Entropia, informazione e rischio: la matematica tra Silicon Valley e Wall Street

Stefano Marmi

http://homepage.sns.it/marmi/

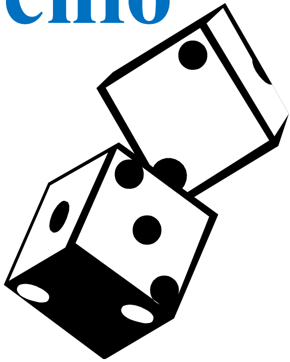
Scuola Normale Superiore

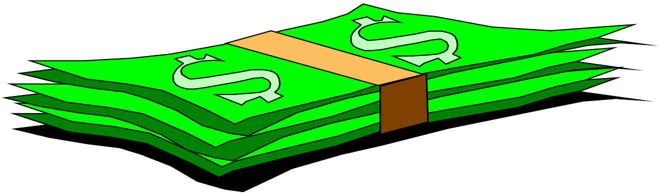
65° Corso di Orientamento Universitario

San Miniato , 10 settembre 2009

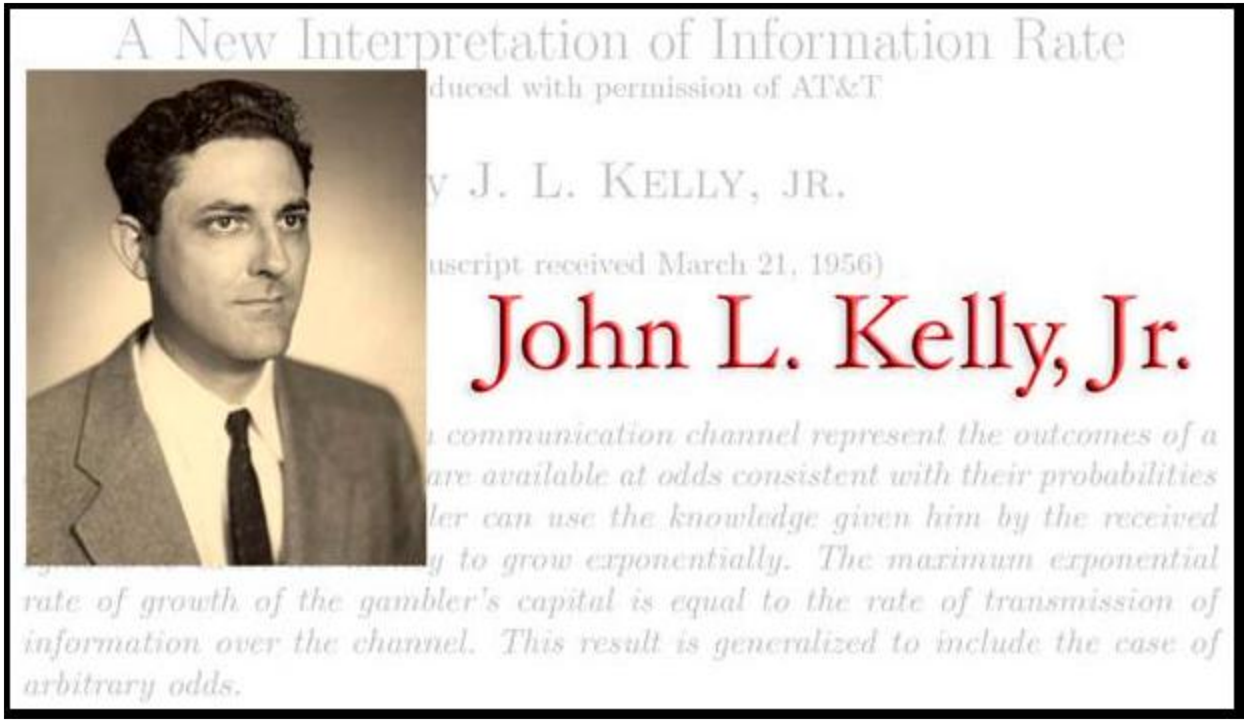
La scommessa, l'informazione e il rischio

Claude Elwood Shannon (1916-2001)





John Larry Kelly Jr.
(1923-1965)



A New Interpretation of Information Rate

duced with permission of AT&T

y J. L. KELLY, JR.

uscript received March 21, 1956)

John L. Kelly, Jr.

communication channel represent the outcomes of a
are available at odds consistent with their probabilities
ler can use the knowledge given him by the received
y 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.

