

# Teoria dell'Informazione e Portafogli Universali

*Alessandro Trenta*  
Relatore: *Stefano Marmi*

Scuola Normale Superiore

- 1 Il mercato azionario
- 2 Portafogli Universali e Stime di probabilità
- 3 Portafogli Universali di Cover

## Cenni storici

- Shannon, "A Mathematical Theory of Communication", 1948. Articolo rivoluzionario sulla teoria dell'informazione.
- Anni 50-60, applicazione di questi nuovi concetti al mercato (ad esempio Kelly)
- Algoet, Cover, "AOP e AEP for log-optimum investments", 1988.
- Anni 90, Portafoglio universale di Cover.
- Algoet, "Universal Schemes for prediction, gambling and portfolio selection", 1992.

# Sistemi ergodici

## Definizione

Un sistema dinamico misurabile è una quadrupla  $(\Omega, \mathcal{F}, \mathbf{P}, T)$  dove:

- $\Omega$  è un insieme.
- $\mathcal{F}$  è una  $\sigma$ -algebra su  $\Omega$ .
- $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$  è una misura di probabilità.
- $T : \Omega \rightarrow \Omega$  è una trasformazione misurabile che preserva la misura, cioè  $\forall A \in \mathcal{F} \mathbf{P}(T^{-1}(A)) = \mathbf{P}(A)$

## Definizione

Un sistema dinamico misurabile è detto ergodico se per ogni  $A \in \mathcal{F}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq t < n} \chi_A \circ T^t = \mathbf{P}(A) \quad \text{q.c.}$$

# Teorema ergodico generalizzato di Breiman

## Teorema

Sia  $\{g_t\}$  una successione di variabili aleatorie reali definite su uno spazio di probabilità  $(\Omega, \mathcal{F}, \mathbf{P})$  e sia  $T$  una trasformazione ergodica sullo stesso spazio. Se  $g_t \rightarrow g$  q.c. e  $\{g_t\}$  è dominata in  $L^1$ , cioè  $\mathbf{E}[\sup_t |g_t|] < \infty$ , allora

$$\frac{1}{n} \sum_{0 \leq t < n} g_t(T^t(\omega)) \rightarrow \mathbf{E}[g] \quad \text{q.c.} \quad (1)$$

Inoltre, se  $\mathbf{E}[\inf_t g_t] > -\infty$ , allora

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq t < n} g_t(T^t(\omega)) \geq \mathbf{E}[\liminf_{t \rightarrow \infty} g_t] \quad (2)$$

# Il mercato azionario

Il mercato azionario viene modellizzato come:

- $m$  titoli azionari.
- $\mathbf{X} = (X_1, X_2, \dots, X_m)$ ,  $X_i \geq 0, \forall i$  vettore dei rendimenti:  $X_i$  è il rapporto tra i prezzi di chiusura dell' $i$ -esimo titolo tra due periodi di investimento.
- $F(\mathbf{x}), f(\mathbf{x})$ : funzione di ripartizione e densità della distribuzione di  $\mathbf{X}$

Consideriamo quindi un processo stocastico  $\{\mathbf{X}_t\}_{t \in \mathbb{N}, \mathbb{Z}}$  in cui  $\mathbf{X}_t$  è il vettore dei rendimenti al periodo  $t$ .

## Problema

Vogliamo trovare il miglior portafoglio

$\mathbf{b}_t \in \mathcal{B} = \{\mathbf{b} : b_i \geq 0, \sum_i b_i = 1\}, \forall t \in \mathbb{N}$ , ossia il miglior modo di suddividere il capitale ad ogni periodo di investimento per massimizzare il valore atteso del capitale futuro.

Il fattore con cui il capitale aumenta al periodo  $t$  è dato da  $\mathbf{b}_t^T \mathbf{X}_t$ ,  
il capitale dopo  $n$  periodi con  $S_0 = 1$  è  $S_n = \prod_{t=1}^n \mathbf{b}_t^T \mathbf{X}_t$ .

# Tasso di raddoppio ottimale

## Definizione

Il tasso di raddoppio associato a un portafoglio  $\mathbf{b}$  e a una distribuzione  $F(\mathbf{x})$  è definito come

$$W(\mathbf{b}, F) = \int_{\mathbb{R}^m} \log(\mathbf{b}^T \mathbf{X}) dF(\mathbf{x})$$

Il tasso di raddoppio log-ottimale per una distribuzione di probabilità  $F(\mathbf{x})$  è definito come  $W^*(F) = \sup_{\mathbf{b}} W(\mathbf{b}, F)$ .

Un portafoglio che raggiunge l'estremo superiore è detto portafoglio log-ottimale e si indica con  $\mathbf{b}^*$ .



# Proprietà

- $W(\mathbf{b}, F)$  è concavo in  $\mathbf{b}$  e lineare in  $F$ ,  $W^*(F)$  è convesso in  $F$ .
- L'insieme dei portafogli log-ottimali rispetto a una data distribuzione  $F$  è convesso.

## Condizioni di Kuhn-Tucker

Il portafoglio log-ottimale  $\mathbf{b}^*$  per il mercato azionario  $\mathbf{X}$  con distribuzione  $F$  soddisfa le seguenti condizioni necessarie e sufficienti:

$$\begin{aligned} \mathbf{E} \left[ \frac{X_i}{\mathbf{b}^{*T} \mathbf{X}} \right] &= 1 && \text{se } b_i^* > 0 \\ &\leq 1 && \text{se } b_i^* = 0 \end{aligned}$$

► Dimostrazione

# Mercati Stazionari

$\{\mathbf{X}_t\}_{-\infty < t < \infty}$ ,  $t \in \mathbb{Z}$  processo stocastico stazionario che rappresenta i rendimenti azionari degli  $m$  asset.

Ci occupiamo a questo punto di portafogli ottimali che si basino solo sulle informazioni passate disponibili ad un certo periodo.

Queste sono rappresentate da una filtrazione  $\mathcal{F}_t$  data da  $\sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1}) = \mathcal{F}_t$ . Ad ogni passo cerchiamo un portafoglio  $\mathbf{b}_t^* = \mathbf{b}^*(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$  che sia  $\mathcal{F}_t$ -misurabile e che massimizzi la speranza condizionale del miglior tasso di raddoppio rispetto alle informazioni in  $\mathcal{F}_t$ , che indichiamo come

$$w_t^* = \mathbf{E}[\log(\mathbf{b}_t^{*T} \mathbf{X}_t) | \mathcal{F}_t] = \sup_{\mathbf{b} = \mathbf{b}(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})} \mathbf{E}[\log(\mathbf{b}^T \mathbf{X}_t) | \mathcal{F}_t]$$

Per le proprietà della speranza condizionale il suo valore atteso è esattamente il tasso di raddoppio log-ottimale per il periodo di investimento  $n$  tra tutti i portafogli basati unicamente sul passato

$$W_n^* = \mathbf{E}[w_n^*] = \mathbf{E}[\log(\mathbf{b}_n^{*T} \mathbf{X}_n)] = \sup_{\mathbf{b}=\mathbf{b}(\mathbf{X}_0, \dots, \mathbf{X}_{n-1})} \mathbf{E}[\log(\mathbf{b}^T \mathbf{X}_n)]$$

Lo indichiamo anche come  $W^*(\mathbf{X}_n | \mathbf{X}_0, \dots, \mathbf{X}_{n-1})$ .

