that
the process \( \{X_t\}_{0 \leq t < \infty} \) is Markov with finite order \( k \). In particular, this will be true if the shifted information fields \( \{\mathcal{F}_t\}_{0 \leq t < \infty} \) stabilize after a finite time \( k \), that is, if \( \mathcal{F}_k = \mathcal{F}_{k+1} = \ldots = \mathcal{F}_t = \ldots = \mathcal{F}_\infty \).

The general case will now be reduced to this special one using a sandwich argument. We define diminished information fields \( \mathcal{F}_t(k) \subseteq \mathcal{F}_t \) for \( 0 \leq t < \infty \) and \( 0 \leq k < \infty \) and expanded fields \( \mathcal{F}_t(\infty) \supseteq \mathcal{F}_t \) for \( 0 \leq t < \infty \), as follows:

\[
\mathcal{F}_t(k) = T^{-t} \mathcal{F}_{t \wedge k} = \begin{cases} T^{-t} \mathcal{F}_t & \text{for } 0 \leq t < k \\ T^{-t} \mathcal{F}_k & \text{for } k \leq t < \infty \end{cases}
\]

and

\[
\mathcal{F}_t(\infty) = T^{-t} \mathcal{F}_\infty \quad \text{for } 0 \leq t < \infty.
\]

Clearly \( \mathcal{F}_t(\infty) = T^t \mathcal{F}_t(\infty) = \mathcal{F}_\infty \) and \( \mathcal{F}_t(k) = T^t \mathcal{F}_t(k) = \mathcal{F}_{t \wedge k} \) for all \( 0 \leq t < \infty \) and all \( 0 \leq k < \infty \). To clarify these definitions it is instructive to examine what they mean if \( \mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1}) \) and \( \mathcal{F}_t = \sigma(X_{t-1}, \ldots, X_2, X_1) \) for all \( 0 \leq t < \infty \). In that case it is appropriate to call \( \mathcal{F}_t(k) \) the \( k \)-past and \( \mathcal{F}_t(\infty) \) the infinite past:

\[
\mathcal{F}_t(k) = \begin{cases} \sigma(X_0, X_1, \ldots, X_{t-1}) = \mathcal{F}_t & \text{if } 0 \leq t < k, \\ \sigma(X_{t-k}, \ldots, X_{t-1}) & \text{if } k \leq t < \infty \end{cases}
\]

for all \( 0 \leq k < \infty \) and

\[
\mathcal{F}_t(\infty) = \sigma(\ldots, X_{t-1}, X_0, X_1, \ldots, X_{t-1}) \quad \text{for all } 0 \leq t < \infty.
\]

Let \( S_n^*(k) \) and \( S_n^*(\infty) \) denote the capital growth over \( n \) periods under the strategies \( b_t^*(k) \) and \( b_t^*(\infty) \) which are the log-optimum choices based on the diminished and expanded \( \sigma \)-fields \( \mathcal{F}_t(k) \) and \( \mathcal{F}_t(\infty) \), respectively. Since

\[
\mathcal{F}_t(k) \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t(\infty) \quad \text{for all } 0 \leq t < \infty,
\]

the asymptotic optimality principle of section 2 implies that

\[
\limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n^*(k)}{S_n^*(\infty)} \right) \leq 0 \quad \text{and} \quad \limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{S_n^*(k)}{S_n^*(\infty)} \right) \leq 0 \quad \text{a.s.}
\]

We have been careful to define the approximating fields \( \mathcal{F}_t(k) \) and \( \mathcal{F}_t(\infty) \) in such a way that the shifted fields \( \mathcal{F}_t(k) = T^t \mathcal{F}_t(k) \) and \( \mathcal{F}_t(\infty) = T^t \mathcal{F}_t(\infty) \) stabilize after a finite time. Specifically,

\[
\mathcal{F}_t(k) = \mathcal{F}_{t+1}(k) = \ldots = \mathcal{F}_t(k) = \ldots = \mathcal{F}_\infty \quad \text{for all } k \leq t < \infty
\]

and

\[
\mathcal{F}_0(\infty) = \mathcal{F}_1(\infty) = \ldots = \mathcal{F}_t(\infty) = \ldots = \mathcal{F}_\infty \quad \text{for all } 0 \leq t < \infty.
\]
But Birkhoff’s ergodic theorem implies that
\[
\frac{1}{n} \log S_n^*(k) \to \bar{W}_k^* \quad \text{a.s. and} \quad \frac{1}{n} \log S_n^*(\infty) \to \bar{W}_\infty^* \quad \text{a.s.}
\]
and we may conclude that
\[
\bar{W}_k^* \leq \liminf_{n \to \infty} \frac{1}{n} \log S_n^* \leq \limsup_{n \to \infty} \frac{1}{n} \log S_n^* \leq \bar{W}_\infty^* \quad \text{a.s.}
\]
for all \(0 \leq k < \infty\). We now let \(k \uparrow \infty\). Since \(\bar{W}_k^* \uparrow \bar{W}_\infty^*\) by hypothesis, we obtain
\[
\frac{1}{n} \log S_n^* \to \bar{W}_\infty^* \quad \text{a.s.}
\]
If \(\bar{W}_k^*\) is finite for all \(0 \leq k \leq \infty\) then \(L^1\)-convergence will hold wherever we argued almost sure convergence.
QED.

**The long run averages of the predicted log returns**

The true log return \(\log \rho(b_t, X_t)\) will in general differ from the conditional expected log return \(E\{\log \rho(b_t, X_t)|\mathcal{F}_t\}\) because of some residual randomness. For any permissible strategy \(\{b_t\}_{0 \leq t < \infty}\) let \(\tilde{S}_n\) denotes the capital growth over \(n\) periods in a hypothetical world where we always exactly realize the conditional expected log return:
\[
\tilde{S}_n = \prod_{0 \leq t < n} \exp \left[ E \{\log \rho(b_t, X_t)|\mathcal{F}_t\} \right]
\]
or equivalently
\[
\frac{1}{n} \log \tilde{S}_n = \frac{1}{n} \sum_{0 \leq t < n} E \{\log \rho(b_t, X_t)|\mathcal{F}_t\}.
\]
If we invest according to the log-optimum strategy \(\{b_t^*\}_{0 \leq t < \infty}\) then the corresponding quantity is denoted by \(\tilde{S}_n^*\):
\[
\tilde{S}_n^* = \prod_{0 \leq t < n} \exp \left[ E \{\log \rho(b_t^*, X_t)|\mathcal{F}_t\} \right]
\]
or equivalently
\[
\frac{1}{n} \log \tilde{S}_n^* = \frac{1}{n} \sum_{0 \leq t < n} w_t^* = \frac{1}{n} \sum_{0 \leq t < n} E \{\log \rho(b_t^*, X_t)|\mathcal{F}_t\}.
\]
Clearly \(\tilde{S}_n \leq \tilde{S}_n^*\) for all \(0 \leq n < \infty\) so that
\[
\limsup_{n \to \infty} \frac{1}{n} \log \left( \frac{\tilde{S}_n}{\tilde{S}_n^*} \right) \leq 0 \quad \text{a.s.}
\]
We can prove an asymptotic equipartition property for the hypothetical capital growth $\tilde{S}_n^*$, using a sandwich argument completely analogous to that in the previous theorem.

**Theorem 4.5** Suppose that $\bar{W}_k^* \not\asymp \bar{W}_\infty^*$, $\bar{w}_\infty^* \in L \log L$, and $\bar{w}_k^* \in L \log L$ for sufficiently large $k$. Then

$$\frac{1}{n} \log \tilde{S}_n^* = \frac{1}{n} \sum_{0 \leq t < n} w_t^* \to \bar{W}_\infty^* \text{ a.s. and in } L^1$$

or equivalently

$$\tilde{S}_n^* = \exp[n(\bar{W}_\infty^* + o(1))] \text{ where } o(1) \to 0 \text{ a.s. and in } L^1.$$ 

**PROOF:**

For $0 \leq k \leq \infty$, we define

$$\tilde{S}_n^*(k) = \prod_{0 \leq t < n} \exp[\mathbb{E} \{ \log \rho(b_t^*(k), X_t) | \mathcal{F}_t \}]$$

where $b_t^*(k)$ is the log-optimum portfolio for period $t$ that is based on the $\sigma$-field $\mathcal{F}_t(k)$. If $\bar{w}_k^* \in L \log L$ then using Theorem 4.2 we can prove that

$$\frac{1}{n} \log \tilde{S}_n^*(k) = \frac{1}{n} \sum_{0 \leq t < n} \mathbb{E} \{ \log \rho(b_t^*(k), X_t) | \mathcal{F}_t \} \to \bar{W}_k^* \text{ a.s. and in } L^1.$$ 

Indeed, $\mathbb{E} \{ \log \rho(b_t^*(k), X_t) | \mathcal{F}_t \} = \mathbb{E} \{ \log \rho(b_t^*(0), X_t) | \mathcal{F}_t \} \circ T_t$ for all $t \geq k$, and $\bar{W}_k^*$ is the expectation of the random variable $\log \rho(b_k^*(0), X_0)$ and its conditional expectation $\bar{w}_k^* = \mathbb{E} \{ \log \rho(b_k^*(0), X_0) | \mathcal{F}_k \}$. Similarly, if $\bar{w}_\infty^* \in L \log L$, then

$$\frac{1}{n} \log \tilde{S}_n^*(\infty) = \frac{1}{n} \sum_{0 \leq t < n} \mathbb{E} \{ \log \rho(b_t^*(\infty), X_t) | \mathcal{F}_t \} \to \bar{W}_\infty^* \text{ a.s. and in } L^1$$

since $\mathbb{E} \{ \log \rho(b_t^*(\infty), X_t) | \mathcal{F}_t \} = \mathbb{E} \{ \log \rho(b_\infty^*(0), X_0) | \mathcal{F}_t \} \circ T_t$ for all $0 \leq t < \infty$ and $\bar{W}_\infty^*$ is the expectation of the random variable $\log \rho(b_\infty^*(0), X_0)$ and its conditional expectation $\bar{w}_\infty^* = \mathbb{E} \{ \log \rho(b_\infty^*(0), X_0) | \mathcal{F}_\infty \}$. Thus the following chain of inequalities holds with probability one:

$$\lim_{n \to \infty} \frac{1}{n} \log \tilde{S}_n^*(k) \leq \liminf_{n \to \infty} \frac{1}{n} \log \tilde{S}_n^* \leq \limsup_{n \to \infty} \frac{1}{n} \log \tilde{S}_n^* \leq \lim_{n \to \infty} \frac{1}{n} \log \tilde{S}_n^*(\infty)$$

or equivalently

$$\bar{W}_k^* \leq \liminf_{n \to \infty} \frac{1}{n} \log \tilde{S}_n^* \leq \limsup_{n \to \infty} \frac{1}{n} \log \tilde{S}_n^* \leq \bar{W}_\infty^* \text{ a.s.}$$

Since $\bar{W}_k^* \not\asymp \bar{W}_\infty^*$ by assumption the theorem follows.

QED.
SECTION 5.

DECOMPOSITION INTO
IDEAL REFERENCE LEVEL
AND GENERALIZED ENTROPY

Sometimes a divide and conquer strategy works to prove the generalized A.E.P. where a direct approach fails. In particular if we decompose the return function \( \rho(b, x) \) into the product of two factors then we have the liberty of considering two terms separately. In this section we shall assume that there exists a separable metrizable space \( \mathcal{U} \) and Borel measurable functions \( u : \mathcal{X} \rightarrow \mathcal{U} \), \( \varphi : \mathcal{X} \rightarrow \mathbb{R}_+ \), and \( \lambda : \mathcal{B} \times \mathcal{U} \rightarrow \mathbb{R}_+ \) such that

\[
\rho(b, x) = \varphi(x) \cdot \lambda(b, u(x)).
\]

The factor \( \varphi(X) \) describes the behaviour of the aggregate market as a whole and \( \lambda(b, u(X)) \) indicates how well each strategy \( b \) performs relative to that global reference. Notice that the relative return \( \lambda(b, u(X)) \) depends on the random outcome \( X \) only through the random variable \( U = u(X) \), which we call the normalized outcome. In order that \( \rho(b, x) \) be concave in \( b \in \mathcal{B} \) for fixed \( x \in X \), we must assume that \( \lambda(b, u) \) is concave in \( b \in \mathcal{B} \) for fixed \( u \in \mathcal{U} \). To obtain a real simplification we shall assume that \( \log \lambda(b, U) \) is bounded above by an integrable function independent of \( b \in \mathcal{B} \). Any normalizing factor can be absorbed in \( \varphi(X) \), so there will be no loss of generality if we actually assume that \( 0 \leq \lambda(b, u) \leq 1 \). No conditions are imposed on the function \( \varphi : \mathcal{X} \rightarrow \mathbb{R}_+ \) except measurability, but \( \lambda : \mathcal{B} \times \mathcal{U} \rightarrow [0, 1] \) is assumed to be upper semicontinuous. It is clear that \( \sup_{b \in \mathcal{B}} \rho(b, x) \) is a natural candidate for the function \( \varphi(x) \).

