# Dynamical systems, information and time series

Stefano Marmi
Scuola Normale Superiore
http://homepage.sns.it/marmi/

Lecture 3- European University Institute

October 9, 2009

- Lecture 1: An introduction to dynamical systems and to time series. Periodic and quasiperiodic motions. (Tue Jan 13, 2 pm - 4 pm Aula Bianchi)
- Lecture 2: A priori probability vs. statistics: ergodicity, uniform distribution of orbits. The analysis of return times. Kac inequality. Mixing (Sep 25)
- Lecture 3: Shannon and Kolmogorov-Sinai entropy. Randomness and deterministic chaos. Relative entropy and Kelly's betting. (Oct 9)
- Lecture 4: Time series analysis and embedology: can we distinguish deterministic chaos in a noisy environment? (**Tuesday, Oct 27, 11am-1pm**)
- Lecture 5: Fractals and multifractals. (Nov 6)

Today's references:

- Benjamin Weiss: "Single Orbit Dynamics", AMS 2000.
- Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379–423; 623–656.
- Kelly Jr., J. L., 1956: A new interpretation of information rate. Bell Sys. Tech. J., 35 (4)
- L. Breiman "Optimal Gambling Systems for Favorable Games" (1961)
- W. Poundstone "Fortune's Formula: The Untold Story of the Scientific Betting System That Beat the Casinos and Wall Street" Hill and Wang, New York, 2005, 400 pages
- Thorp, E. O., 2006: Handbook of Asset and Liability Management:Theory and Methodology, Vol. 1, chap. 9 *The Kelly criterion in blackjack, sports betting, and the stock market*. Elsevier.

The slides of all lectures will be available at my personal webpage: http://homepage.sns.it/marmi/

# An overview of today’s lecture

- Entropy
- Kolmogorov-Sinai entropy
- Bernoulli schemes and topological Markov chains
- Risk management and information theory: the Kelly criterion

# Ergodic theory

The focus of the analysis is mainly on the *asymptotic ditribution of the orbits*, and not on transient phenomena.

Ergodic theory is an attempt to study the *statistical behaviour of orbits* of dynamical systems restricting the attention to their asymptotic distribution.

One waits until all transients have been wiped off and looks for an *invariant probability measure describing the distribution of typical orbits*.

# Ergodic theory: the setup (measure preserving transformations, stationary stochastic process)

X phase space, $\mu$ probability measure on X

$\Phi : X \to \mathbf{R}$ **observable**, $\mu(\Phi) = \int_X \Phi \, d\mu$ **expectation value of** $\Phi$

A measurable subset of X (**event**). A dynamics $T : X \to X$ induces a time evolution:

on observables $\Phi \to \Phi \, T$

on events $A \to T^{-1}(A)$

T **is measure-preserving** if $\mu(\Phi) = \mu(\Phi \, T)$ for all $\Phi$, equivalently $\mu(A) = \mu(T^{-1}(A))$ for all A

# Entropy

In probability theory, *entropy* quantifies the uncertainty associated to a random process

Consider an experiment with mutually esclusive outcomes $A=\{a_{1,} \ldots, a_k\}$

- Assume that the probability of $a_i$ is $p_i$ , $0 \leq p_i \leq 1$, $p_1 + \ldots + p_k = 1$
- If $a_1$ has a probability very close to 1, then in most experiments the outcome would be $a_1$ thus the result is not very uncertain. One doea not gain much information from performing the experiment.
- One can quantify the "surprise" of the outcome as

$$\text{information} = -\log(\text{probability})$$

- (the intensity of a perception is proportional to the logarithm of the intensity of the stimulus)

Suppose that one performs an experiment which we will denote $\alpha$ which has $m \in \mathbb{N}$ possible mutually esclusive outcomes $A_1, \ldots, A_m$ (e.g. throwing a coin $m = 2$ or a dice $m = 6$). Assume that each possible outcome $A_i$ happens with a probability $p_i \in [0,1]$, $\sum_{i=1}^m p_i = 1$ (in an experimental situation the probability will be defined statistically). In a probability space $(X, \mathcal{A}, \mu)$ this corresponds to the following setting : $\alpha$ is a finite *partition* $X = A_1 \cup \ldots \cup A_m \bmod(0)$, $A_i \in \mathcal{A}$, $\mu(A_i \cap A_j) = 0$, $\mu(A_i) = p_i$.

Returning to our "experiment", we define on $X$ a function $I(\alpha)$ called *information relative to the partition* $\alpha$ which, evaluated at the point $x$, expresses the amount of information we get from the knowledge of the element $A_i$ of $\alpha$ to which $x$ belongs. It is natural to ask that $I$ depends only on the probability of $A_i$ so that $I(\alpha) = \sum_{k=1}^m \phi(p_i)\chi_{A_i}$ for some function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$; it is natural to require that $\phi$ is *decreasing* since the information is bigger if we can locate $x$ in a smaller set. Finally we assume that, if $\alpha$ and $\beta$ are *independent*, then the information gained from the knowledge of the position of $x$ with respect to both partitions is obtained summing the information relative to each partition : $I(\alpha \vee \beta) = I(\alpha) + I(\beta)$ . To fulfill this last requirement on $\phi$ we must impose that $\phi(ab) = \phi(a) + \phi(b)$ $\forall a, b \in (0,1)$. It is then clear that $\phi(t)$ must be a constant multiple of $-\log t$.