Cumulativamente, possiamo introdurre il tasso di raddoppio log-ottimale cumulato

$$W^*(\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{n-1}) = \sup_{\mathbf{b}_0, \dots, \mathbf{b}_{n-1}} \mathbf{E}[\log(S_n)] \quad (3)$$

# Chain-Rule e Tasso di raddoppio asintotico

Vale la chain-rule  $W^*(\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{n-1}) = \sum_{t=0}^{n-1} W_t^*$ .

Per processi stocastici, dove esiste il limite, possiamo definire il tasso di raddoppio asintotico come

$$W_{\infty}^* = \lim_{n \rightarrow \infty} \frac{W^*(\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{n-1})}{n}$$

## Lemma

Per un mercato stazionario, il tasso di raddoppio asintotico esiste ed è uguale a

$$W_{\infty}^* = \lim_{n \rightarrow \infty} W^*(\mathbf{X}_n | \mathbf{X}_{n-1}, \dots, \mathbf{X}_0)$$

# Principio di Ottimalità Asintotica

Il prossimo enunciato mostra come una strategia su un processo stazionario che sia log-ottimale ad ogni passo rispetto alle informazioni disponibili in quel periodo di investimento è migliore di tutte le altre strategie che si basano sulle stesse informazioni.

Sia  $\mathbf{P}_t$  una distribuzione di probabilità regolare di  $\mathbf{X}_t$  condizionata a  $\mathcal{F}_t$  e sia quindi  $\mathbf{b}_t^* = \mathbf{b}^*(\mathbf{P}_t)$  un portafoglio  $\mathcal{F}_t$  misurabile che massimizza il ritorno log-ottimale condizionato  $w_t^*$ , il cui valore atteso è il tasso di raddoppio ottimale  $W_t^*$ .

## Teorema (AOP)

Siano i capitali accumulati dopo  $n$  giorni ottenuti applicando il portafoglio  $\mathbf{b}_t^*$  e un qualsiasi altro  $\mathbf{b}_t$  rispettivamente

$$S_n^* = \prod_{0 \leq t < n} \mathbf{b}_t^{*T} \mathbf{X} \quad \text{e} \quad S_n = \prod_{0 \leq t < n} \mathbf{b}_t^T \mathbf{X}$$

Allora  $\left\{ \frac{S_n}{S_n^*}, \mathcal{F}_n \right\}_{0 \leq n < \infty}$  è una supermartingala non negativa convergente quasi certamente a una variabile aleatoria  $Y$  con  $\mathbf{E}[Y] \leq 1$ . Inoltre  $\mathbf{E} \left[ \frac{S_n}{S_n^*} \right] \leq 1$  per ogni  $n$  e

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) \leq 0 \quad \text{quasi certamente}$$

► Salta

# Dimostrazione: parte 1

- Inizialmente  $\frac{S_0}{S_0^*} = 1$ . Il rapporto  $\frac{S_n}{S_n^*} = \prod_{0 \leq i < n} \frac{\mathbf{b}_i^T \mathbf{X}}{\mathbf{b}_i^{*T} \mathbf{X}}$  è  $\mathcal{F}_n$ -misurabile e siccome  $\mathbf{b}_n^*$  è log-ottimale condizionato a  $\mathcal{F}_n$  possiamo scrivere la condizione di Kuhn-Tucker

$$\mathbf{E} \left[ \frac{\mathbf{b}_n^T \mathbf{X}}{\mathbf{b}_n^{*T} \mathbf{X}} \middle| \mathcal{F}_n \right] \leq 1 \quad (4)$$

- Si ottiene subito che

$$\mathbf{E} \left[ \frac{S_{n+1}}{S_{n+1}^*} \middle| \mathcal{F}_n \right] = \mathbf{E} \left[ \frac{S_n}{S_n^*} \frac{\mathbf{b}_n^T \mathbf{X}}{\mathbf{b}_n^{*T} \mathbf{X}} \middle| \mathcal{F}_n \right] = \frac{S_n}{S_n^*} \mathbf{E} \left[ \frac{\mathbf{b}_n^T \mathbf{X}}{\mathbf{b}_n^{*T} \mathbf{X}} \middle| \mathcal{F}_n \right] \leq \frac{S_n}{S_n^*}$$

## Dimostrazione: parte 2

- Il primo punto segue dal teorema di convergenza di Lévy per martingale e dal lemma di Fatou.
- Dalla disuguaglianza di Markov

$$\mathbf{P} \left\{ \frac{S_n}{S_n^*} \geq r_n \right\} \leq \frac{1}{r_n} \mathbf{E} \left[ \frac{S_n}{S_n^*} \right] \leq \frac{1}{r_n}$$

- Scegliendo  $r_n$  in modo che  $\sum_n \frac{1}{r_n} < \infty$  e usando il lemma di Borel-Cantelli otteniamo che

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{S_n}{S_n^*} \right) \leq 0 \quad \text{quasi certamente}$$



# Processi Ergodici e Stazionari

Sia  $\mathbf{X}(\omega) \in \mathbb{R}_+^m$  un vettore di rendimenti e sia  $T$  una trasformazione invertibile e metricamente transitiva che preserva la misura. Sia quindi  $\mathbf{X}_t(\omega) = \mathbf{X}(T^t(\omega))$  e consideriamo il processo su  $\mathbb{Z}$ .

Il fatto che il processo è stazionario ci permette di traslare temporalmente la successione e i portafogli: se  $\mathbf{b}_t^*$  è un portafoglio log-ottimale per il  $t$ -esimo periodo basato sul  $t$ -passato

$\mathcal{F}_t = \sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$  e  $\bar{\mathbf{b}}_t^*$  l'analogo per il giorno 0 con  $t$ -passato  $\bar{\mathcal{F}}_t = T^t \mathcal{F} = \sigma(\mathbf{X}_{-1}, \dots, \mathbf{X}_{-t})$ , otteniamo che

$$\bar{W}_t^* = W^*(\mathbf{X}_0 | \mathbf{X}_{-1}, \dots, \mathbf{X}_{-t}) = W_t^* = W^*(\mathbf{X}_t | \mathbf{X}_{t-1}, \dots, \mathbf{X}_1)$$

Se  $\bar{\mathbf{b}}_{\infty}^*$  è il portafoglio log-ottimale per il periodo zero, basandoci sulle informazioni della  $\sigma$  algebra terminale  $\bar{\mathcal{F}}_{\infty} = \sigma(\mathbf{X}_{-1}, \mathbf{X}_{-2}, \dots)$ , allora  $W_t^* = \bar{W}_t^*$  è debolmente crescente con limite  $\bar{W}_{\infty}^* = \mathbf{E}[\log(\bar{\mathbf{b}}_{\infty}^{*T} \mathbf{X}_0)]$ , che è il tasso di raddoppio ottimale dato il passato infinito  $\bar{W}_{\infty}^* = W^*(\mathbf{X}_0 | \mathbf{X}_{-1}, \mathbf{X}_{-2}, \dots)$ . Quello che vogliamo fare è quindi di utilizzare il fatto che per un processo ergodico le osservazioni sul passato ci indicano l'andamento futuro del processo. Mostriamo quindi che una strategia costituita da portafogli log-ottimali date le informazioni disponibili in quel periodo asintoticamente è equivalente a una strategia che sfrutti una conoscenza totale del processo.

