IIT - DELHI

MAL 390

REPORT

Information Theory

and its

Application in Stock Market

Submitted by:
Kushagra GUPTA (2012MT50599)
Madhav TAPARIA (2012MT50602)
Saransh MAHAJAN (2012MT50617)

April 15, 2015

Acknowledgement

We would like to show our gratitude to Dr. Niladri Chatterjee for giving us a good guideline for assignment throughout numerous consultations. It provided us a good platform to learn a lot. We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in writing this assignment.

Chapter 1 : Introduction to Information Theory

April 15, 2015

Information theory is based on probability theory and statistics. The most important quantities of information are entropy, the information in a random variable, and mutual information, the amount of information in common between two random variables. The former quantity gives the limit on how far message data can be compressed, while the latter can be used to find the communication rate across a channel. Mutual information is a special case of a more general quantity called relative entropy, which is a measure of the distance between two probability distributions. All these quantities are closely related and share a number of simple properties.

In later sections, we show how these quantities arise as natural answers to a number of questions in communication, statistics, complexity, and gambling. That will be the ultimate test of the value of these definitions.

1 ENTROPY

We first introduce the concept of entropy, which is a measure of the uncertainty of a random variable. Let X be a discrete random variable with alphabet X and probability mass function $p(x) = P\{X = x\}, x \in X$.

Definition : The entropy $H(X)$ of a discrete random variable X is defined by

$$H(X) = \sum p(x) \log p(x),\ where\ x \in X$$

The logarithm is to the base 2 and entropy is expressed in bits. For example, the entropy of a fair coin toss is 1 bit. We will use the convention that $0log0 = 0$, which is easily justified by continuity since $xlogx \to 0\ as\ x \to 0$. Adding terms of zero probability does not change the entropy. If the base of the logarithm is b, we denote the entropy as $H_b(X)$. Entropy is a functional of the distribution of X. It does not depend on the actual values taken by the random variable X, but only on the probabilities.

We denote expectation by E. Thus, if $X \sim p(x)$, the expected value of the random variable $g(X)$ is written

$$E_p g(X) = \sum g(x)p(x),\ where\ x \in X$$

Remark The entropy of X can also be interpreted as the expected value of the random variable $log\frac{1}{p(X)}$, where X is drawn according to probability mass function p(x). Thus,

$$H(X) = E_p log\frac{1}{p(X)}$$

Lemma 1 : $H(X) \geq 0$.

$$Proof : 0 \leq p(x) \leq 1 \ implies \ that \ log\frac{1}{p(X)} \geq 0$$

Lemma 2 : $H_b(X) = (log_b a)H_a(X)$.

$$Proof : log_b p = log_b a \ log_a p.$$

The second property of entropy enables us to change the base of the logarithm in the definition. Entropy can be changed from one base to another by multiplying by the appropriate factor.

Example 1 : Let X be a random variable which takes only two values 1 and 0 with probabilities p and 1-p respectively. Then,

$$H(X) = \text{-}p \ logp \text{ - } (1\text{-}p) \ log(1\text{-}p) = H(p) \qquad .(by \ definition)$$

In particular, $H(X) = 1$ bit when p = $\frac{1}{2}$. Some basic properties of entropy: It is a concave function of the distribution and equals 0 when p = 0 or 1. This makes sense, because when p = 0 or 1, the variable is not random and there is no uncertainty. Similarly, the uncertainty is maximum when p = $\frac{1}{2}$, which also corresponds to the maximum value of the entropy.

Example 2 : Consider a random variable that has a uniform distribution over 32 outcomes. To identify an outcome, we need a label that takes on 32 different values. Thus, 5-bit strings suffice as labels. The entropy of this random variable is

$$H(X) = -\sum_{i=1}^{32} p(i) * log \ p(i) = \sum_{i=1}^{32} \frac{1}{32} * log\frac{1}{32} = log32 = 5 \ bits$$

Example 3: Let X takes the colour red with probability 12 , blue with probability 14 , green with probability 18 and pink with probability 18. The entropy of X is

$$H(X) = -12log(12) - 14log(14) - 18log(18) - 18log(18) = 74 \ bits.$$

]

2 Joint Entropy and Conditional Entropy

We defined the entropy of a single random variable. Now we extend the definition to a pair of random variables.

Definition The joint entropy H(X,Y) of a pair of discrete random variables (X, Y) with a joint distribution p(x, y) is defined as :

$$H(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) * log(p(x,y))$$

which can also be expressed as

$$H(X,Y) = Elog\ p(X,Y).$$

We also define the conditional entropy of a random variable given another as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable.

Definition If (X, Y) $\sim$ p(x, y), the conditional entropy$H(Y|X)$ is defined as

$$H(X,Y) = \sum_{x \in X} p(x) * H(Y|X = x)$$

$$= \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) * log(p(y|x))$$

$$= \sum_{x \in X} \sum_{y \in Y} p(x,y) * log(p(y|x))$$

$$= -E\ log\ p(Y|X).$$

The naturalness of the definition of joint entropy and conditional entropy is exhibited by the fact that the entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other. This is stated in the following theorem.

Theorem 1: (Chain rule)

$$H(X,Y) = H(X) + H(Y|X).$$

3 Relative Entropy and Mutual Information

The entropy of a random variable is a measure of the uncertainty of the random variable; it is a measure of the amount of information required on the average to describe the random variable. In this section we introduce two related concepts: relative entropy and mutual information.

The relative entropy is a measure of the distance between two distributions. In statistics, it arises as an expected logarithm of the likelihood ratio. The relative entropy $D(p||q)$ is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p. For example, if we knew the true distribution p of the random variable, we could construct a code with average description length H(p). If, instead, we used the code for a distribution q, we would need $H(p) + D(p||q)$ bits on the average to describe the random variable.

Definition: The relative entropy or KullbackLeibler distance between two probability mass functions p(x) and q(x) is defined as

$$D(p||q) = \sum_{x \in X} p(x) * log(p(x)q(x))$$

$$= E_p \ log(p(x)q(x)).$$

Relative entropy is always nonnegative and is zero if and only if p = q. However, it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality. Nonetheless, it is often useful to think of relative entropy as a distance between distributions.

We now introduce mutual information, which is a measure of the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one random variable due to the knowledge of the other.

Definition: Consider two random variables X and Y with a joint probability mass function p(x, y) and marginal probability mass functions p(x) and p(y). The mutual information I (X; Y) is the relative entropy between the joint distribution and the product distribution p(x)p(y):

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) * log(\frac{p(x,y)}{p(x)p(y)})$$

$$= D(p(x,y)||p(x)p(y))$$

$$= E_p(x,y) log(\frac{p(x,y)}{p(x)p(y)}).$$

4 Relationship between Entropy and Mutual Information

$$I(X;Y) = H(X) + H(Y)H(X,Y).$$

The mutual information I (X; Y) is the reduction in the uncertainty of X due to the knowledge of Y. By symmetry, it also follows that

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

Finally, we note that

$$I(X;X) = H(X)H(X|X) = H(X).$$

Thus, the mutual information of a random variable with itself is the entropy of the random variable. This is the reason that entropy is sometimes referred to as self-information.

Theorem (Chain rule for entropy) Let$X_1, X_2, ..., X_n$be drawn according to $p(x_1, x_2, ..., x_n)$. Then,

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, ..., X_1)$$

Definition: The conditional mutual information of random variables X and Y given Z is defined by

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

$$= E_{p(x,y,z)} log(p(X,Y|Z)p(X|Z)p(Z)).$$

Mutual information also satisfies a chain rule. We define a conditional version of the relative entropy.

