A Bound on the Financial Value of Information

ANDREW R. BARRON, MEMBER, IEEE, AND
THOMAS M. COVER, FELLOW, IEEE

Abstract—It will be shown that each bit of information at most doubles the resulting wealth in the general stock market setup. This information bound on the growth of wealth is actually attained for certain probability distributions on the market investigated by Kelly. The bound will be shown to be a special case of the result that the increase in exponential growth of wealth achieved with true knowledge of the stock market distribution \( F \) over that achieved with incorrect knowledge \( G \) is bounded above by \( D(F \| G) \), the entropy of \( F \) relative to \( G \).

I. INTRODUCTION

Let \( X \geq 0, X \in \mathbb{R}^m \) denote a random stock market vector, with the interpretation that \( X_i \) is the ratio of the price of the \( i \)th stock at the end of an investment period to the price at the beginning. Let \( B = \{ b \in \mathbb{R}^m; b_i \geq 0, \sum_i b_i = 1 \} \), be the set of all portfolios \( b \), where \( b_i \) is the proportion of wealth invested in the \( i \)th stock. The resulting wealth is

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

This is the resulting wealth from a unit investment allocated to the \( m \) stocks according to the portfolio \( b \).

II. DOUBLING RATE

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

\[
W(X) = \max_{b \in B} \int b^T x dF(x).
\]

The units for \( W \) are “doubles per investment.” All logarithms in this correspondence are to the base 2. Let \( b^* = b^*(F) \) denote a portfolio achieving \( W(X) \). Note that \( W(X) \) is a real number, a functional of \( F \); the apparent dependence of \( W \) on \( X \) is for notational convenience.

Necessary and sufficient conditions for \( b \) to maximize \( E \log b^T X \) are

\[
\begin{align*}
E_{\frac{X_i}{b^T X}} &= 1, & & \text{for } b_i > 0 \\
E_{\frac{X_i}{b^T X}} &\leq 1, & & \text{for } b_i = 0.
\end{align*}
\]

These are the Kuhn–Tucker conditions characterizing \( b^*(F) \) (see Bell and Cover [3], Cover [4], and Finkelstein and Whitley [5]).

If current wealth is reallocated according to \( b^* \) in repeated independent investments against stock vectors \( X_1, X_2, \ldots \) independent identically distributed (i.i.d.) according to \( F(x) \), then the wealth \( S_n^* \) at time \( n \) is given by

\[
S_n^* = \prod_{i=1}^n b^* X_i.
\]

The strong law of large numbers for products yields

\[
(S_n^*)^{1/n} = 2^{W(X)} \rightarrow 2^W,
\]

with probability one. Moreover, no other portfolio achieves a higher exponent (Breiman [1]; Algoet and Cover [9]).

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

\[
W(X|Y) = \max_{b(y)} \int \log b^T y x dF(x, y),
\]

and let \( b^*(Y) = b^*(F_{X|Y}) \) be the portfolio achieving \( W(X|Y) \). It can be shown that \( b^*(Y) \) maximizes the conditional expected logarithm of the wealth \( E[\log b^T Y | Y = y] \).

In repeated investments against \( X_i, X_{i+1}, \ldots, X_n \) where \( (X_i, Y) \) are i.i.d. \( \sim F(x, y) \), and \( b^*(Y) \) is the portfolio used at investment time \( i \) given side information \( Y_i \), we have resulting wealth

\[
S_i^{**} = \prod_{i=1}^n b^* (Y) X_i
\]

with asymptotic behavior

\[
(S_n^{**})^{1/n} \rightarrow 2^{W(X|Y)}
\]

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

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

with probability one.

Let the difference between the maximum expected logarithm of wealth with \( Y \) and without \( Y \) be

\[
\Delta = W(X|Y) - W(X).
\]

Thus \( \Delta \) is the increment in doubling rate due to the side information \( Y \). It is this difference that we wish to bound. As an example, if \( \Delta = 1 \) then the information \( Y \) yields an additional doubling of the capital in each investment period. Finally, we observe from (6) and (2) that \( \Delta \geq 0 \). Information never hurts. Kelly [6] identified \( \Delta \) with the mutual information for a “horse-race” stock market, a result we will generalize here.

