

Teoria dell'Informazione e Portafogli Universali

Alessandro Trenta

18 giugno 2020

Indice

| | | |
|----------|--|-----------|
| 1 | Prime definizioni | 5 |
| 1.1 | Entropia per variabili aleatorie discrete | 5 |
| 1.2 | Processi Stocastici e Tasso di Entropia | 8 |
| 1.3 | Proprietà di Equipartizione Asintotica (AEP) | 10 |
| 1.4 | Corse dei Cavalli e Kelly Gambling | 11 |
| 1.5 | Corse dipendenti e Tasso di Entropia | 13 |
| 1.6 | Variabili aleatorie continue | 14 |
| 2 | Il Mercato Azionario | 15 |
| 2.1 | Definizioni | 15 |
| 2.2 | Condizioni di Kuhn-Tucker | 17 |
| 2.3 | Mercati stazionari, AEP e AOP | 19 |
| 2.3.1 | Risultati su portafogli ottimali e generalità | 19 |
| 2.3.2 | Principio di Ottimalità Asintotica | 22 |
| 2.3.3 | Proprietà di Equipartizione Asintotica | 23 |
| 3 | Portafogli Universali di Cover | 27 |
| 3.1 | Portafogli universali a orizzonte finito | 28 |
| 3.2 | Portafoglio Universale a orizzonte infinito | 32 |
| 4 | Algoritmi di Predizione Universali | 37 |
| 4.1 | Raggiungimento del tasso di raddoppio ottimale | 37 |
| 4.2 | Metodo delle Frequenze Empiriche | 40 |
| 4.3 | Algoritmo di Lempel-Ziv | 42 |
| 4.4 | Codici e Scommesse | 44 |
| 4.4.1 | I codici | 44 |
| 4.4.2 | Dualità Codici-Scommesse | 46 |
| 5 | Sperimentazioni | 49 |
| 5.1 | Metodi di scelta dei portafogli | 49 |
| 5.1.1 | Primo ordine | 49 |
| 5.1.2 | Secondo Ordine | 49 |
| 5.2 | Metodi e risultati sperimentali | 50 |
| 5.2.1 | Quantità di analisi | 50 |

| | | |
|----------|--|-----------|
| 5.2.2 | Metodo empirico sui rendimenti | 51 |
| 5.2.3 | Metodo empirico sui rapporti rendimento-incertezza | 54 |
| 5.2.4 | Metodo ad albero (Lempel-Ziv) | 55 |
| 5.2.5 | Metodo del portafoglio universale di Cover | 57 |
| 5.2.6 | Metodo Nearest-Neighbour | 58 |
| A | Definizioni | 61 |
| A.1 | Speranza condizionale e distribuzioni condizionate | 61 |
| A.2 | Martingale | 63 |
| A.3 | Sistemi dinamici misurabili | 63 |
| B | Codici | 65 |
| B.1 | Metodo delle frequenze empiriche | 65 |

Capitolo 1

Prime definizioni

1.1 Entropia per variabili aleatorie discrete

Sia dato uno spazio di probabilità Ω e una variabile aleatoria discreta X su Ω a valori in \mathcal{X} con densità di probabilità $p(x) = \Pr\{X = x\}$ per $x \in \mathcal{X}$.

Notazione. D'ora in poi quando si utilizza il logaritmo questo sarà sempre il logaritmo in base 2 e lo indicheremo con $\log(x) = \log_2(x)$

Definizione 1.1. L'entropia di Shannon della variabile aleatoria X è definita da

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

L'entropia è una misura dell'incertezza di una variabile aleatoria.

Nota. L'entropia di X può anche essere interpretata come il valore atteso della variabile aleatoria $\log\left(\frac{1}{p(X)}\right)$ rispetto alla densità di X .

$$H(X) = \mathbf{E}_p \left[\log \left(\frac{1}{p(X)} \right) \right]$$

Poiché $0 \leq p(x) \leq 1$ segue banalmente che $H(X) \geq 0$.

Nota. In generale, quando scriveremo $H(p_1, p_2, \dots, p_m)$ indicheremo $-\sum_{i=1}^m p_i \log(p_i)$ come se avessimo una variabile aleatoria X distribuita secondo il vettore (p_1, \dots, p_m) .

Associato al concetto di entropia vi sono quelli di entropia congiunta e entropia condizionale di due variabili aleatorie discrete X, Y a valori rispettivamente in \mathcal{X} e \mathcal{Y} .

Definizione 1.2. L'entropia congiunta $H(X, Y)$ di due variabili aleatorie discrete è definita come

$$\begin{aligned} H(X, Y) &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \\ &= -\mathbf{E}[\log p(X, Y)] \end{aligned}$$

dove $p(x, y)$ è la densità di probabilità congiunta.

Definizione 1.3. L'entropia condizionale di Y rispetto a X è definita come

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log(p(y|x)) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log(p(y|x)) \\ &= -\mathbf{E}_{p(x,y)}[\log(p(Y|X))] \end{aligned}$$

dove $p(y|x)$ è la densità di probabilità di Y condizionata a X .