Definition For joint probability mass functions p(x, y) and q(x, y), the conditional relative entropy$D(p(y|x)||q(y|x))$ is the average of the relative entropies between the conditional probability mass functions $p(y|x) and q(y|x)$ averaged over the probability mass function p(x). More precisely,

$$D(p(y|x)||q(y|x)) = \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) * log(\frac{p(y|x)}{q(y|x)})$$

$$= E_p(x,y) log(\frac{p(Y|X)}{q(Y|X)}).$$

The notation for conditional relative entropy is not explicit since it omits mention of the distribution p(x) of the conditioning random variable. However, it is normally understood from the context.

The relative entropy between two joint distributions on a pair of random variables can be expanded as the sum of a relative entropy and a conditional relative entropy.

Theorem (Chain rule for relative entropy)

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x))$$

5 Asymptotic Equipartition Property

In information theory, the analog of the law of large numbers is the asymptotic equipartition property (AEP). It is a direct consequence of the weak law of large numbers. The law of large numbers states that for independent, identically distributed (i.i.d.) random variables, $\frac{1}{n}\sum_{i=1}^{n} X_i$is close to its expected value EX for large values of n. The AEP states that $\frac{1}{n}log\frac{1}{p(X_1,X_2,...,X_n)}$ is close to the entropy H, where $X_1, X_2, ..., X_n$. are i.i.d. random variables and $p(X_1, X_2, ..., X_n)$ is the probability of observing the sequence $X_1, X_2, ..., X_n$. Thus, the probability $p(X_1, X_2, ..., X_n)$ assigned to an observed sequence will be close to 2^{-nH} .

This enables us to divide the set of all sequences into two sets, the typical set, where the sample entropy is close to the true entropy, and the non-typical set, which contains the other sequences.

First, as an example. Let the random variable $X \in (0, 1)$ have a probability mass function defined by p(1) = p and p(0) = q. If $X_1, X_2, ..., X_n$ are i.i.d. according to p(x), the probability of a sequence $x_1, x_2, ..., x_n\ is\ \prod_{i=1}^{n} p(x_i)$. For example, the probability of the sequence (1, 0, 1, 1, 0, 1) is $p^{\sum X_i} q^{n-\sum X_i} = p^4q^2$. Clearly, it is not true that all 2^n sequences of length n have the same probability.

However, we might be able to predict the probability of the sequence that we actually observe. We ask for the probability $p(X_1, X_2, ..., X_n)$ of the outcomes $X_1, X_2, ..., X_n$, where $X_1, X_2, ..., X_n$ are i.i.d. $\sim p(x)$. This is insidiously self-referential, but well defined nonetheless. Apparently, we are asking for the probability of an event drawn according to the same probability distribution. Here it turns out that $p(X_1, X_2, ..., X_n)$ is close to 2^{nH} with high probability.

Definition (Convergence of random variables). Given a sequence of random variables, $X_1, X_2, ...$, we say that the sequence $X_1, X_2, ...$ converges to a random variable X:
1. In probability if for every $\epsilon > 0, Pr\{|X_n - X| > \epsilon\} \to 0$
2. In mean square if $E(X_n - X)^2 \to 0$
3. With probability 1 (also called almost surely) if $Pr\{lim_{n\to\infty} X_n = X\} = 1$.

Asymptotic Equipartition Property Theory
Almost all events are almost equally surprising. Theorem (AEP) If $X_1, X_2, ...$ are i.i.d. $\sim p(x)$, then $-\frac{1}{n}log\ p(X_1, X_2, , X_n) \to H(X)$ in probability.

Entropy Rates of a Stochastic Process
The asymptotic equipartition property establishes that nH(X) bits suffice on the average to describe n independent and identically distributed random variables. But what if the random variables are dependent? In particular, what if the random variables form a stationary process? We will show, just as in the i.i.d. case, that the entropy $H(X_1, X_2, ..., X_n)$ grows (asymptotically) linearly with n at a rate H(X), which we will call the entropy rate of the process.

6 Markov Chains

Definition A stochastic process is said to be stationary if the joint distribution of any subset of the sequence of random variables is invariant with respect to shifts in the time index; that is,

$$Pr\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\} = Pr\{X_{1+l} = x_1, X_{2+l} = x_2, ..., X_{n+l} = x_n\}$$

for every n and every shift l and for all $x_1, x_2, ..., x_n \in X$.
Definition: A discrete stochastic process $X_1, X_2, ...$ is said to be a Markov chain or a Markov process if for n = 1, 2, . . . ,

$$Pr(X_{n+1} = x_{n+1}|X_n = x_n, X_{n1} = x_{n1}, ..., X_1 = x_1) = Pr\{X_{n+1} = x_{n+1}|X_n = x_n\}$$

for all $x_1, x_2, ..., x_n, x_{n+1} \in X$. In this case, the joint probability mass function of the random variables can be written as

$$p(x_1, x_2, ..., x_n) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_n|x_{n1}).$$

Entropy Rate
If we have a sequence of n random variables, a natural question to ask is: How does the entropy of the sequence grow with n? We define the entropy rate as this rate of growth as follows. Definition: The entropy of a stochastic process $\{X_i\}$ is defined by

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, ..., X_n)$$

when the limit exists.

Illustration 1: Suppose there is a faulty keyboard that has m equally likely output letters. It can produce m^n sequences of length n, all of them equally likely. Hence $H(X_1, X_2, ..., X_n) = log\ m^n$ and the entropy rate is $H(X) = \log m$ bits per symbol.

Illustration 2: $X_1, X_2, \ldots$ are i.i.d. random variables. Then,

$$H(X) = \lim \frac{1}{n} H(X_1, X_2, ..., X_n) = \lim \frac{1}{n}(nH(X_1)) = H(X_1)$$

Chapter 2 : Information Theory in Gambling

April 15, 2015

Before proceeding to the application of Information theory in Stock Market, we must first look into its application in gambling. Stock market is much like gambling in many ways. Our study of gambling will form the basis for the Stock Market.

Here we will look upon the famous "Kelly Criterion" and its derivation.Then we will see an example of horse race which is a special case of stock market investment. Further we will look upon how returns and betting strategies change if we get some extra information.

1 Kelly Criterion

Kelly criterion is a formula used to determine the optimal size of a series of bets. In most gambling scenarios, and some investing scenarios under some simplifying assumptions, the Kelly strategy will do better than any essentially different strategy in the long run.

Kelly's insight suggested to maximize the expectation of the logarithm of the capital, rather than the expected profit from each bet. This is important, since in the latter case, one would be led to gamble all he had when presented with a favorable bet, and if he lost, would have no capital with which to place subsequent bets. Kelly realized that it was the logarithm of the gambler's capital which is additive in sequential bets, and "to which the law of large numbers applies."

For example, consider the example of coin toss. The probability of head-scoming in the toss is p. Suppose you have wealth 1 and you invest a fraction x of it. If the outcome is heads, you double whatever you invested otherwise you lose it. So in this case the expected return is

$$= p(1 + 2x) + (1 - p)(1 - x)$$

$$= x(3p - 1) + 1$$

So if $p > 1/3$ and we maximise the expectation, x would always come out to be 1, i.e. we always invest all our wealth in the gamble. But on a long run this is not a good option, because even one loss could make us lose all our wealth and we would not have any to invest further. So instead we should maximise the log of expectations.