Let \( \mathcal{P} \) and \( \mathcal{Q} \) denote the spaces of probability measures on \( \mathcal{X} \) and \( \mathcal{U} \), and let \( P \in \mathcal{P} \) and \( Q \in \mathcal{Q} \) be the probability distributions of the random elements \( X \) and \( U = u(X) \), respectively. Thus \( Q \) is the image measure of \( P \), i.e., \( Q = P \circ u^{-1} \), or

\[
Q\{U \in B\} = P\{u(X) \in B\} = P(u^{-1}(B)) \quad \text{for all } B \in \mathcal{B}_U.
\]

Observe that \( \mathcal{P} \) and \( \mathcal{Q} \) are separable metrizable spaces when equipped with their weak topologies. The Borel measurable function \( u : (\mathcal{X}, \mathcal{B}_{\mathcal{X}}) \rightarrow (\mathcal{U}, \mathcal{B}_{\mathcal{U}}) \) induces a Borel measurable function

\[
\bar{u} : (\mathcal{P}, \mathcal{B}_{\mathcal{P}}) \rightarrow (\mathcal{Q}, \mathcal{B}_{\mathcal{Q}})
\]

\[
P \mapsto Q = \bar{u}(P) = P \circ u^{-1}.
\]
In fact \( \tilde{u} : \mathcal{P} \to \mathcal{Q} \) is continuous if \( u : \mathcal{X} \to \mathcal{U} \) is continuous (see Schwartz [73], prop. 3, p. 372).

The log return \( w(b, X) = \log \rho(b, X) \) can be written in the form

\[
w(b, X) = r(X) - h(b, U)
\]

where

\[
r(x) = \log \varphi(x)
\]

and

\[
h(b, u) = -\log \lambda(b, u).
\]

Taking expectations gives

\[
w(b, P) = r(P) - h(b, Q)
\]

where

\[
r(P) = \mathbb{E}_P \{ r(X) \} = \mathbb{E}_P \{ \log \varphi(X) \}
\]

and

\[
h(b, Q) = \mathbb{E}_Q \{ h(b, U) \} = \mathbb{E}_Q \{ -\log \lambda(b, U) \}.
\]

The quantity \( r(P) \) is interpreted as an ideal reference level for the expected log return. It is an inherent property of the market as a whole, over which the investor has no control. And \( h(b, Q) \) represents the loss of log return with respect to that ideal reference level. Notice that \( h(b, Q) \) depends on \( P \) only through the marginal distribution \( Q \) of the normalized outcome \( U \). Similarly, the maximum expected log return and the minimum loss thereof with respect to the reference level can be written as

\[
w^*(P) = \sup_{b \in \mathcal{B}} \mathbb{E}_P \{ \log \rho(b, X) \} = r(P) - h^*(Q)
\]

and

\[
h^*(Q) = \inf_{b \in \mathcal{B}} h(b, Q) = \inf_{b \in \mathcal{B}} \mathbb{E}_P \{ -\log \lambda(b, U) \}.
\]

The advantages of the decomposition now become clear. The reference level \( r(P) \) may be a very irregular function, possibly infinite, of the distribution \( P \in \mathcal{P} \). However this quantity is irrelevant for our discussion because our actions cannot affect its value. When trying to achieve the maximum expected log return

\[
w^*(P) = \sup_{b \in \mathcal{B}} [r(P) - h(b, Q)],
\]
we shall always minimize the expected loss \( h(b, Q) \), even if \( r(P) \) is infinite. We also write \( B^*(Q) \) for the convex set of log-optimum actions (instead of \( B^*(P) \)). By subtracting out the reference level \( r(P) \) from the expected log return \( w(b, P) = E_P \{ \log \rho(b, X) \} \), we remove irregularities and are left with a more tractable optimization problem, namely minimizing a nonnegative lower semicontinuous function \( h(b, Q) = E_Q \{ -\log \lambda(b, U) \} \).

**Theorem 5.1** Suppose \( \mathcal{U} \) and \( \mathcal{B} \) are separable metrizable spaces with \( \mathcal{B} \) compact, and \( \lambda : \mathcal{B} \times \mathcal{U} \to [0, 1] \) is upper semicontinuous. Let \( B^*(Q) \) denote the set of elements \( b^* \in \mathcal{B} \) attaining the infimum \( h^*(Q) = \inf_{b \in \mathcal{B}} E_Q \{ -\log \lambda(b, U) \} \).

1. Then \( h(b, Q) = E_Q \{ -\log \lambda(b, U) \} \) is nonnegative, lower semicontinuous on \( \mathcal{B} \times \mathcal{Q} \), and affine in \( Q \in \mathcal{Q} \) for fixed \( b \in \mathcal{B} \). And \( h^*(Q) \) is nonnegative, concave, and lower semicontinuous on \( \mathcal{Q} \). Furthermore \( B^* \) is a closed-valued correspondence from \( \mathcal{Q} \) to \( \mathcal{B} \), and there exists a Borel measurable selection function \( b^* : (\mathcal{Q}, \mathcal{B}_Q) \to (\mathcal{B}, \mathcal{B}_\mathcal{B}) \) such that \( b^*(Q) \in B^*(Q) \) for all \( Q \in \mathcal{Q} \).

2. Suppose moreover that \( h^* : \mathcal{Q} \to \mathbb{R}_+ \) is (not only lower semicontinuous but also upper semicontinuous and hence) continuous. (This will be the case if \( \lambda : \mathcal{B} \times \mathcal{U} \to (0, 1] \) is continuous and bounded away from 0.) Then the compact-valued correspondence \( B^* \) from \( \mathcal{Q} \) to \( \mathcal{B} \) is closed, and any selection \( b^*(Q) \in B^*(Q) \) is continuous at any point \( Q \) for which \( B^*(Q) = \{ b^*(Q) \} \) is a singleton set.

**Proof:**

This follows immediately from Theorems 4, 5, 6 and 7 in section 1. QED.

**Sequential investment**

Let us now go back to the sequential investment problem. Again assume that the underlying probability space \((\Omega, \mathcal{F}, P)\) is perfect and equipped with an invertible measure preserving transformation \( T \) that is metrically transitive. The random process \( X_t = X \circ T^t \) \((-\infty < t < \infty)\) is then stationary ergodic. Let \( \{ \mathcal{F}_t \}_{0 \leq t < \infty} \) be a sequence of information fields such that the shifted fields \( \mathcal{F}_t = T^t \mathcal{F}_0 \) increase monotonically towards a limiting \( \sigma \)-field \( \mathcal{F}_\infty \). And let \( Q_t \) and \( Q_t \) denote regular conditional probability distributions of \( U_t = u(X_t) \) given \( \mathcal{F}_t \) and of \( U_0 = u(X_0) \) given \( \mathcal{F}_0 \), respectively. We also assume that a Borel measurable selection \( b^* : (\mathcal{Q}, \mathcal{B}_Q) \to (\mathcal{B}, \mathcal{B}_\mathcal{B}) \) exists such that \( b^*(Q) \in B^*(Q) \) for all \( Q \) in \( \mathcal{Q} \).
The log-optimum portfolio $\tilde{b}_i^* = b^*(\tilde{Q}_i)$ achieves the minimum conditional expected loss $\tilde{h}_i^*$ for period 0 given $\tilde{F}_i$:

$$\tilde{h}_i^* = h^*(Q_i) = \inf_{b \in \tilde{F}_i} \mathbb{E} \{-\log \lambda(b, U_0)|\tilde{F}_i\} = \mathbb{E} \{-\log \lambda(\tilde{b}_i^*, U_0)|\tilde{F}_i\} = \mathbb{E} \{\tilde{h}_i^*|\tilde{F}_i\}. $$

Let $\tilde{h}_i^*$ denote the loss of log return under the log-optimum strategy $\tilde{b}_i^*$:

$$\tilde{h}_i^* = -\log \lambda(\tilde{b}_i^*, U_0),$$

and $\tilde{H}_i^*$ the minimum expected loss of log return under decisions based on $\tilde{F}_i$:

$$\tilde{H}_i^* = \inf_{b \in \tilde{F}_i} \mathbb{E} \{-\log \lambda(b, U_0)\} = \mathbb{E} \{-\log \lambda(\tilde{b}_i^*, U_0)\} = \mathbb{E} \{\tilde{h}_i^*\} = \mathbb{E} \{\tilde{h}_i^*\}. $$

Thus $\tilde{H}_i^*$ is the expectation both of the true loss of log return $\tilde{h}_i^*$ and its conditional expectation $\tilde{h}_i^*$. The true log return for period 0 under the log-optimum strategy $\tilde{b}_i^*$ can now be written as

$$\tilde{w}_i^* = \log \rho(\tilde{b}_i^*, X_0) = r - \tilde{h}_i^*$$

where

$$r = \log \varphi(X_0).$$

The maximum conditional expected log return given $\tilde{F}_i$ is written as

$$\tilde{w}_i^* = \sup_{b \in \tilde{F}_i} \mathbb{E} \{\log \rho(b, X_0)|\tilde{F}_i\} = \mathbb{E} \{\log \rho(\tilde{b}_i^*, X_0)|\tilde{F}_i\} = \mathbb{E} \{\tilde{w}_i^*|\tilde{F}_i\} = \tilde{r}_i - \tilde{h}_i^*$$

where

$$\tilde{r}_i = \mathbb{E} \{r|\tilde{F}_i\} = \mathbb{E} \{\log \varphi(X_0)|\tilde{F}_i\}$$

is the reference level for the conditional expected log return given $\tilde{F}_i$. Finally, the maximum expected log return for period 0 under decisions based on $\tilde{F}_i$ can be expressed as

$$\tilde{W}_i^* = \sup_{b \in \tilde{F}_i} \mathbb{E} \{\log \rho(b, X_0)\} = \mathbb{E} \{\log \rho(\tilde{b}_i^*, X_0)\} = \mathbb{E} \{\tilde{w}_i^*\} = \mathbb{E} \{\tilde{r}_i\} = R - \tilde{H}_i^*$$

where $R$ is the expected reference level for the log return during any period:

$$R = \mathbb{E} \{r\} = \mathbb{E} \{\tilde{r}_i\} = \mathbb{E} \{\log \varphi(X)\}$$

These were the quantities associated with the investor at time $t$ after he has been translated back to period 0. But similar quantities can be defined for the
unshifted investor. The log-optimum portfolio \( b_t^* = b^*(Q_t) \) achieves the minimum conditional expected loss \( h_t^* \) for period 0 given \( \mathcal{F}_t^* \):

\[
h_t^* = h^*(Q_t) = \inf_{b \in \mathcal{H}_t} \mathbb{E}\{ -\log \lambda(b, U_t) | \mathcal{F}_t \} = \mathbb{E}\{ -\log \lambda(b_t^*, U_t) | \mathcal{F}_t \} = \mathbb{E}\{ h_t^* | \mathcal{F}_t \}
\]

where \( h_t^* \) denotes the loss of log return for period \( t \) under the log-optimum strategy \( b_t^* \):

\[
h_t^* = -\log \lambda(b_t^*, U_t).
\]

Let \( H_t^* \) be the minimum expected loss of log return for period \( t \) given \( \mathcal{F}_t^* \):

\[
H_t^* = \inf_{b \in \mathcal{H}_t} \mathbb{E}\{ -\log \lambda(b, U_t) \} = \mathbb{E}\{ -\log \lambda(b_t^*, U_t) \} = \mathbb{E}\{ h_t^* \} = \mathbb{E}\{ h_t^* \}
\]

so that \( H_t^* \) is the expectation both of the true loss of log return \( h_t^* \) and its conditional expectation \( h_t^* \). The true log return for period \( t \) under the log-optimum strategy \( b_t^* \) is given by

\[
w_t^* = \log \rho(b_t^*, X_t) = r_t - h_t^*
\]

where

\[
r_t = \log \varphi(X_t).
\]

For the maximum conditional expected log return for period \( t \) given \( \mathcal{F}_t \) we get

\[
w_t^* = \sup_{b \in \mathcal{H}_t} \mathbb{E}\{ \log \rho(b, X_t) | \mathcal{F}_t \} = \mathbb{E}\{ \log \rho(b_t^*, X_t) | \mathcal{F}_t \} = \mathbb{E}\{ w_t^* | \mathcal{F}_t \} = r_t - h_t^*
\]

where

\[
r_t = \mathbb{E}\{ r_t | \mathcal{F}_t \} = \mathbb{E}\{ \log \varphi(X_t) | \mathcal{F}_t \}
\]

is the reference level for the conditional expected log return given \( \mathcal{F}_t \). Finally the maximum expected log return for period \( t \) given \( \mathcal{F}_t \) is

\[
W_t^* = \sup_{b \in \mathcal{H}_t} \mathbb{E}\{ \log \rho(b, X_t) \} = \mathbb{E}\{ \log \rho(b_t^*, X_t) \} = \mathbb{E}\{ w_t^* \} = \mathbb{E}\{ w_t^* \} = R - H_t^*.
\]

With these assumptions and definitions, we can now state an important theorem. We require some strong hypotheses, but these will later be shown to hold for a market with a finite number of stocks.