# Entropy

The entropy associated to the experiment is

$H=-\sum p_i \log p_i$

Since

information = - Log (probability)

*entropy is simply the expectation value of the information produced by the experiment*

# Uniqueness of entropy

$$\Delta^{(m)} = \{(x_1, \dots, x_m) \in \mathbb{R}^m \mid x_i \in [0,1],\ \sum_{i=1}^{m} x_i = 1\}$$

**Definition 4.15** *A continuous function* $H^{(m)} : \Delta^{(m)} \to [0, +\infty]$ *is called an* entropy *if it has the following properties :*

*(1) symmetry :* $\forall\, i, j \in \{1, \dots, m\}$ $H^{(m)}(p_1, \dots, p_i, \dots, p_j, \dots, p_m) = H(p_1, \dots, p_j, \dots, p_i, \dots, p_m)$ ;

*(2)* $H^{(m)}(1, 0, \dots, 0) = 0$ ;

*(3)* $H^{(m)}(0, p_2, \dots, p_m) = H^{(m-1)}(p_2, \dots, p_m)$ $\forall\ m \geq 2,\ \forall\ (p_2, \dots, p_m) \in \Delta^{(m-1)}$ ;

*(4)* $\forall\ (p_1, \dots, p_m) \in \Delta^{(m)}$ *one has* $H^{(m)}(p_1, \dots, p_m) \leq H^{(m)}\left(\frac{1}{m}, \dots, \frac{1}{m}\right)$ *where equality is possible if and only if* $p_i = \frac{1}{m}$ *for all* $i = 1, \dots, m$ ;

*(5) Let* $(\pi_{11}, \dots, \pi_{1l}, \pi_{21}, \dots, \pi_{2l}, \dots, \pi_{m1}, \dots, \pi_{ml}) \in \Delta^{(ml)}$ *; for all* $(p_1, \dots, p_m) \in \Delta^{(m)}$ *one must have*

$$H^{(ml)}(\pi_{1l}, \dots, \pi_{1l}, \pi_{21}, \dots, \pi_{ml}) = H^{(m)}(p_1, \dots, p_m) + \sum_{i=1}^{m} p_i H^{(l)}\left(\frac{\pi_{i1}}{p_i}, \dots, \frac{\pi_{il}}{p_i}\right).$$

**Theorem 4.16** *An entropy is necessarily a positive multiple of*

$$H(p_1, \dots, p_m) = -\sum_{i=1}^{m} p_i \log p_i\ .$$

# Entropy, coding and data compression

What does entropy measure?

Entropy quantifies the information content (namely the amount of randomness of a signal)

Entropy : a completely random binary sequence has entropy= $\log_2 2 = 1$ and cannot be compressed

Computer file= infinitely long binary sequence

Entropy = *best possible compression ratio*

Lempel-Ziv algorithm (Compression of individual sequences via variable rate coding, IEEE Trans. Inf. Th. 24 (1978) 530-536): does not assume knowledge of probability distribution of the source and achieves asymptotic compression ratio=entropy of source

# Entropy of a dynamical system (Kolmogorov-Sinai entropy)

Given two partitions $\mathcal{P}$ and $\mathcal{Q}$

$\mathcal{P} \vee \mathcal{Q}$ the **join** of $\mathcal{P}$ and $\mathcal{Q}$

$B \cap C$ where $B \in \mathcal{Q}$ and $C \in \mathcal{Q}$

$T : X \to X$ measure preserving

$$\mathcal{P}_n = \mathcal{P} \vee T^{-1}\mathcal{P} \vee \cdots \vee T^{-(n-1)}\mathcal{P}$$

$$h(T, \mathcal{P}) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{P}_n) \qquad h(T) = \sup_{\mathcal{P}} h(T, \mathcal{P})$$

# Properties of the entropy

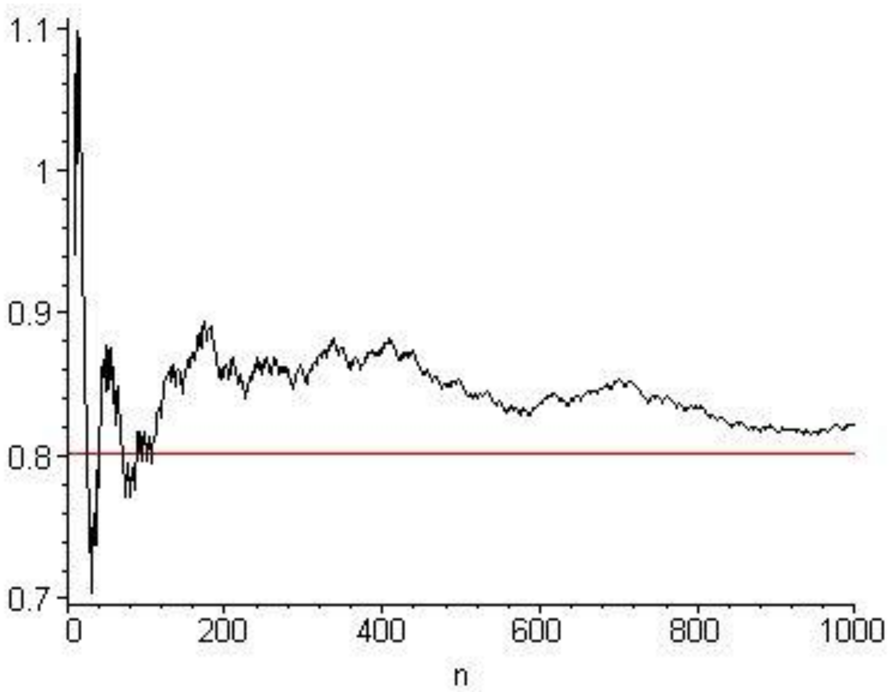
Let T:X→X, S:Y→Y be measure preserving (T preserves μ, S preserves ν)