Nota. L'entropia condizionale di X rispetto a se stessa $H(X|X)$ è nulla, questo segue facilmente dal fatto che $p(x|x) = 1$.

Queste quantità sono strettamente legate tra loro come mostra il seguente lemma:

Lemma 1.1 (Chain rule).

$$H(X, Y) = H(X) + H(Y|X)$$

Dimostrazione. Da $p(x, y) = p(x)p(y|x)$ segue che

$$\log(p(X, Y)) = \log(p(X)) + \log(p(Y|X))$$

Prendendo il valore atteso in entrambi i membri otteniamo la tesi. □

Nota. Più in generale, si può condizionare rispetto a più variabili. Si ottiene con lo stesso procedimento che, ad esempio

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z)$$

Introduciamo altre due quantità utili:

Definizione 1.4. L'entropia relativa o distanza di Kullback-Leibler (a valori in $[0, \infty]$) tra due densità di probabilità discreta $p(x)$ e $q(x)$ è definita come

$$\begin{aligned} D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right) \\ &= \mathbf{E}_p \left[\log \left(\frac{p(x)}{q(x)} \right) \right] \end{aligned}$$

dove usiamo la convenzione che $0 \log \frac{0}{0} = 0$, $0 \log \frac{0}{q} = 0$ e $p \log \frac{p}{0} = \infty$.

Nota. Nonostante il nome, la distanza di Kullback-Leibler non è simmetrica e quindi non è una vera distanza tra distribuzioni.

Definizione 1.5. Siano date due variabili aleatorie X, Y con densità di probabilità congiunta $p(x, y)$ e densità marginali $p(x)$ e $p(y)$. L'informazione mutua $I(X; Y)$ è l'entropia relativa tra la densità congiunta e il prodotto delle marginali. In formule:

$$\begin{aligned} I(X; Y) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ &= D(p(x, y) || p(x)p(y)) \\ &= \mathbf{E}_{p(x, y)} \log \frac{p(X, Y)}{p(X)p(Y)} \end{aligned}$$

Dalla definizione è immediato notare che, per simmetria, $I(X; Y) = I(Y; X)$.

L'informazione mutua rappresenta la quantità di informazione che ciascuna variabile aleatoria fornisce dell'altra, ossia quanto sono distanti dall'essere indipendenti. Possiamo inoltre riscriverla come

$$\begin{aligned} I(X; Y) &= \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ &= \sum_{x, y} p(x, y) \log \frac{p(x|y)}{p(x)} \\ &= - \sum_{x \in \mathcal{X}} p(x) \log(p(x)) - \left(- \sum_{x, y} p(x, y) \log(p(x|y)) \right) \\ &= H(X) - H(X|Y) \end{aligned}$$

Una diretta conseguenza di questo e della simmetria è che

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X, Y)$$

Nota. Da queste formule si ottiene che $I(X; X) = H(X) - H(X|X) = H(X)$. L'entropia è quindi, in un certo senso, la quantità di informazione che una variabile aleatoria dà rispetto a se stessa.

Lemma 1.2. *L'entropia relativa è sempre non negativa. In formule: $D(p||q) \geq 0$*

Dimostrazione. Sia $A = \{x : p(x) > 0\}$, allora:

$$\begin{aligned} -D(p||q) &= - \sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} \\ &= \sum_{x \in A} p(x) \log \frac{q(x)}{p(x)} \leq \log \left(\sum_{x \in A} p(x) \frac{q(x)}{p(x)} \right) \\ &= \log \left(\sum_{x \in A} q(x) \right) \leq \log \left(\sum_{x \in \mathcal{X}} q(x) \right) \\ &= \log(1) = 0 \end{aligned}$$

Dove nel secondo passaggio abbiamo utilizzato la disuguaglianza di Jensen sulla funzione concava $\log(x)$. Sappiamo inoltre che l'uguaglianza si ha solo nel caso in cui $\frac{p(x)}{q(x)}$ è costante ovunque, cioè $p(x) = c \cdot q(x) \forall x \in \mathcal{X}$. Dal fatto che entrambe sono normalizzate a somma 1 si ottiene $p(x) = q(x)$. \square

Corollario 1.3. *Per ogni coppia di variabili aleatorie discrete X, Y vale $I(X; Y) \geq 0$ con uguaglianza se e solo se le variabili sono indipendenti*

Dimostrazione. Segue direttamente dalla definizione di entropia mutua e dal lemma precedente \square

La chain rule per l'entropia può essere estesa a un numero generico di variabili aleatorie discrete, come mostra il seguente risultato.

Teorema 1.4. *Siano X_1, X_2, \dots, X_n variabili aleatorie discrete sullo stesso spazio di probabilità con densità della legge $p(x_1, x_2, \dots, x_n)$. Allora*

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

Dimostrazione. Si ottiene per ricorsione dal lemma 1.1. \square

Lemma 1.5. *Date X e Y variabili aleatorie discrete*

$$H(X|Y) \leq H(X)$$

con uguaglianza se e solo se sono indipendenti.