Il criterio di Kelly: lo scommettitore con informazioni riservate

- C.E. Shannon “A Mathematical Theory of Communications” Bell Tech. J. (1948)
- J. L. Kelly “A New Interpretation of Information Rate” Bell. Tech. J. (1956)
- 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

Entropia

Nella teoria della probabilità, l'*entropia* misura l'incertezza associata a una variabile aleatoria

Consideriamo un esperimento con possibili risultati $A=\{a_{1,} \ldots, a_k\}$

- Supponiamo che la probabilità dell'esito a_i sia p_i , $0 \leq p_i \leq 1$, $p_1 + \ldots p_k = 1$
- Se per esempio l'esito a_1 si presenta con una probabilità molto vicina a 1, allora nella maggior parte degli esperimenti l'esito sarebbe proprio a_1 e dunque non sarebbe molto incerto. Non si guadagna molta informazione dall'esperienza.
- Possiamo misurare la "sorpresa" dell'esito come

  informazione= −log (probabilità)
- (l'intensità della percezione è proporzionale al logaritmo della grandezza dello stimolo)

Entropia come informazione prodotta dall'esperienza

L'entropia associata all'esperimento è allora

$H = -\sum p_i \log p_i$

Poiché abbiamo definito

l'informazione = - Log (probabilità)

l'entropia è semplicemente il valore atteso dell'informazione dall'esperimento

Entropia, codifica, compressione di dati

Cosa misura l'entropia?

L'entropia misura il contenuto di informazione (o quanto è *random* un segnale)

L'entropia misura il massimo rapporto di compressione: una successione binaria completamente *random* ha entropia uguale a $\log_2 2 = 1$ e non può essere compressa (lancio di una moneta *non truccata)*

Computer file= successione binaria (infinitamente) lunga

Entropia = migliore rapporto di compressione possibile

Algoritmo di Lempel-Ziv (Compression of individual sequences via variable rate coding, IEEE Trans. Inf. Th. 24 (1978) 530-536): non richiede la conoscenza della distribuzione di probabilità della sorgente e riesce a raggiungere il tasso di compressione asintotico = entropia della sorgente

Dinamica, probabilità, statistica e il problema dell'induzione

- (Ammesso che esista) la probabilità di un evento non è quasi mai nota a priori.
- La sola possibilità è di usare al suo posto le frequenze calcolate osservando con quale frequenza l'evento si è presentato in passato
- Il problema del *backtesting*
- Il problema dell'*ergodicità*, della *storicità* e dei *punti tipici*: da una singola serie di osservazioni vorrei essere in grado di calcolare la probabilità
- *Il pollo di Bertrand Russell's* (tacchino per gli americani)

Bertrand Russell

(The Problems of Philosophy,
Home University Library, 1912. Chapter VI On Induction) Available at the page
http://www.ditext.com/russell/rus6.html

Domestic animals expect food when they see the person who feeds them. We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken.

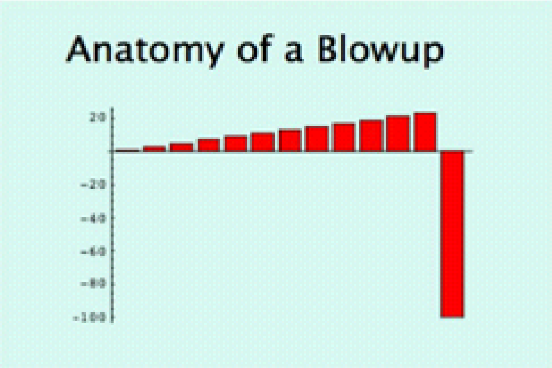


Figure 1 *My classical metaphor: A Turkey is fed for a 1000 days—every days confirms to its statistical department that the human race cares about its welfare "with increased statistical significance". On the 1001st day, the turkey has a surprise.*

