Information Theory and Allocating Resources

John Albers

Abstract

The world of information theory is inextricable linked to the world of gambling. Here, gambling takes on the meaning of any situation where resources are distributed among various investments with a hope of maximizing the total return. A look at how the doubling rate of a gambling scheme can be related to the amount of side information about the gamble present is discussed, as well as a discussion on how the increase of side information directly relates to an increase in the upper bound of the doubling rate. The optimal gambling scheme under "horse race" situations is shown to be Kelly gambling and algorithms for the best allocation of resources under such circumstances are discussed.

The risk of a gamble can be minimized if the gambler knows something about the true distribution of outcomes. In situations where the true distribution is not known, the gambler must make an assumption on the distribution of the outcomes and pay the penalty of extra risk in the form of a Kullback-Liebler distance between the true and estimated distribution. This risk can be bound tighter by making a connection between the KL-divergence and the Hellinger distance.

Finally, a discussion of information theory's relation to investments in stock markets is presented. It is shown that the there is an optimal portfolio strategy that will maximize the doubling rate of investments. This optimal portfolio strategy is impossible to obtain without extraordinary side information, but algorithms exist to come reasonably close.

I. Introduction

In 1948, Claude Shannon founded the basis of Information Theory with his paper 'The Mathematical Theory of Communication'. One of the most remarkable achievements he made was his linking of the thermodynamic concept of entropy, to the measure of information. It seems that in mathematics, important concepts have a way of reappearing. How many times have you been working an a problem and found comfort when a π or Euler's number shows up in the answer. If we attempt to make observations about the world of gambling from a mathematical point of view, we might have a beacon that we are heading in the right direction if we stumble upon familiar concepts. Shannon showed that there is a statistical nature to information, and the world of gambling is heavily dominated by statistics, so it may seem intuitive that the two worlds can be linked. However, some people feel that Shannon's view of information is incomplete. We may need to look at the timeliness of data, and the spatial aspect of information [1]. But regardless of what information is, linking Shannon's measure of information, entropy, to gambling would certainly allow us more insight into the nature of gambling. It would also in some ways verify the importance of this measure of information, strengthening it as a theme that continuously reappears in the universe.

II. Gambling

Suppose we are in a situation where we have a finite amount of a given resource, and a decision must be made on how to allocate that resource so as to maximize the return of some other desirable resource. Possible situations fitting this criteria are a fisherman allocating different lengths of time to different locations on a lake, hoping to maximize the amount of fish he catches, a farmer deciding which crop to plant that will give the best return on the market, or a player placing bets at a roulette wheel, trying to maximize their monetary return.

There is not really that much a difference in these activities, for although we call one fishing, the other farming, and the last one gambling, they all have something to do with chance, and they all seem like there should be a best strategy, if only a little bit more information was available. If someone is allocating capital with a quick expected return, we typically call this gambling; and if the expected return is longer we typically call this investing. Any decision where the best action is not known to the "acter," is a gamble.

A. Maximizing the Doubling Rate

Shannon's work in communication and information theory is so remarkable because it gives an upper limit on the rates of information transmission, which is, channel capacity. Channel capacity is of such interest because it tells you the absolute best you can do. The analogue of rate in gambling may have an absolute best as well. A common measure of how good a certain gambling scheme is, is to look at how quickly wealth is doubled.

Let us set up a gambling scenario based around roulette and see how doubling rate is derived. In gambling, there is the odds of a certain outcome. These odds relate to the payout if that outcome occurs. In roulette, there is a one to one payout for betting on either red or black. This means that for each one dollar you bet, you will receive one dollar back, plus your initial investment. The percentage return on a bet on an outcome i is referred to as o_i. A gambler's total capital is limited (assuming they don't play on credit), and the percentage of their capital that the gambler is willing to wager on an outcome i is referred to as b_i. The probability of any outcome i occurring is referred to as p_i. The increase in a gamblers wealth after a race can be expressed as how much was bet, times the percentage return.