$$\mathit{If}\ n \geq 1,\ \mathit{then}\ h(T^n) = n\,h(T)$$

$$\mathit{If}\ T\ \ \mathit{is\ invertible},\ \mathit{then}\ h(T^{-1}) = h(T)$$

If S is a factor of T then h(S,ν)≤h(T,μ)
If S and T are isomorphic then h(S,ν)=h(T,μ)
On XxY one has h(TxS,μxν)= h(T,μ)xh(S,ν)

# Shannon-Breiman-McMillan




Let $\mathscr{P}$ be a generating partition

Let P(n,x) be the element of

$$\bigvee_{i=0}^{n-1} T^{-i}\mathscr{P}$$

which contains x

The SHANNON-BREIMAN-MCMILLAN theorem says that for a.e. x one has

$$h(T,\mu) = -\lim_{n\to\infty} \frac{\operatorname{Log} \mu(P(n,x))}{n}$$

# Asymptotic equipartition property

*Suppose that* $\mathcal{P}$ *is a finite generating partition of* $X$. *For every* $\varepsilon > 0$ *and* $n \geq 1$ *there exist subsets in* $\mathcal{P}_n$, *which are called* $(n, \varepsilon)$*-typical subsets, satisfying the following:*
*(i) for every typical subset* $\mathcal{P}_n(x)$,

$$2^{-n(h+\varepsilon)} < \mu(\mathcal{P}_n(x)) < 2^{-n(h-\varepsilon)} \ ,$$

*(ii) the union of all* $(n, \varepsilon)$*-typical subsets has measure greater than* $1 - \varepsilon$, *and*
*(iii) the number of* $(n, \varepsilon)$*-typical subsets is between* $(1-\varepsilon)2^{n(h-\varepsilon)}$ *and* $2^{n(h+\varepsilon)}$.

These formulas assume that the entropy is measured in bits, i.e. using the base 2 logarithm

# Entropy of Bernoulli schemes

Let $N \geq 2$, $\Sigma_N = \{1, \ldots N\}^{\mathbb{Z}}$.

$d(x,y) = 2^{-a(x,y)}$ where $a(x,y) = \inf\{|n|\,,\, n \in \mathbb{Z}\,,\, x_n \neq y_n\}$

*shift* $\sigma : \Sigma_N \to \Sigma_N$ $\sigma((x_i)_{i\in\mathbb{Z}}) = (x_{i+1})_{i\in\mathbb{Z}}$

The topological entropy of $(\Sigma_N, \sigma)$ is $\log N$

$(p_1, \ldots, p_N) \in \Delta^{(N)}$ $\nu(\{i\}) = p_i$

**Definition 4.26** *The* Bernoulli scheme $BS(p_1, \ldots, p_N)$ *is the measurable dynamical system given by the shift map* $\sigma : \Sigma_N \to \Sigma_N$ *with the (product) probability measure* $\mu = \nu^{\mathbb{Z}}$ *on* $\Sigma_N$.

**Proposition 4.27** *The Kolmogorov–Sinai entropy of the Bernoulli scheme* $BS(p_1, \ldots, p_N)$ *is* $-\sum_{i=1}^{N} p_i \log p_i$.

# Topological Markov chains or subshifts of finite type

$$\Sigma_A = \{x \in \Sigma_N \,,\, (x_i, x_{i+1}) \in \Gamma \, \forall i \in \mathbb{Z}\} \qquad \Gamma \subset \{1, \ldots N\}^2$$

$\Sigma_A$ is a compact shift invariant subset of $\Sigma_N$

$A = A_\Gamma$ the $N \times N$ matrix with entries $a_{ij} \in \{0, 1\}$

$$a_{ij} = \begin{cases} 1 & \iff (i,j) \in \Gamma \\ 0 & \text{otherwise} \end{cases}$$

The restriction of the shift $\sigma$ to $\Sigma_A$ is denoted $\sigma_A$

$A^m = (a_{ij}^m)$ and $a_{ij}^m > 0$ for all $i, j$

# Entropy of Markov chains

**Theorem 4.35 (Perron–Frobenius, see [Gan])** If $A$ is primitive then there exists an eigenvalue $\lambda_A > 0$ such that :

(i) $|\lambda_A| > \lambda$ for all eigenvalues $\lambda \neq \lambda_A$ ;

(ii) the left and right eigenvectors associated to $\lambda_A$ are strictly positive and are unique up to constant multiples ;

(iii) $\lambda_A$ is a simple root of the characteristic polynomial of $A$.

the topological entropy of $\sigma_A$ is $\log \lambda_A$ (clearly $\lambda_A > 1$ since all the integers $a_{ij}^m > 0$)

Let $P = (P_{ij})$ be an $N \times N$ matrix such that

(i) $P_{ij} \geq 0$ for all $i, j$, and $P_{ij} > 0 \iff a_{ij} = 1$ ;

(ii) $\sum_{j=1}^{N} P_{ij} = 1$ for all $i = 1, \ldots, N$ ;

(iii) $P^m$ has all its entries strictly positive.

