Gambling with Information Theory

February 7, 2016

Coordinator:
Dr. C. (Christian) Schaffner

Student:
Govert Verkes
10211748

Course:
Information and Communication

Course number:
5062INCO6Y

Abstract

This paper explains a betting strategy also referred to as Kelly's criterion. It basically explains that when repeating a bet and reinvesting, it is better to bet according to the probability distribution and not let your bet depend on the odds. This paper also explains the value of side information, and its relation with the information theory. This paper will also introduce the notion of stochastic processes and their entropy rate and how these notions are related to betting where a new outcome depends on previous outcomes.

1 Introduction

As long as people gamble, they try to develop strategies that make them win, may it be by means of superficial, psychological, or more mathematical strategies. One of these mathematical strategies was the theory described by J. Kelly in 1956 [2]. Kelly never used it to gamble himself, but his strategy was used by his famous associate Shannon, the father of information theory. Shannon and his wife went to Las Vegas and made a lot of money using Kelly's findings. Even nowadays his findings are still used by very successful investors like Warren Buffett [3].

This paper explains the findings from Kelly's original paper, which shows what the optimal betting strategy is for a bet with fair odds. What makes Kelly's findings interesting is that they go strongly against one's intuition. Where one's intuition would say that betting depends on the odds, Kelly proved that when there is no betting fee and when one repeats the bet a lot of times, one has to bet proportional to the probability distribution of the outcomes.

This paper also briefly explains the notion of *stochastic processes* and their *entropy rate* and the relation between the entropy rate and *dependent* bets, i.e. bets that depend on previous outcomes.

2 Gambler with private channel

Suppose we have a gambler with a private channel, it is a channel over which the results, for example of a sports match, would be transmitted to the gambler before the game started, so that the gambler can still bet at original odds. This section mostly follows the explanation of Kelly's original paper [2].

2.1 Binary private channel

Let us say the gambler has a noiseless channel, i.e. the results transmitted are always succesfully transmitted. Moreover the odds for the sport match are equal, i.e. both teams have 2-for-1 odds. In this case the gambler will bet as much money as he can, because he is certain that he will get his money doubled since it is a noiseless channel. Suppose the gambler bets on N successive sport matches, then his begin capital will grow with a factor 2^N. So if V_0 is the gambler's starting capital then his capital after N races will be, $V_N = 2^N V_0$. Before going further we define the exponential rate of growth, to compute the gambler's exponential profit.

Definition 2.1. (Exponential rate of growth)

$$G = \lim_{N\to\infty} \frac{1}{N} \log \frac{V_N}{V_0} \tag{2.1}$$

The exponential rate of growth tells how much the gambler's capital grows exponentially with respect to N, i.e. $V_N = 2^{NG}$. So for the gambler with a noiseless channel $G = 1$. Now suppose the gambler has a noisy channel, again the results are transmitted to the gambler while he can still bet with the original odds, but this time the result is only correctly transmitted with a probability p and incorrectly transmitted with a probability q. If the gambler bets on the transmitted result his expected capital after N races will grow proportional to p (assuming p is greater than q).

$$\mathbb{E}[V_N] = (2p)^2 V_0 \tag{2.2}$$

However, even though this maximizes the gambler's capital after N races, he will lose all of his money with probability q every bet, and if he continues indefinitely the probability that he loses all his money becomes one. Let us see what happens if the gambler bets only a fraction of his capital, say ℓ. In that case his capital after N races would be,

$$V_N = (1+\ell)^W (1-\ell)^L V_0 \tag{2.3}$$

Where W is the number of time the gamblers wins and L the number of times he loses. Therefore the exponential rate of growth is,

$$\begin{aligned} G &= \lim_{N\to\infty} \frac{1}{N} \log(1+\ell)^W (1-\ell)^L && (2.4)\\ &= \lim_{N\to\infty} \frac{W}{N} \log(1+\ell) + \frac{L}{N} \log(1+\ell) && (2.5)\\ &= p\log(1+\ell) + q\log(1-\ell) && (2.6) \end{aligned}$$

In order to maximize the gambler's capital after N bets, we need to maximize G with respect to ℓ. Any maximization problem like 2.6 can be solved using lemma 2.1.