III. MUTUAL INFORMATION AND RELATIVE ENTROPY

The relative entropy (or Kullback–Leibler information number) of probability distributions \( F \) and \( G \) is

\[
D(F \| G) = \int \log (f/g) dF
\]

where \( f \) and \( g \) are the respective densities with respect to any dominating measure. (Note: \( D \) is infinite if \( g(x) \) is zero on a set of positive probability with respect to \( F \).)

The relative entropy may be interpreted as the error exponent for the hypothesis test \( F \) versus \( G \) (Stein’s lemma; see Chernoff [2]). Another interpretation of the relative entropy for a discrete random vector \( X \sim P \) is that \( D(P \| Q) \) is the expected increase in description length of the Shannon–Fano code based on the incorrect distribution \( Q \).

Let \( X, Y \) be two random variables with joint distribution \( P_{X,Y} \). The relative entropy between the conditional distribution \( P_{X|Y} \) is
and the marginal distribution $P_x$ is the mutual information

$$I(X; Y) = \int D(P_{X|Y} \parallel P_X) P_Y(dy) = D(P_{X|Y} \parallel P_X P_Y).$$

(12)

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

$$I(X; Y) = H(X) - H(X|Y)$$

(13)

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 (13) with (10) to see why a relationship between $\Delta$ and $I$ might be expected.

The mutual information $I$ can also be interpreted as the information rate achievable in communication over the communication channel $P(x, y)$. There is also an interpretation of $I(X; Y)$ in terms of efficient descriptions. 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 decrease in expected description length per bit, 4) the Bayes error exponent for the hypothesis test $(X; Y)$ independent versus $(X, Y)$ dependent.

IV. PORTFOLIOS BASED ON INCORRECT DISTRIBUTIONS

Suppose that it is believed that $X = G(x)$ when in fact $X = F(x)$. Thus the incorrect portfolio $b^*(G)$ is used instead of $b^*(F)$. The doubling rate associated with portfolio $b$ and distribution $F$ can be written

$$W(b, F) = \int \log b x dF(x)$$

(14)

with resulting growth of wealth

$$S_t \triangleq \sum_{t=0}^{\infty} b^* X_t$$

(15)

The decrement in exponent from using $b^*(G)$ is

$$\Delta W(F, G) = W(b^*(F), F) - W(b^*(G), F).$$

(16)

The following theorem is central to our results.

Theorem 1:

$$0 \leq \Delta W(F, G) \leq D(F\parallel G).$$

(17)

Proof: The first inequality $0 \leq \Delta$ follows by the optimality of $b^*(F)$ for the distribution $F$. The second inequality $\Delta \leq B$ is shown to be a consequence of the optimality of $b^*(G)$ for the distribution $G$. Let $F$ and $G$ have densities $f$ and $g$ with respect to some dominating measure. The result $\Delta \leq B$ is trivially true if $D(F\parallel G) = \infty$, so it is henceforth assumed that $D$ is finite (whence $F \ll G$).

Let

$$S_t^* = b^*(F) X_t \quad S_t = b^*(G) X_t$$

(18)

be the wealth factors corresponding to the optimal portfolios with respect to $F$ and $G$. From the Kuhn–Tucker conditions the wealth factor $S_t$ is strictly positive with probability one with respect to $G$ (and with respect to $F$ since $F \ll G$). It follows (again since $F \ll G$) that the set $A = \{x: S_t > 0, f(x) > 0, g(x) > 0\}$ has probability one with respect to $F$. Then

$$\Delta W(F, G) = \int_A \left( \log \frac{S_t^*}{S_t} \right) dF$$

$$= \int_A \left( \log \left( \frac{S_t^*}{S_t} \frac{g}{f} \right) \right) dF$$

$$= \int_A \left( \log \frac{S_t^*}{S_t} \frac{g}{f} \right) dF + D(F\parallel G)$$

$$\leq \int_A \log \frac{S_t^*}{S_t} dF + D(F\parallel G)$$

$$\leq D(F\parallel G)$$

(19)

where the first inequality follows from the concavity of the logarithm and the second from the Kuhn–Tucker conditions for the optimality of $b^*(G)$ for the distribution $G$.