Such a matrix is called a *stochastic matrix*. Applying Perron–Frobenius theorem to $P$ we see that 1 is a simple eigenvalue of $P$ and there exists a normalized eigenvector $p = (p_1, \ldots, p_N) \in \Delta^{(N)}$ such that $p_i > 0$ for all $i$ and

$$\sum_{i=1}^{N} p_i P_{ij} = p_j \ , \quad \forall\, 1 \le i \le N \ .$$

We define a probability measure $\mu$ on $\Sigma_A$ corresponding to $P$ prescribing its value on the cylinders :

$$\mu\left(C\begin{pmatrix} j_0, \ldots, j_k \\ i, \ldots, i+k \end{pmatrix}\right) = p_{j_0} P_{j_0 j_1} \cdots P_{j_{k-1} j_k} \ ,$$

for all $i \in \mathbb{Z}$, $k \ge 0$ and $j_0, \ldots, j_k \in \{1, \ldots, N\}$. It is called the *Markov measure* associated to the stochastic matrix $P$.

the subshift $\sigma_A$ preserves the Markov measure $\mu$.

$$h_\mu(\sigma_A) = -\sum_{i,j=1}^{N} p_i P_{ij} \log P_{ij} \qquad\qquad h_\mu(\sigma_A) \le h_{top}(\sigma_A)$$

# Entropy, coding and data compression

- Computer file= infinitely long binary sequence
- Entropy = best possible compression ratio
- Lempel-Ziv (Compression of individual sequences via variable rate coding, IEEE Trans. Inf. Th. 24 (1978) 530-536): does not assume knowledge of probability distribution of the source and achieves asymptotic compression ratio=entropy of source

Let $X = \{0,1\}^{\mathbb{N}}$ and $\sigma$ be a left-shift map.
Define $R_n$ to be the first return time of the initial $n$-block, i.e.,

$$R_n(x) = \min\{j \geq 1 : x_1 \dots x_n = x_{j+1} \dots x_{j+n}\}.$$

$$x = \overbrace{\boxed{1010}\,01001101100}^{15}\,\boxed{1010}\cdots \Rightarrow R_4(x) = 15.$$

The convergence of $\dfrac{1}{n}\log R_n(x)$ to the entropy $h$ was studied in a relation with data compression algorithm such as the Lempel-Ziv compression algorithm.

The Lempel-Ziv data compression algorithm provide a universal way to coding a sequence without knowledge of source.
Parse a source sequence into shortest words that has not appeared so far:

$$1011010100010\cdots \Rightarrow 1, 0, 11, 01, 010, 00, 10, \ldots$$

For each new word, find a phrase consisting of all but the last bit, and recode the location of the phrase and the last bit as the compressed data.

$$(000, 1)\,(000, 0)\,(001, 1)\,(010, 1)\,(100, 0)\,(010, 0)\,(001, 0)\ldots$$

**Theorem (Wyner-Ziv(1989), Ornstein and Weiss(1993))**

*For ergodic processes with entropy $h$,*

$$\lim_{n\to\infty} \frac{1}{n} \log R_n(x) = h \quad \textit{almost surely.}$$

The meaning of entropy

- Entropy measures the information content or the amount of randomness.
- Entropy measures the maximum compression rate.
- Totally random binary sequence has entropy $\log 2 = 1$. It cannot be compressed further.