Lemma 2.1. The function $H_{\max} = \sum x_i \log y_i$ subject to the constraint $\sum y_i = Y$ is maximized by

$$y_i = \frac{Y}{X} x_i \tag{2.7}$$

Where $X = \sum x_i$.

Proof. To prove it we add Lagrange multipliers and change the base of the log, so we get the following function,

$$J(\mathbf{y}) = \sum x_i \ln y_i + \lambda \sum y_i \tag{2.8}$$

Next up we take the derivative,

$$\frac{\delta J}{\delta y_i} = \frac{x_i}{y_i} + \lambda \tag{2.9}$$

And set it equal to zero,

$$y_i = -\frac{x_i}{\lambda} \tag{2.10}$$

We can calculate λ using the sums X and Y,

$$\lambda = -\frac{\sum x_i}{\sum y_i} = -\frac{X}{Y} \tag{2.11}$$

Using 2.11 and 2.10 we get,

$$y_i = \frac{Y}{X} x_i \tag{2.12}$$

□

Now using 2.7 we can find ℓ that maximizes the exponential rate of growth, namely,

$$\begin{aligned} 2p &= 1 + \ell \\ 2q &= 1 - \ell \end{aligned} \tag{2.13}$$

which gives us,

$$\begin{aligned} G_{\max} &= \max_{\ell} p \log(1+\ell) + q \log(1-\ell) \\ &= 1 + p \log p + q \log q \\ &= 1 - H(X) \end{aligned} \tag{2.14}$$

Where $H(X)$ is the entropy defined by Shannon [4]. Note that using this strategy, i.e. using the ℓ from 2.13, does not maximize the expected value of the gambler's capital after N bets. The expected value would still be maximized by betting his whole capital on the outcome with the highest expected winnings, and for small N it might even be smarter to bet the full capital. However when the gambler is playing for an indefinite amount of time, he will lose his full capital with probability one.

2.2 General private channel

In the previous section we looked at a gambler with a binary private channel. Now let us suppose the gambler has a more general private channel where symbols $x \in X$ are transmitted and symbols $y \in Y$ are received by the gambler. The following definitions will be used:

$p(x)$	probability transmitted symbol is x
$p(y)$	probability received symbol is y
$p(x\|y)$	conditional probability that transmitted symbol is x given the received symbol is y
$p(y\|x)$	conditional probability that received symbol is y given the transmitted symbol is x
$p(x,y)$	joint probability of x and y
o_x	odds paid on the occurrence of transmitted symbol x (o_x-for-1)
$b(x\|y)$	fraction of the gambler's capital that he decides to bet on the occurrence of the transmitted symbol s after receiving r

2.2.1 Fair odds

Let us look at the simple case where there is no betting fee,

$$\sum \frac{1}{o_x} = 1 \tag{2.15}$$

and the odds are fair proportional to the distribution $p(x)$,

$$o_x = \frac{1}{p(x)} \tag{2.16}$$

Because there is no betting fee, we can also assume the gambler bets his full capital,

$$\sum_x b(x|y) = 1 \tag{2.17}$$

Using this notation we can define the gambler's capital after N bets as follows,

$$V_N = \prod_{x,y} [b(x|y)o_x]^{W_{xy}} V_0 \tag{2.18}$$

And so the exponential rate of growth (eq. 2.1) is,

$$G = \lim_{N\to\infty} \sum_{x,y} \frac{W_{xy}}{N} \log b(x|y) o_x \tag{2.19}$$

$$= \sum_{x,y} p(x,y) \log b(x|y) o_x \tag{2.20}$$

$$\stackrel{eq.\ 2.15}{=} \sum_{x,y} p(x,y) \log \frac{b(x|y)}{p(x)} \tag{2.21}$$