We can improve Theorem 1 by normalizing $X$. Let $\tilde{F}$ denote the distribution of $X/\Sigma X_i$. We note that $E(\log b^* X/\tilde{b}^* X)$ depends on the distribution $F(x)$ only through the distribution of $X/\Sigma X_i X_i \sim \tilde{F}$.

Corollary:

$$\Delta W(F, G) \leq D(\tilde{F}\parallel G)$$

Remark: Another relationship between $W$ and $D$ is shown by Mori [13]. The doubling rate $W = W(b^*(F), F)$ is equal to the minimum of $D(F\parallel G)$ over all distributions $G$ for which $E_{X_i} X_i \leq 1$, for $i = 1, 2, \ldots, m$.

V. THE INFORMATION BOUND FOR SIDE INFORMATION

We now ask how $\Delta$ and $I$ are related for the stock market. We have

$$\Delta = E \log \frac{b^*(Y) X}{b^* X}$$

(20)

and

$$I = E \log \frac{f(X, Y)}{f(X)}$$

(21)

where $(X, Y) \sim F(x, y)$. The first involves wealth 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 2:

$$0 \leq \Delta \leq I(X; Y)$$

(22)

Proof: For any $y$, let $P_{X|Y}$ be the conditional distribution for $X$ given that $Y = y$ and let $\tilde{P}_X$ be the marginal distribution for $X$. Also let $\tilde{b}^* = b^*(P_{X|Y})$. Apply Theorem 1, with $P_{X|Y}$ and $P_X$ in place of $F$ and $G$, respectively, to obtain

$$0 \leq E \left[ \log \frac{b^* Y}{b^* X} \right] = Y = \tilde{D}(P_{X|Y} \parallel P_X).$$

(23)

Averaging with respect to the distribution of $Y$ yields

$$0 \leq \Delta \leq I$$

(24)

Remark: An alternative proof of this theorem, based on money ratio tests and Stein’s lemma, appears in [7].

VI. SEQUENTIAL PORTFOLIO ESTIMATION

Here we show that a good sequence of estimates of the true market distribution leads to asymptotically optimal growth rate of wealth. First we generalize Theorem 1 to handle the sequential setting.
Let $X_1, X_2, \cdots, X_n$ be a sequence of random stock vectors with joint probability distribution $P$. The log-optimal sequential strategy uses the portfolio $b^* = b^*(P_{X_1, X_2, \cdots, X_n})$, which maximizes the conditional expected value of $\log b^* X_i$ given that $X_i = x_1, \cdots, X_{i-1} = x_{i-1}$. Suppose that instead of $b^*$, we use portfolios $\tilde{b}_i = b^*(Q_{X_1, X_2, \cdots, X_{i-1}})$ which are optimal for an incorrect distribution $Q_i$ for the sequence $X_1, \cdots, X_n$.

Let $P_{X_1, X_2, \cdots, X_{i-1}}$ and $Q_{X_1, X_2, \cdots, X_{i-1}}$ be the regular conditional distributions associated with $P$ and $Q$, respectively. We compare the resulting wealth

$$\tilde{S}_n = \sum_{i=1}^{n} \tilde{b}_i X_i$$

with the wealth

$$S^*_n = \sum_{i=1}^{n} b^*_i X_i.$$  (26)

**Theorem 3:**

$$0 \leq E \log \frac{S^*_n}{\tilde{S}_n} \leq D(P||Q^*).$$  (27)

**Proof:** Application of Theorem 1 shows that

$$0 \leq E \left( \log \frac{b^*_i X_i}{\tilde{b}_i X_i} \right)_{X_1, \cdots, X_{i-1}}$$

$$\leq D(P_{X_1, X_2, \cdots, X_{i-1}}||Q_{X_1, X_2, \cdots, X_{i-1}}).$$

(28)

Averaging with respect to the distribution of $X_1 = (X_1, \cdots, X_{i-1})$ and then summing for $i = 1, 2, \cdots, n$ yields

$$0 \leq \sum_{i=1}^{n} E \left( \log \frac{b^*_i X_i}{\tilde{b}_i X_i} \right)_{X_1, \cdots, X_{i-1}}$$

$$\leq \sum_{i=1}^{n} E D(P_{X_1, X_2, \cdots, X_{i-1}}||Q_{X_1, X_2, \cdots, X_{i-1}})$$

$$= D(P||Q^*)$$

by the chain rule, completing the proof.