Dimostrazione. Segue direttamente dal fatto che $0 \leq I(X; Y) = H(X) - H(X|Y)$ \square

Intuitivamente, le informazioni date da una variabile aleatoria, possono solo ridurre l'incertezza legata ad un'altra. In generale, condizionare a più variabili aleatorie può solo ridurre l'entropia.

1.2 Processi Stocastici e Tasso di Entropia

Definiamo alcuni dei concetti appena visti al caso di processi stocastici a tempi discreti, ossia a successioni di variabili aleatorie $\{X_i\}_{i \in I}$ dove $I = \mathbb{N}, \mathbb{Z}$. In questa sezione, ogni processo stazionario sarà da intendersi a tempi discreti e verrà indicato con l'alfabeto ad esso associato \mathcal{X} .

Definizione 1.6. Un processo stocastico è detto stazionario se la distribuzione congiunta di qualunque sottoinsieme della sequenza di variabili aleatorie è invariante rispetto alla traslazione temporale. In formule

$$\begin{aligned} & \Pr\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \\ &= \Pr\{X_{1+l} = x_1, X_{2+l} = x_2, \dots, X_{n+l} = x_n\} \end{aligned}$$

per ogni n , per ogni l e per ogni $x_1, x_2, \dots, x_n \in \mathcal{X}$.

Definizione 1.7. Un processo stocastico discreto è detto *catena di Markov* se per $n = 1, 2, \dots$

$$\Pr\{X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_1 = x_1\} = \Pr\{X_{n+1} = x_{n+1} | X_n = x_n\}$$

per ogni $x_1, \dots, x_{n+1} \in \mathcal{X}$.

Definiamo ora la generalizzazione del concetto di entropia per i processi stocastici a tempi discreti:

Definizione 1.8 (Tasso di Entropia). Il tasso di entropia di un processo stocastico o *entropy rate* è definita da

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

laddove il limite esiste.

Possiamo dare anche una definizione alternativa di entropy rate che, nel caso di processi stazionari, si vedrà essere equivalente alla precedente.

Definizione 1.9. Definiamo, dove il limite esiste

$$H'(\mathcal{X}) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1) \tag{1.1}$$

Dimostriamo ora che nel caso di processi stazionari entrambi i limiti esistono e coincidono.

Teorema 1.6. Per un processo stocastico stazionario, $H(X_n | X_{n-1}, \dots, X_1)$ è debolmente decrescente in n e ammette limite $H'(\mathcal{X})$.

Dimostrazione.

$$\begin{aligned} H(X_{n+1} | X_n, \dots, X_2, X_1) &\leq H(X_{n+1} | X_n, \dots, X_2) \\ &= H(X_n | X_{n-1}, \dots, X_1) \end{aligned}$$

dove la disuguaglianza segue dal fatto che il condizionamento a una variabile aleatoria diminuisce l'entropia.

Siccome $H(X_n | X_{n-1}, \dots, X_1)$ è una successione debolmente decrescente a valori non negativi, ammette limite $H'(\mathcal{X})$. \square

Enunciamo ora il lemma della media di Cesàro in vista della prossima dimostrazione.

Lemma 1.7. Sia $\{a_n\}_{n \in \mathbb{N}}$ una successione di numeri reali e supponiamo che a_n converga a l in \mathbb{R} .

$$\lim_{n \rightarrow \infty} a_n = l$$

Allora

$$\lim_{n \rightarrow \infty} \frac{a_1 + a_2 + \dots + a_n}{n} = l$$

Teorema 1.8. *Per un processo stocastico stazionario entrambi i limiti $H(\mathcal{X}), H'(\mathcal{X})$ esistono e sono uguali:*

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

Dimostrazione. Il lemma precedente ci assicura l'esistenza del limite in $H'(\mathcal{X})$. Consideriamo ora (utilizzando la chain rule)

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

Sappiamo però che il termine a destra è la media dei primi n termini della successione del lemma sopra. Utilizzando quindi quest'ultimo risultato e il lemma della media di Cesàro sappiamo che il limite esiste e che coincide con $H'(\mathcal{X})$.

$$H(\mathcal{X}) = \lim_{n \rightarrow \infty} \frac{H(X_1, X_2, \dots, X_n)}{n} = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1) = H'(\mathcal{X})$$

□

Nel caso di catene di Markov, si vede immediatamente che l'entropia rate si scrive come

$$H(\mathcal{X}) = H'(\mathcal{X}) = \lim_{n \rightarrow \infty} H(X_n | X_{n-1}, \dots, X_1) \quad (1.2)$$

$$= \lim_{n \rightarrow \infty} H(X_n | X_{n-1}) = H(X_2 | X_1) \quad (1.3)$$

1.3 Proprietà di Equipartizione Asintotica (AEP)

Introduciamo ora una proprietà per variabili i.i.d. che verrà successivamente estesa e generalizzata.

