Gambling And Entropy

Outline

- There is a strong relationship between the **growth rate** of investment in a horse race and the **entropy** of the horse race.
- The value of **side information** is related to the **mutual information** from the perspective of growth rate.
- We can use a (gambling) game to estimate the entropy of English.

The Horse Race

- p_i, the **probability** that horse i wins the race.
- o_i, the **payoff** for 1 dollars bet on horse i if it wins.
  - a-for-1 vs. r-to-1, equivalent if $r = a - 1$.
- b_i, the **proportion** of money bet on horse i.
- $S(X) = b(X)o(X)$ is the **multiplication factor** by which the gambler's wealth is increased when horse X wins a race.
- $S_n = \prod S(X_i)$ is the multiplication factor of the gambler's wealth **after** n **races**.

The Doubling Rate

- The **doubling rate** of a horse race is defined by

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

Here the o_i is given (decided by the **bookies**).

- Let the horse race outcomes be i.i.d. random variables. Then the wealth of gambler increases exponentially

$$S_n \doteq 2^{nW(b,\, p)},$$

since $\frac{1}{n} \log S_n = \frac{1}{n} \sum_i \log S(X_i) \rightarrow E \log S(X).$

The Optimal Doubling Rate

- Given p and o, the **optimal doubling rate** is the maximum doubling rate over the choice of b.

$$W^*(p) = \max_b W(b, p).$$

- The optimal doubling rate is

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

achieved by the **proportional gambling** $b^* = p$.

- Note that entropy and optimal rate is **complementary**.

Proof

$$
\begin{aligned}
W(b,p) &= \sum p_i \log b_i o_i \\
&= \sum p_i \log \frac{b_i}{p_i} p_i o_i \\
&= \sum p_i \log o_i - H(p) - D(p||b) \\
&\leq \sum p_i \log o_i - H(p),
\end{aligned}
$$

with equality iff $p = b$. See Example 6.1.1 for illustration.

Note: While it is tentative to bet all money on horse i with maximum $p_i o_i$, this strategy has the risk of losing all money.

Fair Odds

- Odds are **fair** if $\sum_i \frac{1}{o_i} = 1$. Betting according to $b_i = \frac{1}{o_i}$ guarantees money back, no more and no less.
- Suppose the odd o is fair. Define the probability $r_i = \frac{1}{o_i}$. For betting b,

$$\begin{aligned} W(b, p) &= \sum p_i \log b_i o_i = \sum p_i \log \frac{b_i}{p_i} \frac{p_i}{r_i} \\ &= D(p||r) - D(p||b). \end{aligned}$$

  Therefore, betting is a **competition** between the gambler's and the bookie's estimates of p .

Uniform Fair Odds

- **Uniform fair odds** are m-for-1 on each horse.
- For uniform fair odds,

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

The sum does *not* depend on p!

Superfair and Subfair Odds

- Odds are **superfair** if $\sum \frac{1}{o_i} < 1$.
  - In this case, one can adopt a "**Dutch book**" strategy, betting $b_i = \frac{1}{o_i}$ and getting $1 + (1 - \sum \frac{1}{o_i}) > 1$ in return. While risk-free, this does not optimize the doubling rate.
- Odds are **subfair** if $\sum \frac{1}{o_i} > 1$.
  - This is the real-life scenario. The organizer takes a cut of all bets. If you have to bet, it's probably better to leave some cash for cab.

Side Information

- Suppose the gambler has some information relevant to the outcome of the race. Does such information help to gambler to win more money?
- Let X be the outcome of race, Y be the information.
- Without Y, the optimal doubling rate is

$$W^*(X) = \max_{b(x)} \sum_x p(x) \log b(x) o(x)$$
$$= \sum_x p(x) \log o(x) - H(X),$$

since we know that proportional gambling is optimal.

Side Information

- With Y, the multiplication factor of a race is changed to

$$S(X|Y) = b(X|Y)o(X)$$

- So the optimal doubling rate is

$$\begin{aligned} W^*(X|Y) &= \max_{b(x|y)} E[\log S] \\ &= \max_{b(x|y)} \sum_{x,y} p(x,y) \log b(x|y)o(x) \\ &= \sum_{y} p(y) \max_{b(x|y)} \sum_{x} p(x|y) \log b(x|y)o(x). \end{aligned}$$

Value of Side Information

- Again the optimal bet is proportional to the conditional probability of x given y, for each y. So the optimal doubling rate is

$$
\begin{aligned}
W^*(X|Y) &= \sum_{x,y} p(x,y) \log p(x|y) o(x) \\
&= \sum_{x} p(x) \log o(x) - H(X|Y).
\end{aligned}
$$

- So the optimal doubling rate increases with side information Y by

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

Red and Black

- Consider the game of betting on the color of the next card of a deck of 26 blacks and 26 reds.
- One can bet sequentially, based on the conditional probability of black and red given the history of cards.
- One can also bet the entire sequence at once. Betting $\frac{1}{C_{26}^{52}}$ on each of the C_{26}^{52} possible output sequences, which are equally likely.
- These two schemes are equivalent.
- The resulting wealth is 9.08.

Simulation of English Text

- Choose the **alphabet**. Use a book.
- For the **first** token, open the book to a random page and choose a random **token** on this page.
- For the **second** token, start at a random position of a random page until the first token is encountered. Record the token next.
- For the **third** token, again start at a random position of another random page until the second token is encountered. Record the token next.
- This approximates English with the **first-order Markov model**. Examples in pp. 134-135.

The Entropy of Simulated English

- As the order becomes higher, the model gets closer to the true English.
  - The entropy of the zero-th order model is $\log 27 = 4.76$ bits.
  - The first order: 4.03 bits.
  - The fourth order: 2.8 bits.
  - What is the limit?
- The dependency between tokens can be backward or long-range. However, the N-gram (Markov) word models do quite well in some applications.

The Entropy of English: Shannon Game

- A human subject is given a sample of English text and asked to **guess** the next letter.
- The experimenter records the **number of guesses** required to figure out the next letter.
- Assume that the subject can be modeled as a computer making **deterministic** choice of guesses based on the history. Then if we have this machine and the numbers of each guess, we can reconstruct the English text, so the entropy of the guess sequence is the entropy of English text.
- The entropy of English is estimated to be 1.3 bits.

The Entropy of English: Gambling Estimate

- A human subject **gambles** on the next letter. The payoff is 27-to-1 for all letters and the space.
- Since sequential betting is equivalent to betting on the entire sequence, the **payoff** after n letters are

$$S_n(X_1, \ldots, X_n) = (27)^n b(X_1, \ldots, X_n)$$

- The expected log wealth satisfies

$$\begin{aligned} E\,\frac{1}{n}\log S_n &= \log 27 + \frac{1}{n} E \log b(X_1, \ldots, X_n) \\ &= \log 27 + \frac{1}{n}\sum p(x^n)\log b(x^n) \le \log 27 - H(\mathcal{X}). \end{aligned}$$

- $\hat{H} = \log 27 - \frac{1}{n}\log S_n$ is an estimate (1.34 bits).