1.1 Derivation of Kelly Criterion for a Simple bet

Suppose you have a simple bet where with probability p you will make a profit of b times what you bet, and otherwise you lose your bet. What is the optimal fraction of our bankroll to bet? So the expectation of logarithmic return is

$$E(\log(X)) = p\log(1 + bx) + (1 - p)\log(1 - x)$$

Differentiating it with respect to x and equaling it to 0 for maximisation.

$$\frac{pb}{1 + bx} - \frac{1 - p}{1 - x} = 0$$

$$pb * (1 - x) - (1 - p) * (1 + bx) = 0$$

$$bx = pb + p - 1$$

Solving we get,

$$x = \frac{pb - (1 - p)}{b}$$

This is the famous Kelly Criterion. Here numerator is balled the edge and b is called the odds. So it is also known as edge over odd rule.

In the previous example of coin toss that we discussed, we have $b = 2$ and consider $p = 1/2$. So the optimal fraction

$$x = \frac{(1/2) * 2 - (1 - (1/2)}{2} = 0.25$$

So this means we would invest one fourth of our wealth at a time. This is a small simualtion where I took four proportions 0.05, 0.25, 0.8 and 1 and tried to simulate the expected wealth after 10,50,100 and 200 coin tosses using a C++ program. Here is the table of wealth if all started with Rs 1.

No of Tosses	P1	P2	P3	P4
10	0.802739	1.80203	0.494265	0
50	2.24166	152.021	0.0226911	0
100	10.2364	180.549	0	0
200	18.8338	2037.38	0	0

From the table we could see that Kelly criterion is much more success ful than other proportions. In fact the wealth increases exponentially over time. On the other hand if we bet all our money, we would most certainly lose it all. So maximising the log of expectation is the real key. This introduces information theory as we are maximising the log of returns.

2 The "Horce Race" Example

We will consider the example of horse race for understanding the futher concepts. Assume that m horses run in a race. Let the i^{th} horse win with probability p_i. If horse i wins, the payoff is o_i for 1, i.e., an investment of one dollar on horse i results in a profit of o_i dollars if horse i wins and 0 dollars if horse i loses. Let b_i be the fraction of the gamblers wealth invested in horse i, where each $b_i \geq 0$ and $\Sigma\ bi = 1$. Then if horse i wins the race, the gambler will receive o_i times the amount of wealth bet on horse i. Let X be the random variable representing which horse wins the race. So after n races, your wealth will be

$$S_n = S_0 \prod_{i=1}^{n} b(X_i)o(X_i)$$

where S_n is the wealth after n races and S_0 is the initial wealth.The factor $S(X) = b(X)o(X)$ is called the wealth relative i.e. the factor by which the gamblers wealth grows if horse X wins the race. The log of expectation of wealth relative is called the doubling rate $W(b,p)$ of a horse race.

$$W(b,p) = E(\log(S(X))) = \sum_{i=1}^{m} p_i \log(b_i o_i)$$

Theorem : Let the race outcomes $X_1, X_2, ..., X_n$, be i.i.d. with distribution $p(x)$. Then the wealth of the gambler using betting strategy b grows exponentially at rate $W(b,p)$, i.e.,

$$S_n = S_0 2^{nW(b,p)}$$

Proof: Functions of independent random variables are also independent, and hence $\log S(X_1), \log S(X_2), ..., \log S(X_n)$ are i.i.d. Then, by the weak law of large numbers,

$$\frac{1}{n} \log \frac{S_n}{S_0} = \frac{1}{n} \sum_{i=1}^{n} \log S(X_i) \longrightarrow E(\log S(X) \quad \text{in probability.}$$

$$\Rightarrow S_n = S_0 2^{nW(b,p)}$$

Now the optimum doubling rate $W^*(p)$ is defined as the maximum doubling rate over all choices of the portfolio b, i.e.,

$$W^*(p) = max \sum_{i=1}^{m} p_i \log(b_i o_i) \quad \forall\ b \text{ such that } \ b_i \geq 0\ \ and \sum b_i = 1$$

To get this value, we maximize $W(b,p)$ as a function of$b$ subject to the constraint $\sum b_i = 1$. Writing the functional with a Lagrange multiplier, we have

$$J(b) = \sum p_i \log(b_i o_i) + \lambda \sum b_i$$

Differentiating w.r.t. b_i and equaling it to zero, we get

$$b_i = -\frac{p_i}{\lambda}.$$

Putting this back in the constraint $\sum b_i = 1$, we get the value of $\lambda = -1$ and thus we get $b_i = p_i$. Now we also need to verify that this is also a maxima but it is a difficult process as second derivatives tend to be complex in this case. So we are skipping the proof. This leads us to our next theorem,

Theorem *(Proportional gambling is log-optimal)* : The optimum doubling rate is given by

$$W^*(p) = \sum p_i \log o_i - H(p)$$

and is achieved by the proportional gambling scheme $b* = p$. Here $H(p)$ is the entropy of p.

Proof:

$$W(b,p) = \sum p_i \log b_i o_i$$

$$W(b,p) = \sum p_i \log(\frac{b_i}{p_i} o_i p_i)$$

$$W(b,p) = \sum p_i \log(o_i) - H(p) + \sum p_i \log(\frac{b_i}{p_i})$$

Here H(p)=$\sum -p_i \log(p_i)$ and the third term is the divergence ($D(p||b)$) of p and b.

$$\Rightarrow W(b,p) \leq \sum p_i \log(o_i) - H(p)$$

and the equality comes when $D(p||b) = 0$ i.e. p=b. Hence Proved.

We now consider a special case when the odds are fair, i.e., $\sum \frac{1}{o_i} = 1$. In this case, we write $r_i = \frac{1}{o_i}$, wherer_i can be interpreted as a probability mass function over the horses. (This is the bookies estimate of the win probabilities.) With this definition, we can write the doubling rate as

$$W(b,p) = \sum p_i \log b_i o_i$$

$$W(b,p) = \sum p_i \log(\frac{b_i p_i}{p_i r_i})$$

$$W(b,p) = D(p||r) - D(p||b)$$

This equation gives another interpretation for the relative entropy distance: the doubling rate is the difference between the distance of the bookies estimate from the true distribution and the distance of the gamblers estimate from the true distribution. Hence the gambler can make money only if his estimate (as expressed by b) is better than the bookies.

Theorem: *(Conservation theorem)*: For uniform fair odds,

$$W^*(p) + H(p) = \log m$$

Thus the sum of the doubling rate and the entropy rate is a constant.

Proof : Since by previous theorem

$$W^*(b,p) = \sum p_i \log o_i - H(p)$$

$$\Rightarrow W^*(b,p) = \sum p_i \log m - H(p)$$

$$\Rightarrow W^*(b,p) = \log m \sum p_i - H(p)$$

$$\Rightarrow W^*(b,p) = \log m - H(p)$$

Thus the sum of the optimal doubling rate and the entropy rate is a constant.

3 Effect of Side Information

Side Information is defined as any extraa information that was not available before. For example consider the case of Horse Race, suppose you get to know the past performances of the horses participating in the race, so we will now try to find that how can this information change the growth rate and optimum betting criterions.

Suppose horse x wins the race with probability $p(x)$ and pay odds of $o(x)$ for 1. Let (X,Y) have joint probability mass function $p(x,y)$. Let $b(x|y) \geq 0$, $\sum b(x|y) = 1$ be an arbitrary conditional betting strategy depending on the side information Y, where $b(x|y)$ is the proportion of wealth bet on horse x when y is observed. As before, let $b(x) \geq 0$, $\sum b(x) = 1$ denote the unconditional betting scheme.