Teorema 1.9 (AEP). *Se X_1, X_2, \dots sono variabili aleatorie discrete i.i.d. secondo $p(x)$ allora*

$$-\frac{1}{n} \log(p(X_1, X_2, \dots, X_n)) \rightarrow H(X) \quad \text{in probabilità}$$

Dimostrazione. Funzioni di variabili aleatorie indipendenti sono variabili aleatorie indipendenti. Quindi, dato che le X_i sono i.i.d., lo sono anche $\log(p(X_i))$. Dalla legge debole dei grandi numeri segue che:

$$\begin{aligned} -\frac{1}{n} \log(p(X_1, X_2, \dots, X_n)) &= -\frac{1}{n} \sum_{i=1}^n \log p(X_i) \\ &\rightarrow -\mathbf{E}[\log(p(X))] \quad \text{in probabilità} \\ &= H(X) \end{aligned}$$

□

In generale, come mostra il seguente teorema, l'*Asymptotic Equipartition Property*, vale per processi stazionari e ergodici.

Teorema 1.10 (AEP: teorema di Shannon-McMillan-Breiman). *Se H è l'entropy rate 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 (FORSE SOLO PER MERCATI DA ARTICOLO).

1.4 Corse dei Cavalli e Kelly Gambling

Consideriamo una corsa alla quale partecipano m cavalli, in cui l' i -esimo vince con probabilità p_i e la sua vittoria dà un pagamento o_i per ogni unità di capitale scommessa. Vogliamo trovare un metodo per dividere il capitale a nostra disposizione in m frazioni b_i , dove $b_i \geq 0$ e $\sum_{i=1}^m b_i = 1$, ciascuna delle quali è scommessa sul corrispondente cavallo. Cerchiamo il modo di massimizzare il nostro capitale risultante dopo ogni gara o asintoticamente.

Possiamo quindi definire il vettore dei pagamenti di una scommessa $\mathbf{o} \in \mathbb{R}^m$ e quelli delle frazioni scommesse $\mathbf{b} \in \mathbb{R}^m$ (detto anche portafoglio). Consideriamo per ogni $i \in \mathbb{N}$ la variabile aleatoria X_i a valori in $\{1, \dots, m\}$ che corrisponde al cavallo vincente alla i -esima gara. Assumiamo che ciascuna di queste abbia legge di probabilità data dalle p_j , $j \in \{1, \dots, m\}$ sopra citate.

Definizione 1.10. Il capitale relativo $S(X_i) = b_{X_i} o_{X_i}$ è il fattore con cui il capitale dello scommettitore aumenta se il cavallo X_i vince alla corsa i . Il capitale totale dopo n corse sarà quindi

$$S_n = \prod_{i=1}^n S(X_i) \tag{1.4}$$

Definizione 1.11. Il tasso di raddoppio (*doubling rate*) di una corsa è definito da

$$W(\mathbf{b}, \mathbf{p}) = \mathbf{E}[\log(S(X))] = \sum_{k=1}^m p_k \log(b_k o_k)$$

Teorema 1.11. *Date variabili aleatorie X_i come sopra, indipendenti e identicamente distribuite secondo $p(x)$, il valore atteso del capitale dello scommettitore che utilizza ad ogni passo una strategia \mathbf{b} cresce esponenzialmente con tasso di raddoppio $W(\mathbf{b}, \mathbf{p})$, cioè*

$$\mathbf{E}[S_n] \doteq 2^{nW(\mathbf{b}, \mathbf{p})}$$

Dimostrazione. Funzioni di variabili aleatorie indipendenti sono variabili aleatorie indipendenti, quindi anche $\log(S_i)$ sono i.i.d.. Dalla legge dei grandi numeri segue che

$$\frac{1}{n} \log(S_n) = \frac{1}{n} \sum_{i=1}^n \log S(X_i) \rightarrow \mathbf{E}[\log(S(X))] \quad \text{in probabilità}$$

Da questo segue la tesi. □

Massimizzare il capitale corrisponde quindi a massimizzare il tasso di raddoppio.

Definizione 1.12. Il tasso di raddoppio ottimale $W^*(p)$ è il massimo tasso di raddoppio al variare di tutti i portafogli \mathbf{b}

$$W^*(p) = \max_{\mathbf{b}} W(\mathbf{b}, \mathbf{p}) = \max_{\mathbf{b}: b_i \geq 0, \sum_i b_i = 1} \sum_{i=1}^m p_i \log(b_i o_i)$$

Il prossimo teorema mostra come il portafoglio ottimale corrisponda a quello che pone pesi proporzionali alle probabilità di vittoria (questo metodo è detto anche *Kelly gambling*).

Teorema 1.12. Il tasso di raddoppio ottimale è dato da

$$W^*(\mathbf{p}) = \sum_{i=1}^m p_i \log(o_i) - H(\mathbf{p})$$

e viene raggiunto dal portafoglio $b_i = p_i \forall i$.