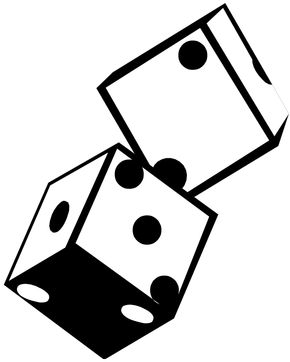
# Risk and how to manage risk

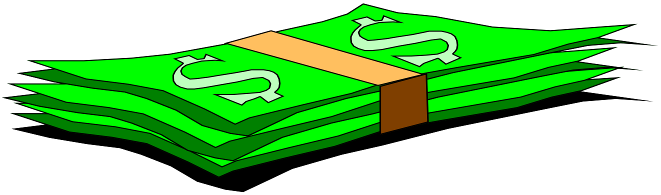
Risk = quantifiable uncertainty

Various cognitive problems

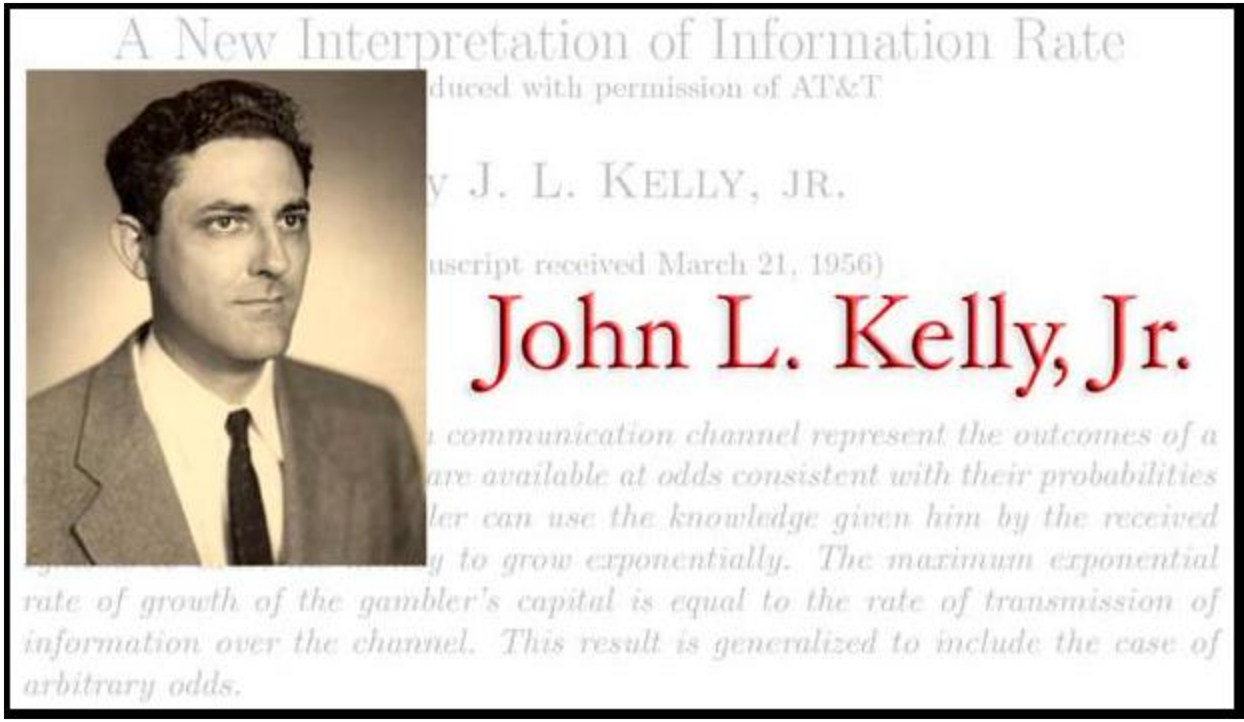
- illusions of safety
- ignorance
- wrong communication
- confusion: unable to elaborate information

Claude Elwood Shannon (1916-2001)





John Larry Kelly Jr.
(1923-1965)



# Kelly Criterion: some history

- Developed by John Kelly, a physicist at Bell Labs, who in 1956 published the paper "A New Interpretation of Information Rate" in the Bell Technical Journal
- Original title "Information Theory and Gambling" was censored
- Kelly used Information Theory to show how much a gambler with inside information should bet to optimize the growth rate of capital
- Breiman showed that Kelly gambling had a higher growth rate of wealth than any other investment scheme and that it minimized the time necessary for the wealth to achieve a distant goal
- In the mid 1960s Shannon gave a lecture on maximizing the growth rate of wealth and gave a geometric Wiener example
- Ed Thorpe used system to compute optimum bets for blackjack and later as manager of a hedge fund on Wall Street. In 1962 he wrote the book

"Beat the Dealer: A Winning Strategy for the Game of Twenty One"

# Kelly’s article

If the input symbols to a communication channel represent the outcomes of a chance event on which bets are available at odds consistent with their probabilities (i.e., “fair" odds), a gambler can use the knowledge given him by the received symbols to cause his money to grow exponentially. The maximum exponential rate of growth of the gambler's capital is equal to the rate of transmission of information over the channel. This result is generalized to include the case of arbitrary odds.

THE GAMBLER WITH A PRIVATE WIRE

Let us consider a communication channel which is used to transmit the results of a chance situation before those results become common knowledge, so that a gambler may still place bets at the original odds.

Without noise the gambler would bet all his capital each time, but what is the optimal fraction of capital to bet when the channel is noisy?

# Kelly’s fractional betting criterion

You don't even have to know what a logarithm is to use the so-called Kelly formula. You should wager this fraction of your bankroll on a favorable bet:

*edge/odds*

The *edge* is how much you expect to win, on the average, assuming you could make this wager over and over with the same probabilities. It is a fraction because the profit is always in proportion to how much you wager. At a racetrack, the edge is diminished by the track take. When your edge is zero or negative, the Kelly criterion says not to bet.

*Odds* means the public or tote-board odds. It measures the profit *if* you win. The odds will be something like 8:1, meaning that a winning wager receives 8 times the amount wagered *plus* return of the wager itself.

http://home.williampoundstone.net/Kelly/Kelly.html

# Kelly's formula: edge/odds

In the Kelly formula, *odds* is not necessarily a good measure of probability. Odds are set by market forces, by everyone else's beliefs about the chance of winning. These beliefs may be wrong. In fact, they have to be wrong for the Kelly gambler to have an edge. The odds do not factor in the Kelly gambler's inside tips.

Example: The tote board odds for Seabiscuit are 5:1. Odds are a fraction -- 5:1 means 5/1 or 5. The 5 is all you need.

The tips convince you that Seabiscuit actually has a 1 in 3 chance of winning. Then by betting $100 on Seabiscuit you stand a 1/3 chance of ending up with $600. On the average, that is worth $200, a net profit of $100. The edge is the $100 profit divided by the $100 wager, or simply 1.