$$S(X) = b_X o_X$$

where in this case, the outcome of the race was X winning [2]. If we think back to our roulette analogy, say our gambler bet everything on red ($b_{red} = 1$), with a payout of 1:1 ($o_{red} = 2$), their increase in wealth would then be $S_{red} = 1 * 2$, a doubling of their wealth, assuming that red was the outcome. But this neglects the two other possible outcomes in this type of bet in roulette, the occurrence of black, and the occurrence of one of the "house edge" squares 0 and 00 (which are considered neither black or red, even or odd). It seems that if we want to get an accurate model of our doubling rate, we will need to take into account all possible outcomes. Thus, we define the doubling rate to be

$$W(\mathbf{b{,}p}) = E[\log(S(X))] = \sum_{k=1}^{m} p_k \log b_k o_k$$

as shown on page 160 in [2]. In our case of the roulette game, using the strategy of betting it all on red every time, gives us a doubling rate of

$$W(\mathbf{b{,}p}) = \frac{18}{38} \log_2(1 * 2) = \frac{18}{38} \approx 0.4737$$

Is this a good doubling rate? Is there a better betting strategy, **b**, we can use at a roulette table? What would the doubling rate look like if we put it all on one particular outcome, say my lucky number 3? In this case we would have the $p_3 = \frac{1}{38}$, and since we are betting it all, $b_3 = 1$, and the standard return on an American Roulette table for a "straight up" bet is 1:35, we get that $o_3 = 36$. Plugging these numbers into the equation, we get that our doubling rate is ≈ 0.1361. Obviously this is not as good a scheme as still betting it all on red. But, I'm feeling lucky, let's see what happens if we were to divide our bet over my two lucky numbers, 3 and 7. Here we would have $p_3 = p_7 = \frac{1}{38}$, but because we are splitting our total capital over two outcomes, we have $b_3 = b_7 = \frac{1}{2}$, we still have odds of 1:35 on both 3 and 7, so $o_3 = o_7 = 36$. When we calculate out new doubling rate we get $\approx .2195$. This is better than a bet on a single number, but still not as good as betting it all on red. Let's "let it ride" and see what happens if we keep increasing the number of unique bets we make.

We can see from Figure 1 that we can get a higher and higher doubling rate the more we spread out our total capital, with a maximum doubling rate achieved when we bet on 13 unique outcomes for the roulette wheel. Note that because the gambler is distributing all of their capital over a subset of all the possible outcomes, there is a chance that they will not win ever round. Thus, as the number of bets goes to infinity, the gamblers probability of going broke goes to 1. [3]

If you were to bet on all outcomes, and the odds were fair, no profit would be made because all winnings would be canceled out the cost of placing a bet on all outcomes. An interesting way around this is to use so called arbitrage schemes which require placing bets with more than one "bookie" [4]. Imagine that bookie A has fair odds for the outcome of some binary, event to be $o_{A,1} = \left(\frac{1}{2}\right)^{-1}, o_{A,2} = \left(\frac{1}{2}\right)^{-1}$ and bookie B has a different odds $o_{B,1} = \left(\frac{1}{3}\right)^{-1}, o_{B,2} = \left(\frac{2}{3}\right)^{-1}$. By making a 100 dollar bet with bookie A on event 2, and a 66.66 dollar bet with bookie B on event 1, you are guaranteed to win 200 dollars while only having to pay 166.66 dollars to play the game.

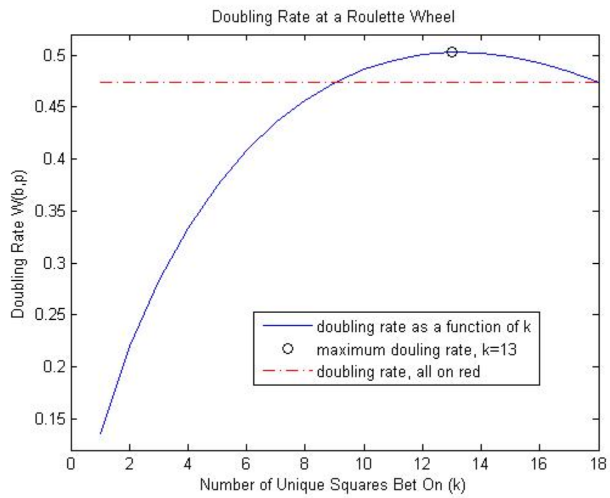

Fig. 1. Splitting the total capital among several outcomes, k, gives a better doubling rate, up to 13 unique bets.