$$= \sum_{x,y} p(x,y) \log b(x|y) + H(X) \tag{2.22}$$

Again we want to maximize G with respect to $b(x|y)$. This can be done using equation 2.7,

$$G_{\max} = \max_{b(x|y)} \sum_{x,y} p(x,y) \log b(x|y) + H(X) \tag{2.23}$$

$$\overset{eq.\ 2.7}{=} \sum_{x,y} p(x,y) \log p(x|y) + H(X) \tag{2.24}$$

$$= H(X) - H(X|Y) \tag{2.25}$$

$$= I(X;Y) \tag{2.26}$$

Thus, we see from 2.23 that the maximum exponential rate of growth for betting proportional to the probability distribution, whilst having side information ($y \in Y$) is equal to the mutual information between X and Y. What becomes clear is that when the gambler does not have side information $G_{\max} = 0$. Therefore we can conclude that if the odds are fair, the only way the gambler could make a profit is by having side information. Without side information the growth rate becomes one and the gambler's capital stays steady.

2.2.2 Unfair odds

Suppose there is still no betting fee. Thus,

$$\sum \frac{1}{o_x} = 1 \tag{2.27}$$

But the odds are not fair with respect tot the probability distribution. Thus it is not necessarily true that,

$$o_x = \frac{1}{p_x} \tag{2.28}$$

Let us define $r_x = \frac{1}{o_x}$, then equation 2.20 becomes,

$$G = \sum_{x,y} p(x,y) \log \frac{b(x|y)}{r_x} \tag{2.29}$$

$$= \sum_{x,y} p(x,y) \log \frac{b(x|y)}{p(x|y)} \frac{p(x|y)}{r_x} \tag{2.30}$$

$$= D(p\|r) - D(p\|b) \tag{2.31}$$

Where $D(p\|q)$ is the relative entropy. The exponential rate of growth is again maximized by taking $\mathbf{b} = \mathbf{p}$. if we assume the mutual information between X and Y is zero, i.e. $b(x|y) = b(x)$ and $p(x|y) = p(x)$, then 2.31 shows that the exponential rate growth is equal to how much better the gambler's guess of the real probability is compared to the bookmaker's guess of the real probability (r_x).

Example 2.1. (Dice rolling with unfair odds) Let us look at gambling with a fair dice with 6 sides $X = \{1,2,3,4,5,6\}$, where the probability for each side is $p(x) = \frac{1}{6}$ and without side information. The odds for each side are,

$$o_x = \begin{cases} 6 & x \in \{1,2,3,4\} \\ 12 & x = 5 \\ 4 & x = 6 \end{cases} \tag{2.32}$$

$$\tag{2.33}$$

Note that $\sum \frac{1}{o_x} = 1$ still holds. One's intuition might tell you that one should put more money on $x = 5$ and less money on $x = 6$, since $x = 5$ gives 12-for-1 odds and $x = 6$ only 4-for-1. However Kelly's findings tell us that we should bet according to the distribution $p(x)$ as long as $\sum \frac{1}{o_x} = 1$. Therefore we should bet $\frac{1}{6}$ of our wealth on every x. Suppose another gambler "Franky" is using a more intuitive strategy and bets $\frac{1}{6}$ of his capital for $x \in \{1,2,3,4\}$, but $\frac{3}{12}$ for $x = 5$ and $\frac{1}{12}$ for $x = 6$. Then using the formula for exponential rate of growth we get,

$$G_{\text{Kelly}} = \frac{4}{6}\log(\frac{1}{6}6) + \frac{1}{6}\log(\frac{1}{6}12) + \frac{1}{6}\log(\frac{1}{6}4) = 0.069 \tag{2.34}$$

$$G_{\text{Franky}} = \frac{4}{6}\log(\frac{1}{6}6) + \frac{1}{6}\log(\frac{3}{12}12) + \frac{1}{12}\log(\frac{1}{12}4) = 0 \tag{2.35}$$

We can actually model this using a little program (see appendix A) and using the odds described above. And indeed using Kelly's strategy one's capital grows and the other one shrinks. However for a low numbers of games Kelly's strategy does not always work, but if you play a lot of games it does.