## Teorema (AEP)

Se la successione dei vettori dei rendimenti degli asset  $\{\mathbf{X}_t\}$  è un processo ergodico e stazionario, il capitale cresce in modo esponenziale q.c. con tasso di raddoppio asintotico costante e pari al massimo ritorno log-ottimale atteso dato il passato infinito. In formule

$$\frac{1}{n} \log(S_n^*) \rightarrow \bar{W}_\infty^* = W^*(\mathbf{X}_0 | \mathbf{X}_{-1}, \mathbf{X}_{-2}, \dots) \quad \text{q.c.}$$

dove

$$\begin{aligned} W^*(\mathbf{X}_0 | \mathbf{X}_{-1}, \mathbf{X}_{-2}, \dots) &= \lim_{t \rightarrow \infty} \uparrow W^*(\mathbf{X}_0 | \mathbf{X}_{-1}, \dots, \mathbf{X}_{-t}) \\ &= \lim_{t \rightarrow \infty} \uparrow W^*(\mathbf{X}_t | \mathbf{X}_{t-1}, \dots, \mathbf{X}_0) \\ &= \lim_{n \rightarrow \infty} \uparrow \frac{1}{n} W^*(\mathbf{X}_0, \dots, \mathbf{X}_{n-1}) \end{aligned}$$

# Dimostrazione: parte 1

- La  $\sigma$ -algebra di informazione  $\mathcal{F}_t = \sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$  è approssimata dall'alto da  $\mathcal{F}_t^{(\infty)}$  e dal basso da  $\mathcal{F}_t^{(k)}$ , definiti come

$$\mathcal{F}_t^{(k)} = T^{-t} \bar{\mathcal{F}}_{t \wedge k} = \begin{cases} \sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1}), & \text{se } 0 \leq t < k \\ \sigma(\mathbf{X}_{t-k}, \dots, \mathbf{X}_{t-1}), & \text{se } k \leq t < \infty \end{cases}$$

$$\mathcal{F}_t^{(\infty)} = T^{-t} \bar{\mathcal{F}}_{\infty} = \sigma(\dots, \mathbf{X}_{-1}, \mathbf{X}_0, \dots, \mathbf{X}_{t-1})$$

- Siano  $\mathbf{b}_t^{(k)}$  e  $\mathbf{b}_t^{(\infty)}$  i corrispondenti portafogli log-ottimali e i rispettivi capitali dopo  $n$  periodi

$$S_n^{(k)} = \prod_{0 \leq t < n} \mathbf{b}_t^{(k)T} \mathbf{X}_t \quad \text{e} \quad S_n^{(\infty)} = \prod_{0 \leq t < n} \mathbf{b}_t^{(\infty)T} \mathbf{X}_t$$

# Dimostrazione: parte 2

- $\frac{1}{n} \log \left( S_n^{(k)} \right) = \frac{1}{n} \log(S_k^*) + \frac{1}{n} \sum_{k \leq t < n} \log(\mathbf{b}_t^{(k)T} \mathbf{X}_t)$  e dal teorema ergodico (con  $g_t = \log \mathbf{b}_t^{(k)T} \mathbf{X}_t$ ) segue che

$$\frac{1}{n} \log S_n^{(k)} \rightarrow W_k^* = \mathbf{E}[\log(\mathbf{b}_k^{*T} \mathbf{X}_k)] \quad \text{q.c.}$$

- Analogamente segue che

$$\frac{1}{n} \log \left( S_n^{(\infty)} \right) = \frac{1}{n} \sum_{0 \leq t < n} \log(\mathbf{b}_t^{(\infty)T} \mathbf{X}_t) \rightarrow \bar{W}_\infty^* \quad \text{q.c.}$$

## Dimostrazione: parte 3

- $\mathcal{F}_t^{(k)} \subseteq \mathcal{F}_t$  e quindi  $\mathbf{b}_t^{(k)}$  è  $\mathcal{F}_t$ -misurabile, così come  $\mathcal{F}_t \subseteq \mathcal{F}_t^{(\infty)}$  e  $\mathbf{b}_t^*$  è  $\mathcal{F}_t^{(\infty)}$ -misurabile.
- Dall'AOP segue che  $\limsup \frac{1}{n} \log \left( \frac{S_n^{(k)}}{S_n^*} \right) \leq 0$  e  
 $\limsup \frac{1}{n} \log \left( \frac{S_n^*}{S_n^{(\infty)}} \right) \leq 0$  q.c.
- otteniamo la catena di disuguaglianze

$$\begin{aligned} W_k^* &= \lim_{n \rightarrow \infty} \frac{1}{n} \log(S_n^{(k)}) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log(S_n^*) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log(S_n^*) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log(S_n^{(\infty)}) = \bar{W}_\infty^* \quad \text{q.c.} \end{aligned}$$

e la tesi segue dal fatto che  $W_k^* = \bar{W}_k^* \uparrow \bar{W}_\infty^*$  per  $k \rightarrow \infty$ .

# Portafogli universali e stime di probabilità

A questo punto mostriamo alcuni approcci pratici per trovare strategie che asintoticamente hanno un tasso di raddoppio log-ottimale pari a  $W_{\infty}^*$ . Per farlo sono necessari due passaggi.

## Problema

- Trovare metodi di stima della distribuzione di probabilità per un processo stocastico e ergodico  $\mathbf{X}_t$  che modella il mercato che convergano a quella effettiva.
- Trovare il portafoglio log-ottimale per queste distribuzioni.

# Portafogli: approssimazioni

Per trovare il portafoglio log-ottimale data una distribuzione si possono usare le seguenti strategie:

- Approssimazione al primo ordine: consiste nell'investire tutto sul titolo con valore atteso maggiore poiché al primo ordine è la strategia migliore. Non tiene conto delle incertezze ed è quindi rischioso.
- Approssimazione al secondo ordine (criterio di Kelly): è ampiamente descritto in [Tho06] e tiene conto delle incertezze e covarianze.
- Metodi ad hoc: ad esempio il Nearest Neighbour in [GUW08]



## Stime di Probabilità

Il teorema generalizzato di Breiman ci permette di dimostrare il seguente enunciato che si trova in [Alg92]

### Teorema

Sia  $\mathbf{X}_t$  un processo ergodico e stazionario, siano quindi  $\mathbf{b}_t^*$  e  $\hat{\mathbf{b}}_t^*$  i portafogli log-ottimali rispetto alla vera distribuzione condizionata fino al tempo  $t$   $\mathbf{P}(\mathrm{d}\mathbf{x}_t|\mathbf{X}^t)$  e da una stima  $\hat{\mathbf{P}}(\mathrm{d}\mathbf{x}_t|\mathbf{X}^t)$  tale che (grazie all'invarianza per traslazione temporale)

$$\hat{\mathbf{P}}(\mathrm{d}\mathbf{x}_0|\mathbf{X}^{-t}) \rightarrow \mathbf{P}(\mathrm{d}\mathbf{x}_0|\mathbf{X}^{-}) \quad \text{debolmente q.c.}$$

Se il mercato è sicuro, ossia se  $\mathbf{E}[\log(X_{t,j})] > -\infty \quad \forall t, 1 \leq j \leq m$ , allora

$$\frac{1}{n} \log \hat{S}_n^* \rightarrow W(\mathbf{X}|\mathbf{X}^{-}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log S_n^*$$

Con qualche accorgimento l'ipotesi di mercato sicuro può essere tolta. Si considerano portafogli log-ottimali leggermente spostati nel semplice in modo che investano frazioni non nulle su ogni titolo. Se questo spostamento tende a 0, si ottiene il medesimo risultato al limite.

Vediamo ora alcuni metodi possibili per trovare queste stime di probabilità consistenti. Utilizzando il teorema ergodico di Breiman si può dimostrare che queste convergono alle distribuzioni effettive.