We know,

$$W^*(X) = \sum p(x) \log o(x) - H(X)$$

and,

$$W^*(X|Y) = max \sum p(x,y) \log b(x|y) o(x)$$

With side information, the maximum value of $W^*(X|Y)$ with side information Y is achieved by conditionally proportional gambling, i.e., $b^*(x|y) = p(x|y)$. Thus

$$W^*(X|Y) = \sum p(x,y) \log p(x|y) o(x)$$

$$W^*(X|Y) = \sum p(x) \log o(x) - H(X|Y)$$

Putting values from earlier equation,

$$W^*(X|Y) = W^*(X) + H(X) - H(X|Y)$$

$$W^*(X|Y) - W^*(X) = H(X) - H(X|Y) = I(X;Y)$$

Hence, gain in doubling rate is equal to the Information Gain caused by the new information. This equation would be further used in Stock Market.

4 Example/Simulation

Consider that we have 8 horses that run in a race. You ave a total of 100 rupees in the start to bet on. You have to bet all your money but we could hedge it. The odds are all equal and fair, i.e., the odds are 8 to 1 on each horse. (You get 8 rupees if you bet 1 rupee on the winning horse and lose the rupee if you bet on the losing horse. The underlying probability of horses winning are $12/80, 13/80, 7/80, 11/80, 10/80, 9/80, 5/80, 13/80$. But these values are unknown to all and are just meant for simulations so that each horse does not wins uniformly.

Now I created a MATLAB programs that simulates the performance of gamblers. The program randomly makes a horse win on each iteration depending on a probability distribution that is not available to the gambler. I have created the following 4 gamblers :

Random gambler : This player randomly hedges its bet. So we could consider it the most unexperienced player. Its success could not be predicted and on a long run it will lose all the money.

Constant gambler : This player always hedges its bet a fixed proportion that is decided in the start. I made it $1/12, 1/6, 1/12, 1/6, 1/12, 1/6, 1/12, 1/6$ repectively. So we could consider it the not learning player. Its success depends on the fact how the money is divided at the start and what is the hidden distribution of winning. So it all comes down to luck. If fixed proportion is in favour of horse with maximum winning probability then he will win or If fixed proportion is in favour of horse with minimum winning probability then he will lose on a long run.
We know

$$W(b,p) = D(p||r) - D(p||b)$$

In this case

$$D(p||r) =$$

$$D(p||b) =$$

So the gambler will make some profit.

Uniform Gambler : This player always hedges its bet a fixed proportion that is equal to inverse of odds. So see from the mathematics that it would never lose or earn anything. It is the safest player.

$$W(b,p) = D(p||r) - D(p||b) = 0 \; beacuse \; r = b$$

Learning gambler : This is the most interesting player. It changes its betting proportions after each race depending on the the outcomes of previous races which it considers as side information. On the first iteration it considers the underlying distributuion as uniform i.e. it bets equally on all the horses. Then after n races the proportion of i^{th} horse becomes

$$\frac{\sum \text{races won by this horse} + 1}{Total\ Races + 8}$$

. The one in numerator and 8 in denominator is apperaing due to Laplace Smoothing because in the start when just races are less, we do not want to bid zero on any horse.
Now this gambler must perform the best as it is using the Side Information but the change in doubling rate is very difficult to calculate as the entropy of distributuion is changin at every instant. But if we consider enough iterations, the betting proportion would tend to the underlying probability, so the doubling rate would come as

$$W(b,p) = D(p||r) - D(p||b)$$

In this case

$$D(p||r) =$$

$$D(p||b) =$$

I am attaching the observations in the form of graphs and the codes used. All the observations are as predicted.

In real life the odds also change after each races and they are also not fair. Moreover one could decide that he doesn't needs to bet all his money. So this all theory needs slight modifications for application in real life.

CODES

Random Player :

```
s=[];
setdemorandstream(1000);

r = 80.*rand(1000,1);

s0=100.0;

b=[1 1 1 1 1 1 1 1];
for i=1:100

    b=1000.*rand(8,1);

    if(r(i)<12)
        s0=s0*(b(1)/sum(b))*8;
    else if(r(i)<25)
            s0=s0*(b(2)/sum(b))*8;
        else if(r(i)<32)
                s0=s0*(b(3)/sum(b))*8;
            else if(r(i)<43)
                     s0=s0*(b(4)/sum(b))*8;
                else if(r(i)<53)
                        s0=s0*(b(5)/sum(b))*8;
                    else if(r(i)<62)
                            s0=s0*(b(6)/sum(b))*8;
                        else if(r(i)<67)
                                s0=s0*(b(7)/sum(b))*8;
                            else
                                s0=s0*(b(8)/sum(b))*8;
                            end
                        end
                    end
                end
            end
        end
    end




     s(i)=s0;


end

plot(s)
```

Constant Player

```
s=[];
setdemorandstream(1000);

r = 80.*rand(1000,1);

s0=100.0;

b=[1 2 1 2 1 2 1 2];
for i=1:1000

    if(r(i)<12)
        s0=s0*(b(1)/sum(b))*8;
    else if(r(i)<25)
            s0=s0*(b(2)/sum(b))*8;
        else if(r(i)<32)
                s0=s0*(b(3)/sum(b))*8;
            else if(r(i)<43)
                     s0=s0*(b(4)/sum(b))*8;
                else if(r(i)<53)
                        s0=s0*(b(5)/sum(b))*8;
                    else if(r(i)<62)
                            s0=s0*(b(6)/sum(b))*8;
                        else if(r(i)<67)
                                s0=s0*(b(7)/sum(b))*8;
                            else
                                s0=s0*(b(8)/sum(b))*8;
                            end
                        end
                    end
                end
            end
        end
    end




     s(i)=s0;


end

plot(s)
```

Uniform Player

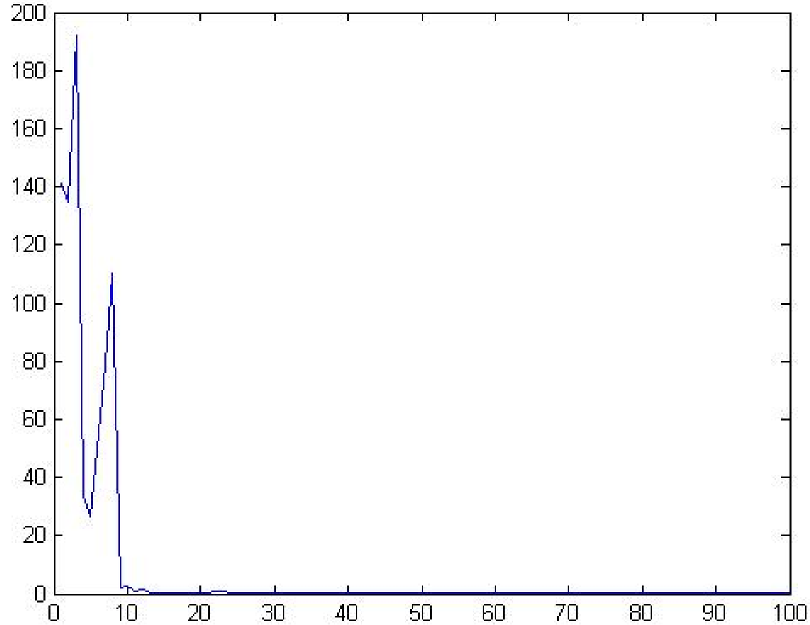
```
s=[];
setdemorandstream(1000);

r = 80.*rand(1000,1);

s0=100.0;

b=[1 1 1 1 1 1 1 1];
for i=1:1000

    if(r(i)<12)
        s0=s0*(b(1)/sum(b))*8;
    else if(r(i)<25)
            s0=s0*(b(2)/sum(b))*8;
        else if(r(i)<32)
                s0=s0*(b(3)/sum(b))*8;
            else if(r(i)<43)
                     s0=s0*(b(4)/sum(b))*8;
                else if(r(i)<53)
                        s0=s0*(b(5)/sum(b))*8;
                    else if(r(i)<62)
                            s0=s0*(b(6)/sum(b))*8;
                        else if(r(i)<67)
                                s0=s0*(b(7)/sum(b))*8;
                            else
                                s0=s0*(b(8)/sum(b))*8;
                            end
                        end
                    end
                end
            end
        end
    end




     s(i)=s0;


end

plot(s)
```