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

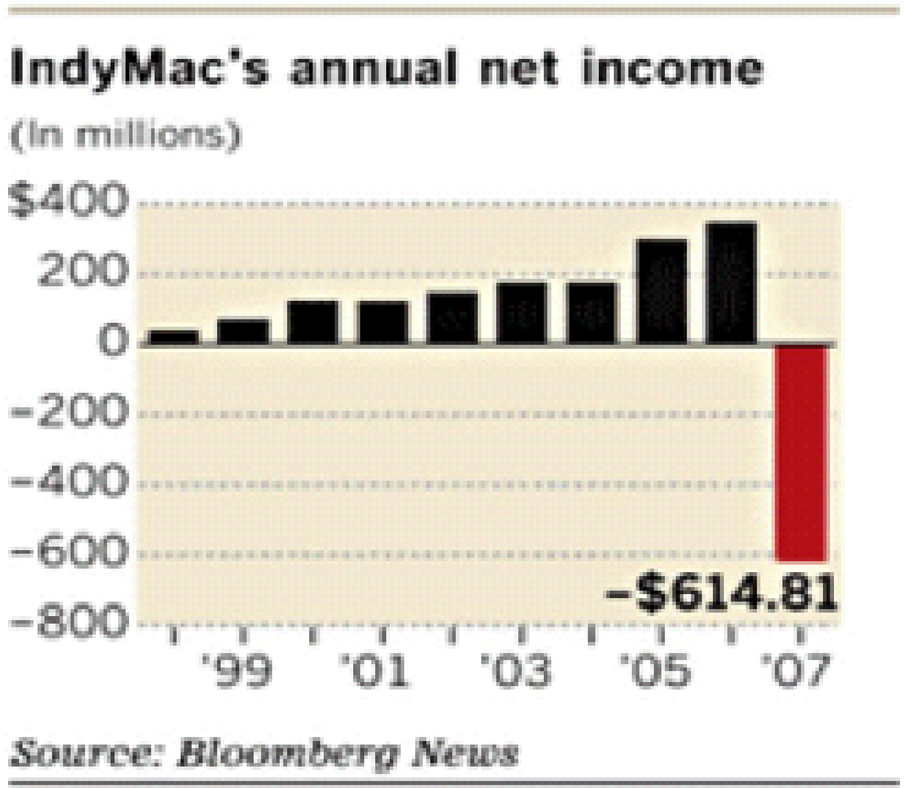


Figure 2 *The graph above shows the fate of close to 1000 financial institutions (includes busts such as FNMA, Bear Stearns, Northern Rock, Lehman Brothers, etc.). The banking system (betting AGAINST rare events) just lost > 1 Trillion dollars (so far) on a single error, more than was ever earned in the history*