**Theorem 5.2** Suppose \( \mathcal{B} \) is compact, \( \lambda: \mathcal{B} \times \mathcal{U} \rightarrow [0, 1] \) is continuous, and the minimum loss of log return \( h^*: \mathcal{Q} \rightarrow \mathcal{R}_+ \) is bounded and continuous. Then
1. The maximum expected log return and the minimum expected loss of log return with respect to the reference level $R$ converge monotonically:

\[ \bar{H}_t \downarrow \bar{H}_\infty^* \quad \text{and} \quad \bar{W}_t^* = R - H_t^* \nearrow \bar{W}_\infty^* = R - \bar{H}_\infty^*. \]

Moreover $H_t^* = \bar{H}_t^*$ and $W_t^* = \bar{W}_t^*$ by stationarity. Thus the Cesàro averages also converge:

\[ \frac{1}{n} \sum_{0 \leq i < n} H_i^* \downarrow \bar{H}_\infty^* \quad \text{and} \quad \frac{1}{n} \sum_{0 \leq i < n} W_i^* \nearrow \bar{W}_\infty^*. \]

2. The sequence $\{\bar{h}_i^*, \bar{\pi}_i\}_{0 \leq i \leq \infty}$ is a bounded nonnegative supermartingale, and

\[ \bar{h}_i^* \rightarrow \bar{h}_\infty^* \quad \text{a.s. and in } L^1. \]

The conditional expected reference level $\{\bar{r}_i, \bar{\pi}_i\}_{0 \leq i \leq \infty}$ is a martingale (strictly speaking only when $R = B \{\bar{r}_i\}$ is finite). In any case,

\[ \bar{r}_i \rightarrow \bar{r}_\infty \quad \text{a.s. (and in } L^1 \text{ if } R \text{ is finite).} \]

Thus $\{\bar{w}_i^*, \bar{\pi}_i\}_{0 \leq i \leq \infty}$ is a submartingale (strictly speaking only when $R$ or $\bar{W}_\infty^*$ is finite), and

\[ \bar{w}_i^* = \bar{r}_i - \bar{h}_i^* \rightarrow \bar{w}_\infty^* = \bar{r}_\infty - \bar{h}_\infty^* \quad \text{a.s. (and in } L^1 \text{ if } R \text{ or } \bar{W}_\infty^* \text{ is finite).} \]

3. The true loss of log return $\bar{h}_i^* = -\log \lambda(\bar{h}_i^*, U_0)$ converges:

\[ \bar{h}_i^* \rightarrow \bar{h}_\infty^* \quad \text{a.s. and in } L^1. \]

Thus the true log return $\bar{w}_i^* = \log \rho(\bar{h}_i^*, X_0)$ also converges:

\[ \bar{w}_i^* = r - \bar{h}_i^* \rightarrow \bar{w}_\infty^* = r - \bar{h}_\infty^* \quad \text{a.s. (and in } L^1 \text{ if } R \text{ or } \bar{W}_\infty^* \text{ is finite).} \]

4. The long run averages of the expected losses converge:

\[ \frac{1}{n} \sum_{0 \leq i < n} h_i^* \rightarrow H_\infty^* \quad \text{a.s. and in } L^1. \]

Suppose that the long run averages of the conditional expected reference levels also converge:

\[ \frac{1}{n} \sum_{0 \leq i < n} r_i \rightarrow R \quad \text{a.s. and possibly in } L^1. \]
(Almost sure and $L^1$ convergence will hold if $R$ is finite and the martingale 
$\{\bar{r}_t\}_{0 \leq t < \infty}$ is $L^1$-dominated, i.e. $E\{\sup_{0 \leq t < \infty} |\bar{r}_t|\} < \infty$, and for this it 
suffices that $E\{\bar{r}_\infty | \log |\bar{r}_\infty|\}$ or $E\{|r| \log |r|\}$ is finite.)
Then the Cesàro averages of the maximum conditional expected log returns 
converge:

$$\frac{1}{n} \sum_{0 \leq t < n} w_t^* = \frac{1}{n} \sum_{0 \leq t < n} [r_t - h_t^*] \to \bar{W}_\infty^* = R - \bar{H}_\infty^* \text{ a.s.}$$

5. Finally we can prove convergence of the average loss of log return per 
period:

$$\frac{1}{n} \sum_{0 \leq t < n} h_t^* \to \bar{H}_\infty^* \text{ a.s. and in } L^1.$$ 

But the sample averages of the reference levels $\mathbf{r}_t = \log \varphi(\mathbf{X}_t)$ converge by 
Birkhoff’s ergodic theorem:

$$\frac{1}{n} \sum_{0 \leq t < n} \mathbf{r}_t \to R \text{ a.s (and in } L^1 \text{ if } R \text{ is finite).}$$

Consequently the generalized A.E.P. holds:

$$\frac{1}{n} \log S_n^* = \frac{1}{n} \sum_{0 \leq t < n} w_t^* = \frac{1}{n} \sum_{0 \leq t < n} \log \rho(b_t^*, \mathbf{X}_t) \to \bar{W}_\infty^*.$$ 

The convergence is a.s., and in $L^1$ if $R$ or $\bar{W}_\infty^*$ is finite.

PROOF:

This theorem summarizes previous results and the proof contains no new 
ideas. We leave it to the reader.

QED.
SECTION 6.
MARKETS WITH FINITELY MANY INVESTMENT OPTIONS

In this section we discuss the case where capital must be distributed over a finite set of elementary investment opportunities $\mathcal{A} = \{1, 2, \ldots, m\}$. A portfolio is a vector of nonnegative weights summing to one, that is, a vector in the simplex $\mathcal{B}$, where

$$\mathcal{B} = \{ \mathbf{b} = (b_j)_{1 \leq j \leq m} \in \mathbb{R}_+^m : b_1 + \ldots + b_m = 1 \}.$$

Let $X^j$ denote the return per dollar invested in stock $j$. These returns are summarized in a vector $X = (X^j)_{1 \leq j \leq m} \in \mathbb{R}_+^m$ so that the total return resulting from investment according to a portfolio $\mathbf{b}$ is given by the inner product

$$\rho(\mathbf{b}, X) = (\mathbf{b} \cdot X) = b_1 X^1 + \ldots + b_m X^m.$$

When we compute the average of the returns $X^j$ over all stocks $j$, then we obtain a quantity $(\eta \cdot X) = (X^1 + \ldots + X^m)/m$ which can be viewed as the total return when one monetary unit is invested according to the uniform portfolio $\eta = (1/m)_{1 \leq j \leq m}$ which allocates an equal amount to each of the $m$ stocks. It will be convenient to factor the outcome $X$ as a product $X = m(\eta \cdot X) U$, where

$$U = \frac{X}{m(\eta \cdot X)} = \left( \frac{X^j}{X^1 + \ldots + X^j + \ldots + X^m} \right)_{1 \leq j \leq m}$$

is called the normalized or scaled outcome. It is a vector in the same direction as $X$, but taking values in the unit simplex $\mathcal{U}$:

$$U \in \mathcal{U} = \{ \mathbf{u} = (u^j)_{1 \leq j \leq m} \in \mathbb{R}_+^m : u_1 + \ldots + u_m = 1 \}.$$

Let $P$ be the probability distribution of $X$ on $\mathbb{R}_+^m$ and $Q$ the marginal distribution of $U$ on $\mathcal{U}$. Informally, $Q$ can be obtained by integrating out the measure $P$ along rays through the origin in $\mathbb{R}_+^m$. If we restrict ourselves to distributions $P$ for which $P\{X = 0\} = 0$ then $U$ is a well-defined continuous function of $X$, and $Q$ is a continuous function of $P$. But a more complete definition is needed if $P\{X = 0\} > 0$. In general we shall define (see figure)

$$U = u(X) = \begin{cases} \frac{\eta \cdot X}{m(\eta \cdot X)} = \frac{X}{X^1 + \ldots + X^m} & \text{if } X = 0; \\ X & \text{otherwise.} \end{cases}$$

55
Given the factorization $(b \cdot x) = \varphi(x) \cdot \lambda(b, u(x))$ where
\[ \varphi(x) = m(\eta \cdot x) = x_1 + \ldots + x^m \quad \text{and} \quad \lambda(b, u) = (b \cdot u), \]
we can write the maximum expected log return
\[ w^*(P) = \sup_{b \in \mathcal{B}} E_P\{\log(b \cdot X)\} = r(P) - h^*(Q) \]
as the difference of an ideal reference level
\[ r(P) = E_P\{\log(m(\eta \cdot X))\} = E_P\{\log(X^1 + \ldots + X^m)\} \]
and a loss of expected log return with respect to that ideal reference level
\[ h^*(Q) = \inf_{b \in \mathcal{B}} E_Q\{-\log(b \cdot U)\}. \]

It is now straightforward to prove the A.E.P. for log-optimum investment in a market with finitely many stocks. It suffices to show that all hypotheses assumed in Theorem 5.2 are satisfied. Clearly the simplex $\mathcal{B}$ is compact and $\lambda(b, u) = (b \cdot u)$ is continuous on $\mathcal{B} \times \mathcal{U}$. In the following theorem we exploit the detailed properties of the inner product to prove that the expected loss of log return $h^*(Q)$ is bounded and continuous on $\mathcal{Q}$. Let $ri(\mathcal{B})$ and $rb(\mathcal{B})$ denote the relative interior and the relative boundary of the simplex $\mathcal{B}$.

**Theorem 6.1** Let $\mathcal{Q}$ be the space of probability measures on the simplex $\mathcal{U}$.
1. The function $h(b, Q) = E_Q\{-\log(b \cdot U)\}$ is lower semicontinuous on the compact product space $\mathcal{B} \times \mathcal{Q}$. In fact $h(b, Q)$ is continuous and finite on $\text{ri}(\mathcal{B}) \times \mathcal{Q}$, and discontinuous at any point of $\text{rb}(\mathcal{B}) \times \mathcal{Q}$ unless the value at that point is $+\infty$. However if $Q \in \mathcal{Q}$ is kept fixed, then $b \mapsto h(b, Q)$ is continuous on $\mathcal{B}$.

2. $\mathcal{Q}$ is convex and compact when equipped with the weak topology, and $h^*(Q) = \inf_{b \in \mathcal{B}} E_Q\{-\log(b \cdot U)\}$ is concave, bounded between 0 and $\log m$, and uniformly continuous in $Q \in \mathcal{Q}$.

**Proof:**

Clearly $h(b, Q) = E_Q\{-\log \lambda(b, U)\}$ and $h^*(Q) = \inf_{b \in \mathcal{B}} h(b, Q)$ are nonnegative and lower semicontinuous since $\mathcal{B}$ is compact and $\lambda(b, u) = (b \cdot u)$ is upper semicontinuous and bounded above by 1 on $\mathcal{B} \times \mathcal{U}$. Any function which is lower semicontinuous is continuous at all points where its value is $+\infty$.

If $K$ is any compact subset of $\text{ri}(\mathcal{B})$, then $-\log(b \cdot u)$ is bounded and uniformly continuous on $K \times \mathcal{U}$, hence $h(b, Q) = E_Q\{-\log(b \cdot U)\}$ is bounded and uniformly continuous on $K \times \mathcal{Q}$. By taking the union over all compact subsets $K$ of $\text{ri}(\mathcal{B})$, we see that $h(b, Q)$ is finite and continuous on $\text{ri}(\mathcal{B}) \times \mathcal{Q}$.

But we claim that $h(b, Q)$ cannot be continuous at any point $(b, Q)$ in $\mathcal{B} \times \mathcal{Q}$ such that $b \in \text{rb}(\mathcal{B})$ and $h(b, Q) < \infty$. In fact, for any $a \in \mathbb{R}_+$, we can exhibit a sequence $\{Q_n\}_n \subseteq \mathcal{Q}$ such that $Q_n \to Q$ in $\mathcal{P}$ and $h(b, Q_n) \to h(b, Q) + a$. It suffices to take

$$Q_n = \left(1 - \frac{1}{n}\right) Q + \frac{1}{n} \delta_{u_n}$$

where $\delta_{u_n}$ is a unit point mass or Dirac measure at $u_n$ and the sequence $\{u_n\}_n \subseteq \mathcal{U}$ is chosen so that $(b \cdot u_n) = \exp(-na)$. Indeed, we then have

$$h(b, Q_n) = \left(1 - \frac{1}{n}\right) h(b, Q) + \frac{1}{n} h(b, \delta_{u_n})$$

with

$$h(b, \delta_{u_n}) = E_Q\{-\log(b \cdot u_n)\} = na.$$ 

Such choice of $u_n$ is possible if $b \in \text{rb}(\mathcal{B})$, i.e., if some component $b^i$ of $b$ vanishes.