Why it goes against our intuition might be because we focus on the maximum amount of winning when we bet more $x = 5$, and overlook the the fact that we will also lose more when $x = 6$.

3 Entropy rate and dependent bets

Bets depending on previous results is a very common example of side information, for example for sport matches where previous results definitely tells us something about future outcomes. Before looking into such dependent bets, let us define a stochastic process and its entropy rate. Most of this section follows chapter 4 from Cover and Thomas [1].

3.1 Stochastic process

A stochastic process is a collection of random variables that evolve, typically over time. Stochastic processes are used to model various phenomena in which quantities vary over time, in a random fashion. They are used in the fields of engineering, economics, etc.

Definition 3.1. (Discrete stochastic process) A discrete stochastic process is a set of random variables, $\{X_t\}_{t\in\mathcal{T}}$ with $\mathcal{T} = \mathbb{N}$ characterized by their joint probability,

$$\Pr(X_1, X_2, \ldots, X_n) \tag{3.1}$$

There are a couple of important properties that a stochastic process can have,

Definition 3.2. (Markov process) A stochastic process is said to be a Markov process when, for $n = 1, 2, \ldots,$

$$\Pr(X_{n+1} \mid X_1, X_2, \ldots, X_n) = \Pr(X_{n+1} \mid X_n) \tag{3.2}$$

i.e. when a new random variable is only dependent on the previous random value.

Definition 3.3. (Stationary process) A stochastic process is said to be a stationary process when, for $n = 1, 2, \ldots$ and for every shift t and every x_i,

$$\begin{aligned}\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)\\ = \Pr(X_{1+t} = x_1, X_{2+t} = x_2, \ldots, X_{n+t} = x_n)\end{aligned} \tag{3.3}$$

i.e. when the probability distribution is invariant to shifts in time.

Example 3.1. (Simple random walk) Suppose we have a stochastic process defined by the following simple random walk,

$$\begin{aligned} Y &= \begin{cases} 1 & \text{with Pr: } \frac{1}{2} \\ -1 & \text{with Pr: } \frac{1}{2} \end{cases} \\ X_n &= \sum_{i=1}^{n} Y_i \end{aligned} \tag{3.4}$$

This stochastic process is a Markov process because a new random variable is only dependent on it predecessor, namely $X_{n+1} = X_n + Y_{i+1}$. However this stochastic process is not stationary,

$$\begin{aligned} &\Pr(X_1 = 2) = 0 \\ &\Pr(X_2 = 2) = \frac{1}{4} \\ &\Pr(X_1 = 2) \neq \Pr(X_2 = 2) \end{aligned} \tag{3.5}$$

And therefore it does not follow equation 3.3.

3.2 Entropy rate

Entropy rate is defined to give the rate in which the entropy grows for a stochastic process. In other words how much the uncertainty grows through time.

Definition 3.4. (Entropy rate) The entropy rate for a process is defined by,

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots X_n) \tag{3.6}$$

Example 3.2. Suppose X is identically independently distributed, then the entropy for every random variable in X has the same entropy. Therefore the entropy rate is equal to the entropy of one random variable,

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= nH(X_1) \\ H(\mathcal{X}) &= H(X_1) \end{aligned} \tag{3.7}$$

There is also a related quantity,

Definition 3.5. (Related quantity for entropy rate)

$$H(\mathcal{X})' = \lim_{n\to\infty} \frac{1}{n} H(X_n | X_1, X_2, \ldots X_n - 1) \tag{3.8}$$

This is the conditional entropy of the last randomX variable given the past.