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

Payoff from mildly OTM UK Sterling Short Option, 1988-2008

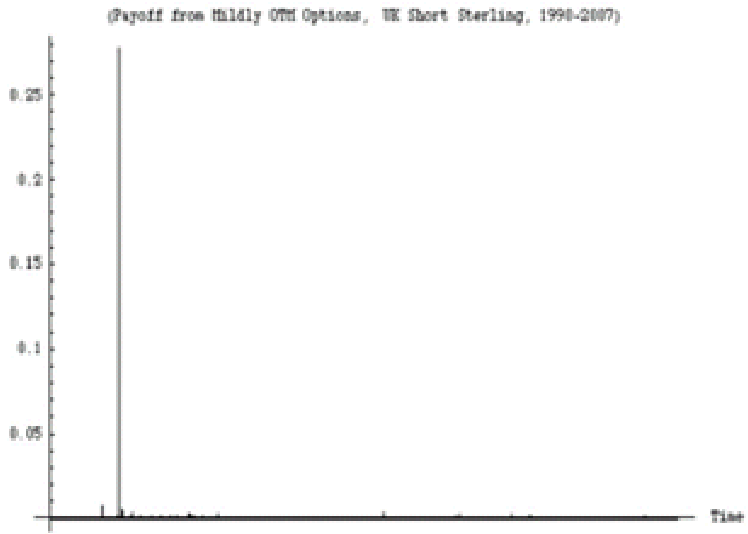


Figure 3 The graph shows the daily variations a derivativesportfolio exposed to U.K. interest rates between 1988 and 2008. Close to 99% of the variations, over the span of 20 years, will be represented in 1 single day—the day the European Monetary System collapsed. As I show in the appendix, this is typical with ANY socio-economic variable (commodity prices, currencies, inflation numbers, GDP, company performance, etc.). No known econometric statistical method can capture the probability of the event with any remotely acceptable accuracy (except, of course, in hindsight, and "on paper"). Also note that this applies to surges on electricity grids and all manner of modern-day phenomena.

http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

Il problema della storicità: e se non ci fosse la media temporale?

La successione di Kolakoski (1965): inizia con 2 e la successione delle lunghezze di stringhe consecutive di 1 e di 2 è la successione stessa: dunque si ha

2

22

2211

221121

221121221

22112122122112…

Un problema aperto

1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 1, 2, 1, 2, 2 …

Non si sa se la densità di 1 esiste ed è =1/2 come sembra plausibile numericamente

La successione può essere generata (tranne il primo 1) iniziando con 22 e applicando le regole di sostituzione dei blocchi 22 →2211, 21 → 221, 12 → 211, 11 → 21 (Lagarias) dunque ha una complessità algoritmica molto bassa. Anche l'entropia è nulla, dunque non è una successione molto random…

Il cigno nero

N. Taleb The Black Swan, 2007

le eccezioni sono tra noi, pronte a sorprenderci e a cambiare la nostra visione del mondo

“mediocristan” vs. “extremistan”

Mediocristan tipico delle grandezze fisiche (altezza, peso, ecc)

Extremistan tipico dei fenomeni sociali e storici

Un lunedì davvero speciale

Venerdì 19 ottobre 2007 la borsa statunitense ha celebrato con una sostanziosa perdita del 2,6% il ventesimo anniversario del terribile “lunedì nero”: il 19 ottobre 1987, l’indice Dow Jones ebbe un crollo del 22,6% in una sola seduta. Il crollo del Dow innescò una serie di perdite nei mercati finanziari di tutto il mondo: nel giro di una settimana o poco più di un quarto della capitalizzazione delle borse mondiali era andata in fumo. Dal 1928 ad oggi la deviazione standard della distribuzione dei rendimenti giornalieri del Dow Jones è circa l’uno percento. Dunque il crollo del 19 ottobre 1987 non avrebbe mai dovuto verificarsi poiché avrebbe avuto, in teoria, una probabilità di verificarsi pari a 1 contro un googol, che è un 1 seguito da 100 zeri! E’ la stessa probabilità che avete di fare sempre testa lanciando una moneta per 332 volte di fila oppure di lanciare un dado e ottenere 6 per 129 volte consecutive. Se vi sembra facile pensate che fare un sei al superenalotto è un po’ meno difficile che fare sempre testa lanciando una moneta per 30 volte di fila.

Incertezza, rischio e previsione:

"How do you want it—the crystal mumbo-jumbo or statistical probability?"

Le considerazioni svolte finora condizionano l'attività economica (banche, VaR, Basilea 2, TFR, fondi, …)

L'importanza dell'analisi statistica e della modellizzazione matematica nelle scelte economiche e finanziarie è cresciuta costantemente negli ultimi 20-30 anni

Imparare a vivere con il rischio e l'incertezza

Rischio = incertezza quantificabile

Oltre ai problemi quantitativi che abbiamo visto e che vedremo tra poco ci sono anche importanti problemi cognitivi:

- illusione della certezza
- ignoranza del rischio
- comunicazione scorretta
- pensiero annebbiato: incapacità di elaborare l'informazione

Gerd Gigerenzer:
Quando i numeri ingannano (2003)

Come agire quando si può fare una scommessa vantaggiosa?

