A BOUND ON THE MONETARY VALUE OF INFORMATION

by

Thomas M. Cover
Stanford University

TECHNICAL REPORT NO. 52
JUNE 1984

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

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A BOUND ON THE MONETARY VALUE
OF INFORMATION

by

Thomas M. Cover

TECHNICAL REPORT NO. 52
June 1984

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

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A Bound on the Monetary Value of Information†

Thomas M. Cover
Stanford University

ABSTRACT

In the general stock market setup it will be shown that each bit of information at most doubles the resulting capital. This information bound on capital growth is actually attained for certain probability distributions on the market.

The proof is based on devising a money-ratio test for establishing whether the side information is or is not independent of the stock market. The increase \( \Delta \) in the exponent of the growth rate of capital due to side information is also the error exponent of the associated money-ratio test. The probability of error exponent \( \Delta \) for this test is necessarily no greater than the mutual information \( I \), the error exponent for the best test as proved in Stein's lemma. This yields the desired inequality.

†This work was partially supported by National Science Foundation Grant ECS82-11588 and Joint Services Electronics Program DAAG29-81-K-0057.
1. Introduction.

Let \( \mathbf{X} \geq 0, \mathbf{X} \in \mathbb{R}^m \) denote a random stock market vector, with the interpretation that \( X_i \) is the worth of a one unit investment in stock \( i \). Let \( B = \{ \mathbf{b} \in \mathbb{R}^m : b_i \geq 0, \sum_{i=1}^m b_i = 1 \} \) be the set of all portfolios \( \mathbf{b} \). Here \( b_i \) is the proportion of capital invested in the \( i \)-th stock. The resulting capital is

\[
S = \sum_{i=1}^m b_i X_i = \mathbf{b}'\mathbf{X}.
\]

This is the capital resulting from a one unit investment allocated to the \( m \) stocks according to the portfolio \( \mathbf{b} \).

2. Doubling Rate.

Now let \( F(\mathbf{x}) \) be the probability distribution function of the stock vector \( \mathbf{X} \). We define the doubling rate \( W(\mathbf{X}) \) for the market by

\[
W(\mathbf{X}) = \max_{\mathbf{b} \in B} \int \log \mathbf{b}'\mathbf{x} \ dF(\mathbf{x}).
\]

The units for \( W \) are “doubles per investment”. All logs in this paper are to the base 2. Let \( \mathbf{b}^* \) denote a portfolio achieving \( W(\mathbf{X}) \). It should be noted that \( W(\mathbf{X}) \) is a real number, a functional of \( F \); the apparent dependence of \( W \) on \( \mathbf{X} \) is a notational convenience.

If the current capital is reallocated according to \( \mathbf{b}^* \) in repeated independent investments against stock vectors \( \mathbf{X}_1, \mathbf{X}_2, \ldots \) independent identically distributed (i.i.d.) according to \( F(\mathbf{x}) \), then the capital \( S_n^* \) at time \( n \) is given by

\[
S_n^* = \prod_{i=1}^n \mathbf{b}^* \mathbf{X}_i.
\]

The strong law of large numbers for products yields
\[(S_n^*)^{1/n} = 2^n \sum_{i=1}^n \log b^"X_i, \rightarrow 2^W, \quad (2.3)\]

with probability one. Moreover, no other portfolio achieves a higher exponent (Breiman [1]). The reason for calling \( W \) the doubling rate is that \( n = 1/W \) investments are required to double the capital -- the capital grows like \( S_n \rightarrow 2^{nW} \).

Now suppose side information \( Y \) is available. Here \( Y \) could be the weather, world events, or the behavior of a correlated market. Again we define the maximum expected log return, but this time we allow the portfolio \( b \) to depend on \( Y \). Let the doubling rate for side information be

\[ W(X \mid Y) = \max_{b(y)} \int \int \log b^t(y)X\ dF(x,y), \quad (2.4) \]

and let \( b^*(y) \) be the portfolio achieving \( W(X \mid Y) \). It can be shown that \( b^*(y) \) maximizes the conditional expected log return \( E\{ \log b^tX \mid Y = y \} \).