The Kelly formula, *edge/odds*, is 1/5. This means that you should bet one-fifth of your bankroll on Seabiscuit.

http://home.williampoundstone.net/Kelly/Kelly.html

# An example

You play a sequence of games, where:

- If you *win*, you get W dollars for each dollar bet
- If you *lose*, you lose your bet

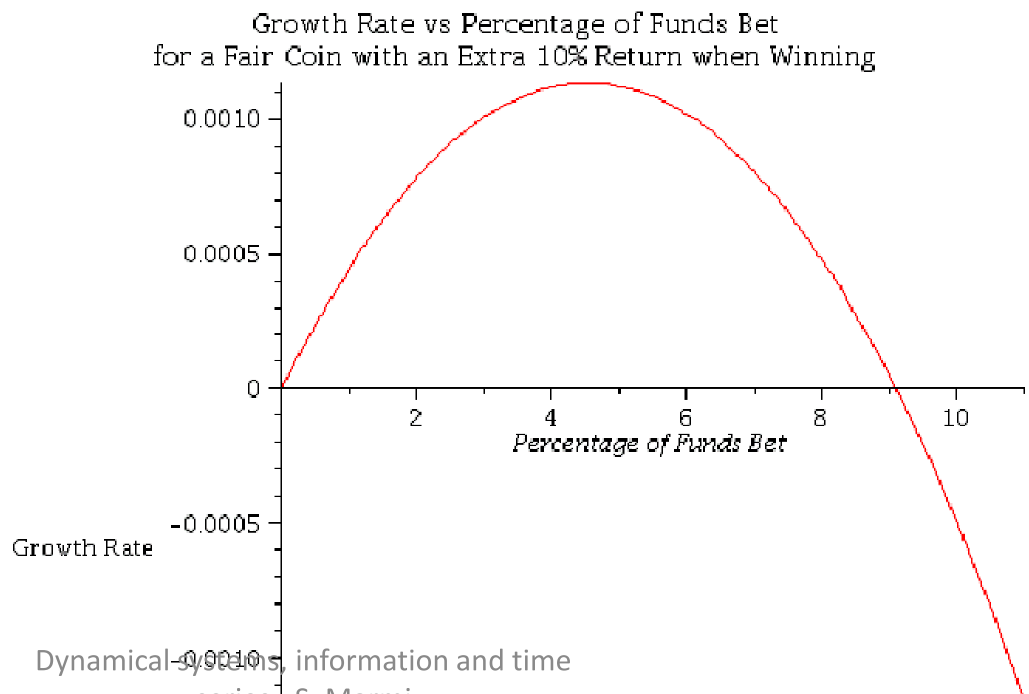
For each game, the probability of winning is p and losing is $q = 1 - p$

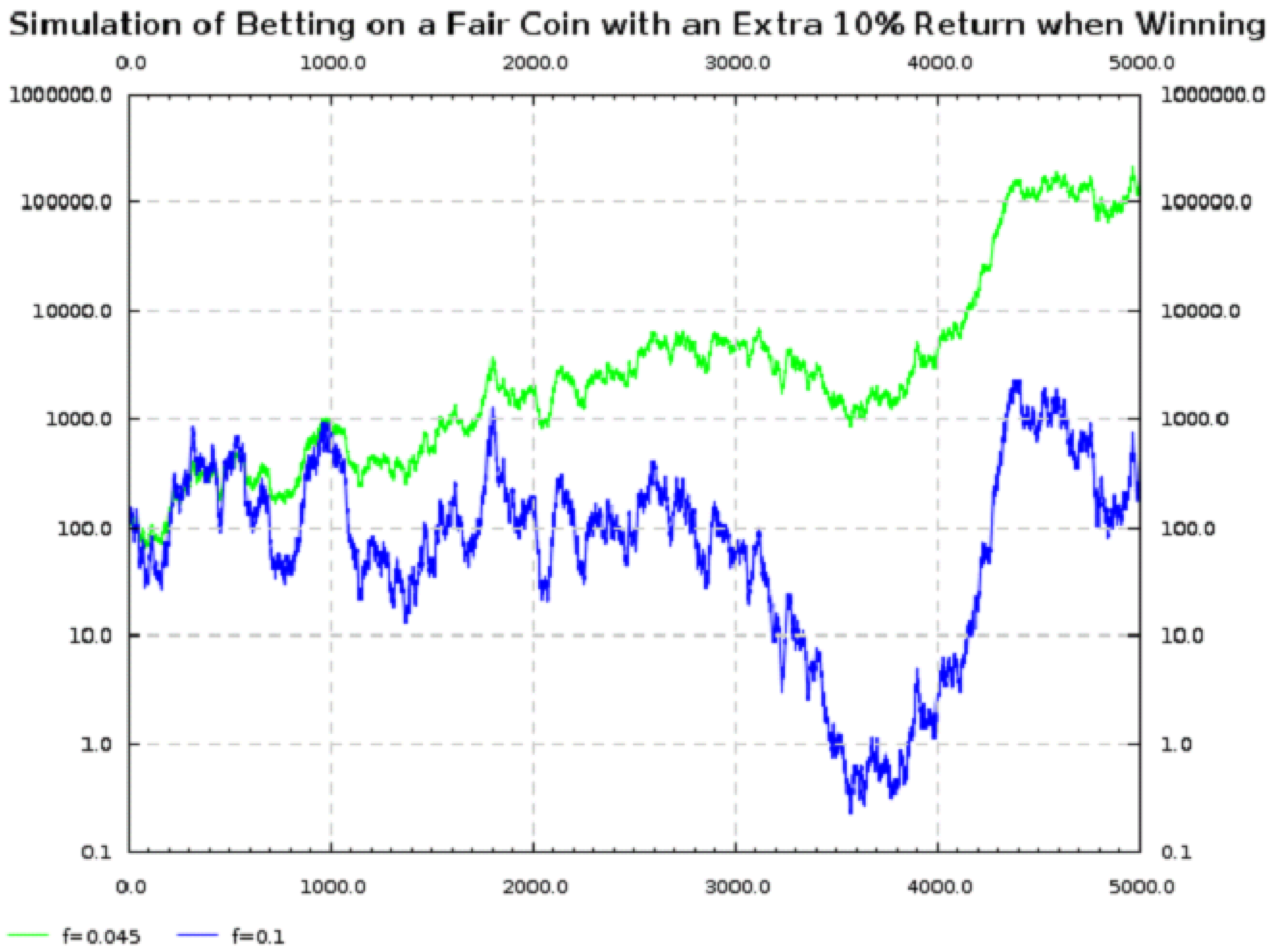
You bet some fixed percentage f of your bankroll B each game, for you have $(1 - f)B$ if you lose and $(W - 1)fB + B$ if you win.