- Supponete di giocare una serie di scommesse in cui si lancia una moneta:

– testa e croce hanno la stessa probabilità = ½

– testa = vinco 1.1 euro per ogni euro scommesso dunque ho 2.1 euro per ogni euro scomesso

– croce= perdo il capitale scommesso

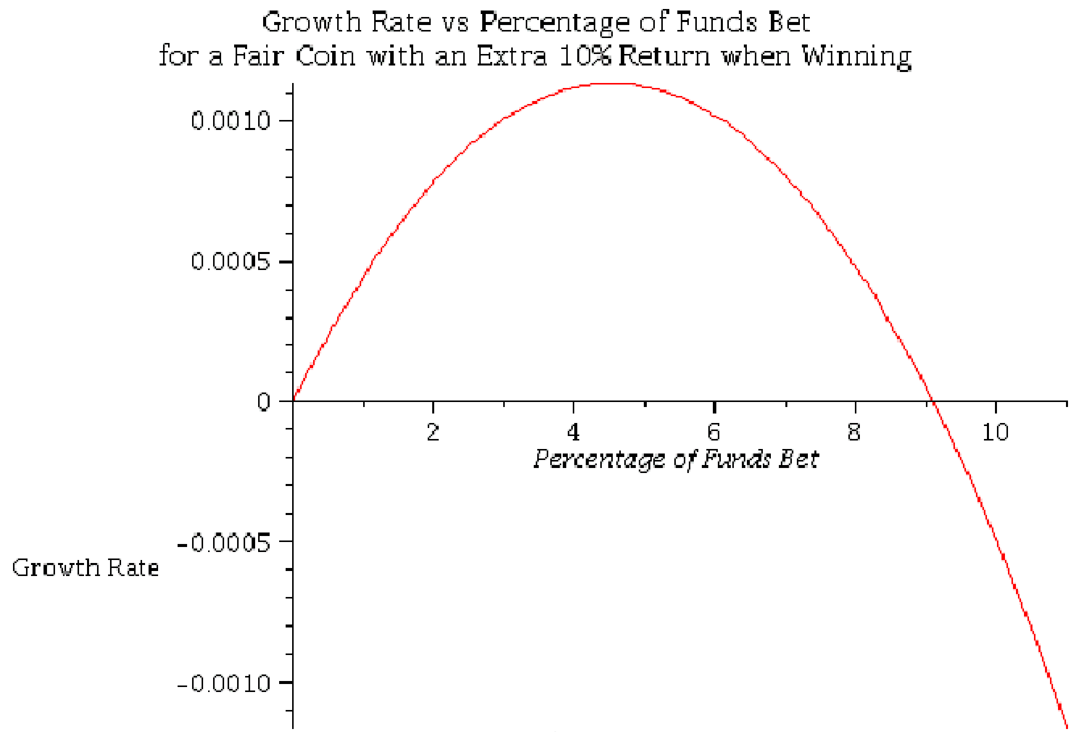
- Che frazione del capitale dovrei rischiare ogni volta??

E' un gioco vantaggioso ma se scommetto tutto ogni volta finirò sicuramente in bancarotta, se scommetto troppo poco perderò un'occasione di guadagno