In repeated investments against \( X_1, X_2, \ldots, X_n \) where \( (X_i, Y_i) \) are i.i.d. \( \sim F(x,y) \), and \( b^*(Y_i) \) is the portfolio used at investment time \( i \) given side information \( Y_i \), we have resulting capital

\[ S_n^{**} = \prod_{i=1}^n \ b^*(Y_i)X_i, \quad (2.5) \]

with asymptotic behavior

\[ (S_n^{**})^{1/n} \rightarrow 2^{W(X \mid Y)}, \quad (2.6) \]

with probability one. It follows that the ratio of capital with side information to that without side information has limit

\[ \left( \frac{S_n^{**}}{S_n^*} \right)^{1/n} \rightarrow 2^{W(X \mid Y) - W(X)}, \quad (2.7) \]

with probability one.
Let the difference in maximum expected log return with and without $Y$ be

$$\Delta = W(X | Y) - W(X).$$  \hspace{1cm} (2.8)

Thus $\Delta$ is the increment in doubling rate due to the side information $Y$. It is this difference that we are trying to bound. As an example, if $\Delta = 1$ then the information $Y$ yields an additional doubling of the capital each investment period. Finally, we observe from (2.1) and (2.4) that $\Delta \geq 0$. Information never hurts.

3. **Mutual Information.**

So much for the doubling rate $W$. Now we define mutual information. Suppose that $(X, Y)$ is drawn according to a joint probability density $f(x, y)$. Then the mutual information $I(X; Y)$ is defined by

$$I(X; Y) = \int \int f(x, y) \log \frac{f(x, y)}{f(x)f(y)} \, dx \, dy = E \log \frac{f(X, Y)}{f(X)f(Y)}. \hspace{1cm} (3.1)$$

Thus $I(X; Y)$ is the expected log likelihood ratio under the hypothesis $H_0$ for the two-hypothesis test $H_0 : f = f(x, y)$ vs. $H_1 : f = f(x)f(y)$. (If $F(x, y)$ has no density, simply replace $\frac{f(x, y)}{f(x)f(y)}$ by the Radon Nikodym derivative of $F(x, y)$ with respect to $F(x)f(y)$ to define $I$ in general.)

Of the many alternative expressions for $I$, the most evocative is the identity

$$I(X; Y) = H(X) - H(X | Y), \hspace{1cm} (3.2)$$

where $H(X)$ is the entropy of $X$ and $H(X | Y)$ is the conditional entropy. Thus $I$ is the amount the entropy of $X$ is decreased by knowledge of $Y$. One can compare (3.2) with (2.8) to see why a relationship between $\Delta$ and $I$ might be expected.
The most well known interpretation of mutual information is in Shannon's channel capacity theorem, in which it is proved that \( I \) bits can be communicated by \( Y \) to a receiver \( X \). \( I \) is called the channel capacity (in bits per unit time) for the communication channel \( f(x, y) \).

There is an interpretation of \( I(X; Y) \) in terms of efficient descriptions. (See Gallager [2].) Since \( H(X) \) bits are required to describe the value of the random variable \( X \) (if \( X \) is discrete), and since \( H(X | Y) \) bits are required to describe \( X \) given knowledge of \( Y \), the decrement in the expected description length of \( X \) is given by \( H(X) - H(X | Y) = I(X; Y) \).

In summary, the mutual information \( I(X; Y) \) is 1) the decrease in the entropy of \( X \) when \( Y \) is made available, 2) the number of bits by which the expected description length of \( X \) is reduced by knowledge of \( Y \), 3) the rate in bits at which \( Y \) can communicate with \( X \) by appropriate choice of \( Y \), 4) the error exponent for the hypothesis test \((X, Y)\) independent vs. \((X', Y)\) dependent.

4. Preliminaries to the Theorem.

We shall need Stein's lemma, as well as an analysis of the log optimal portfolio, in order to prove the theorem in the next section.