# L'importanza dei codici

- Lo stretto legame tra teoria dell'informazione e il mondo della finanza e delle scommesse risulta evidente confrontando i concetti di entropia (o tasso di entropia) e tasso di raddoppio.
- L'entropia di Shannon ( $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$ ) è intuitivamente la quantità di informazione media fornita da un'osservazione della variabile aleatoria  $X$ .
- La proprietà di equipartizione asintotica per le codifiche [CT06] garantisce che per abbastanza osservazioni di un processo stocastico, quasi tutte le stringhe di valori appartengono a un insieme detto insieme tipico.
- Una conseguenza di questo fatto è la possibilità di avere compressione di dati. A lungo andare le stringhe sono prevedibili e appartengono a questo insieme limitato.

# L'importanza dei codici

- Le codifiche e gli algoritmi di compressione sono ottimi metodi per ottenere stime di probabilità e sono concetti estremamente collegati agli investimenti e al gambling.
- Consideriamo ad esempio il caso particolare del gambling sulle corse di cavalli: abbiamo  $m$  cavalli e ad ogni turno dividiamo il capitale scommettendolo su di essi. In questo caso le variabili aleatorie sono discrete, possiamo utilizzare molti dei concetti visti per il mercato adattandoli a questo.
- Grazie all'AEP, i processi  $X_t$  che indicano il cavallo vincente al tempo  $t$  saranno codificabili e comprimibili con rapporto di compressione ottimale  $H(\mathcal{X})$ . Più il tasso di entropia è basso, più ho prevedibilità e maggiore è la mia possibilità di guadagno.
- Vale la legge di conservazione:  $W_\infty^* + H(\mathcal{X}) = \log m$ . [► Dettagli](#)

## Metodo delle frequenze Empiriche

- Approssimazione del processo a una  $k$ -catena di Markov.
- Si costruiscono degli stati basati sui rendimenti, la probabilità di finire uno stato dato i  $k$  precedenti viene costruito sulla base delle frequenze empiriche osservate, aggiungendo un peso a priori dato a tutte le transizioni.
- Se indichiamo questi stati con dei simboli, i  $k$  stati passati con  $J^{-k}$  e il successivo con  $j$  scriviamo

$$\hat{P}_t(j|J^{-k}) = \frac{1 + c_t(j|J^{-k})}{m + c_t(J^{-k})}$$

dove  $c_t(j|J^{-k})$  è il numero di occorrenze fino al tempo  $t$  della transizione allo stato  $j$  a partire dal contesto  $J^{-k}$  e  $c_t(J^{-k})$  sono le occorrenze del contesto  $J^{-k}$ .

# Algoritmo di Lempel-Ziv

- Algoritmo di compressione descritto in [LZ78]. Comprime in modo ottimale una sorgente di dati, ossia con rapporto di compressione pari all'entropia della distribuzione dei caratteri.
- La successione viene suddivisa in frasi inserendo virgole tra i simboli, in modo che ogni frase consiste di una stringa di lunghezza massima possibile che è già comparsa in una frase precedente seguita da un simbolo "nuovo" e un'altra virgola.
- Questo algoritmo si presta a essere realizzato con una struttura ad albero incrementale.
- Esempio: 0101000100 che ha lunghezza  $n = 10$  genera , 0, 1, 01, 00, 010, 0 e contiene  $\nu_n = 5$  frasi complete.

- Ogni nodo ha come figli possibili gli  $m$  stati o simboli di un alfabeto. Inizialmente consiste di un solo nodo, la radice.
- La ricerca di una nuova frase inizia dalla radice: ci muoviamo all'interno dell'albero leggendo i simboli dati dalla sorgente.
- Quando ci muoviamo in una direzione mai visitata prima, aggiungiamo il nodo e ricominciamo dalla radice.
- Sia  $T_n$  l'albero al passo  $n$  e sia  $\bar{T}_n$  l'albero completato aggiungendo i nodi mai visitati adiacenti a quelli di  $T_n$ .
- Sia  $z_t = f(j^t)$  il nodo di  $T_t$  dove l'algoritmo di ricerca è arrivato dopo aver letto  $(j_0, \dots, j_{t-1})$  e prima di leggere  $j_t$ . Definiamo quindi

$$c_n(j|z) = \#\{t : 0 \leq t < n, (j_t, z_t) = (j, z)\}$$
$$c_n(z) = \sum_{1 \leq j \leq m} c_n(j|z) = \#\{t : 0 \leq t < n, z_t = z\}$$

- $c_n(j|z)$  è il numero di passaggi durante tutte le ricerche dal nodo  $z$  al figlio  $j$  e  $c_n(z)$  è il numero di volte in cui siamo passati per il nodo  $z$ .
- $1 + c_n(z)$  è il numero di nodi nel sotto-albero radicato in  $z$ , a meno che  $z$  sia un predecessore di  $z_n$ , l'ultimo nodo visitato, nel qual caso  $c_n(z)$  è il numero di nodi nel sotto-albero.
- Contando anche le foglie dell'albero completato  $\bar{T}_n$ , definiamo

$$\begin{aligned}\gamma_n(z_n) &= 1 + (m - 1)(c_n(z_n) + 1) \\ \gamma_n(j|z_n) &= 1 + (m - 1)(c_n(j|z_n) + 1) \\ \hat{P}(j|j^n) &= \frac{\gamma_n(j|z_n)}{\gamma_n(z_n)} = \frac{m + (m - 1)c_n(j|z_n)}{m + (m - 1)c_n(z_n)}\end{aligned}$$



## Metodi ad Hoc: Rapporti rendimento-incertezza

- Per ogni periodo di investimento  $t$ , guardiamo gli ultimi  $\tau$  vettori di log-rendimento  $\log(\mathbf{X}_{t-\tau}), \dots, \log(\mathbf{X}_{t-1})$  e calcoliamo media e deviazione standard empirica, rispettivamente  $\hat{\mu}$  e  $\hat{\sigma}$ .
- Calcoliamo  $\frac{\log(X_{t-1,j}) - \hat{\mu}_j}{\hat{\sigma}_j}$  e, dopo aver definito degli stati, cerchiamo a quale appartiene.
- Per ognuno di questi stati consideriamo le loro occorrenze passate e i rendimenti di quando ci trovavamo in essi.
- Utilizziamo i rendimenti storici collegati allo stato attuale e usiamo questi per trovare una stima del rendimento futuro, con le approssimazioni citate prima.

## Metodi ad Hoc: Nearest Neighbour

Il metodo è ampiamente descritto in [GUW08]. Qui viene mostrata una versione leggermente modificata.

- Due parametri:  $k$ , la dimensione della finestra che teniamo in considerazione e  $s_I$  ossia la lunghezza di ogni segmento nel passato in cui cerchiamo il vicino a distanza minore dalla finestra attuale.
- Al periodo di investimento  $n > s_I + 1$ , per ogni  $1 \leq i \leq \lfloor \frac{n-1}{s_I} \rfloor$ , definiamo quindi  $N_i = \arg \min_{(i-1)s_I + k \leq j < is_I} \|\mathbf{X}_{j-k}^{j-1} - \mathbf{X}_{n-k}^{n-1}\|$ . Questa ricerca viene fatta solo nei 10 anni appena precedenti al periodo attuale.
- Scegliamo il portafoglio con i dati storici di questi  $N_i$  approssimandoli come sopra o come
$$\hat{\mathbf{b}} = \arg \max_{\mathbf{b} \in \mathcal{B}} \prod_{1 \leq i \leq \lfloor \frac{n-1}{s_I} \rfloor} \mathbf{b}^T \mathbf{X}_{N_i}$$

## Prove e risultati sperimentali

- Sono stati testati i metodi descritti in precedenza sui dati reali di mercato.
- Per il metodo empirico sui rendimenti, l'algoritmo di Lempel-Ziv e il metodo sui rapporti rendimento-incertezza sono stati usati i dati storici dei titoli IBM e Coca-Cola (KO).
- Per il metodo Nearest Neighbour sono stati usati i dati storici di IBM, Coca-Cola (KO) e General Electric (GE).
- In tutti i casi il periodo considerato è di circa 14600 giorni (oppure circa 3000 settimane o 700 mesi) e il portafoglio iniziale  $S_0 = 1$ .