The right value of f is called the Kelly Criterion.

Suppose we bet $1 on a fair coin, but one which pays $2.10 if it comes up heads? What fraction of your bankroll should you bet? The odds are in your favor, but if you bet all your money on each game, you will eventually lose a game and be bankrupt. If you bet too little, you will not make as much money as you could

Bet too much and we lose, even with the odds in our favor!






http://www.cse.ust.hk/~skiena/510/lectures/lecture25.pdf

You play a sequence of $n = w + l$ games. Each game, you either win W for each dollar bet with probability p or lose your bet with probability $q = 1 - p$.

If after n games, you have won $w$ and lost $l$ games, your total bankroll $B(n)$ is

$\text{Log } B(n) = w \text{ Log}(1+fW) + l \text{ Log}(1-f) + \text{Log } B(0)$

the exponential growth rate of your bankroll clearly is

$[w \text{ Log}(1+fW) + l \text{ Log}(1-f)]/n$

And by the ergodic theorem, or the strong law of large numbers, we almost surely have the limit

Exponential growth rate = $p \text{ Log}(1+fW) + q \text{ Log}(1-f)$

The f that maximizes G is easy to determine:

$$f = (pW - q)/W = p - q/W$$

# Kelly's criterion for gambling

The Kelly criterion $f = (pW - q)/W = \text{edge}/\text{odds}$

The odds are how much you will win if you win, e.g. the tote-board odds at a racetrack.

The edge is how much you expect to win, e.g. p is your inside knowledge of which horse will win.

The corresponding exponential growth rate is

$$G=-h+\text{Log}(1+W)-q\text{Log}W$$

Where h is the entropy.

If $pW - q = 0$, you have no advantage and shouldn't bet

anything, so $f = 0$. If $q = 0$, then $f = 1$ and you should borrow to bet all you possibly could!

# Horse races

$p_i$ = probability that the i-th horse wins the race;

$b_i$ = fraction of wealth I bet on the i-th horse

$o_i$ =odds (payoff) of the i-th horse (if $o_i = q$ the i-th horse pays q times the amount bet).

The bets are *fair* if the bookmaker makes no profit, i.e. $1 = \sum 1/o_i$

Exponential growth rate of capital : $W(b,p) = \sum p_i \log (b_i o_i)$

Is maximum if $b_i = p_i$ for all i which gives

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

This is the distance between the entropies of the estimated probability measures of the bettor and of the bookmaker of the true (unknown) probability distribution

# A race with 2 horses

Two horses have probabilities $p_1 > p_2$ of winning the race. But the first horse is less popular than the second and the bets are evenly distributed between the two horses, thus the bookmakers give the same odds (2-1) to both horses

The optimal bet (Kelly) is given by $b_1 = p_1$, $b_2 = p_2$.

And the corresponding capital growth rate is

$W(p) = \sum p_i \log o_i - H(p) = 1 - H(p)$

Here 1=entropy estimated (wrongly) by the bookmaker

THUS THE CAPITAL GROWTH RATE = DIFFERENCE BETWEEN THE BOOKMAKER ESTIMATE OF THE ENTROPY AND THE TRUE ENTROPY

And after n bets the expected bankroll is

$B_n = 2^{n(1-H(p))} B_0$

# What is an efficient market? (stockmarket, horse races, sports betting, etc)

An efficient capital market is a market which is efficient in processing information: the prices of securities observed at any time are based on “correct” evaluation of all information available at that time. Prices “fully reflect” available information.
The prices are always “fair”, they are good indicators of value