There are two variables in the argument of the doubling rate function, **b** and **p**. Until now I have neglected to explain the meaning of these symbols. **b** refers to the betting strategy **b** being used to double the capital, and $W^*(\mathbf{p})$ refers to the optimal betting strategy.

B. *Kelly Gambling*

This implies that there is an optimal betting strategy. Kelly shows that the optimal betting strategy is related to the rate of transmission defined by Shannon.

When a gambler is making their bets, they need to come up with a betting strategy **b**. In the case of the roulette wheel, let's assume that the gambler makes their betting strategy by trying to observe a bias in the wheel due to mechanical error (the wheel is wobbly, the wood is warped, etc). In fact, let's say that they are somehow in communication with a computer that can observe the wheel as it is spun, use the wheel's bias to calculate the winning number and transmit binary message declaring if the outcome will be red or black (for the sake of this example, we also assume that the house edge squares of 0 and 00 have been removed so that $p_{red} = \frac{1}{2}$ and $p_{black} = \frac{1}{2}$). The gambler can then receive this message and make a bet before the dealer calls "no more bets."

If the computer can perfectly transmit its message to the gambler, and we assume this computer is never wrong, obviously the gambler should simply follow the computer's guidelines, and double their money with each bet. However, what if there is noise between the computer and the gambler? Then we must take into account the transition probabilities of the noise, $p(r|s)$, the probability that bit "r" was received given that bit "s" was sent. Kelly writes his growth rate as

$$G = \lim_{N\to\infty} \frac{1}{N} \log_2 \frac{V_N}{V_0}$$

where V_N is the gamblers capital after N bets and V_0 is the gamblers starting capital [3]. Kelly then writes that the growth rate can be written in more familiar terms as

$$G = \sum_{r,s} p(s,r) \log_2 \alpha_s a(s|r)$$

where $a(s|r)$ is fraction of the gamblers total wealth that will be bet on s (in our case either black or red) given that r was sent (in this case either a 1 or a 0), α_s is the payout for the bet that the computer tells the gambler to make (2:1 in both cases) and p(s,r) is the joint probability of s transmitted and r being received.

We assume, that we have fair odds, that is $\alpha_s = \frac{1}{p(s)}$, which we do if we remove the house edge, and then we can write the growth rate as

$$G = \sum_{r,s} p(s,r) \log_2 \frac{a(s|r)}{p(s)}$$

$$G = \sum_{r,s} p(s,r) \log_2 a(s|r) + H(X)$$

performing a maximizing on G, Kelly gets that

$$G_{max} = H(X) - H(X|Y)$$

where here X is referring to sent message and Y is referring to the received message. [3].

There is an interesting observation made by Cover and Erkip that shows that for each extra bit of information you use to describe the outcome, you get an extra bit of increase in the doubling rate [5]. In our roulette example, sending extra bits would tell us more about the outcome. For example if the first bit of a sequence could relate to the outcome being red or black, the second bit in the sequence could relate to being even or odd, and so on. Cover and Erkip do this in three steps; first they show that if we have a situation we are betting on, and we can send a rate of R bits to a gambler then the maximum increase in doubling rate, $\Delta(R)$ is equal to R. This is a fairly obvious use of the results proved by Kelly. They then go on to say that we imagine whatever is sending us information about the gamble notices something correlated to the outcome of the gamble (the wheel is wobbly, the wood is warped, etc.) This information is sent to the gambler and $\Delta(R)$ becomes limited by the mutual information between the observation and the gamble [5]. Finally they show that when the investor then has side information, S, they can use the side information along with the bits of data sent about V to increase their doubling rates bound to be the mutual information between $\tilde{V}$ and X given the side information, $I(\tilde{V}; X|S)$ where $\tilde{V}$ comes from the Markov chain $\tilde{V} \to V \to X$ [5], [6].