Dimostrazione. Riscriviamo $W(\mathbf{b}, \mathbf{p})$ come

$$\begin{aligned} W(\mathbf{b}, \mathbf{p}) &= \sum_{i=1}^m p_i \log(b_i o_i) \\ &= \sum_{i=1}^m p_i \log\left(\frac{b_i}{p_i} p_i o_i\right) \\ &= \sum_{i=1}^m p_i \log(o_i) - H(\mathbf{p}) - D(\mathbf{p}||\mathbf{b}) \\ &\leq \sum_{i=1}^m p_i \log(o_i) - H(\mathbf{p}) \end{aligned}$$

Dal lemma 1.2 segue che l'uguaglianza si ha quando $p_i = b_i \forall i$. □

Consideriamo in particolare il caso in cui il pagamento è equo, ossia quando $\sum_i \frac{1}{o_i} = 1$. Possiamo quindi scrivere $\frac{1}{o_i} = r_i$ e interpretare quest'ultimi come la densità di legge stimata dal bookmaker per ogni cavallo. In tal caso possiamo riscrivere il tasso di raddoppio come

$$\begin{aligned} W(\mathbf{b}, \mathbf{p}) &= \sum_{i=1}^m p_i \log(b_i o_i) \\ &= \sum_{i=1}^m p_i \log\left(\frac{b_i p_i}{p_i r_i}\right) \\ &= D(\mathbf{p}||\mathbf{r}) - D(\mathbf{p}||\mathbf{b}) \end{aligned}$$

Questo risultato si interpreta facilmente col fatto che lo scommettitore guadagna se e solo se riesce a stimare meglio \mathbf{p} rispetto al bookmaker.

Nel caso di pagamento uniforme e equo $o_i = m$, si ottiene allora che

$$W^*(\mathbf{p}) = D\left(\mathbf{p} \parallel \frac{1}{m}\right) = \log(m) - H(\mathbf{p})$$

Vale in questo caso il cosiddetto teorema di conservazione ossia che $W^*(\mathbf{p}) + H(\mathbf{p}) = \log(m)$.

1.5 Corse dipendenti e Tasso di Entropia

Analizziamo ora il caso di una successione di corse dipendenti ognuna dalle precedenti. Consideriamo quindi un processo stocastico $\{X_i\}$ che rappresenta i risultati delle varie corse. Ad ogni turno possiamo scegliere un portafoglio $\mathbf{b}(\cdot|X_{k-1}, \dots, X_1)$ dipendente dal passato.

In questo caso il tasso di raddoppio ottimale si ottiene come

$$W^*(X_k|X_{k-1}, \dots, X_1) = \max_{\mathbf{b}(\cdot|X_{k-1}, \dots, X_1)} \mathbf{E}[\log(S(X_k))|X_{k-1}, \dots, X_1]$$

Nel caso di pagamento uniforme ed equo questa può essere riscritta (in modo analogo ai casi precedenti) come

$$W^*(X_k|X_{k-1}, \dots, X_1) = \log(m) - H(X_k|X_{k-1}, \dots, X_1)$$

e tale massimo è raggiunto con il portafoglio $b^*(x_k|x_{k-1}, \dots, x_1) = p(x_k|x_{k-1}, \dots, x_1)$. Dopo n corse il capitale è dato da

$$S_n = \prod_{i=1}^n S(X_i)$$

e l'esponente nel valore atteso di S_n si ottiene prendendo il logaritmo:

$$\begin{aligned} \frac{1}{n} \mathbf{E}[\log(S_n)] &= \frac{1}{n} \sum_{i=1}^n \mathbf{E}[\log(S(X_i))] \\ &= \frac{1}{n} \sum_{i=1}^n (\log(m) - H(X_i|X_{i-1}, \dots, X_1)) \\ &= \log(m) - \frac{H(X_1, X_2, \dots, X_n)}{n} \end{aligned}$$

Se il processo stocastico ha un tasso di entropia pari a $H(\mathcal{X})$ si ottiene prendendo il limite che

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}[\log(S_n)] + H(\mathcal{X}) = \log(m)$$

1.6 Variabili aleatorie continue

Estendiamo molti dei concetti già presentati al caso di variabili aleatorie continue dotate di densità su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$. Nell'appendice sono elencate alcune proprietà e definizioni di concetti usati in questo paragrafo.

Definizione 1.13. L'entropia di una variabile aleatoria continua a valori in \mathbb{R} con densità $f(x)$, tale che $f(x) \log f(x) \in L^1$ è definita da:

$$H(X) = - \int_{\mathbb{R}} f(x) \log(f(x)) dx$$

Definizione 1.14. L'entropia congiunta di un insieme di variabili aleatorie continue X_1, X_2, \dots, X_n con densità $f(\mathbf{x})$, con $f(\mathbf{x}) \log f(\mathbf{x}) \in L^1$ è definita da:

$$H(X_1, X_2, \dots, X_n) = - \int_{\mathbb{R}^n} f(\mathbf{x}) \log(f(\mathbf{x})) d\mathbf{x}$$

Definizione 1.15. L'entropia condizionale di due variabili aleatorie X, Y con densità congiunta $f(x, y)$ e condizionale $f_{X|Y}(x, Y = y)$, tale che $f(x, y) \log(f_{X|Y}(x, y)) \in L^1$ è definita come:

$$H(X|Y) = - \int_{\mathbb{R}^2} f(x, y) \log(f_{X|Y}(x, Y = y)) dx dy$$

In particolare si ha che la densità condizionale può essere espressa come $f_{X|Y}(x, y) = \frac{f(x, y)}{f_Y(y)}$ (rif. Appendice).