# Metodo empirico sui rendimenti

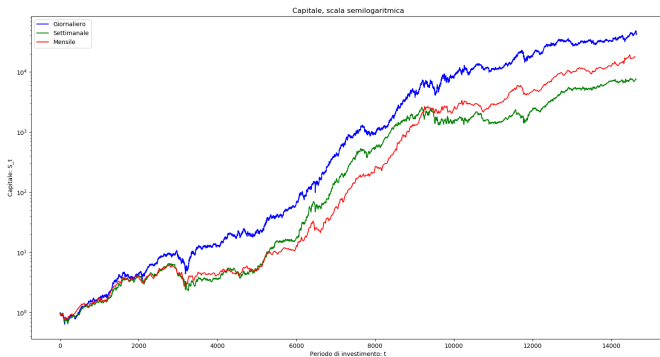


Figura:

## Metodo empirico sui rendimenti

Quantità	Giornaliero	Settimanale	Mensile
$S_n$	43025.29	7584.782	18071.66
$\mu_{\text{ann}}$	21.70%	18.56%	19.89%
$\sigma_{\text{ann}}$	0.2460	0.2394	0.2243
$\mu_{\text{geom}}$	0.0007291	0.003057	0.0141831
$\max dd$	59.20%	64.83%	55.25%
$CR$	0.0229	0.0179	0.0225
$SR$	0.8820	0.7754	0.8867

# Algoritmo di Lempel-Ziv

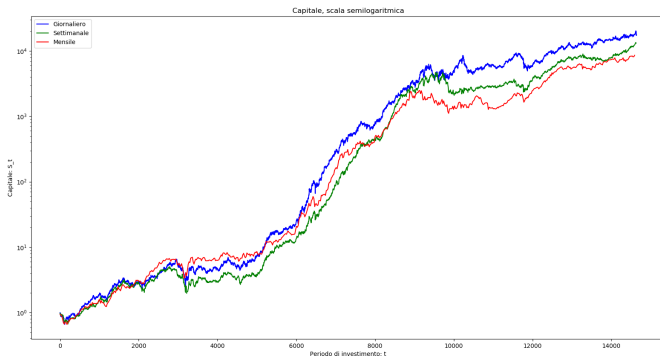


Figura:

# Algoritmo di Lempel-Ziv

Quantità	Giornaliero	Settimanale	Mensile
$S_n$	17806.67	13149.88	8605.029
$\mu_{\text{ann}}$	20.12%	19.32%	18.71%
$\sigma_{\text{ann}}$	0.2446	0.2303	0.2292
$\mu_{\text{geom}}$	0.0006688	0.0032455	0.0131025
$\max dd$	62.20%	60.89%	55.40%
$CR$	0.0202	0.0198	0.0211
$SR$	0.8226	0.8387	0.8164

# Metodo dei rapporti rendimento-incertezza

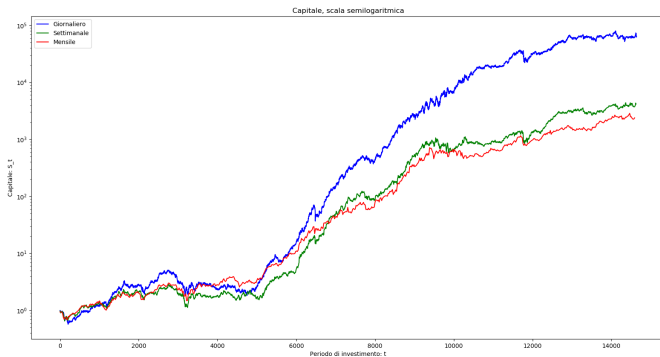


Figura:



## Metodo dei rapporti rendimento-incertezza

Quantità	Giornaliero	Settimanale	Mensile
$S_n$	64102.81	4166.649	2384.919
$\mu_{\text{ann}}$	22.37%	17.44%	16.59%
$\sigma_{\text{ann}}$	0.2453	0.2371	0.2357
$\mu_{\text{geom}}$	0.0007563	0.0028516	0.0011236
$\max dd$	63.64%	60.84%	50.82%
$CR$	0.0220	0.0179	0.0204
$SR$	0.9119	0.7356	0.7039

# Metodo Nearest Neighbour

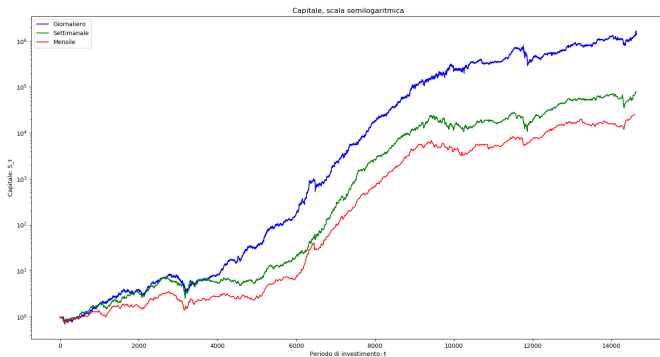


Figura:

## Metodo Nearest Neighbour

Quantità	Giornaliero	Settimanale	Mensile
$S_n$	1485001	77065.69	25047.91
$\mu_{\text{ann}}$	28.13%	22.97%	20.65%
$\sigma_{\text{ann}}$	0.2564	0.2541	0.2327
$\mu_{\text{geom}}$	0.0009712	0.0098517	0.0146589
$\max dd$	64.46%	66.20%	59.75%
$CR$	0.0272	0.0217	0.0216
$SR$	1.0975	0.9404	0.8876

# Portafogli universali di Cover

I portafogli universali di Cover sono metodologie che sul lungo periodo hanno lo stesso rendimento asintotico del miglior portafoglio costantemente ribilanciato, ossia che ad ogni periodo mantiene costante la parte di capitale investita in ogni titolo. Il primo che presentiamo fornisce una strategia dipendente dal numero  $n$  di periodi di investimento considerati, la seconda invece non presenta tale limite. In entrambi i casi però, non viene fatta nessuna assunzione sulla distribuzione di probabilità che è completamente ignota.

# Portafoglio universale a orizzonte finito

## Teorema

Per una successione di vettori di stock  $\mathbf{x}^n = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , con  $\mathbf{x}_i \in \mathbb{R}_+^m$  di lunghezza  $n$  e  $m$  asset, sia  $S_n^*(\mathbf{x}^n)$  il capitale ottenuto miglior portafoglio costantemente ribilanciato su  $\mathbf{x}^n$  e sia  $\hat{S}_n(\mathbf{x}^n)$  quello derivante da una qualsiasi strategia  $\hat{\mathbf{b}}_i(\cdot)$  su  $\mathbf{x}^n$ . Allora

$$\max_{\hat{\mathbf{b}}_i(\cdot)} \min_{\mathbf{x}_1, \dots, \mathbf{x}_n} \frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} = V_n$$

dove

$$V_n = \left[ \sum_{n_1+n_2+\dots+n_m=n} \binom{n}{n_1, n_2, \dots, n_m} 2^{-nH(\frac{n_1}{n}, \dots, \frac{n_m}{n})} \right]^{-1} \approx n^{-\frac{m-1}{2}}$$

# Dimostrazione: Parte 1

Dimostriamo il teorema nel caso  $m = 2$  asset.