To prove that $b \mapsto h(b, Q)$ is continuous on $\mathcal{B}$ it suffices to show that $h(b_n, Q) \to h(b, Q)$ whenever $b_n \to b$ in $\mathcal{B}$. But this follows from the dominated convergence theorem. Indeed, the functions $-\log(b_n \cdot U)$ converge:

$$-\log(b_n \cdot U) \to -\log(b \cdot U)$$

$\mathcal{Q}$-a.s.
and they are dominated by an integrable function:

\[-\log(b_n \cdot U) \leq -\log(\frac{1}{2} b \cdot U) = \log 2 - \log(b \cdot U)\]

eventually for all \( n \geq n_0 \),

since \( b^j_n \geq b^j / 2 \) for all \( 1 \leq j \leq m \) and all \( n \geq n_0 \) if \( n_0 \) is sufficiently large.

Clearly \( h^* : \mathcal{Q} \to \mathbb{R}_+ \) is concave, since \( h^*(Q) = \inf_{b \in \mathcal{B}} h(b, Q) \) is the infimum of a family of functions \( h(b, Q) \) (indexed by \( b \in \mathcal{B} \)) which are affine on \( \mathcal{Q} \). In fact, \( h(b, Q) \) is continuous in \( b \in \mathcal{B} \) for fixed \( Q \in \mathcal{Q} \), so that the infimum over \( \mathcal{B} \) is the same as the infimum over \( r_\mathcal{B}(\mathcal{B}) \). But \( Q \mapsto h(b, Q) \) is bounded and continuous if \( b \in r_\mathcal{B}(\mathcal{B}) \), so that

\[
h^*(Q) = \inf_{b \in \mathcal{B}} h(b, Q) = \inf_{b \in r_\mathcal{B}(\mathcal{B})} h(b, Q)
\]

is bounded and upper semicontinuous. Since the uniform \( \eta \in \mathcal{B} \) is one of the competing portfolios and \( h(\eta, Q) = E_Q(-\log(\eta \cdot U)) = \log m \) we may conclude that \( h^*(Q) \leq \log m \).

QED.

Although we will not need this fact, the reader may observe that the function \( b \mapsto h(b, Q) = E_Q(-\log(b \cdot U)) \) is convex on \( \mathcal{B} \). In fact, let \( \mathcal{L}(Q) \) denote the linear hull of the support of \( Q \), that is, the smallest linear subspace \( \mathcal{L} \) of \( \mathbb{R}^m \) such that \( Q\{U \in \mathcal{L}\} = 1 \). Clearly \( \mathcal{L}(Q) \) can also be defined as the range space of the \( m \times m \) correlation matrix \( E_Q(U \otimes U) \) of \( U \). Using Jensen’s inequality for the strictly concave log function, one may show that the function \( b \in \mathbb{R}^m_+ \mapsto h(b, Q) = E_Q(-\log(b \cdot U)) \) is constant along fibers perpendicular to \( \mathcal{L}(Q) \), and strictly convex when restricted to the domain \( \mathcal{L}(Q) \cap \mathbb{R}^m_+ \). Thus the set \( B^*(Q) \) of log-optimum portfolios is the set of elements of \( \mathcal{B} \) which have the same projection on \( \mathcal{L}(Q) \) as \( b^*(Q) \).

**Horse races and entropy**

We now discuss a special type of investment game, introduced in the classical paper of Kelly [56]. Suppose \( m \) horses run a race, and \( J \) is the identity of the winner. \( J \) is a \( \{1, 2, \ldots, m\} \)-valued random variable whose distribution can be identified with a probability vector \( q = (q^j)_{1 \leq j \leq m} \) in the simplex \( \mathcal{U} \). Suppose furthermore that odds \( o = (o^j)_{1 \leq j \leq m} \in \mathbb{R}^m_+ \) are posted. These odds are called fair if \( o^j = m \) for \( 1 \leq j \leq m \), or more generally if they are derived from a probability measure \( \mu = (\mu^j)_{1 \leq j \leq m} \) on \( \{1, 2, \ldots, m\} \) by the rule \( o^j = 1/\mu^j \). We allow the odds \( o = (o^j)_{1 \leq j \leq m} \) to be any random vector in \( \mathbb{R}^m_+ \), for complete generality.

Placing a bet of one dollar on a horse \( j \) leads to a complete loss, unless \( j \) turns out to be the winner \( J \), and the return is given by \( o^J \) in that case. Gambling
on the outcome of a horse race is thus a special type of investment game, with return vector of the form \( X = o^j e_j \), where \( e_j \) is the \( j \)th unit vector in \( \mathbb{R}^m_+ \). The distinguishing feature that characterizes horse races within the class of all possible stock market problems is the fact that the outcome \( X \) is a vector in \( \mathbb{R}^m_+ \) oriented along one of the coordinate axes, or equivalently, the normalized outcome \( U \) is an extreme point of the simplex \( U \). Notice that \( U = e_j \) is the unit vector which indicates the winner \( J \). Since \( U \) is a simplex, there is a one-to-one correspondence between probability measures \( Q \) on the extreme points of \( U \) and probability vectors \( q \) in \( U \). With a vector \( q = (q^i)_{1 \leq i \leq m} \in U \), we can associate a measure \( Q_q \) on the set of extreme points \( \{ e_j \mid 1 \leq j \leq m \} \) of \( U \), namely \( Q_q = \sum_{1 \leq i \leq m} q^i \delta_{e_i} \). Conversely, with a distribution \( Q \) on \( U \), we can associate its barycenter \( q = E_Q\{ U \} = \int_U u \cdot Q(du) \).

Let \( P \) be the distribution of \( X = o^j e_j \) on \( \mathbb{R}^m_+ \), and \( q = (q^i)_{1 \leq i \leq m} \) the probability vector describing the distribution of \( J \) on \( \{ 1, 2, \ldots, m \} \). Suppose the gambler places bets \( b = (b^i)_{1 \leq i \leq m} \) on the horses \( j, 1 \leq j \leq m \). When the race is over and the winner \( J \) is known, the gambler will collect a return \( (b \cdot X) = b^j o^j \).

The expected log return is given by

\[
\begin{align*}
    w(b, P) &= E_P\{ \log(b \cdot X) \} = E_P\{ \log(b^j o^j) \} \\
    &= E_q\{ \log(b^j/q^j) \} + E_P\{ \log(q^j o^j) \} \\
    &= -D(q\|b) + E_P\{ \log(q^j o^j) \}.
\end{align*}
\]

Recall that \( D(q\|b) \geq 0 \), with equality iff \( b = q \), since \( b \) and \( q \) are normalized probability vectors. Thus the maximum expected log return is uniquely achieved by ignoring the odds and placing proportional bets \( b^* = q \), since \( D(q\|b^*) = 0 \) in that case. This means that the amount \( b^*o^j \) invested in each horse \( j \) should be proportional to the probability of its victory. For the maximum expected log return \( w^*(P) \) we may write

\[
\begin{align*}
    w^*(P) &= E_P\{ \log(q^j o^j) \} \\
    &= E_P\{ \log o^j \} - E_q\{ \log(1/q^j) \} \\
    &= E_P\{ \log o^j \} - h^*(q) = r(P) - h^*(q).
\end{align*}
\]

Notice that the log-optimum portfolio \( b^*(P) \) and the maximum expected log return \( w^*(P) \) are computable in explicit form. In particular, if the odds \( o = (o^j)_{1 \leq i \leq m} \) are derived from a fixed probability vector \( \mu \) by the rule \( o^j = 1/\mu^j \), then \( w^*(P) \) is equal to the Kullback-Leibler divergence \( D(q\|\mu) \). The ideal reference level \( r(P) = E_P\{ \log o^j \} \) is equal to the expected log of the odds, and thus equal to \( \log m \) in the case of fair odds derived from the uniform distribution.
\( \eta = (1/m)_{1 \leq i \leq m} \). The minimum loss of expected log return with respect to that reference level is given by the Shannon entropy

\[
h^\ast(q) = E_Q\{\log(1/q^i)\} = -\sum_{1 \leq i \leq m} q^i \log q^i.
\]

Since the minimum loss of expected log return \( h^\ast(Q) = \inf_{b \in \mathbb{B}} E_Q\{-\log(b \cdot U)\} \)
reduces to the Shannon entropy \( h^\ast(Q_q) = h^\ast(q) = -\sum_{1 \leq i \leq m} q^i \log q^i \) in case \( Q \) is a distribution \( Q_q \) on the extreme points of the simplex \( \mathcal{U} \), it is appropriate to call \( h^\ast(Q) \) the generalized entropy of \( Q \) when \( Q \) is an arbitrary probability measure on the simplex \( \mathcal{U} \). We have shown that the generalized entropy \( h^\ast(Q) \) has the same desirable properties as its restriction \( h^\ast(q) \): \( h^\ast(Q) \) is concave, bounded and uniformly continuous in \( Q \), and \( 0 \leq h^\ast(Q) \leq \log m \). In fact, it is straightforward to show that

\[
0 \leq E_Q\{-\log(\max_{1 \leq i \leq m} U^i)\} \leq h^\ast(Q) = \inf_{b \in \mathbb{B}} E_Q\{-\log(b \cdot U)\} \leq E_Q\{-\log(\eta \cdot U)\} = \log m.
\]

When we translate this back into a statement about the maximum expected log return \( w^\ast(P) \), we obtain

\[
R = E_P\{\log(X_1 + \ldots + X_m)\} \geq E_P\{\log(\max_{1 \leq i \leq m} X^i)\} \\
\geq w^\ast(P) = R - h^\ast(Q) = \sup_{b \in \mathbb{B}} E_P\{\log(b \cdot X)\} \\
\geq E_Q\{\log(\eta \cdot X)\} = R - \log m.
\]

The maximum expected log return \( w^\ast(P) \) is convex but neither bounded nor continuous on the space of probability measures on \( \mathcal{R}^m_+ \), equipped with the weak topology. However if we restrict our attention to distributions \( P \) whose support is contained in a closed subset \( K \) of \( \mathcal{R}^m_+ \), then \( w^\ast(P) \) is lower semicontinuous and bounded below iff \( K \) is bounded away from \((0,0,\ldots,0)\), upper semicontinuous and bounded above iff \( K \) is bounded, and bounded and uniformly continuous iff \( K \) is bounded and bounded away from \((0,0,\ldots,0)\).

**Sequential investment**

Suppose \( \{X_t = X \circ T^t\}_{-\infty < t < \infty} \) is a stationary ergodic stock market process and \( \{\mathcal{F}_t\}_{0 \leq t < \infty} \) a sequence of information fields such that the shifted fields \( \mathcal{F}_t = T^t \mathcal{F}_t \) increase monotonically towards a limiting \( \sigma \)-field \( \mathcal{F}_\infty \). The limiting capital growth rate per period can then be written as the difference \( \bar{W}_\infty^* = R - \bar{H}_\infty^* \) of the ideal reference level \( R = \mathbb{E}\{\log(X_1 + \ldots + X^m)\} \) and the generalized entropy rate

\[
\bar{H}_\infty^* = \lim_{t \to \infty} \mathbb{E}\{h^\ast(Q_t)\} = \mathbb{E}\{h^\ast(Q_\infty)\}.
\]
Here $\bar{Q}_t$ is the conditional distribution of $U_0 = U$ given $\mathcal{F}_t$. There are two reasons that the generalized entropy rate $\bar{H}_\infty^*$ may be strictly positive. First, if more than one stock $j$ yields a nonzero return $X^j \neq 0$, then $\max_{1 \leq j \leq m} X^j < X^1 + \ldots + X^m$ and the first in the above chain of inequalities will be strict. The second inequality is strict if there is a chance that the log-optimum investment strategy $\bar{b}_\infty^*$ allocates some funds to stocks which yield less than the maximum return $\max_{1 \leq j \leq m} X^j$. Thus zero loss with respect to the ideal reference level $R = \mathbb{E}\{\log(X^1 + \ldots + X^m)\}$ can be achieved in the limit only if we are in the horse race case, and it must be possible to predict with certainty which horse will win on the basis of the infinite past (i.e., the limiting information field $\mathcal{F}_\infty$).
SECTION 7.