First, Stein's lemma. Let \( f \) and \( g \) be two probability densities with respect to a common dominating measure \( \nu \). For \( A \subseteq \mathbb{R}^n \), let

\[
P_f^n(A) = \int_A \prod_{i=1}^n f(x_i) d\nu
\]

(4.1)

and

\[
P_g^n(A) = \int_A \prod_{i=1}^n g(x_i) d\nu
\]

(4.2)

Thus \( P_f^n(A) \) is the probability that \((X_1, X_2, \ldots, X_n)\) is in \( A \) given that
$X_i$, $i = 1, 2, ..., n$ are drawn i.i.d. according to $f(x)$. Let

$$\beta_n(\epsilon) = \min_{A \subseteq \mathbb{R}^n} \frac{1}{\epsilon} \log P^n_x(A)$$

(4.3)

The following lemma was proved by Stein in 1952 (unpublished). A proof appears in Chernoff (1956). (See also Csiszár and Körner (1981)).

**Lemma (Stein):** For any $0 < \epsilon < 1$,

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n(\epsilon) = D(f \parallel g)$$

(4.4)

where

$$D(f \parallel g) = \int f \ln \frac{f}{g} d\nu$$

(4.5)

The interpretation in the two-hypothesis testing problem is that there is no set $A$ having large probability under hypothesis $X_i$ i.i.d. $\sim g$ and exponentially small probability under hypothesis $X_i$ i.i.d. $\sim f$ with exponent better than $D(f \parallel g)$. Thus if the probability of error of the first kind is held $\leq \epsilon$, the probability of error of the second kind cannot tend to zero faster than $e^{-nD}$, in the sense of (4.4).

Incidentally, a set $A$ achieving this bound is

$$A = \{ x \in \mathbb{R}^n : \frac{1}{n} \ln \frac{f(x)}{g(x)} \geq 1 + \epsilon' \}.$$  The fact that any other set $B$ satisfying $P^n_B(B) \geq 1 - \epsilon$ must significantly intersect $A$ leads to a proof of the lemma.

We now consider the behavior of

$$E \frac{S}{S^*} = E \frac{b^T X}{b^* X}$$

(4.6)

where $b^*$ maximizes $E \ln b^T X$, over all portfolios, and $b$ is any other portfolio. From the Kuhn-Tucker conditions, it is necessary and sufficient that $b^*$
satisfy

\[ E \frac{X_i}{b^*X} = 1, \quad b_i^* > 0 \]

\[ \leq 1, \quad b_i^* = 0. \quad (4.7) \]

See Bell and Cover [5], Cover [6], or Finkelstein and Whitley [7].

We thus have, for any other portfolio \( b \), \( \sum b_i = 1, \quad b_i = 0 \),

\[ E \frac{S}{S^*} = E \frac{\sum b_i X_i}{b^*X} \]

\[ = \sum_{i=1}^{m} b_i E \frac{X_i}{b^*X} \]

\[ \leq \sum_{i=1}^{m} b_i = 1. \quad (4.8) \]

5. The Information Bound.

We ask how \( \Delta \) and \( I \) are related for the stock market. We have

\[ \Delta = E \log \frac{b^*(Y)X}{b^*X} \quad (5.1) \]

and

\[ I = E \log \frac{f(X,Y)}{f(X)f(Y)}, \quad (5.2) \]

where \( (X,Y) \sim F(x,y) \). The first involves money and depends on the values \( X \) takes on. The second involves information and depends on \( X \) and \( Y \) only through the density \( f(x,y) \).

The following theorem establishes that the increment \( \Delta \) in the doubling rate resulting from side information \( Y \) is less than or equal to the mutual information \( I \).
Theorem:

\[ 0 \leq \Delta \leq I(X;Y). \quad (5.3) \]

Proof: We do not offer an algebraic proof for this result, although a proof along the lines of the data processing theorem seems possible. Instead, we use the capital returns \( S^{**}_n \) and \( S^*_n \) defined in (2.2) and (2.5) to form a "money ratio" test for establishing the independence or dependence of the market \( X \) and the side information \( Y \), and show that the exponent of the probability of error of this test is \( \Delta \). By Stein's lemma, the best exponent in a hypothesis test is \( I \). Hence \( \Delta \leq I \).