- Scriviamo  $S_n(\mathbf{x}^n) = \prod_{i=1}^n \mathbf{b}_i^T \mathbf{x}_i$  che è un prodotto di somme come una somma di prodotti. Espandendo infatti otteniamo

$$S_n(\mathbf{x}^n) = \sum_{j^n \in \{1,2\}^n} \prod_{i=1}^n b_{ij_i} x_{ij_i} = \sum_{j^n \in \{1,2\}^n} \prod_{i=1}^n b_{ij_i} \prod_{i=1}^n x_{ij_i}$$

dove  $b_{ij_i}$  è la frazione di capitale investita il giorno  $i$  sul titolo  $j_i \in \{1, 2\}$ .

- Definiamo  $x(j^n) = \prod_{i=1}^n x_{ij_i}$  e  $w(j^n) = \prod_{i=1}^n b_{ij_i}$ , da cui  $S_n(\mathbf{x}^n) = \sum_{j^n \in \{1,2\}^n} w(j^n) x(j^n)$ .

## Dimostrazione: parte 2

- Applichiamo le stesse notazioni a  $S_n^*$  e otteniamo che
$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} = \frac{\sum_{j^n \in \{1,2\}^n} \hat{w}(j^n) x(j^n)}{\sum_{j^n \in \{1,2\}^n} w^*(j^n) x(j^n)}.$$
- Applichiamo il seguente lemma:

Dati numeri reali  $p_1, p_2, \dots, p_m \geq 0$  e  $q_1, q_2, \dots, q_m \geq 0$  vale

$$\frac{\sum_{i=1}^m p_i}{\sum_{i=1}^m q_i} \geq \min_i \frac{p_i}{q_i}$$

dove poniamo che  $\frac{x}{0} = \infty$ .

da cui otteniamo che

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq \min_{j^n} \frac{\hat{w}(j^n) x(j^n)}{w^*(j^n) x(j^n)} = \min_{j^n} \frac{\hat{w}(j^n)}{w^*(j^n)} \quad (5)$$

## Dimostrazione: Parte 3

- La miglior strategia sulle successioni  $j^n$  corrisponde a quella che pone pesi proporzionali a quelli posti dal miglior portafoglio costantemente ribilanciato  $w^*(j^n)$ .
- $w^*(j^n) = \max_{0 \leq b \leq 1} b^{k(j^n)}(1-b)^{n-k(j^n)} = \left(\frac{k(j^n)}{n}\right)^{k(j^n)} \left(\frac{n-k(j^n)}{n}\right)^{n-k(j^n)}$  con portafoglio  $\mathbf{b}^* = \left(\frac{k}{n}, \frac{n-k}{n}\right)$  dove  $k(j^n)$  è il numero di 1 in  $j^n$ .
- Dobbiamo rinormalizzare poiché questi non costituiscono una distribuzione di probabilità. Definiamo quindi

$$\begin{aligned} \frac{1}{V_n} &= \sum_{j^n \in \{1,2\}^n} \left(\frac{k(j^n)}{n}\right)^{k(j^n)} \left(\frac{n-k(j^n)}{n}\right)^{n-k(j^n)} \\ &= \sum_{k=0}^n \binom{n}{k} \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} \end{aligned}$$



## Dimostrazione: Parte 4

- Definiamo inoltre

$$\hat{w}(j^n) = V_n \left( \frac{k(j^n)}{n} \right)^{k(j^n)} \left( \frac{n - k(j^n)}{n} \right)^{n - k(j^n)} \quad (6)$$

dove  $k(j^n)$  è il numero di volte in cui 1 compare in  $j^n$ .

- Dalla disequazione 5 segue che

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq \min_k \frac{V_n \left( \frac{k}{n} \right)^k \left( \frac{n-k}{n} \right)^{n-k}}{(b^*)^k (1 - b^*)^{n-k}} \geq V_n$$

- Abbiamo quindi dimostrato che esiste una strategia con limite inferiore sul rapporto  $V_n$ . Dimostriamo ora che non possiamo fare di meglio.

## Dimostrazione: Parte 5

- Chiamiamo vettori di rendimenti estremali quelli della forma

$$\mathbf{x}_i(j_i) = \begin{cases} (1, 0)^t & \text{se } j_i = 1, \\ (0, 1)^t & \text{se } j_i = 2, \end{cases}$$

e sia  $\mathcal{K}$  l'insieme di questi.

- Per loro definizione, ad ogni periodo, un solo asset ha ritorno non nullo, ne segue che  $S_n(\mathbf{x}^n(j^n)) = \prod_i b_{j_i} = w(j^n)$ , da cui inoltre  $\sum_{j^n \in \{1,2\}^n} S_n(\mathbf{x}^n(j^n)) = 1$ .
- Si ha che  $\mathbf{b}^*(\mathbf{x}^n(j^n)) = \left( \frac{n_1(j^n)}{n}, \frac{n_2(j^n)}{n} \right)^t$  che ottiene un capitale

$$S_n^*(\mathbf{x}^n(j^n)) = \left( \frac{n_1(j^n)}{n} \right)^{n_1(j^n)} \left( \frac{n_2(j^n)}{n} \right)^{n_2(j^n)} = \frac{\hat{w}(j^n)}{V_n}$$

## Dimostrazione: Parte 6

- Si ha quindi che

$$\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n) = \frac{1}{V_n} \sum_{j^n \in \{1,2\}^n} \hat{w}(j^n) = \frac{1}{V_n}$$

- Concludiamo con

$$\begin{aligned} \min_{\mathbf{x}^n \in \mathcal{K}} \frac{S_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} &\leq \sum_{\tilde{\mathbf{x}}^n \in \mathcal{K}} \frac{S_n^*(\tilde{\mathbf{x}}^n)}{\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n)} \frac{S_n(\tilde{\mathbf{x}}^n)}{S_n^*(\tilde{\mathbf{x}}^n)} \\ &= \sum_{\tilde{\mathbf{x}}^n \in \mathcal{K}} \frac{S_n(\tilde{\mathbf{x}}^n)}{\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n)} \\ &= \frac{1}{\sum_{\mathbf{x}^n \in \mathcal{K}} S_n^*(\mathbf{x}^n)} \\ &= V_n \end{aligned}$$

# Osservazioni

- I risultati in [OC98] indicano che

$$V_n \sim \left( \sqrt{\frac{2}{n}} \right)^{m-1} \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{\pi}}$$

da cui  $\frac{1}{2\sqrt{n+1}} \leq V_n \leq \frac{2}{\sqrt{n+1}}$  per  $m = 2$ .