SHANNON-McMILLAN-BREIMAN THEOREM

The names of Shannon [48], McMillan [53] and Breiman [57,60] are attached to the following classical theorem of information theory. Suppose \( \{X_t\}_{-\infty < t < \infty} \) is a stationary ergodic process with values in a finite or countable set \( \mathcal{X} \), and 
\[
p(x_t|X_{t-1}, \ldots, x_0)
\]
and 
\[
p(x_0, \ldots, x_{n-1})
\]
denote the conditional and joint probability mass functions of the process. Then 
\[
-\frac{1}{n} \log p(X_0, X_1, \ldots, X_{n-1}) = -\frac{1}{n} \sum_{0 \leq i < n} \log p(X_t|X_{t-1}, \ldots, X_0) \to H \quad \text{a.s.}
\]
where \( H = \mathbb{E} \{-\log p(X_0|X_{-1}, X_{-2}, \ldots)\} \) is the entropy rate of the process. This statement is often called the asymptotic equipartition property because it asserts the existence of a set of roughly \( \exp(nH) \) typical sequences in \( \mathcal{X}^n \), all having roughly equal probability \( \exp(-nH) \), for large \( n \).

Barron [84] has recently generalized this almost sure convergence theorem to processes with densities. Let us assume that \( \mathcal{X} \) is a standard Borel space, and 
\[
p(X_t|X_{t-1}, \ldots, X_0)
\]
and 
\[
p(X_0, \ldots, X_{n-1})
\]
denote the conditional and joint likelihood ratios of the process with respect to a fixed reference measure \( m \) on \((\mathcal{X}, \mathcal{B}_\mathcal{X})\):
\[
p(x_t|x_{t-1}, \ldots, x_0) = \frac{\mathbb{P}\{X_t \in dx_t|X_{t-1} = x_{t-1}, \ldots, X_0 = x_0\}}{m\{X_t \in dx_t\}}
\]
and 
\[
p(x_0, \ldots, x_{n-1}) = \frac{\mathbb{P}\{X_0 \in dx_0, \ldots, X_{n-1} \in dx_{n-1}\}}{m\{X_0 \in dx_0\} \ldots m\{X_{n-1} \in dx_{n-1}\}}
\]
\[
= \prod_{0 \leq i < n} p(x_i|x_{i-1}, \ldots, x_0).
\]
The average log-likelihood ratio, when evaluated at the true observations, converges to a limiting relative entropy rate \( \tilde{\Delta}^*_\infty \) with probability one:
\[
-\frac{1}{n} \log p(X_0, X_1, \ldots, X_{n-1}) = -\frac{1}{n} \sum_{0 \leq i < n} \log p(X_t|X_{t-1}, \ldots, X_0) \to \tilde{\Delta}^*_\infty \quad \text{a.s.}
\]
where
\[
\tilde{\Delta}^*_\infty = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\{(1/n) \log p(X_0, X_1, \ldots, X_{n-1})\}
\]
\[
= \lim_{t \to \infty} \mathbb{E}\{\log p(X_t|X_{t-1}, \ldots, X_0)\}
\]
\[
= \mathbb{E}\{\log p(X_0|X_{-1}, X_{-2}, \ldots)\}.
\]

62
Barron's essential contribution was a martingale inequality showing that the shifted random variables \( \{ \log p(X_0|X_{-1}, \ldots, X_{-i}) \}_{0 \leq i < \infty} \) are \( L^1 \)-dominated. He further proves that this sequence converges almost surely. The Cesàro averages \( (1/n) \sum_{0 \leq i < n} \log p(X_i|X_{i-1}, \ldots, X_0) \) then converge almost surely towards the expectation of the limit, by the extension of Birkhoff's ergodic theorem that was originally used by Breiman [57,60].

This section provides a new proof of the Shannon-McMillan-Breiman theorem. We show that this theorem is equivalent with the generalized A.E.P. for log-optimum investment in a stationary ergodic stock market of a very special type, namely gambling on the outcome of a horse race. First we shall describe this investment game in some detail.

**Gambling as an investment problem**

Let \( \mathcal{X} \) be the space of horses (not necessarily countable). We assume that \( \mathcal{X} \) is a universally measurable subset of a Polish space so that every probability measure on \( \mathcal{X} \) is perfect and regular conditional probability distributions exist. A bookie publicizes odds by posting a reference measure \( m \) and the gambler decides how to distribute his capital over the horses by choosing a betting measure \( b \) on \( (\mathcal{X}, \mathcal{B}_\mathcal{X}) \). The set \( \mathcal{B} \) of possible actions therefore coincides with the space of probability measures on \( (\mathcal{X}, \mathcal{B}_\mathcal{X}) \). The return function \( \rho(b, x) \) remains to be defined; this will be done via a limiting process.

Suppose the bookie and the gambler agree on some partition \( \alpha \) of \( \mathcal{X} \) into a finite or countable collection of measurable subsets. Agreement about \( \alpha \) is reached before the race. When the race is over and the identity \( X \in \mathcal{X} \) of the winning horse is known, the bookie may collect bets placed on loosing horses but has the obligation to pay out the amount \( b(\alpha(X)) \) multiplied by the odds \( 1/m(\alpha(X)) \). Here \( \alpha(X) \) denotes the atom of the partition \( \alpha \) into which the winner \( X \) is classified. The return to the gambler is therefore given by the likelihood ratio

\[
\rho_\alpha(b, X) = \frac{b(\alpha(X))}{m(\alpha(X))}.
\]

Let \( w_\alpha(b, P) \) denote the expected log return of betting strategy \( b \) when \( X \) has
distribution $P$:

$$w^\alpha(b, P) = E_P \{ \log \rho^\alpha(b, X) \} = E_P \left\{ \log \left( \frac{b(\alpha(X))}{m(\alpha(X))} \right) \right\}$$

$$= \sum_{A \in \alpha} P\{A\} \cdot \log \left( \frac{b(A)}{m(A)} \right)$$

$$= \sum_{A \in \alpha} P\{A\} \cdot \log \left( \frac{P\{A\}}{m(A)} \right) - \sum_{A \in \alpha} P\{A\} \cdot \log \left( \frac{P\{A\}}{b(A)} \right)$$

$$= D(P_\alpha || m_\alpha) - D(P_\alpha || b_\alpha).$$

Here $P_\alpha$, $m_\alpha$, and $b_\alpha$ denote the restrictions of the measures $P$, $m$ and $b$ to the finite or countable partition $\alpha \subseteq 2^X$, and the summations $\sum_{A \in \alpha}$ extend over all atoms $A$ of $\alpha$. The Kullback-Leibler divergence $D(P_\alpha || b_\alpha)$ is nonnegative, and equal to zero iff $b_\alpha = P_\alpha$. Thus the maximum expected log return $w^{\alpha^*}(P)$ is given by

$$w^{\alpha^*}(P) = D(P_\alpha || m_\alpha)$$

and is achieved by any bet measure $b$ for which $b_\alpha = P_\alpha$. There may be many such gambling strategies $b$ for a given $\alpha$, but there is only one which is log-optimum for all measurable partitions $\alpha$, namely the proportional betting scheme $b^*(P) = P$. According to this rule the gambler should ignore the odds posted by the bookie and distribute his funds over various pieces of the outcome space $X$ in proportion of their probability of occurrence:

$$b^*(A) = P\{A\} \quad \text{for all } A \in 2^X.$$

The maximum expected log return $w^{\alpha^*}(P) = D(P_\alpha || m_\alpha)$ increases when the partition $\alpha$ is refined: if $\beta$ is a finer measurable partition than $\alpha$ then

$$D(P_\beta || m_\beta) = \sum_{B \in \beta} P\{B\} \cdot \log \left( \frac{P\{B\}}{m\{B\}} \right)$$

$$= \sum_{A \in \alpha} P\{A\} \cdot \log \left( \frac{P\{A\}}{m\{A\}} \right) + \sum_{A \in \alpha} P\{A\} \cdot \sum_{B \in \beta : B \subseteq A} P\{B\} | A \cdot \log \left( \frac{P\{B\} | A}{m\{B\} | A} \right)$$

$$= D(P_\alpha || m_\alpha) + \sum_{A \in \alpha} P\{A\} \cdot D(P_\beta | A || m_\beta | A)$$

$$= D(P_\alpha || m_\alpha) + E_P \{ D(P_\beta | \alpha(X) || m_\beta | \alpha(X)) \} \geq D(P_\alpha || m_\alpha).$$

Here $P_\beta | A = (P\{B\} | A)_{B \in \beta : B \subseteq A}$ and $m_\beta | A = (m\{B\} | A)_{B \in \beta : B \subseteq A}$ are vectors of conditional probabilities. The limit of $D(P_\alpha || m_\alpha)$ as the finite or countable partition $\alpha$ is more and more refined is, by definition, the Kullback-Leibler divergence $D(P || m)$:

$$D(P_\alpha || m_\alpha) \searrow D(P || m) = \sup_{\alpha} D(P_\alpha || m_\alpha) \quad \text{as } \alpha \uparrow.$$
In fact, let us consider the true log returns under the log-optimum strategy \( b^*(P) \):

\[
\log \rho^\alpha(b^*(P), X) = \log \left( \frac{P\{\alpha(X)\}}{m\{\alpha(X)\}} \right).
\]

The reader may verify that \( \{\log \rho_\alpha(b^*(P), X), \sigma(\alpha)\}_\alpha \) is a submartingale indexed by the directed sets of finite or countable partitions \( \alpha \), at least if \( D(P\|m) < \infty \).

Since the expected log return \( w^*(P) = D(P_\alpha\|m_\alpha) \) increases when \( \alpha \) is refined, it will be in the interest of the gambler to insist on as fine a level of granularity for partitioning \( X \) as is practically feasible before the race. The bookie has a similar incentive if he believes that the odds measure \( m \) truly reflects the distribution of the outcome \( X \). For these reasons we shall define the return function \( \rho(b, x) \) as the limit of the likelihood ratios \( \rho^\alpha(b, x) = b\{\alpha(X)\}/m\{\alpha(X)\} \)
along the directed set of finite or countable partitions \( \alpha \) when \( \alpha \) is more and more refined:

\[
\rho(b, x) = \lim_{\alpha \uparrow} \rho^\alpha(b, x) = \lim_{\alpha \uparrow} \frac{b\{\alpha(X)\}}{m\{\alpha(X)\}}.
\]

Alternatively,

\[
\rho(b, x) = \frac{b(dx)}{m(dx)} = \lim_{A \downarrow \{x\}} \frac{b(A)}{m(A)}
\]
is the limit of the likelihood ratio \( b\{A\}/m\{A\} \) along the filter of neighborhoods \( A \) shrinking to \( \{x\} \). Clearly \( \rho(b, x) \) is measurable on \( B \times X \), and affine in \( b \in B \) for fixed \( x \in X \). If \( b \ll m \) then

\[
\rho(b, x) = \frac{db}{dm}(x)
\]
is the Radon-Nikodym derivative of \( b \) with respect to the dominating measure \( m \). In general,

\[
\rho(b, x) = \begin{cases} 
\frac{db}{dm}(x) & \text{if } x \in \text{Support}(b_a) \\
\infty & \text{if } x \in \text{Support}(b_s)
\end{cases}
\]

where \( b = b_a + b_s \) is the Lebesgue decomposition of the measure \( b \) relative to \( m \) into an absolutely continuous part \( b_a \ll m \) and a singular part \( b_s \perp m \).