C. *Directed Information*

Now that we know that the Kelly Criterion is the optimal betting strategy, an obvious question arise: How much should you bet, using the Kelly betting scheme? As previously stated, wagering all of the capital available at each bet without using bet canceling bets will surely result in the gambler going broke. The tool that is developed to answer this question comes from Massey's directed information. Massey looked at how information flowed through a discrete memoryless channel and paid special attention to causal systems. Massey defines directed information to be

$$I(X^N \to Y^N) = \sum_{n=1}^{N} I(X^n; Y^n | Y^{n-1})$$

where X^N and Y^N are the input and output sequences, respectively [7]. He concludes in his paper

> ...when feedback is present, the directed information $I(X^N \to Y^N)$ gives a better upper bound on the information that the output sequence says about the input sequence than typical mutual information. [7]

This directed information is then built up by Permuter et al. when they showed that directed information is a better bound on growth rate than Kelly's bound of typical mutual information.

They start by looking at causally conditioned probability mass functions, which is defined as

$$p(x^n \widehat{||} y^{n-d}) \triangleq \prod_{i=1}^{n} p(x_i | x^{i-1}, y^{i-d})$$

from [8]. (Here I am using $\widehat{||}$ instead of the more traditional $||$ for causal conditioning so as to not create confusion with Kullback-Liebler distances which show up later in this paper.) They then apply this causal conditioning to causally conditional entropy and also create a new definition of growth rate based around the idea of limiting our side information about an outcome to that or previous outcomes and other side information that we obtained causally.

$$H(X^n \widehat{||} Y^n) \triangleq E\left[\log p(X^n \widehat{||} Y^n)\right] = \sum_{i=1}^{n} H(X_i | X^{i-1}, Y^i)$$

$$\frac{1}{n}W(X^n\widehat{||}Y^n) \triangleq \frac{1}{n}E\left[\log S(X^n\widehat{||}Y^n)\right]$$

where $S(X^n\widehat{||}Y^n)$ is the wealth after n races, with previous outcomes x^n and side information y^n available [9]. When dealing with fair odds, and uniformly distributed payoffs, (given $\mathcal{X} = \{1, 2...m\}$ horses racing, the payoff, o_i=$\frac{1}{m}$) they show that the maximum growth rate is equal to

$$\frac{1}{n}W(X^n\widehat{||}Y^n) = \log|\mathcal{X}| - \frac{1}{n}H(X^n\widehat{||}Y^n)$$

This shows that the maximum growth rate and the causally conditioned entropy are bound by a constant, which is related to the number of "horses in the race" [9]. This example assumed fair odds and a uniform distribution on the payoffs, so it should be pretty intuitive that in this case the gambler will be betting all their money. They then look at the two cases where the payoffs are not uniform, super fair ($\sum_x \frac{1}{o(x)} \leq 1$) and sub-fair ($\sum_x \frac{1}{o(x)} > 1$). It is fairly well known that in the case of super fair odds (which can be forced by schemes such as arbitrage) that betting all capital is the best approach [10]. They rely on an algorithm described by Kelly to show that under sub-fair odds, using conditional distributions, a solution to how much money should be bet can be found.

D. Risk

Whenever we are allocating resources where the return is not deterministic (and we are not dealing with super fair odds), we have a risk associated with our investment. It makes sense to say that the higher the risk involved in an investment, the higher the possible return will need to be to entice an investor to allocate their resources. This shows that there is a relationship between the risk of an investment and the maximum expected return. In cases where we know the distribution of the outcome, it is possible to create models for risk that are related to the mean, and more importantly, the variance of the distribution. [11].

There are times when we wish to allocate resources, but we do not know the distribution of the outcome, thus we cannot create a risk model using the variance of the distribution. The typical approach then, is to assume or create a distribution. We incur a risk in assuming a distribution for the outcome, but this must be done so that we can better judge how to allocate resources. We then define some "loss function" saying how wrong we were in our guess. Berger describes this by saying:

> If a particular actions a_1 is taken, and θ_1 turns out to the true state of nature, then a loss function $L(\theta_1, a_1)$ will be incurred. [12]

This loss function will in general vary from by situation. Several examples of loss functions are: square loss, $L(\theta_1, a_1) = (\theta_1 - a_1)^2$, absolute loss, $L(\theta_1, a_1) = |\theta_1 - a_1|$, and relative entropy loss, $L(\theta_1, a_1) = \sum \theta_1 \ln \frac{\theta_1}{a_1}$.[13] This relative entropy loss is nothing more than the same relative entropy defined by Claude Shannon. [14] A method described by Haussler and Opper shows that we can place reasonable bounds on the risk when we are using a relative entropy loss function. [15].