Definizione 1.16. L'entropia relativa o distanza di Kullback-Leibler tra due densità $f(\mathbf{x})$ e $g(\mathbf{x})$, tali che $f(\mathbf{x}) \log \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) \in L^1$ è definita da:

$$D(f||g) = \int_{\mathbb{R}^n} f(\mathbf{x}) \log \left(\frac{f(\mathbf{x})}{g(\mathbf{x})} \right) d\mathbf{x}$$

Definizione 1.17. L'informazione mutua tra due variabili aleatorie continue X e Y di densità congiunta $f(x, y)$ e marginali $f_X(x), f_Y(y)$, tali che $f(x, y) \log \left(\frac{f(x, y)}{f_X(x)f_Y(y)} \right) \in L^1$ è definita come

$$\begin{aligned} I(X; Y) &= \int_{\mathbb{R}^2} f(x, y) \log \left(\frac{f(x, y)}{f_X(x)f_Y(y)} \right) dx dy \\ &= D(f(x, y)||f_X(x)f_Y(y)) \end{aligned}$$

Grazie alle proprietà del logaritmo valgono tutti i lemmi e i teoremi già verificati per le variabili discrete. [da un controllo veloce](#) sì, [fare controllo approfondito](#)

Capitolo 2

Il Mercato Azionario

2.1 Definizioni

In questo capitolo presenteremo un'applicazione del criterio di Kelly (*Kelly gambling*) al mercato azionario. Seguiremo da vicino la presentazione contenuta nel libro di Cover e Thomas, "*Elements of Information Theory*" [CT06]. Esso viene modellizzato come un vettore di rendimenti azionari $\mathbf{X} = (X_1, X_2, \dots, X_m)$, $X_i \geq 0, \forall i$ dove m è il numero di asset (o titoli) considerati. Ogni X_i è il rapporto tra il prezzo di chiusura in due giorni consecutivi e tipicamente questo valore è vicino a 1.

Assumiamo che \mathbf{X} sia distribuito con funzione di ripartizione $F(\mathbf{x})$ e densità $f(\mathbf{x})$.

Un portafoglio è una suddivisione del proprio capitale in frazioni sulle varie azioni ed è rappresentato come un vettore $\mathbf{b} = (b_1, b_2, \dots, b_m)$ dove $b_i \geq 0 \forall i$ e $\sum_{i=1}^m b_i = 1$.

Usando un portafoglio \mathbf{b} con un vettore di rendimenti \mathbf{X} il fattore di aumento del capitale da un giorno all'altro sarà $S = \mathbf{b}^t \mathbf{X}$.

Definizione 2.1. Il tasso di raddoppio associato a un portafoglio \mathbf{b} e a una distribuzione $F(\mathbf{x})$ è definito come

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

Definizione 2.2. Il tasso di raddoppio 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}^* .

Teorema 2.1. Siano $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ i.i.d. vettori di rendimenti azionari per n giorni con distribuzione data da $F(\mathbf{x})$. Sia quindi

$$S_n^* = \prod_{i=1}^n \mathbf{b}^{*t} \mathbf{X}_i$$

il capitale dopo n giorni di utilizzo del portafoglio log-ottimale. Allora

$$\frac{1}{n} \log(S_n^*) \rightarrow W^* \quad \text{con probabilità 1}$$

Dimostrazione. Dalla legge forte dei grandi numeri

$$\begin{aligned} \frac{1}{n} \log(S_n^*) &= \frac{1}{n} \sum_{i=1}^n \log(\mathbf{b}^{*t} \mathbf{X}_i) \\ &\rightarrow W^* \quad \text{con probabilità 1} \end{aligned}$$

da cui otteniamo anche che $\mathbf{E}[S_n^*] \doteq 2^{nW^*}$. □

Vogliamo caratterizzare il portafoglio log-ottimale in qualche modo. Per fare questo iniziamo con un lemma preliminare

Lemma 2.2. $W(\mathbf{b}, F)$ è concavo in \mathbf{b} e lineare in F , $W^*(F)$ è convesso in F .

Dimostrazione. Dalla concavità del logaritmo sappiamo che

$$\log((\lambda \mathbf{b}_1 + (1 - \lambda) \mathbf{b}_2)^t \mathbf{X}) \geq \lambda \log(\mathbf{b}_1^t \mathbf{X}) + (1 - \lambda) \log(\mathbf{b}_2^t \mathbf{X})$$

Prendendo il valore atteso da entrambi i membri della disuguaglianza otteniamo che

$$W(\lambda \mathbf{b}_1 + (1 - \lambda) \mathbf{b}_2, F) \geq \lambda W(\mathbf{b}_1, F) + (1 - \lambda) W(\mathbf{b}_2, F)$$