- Possiamo riscrivere il portafoglio in modo incrementale come

$$\hat{b}_{i,1}(\mathbf{x}^{i-1}) = \frac{\sum_{j^{i-1} \in M^{i-1}} \hat{w}(j^{i-1}) x(j^{i-1})}{\sum_{j^i \in M^i} \hat{w}(j^i) x(j^{i-1})}$$

dove  $\hat{w}(j^i) = \sum_{j^n: j^i \subseteq j^n} w(j^n)$  e  $x(j^{i-1}) = \prod_{k=1}^{i-1} x_{kj_k}$

## Portafoglio Universale a orizzonte infinito

Questa seconda strategia è indipendente da  $n$ . Idealmente corrisponde a mescolare insieme le strategie di infiniti manager ognuno con un proprio portafoglio costantemente ribilanciato. Sia  $\mathcal{B} = \{\mathbf{b} \in \mathbb{R}^m : b_i \geq 0, \sum_i b_i = 1\}$  l' $m$  semplice, ossia lo spazio dei portafogli ammissibili. Daremo quindi un certo peso a ogni investitore, rappresentato da una misura su  $\mathcal{B}$   $\mu(\mathbf{b})$ , dove  $d\mu(\mathbf{b})$  sarà la quantità di capitale investita nei portafogli nell'intorno  $d\mathbf{b}$  del portafoglio  $\mathbf{b}$ .

Definiamo  $S_n(\mathbf{b}, \mathbf{x}^n) = \prod_{i=1}^n \mathbf{b}^T \mathbf{x}_i$  e consideriamo il portafoglio

$$\hat{b}_{i+1}(\mathbf{x}^i) = \frac{\int_{\mathcal{B}} \mathbf{b} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})} \quad (7)$$

## Osservazione

$$\hat{\mathbf{b}}_{i+1}^T(\mathbf{x}^i)\mathbf{x}_{i+1} = \frac{\int_{\mathcal{B}} \mathbf{b}^T \mathbf{x}_{i+1} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})} = \frac{\int_{\mathcal{B}} S_{i+1}(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})}$$

Essendo questo prodotto telescopico otteniamo

$$\hat{S}_n(\mathbf{x}^n) = \prod_{i=1}^n \hat{\mathbf{b}}_i^T(\mathbf{x}^{i-1})\mathbf{x}_i = \int_{\mathbf{b} \in \mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n) d\mu(\mathbf{b})$$

Il capitale finale è quindi una media dei vari capitali.

Ricordiamo la distribuzione di Dirichlet $(\frac{1}{2}, \frac{1}{2})$

$$d\mu(\mathbf{b}) = \frac{\Gamma(\frac{m}{2})}{[\Gamma(\frac{1}{2})]^m} \prod_{j=1}^m b_j^{-\frac{1}{2}} d\mathbf{b}$$

Sia  $S_n^*(\mathbf{x}^n)$  il capitale ottenuto usando il miglior portafoglio costantemente ribilanciato e sia  $\hat{S}_n(\mathbf{x}^n)$  quello dato dal portafoglio misto  $\hat{\mathbf{b}}(\cdot)$ , allora

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq \min_{j^n} \frac{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})}{\prod_{i=1}^n b_{j_i}^*}$$

Per il portafoglio universale  $\hat{\mathbf{b}}_i(\cdot)$ ,  $i = 1, 2, \dots$ , con  $m = 2$  stock e  $d\mu(\mathbf{b})$  la distribuzione Dirichlet $(\frac{1}{2}, \frac{1}{2})$ , si ha

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq \frac{1}{2\sqrt{n+1}}$$

per ogni  $n$  e ogni successione  $x^n$ .

# Osservazioni

- $\frac{\hat{S}_n}{S_n^*} \geq \frac{1}{\sqrt{2\pi}} V_n$ . Rendere l'orizzonte infinito è costato solo un fattore  $\frac{1}{\sqrt{2\pi}}$ .
- Asintoticamente  

$$\frac{1}{n} \ln(\hat{S}_n(\mathbf{x}^n)) - \frac{1}{n} \ln(S_n^*(\mathbf{x}^n)) \geq \frac{1}{n} \ln\left(\frac{V_n}{\sqrt{2\pi}}\right) \rightarrow 0.$$
- La scelta della distribuzione di Dirichlet è dovuta al fatto che pone molto peso sui portafogli estremali. Al primo ordine, infatti, scrivendo  $\mathbf{x} = \mathbf{e} + \mathbf{r}$  con  $\mathbf{e} = (1, \dots, 1)^t$ , si ha che

$$\log(\mathbf{b}^t \mathbf{x}) = \log(1 + \mathbf{b}^t \mathbf{r}) = \mathbf{b}^t \mathbf{r} + o(\mathbf{b}^t \mathbf{r})$$

e conviene quindi puntare tutto sull'asset che assicura rendimento maggiore.



# Portafoglio Universale di Cover

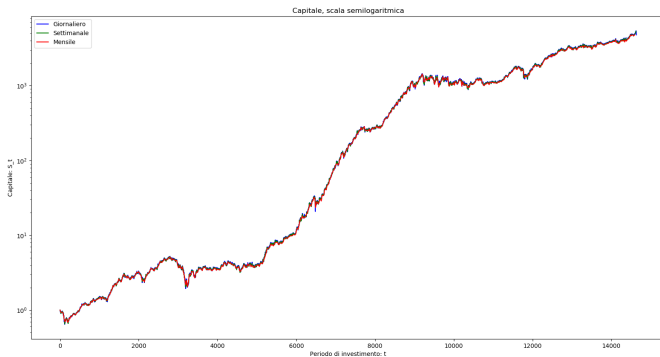


Figura:

# Portafoglio Universale di Cover

Quantità	Giornaliero	Settimanale	Mensile
$S_n$	4764.502	5324.182	4777.275
$\mu_{\text{ann}}$	16.93%	17.04%	16.84%
$\sigma_{\text{ann}}$	0.2058	0.2003	0.1942
$\mu_{\text{geom}}$	0.0005786	0.0029356	0.0122462
$\max dd$	63.01%	60.62%	59.91%
$CR$	0.0168	0.0176	0.0176
$SR$	0.8225	0.8505	0.8669



Paul H. Algoet and Thomas M. Cover.

Asymptotic optimality and asymptotic equipartition properties of log-optimum investment.

*Ann. Probab.*, 16(2):876–898, 04 1988.



Paul Algoet.

Universal schemes for prediction, gambling and portfolio selection.

*Ann. Probab.*, 20(2):901–941, 04 1992.



T. M. Cover and E. Ordentlich.

Universal portfolios with side information.

*IEEE Transactions on Information Theory*, 42(2):348–363, March 1996.



Thomas M. Cover and Joy A. Thomas.

*Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing).*

Wiley-Interscience, USA, 2006.



László Györfi, Frederic Udina, and Harro Walk.

Experiments on universal portfolio selection using data from real markets.

02 2008.



A. Klenke.

*Probability Theory: A Comprehensive Course.*

Universitext. Springer London, 2013.



A. Lempel and J. Ziv.

Compression of individual sequences via variable-rate coding.

*IEEE Transactions on Information Theory*, 24(5):530–536, Sep. 1978.



A. C. G. Mennucci and S. K. Mitter.

*Probabilità e informazione.*

Edizioni della Normale, Pisa, Italia, 2008.



Erik Ordentlich and Thomas M. Cover.

The cost of achieving the best portfolio in hindsight.

*Mathematics of Operations Research*, 23(4):960–982, 1998.



E. O. Thorp.

The kelly criterion in blackjack, sports betting and the stock market.

*Handbook of Asset and Liability Management*, 1:385–428, 2006.