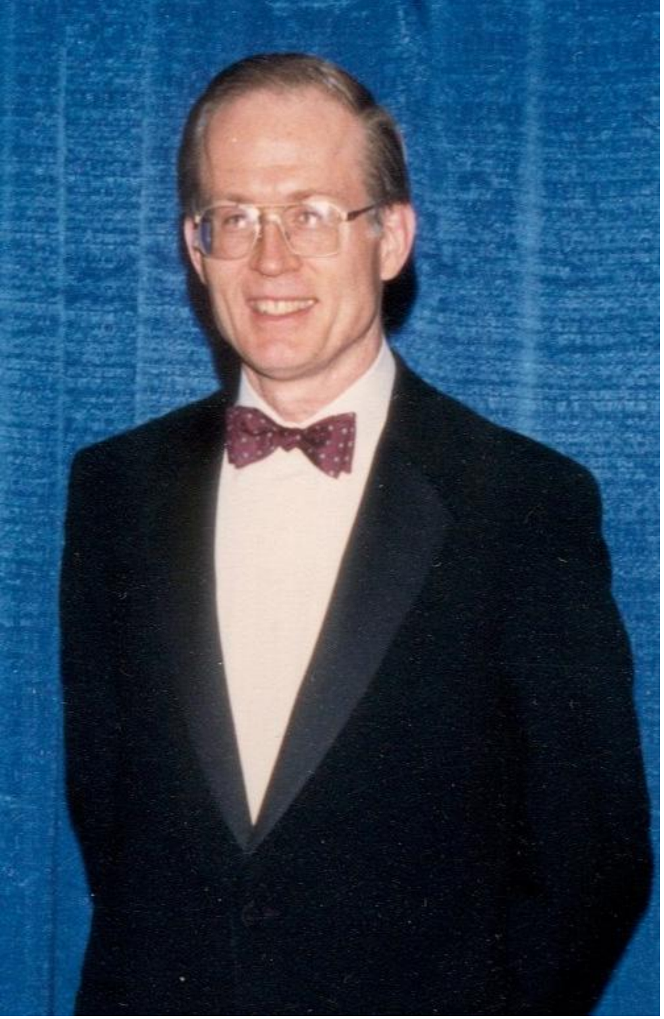
The concept of market efficiency had been anticipated at the beginning of the century:

**Bachelier** (1900) “Les influences qui déterminent les mouvements de la Bourse sont innombrables, des événements passés, actuels ou même escomptables, ne présentant souvent aucun rapport apparent avec ses variations, se répercutent sur son cours”
…”Si le marché, en effet, ne prévoit pas les mouvements, il les considère comme étant plus ou moins probables, et cette probabilité peut s’évaluer mathématiquement.”

# Weak vs. strong efficiency

More formally: a capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set, $\Theta_t$ , if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set, $\Theta_t$, implies that it is impossible to make economic profits by trading on the basis of $\Theta_t$.

The weak form of the efficient market hypothesis claims that prices <u>fully reflect the information implicit in the sequence of **past prices**</u>. The semi-strong form of the hypothesis asserts that <u>prices reflect all relevant information that is **publicly** available</u>, while the strong form of market efficiency asserts <u>information that is known to **any** participant is reflected in market prices.</u>



**Fischer Sheffey Black** (January 11, 1938 – August 30, 1995) was an American economist, best known as one of the authors of the famous Black-Scholes equation.

"However, we might define an efficient market as one in which price is within a factor of 2 of value, i.e. the price is more than half of value and less than twice value. The factor of 2 is arbitrary, of course. Intuitively, though, it seems reasonable to me, in the light of sources of uncertainty about value and the strength of the forces tending to cause price to return to value. By this definition, I think almost all markets are efficient almost all of the time. 'Almost all' means at least 90% "

F. Black, Noise, Journal of Finance (1986) p. 533.

THE JOURNAL OF FINANCE • VOL. XLI, NO. 3 • JULY 1986

# Noise

FISCHER BLACK*

**ABSTRACT**

The effects of noise on the world, and on our views of the world, are profound. Noise in the sense of a large number of small events is often a causal factor much more powerful than a small number of large events can be. Noise makes trading in financial markets possible, and thus allows us to observe prices for financial assets. Noise causes markets to be somewhat inefficient, but often prevents us from taking advantage of inefficiencies. Noise in the form of uncertainty about future tastes and technology by sector causes business cycles, and makes them highly resistant to improvement through government intervention. Noise in the form of expectations that need not follow rational rules causes inflation to be what it is, at least in the absence of a gold standard or fixed exchange rates. Noise in the form of uncertainty about what relative prices would be with other exchange rates makes us think incorrectly that changes in exchange rates or inflation rates cause changes in trade or investment flows or economic activity. Most generally, noise makes it very difficult to test either practical or academic theories about the way that financial or economic markets work. We are forced to act largely in the dark.

# Shannon makes money even with geometric random walks with no drift

Suppose that the value of a stock follows a geometric random walk with *no drift:* each day the stock price

– doubles if we "win": W = 1 return = +100%

– halves if we "lose": L = ½ return = -50%

- The arithmetic mean of the returns is 25% but due to the extreme volatility the geometric mean is 0%

# Shannon vs. buy and hold

p = ½, W = 1, L = 0.5. Then f = .5 and

G = 1.0607

If for example we start with B = 100:

– 1 bet: we bet 50 and lose (25) . B is now =75

– 2 bet: we bet ½ of the new B i.e. 37.50. We win. B becomes 37.50+2*37.50= 112.50

If we had just followed a *buy and hold* strategy and stayed fully invested all the time the result would have been:

– after the first bet B would have been =50

– after the second bet B =100 NO GAIN AT ALL!

# Shannon and range bound markets

Thus, even in a “range bound” market, with the stock simply oscillating around a mean values, following a geometric random walk without drifts, after n days the expected capital gain following Shannon’s advice is $(1.0607)^n$