Sia ora $\mathbf{b}^*(\lambda F_1 + (1 - \lambda) F_2)$ il portafoglio log-ottimale. Ora

$$\begin{aligned} W^*(\lambda F_1 + (1 - \lambda) F_2) &= W(\mathbf{b}^*(\lambda F_1 + (1 - \lambda) F_2), \lambda F_1 + (1 - \lambda) F_2) \\ &= \lambda W(\mathbf{b}^*(\lambda F_1 + (1 - \lambda) F_2), F_1) \\ &\quad + (1 - \lambda) W(\mathbf{b}^*(\lambda F_1 + (1 - \lambda) F_2), F_2) \\ &\leq \lambda W^*(\mathbf{b}^*(F_1), F_1) + (1 - \lambda) W^*(\mathbf{b}^*(F_2), F_2) \end{aligned}$$

dove l'ultima disuguaglianza segue dal fatto che $\mathbf{b}^*(F_1)$ massimizza $W(\mathbf{b}, F_1)$ e $\mathbf{b}^*(F_2)$ massimizza $W(\mathbf{b}, F_2)$. □

Lemma 2.3. L'insieme dei portafogli log-ottimali rispetto a una data distribuzione F è convesso.

Dimostrazione. Supponiamo che sia \mathbf{b}_1 che \mathbf{b}_2 siano log-ottimali per F . Dalla concavità di $W(\mathbf{b}, F)$ rispetto a \mathbf{b} abbiamo

$$W(\lambda \mathbf{b}_1 + (1 - \lambda) \mathbf{b}_2, F) \geq \lambda W(\mathbf{b}_1, F) + (1 - \lambda) W(\mathbf{b}_2, F) = W^*(F)$$

Quindi anche $\lambda \mathbf{b}_1 + (1 - \lambda) \mathbf{b}_2$ è log-ottimale. □

2.2 Condizioni di Kuhn-Tucker

Vogliamo a questo punto caratterizzare il portafoglio log-ottimale di un investimento su singolo periodo. Per fare ciò utilizziamo le condizioni di Kuhn-Tucker. Sia quindi $\mathcal{B} = \{\mathbf{b} \in \mathbb{R}^m : b_i \geq 0, \sum_{i=1}^m b_i = 1\}$ l'insieme dei portafogli. Vale il seguente teorema

Teorema 2.4. *Il portafoglio log-ottimale \mathbf{b}^* per il mercato azionario \mathbf{X} con distribuzione F soddisfa le seguenti condizioni necessarie e sufficienti:*

$$\mathbf{E} \left[\frac{X_i}{\mathbf{b}^{*t} \mathbf{X}} \right] = 1 \quad \text{se } b_i^* > 0$$

$$\leq 1 \quad \text{se } b_i^* = 0$$

Dimostrazione. Per comodità verrà usato il logaritmo naturale invece di quello binario. A meno di costanti il risultato è analogo.

Il tasso di raddoppio $W(\mathbf{b}) = \mathbf{E}[\ln(\mathbf{b}^t \mathbf{X})]$ è concavo in \mathbf{b} . Da questo segue che \mathbf{b}^* è log-ottimale se e solo se la derivata direzionale di $W(\cdot)$ nella direzione da \mathbf{b}^* a qualsiasi altro portafoglio \mathbf{b} è non positiva. Ponendo quindi $\mathbf{b}_\lambda = (1 - \lambda)\mathbf{b}^* + \lambda\mathbf{b}$ per $0 \leq \lambda \leq 1$ abbiamo:

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

Dove nel secondo passaggio abbiamo applicato il teorema di convergenza dominata. Si ha infatti che per λ piccoli

$$\left| \frac{1}{\lambda} \log \left(1 + \lambda \left(\frac{\mathbf{b}^t \mathbf{X}}{\mathbf{b}^{*t} \mathbf{X}} - 1 \right) \right) \right| \leq \left| \frac{1}{\lambda} \lambda \left(\frac{\mathbf{b}^t \mathbf{X}}{\mathbf{b}^{*t} \mathbf{X}} - 1 \right) \right|$$

$$\leq 2$$

poiché $\frac{\mathbf{b}^t \mathbf{X}}{\mathbf{b}^{*t} \mathbf{X}} \leq 1$ (portafoglio log-ottimale) e per $x \geq -\frac{1}{2}$ si ha $|\log(1 + x)| \leq |x|$. Come funzione dominante si può quindi prendere 2 (integrabile).

La condizione si riconduce quindi a

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

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

Le condizioni di Kuhn-Tucker valgono per ogni portafoglio in \mathcal{B} . Consideriamo quindi quelli agli estremi, ossia 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 $b_j^* > 0$. Possiamo quindi riscrivere le condizioni in funzione di questo ottenendo la tesi:

$$\mathbf{E} \left[\frac{X_i}{\mathbf{b}^{*t} \mathbf{X}} \right] = 1 \quad \text{se } b_i^* > 0$$

$$\leq 1 \quad \text{se } b_i^* = 0$$

□

Il seguente risultato mostra una utile conseguenza

Teorema 2.5. *Sia $S^* = \mathbf{b}^{*t} \mathbf{X}$ il capitale aleatorio risultante da un investimento su singolo periodo con un'allocazione data dal portafoglio log-ottimale. Sia quindi $S = \mathbf{b}^t \mathbf{X}$ il capitale aleatorio ottenuto investendo secondo un qualsiasi altro portafoglio. Allora:*

$$\mathbf{E} \left[\ln \left(\frac{S}{S^*} \right) \right] \leq 0 \quad \forall S \iff \mathbf{E} \left[\frac{S}{S^*} \right] \leq 1 \quad \forall S$$