Learning Player

```
s=[];
setdemorandstream(1000);

r = 80.*rand(1000,1);

s0=100.0;

b=[1 1 1 1 1 1 1 1];

for i=1:1000

    if(r(i)<12)
        s0=s0*(b(1)/sum(b))*8;
        b(1)=b(1)+1;
    else if(r(i)<25)
            s0=s0*(b(2)/sum(b))*8;
            b(2)=b(2)+1;
        else if(r(i)<32)
                s0=s0*(b(3)/sum(b))*8;
                b(3)=b(3)+1;
                else if(r(i)<43)
                     s0=s0*(b(4)/sum(b))*8;
                     b(4)=b(4)+1;
                else if(r(i)<53)
                        s0=s0*(b(5)/sum(b))*8;
                        b(5)=b(5)+1;
                    else if(r(i)<62)
                            s0=s0*(b(6)/sum(b))*8;
                            b(6)=b(6)+1;
                        else if(r(i)<67)
                                s0=s0*(b(7)/sum(b))*8;
                                b(7)=b(7)+1;
                            else
                                s0=s0*(b(8)/sum(b))*8;
                                b(8)=b(8)+1;

                            end
                        end
                    end
                end
            end
        end
    end
s(i)=s0;
end
plot(s);
```

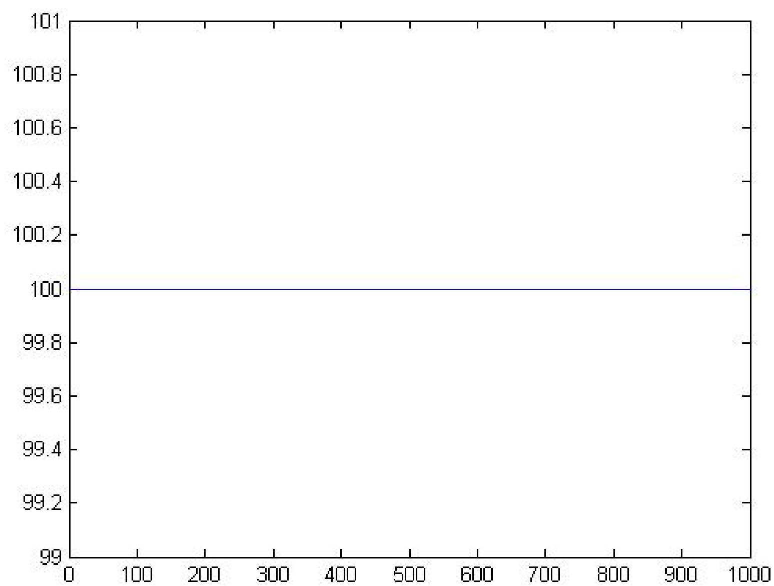
GRAPHS

1. **Random Gambler :**

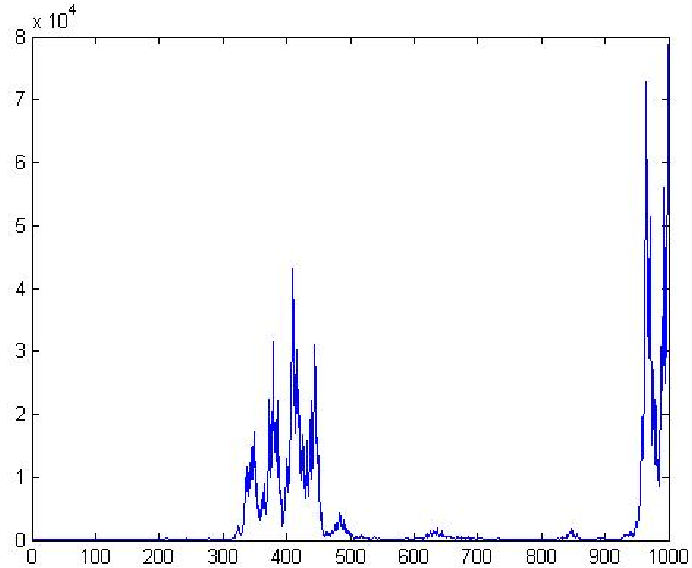


As guessed, the amount tends to zero as the number of iteration increases.

2. Uniform Gambler

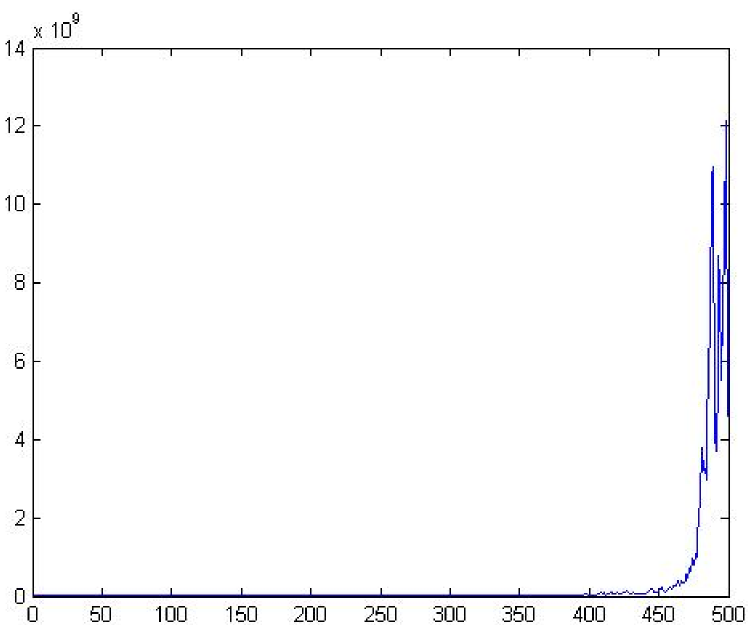


3. Constant Gambler



As suggested in theory, this gambler makes profit.

4. Learning Gambler



This is also according to the theory. In just 500 iterations it reaches around 10^10, which is much larger than all the other gamblers.

Chapter 3 : Information Theory in Stock Market

April 15, 2015

1 Introduction

It was the goal of individual, nations and societies since the dawn of human race to accumulate wealth, sometimes to satisfy basic needs for existence, and sometimes just for the sake of wealth accumulation.

In recent century, financial markets became increasingly attractive for individuals and institutions alike, and are crucial part of entire economy, due to their role in facilitating the raise of capital, hedging, accumulation of wealth, and international trade. Stock market is closely related to the concept of information theory.