The expected log return under portfolio \( b \) when the outcome \( X \) has distribution \( P \) is given by

\[
w(b, P) = E_P\{\log \rho(b, X)\} = E_P \left\{ \log \left( \frac{b\{X \in dx\}}{m\{X \in dx\}} \right) \right\}
\]

\[
= E_P \left\{ \log \left( \frac{P\{X \in dx\}}{m\{X \in dx\}} \right) \right\} - E_P \left\{ \log \left( \frac{P\{X \in dx\}}{b\{X \in dx\}} \right) \right\}
\]

\[
= D(P\|m) - D(P\|b).
\]
Hence the maximum expected log return is given by

$$w^*(P) = D(P||m)$$

and is uniquely achieved by $b^*(P) = P$. Observe that $b^*(P) = P$ is a continuous and therefore certainly a measurable function of the distribution $P$. Also $w^*(P) = D(P||m)$ is nonnegative, convex, and lower semicontinuous in $P$. The log return under the log-optimum strategy $b^*(P) = P$ is the density of the distribution $P$ with respect to the reference measure $m$, when evaluated at the actual outcome $X$:

$$\log \rho(b^*(P), X) = \log p(X) \text{ where } p(x) = \frac{P\{X \in dx\}}{m\{X \in dx\}}.$$

The sequential problem

We now consider the sequential gambling problem. Suppose the underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is equipped with an invertible measure preserving and metrically transitive transformation $T$, so that $\{X_t = X \circ T^t\}_{-\infty < t < \infty}$ is a stationary ergodic process. Let $\{\mathcal{F}_t\}_{0 \leq t < \infty}$ be a sequence of information fields such that the shifted fields $\mathcal{F}_t = T^t \mathcal{F}_0$ increase monotonically towards a limiting $\sigma$-field $\mathcal{F}_\infty$. Regular conditional probability distributions of $X_t$ given $\mathcal{F}_t$ and of $X_0$ given $\mathcal{F}_t$ exist and are denoted by $P_t$ and $\bar{P}_t$, respectively. The return for period $t$ under the log-optimum strategy $b^*_t = P_t$ is given by the conditional density of $X_t$ given $\mathcal{F}_t$:

$$\rho(b^*_t, X_t) = \frac{dP_t}{dm}(X_t) = p(X_t | \mathcal{F}_t)$$

where

$$p(x_t | \mathcal{F}_t) = \frac{P_t\{X_t \in dx_t\}}{m\{X_t \in dx_t\}} = \frac{P\{X_t \in dx_t | \mathcal{F}_t\}}{m\{X_t \in dx_t\}}.$$

For the log return $w^*_t$ under the log-optimum strategy $b^*_t = P_t$, its conditional expectation $w^*_t = \mathbb{E}\{w^*_t | \mathcal{F}_t\}$, and its unconditional expectation $W^*_t = \mathbb{E}\{w^*_t\} = \mathbb{E}\{w^*_t\}$, we can therefore write

$$w^*_t = \log \rho(b^*_t, X_t) = \log p(X_t | \mathcal{F}_t);$$

$$w^*_t = \mathbb{E}\{\log p(X_t | \mathcal{F}_t) | \mathcal{F}_t\} = D(P_t || m);$$

$$W^*_t = \mathbb{E}\{\log p(X_t | \mathcal{F}_t)\} = \mathbb{E}\{D(P_t || m)\} = \Delta^*_t.$$

The maximum conditional expected log return $w^*_t$ is equal to the Kullback-Leibler divergence $D(P_t || m)$ between the conditional distribution $P_t$ and the reference measure $m$, and the maximum unconditional expected log return is given by
the relative entropy $\Delta^*_i = \mathbb{E} \left\{ D(P_i \| m) \right\}$. Similar expressions hold for shifted quantities. The return for period 0 under the log-optimum strategy $\tilde{b}^*_i = \tilde{P}_i$ based on the shifted information field $\mathcal{F}_i$ is given by

$$\rho(\tilde{b}^*_i, X_0) = \frac{d\tilde{P}_i}{dm}(X_0) = p(X_0|\mathcal{F}_i)$$

where $p(x_0|\mathcal{F}_i)$ is the conditional density of $X_0$ given $\mathcal{F}_i$:

$$p(x_0|\mathcal{F}_i) = \frac{\tilde{P}_i\{X_0 \in dx_0\}}{m\{X_0 \in dx_0\}} = \frac{\mathbb{P}\{X_0 \in dx_0|\mathcal{F}_i\}}{m\{X_0 \in dx_0\}}.$$

Thus we can write formulas for the log return during period 0 under the log-optimum strategy based on $\mathcal{F}_i$, for its conditional expectation, and for its unconditional expectation:

$$\tilde{w}^*_i = \log p(X_0|\mathcal{F}_i);$$

$$\tilde{w}_i = \mathbb{E} \left\{ \tilde{w}^*_i | \mathcal{F}_i \right\} = \mathbb{E} \left\{ \log p(X_0|\mathcal{F}_i) | \mathcal{F}_i \right\} = D(\tilde{P}_i \| m);$$

$$\tilde{W}^*_i = \mathbb{E} \left\{ \tilde{w}^*_i \right\} = \mathbb{E} \left\{ \tilde{w}_i \right\} = \mathbb{E} \left\{ \log p(X_0|\mathcal{F}_i) \right\} = \mathbb{E} \left\{ D(\tilde{P}_i \| m) \right\} = \tilde{\Delta}^*_i.$$  

The asymptotic equipartition property for log-optimum investment reduces to a generalization of the Shannon-McMillan-Breiman theorem in case the investment game consists of gambling on the outcome of a horse race.

**Theorem 7.1** Let $\{X_i\}_{-\infty \leq i \leq \infty}$ be a stationary ergodic process taking values in a standard space and let $\{\mathcal{F}_i\}_{0 \leq i \leq \infty}$ be a sequence of information fields such that the shifted fields $\mathcal{F}_i = T^i \mathcal{F}_i$ increase monotonically towards a limiting $\sigma$-field $\mathcal{F}_\infty$. Then

1. $\Delta^*_i = \tilde{\Delta}^*_i$ by stationarity and $\tilde{\Delta}^*_i \nearrow \tilde{\Delta}^*_\infty$. Hence
   $$\frac{1}{n}\mathbb{E} \left\{ \log p(X_0, \ldots, X_{n-1}) \right\} = \left(\frac{1}{n}\right) \sum_{0 \leq i \leq n} \Delta^*_i \nearrow \tilde{\Delta}^*_\infty.$$  

2. $\{D(\tilde{P}_i \| m), \mathcal{F}_i\}_{0 \leq i \leq \infty}$ is a submartingale (strictly speaking only if $\tilde{\Delta}^*_\infty < \infty$) and $D(\tilde{P}_i \| m) \to D(\tilde{P}_\infty \| m)$ a.s. (and in $L^1$ if $\tilde{\Delta}^*_\infty < \infty$).

3. $\log p(X_0|\mathcal{F}_i) \to \log p(X_0|\mathcal{F}_\infty)$ a.s. (and in $L^1$ if $\tilde{\Delta}^*_\infty < \infty$).

4. $\frac{1}{n} \sum_{0 \leq i \leq n} D(\tilde{P}_i \| m) \to \tilde{\Delta}^*_\infty$ a.s and in $L^1$, at least if $D(\tilde{P}_\infty \| m) \in L \log L$.

5. $\frac{1}{n} \sum_{0 \leq i \leq n} \log p(X_i|\mathcal{F}_i) \to \tilde{\Delta}^*_\infty$ a.s. (and in $L^1$ if $\tilde{\Delta}^*_\infty < \infty$).
PROOF:

These assertions are merely specializations of results proved earlier, for log-optimum investment in general, to the case where the investment game consists of gambling on the outcome of a horse race. The only point that requires proof is that $\tilde{\Delta}_i^* \not\to \tilde{\Delta}_\infty^*$. But this follows from Fatou's lemma and the lower semicontinuity of the function $P \mapsto D(P||m)$ as $P$ ranges over the space $P$ of probability measures on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ with the weak topology.

QED.

Remark 1. We obtain the classical Shannon-McMillan-Breiman theorem if we choose $\mathcal{F}_t = \sigma(X_0, X_1, \ldots, X_{t-1})$:

$$\frac{1}{n} \sum_{0 \leq t < n} \log p(X_t|X_0, X_1, \ldots, X_{t-1}) = \frac{1}{n} \log p(X_0, X_1, \ldots, X_{n-1}) \to \tilde{\Delta}_\infty^* \quad \text{a.s.}$$

where

$$\tilde{\Delta}_\infty^* = \mathbb{E}\{\log p(X_0, \ldots, X_{-2}, X_{-1})\}$$

and

$$p(x_0, x_1, \ldots, x_{n-1}) = \frac{P\{X_0 \in dx_0, \ldots, X_{n-1} \in dx_{n-1}\}}{m\{X_0 \in dx_0\} \ldots m\{X_{n-1} \in dx_{n-1}\}}.$$  

Remark 2. Suppose the return function has the form

$$\rho(b, x) = \frac{b\{X \in dx\}}{\ell\{X \in dx\}} \left( = \frac{db}{dm}(x) \quad \text{if} \ b \ll \ell \right)$$

where $\ell$ is an $\sigma$-finite measure on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$. Then there exists a probability measure $m$ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ such that $\ell$ and $m$ are mutually absolutely continuous, and the return function can be written as

$$\rho(b, x) = \rho'(b, x) \cdot \frac{dm}{d\ell}(x) \quad \text{where} \quad \rho'(b, x) = \frac{b\{X \in dx\}}{m\{X \in dx\}}.$$  

Notice that $\rho'(b, x)$ is a return function of the type we have been dealing with all along. Moreover the log returns $\log \rho(b, x)$ and $\log \rho'(b, x)$ differ only by a quantity which is independent of the portfolio $b$:

$$\log \rho(b, x) = \log \rho'(b, x) + \log \left(\frac{dm}{d\ell}(x)\right).$$

Hence the return functions $\rho$ and $\rho'$ are completely equivalent as far as the guidance they provide for selecting log-optimum portfolios. It is clear that the asymptotic equipartition property continues to hold even if the reference measure is $\sigma$-finite.
Remark 3. All likelihood ratios we have considered so far were taken with respect to (independent copies of) a fixed reference measure \( m \) on \((\mathcal{X}, \mathcal{B}_\mathcal{X})\). However the A.E.P. remains valid for a Markovian reference measure. Indeed, suppose that \( \{\mathcal{G}_t\}_{0 \leq l < \infty} \) is a second sequence of sub-\( \sigma \)-fields of \( \mathcal{F} \) such that the shifted fields \( \tilde{\mathcal{G}}_t = T^l \mathcal{G}_t \) increase monotonically towards a limiting \( \sigma \)-field \( \tilde{\mathcal{G}}_\infty \). Let \( \hat{\mathcal{M}} \) be a perfect probability measure on \((\Omega, \mathcal{F})\) such that \( \{X_t\}_{-\infty \leq t \leq \infty} \) is stationary ergodic under \( \hat{\mathcal{M}} \), and let

\[
p_t = \frac{\mathcal{P}\{X_t \in d\omega | \mathcal{F}_t\}}{\hat{\mathcal{M}\{X_t \in d\omega | \mathcal{G}_t\}}} \quad \text{for } 0 \leq t < \infty
\]

and

\[
\bar{p}_t = \frac{\mathcal{P}\{X_0 \in d\omega_0 | \mathcal{F}_t\}}{\hat{\mathcal{M}\{X_0 \in d\omega_0 | \tilde{\mathcal{G}}_t\}}} \quad \text{for } 0 \leq t \leq \infty.
\]

Assume \( X_0 \) and \( \tilde{\mathcal{G}}_t \) be conditionally independent given \( \tilde{\mathcal{G}}_t \), for some \( 0 \leq l < \infty \) and all \( l \leq t < \infty \). (If \( \mathcal{G}_t = \sigma(X_0, X_1, \ldots, X_{t-l}) \) for all \( 0 \leq t < \infty \), then this condition means that \( \{X_t\}_{-\infty \leq t \leq \infty} \) is Markov with finite order \( l \) under \( \hat{\mathcal{M}} \).) It can then be shown that

\[
\mathbb{E}\{\log \bar{p}_k\} \leq \lim \inf_{n \to \infty} \frac{1}{n} \log S^*_n \leq \lim \sup_{n \to \infty} \frac{1}{n} \log S^*_n \leq \mathbb{E}\{\log \bar{p}_\infty\} \quad \text{a.s.}
\]

for all \( l \leq k < \infty \), and \( \mathbb{E}\{\log \bar{p}_k\} \neq \mathbb{E}\{\log \bar{p}_\infty\} \) a.s. Consequently

\[
\frac{1}{n} \sum_{0 \leq t < n} \log p_t \to \mathbb{E}\{\log \bar{p}_\infty\} \quad \text{a.s. (and in } L^1 \text{ if } \mathbb{E}\{\log \bar{p}_\infty\} \text{ is finite).}
\]
SECTION 8.

ERGODIC DECOMPOSITION FOR STATIONARY AND A.M.S. MARKETS

Finally we shall prove the generalized asymptotic equipartition property for a stock market which is stationary or even asymptotically mean stationary (a.m.s.) but not necessarily ergodic. Although capital will grow exponentially in the long run with highest possible rate, this rate is now a random variable whose value may depend on the particular realization $\omega \in \Omega$.

Log-optimum investment for a stationary market

Suppose $T$ is an invertible and measure preserving but not necessarily metrically transitive transformation of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Thus the stock market process $\{X_t = X \circ T^t\}_{-\infty < t < \infty}$ is stationary but not necessarily ergodic. If $g$ is a random variable whose expectation $\mathbb{E}\{g\}$ is well defined, then Birkhoff's ergodic theorem asserts that

$$
\frac{1}{n} \sum_{0 \leq t < n} g \circ T^t \to \mathbb{E}\{g|I\} \quad \text{a.s. (and in } L^1 \text{ if } \mathbb{E}\{g\} \text{ is finite).}
$$

Here $I$ denotes the $\sigma$-field of almost surely $T$-invariant events:

$$
I = \{F \in \mathcal{F} : T^{-1}F = F \ \text{a.s.}\}.
$$

A random variable $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measurable with respect to $I$ iff $g$ is $T$-invariant almost surely, that is, iff $g = g \circ T$ a.s. Notice that $T$ is metrically transitive iff $I$ is trivial, that is, iff all $T$-invariant events have probability equal to 0 or 1, or equivalently, iff all $T$-invariant random variables are constant almost surely (equal to their expectation, if that is well-defined). It is easy to adapt our previous proof of the asymptotic equipartition property to the general stationary case.