# Grazie per l'attenzione

## Dimostrazione: parte 1

Dalla concavità in  $\mathbf{b}$  del tasso di raddoppio  $\mathbf{b}^*$  è log-ottimale  $\iff$  la derivata direzionale in qualsiasi direzione è  $\leq 0$

$$\begin{aligned}\left. \frac{d}{d\lambda} \mathbf{E}[\ln(\mathbf{b}_\lambda^T \mathbf{X})] \right|_{\lambda=0+} &= \lim_{\lambda \downarrow 0} \frac{1}{\lambda} \mathbf{E} \left[ \ln \left( \frac{(1-\lambda)\mathbf{b}^{*T} \mathbf{X} + \lambda \mathbf{b}^T \mathbf{X}}{\mathbf{b}^{*T} \mathbf{X}} \right) \right] \\ &= \mathbf{E} \left[ \lim_{\lambda \downarrow 0} \frac{1}{\lambda} \ln \left( 1 + \lambda \left( \frac{\mathbf{b}^T \mathbf{X}}{\mathbf{b}^{*T} \mathbf{X}} - 1 \right) \right) \right] \\ &= \mathbf{E} \left[ \frac{\mathbf{b}^T \mathbf{X}}{\mathbf{b}^{*T} \mathbf{X}} \right] - 1\end{aligned}$$

Il secondo passaggio segue dalla convergenza dominata con  $g(\mathbf{X}) = 2$ .

## Dimostrazione: parte 2

La condizione si riscrive come

$$\mathbb{E} \left[ \frac{\mathbf{b}^T \mathbf{X}}{\mathbf{b}^{*T} \mathbf{X}} \right] \leq 1$$

per ogni  $\mathbf{b} \in \mathcal{B}$ . Se il segmento tra  $\mathbf{b}^*$  e  $\mathbf{b}$  può essere esteso oltre  $\mathbf{b}^*$  allora abbiamo uguaglianza nella formula, altrimenti no.

Consideriamo i  $\mathbf{b}_j = (\mathbf{b} : b_j = 1, b_i = 0, i \neq j)$ . In questo caso il segmento tra  $\mathbf{b}^*$  e  $\mathbf{b}_j$  può essere esteso se e solo se  $b_j^* > 0$ . Da questo si ottiene la tesi

◀ Indietro



## Dimostrazione: Parte 1

- In modo completamente analogo a prima possiamo scrivere

$$S_n^*(\mathbf{x}^n) = \sum_{j^n \in \{1, \dots, m\}^n} w^*(j^n) x(j^n)$$

- Similmente scriviamo

$$\begin{aligned} \hat{S}_n(\mathbf{x}^n) &= \int_{\mathcal{B}} \prod_{i=1}^n \mathbf{b}^T \mathbf{x}_i d\mu(\mathbf{b}) \\ &= \sum_{j^n \in \{1, \dots, m\}^n} \int_{\mathcal{B}} \prod_{i=1}^n b_{j_i} x_{ij_i} d\mu(\mathbf{b}) \\ &= \sum_{j^n \in \{1, \dots, m\}^n} \hat{w}(j^n) x(j^n) \end{aligned}$$

## Dimostrazione: Parte 2

- Applicando ora il lemma precedente abbiamo

$$\begin{aligned}\frac{\hat{S}(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} &= \frac{\sum_{j^n} \hat{w}(j^n) x(j^n)}{\sum_{j^n} w^*(j^n) x(j^n)} \\ &\geq \min_{j^n} \frac{\hat{w}(j^n) x(j^n)}{w^*(j^n) x(j^n)} \\ &= \min_{j^n} \frac{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})}{\prod_{i=1}^n b_{j_i}^*}\end{aligned}$$

- Come prima  $\prod_{i=1}^n b_{j_i}^* = \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k} = 2^{-nH\left(\frac{k}{n}\right)}$
- La distribuzione di Dirichlet  $\left(\frac{1}{2}\right)$  è definita come

$$d\mu(\mathbf{b}) = \frac{\Gamma\left(\frac{m}{2}\right)}{\left[\Gamma\left(\frac{1}{2}\right)\right]^m} \prod_{j=1}^m b_j^{-\frac{1}{2}} d\mathbf{b}$$

## Dimostrazione: Parte 3

- Nel caso  $m = 2$ ,  $d\mu(b) = \frac{1}{\pi} \frac{1}{\sqrt{b(1-b)}} db$ ,  $0 \leq b \leq 1$  e  
 $b(j^n) = \prod_{i=1}^n b_{j_i} = b^l (1-b)^{n-l}$  con  $l$  numero di volte in cui  $j_i = 1$ .
- Possiamo calcolare esplicitamente

$$\begin{aligned} \int_B b(j^n) d\mu(\mathbf{b}) &= \int_0^1 b^l (1-b)^{n-l} \frac{1}{\pi} \frac{1}{\sqrt{b(1-b)}} db \\ &= \frac{1}{\pi} \int_0^1 b^{l-\frac{1}{2}} (1-b)^{n-l-\frac{1}{2}} db \\ &= \frac{1}{\pi} B\left(l + \frac{1}{2}, n - l + \frac{1}{2}\right) \end{aligned}$$



$X$  e  $Y$  variabili aleatorie discrete,  $p(x)$ ,  $p(x, y)$ ,  $p(y|x)$  distribuzioni.

## Entropia di Shannon

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

$X$  e  $Y$  variabili aleatorie discrete,  $p(x)$ ,  $p(x, y)$ ,  $p(y|x)$  distribuzioni.

## Entropia di Shannon

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

## Entropia Congiunta

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$

$X$  e  $Y$  variabili aleatorie discrete,  $p(x)$ ,  $p(x, y)$ ,  $p(y|x)$  distribuzioni.

## Entropia di Shannon

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)$$

## Entropia Congiunta

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y)$$

## Entropia Condizionale

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x)$$

# Proprietà

Valgono le seguenti proprietà:

- L'entropia è sempre non negativa.
- $H(X, Y) = H(X) + H(Y|X)$ .
- $H(X|Y) \leq H(X)$  con uguaglianza  $\iff$  indipendenza

Questi concetti possono essere generalizzati in modo analogo a  $n$  variabili aleatorie discrete  $X_1, \dots, X_n$ , così come le proprietà.

## Chain Rule

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1)$$



# Processi Stocastici

Nel caso di processi stocastici a tempi discreti  $\{X_t\}_{t \in I}$  con  $I = \mathbb{N}, \mathbb{Z}$  indicato con  $\mathbf{X}$  possiamo definire, quando il limite esiste

## Tasso di Entropia

$$H(\mathbf{X}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$$

## Lemma

Per processi stazionari il limite esiste ed è anche uguale a

$$\lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1)$$

# Proprietà di Equipartizione asintotica

## Teorema (AEP)

Se  $H$  è il tasso di entropia di un processo ergodico stazionario  $\{X_n\}$ , allora

$$-\frac{1}{n} \log p(X_0, X_1, \dots, X_{n-1}) \rightarrow H \quad \text{con probabilità 1}$$

La dimostrazione nel caso di variabili i.i.d. è una diretta conseguenza delle proprietà del logaritmo e della legge debole dei grandi numeri

# Conseguenze

- Fissati  $\epsilon, n$  esiste un insieme, detto insieme tipico  $A_\epsilon^{(n)}$ , definito come l'insieme dei  $x_1, x_2, \dots, x_n$  tali che  $2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}$ .
- Per l'AEP,  $\mathbf{P}\{A_\epsilon^{(n)}\} > 1 - \epsilon$  per  $n$  abbastanza grande. In pratica, a lungo andare, quasi tutte le successioni che si osservano appartengono all'insieme tipico.
- $(1 - \epsilon)2^{n(H(X)-\epsilon)} \leq |A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$ .
- Questo è il motivo principale per cui è possibile avere compressione di codici. Su periodi lunghi, quasi tutte le frasi da codificare apparterranno a questi insiemi tipici che possono essere compressi in modo ottimale. Il resto delle stringhe viene codificato in modo banale, ma hanno probabilità trascurabile.