Let's say that we don't know the distribution of an outcome. One very logical attempt at determining what the distribution is would be to keep track of previous outputs to build an estimation. In theory, as the number of observations increases, we should get a better and better approximation of the real distribution. Of course, there will probably always be some error between the real and estimated distribution. There are several ways of representing this difference, and an often used method is the Kullback-Liebler distance, or relative entropy [2]. Kullback-Liebler distances do not satisfy triangle inequality [16], and can't be considered a true metric which somewhat limits their abilities when trying to make bounds. There are, however, other means for measuring the error between two distributions and one such distance that is related to Kullback-Liebler distance can be shown to be a true metric [17], [15]. Haussler et al.s approach is to look at the average Bayes risk of a scenario. (The Bayes risk is the "minimum average-case risk over all possible true distributions" of the outcome. [15])

Assume then, that there is a distribution picked by nature, at random, out of a set of distributions. The probability of a distribution, θ being picked is P_θ. This is what we want to estimate, and after n observations on θ, we have an estimator $\hat{P}_n$. Haussler et al. then show that the risk given after the Nth observation is

$$r_{N,\hat{P}}(\theta) = \int dP_\theta^{n-1} L(P_\theta, \hat{P}_N)$$

where, the loss function being used here is the Kullback-Liebler distance (relative entropy) between P_θ and $\hat{P}_N$. They then go on to show that the cumulative risk after N observations can be expressed as

$$R_{n,\hat{P}}(\theta) = \sum_{n=1}^{N} r_{n,\hat{P}}(\theta) = D(P_\theta^n || \hat{P})$$

Using the definition of Bayes Risk [12] and letting $\hat{P}$ be a Bayes Strategy, it is shown that

$$R_{n,\mu}^{Bayes} = \int_\Theta d\mu(\theta^*) D(P_{\theta^*}^n || M_{n,\mu}) = I(\Theta^*; Y^n)$$

where Θ is the total set of possible probabilities from which θ^* was chosen from using the strategy μ; and $M_{n,\mu}$ is the marginal distribution on Y^n, which is the sequences of observations[15]. Then by making a few assumptions, they show that there is a relation between Kullback-Liebler distances and a special case of the Hellinger distance, which can be viewed as a true metric [15], [18]. Using Laplace transforms on various distances measures (most importantly the Hellinger distance) they are able to achieve upper and lower bounds on the mutual information, $I(\Theta^*; Y^n)$, which is the Bayes risk.

III. Economics

Investing in markets is really no different than gambling. The goal is still to maximize some return of the allocation of resources on random variables. There are some differences though in how we must think about the problem. When one is gambling at a roulette wheel, one knows rather quickly whether they made a correct wager or not. When dealing with investing in markets however, the lag time between investment and payoff is greatly increased. Something interesting to note is that unlike roulette, where the game effectively ends once the ball has stopped moving, under normal circumstances, the investment in markets doesn't end until the investor decides to pull their money. This means that nothing is lost until the investor decides to lose it. Also when dealing with investing, one isn't as likely to talk about individual wagers, as they are to be concerned with the distribution of their wealth over a portfolio of investments.

Another key difference between gambles and investments, is that in investments, the probability of an outcome is either known (as in roulette) or if not known, can be inferred by the odds given by the track or bookie. The most popular method for determining the ratio of risk to reward for investments in the stock market is the Capital Asset Pricing Model (CAPM). Although there is controversy as to what factors should be included when calculating the CAPM [19] it gives us a good start to seeing how stock market investments relate to information theory.