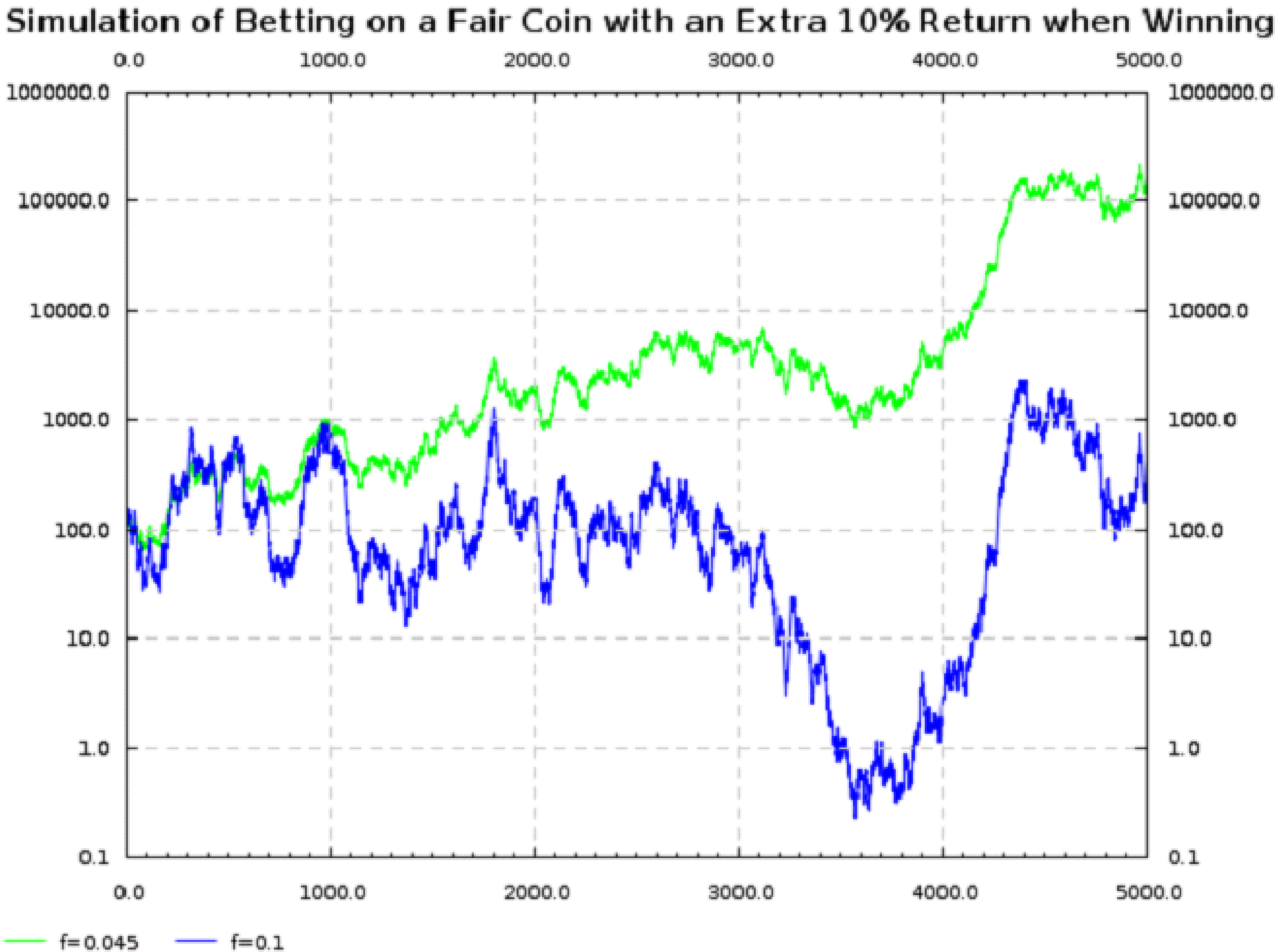
Supponiamo di scommettere $1 su una moneta *fair*, ma che il bookmaker restituisca $2.10 quando viene testa (perdiamo tutto se viene croce)

Quale frazione del capitale è giusto rischiare ogni volta?

Se si scommette troppo si finisce con perdere il capitale anche se la scommessa è favorevole !!!!!!!!!!



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



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

L'articolo di Kelly

« ... 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? ... »

Il criterio di Kelly

« ... 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

Il criterio di Kelly

« ... 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

Il criterio di Kelly per una scommessa semplice

Consideriamo una serie di scommesse nelle quali per ogni euro di capitale scommesso (capitale di rischio) se ne guadagnano W in caso di vincita e se ne perdono L in caso di perdita

Indichiamo con f la frazione del capitale disponibile che viene puntata ogni volta. Se la probabilità di vincere è p e quella di perdere è q=1-p il valore atteso del rapporto tra il capitale finale e quello iniziale è

$$G = (1 + fW)^p * (1 – fL)^q$$

Massimizzare la media geometrica delle vincite

G coincide con la media geometrica del capitale nel caso di un numero infinito di scommesse ripetute: infatti se n=w+l, w=numero di scommesse vinte, l= numero di scommesse perse

$$G = \lim_{n\to\infty} ((1 + fW)^{w/n} * (1 - fL)^{l/n})$$

$$= (1 + fW)^p * (1 - fL)^q$$

poichè $w/n \to p$, $l/n \to q$

Mi aspetto dunque che un capitale iniziale B diventi dopo n scommesse

$$B_n = G^n * B$$

Il criterio di Kelly

G è massimo scegliendo

$f = (pW - qL)/WL = p/L - q/W$

Questo è il criterio di Kelly per questo tipo di scommessa

Se p=q e L=W allora f=0 (non scommettere)