Let \( (X_i, Y_i) \) be i.i.d. random pairs drawn according to the distribution \( G(x, y) \), where \( X \in \mathbb{R}^m \), \( X \geq 0 \), is the random stock vector and \( Y \) is the side information. Consider the two hypotheses

\[ H_0 : G(x, y) = F_{XY}(x, y) \]

and

\[ H_1 : G(x, y) = F_X(x)F_Y(y). \quad (5.4) \]

Thus under hypothesis \( H_1 \), \( X \) and \( Y \) are independent with the same marginals as the joint distribution \( F_{XY}(x, y) \). Let \( \alpha_n \) and \( \beta_n \) be the probabilities of error under hypotheses \( H_0 \) and \( H_1 \), respectively, for a given decision rule based on \( (X_1, Y_1), \ldots, (X_n, Y_n) \). If \( \alpha_n \leq \alpha < 1 \), for all \( n \) sufficiently large, then by Stein's lemma, it follows that

\[ \lim_{n \to \infty} \left( -\frac{1}{n} \log \beta_n \right) \leq D(F_{XY} \parallel F_X \times F_Y) \]

\[ = I(X;Y), \quad (5.5) \]

for any statistical decision rule based on sample size \( n \). (In fact, the likelihood ratio test achieves equality.)
We now devise a decision rule based on the money ratio. Let the portfolio \( b^* \) achieve \( W(X) \) and let \( b^*(y) \) achieve \( W(X | Y) \). Let

\[
S^*_n = \prod_{i=1}^{n} b^* X_i \tag{5.6}
\]

and

\[
S^{**}_n = \prod_{i=1}^{n} b^*(Y_i) X_i , \tag{5.7}
\]

be the resulting random capitals. Both \( S^*_n \) and \( S^{**}_n \) are functions of the data \((X_1, Y_1), \ldots, (X_n, Y_n)\), so a test can be based on these statistics. For \( \epsilon > 0 \), consider the decision rule

\[
\frac{S^{**}_n}{S^*_n} > 2^{n(\Delta - \epsilon)} , \quad \text{Decide } H_0
\]

\[
< 2^{n(\Delta - \epsilon)} , \quad \text{Decide } H_1 . \tag{5.8}
\]

We first consider \( \alpha_n \), the probability of error of this decision rule under \( H_0 \). After taking logarithms in (5.8), we have

\[
\frac{1}{n} \log \left( \frac{S^{**}_n}{S^*_n} \right) = \frac{1}{n} \sum_{i=1}^{n} \log b^*(Y_i) X_i - \frac{1}{n} \sum_{i=1}^{n} \log b^* X_i ;
\]

\[
\to W(X | Y) - W(X) = \Delta , \tag{5.9}
\]

with probability one, under hypothesis \( H_0 \), by the strong law of large numbers. Thus

\[
\alpha_n = P \left\{ \frac{1}{n} \log \frac{S^{**}_n}{S^*_n} < \Delta - \epsilon \mid H_0 \right\} \to 0 , \tag{5.10}
\]

for all \( \epsilon > 0 \), and \( \alpha_n \) satisfies the conditions for Stein's lemma.
Now $\beta_n$, the probability of error under $H_1$, is given by

$$
\beta_n = P \left\{ \frac{S_n^{**}}{S_n^*} > 2^{(\Delta - \epsilon)} \mid H_1 \right\} \leq \frac{E\{S_n^{**} / S_n^* \mid H_1\}}{2^{(\Delta - \epsilon)}}, \tag{5.11}
$$

by Markov's inequality and the nonnegativity of $S_n^{**} / S_n^*$. We now show that $E\{S_n^{**} / S_n^* \mid H_1\} \leq 1$. We have

$$
E\{S_n^{**} / S_n^* \mid H_1\} = E \left\{ \frac{\Pi_b^{*}(Y_i)X_i}{\Pi_b^{*}X_i} \mid H_1 \right\}
$$

$$
\overset{(a)}{=} \prod_{i=1}^{n} E \left\{ \frac{b^{*}(Y_i)X_i}{b^{*}X_i} \mid H_1 \right\}
$$

$$
\overset{(b)}{=} \prod_{i=1}^{n} E \left\{ b^{*}(Y_i) \mid H_1 \right\} E \left\{ \frac{X_i}{b^{*}X_i} \mid H_1 \right\}
$$

$$
\overset{(c)}{=} \prod_{i=1}^{n} E \left\{ b^{*}(Y_i) \right\}
$$

$$
= 1, \tag{5.12}
$$

where (a) follows from the independence of the pairs $(X_i, Y_i)$, $i = 1, 2, \ldots, n$, (b) follows from the independence of $X_i$ and $Y_i$ under hypothesis $H_1$, and (c) follows from the Kuhn-Tucker conditions in (4.7). Consequently, from (5.11) and (5.12),
\[
\beta_n \leq 2^{-n(\Delta - \epsilon)}, \quad \text{for all } n.
\] (5.13)