As discussed before, a key measure of information is entropy. The duality between the growth rate of wealth in the stock market and the entropy rate of the market is striking. We will look at the growth rate optimal portfolio strategies.

2 Stock Market : Definitions

A stock market is represented as a vector of stocks $\mathbf{X} = (X_1, X_2, ...X_m)$, $X_i \geq 0$ $i = 1, 2, ...m$ where m is the number of stocks and the price relative X_i is the ratio of the price at the end of the day to the price at the beginning of the day. So typically, X_i is near 1. For example,$X_i = 1.05$ means that the i^{th} stock went up by 5 percent that day.

Now let us assume that $X \sim F(x)$, where F(x) is the joint distribution of the vector of price relatives. A portfolio $b = (b_1, b_2, ..., b_m)$, $b_i \geq 0$, $\sum_{i=1}^{m} b_i = 1$ is an allocation of wealth across the stocks. Here b_i is the fraction of one's wealth invested in stock i. For example, we can have for m=3, the distribution amongst 3 stocks as the fraction 1/2,1/4,1/4 . Clearly, if we use the portfolio $\mathbf{b}$ and the stock vector $\mathbf{X}$, the wealth relative (that is, ratio of the wealth at the end of the day to the wealth at the beginning of the day) becomes $S = \mathbf{b}^t\mathbf{X} = \sum_{i=1}^{m} b_i X_i$

Now, S is a measure of our wealth relative as previously defined, so obviously we would want to maximise this in some way. But the problem is that S is a random variable, because the vector of stocks is a random variable; the distribution of S depends on the portfolio **b**. Hence there is a controversy regarding the best choice for the distribution of S.

3 Markowitz theory

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

The meanvariance approach is the basis of the Sharpe-Markowitz theory of investment in the stock market . It is shown in figure 1. The figure illustrates the set of achievable meanvariance pairs using various portfolios. The set of portfolios on the boundary of this region corresponds to the undominated portfolios,undominated because these are the portfolios that have the highest mean for a given variance. This boundary is called the **efficient frontier**, and if one is interested only in mean and variance, one should operate along this boundary.

In general, we can also have risk free assets(i.e assets, which unlike the risky stocks are not random variables and give returns which are known well in advanced).Examples of such assets are cash or Treasury bonds, which provide a fixed interest rate with zero variance. This stock corresponds to a point on the Y axis in the figure. By combining the risk-free asset with various stocks, one obtains all points below the tangent from the risk-free asset to the efficient frontier.Now the efficient frontier gets changed by the introduction of risk free assets and this line becomes the new efficient frontier.

4 CAPM

The concept of the efficient frontier also implies that there is a true price for a stock corresponding to its risk. This theory of stock prices, called the capital asset pricing model (CAPM), is used to decide whether the market price for a stock is too high or too low. Looking at the mean of a random variable gives information about the long-term behavior of the sum of i.i.d. versions of the random variable. But in the stock market, one normally reinvests every day, so that the wealth at the end of n days is the product of factors, one for each day of the market. The behavior of the product is determined not by the expected value but by the expected logarithm. This is because we expect that after every day, our wealth will increase in some way that resembles the wealth getting "added" This leads us to define the growth rate as follows:

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

$$W(\mathbf{b}, F) = \int log \mathbf{b}^t \mathbf{x} dF(\mathbf{x}) = E(log \mathbf{b}^t \mathbf{X}) \tag{1}$$

If additionally the logarithm is to base 2, the growth rate is also called the doubling rate, as will shortly become clear.

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

$$W^*(F) = max_b W(\mathbf{b}, F) \tag{2}$$

where the maximum is over all possible portfolios $b_i \geq 0$, $\sum_{i=1}^{m} b_i = 1$

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

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

Theorem : Let $X_1, X_2, ..., X_n$ be i.i.d. according to F(x). Let

$$S_n{}^* = \prod_{i=1}^{n} b^{*t} \mathbf{X}_i \tag{3}$$

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

$$\frac{1}{n} log S_n{}^* \rightarrow W^* \text{ with probability 1.} \tag{4}$$

Proof : By the strong law of large numbers,

$$\frac{1}{n} log S_n{}^* = \frac{1}{n} \sum_{i=1}^{n} log \mathbf{b}^{*t} \mathbf{X}_i \tag{5}$$

$$\rightarrow W^* \text{ with probability 1.} \tag{6}$$

Hence, $S_n{}^* = 2^{nW^*}$

Lemma : W(**b**, F) is concave in **b** and linear in F. $W(F)$ is convex in F.

The proof is pretty straightforward, relying on the facts that integral for growth rate is linear in F, log is a concave function and the linearity of W(**b**,F) on F.

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

Again, the proof is simple and depends on the previous lemma.

5 Kuhn Tucker Characterization of the Log Optimal Portfolio

The determination of **b** that achieves $W(F)$ is a problem of maximization of a concave function W(**b**, F) over a convex set B. The maximum may lie on the boundary. The standard KuhnTucker are used conditions to characterize the maximum.

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

$$E\left(\frac{X_i}{\mathbf{b}^{*t}}\right) = 1 \text{ if } {b_i}^* > 0 \tag{7}$$

$$E\left(\frac{X_i}{\mathbf{b}^{*t}}\right) \leq 1 \text{ if } {b_i}^* = 0 \tag{8}$$

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

$$E\ \ln\frac{S}{S^*} \leq 0 \text{ for all S} \iff E\frac{S}{S^*} \leq 1 \text{ for all S} \tag{9}$$

The proof is simple and makes use of Jensen's inequality.

Now,as mentioned before, maximizing the expected logarithm was motivated by the asymptotic growth rate. But we have just shown that the log-optimal portfolio, in addition to maximizing the asymptotic growth rate, also maximizes the expected wealth relative $E(\frac{S}{S})$ for one day.

Also, because of the above characterizations, we have that the expected proportion of wealth in each stock under the log-optimal portfolio is unchanged from day to day. Consider the stocks at the end of the first day. The initial allocation of wealth is $\mathbf{b}^*$.

The proportion of the wealth in stock i at the end of the day is $\frac{{b_i}^* X_i}{\mathbf{b}^{*t}\mathbf{X}}$ and the expected value of this proportion is

$$E\frac{{b_i}^* X_i}{\mathbf{b}^{*t}\mathbf{X}} = {b_i}^* E\frac{X_i}{\mathbf{b}^{*t}\mathbf{X}} = {b_i}^* \tag{10}$$

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

6 Asymptotic Optimality of the Log Optimal Portfolio

The motivation of the log-optimal portfolio has been shown to be in terms of the long-term behavior of a sequence of investments in a repeated independent versions of the stock market. Now we proceed to show that with probability 1, the conditionally log-optimal investor will not do any worse than any other investor who uses a causal investment strategy.

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

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i{}^t \mathbf{X}_i \tag{11}$$

be the wealth after n days for an investor who uses portfolio bi on day i.

Also let

$$W^* = max_b W(\mathbf{b}, F) = max_b E \text{ log} \mathbf{b}^t \mathbf{X} \tag{12}$$

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

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

From the definition of W, it follows immediately that the log-optimal portfolio maximizes the expected log of the final wealth.