Theorem 8.1 Let $\{X_t\}_{-\infty < t < \infty}$ be a stationary but not necessarily ergodic process and $\{\mathcal{F}_t\}_{0 \leq t < \infty}$ a sequence of information fields such that the shifted fields $\mathcal{F}_t = T^t \mathcal{F}_0$ increase monotonically towards a limiting $\sigma$-field $\mathcal{F}_\infty$. If $\tilde{W}_t^* \xrightarrow{\mathcal{F}_\infty} \tilde{W}_\infty^*$ then the asymptotic equipartition property holds:

$$
\frac{1}{n} \log S_n^* = \frac{1}{n} \sum_{0 \leq t < n} \log \rho(b^*_t, X_t) \to \mathbb{E}\{\tilde{w}_\infty^*|I\} \quad \text{a.s.}
$$
or equivalently,

\[ S^*_n = \exp[n(\mathbb{E} \{ \tilde{w}^*_\infty | I \} + o(1))] \quad \text{where } o(1) \to 0 \text{ a.s.} \]

where

\[ \tilde{w}^*_\infty = \log \rho(\tilde{b}^*_\infty, X_0). \]

There is also convergence in \( L^1 \) if \( \tilde{W}^*_\infty = \mathbb{E} \{ \tilde{w}^*_\infty \} = \sup_{b \in \mathcal{R}_+} \mathbb{E} \{ \log \rho(b, X_0) \} \) is finite.

**Proof:**

Straightforward generalization of Theorem 4.4.

QED.

**Ergodic decomposition for stationary markets**

Let \( \mathcal{S}_T \) denote the collection of \( T \)-invariant probability measures on \((\Omega, \mathcal{F})\), that is, those probability measures \( P \) on \((\Omega, \mathcal{F})\) such that \((\Omega, \mathcal{F}, P, T)\) is a stationary dynamical system, or \( P(F) = P(T^{-1}F) \) for all \( F \in \mathcal{F} \). Then \( \mathcal{S}_T \) is a convex set and the set \( \mathcal{E}_T \) of extreme points of \( \mathcal{S}_T \) coincides precisely with the collection of probability measures \( \epsilon \) on \((\Omega, \mathcal{F})\) such that \((\Omega, \mathcal{F}, \epsilon, T)\) is a stationary ergodic dynamical system. We equip \( \mathcal{E}_T \) with the smallest \( \sigma \)-field \( \Sigma_T \) such that the functions \( \epsilon \mapsto \epsilon(F) \) are measurable on \((\mathcal{E}_T, \Sigma_T)\) for all \( F \in \mathcal{F} \).

A stationary measure \( P \in \mathcal{S}_T \) admits an ergodic decomposition if there exists a probability measure \( \pi \) on \((\mathcal{E}_T, \Sigma_T)\) which admits \( P \) as barycenter, that is,

\[ P = \int_{\mathcal{E}_T} \epsilon \cdot \pi(d\epsilon) \quad \text{or} \quad P(F) = \int_{\mathcal{E}_T} \epsilon(F) \cdot \pi(d\epsilon) \quad \text{for all } F \in \mathcal{F}, \]

and hence,

\[ E_P\{g\} = \int_{\Omega} g \cdot dP = \int_{\mathcal{E}_T} \left( \int_{\Omega} g(\omega) \cdot \epsilon(d\omega) \right) \cdot \pi(d\epsilon) \]

for all \( g : (\Omega, \mathcal{F}) \to (\mathcal{R}_+, \mathcal{B}_{\mathcal{R}_+}) \). A sample point with distribution \( P \) on \((\Omega, \mathcal{F})\) can then be obtained by a two-stage operation: first we select a stationary ergodic mode \( \epsilon \) with distribution \( \pi \) on \((\mathcal{E}_T, \Sigma_T)\), and then a sample point \( \omega \) with distribution \( \epsilon \) on \((\Omega, \mathcal{F})\). Thus as the system evolves over time it will be locked in some ergodic mode \( \epsilon \) which was selected a priori according to the distribution \( \pi \) on \((\mathcal{E}_T, \Sigma_T)\).

If \( P \in \mathcal{S}_T \) and \( g : (\Omega, \mathcal{F}) \to (\mathcal{R}_+, \mathcal{B}_{\mathcal{R}_+}) \) then the ergodic theorem asserts that

\[ \frac{1}{n} \sum_{0 \leq t < n} g \circ T^t \to E_P\{g|I\} \quad \text{P-a.s. (and in } L^1(P) \text{ if } E_P\{g\} \text{ is finite).} \]
The limiting random variable \( E_P\{g|I\} \) can be interpreted as the expectation \( E_\varepsilon\{g\} \) under the randomly selected stationary ergodic mode \( \varepsilon \in \mathcal{E}_T \), and the expectation \( E_P\{g\} \) is obtained by averaging \( E_\varepsilon\{g\} \) over all ergodic modes \( \varepsilon \):

\[
E_P\{g\} = \int_{\mathcal{E}_T} E_\varepsilon\{g\} \cdot \pi(d\varepsilon).
\]

Suppose the measurable space \((\Omega, \mathcal{F})\) is countably separated and every probability measure on \((\Omega, \mathcal{F})\) is perfect. This will be the case if \((\Omega, \mathcal{F})\) can be identified with a universally measurable subset of a Polish space, equipped with its Borel \( \sigma \)-field. Maitra [77] has shown that \( \mathcal{S}_T \) is then a Choquet simplex, that is, every stationary measure \( P \in \mathcal{S}_T \) is a mixture of extreme points in \( \mathcal{E}_T \), in a unique way. In other words, every stationary measure \( P \) on \((\Omega, \mathcal{F})\) then admits a unique decomposition into stationary ergodic modes, say \( P = \int_{\mathcal{E}_T} \varepsilon \cdot \pi(d\varepsilon) \) for some probability measure \( \pi \) on \((\mathcal{E}_T, \Sigma_T)\). Moreover \((\mathcal{E}_T, \Sigma_T)\) itself is then isomorphic with a universally measurable subset of a Polish space, so that every probability measure on \((\mathcal{E}_T, \Sigma_T)\) is perfect.

We now apply these results to log-optimum investment in a stationary stock market. Let \( T \) be a measure preserving transformation of a perfect probability space \((\Omega, \mathcal{F}, \mathcal{P})\). We also assume that \((\Omega, \mathcal{F})\) is isomorphic with a universally measurable subset of a Polish space so that the \( T \)-invariant measure \( \mathcal{P} \) admits an ergodic decomposition, say \( \mathcal{P} = \int_{\mathcal{E}_T} \varepsilon \cdot \pi(d\varepsilon) \) for some probability measure \( \pi \) on \((\mathcal{E}_T, \Sigma_T)\). We know that capital will grow exponentially in the long run with limiting rate \( \mathcal{E}\{\mathcal{W}^*_\infty|I\} \), which can now be interpreted as the maximum growth rate \( \mathcal{W}^*_\infty(\varepsilon) \) of the random stationary ergodic mode \( \varepsilon = \varepsilon(\omega) \):

\[
\frac{1}{n} \log \mathcal{S}^*_n = \frac{1}{n} \sum_{0 \leq t < n} \log \rho(b^*_t, X_t) - \mathcal{E}\{\mathcal{W}^*_\infty|I\} = \mathcal{W}^*_\infty(\varepsilon) \quad \text{a.s.}
\]

The average growth rate \( \mathcal{W}^*_\infty \) is obtained by averaging over ergodic modes \( \varepsilon \):

\[
\mathcal{W}^*_\infty = \int_{\mathcal{E}_T} \mathcal{W}^*_\infty(\varepsilon) \cdot \pi(d\varepsilon).
\]

To prove these assertions it suffices to observe that the sandwich trick reduces the proof of the A.E.P. to applications of the ergodic theorem, which now becomes

\[
\frac{1}{n} \sum_{0 \leq t < n} g \circ T^t \rightarrow \mathcal{E}\{g|I\} = E_\varepsilon\{g\} \quad \text{a.s.} \quad \text{and in } L^1 \text{ if } \mathcal{E}\{g\} \text{ is finite}
\]

for any random variable \( g : (\Omega, \mathcal{F}) \rightarrow (\mathcal{R}, \mathcal{B}_\mathcal{R}) \) with a well-defined expectation.
Asymptotically mean stationary stock market processes

A dynamical system \((\Omega, \mathcal{F}, \mathbf{P}, T)\) is called asymptotically mean stationary (a.m.s.) if the Cesàro averages of the measures \(\mathbf{P} \circ T^t\) converge, that is, if

\[
\bar{\mathbf{P}}(F) = \lim_{n \to \infty} \frac{1}{n} \sum_{0 \leq t < n} \mathbf{P}(T^{-t}F) \quad \text{exists for all } F \in \mathcal{F}.
\]

The Vitali-Hahn-Saks theorem then implies that \(\bar{\mathbf{P}}\) is a probability measure. The measure \(\bar{\mathbf{P}}\) is \(T\)-invariant or stationary (i.e. \(\bar{\mathbf{P}}(F) = \bar{\mathbf{P}}(T^{-1}F)\) for \(F \in \mathcal{F}\)), and is called the stationary mean of \(\mathbf{P}\). Gray and Kiefer [80] show that a probability measure \(\mathbf{P}\) is a.m.s. with respect to the transformation \(T\) iff there exists a \(T\)-invariant probability measure \(\bar{\mathbf{P}}\) on \((\Omega, \mathcal{F})\) satisfying one of the following properties:

- \(\bar{\mathbf{P}}\) asymptotically dominates \(\mathbf{P}\), in the sense that

  if \(F \in \mathcal{F}\) and \(\bar{\mathbf{P}}(F) = 0\), then \(\lim_{n \to \infty} \mathbf{P}(T^{-n}F) = 0\).

- \(\bar{\mathbf{P}}\) dominates \(\mathbf{P}\) on the \(\sigma\)-field \(I\) of almost surely \(T\)-invariant events:

  \[
  \bar{\mathbf{P}}|_I \gg \mathbf{P}|_I \quad \text{where} \quad I = \{F \in \mathcal{F} : T^{-1}F = F\}.
  \]

- \(\bar{\mathbf{P}}\) dominates \(\mathbf{P}\) on the tail \(\sigma\)-field \(\mathcal{F}_{\infty} = \bigcap_{0 \leq t < \infty} T^{-t}\mathcal{F}:

  \[
  \bar{\mathbf{P}}|_{\mathcal{F}_{\infty}} \gg \mathbf{P}|_{\mathcal{F}_{\infty}}.
  \]

If \(\mathbf{P}\) is a.m.s. then its stationary mean \(\bar{\mathbf{P}}\) is a stationary measure satisfying these three properties. Moreover \(\mathbf{P}\) and \(\bar{\mathbf{P}}\) have the same restrictions to the \(\sigma\)-field \(I\) of \(T\)-invariant events so that \(\mathbf{P}(F) = \bar{\mathbf{P}}(F)\) for all \(F \in I\) and \(\mathbb{E}\{g\} = \bar{\mathbb{E}}\{g\}\) for all \(T\)-invariant random variables \(g = g \circ T : (\Omega, \mathcal{F}) \to (\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})\). We use \(\bar{\mathbb{E}}\{\cdot\}\) to denote expectations with respect to the stationary mean \(\bar{\mathbf{P}}\).

The class of asymptotically mean stationary systems is the widest class of dynamical systems for which ergodic theory applies. A probability measure \(\mathbf{P}\) is a.m.s. with stationary mean \(\bar{\mathbf{P}}\) iff the following strong law of large numbers holds for all random variables \(g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})\) such that the expectation \(\bar{\mathbb{E}}\{g\}\) is well-defined:

\[
\frac{1}{n} \sum_{0 \leq t < n} g \circ T^t \to \bar{\mathbb{E}}\{g|I\} = \mathbb{E}\{g|I\} \quad \bar{\mathbf{P}}\text{-a.s. and } \mathbf{P}\text{-a.s.}
\]
In fact $\mathbf{P}$ is a.m.s. with stationary mean $\bar{\mathbf{P}}$ iff this strong law holds for all bounded measurable functions $g$ iff relative frequencies converge, that is, iff such almost sure convergence of sample averages holds for all indicator functions $g = 1_F$ of measurable subsets $F \in \mathcal{F}$. (See Gray [83].)

**Theorem 8.2** The asymptotic equipartition property for log-optimum investment continues to hold if the dynamical system $(\Omega, \mathcal{F}, \mathbf{P}, T)$ is a.m.s. (with $T$ not necessarily invertible).