The CAPM looks at the variance and mean of a possible investment and weighs it against a "market portfolio", which can contain a debatable number of factors [19], [2]. This gives us some idea of how to build a portfolio. This portfolio will have some distribution of stocks $F(\mathbf{x})$ that should in theory minimize the risk of the portfolio. We can now look at the doubling rate (previously explained in this paper and in [2]) for the portfolio as

$$W(\mathbf{b}, F) = \int \log_2 \mathbf{b}^T \mathbf{x} \, dF(\mathbf{X})$$

where $\mathbf{b}$ is the stock market portfolio we developed from the CAPM.

It should be pretty intuitive that side information will enhance our doubling rate (ask Martha Stewart about this), but how much of an increase in the doubling rate as a result of one more bit of information is not as obvious. Cover and Barron show that the doubling rate can increase by up to one bit, for each extra bit of information [20]. They do this by looking at the doubling rate defined above and make an assumption that wealth is optimally redistributed at the end of each day on the stocks in the portfolio, we have the optimal portfolio, $\mathbf{b}^*$. If you allow side information Y to become available, they show that

$$\Delta = E \log_2 \frac{\mathbf{b}^{*T(Y)} \mathbf{X}}{\mathbf{b}^{*T} \mathbf{X}}$$

$$I = E \log_2 \frac{f(\mathbf{X}, Y)}{f(\mathbf{X}) f(Y)}$$

where Δ is the increase in doubling rate due to the presence of some side information Y. They finally show that

$$0 \leq \Delta \leq I(\mathbf{X}; Y)$$

which shows that the increase in doubling rate cannot exceed the mutual information between the stocks and the side information, but also that doubling rate cannot be hurt by side information.

This is all very interesting to read about, but can one in practice actually come up with the optimal portfolio $\mathbf{b}^*$? Sadly no, but with a few assumptions, one can get very close [2], using the same approach that we laid out for the Kelly Criterion (giving the investor access to some side information about the stocks in their portfolio). Cover builds upon his own work from [21] that constantly redistributing wealth over a portfolio achieves an optimal growth rate under ideal conditions to create an algorithm that can very well mimic the properties of the "best state-constant rebalanced portfolio" [22]. The algorithm for so called universal portfolios shows a connection to universal data compression used in more traditional information theory applications. They show that the algorithm produces a better result in the presence of side information, than when there is no side information available. It should also be noted that Cover states that the algorithm begins to fail when ever there is a fee involved with the investing. In fact most of the work done in portfolio optimization fails to deal with commission. Because these strategies all tend to call for redistribution wealth constantly (at the end of each day rewarding the stocks that are less risky), a commission would be incurred everyday [22]. One questions whether or not any of these algorithms are feasible in the real world.

A. *Conclusion*

A joke about Martha Stewart and insider trading may not have been in the best taste, but it was there for a purpose. In order to get any edge in all the gambling schemes, some extra information besides the given outcomes is needed. There usually is some pretty hefty force preventing the gathering of side information. In casino games, side information may take the form of bias in a roulette wheel, or the cards already played in a blackjack game. Both of these are easily combated by the casino as roulette wheels are meticulously realigned, and counting cards will get you removed from the casino. When dealing with events like horse races, side information becomes more nebulous. What information is actually relevant to the outcome of the race? The horse's parents, the affinity of the animal to a certain weather condition, but how to you quantify this information?

In the world of stock markets, side information like insider trading is illegal. Strategies to create side information by view tendencies in the market become increasingly complicated the more factors they take into account. Even CAPM seems to fail in empirical tests [19]. Pricing models that use risk (such as CAPM) can vary wildly from model to model and a "good buy" under one model may be nothing special in another model [23]. Obviously if anyone developed a truly good strategy in gambling or investing, they wouldn't publish it in an academic paper as they would make it public knowledge and they would lose their edge over the rest of the gambling public.

Although we can achieve some bounds on effects of side information on a gamble, it seems that there should be more work done on how to quantify side information. Even Cover admits this in [22]. Because the equations for doubling rate are concerned with a percentage of a total amount of resource to be allocated, and an arbitrary payoff as a ratio to the resources allocated, without caring what the resource is, it can be used as a measure for all kinds of allocation scenarios. A limit to this use of the doubling rate is that a probability for the outcome must be known. If the probability of the outcome is not known, it can be estimated, and the allocator will simply have to be aware of the error they are incurring.