Now it can be proven that for stationary stochastic processes the related quantity and the entropy rate both exist and are equal,

Theorem 3.1. For a stationary process (eq. 3.3) both the entropy rate and the related quantity for entropy rate exist and are equal,

$$H(\mathcal{X}) = H(\mathcal{X})' \tag{3.9}$$

3.3 Dependent bets

Dependent bets are bets were we use previosu outcomes to determine the strategy for future betting. Examples of dependent bets can be seen in betting on sport matches or investing in stock markets. For example a team that wins more is more likely to win again. A stock that fluctuates is likely to fluctuate in the future. Thus, we can use these previous results in our betting strategy. Suppose the matches are a stochastic process where the sequence $\{X_i\}$ consists of all the match results, and that the odds are uniform, i.e. for a match between m teams the odds are m-for-1. Because the gambler's bet depends on previous results we can define $b(X_n|X_1, X_2, \ldots, X_{n-1})$. Then using the definition (eq. 2.1) the exponential rate of growth is,

$$\begin{aligned} G_{\max} &= \max_b \sum_x p(X_1, X_2, \ldots, X_n) \log(b(X_n|X_{n-1}, X_{n-2}, \ldots, X_1)m) \\ &= \log m - H(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) \end{aligned} \tag{3.10}$$

if $\{X_i\}$ is a stationary process, then using equation 3.9 we can rewrite 3.10,

$$G_{\max} + H(\mathcal{X}) = \log m \tag{3.11}$$

This shows that the exponential rate of growth plus the entropy rate of the stochastic process $\{X_i\}$ is a constant. Therefore if the entropy rate for a stochastic process is very low the maximum exponential rate of growth becomes larger.

4 Conclusion

It is very interesting how Kelly's findings show that for fair odds, the best strategy in the long run is to bet proportional to the probability distribution. So even when the odds are very good for one outcome, you should not bet disproportionally on this outcome. This can be understood by the fact that in the long when you do bet disproportional amount on one specific outcome you also lose more in other outcomes. Note that Kelly's strategy only applies when you reinvest your money in repeated bets. For example when one goes to the casino every day with the plan to spend just 10 euro on one bet, then it is better to bet on the outcome with the highest expected winning, instead of using Kelly's strategy.

Another aspect for gambling that has been addresses is the betting using information of previous outcomes. When betting on sport matches or investing on the stock market, it is rational to base your strategy on previous results. For bets that depend on previous results, it can be shown that the exponential rate of growth plus the entropy rate for the results is a constant. Therefore if the lower the entropy rate for a stochastic process is, the larger the maximum exponential rate of growth. This is a intuitive conclusion if you consider that the entropy rate defines the amount of uncertainty in a stochastic process.

References

[1] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[2] J. Kelly. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3):185–189, 1956.

[3] M. Pabrai. *The Dhandho Investor: The Low-Risk Value Method to High Returns*. Wiley, 2011.

[4] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Appendices

A Similation in C++ of rolling a dice with unfair odds

```
#include <cstdio>
#include <cstdlib>
#include <time.h>

int main() {
    srand(time(NULL));

    float kelly_capital = 1.0;
    float normal_capital = 1.0;

    // The odds-for-1
    float odds[6] = {6, 6, 6, 6, 12, 4};

    // Bets following Kelly's findings
    float kelly_bets[6] = {1/6., 1/6., 1/6., 1/6., 1/6., 1/6.};

    // Bets not following Kelly's findings
    float normal_bets[6] = {1/6., 1/6., 1/6., 1/6., 3/12., 1/12.};

    // Simulate 100 dice rolls
    for (int i = 0; i < 100; i++) {

        // Simulate dice roll 1-6
        int r = rand() % 6;

        kelly_capital *= odds[r] * kelly_bets[r];
        normal_capital *= odds[r] * normal_bets[r];

    }

    printf("Kelly:%f vs. Normal:%f\n", kelly_capital, normal_capital);
}
```