Se invece p=q ma L<W allora conviene scommettere! *Se addirittura $L \to 0$ allora si può avere $f \to \infty$ dunque conviene indebitarsi illimitatamente e scommettere …*

…chissà se lo sapevano i *risk manager* di Bear Stearns, Lehman Brothers, American Insurance Group, Freddie Mac, ecc. ecc…?

Kelly come criterio di risk management

- Nel caso ($W = 2$, $p = ½$, $L=1$)

$f = .5 - .5 / 2 = .25$

$G = 1.0607$

– Dopo 10 scommesse (supponendo $B = 1$)

- il valore atteso della ricchezza finale è = 3.25
- il valore mediano è = 1.80

Se invece si scommette tutto il capitale disponibile ogni volta ($f = 1$), dopo 10 scommesse si ha

- il valore atteso della ricchezza finale è = 57.67 … **MA**
- il valore mediano è = 0 (la probabilità di farcela è 2^{-10})

Le corse dei cavalli

La probabilità che l'i-esimo cavallo vinca la corsa è p_i

La frazione del capitale scommessa sull'i-esimo cavallo è b_i

La quota alla quale viene dato l'i-esimo cavallo è o_i (se $o_i = q$ l'i-esimo cavallo paga q volte la posta). Le scommesse sono *fair* se il bookmaker non trattiene nulla, ovvero $1 = \sum 1/o_i$

Allora il tasso di crescita esponenziale del capitale è

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

che è massima scegliendo $b_i = p_i$ per ogni i nel qual caso si ottiene

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

Questa formulazione si presta a importanti considerazioni in termini di teoria dell'informazione, ad esempio come distanza tra la stima che lo scommettitore e il bookmaker fanno della vera distribuzione di probabilità degli esiti delle corse

La corsa con 2 cavalli

Due cavalli, con probabilità $p_{1>}$ p_2 di vincere

Il primo però è un po'meno di moda del secondo così le scommesse si equiripartiscono e i bookmaker assegnano la stessa quota (2-1) a entrambi.

La scommessa ottimale (Kelly) si ha scegliendo $b_1 = p_1$, $b_2 = p_2$.

e il tasso di crescita del capitale corrispondente è

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

qui 1=entropia stimata (erroneamente) dal bookmaker

DUNQUE IL TASSO DI CRESCITA DEL CAPITALE = DIFFERENZA TRA L'ENTROPIA STIMATA DAL BOOKMAKER E QUELLA VERA

così dopo n scommesse il capitale atteso sarà

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

Cos'è un mercato efficiente (borsa, sala corse, ecc)?