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

$$E \text{ log} S_n{}^* = nW^* \geq E \text{ log} S_n \tag{13}$$

Theorem : Let $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be a sequence of i.i.d. stock vectors drawn according to F($\mathbf{x}$). Let $S_n{}^* = \prod_{i=1}^{n} \mathbf{b}^{*t} \mathbf{X}_i$ be the wealth resulting from any other causal portfolio. Then

$$lim \text{ sup}_{n\to\infty} \frac{1}{n} log \frac{S_n}{S_n{}^*} \leq 0 \text{ with probability 1.} \tag{14}$$

Using the KKT conditions, the log optimality of ${S_n}^*$ and the Markov's inequality, and also the Borel-Cantelli lemma we arrive at the proof of this theorem.

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

7 Side Information and the Growth Rate

It has been earlier shown in the horse race case that side information can be used to increase the growth rate by the mutual information I(X;Y), where X denotes the horse race and Y the side information. Now, let's look at the same result for the stock market. Here, I (X; Y) is an upper bound on the increase in the growth rate, with equality if X is a horse race. The obvious difference between the horse race case and the stock market is that in horse race, the payoff of just one horse is there and the rest of the money invested in different horses goes to 0, while in the stock market it is not so extreme, and usually none of the stocks go to 0.

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

$$\Delta W = W(\mathbf{b}_f, F) - W(\mathbf{b}_g, F) \leq D(f\|g). \tag{15}$$

Here $D(f\|g)$ is the divergence of f from g.

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

$$\Delta W \leq I(\mathbf{X}; Y). \tag{16}$$

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

$$\Delta W_{Y=y} \leq D(f(\mathbf{x}|Y=y)||f(\mathbf{x})) = \int_x f(\mathbf{x}|Y=y) log \frac{f(\mathbf{x}|Y=y)}{f(\mathbf{x})} d\mathbf{x} \tag{17}$$

Taking the average over all values of y, we have

$$\Delta W \leq= \int_y f(y) \int_x f(\mathbf{x}|Y=y) log \frac{f(\mathbf{x}|Y=y)}{f(\mathbf{x})} d\mathbf{x} dy \tag{18}$$

$$\Delta W \leq \int_y \int_x f(y) f(\mathbf{x}|Y=y) log \frac{f(\mathbf{x}|Y=y)f(y)}{f(\mathbf{x})f(y)} d\mathbf{x}dy \tag{19}$$

$$\Delta W \leq \int_y \int_x f(\mathbf{x}, y) log \frac{f(\mathbf{x}, y)}{f(\mathbf{x})f(y)} d\mathbf{x}dy \tag{20}$$

$$\Delta W \leq I(\mathbf{X}; Y). \tag{21}$$

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

8 Some examples

8.1 Example 1

We have seen earlier that for the log optimal portfolio, it makes sense to use the same constant rebalanced portfolio for even the next 'n' time steps.

In order to show this, consider a stock whose price remains constant, and another stock whose price alternately doubles and halves.We will show that in this scenario, the wealth increases exponentially. In this example at the end of each day we take the wealth and redistribute it into all the stocks based on a given distribution $\mathbf{b}$. This is called a constant rebalanced portfolio (CRP). Now if we let $\mathbf{b} = [1/2, 1/2]$,then the wealth grows as

$$S_1 = 1 \tag{22}$$

$$S_2 = S_1 \left(\frac{1}{2}.1 + \frac{1}{2}.\frac{1}{2} \right) = \frac{3}{4} S_1 \tag{23}$$

$$S_3 = S_2 \left(\frac{1}{2}.1 + \frac{1}{2}.2 \right) = \frac{3}{2} S_2 = \frac{9}{8} S_1 \tag{24}$$

Continuing in this manner, we find that in general we have that

$$S_{t+2} = \frac{3}{2}.\frac{3}{4} S_t = \frac{9}{8} S_t \tag{25}$$

Hence, it can be seen that after every two days the money grows by $\frac{1}{8}_{th}$. This implies that the money is growing exponentially fast.

8.2 Example 2

If an investor purchased 1 share of each company, the return in 10 years would be 9the return on investment would be equal to 19

(The data for the research was obtained from finance.yahoo.com, and the stock price reflects the splits and dividends adjustment.)

According to CRP, which is based on Information Theory, at the end of each trading day redistribute the wealth to all stocks given the vector:

$$\mathbf{b} = (b_1, b_2, ...b_m) \tag{26}$$

To observe how CRP will affect the wealth over time, we take 10 stocks in our portfolio and redistribute the wealth at the end of each day.

$$\mathbf{b} = (\frac{1}{10}, \frac{1}{10}, ...\frac{1}{10}) \tag{27}$$

If we disregard the commissions and fees, then at the end of the 2009, our portfolio return is 39where our wealth were equally distributed among 10 stocks.

However, if we would allocate all of our wealth in 3 stocks: BA, CAT and MCD in 3/1/2000, we can observe that our CRP strategy would not be superior to buy and hold, because the return on investment on three stocks is higher than the return on CRP.

When we assign equal weights to each stock in our portfolio, it is called uniform rebalanced portfolio. Since the weights can be easily adjusted, we are not constrained to keep equal amount of wealth in each stock. Simply adjusting our portfolio b:

$$\mathbf{b} = [0.2, 0, 0.1, 0.2, 0.3, 0, 0, 0.2, 0, 0] \tag{28}$$

$$\mathbf{b} = [KO, AA, WMT, CAT, BA, T, PFE, MCD, MSFT, GE] \tag{29}$$

Rebalancing our portfolio at the end of each trading day, we will achieve 218 percent return on our initial investment.

This clearly beats the best buy and hold strategy in the portfolio, which would be Caterpillar (NYSE: CAT). Another angle that CRP will be a better option than buy and hold or any other strategies is when we have picked losers in our portfolio. Consider a scenario where we picked two stocks in our portfolio, which were Alcoa (NYSE: AA), and General Electric (NYSE: GE). If we had decided to equally divide our wealth into two stocks and apply buy and hold strategy, then at the end of the investment period we would lose 55 percent of our wealth. Now if we use CRP and apply various weights, then our return on investment will still be negative but CRP outperforms buy and hold strategy. The following table summarizes portfolio return when different weights are applied.

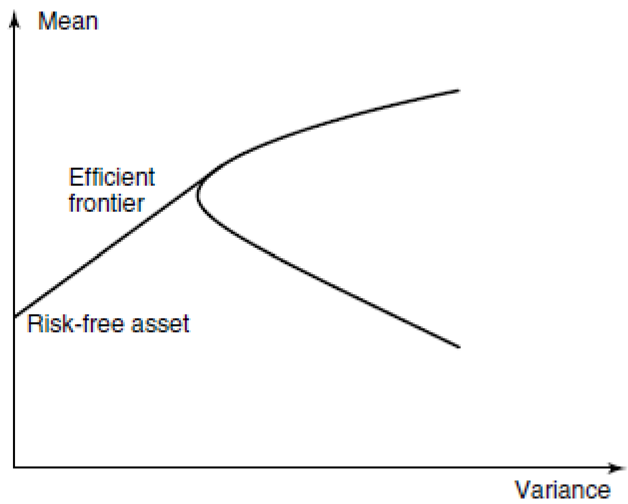
As we can see from the table, the best result is achieved when two stocks are constantly rebalanced at 60 percent to Alcoa stock and 40 percent to GE ratio. The following graphs represent the wealth increase or decrease for buy and hold vs. CRP.

We can see from the graph that CRP outperforming the Buy and Hold strategy in both scenarios, that is maximizes the revenue and minimizes the risk. There are also two important things can be derived by analyzing the graph: - There exists a portfolio b, which is the optimal weight and if one uses it, he or she can maximize the return on investment. - Second observation indicates that CRP follows the trend and yields better return when the portfolio is trending up, and minimizes the loss when the portfolio trends down.