Thus the money ratio test has error exponent \( \Delta - \epsilon \). We have shown \( \alpha_n \to 0 \) in (5.10). By Stein's lemma, the best exponent for \( \beta_n \) for any statistical test is \( I(X; Y) \). Thus
\[
\Delta - \epsilon \leq I(X; Y).
\] (5.14)

Since \( \epsilon > 0 \) is arbitrary, the theorem is proved.


We first give an example due to Kelly (1956) in which \( \Delta = I \). Here the stock market is a horse-race, which, in the setup of (1.1), consists of a probability mass function \( P \{ X = O_i \mathbf{e}_i \} = p_i, \quad i = 1, 2, \ldots, m \), where \( \mathbf{e}_i \) is a unit vector with a 1 in the \( i \)-th place and 0's elsewhere, \( O_i \) equals the win odds (\( O_i \) for 1), and \( p_i \) is the probability that the \( i \)-th horse wins the race. Then
\[
W(X) = \max_{\mathbf{b}} E \log \mathbf{b}'X
\]
\[
= \max_{\mathbf{b}} \sum_{i=1}^m p_i \log b_i O_i
\]
\[
= \sum p_i \log O_i - H(X),
\] (6.1)

where \( H(X) = -\sum_{i=1}^m p_i \log p_i \). Also, \( \mathbf{b}^* = \mathbf{p} \), i.e., the optimal portfolio is to bet in proportion to the win probabilities, regardless of the odds.

For side information \( Y \), where \( (X, Y) \) has a given distribution, a similar calculation yields
\[
W(X \mid Y) = \sum p_i \log O_i - H(X \mid Y),
\] (6.2)
and
\[ b_i^*(y) = P(x = e_i \mid y), \quad i = 1, 2, \ldots, m. \]

Here the optimal portfolio is to bet in proportion to the conditional probabilities, given \( Y \). Subtracting (6.1) from (6.2), we have
\[ \Delta = W(X \mid Y) - W(X) = H(X) - H(X \mid Y) = I(X; Y). \quad (6.3) \]

Consequently, the information bound on \( \Delta \) is tight.

Of course, it sometimes happens that the information \( Y \) about the market is useless for investment purposes. The next example has \( \Delta = 0, \ I = 1 \). Let \( X = (1, 1/2) \) with probability \( 1/2 \), and \( X = (1, 3/4) \), with probability \( 1/2 \). Let \( Y = X \). An investment in the first stock always returns one's money, but an investment in the second stock may cut the investment capital to either \( 1/2 \) or \( 3/4 \) depending on the outcome \( X \). It would be foolish to invest in the second stock, since the first stock dominates its performance. Thus \( b^* = b^*(y) = (1, 0), \) for all \( y \), and \( \Delta = 0 \). On the other hand, since the outcomes of \( X \) are equally likely, and \( Y = X \), we see
\[ I(X; Y) = I(X; X) = H(X) - H(X \mid X) \]
\[ = H(X) = 1 \text{ bit}. \]

Thus a bit of information is available, but \( \Delta = 0 \) and the rate of return is not improved.

7. Conclusions.

We offer one final interpretation. Recall that \( H(X) - H(X \mid Y) = I(X; Y) \) is the decrement in the expected description length of \( X \) due to the side information \( Y \). Hence the inequality \( \Delta \leq I \) has the interpretation that the increment in the doubling rate of the market \( X \) is less than the decrement in the description rate of \( X \).
Acknowledgment

I would like to thank R.O. Duda for a lunchtime remark that led to the statement of this theorem.
References