Suppose $X_1, X_2, \cdots$ are independent with unknown density $p(x)$. Clearly, the optimal portfolio $b^*$ does not depend on the time or on the past. However, if $p(x)$ is unknown, a series of estimators of the distribution $P_{b^*}(x)$ corresponding to density estimators $\hat{P}_i(x)$ based on the past $X_1-t$ may be used to obtain asymptotically optimal portfolios $\tilde{b}_i - b^*(\hat{P})$. It is often the case (see Barron [11], [12]) that there exists a sequence of estimators $\hat{P}_i$, converging to $P$ in the sense that $ED(P_{b^*}(\hat{P}) = 0$, at least in the Cesaro sense, i.e.,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E D(P_{b^*}(\hat{P}) = 0).$$  (30)

In this case Theorem 3 applies with $Q_{X_1, X_2, \cdots, X_{i-1}}$ given by the estimator $\hat{P}_i$ to yield

$$\lim_{n \to \infty} \frac{1}{n} E \log \frac{S^*_n}{\tilde{S}_n} = 0.$$  (31)

It follows that the actual wealth $\tilde{S}_n$ is close to the log-optimal wealth $S^*$ as shown in the following theorem.

**Theorem 4:** Let $X_1, X_2, \cdots$ be i.i.d. $\sim P$. Let $\tilde{P}_i$ be a sequence of estimators of the true distribution $P$ such that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E D(P_{b^*}(\hat{P}) = 0)$$

(32)

Let $S^*_n = \sum_{i=1}^{n} b^*(P) X_i$ be the optimal wealth sequence. Then

$$\tilde{S}_n = S^*_n 2^{o(1)}$$

(33)

where $o(1) \to 0$ in probability.

Consequently, if $S^*_n$ has an exponential growth rate $\tilde{e}^n$, then $\tilde{S}_n$ has the same asymptotic exponent.

**Proof:** To see that $\tilde{S}_n/S^*_n \to 2^{o(1)}$ in probability, first observe that by Markov’s inequality

$$P\left( \frac{\tilde{S}_n}{S^*_n} > 2^n \right) < 2^{-n} E \left( \frac{S^*_n}{\tilde{S}_n} \right)$$

(37)

where the inequality $E(\tilde{S}_n/S^*_n) \leq 1$ follows from the Kuhn–Tucker conditions for the optimality of $b^*$ (see Bell and Cover [8]).

On the other hand, using the notation $p^+ = \max \{0, y\}$, $y^- = \max \{0, -y\}$,

$$P\left( \frac{S^*_n}{\tilde{S}_n} > 2^n \right) = P\left( \frac{S^*_n}{\tilde{S}_n} > nt \right)$$

$$\leq \frac{1}{nt} E \left( \frac{S^*_n}{\tilde{S}_n} \right)$$

$$\leq 1 - \frac{1}{n} E \left( \frac{S^*_n}{\tilde{S}_n} \right)$$

(38)

where the first inequality follows from Markov’s inequality and the second from

$$E(\log S^*_n/\tilde{S}_n) = E \log \max \{ S^*_n/\tilde{S}_n \}$$

$$\leq E \log (1 + S^*_n/\tilde{S}_n)$$

$$\leq \log (1 + E(\tilde{S}_n/S^*_n)) \leq \log 2 = 1$$

(39)

by the concavity of the logarithm and the Kuhn–Tucker conditions. Combining (31), (37) and (38), we have $\tilde{S}_n/S^*_n \to 2^{o(1)}$ in probability, as claimed.