One of the weaknesses of CRP offered by Cover is that it tends to follow the trend. That is if stock goes up, CRP will beat the best performing stock, and if the portfolio goes down it will not perform worse than other strategies. So its risk and reward trade off CRP is more geared towards the reward side.

9 Conclusion

As far as we have discussed about this topic we can see that this topic is very interesting and important to the economic and financial industry. The outstanding works by many researchers have pushed information theory and the development of investing on a stock market moving forward. However, the research in this field of information theory is far from settled. It is still wide open and has a lot of potential. The more we know about this topic the more return we can make from our investment and the better we can improve the economic and financial industry.



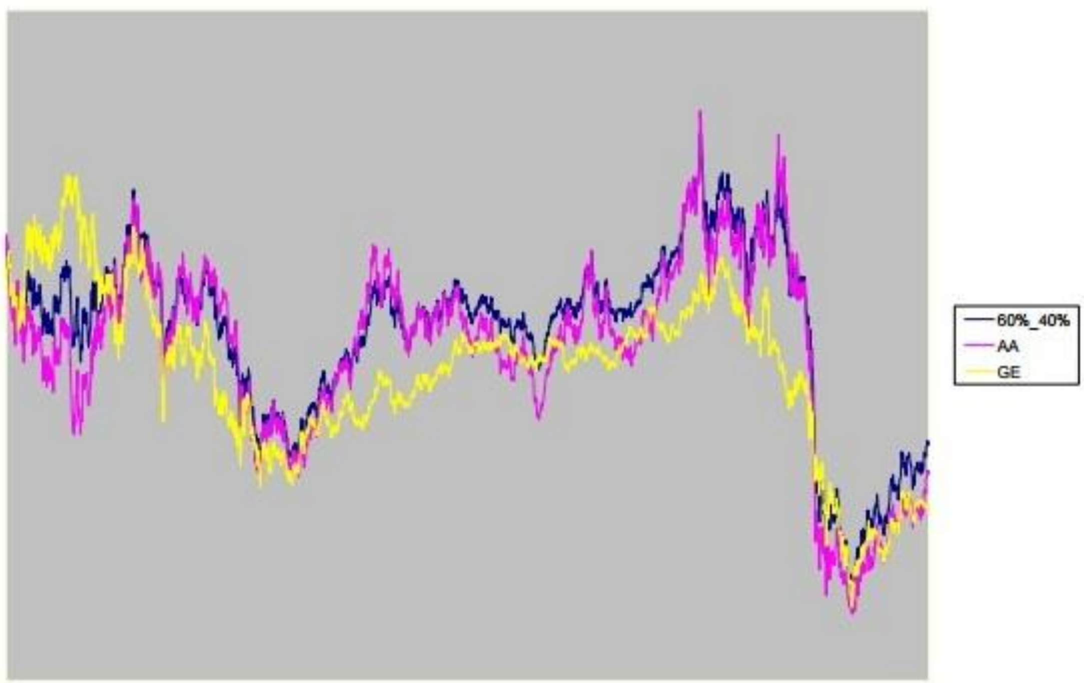
Stock 2 - Price doubles / halves alternatingly every two days

	Stock 1's p_t	Stock 2's p_t
Day 1	1	$\frac{1}{2}$
Day 2	1	2
Day 3	1	$\frac{1}{2}$
Day 4	1	2
$\vdots$	$\vdots$	$\vdots$

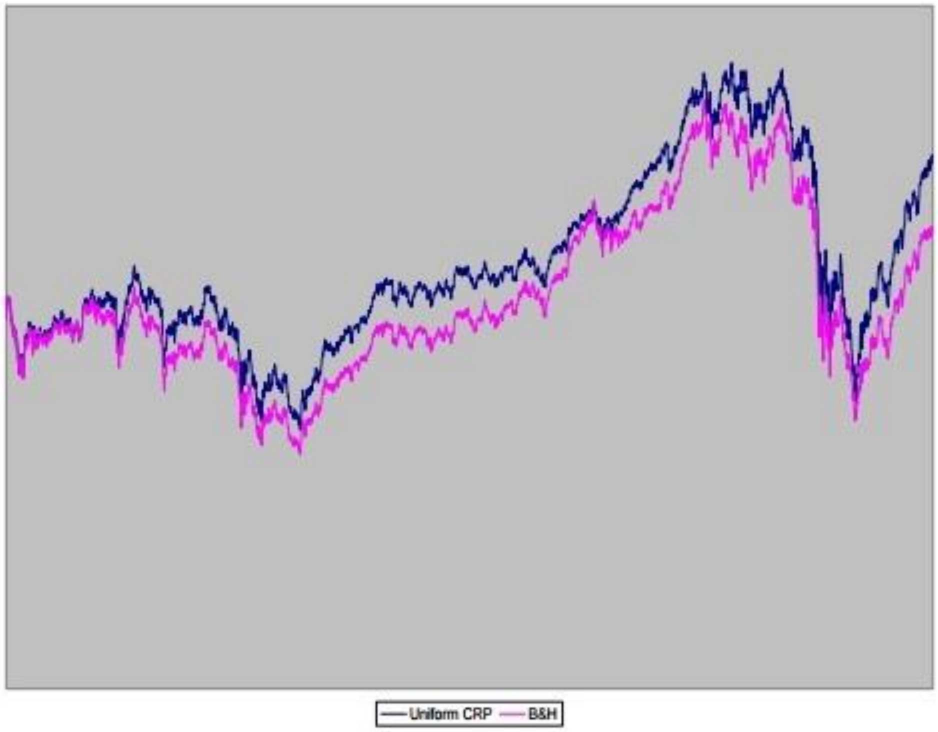
Buy and Hold Single Stock Strategy			
Stock	**Stock Adjusted price (1/03/2000)**	**stock price (12/31/2009)**	**Return %**
KO	$ 45.03	$ 57.00	27%
AA	$ 32.59	$ 16.12	-51%
WMT	$ 59.46	$ 53.45	-10%
CAT	$ 18.61	$ 56.60	204%
BA	$ 33.29	$ 54.13	63%
T	$ 30.06	$ 27.62	-8%
PFE	$ 23.57	$ 18.19	-23%
MCD	$ 32.57	$ 62.44	92%
MSFT	$ 47.64	$ 30.48	-36%
GE	$ 37.36	$ 15.13	-60%

AA	GE	ROI
10%	90%	-55.32%
20%	80%	-51.56%
30%	70%	-48.38%
40%	60%	-45.96%
50%	50%	-44.38%
60%	40%	-43.74%
70%	30%	-44.08%
80%	20%	-45.38%
90%	10%	-47.58%

CRP vs. Buy&Hold



Uniform CRP vs. Buy & Hold



References

[1] T. Cover and J. Thomas, Elements of Information Theory, Wiley, New York, 2 edition, 2006.

[2] Harry Markowitz, Portfolio selection, Journal of Finance, vol. 7, no. 1, pp. 7791, 1952.

[3] Eugene F Fama and Kenneth R French, The capital asset pricing model: Theory and evidence, Journal of Economic Perspectives, vol. 18, no. 3, pp. 2546, 2004.

[4] Yongjia Xu, Yongzeng Lai, and Xiaojing Xi, Optimal portfolios with power and log utilities, in Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on, april 2011, pp. 1271 1275.

[5] J Kelly, A new interpretation of information rate, IEEE Transactions on Information Theory, vol. 2, no. 3, pp. 185189, 1956.