If we are using the doubling rate equation as a measure of how good a non-traditional gamble (neither stock market investing, nor "horse race" type gambles) there could be a lot of potential side information available, whose use would be acceptable for the situation. One example might be in marketing a product to the public. Where the resource being allocated would be capital in the form of advertisements, with the payoff being somehow linked to uses of the product, with a distribution being spread over various marketing demographics. Here side information about the demographics would be widely available and could lead to a model of using side information to increasing the strength of a marketing campaign.

The results from Figure 1 are some what interesting, but show how easily one can fall into the gambler's fallacy. Although the doubling function is a standard measure for how effective a gambling strategy is, it should be remembered it is an expectation function based on probability. If the gambler is betting it all, every time, it only takes one bad round to totally bankrupt them. It would be interesting to see if laws of typicality could be applied to the creation of a roulette system, although I would be incredibly dubious of any roulette system, made available to the public, that claimed to work more effectively than random chance.

The obvious casino games where one could really exploit the Kelly betting scheme would be card games where the deck is not reshuffled after each round, or where a player has a view of some of the other players cards. Blackjack and poker are the two that come to my mind, but as previously mentioned, because blackjack is played against the house, casinos reserve the right to remove you from the game if you become to successful in your attack. Also as previously stated, some work would need to be done by the schemes creator to properly quantify the side information so that it could be usable to the gambler.

REFERENCES

[1] Jerzy Konorski and Wojciech Szpankowski. What is information? 2008.
[2] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.
[3] J.L. Kelly. A new interpreation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.
[4] Donald B. Hausch and William T. Ziemba. Arbitrage strategies for cross-track betting on major horse races. *The Journal of Business*, 63:61–78, 1990.
[5] Thomas M. Cover and Elza Erkip. Information efficieny in investment.
[6] Thomas M. Cover and Elza Erkip, editors. *The relation of investment rate and description growth rate*, 1995.
[7] James L. Massey. Causality, feedback and directed information. *Proceedings 1990 International Symposium on Information Theory and its Applications*, 1990.
[8] G. Kramer. *Directed information for channels with feedback*. PhD thesis, Swiss Federal Insitute of Technology, 1998.
[9] Haim Permuter, Young-Han Kim, and Tsachy Weissman. On directed information and gambling, July 2008.
[10] John Haigh. The kelly criterion and bet comparisons in spread betting. *Journal of The Royal Statistical Society*, 49(4), 531–539.
[11] Anders Wilhelmsson. Value at risk with time varying variance, skewness and kurtosis - the nig-acd model. *Econometrics Journal*, 12:82–104, 2009.
[12] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 2 edition, 1985.
[13] Nicol Cesa-Binachi and Gbor Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.
[14] Claude E. Shannon and Warren Weaver. The mathematical theory of communication. *The Bell Systems Technical Journal*, 27:623–656, 1948.
[15] David Haussler and Manfred Opper. Mutual information, metric entropy and cumulative relative entropy risk. *The Annals of Statistics*, 25:2451–2492, 1997.
[16] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
[17] I. Csiszr. I-divergence geomety of probability distibutions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.
[18] Douglas G. Simpson. Hellinger deviance tests: Efficiency, breakdown points, and examples. *Journal of the American Statistical Association*, 84(405):107–113, 1989.
[19] Eugene F. Fama and Kenneth R. French. The capital asset pricing model: Theory and evidence. *The Journal of Economic Perspectives*, 18(3):25–46, 2004.
[20] Andrew R. Barron and Thomas M. Cover. A bound on the financial value of information. *IEEE Transactions on Information Theory*, 34(5), 1988.
[21] R. Bell and Thomas M. Cover. Competitive optimality of logarithmic investment,.
[22] Erik Ordentlich and Thomas M. Cover. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(4), 1996.
[23] Michael J Brennan and Yihong Xia. Asset pricing anomalies. *The Review of Financial Studies*, 14(2001):905–942, 2001.