**Proof**:

For concreteness let us first consider the special case where $(\Omega, \mathcal{F})$ is of the form $\otimes_{0 \leq t < \infty}(\mathcal{X}, \mathcal{B}_\mathcal{X})$ with $\mathcal{X}$ Polish, $T$ is the left shift, and $\mathcal{F}_t = \sigma(X_0, \ldots, X_{t-1})$ for all $0 \leq t < \infty$. The stationary mean $\bar{\mathbf{P}}$ may be viewed as a measure on $\otimes_{-\infty < t < \infty}(\mathcal{X}, \mathcal{B}_\mathcal{X})$, and $\bar{\mathbf{P}}$ is stationary with respect to the shift transformation

$$\bar{T}(\ldots, x_{-1}, x_0, x_1, \ldots) = (\ldots, x_0, x_1, x_2, \ldots)$$

which is invertible on this two-sided sequence space. Let $b^*_t$ denote a log-optimum portfolio selection for period $t$ based on $\mathcal{F}_t = \sigma(X_0, \ldots, X_{t-1})$ under the true measure $\mathbf{P}$, and $\tilde{b}_k^*$ a log-optimum portfolio for period 0 given $\mathcal{F}_k = \sigma(X_{-k}, \ldots, X_{-1})$ under the stationary mean $\bar{\mathbf{P}}$. The principle of asymptotic optimality gives

$$\lim_{n \to \infty} \sup_{k} \frac{1}{n} \log \left( \frac{S_k^*(k)}{S_k^*(\infty)} \right) \leq 0 \quad \mathbf{P}\text{-a.s.}$$

and

$$\lim_{n \to \infty} \sup_{k} \frac{1}{n} \log \left( \frac{S_k^*(k)}{S_k^*(\infty)} \right) \leq 0 \quad \bar{\mathbf{P}}\text{-a.s.}$$

where $S_k^*(k) = \prod_{0 \leq t < n} \rho(b^*_t, X_t)$, $S_k^*(\infty) = \prod_{0 \leq t < n} \rho(\tilde{b}_{k,t}^* \circ \bar{T}^t, X_t)$, and $S_k^*(\infty) = \prod_{0 \leq t < n} \rho(\tilde{b}_{\infty,t}^* \circ \bar{T}^t, X_t)$. The ergodic theorem for a.m.s. measures yields

$$\frac{1}{n} \log S_k^*(k) \to \mathbb{E}\{\log \rho(\tilde{b}_k^*, X_0)|I\} \quad \bar{\mathbf{P}}\text{-a.s. \ and \ P}\text{-a.s.}$$

and

$$\frac{1}{n} \log S_k^*(\infty) \to \mathbb{E}\{\log \rho(\tilde{b}_{\infty}^*, X_0)|I\} \quad \bar{\mathbf{P}}\text{-a.s.}$$

Combining these results we obtain

$$\mathbb{E}\{\log \rho(\tilde{b}_k^*, X_0)|I\} \leq \lim_{n \to \infty} \frac{1}{n} \log S_k^* \quad \mathbf{P}\text{-a.s.}$$

and

$$\lim_{n \to \infty} \sup_{k} \frac{1}{n} \log S_k^* \leq \mathbb{E}\{\log \rho(\tilde{b}_{\infty}^*, X_0)|I\} \quad \bar{\mathbf{P}}\text{-a.s.}$$
The last inequality also holds \( \mathbb{P} \)-a.s. since both sides are invariant random variables. Now

\[
\mathbb{E}\{\log \rho(\delta^*_k, X_0) | I\} / \mathbb{E}\{\log \rho(\delta^*_\infty, X_0) | I\} \ \mathbb{P}\text{-a.s.}
\]

and hence \( \mathbb{P} \)-a.s. since \( \mathbb{E}\{\log \rho(\delta^*_k, X_0) | I\} \) is invariant for all \( 0 \leq k \leq \infty \). We may conclude that

\[
\frac{1}{n} \log S^*_n \to \mathbb{E}\{\log \rho(\delta^*_\infty, X_0) | I\} \ \mathbb{P}\text{-a.s.}
\]

The generalization to arbitrary dynamical systems \((\Omega, \mathcal{F}, \mathbb{P}, T)\) such that \( \mathbb{P} \) is perfect and a.m.s. with respect to \( T \) but \( T \) is not necessarily invertible is straightforward, except for some technical details. Let \( \hat{T} \) denote the left shift on the one-sided sequence space \((\hat{\Omega}, \hat{\mathcal{F}}) = \bigotimes_{0 \leq t < \infty}(\Omega, \mathcal{F})\):

\[
\hat{T}(\omega_0, \omega_1, \ldots, \omega_{n-1}, \ldots) = (\omega_1, \omega_2, \ldots, \omega_n, \ldots).
\]

The measure \( \mathbb{P} \) induces a measure \( \hat{\mathbb{P}} \) on \((\hat{\Omega}, \hat{\mathcal{F}})\), defined by its behavior on cylinder sets as follows:

\[
\hat{\mathbb{P}}\{\omega_0 \in F_0, \ldots, \omega_K \in F_K\} = \mathbb{P}\{\omega_k = T^k \omega_0 \text{ and } \omega_k \in F_k \text{ for } 0 \leq k \leq K\}
\]

for all \( 0 \leq K < \infty \) and \( F_0, \ldots, F_K \in \mathcal{F} \). Notice that \((\Omega, \mathcal{F})\) is embedded as a measurable subspace in \((\hat{\Omega}, \hat{\mathcal{F}})\) if we identify \( \omega \in \Omega \) with \( \hat{\omega} = (\omega_t)_{0 \leq t < \infty} \in \hat{\Omega} \), where \( \omega_t = T^t \omega \) and trivially extend \( \mathcal{F} \) to a sub-\(\sigma\)-field of \( \hat{\mathcal{F}} \). The advantage of extending \((\Omega, \mathcal{F})\) to \((\hat{\Omega}, \hat{\mathcal{F}})\) is that \( \hat{T} \) is now a left shift on a one-sided sequence space.

Kolmogorov’s extension theorem for consistent families of finite-dimensional marginal distributions holds if all the marginals are perfect, and the unique extension measure is then also perfect (see Ryll-Nardzewski [53] and Musial[80]). Thus \( \hat{\mathbb{P}} \) is perfect and it is easily verified that \( \hat{\mathbb{P}} \) is a.m.s. with respect to the shift transformation \( \hat{T} \) on \((\hat{\Omega}, \hat{\mathcal{F}})\). Moreover its stationary mean \( \hat{\mathbb{P}} \) can be viewed as a perfect measure on the two-sided sequence space \((\hat{\Omega}, \hat{\mathcal{F}}) = \bigotimes_{-\infty < t < \infty}(\Omega, \mathcal{F})\), when \( \hat{\mathbb{P}} \) is defined on cylinder sets as follows:

\[
\hat{\mathbb{P}}\{\omega_N \in F_0, \ldots, \omega_{N+K} \in F_K\} = \\
\lim_{n \to \infty} \frac{1}{n} \sum_{0 \leq t < n} \mathbb{P}\{\omega_{t+k} \in F_k \text{ and } \omega_{t+k} = T^k \omega_t \text{ for } 0 \leq k \leq K\}
\]

for all \(-\infty < N < \infty, 0 \leq K < \infty, \) and \( F_0, \ldots, F_K \in \mathcal{F} \). Notice that \( \hat{\mathbb{P}} \) is stationary with respect to the shift transformation

\[
\hat{T}(\ldots, \omega_{-1}, \omega_0, \omega_1 \ldots) = (\ldots, \omega_0, \omega_1, \omega_2, \ldots)
\]
and $\mathcal{T}$ is invertible on $(\mathcal{O}, \mathcal{F})$.

Now let $\{\mathcal{F}_t\}_{t \leq t < \infty}$ be a sequence of sub-$\sigma$-fields of $\mathcal{F}$. For $0 \leq k < \infty$ let $\mathcal{F}_k$ denote the sub-$\sigma$-field of $\mathcal{F}$ defined by the condition that a generic random variable $g = g(\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ on $(\mathcal{O}, \mathcal{F})$ be $\mathcal{F}_k$-measurable iff $g \circ \mathcal{T}^k$ is $\mathbb{P}$-a.s. equal to $g_k(\omega_0, T\omega_0, \ldots)$ for some random variable $g_k = g_k(\omega_0, \omega_1, \ldots)$ such that $\omega_0 \mapsto g_k(\omega_0, T\omega_0, \ldots)$ is $\mathcal{F}_k$-measurable on $(\mathcal{O}, \mathcal{F})$. We assume that $\mathcal{F}_k$ increases monotonically towards a limiting $\sigma$-field $\mathcal{F}_\infty$:

$$\mathcal{F}_k \searrow \mathcal{F}_\infty = \bigvee_{0 \leq k < \infty} \mathcal{F}_k = \sigma(\bigcup_{0 \leq k < \infty} \mathcal{F}_k).$$

Let $b_t^*$ denote a log-optimum portfolio choice for period $t$ based on $\mathcal{F}_t$ under the true distribution $\mathbb{P}$ and $\bar{b}_k^*$ a log-optimum portfolio for period $0$ based on $\mathcal{F}_k$, under the stationary mean $\bar{\mathbb{P}}$. The proof of the asymptotic equipartition property

$$\frac{1}{n} \log \mathcal{L}_n^* \rightarrow \mathbb{E}\{\log \rho(\bar{b}_\infty^*, X_0) | I\} \quad \mathbb{P}\text{-a.s.}$$

now proceeds in exactly the same fashion as for the special case that we considered as a preliminary.

QED.

If the stationary mean $\bar{\mathbb{P}}$ admits an ergodic decomposition, say

$$\bar{\mathbb{P}} = \int_{\mathcal{E}_T} \epsilon \cdot \bar{\pi}(d\epsilon)$$

for some measure $\bar{\pi}$ on $(\mathcal{E}_T, \Sigma_T)$, then we may interpret the limiting random variable $\mathbb{E}\{\log \rho(\bar{b}_\infty^*, X_0) | I\}$ as the expectation $\bar{W}_\infty^* \epsilon \in \mathcal{E}_T$ which was randomly selected in $\mathcal{E}_T$ according to the distribution $\bar{\pi}$. The process will evolve towards this stationary ergodic mode $\epsilon = \epsilon(\omega)$ and will be indistinguishable from it in the limit. The average growth rate is given by

$$\bar{W}_\infty^* = \int_{\mathcal{E}_T} \bar{W}_\infty^* \epsilon \cdot \bar{\pi}(d\epsilon).$$
APPENDIX.
LOWER SEMicontinuous FUNCTIONS

Let $f : \mathcal{Z} \to \mathcal{R}$ be an extended real valued function defined on a separable metric space $\mathcal{Z}$. Then $f$ is called lower semicontinuous (l.s.c.) if any one of the following equivalent conditions holds.

- $\liminf_{n \to \infty} f(z_n) \geq f(z)$ whenever $z_n \to z$ in $\mathcal{Z}$.

- $f(z) = \lim_{N \to \infty} \inf_{z' \in N} f(z')$ where the limit is taken along the filter of shrinking neighborhoods of $z$.

- $\{f \leq r\}$ is a closed subset of $\mathcal{Z}$, for all $r \in \mathcal{R}$.

- $\{f > r\}$ is an open subset of $\mathcal{Z}$, for all $r \in \mathcal{R}$.

- The epigraph $\mathcal{E}(f)$ of $f$ is a closed subset of $\mathcal{Z} \times \mathcal{R}$. The epigraph $\mathcal{E}(f)$ is the subset $\mathcal{E}(f) = \{(z,r) \in \mathcal{Z} \times \mathcal{R} : r \geq f(z)\}$ of $\mathcal{Z} \times \mathcal{R}$.

A function $f : \mathcal{Z} \to \mathcal{R}$ is called upper semicontinuous (u.s.c.) iff $-f$ is l.s.c. The following properties of lower semicontinuous functions are well-known.

- The supremum of any family of l.s.c. functions on $\mathcal{Z}$ is l.s.c.

- An extended-real-valued function $f : \mathcal{Z} \to \mathcal{R}_+$ is l.s.c. and bounded below iff there exists an increasing sequence of bounded continuous real-valued functions $f_n : \mathcal{Z} \to \mathcal{R}$ such that $f_n \uparrow f$.

- The indicator function $1_G$ of a subset $G \subseteq \mathcal{Z}$ is l.s.c. (resp. u.s.c.) iff $G$ is open (resp. closed).

- A function $f : \mathcal{Z} \to \mathcal{R}$ is continuous iff $f$ is both l.s.c. and u.s.c.

- If $f : \mathcal{Z} \to \mathcal{R}$ is l.s.c., then $f$ achieves its infimum on every nonempty compact subset $K \neq \emptyset$ of $\mathcal{Z}$, i.e., there exists $z_K \in K$ such that

$$\inf_{z \in K} f(z) = \min_{z \in K} f(z) = f(z_K).$$
REFERENCES


Gray, R.M., Lecture notes on information and ergodic theory. Stanford University, Spring 1983.


Ramachandran, D., Perfect measures, Part I (Basic Theory) and Part II (Special Topics). ISI Lecture Notes 5 and 7, Macmillan Company of India, 1979.