**VII. EXAMPLES**

We first give an example due to Kelly [6] in which $\Delta = I$. Here the stock market is a horse race, which, in the setup of (1), consists of a probability mass function $P(X = O, e_i) = p_i$, $i = 1, 2, \cdots, m$, where $e_i$ is a unit vector with 1 in the $i$th place and 0’s elsewhere, $O$ equals the win odds ($O$, for 1), and $p_i$ is the probability that the $i$th horse wins the race.

$$W(X) = \max_b E \log b X$$

$$= \max_b \sum_{i=1}^{m} p_i \log b_i O_i$$

(40)

$$= \sum p_i \log O_i - H(X)$$

where $H(X) = -\sum_{i=1}^{m} p_i \log p_i$. Also $b^* = p$, 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

\[ W(X|Y) = \sum_p p \log Q_i - H(X|Y) \]  

(41)

and

\[ b^* = p(X = o_i | y), \quad i = 1, 2, \ldots, m. \]

Here the optimal portfolio is to bet in proportion to the conditional probabilities, given Y. Subtracting (40) from (41), we have

\[ \Delta = W(X|Y) - W(X) = H(X) - H(X|Y) = I(X; Y). \]  

(42)

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 \( \Delta = 0, \quad I = 1 \). Let \( X = (1, 1/2) \) with probability 1/2, and \( X = (1, 3/4) \) with probability 1/2, and \( X = (1, 3/4) \) with probability 1/2. Let \( Y = X \). An investment in the first stock always returns the investment, 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

\[ I(X; Y) = I(X; X) = H(X) - H(X|X) \]

\[ = H(X) = 1 \quad \text{bit} \]  

(43)

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

VIII. CONCLUSION

We offer one final interpretation. Recall that \( H(X) - H(X|Y) = I(X; Y) \) 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 \( X \).

IX. ACKNOWLEDGMENT

We would like to thank R. O. Duda for speculations that led to the statement of Theorem 2.

REFERENCES


Code Construction for the Noiseless Binary Switching Multiple-Access Channel

PETER VANROOOSE

Abstract — The noiseless coding problem is considered for a recently introduced discrete memoryless multiple-access channel that is a counterpart to the well-known binary adder channel. Upper and lower bounds on the number of codewords in a uniquely decodable code pair are given, from which the zero-error capacity region of this channel is derived. This region coincides with the classical capacity region of this channel. The proof uses the new notion of second-order distance of a code. For several values of n and k, good code pairs of block length n are constructed with the first code being [n, k]-linear. Some of these are found to be optimal. Furthermore, some convolutional codes are investigated that yield additional good rate pairs.

I. THE BINARY SWITCHING MULTIPLE-ACCESS CHANNEL

Consider the classical two-access communication situation of Fig. 1, where two separate senders attempt to communicate to a third user, the receiver. The channel accepts two binary input streams, transmitted at the same symbol rate and divided into blocks of the same length n (assuming bit and block synchronization), and outputs a ternary stream according to the bitwise deterministic transitions depicted in Fig. 2: y = x1 + x2, where division by 0 gives \( \infty \). We call this channel the binary switching multiple-access channel (BS-MAC). Note that this model is completely noiseless. We will consider the problem where the decoder has to reconstruct the two messages error-free. This of course restricts the information rate of the input streams.

![Fig. 1. Two-access communication system.](image)

The only other deterministic binary two-input multiple-access channel (2-MAC) with ternary output is the binary adder channel (BAC), where the channel operation is \( y = x_1 + x_2 \). Noiseless coding for this channel has been studied by various authors (e.g. [1]–[5]). Surprisingly, the channels differ in many ways which will become clear in this correspondence.

Its relationship to the BAC was the first reason for studying the BS-MAC. Because both channels have the same input and

Manuscript received August 19, 1987; revised January 25, 1988. This work was supported in part by the Belgian Program for the Reinforcement of the Scientific Potential in the New Technologies and in part by the Ministry of Science Policy, Belgium, under Project GOA 83/88-53. This material was partially presented at the IEEE International Symposium on Information Theory, Ann Arbor, MI, October 1986, at the 7th Symposium on Information Theory in the Benelux, Deventer, The Netherlands, May 1987, and at the 3rd Joint Soviet-Swedish International Workshop on Information Theory, Sochi, USSR, May 1987. The author is with the Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200 B, B-3000 Heverlee, Belgium.

IEEE Log Number 8824064.