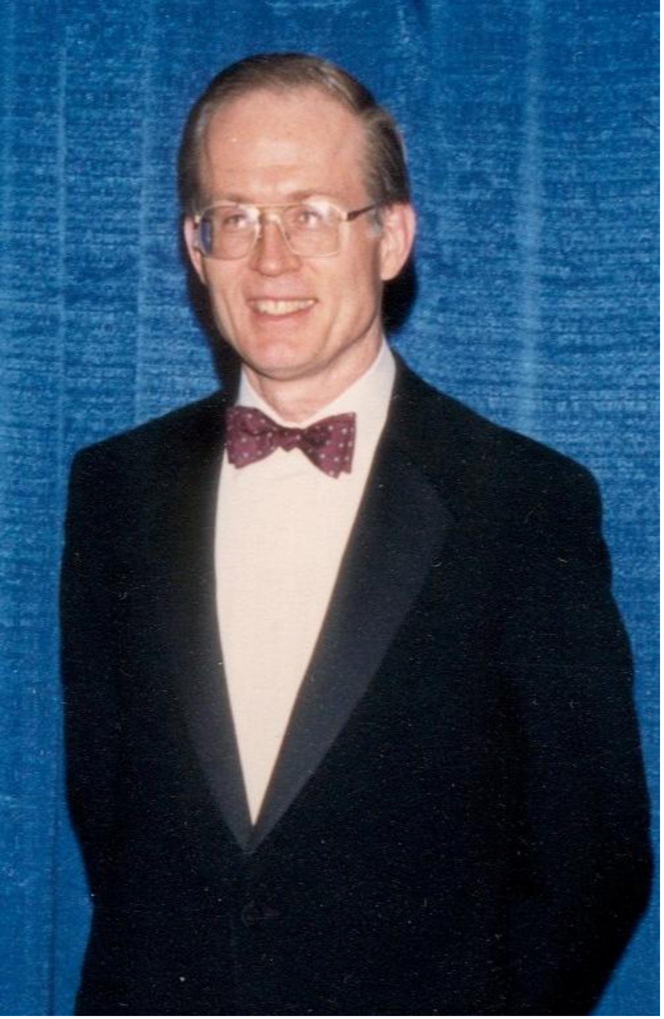
Un mercato è efficiente quando è efficiente nell'elaborazione delle informazioni: i prezzi dei beni (azioni, quote del bookmaker, obbligazioni, materie prime, ecc) osservati in ogni istante di tempo sono il risultato di una valutazione "corretta" di tutta l'informazione disponibile al momento. I prezzi "riflettono pienamente" tutta l'informazione disponibile, sono sempre "fair", cioè buone indicazioni dei valori in gioco.

Bachelier (1900) scrive che "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."

Efficienza forte e debole

Un mercato è efficiente rispetto a un "insieme" di informazioni Θ_t se i prezzi non cambierebbero rivelando queste informazioni a tutti gli agenti → non è possibile fare profitti utilizzando Θ_t per il trading

La forma debole dell'ipotesi dei mercati efficienti richiede che i prezzi rispecchino pienamente l'informazione implicita nella successione dei prezzi passati. La forma semi-forte asserisce che i prezzi rispecchiano tutta l'informazione pubblicamente disponibile mentre nella forma forte i prezzi riflettono anche l'informazione non pubblicamente disponibile ma conosciuta da almeno un agente.



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 sfrutta Bachelier

- Claude Shannon ebbe l'idea di utilizzare il criterio di Kelly per guadagnare da una variazione aleatoria dei prezzi delle azioni

Supponiamo che il valore di un'azione segua un moto Browniano "geometrico": alla fine di ogni giorno

– Se "vinciamo" W = 1 cioè il prezzo raddoppia

– Se "perdiamo" L = ½ cioè l'azione dimezza il suo valore

Shannon e il cassettista

Supponiamo $p = ½$, $W = 1$, $L = 0.5$. Allora $f = .5$ e $G = 1.0607$

Iniziamo con B = 100. Seguiamo il consiglio di Shannon e:

– scommessa1: puntiamo 50 e perdiamo (25) . B ora vale 75

– scommessa 2: puntiamo ½ del nuovo capitale B cioè 37.50. Vinciamo. B diventa 37.50+2*37.50= 112.50

Se invece seguiamo una strategia da "cassettista" (buy and hold) puntando costantemente l'intero capitale a disposizione

– dopo la scommessa1, B sarebbe uguale a 50

– dopo la scommessa 2, B sarebbe 100

Shannon e i mercati oscillanti

Se il primo giorno l'azione raddoppia e il secondo l'azione si dimezza, un investitore "cassettista" che semplicemente è rimasto pienamente investito tutto il tempo non avrebbe nè guadagnato nè perso

Shannon invece guadagna: dopo due giorni il suo capitale inziale di 100 è diventanto 112.50

Anche nel caso di un mercato "range bound" nel quale l'azione semplicemente oscilla intorno al suo valore, seguendo un moto browniano geometrico senza drift, in n giorni il guadagno atteso seguendo la strategia di Shannon è $(1.0607)^n$

GRAZIE!