Dimostrazione. Dal teorema precedente sappiamo che se \mathbf{b}^* è log-ottimale (e quindi $\mathbf{E} \left[\log \left(\frac{S}{S^*} \right) \right] \leq 0$)

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

per ogni i . Moltiplicando questa disuguaglianza per b_i e sommando su i otteniamo

$$\sum_{i=1}^m b_i \mathbf{E} \left[\frac{\mathbf{b}^t \mathbf{X}}{\mathbf{b}^{*t} \mathbf{X}} \right] \leq \sum_{i=1}^m b_i = 1$$

che è equivalente a

$$\mathbf{E} \left[\frac{\mathbf{b}^t \mathbf{X}}{\mathbf{b}^{*t} \mathbf{X}} \right] = \mathbf{E} \left[\frac{S}{S^*} \right] \leq 1$$

Infine, l'equivalenza è una conseguenza diretta della disuguaglianza di Jensen

$$\mathbf{E} \left[\log \left(\frac{S}{S^*} \right) \right] \leq \log \mathbf{E} \left[\frac{S}{S^*} \right] \leq \log(1) = 0$$

□

Questo teorema ci dice che massimizzare il tasso di raddoppio corrisponde in realtà a massimizzare il rapporto tra capitali.

Come ulteriore conseguenza della caratterizzazione di Kuhn-Tucker abbiamo che la parte di capitale ottenuta dopo ogni giornata da un asset è la stessa porzione che era investita nello stesso.

$$\mathbf{E} \left[\frac{b_i^* X_i}{\mathbf{b}^{*t} \mathbf{X}} \right] = b_i^* \mathbf{E} \left[\frac{X_i}{\mathbf{b}^{*t} \mathbf{X}} \right] = b_i^*$$

Questa è una evidente analogia con il Kelly gambling, in quanto la parte reinvestita è sempre pari a quella ottenuta, mantenendo le proporzioni con il guadagno atteso.

2.3 Mercati stazionari, AEP e AOP

2.3.1 Risultati su portafogli ottimali e generalità

Il teorema in appendice assicura l'esistenza della probabilità condizionale in partenza se l'arrivo è \mathbb{R} coi boreliani, si potrebbe quindi togliere l'ipotesi dello spazio di probabilità perfetto.

Iniziamo con qualche definizione preliminare. In questa sezione seguiremo quanto fatto da Algoet e Cover in [AC88]. Sia dato un processo stocastico stazionario $\{\mathbf{X}_t\}_{-\infty < t < \infty}$, $t \in \mathbb{Z}$ che rappresenta i rendimenti azionari degli asset. Ad ogni passo scegliamo un portafoglio che sia log-ottimale basandoci sul passato. Al periodo di investimento n avremo quindi a disposizione informazioni rappresentate da una σ -algebra \mathcal{F}_t tale che $\sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1}) \subseteq \mathcal{F}_t$ e scegliamo un portafoglio che si \mathcal{F}_t misurabile e indichiamo quello log-ottimale come $\mathbf{b}_t^* = \mathbf{b}^*(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$, ossia quello che massimizza

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

Il capitale dopo n passi sarà dato da $S_n = \prod_{t=1}^n \mathbf{b}_t^*(\mathbf{X}_{t-1}, \dots, \mathbf{X}_0) \mathbf{X}_t$.

Definizione 2.3. Il tasso di raddoppio log-ottimale al passo n è dato da

$$W_n^* = \mathbf{E}[w_n^*] = \mathbf{E}[\log(\mathbf{b}_n^{*t} \mathbf{X}_n)] = \sup_{\mathbf{b}=\mathbf{b}(\mathbf{X}_0, \dots, \mathbf{X}_{n-1})} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X}_n)] = W^*(\mathbf{X}_n | \mathbf{X}_0, \dots, \mathbf{X}_{n-1}) \quad (2.1)$$

dove la seconda uguaglianza segue dalle proprietà della speranza condizionale.

Definizione 2.4. Il tasso di raddoppio ottimale per i primi n passi è dato da

$$W^*(\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_{n-1}) = \sup_{\mathbf{b}_0, \dots, \mathbf{b}_{n-1}} \mathbf{E}[\log(S_n)] \quad (2.2)$$

Il logaritmo di S_n si scompone in una somma. Considerando anche che ad ogni passo scelgo un portafoglio indipendente dai precedenti (ma dipendente dai vettori degli stock), posso scrivere che l'estremo superiore nell'equazione 2.2 al variare della successione dei portafogli è la somma dei singoli estremi superiori in 2.1, otteniamo la chain rule:

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

Definizione 2.5. Il tasso di raddoppio asintotico W_∞^* è definito da

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

se il limite esiste.

Teorema 2.6. Per un mercato stazionario, il tasso di raddoppio asintotico esiste ed è uguale a

$$W_\infty^* = \lim_{n \rightarrow \infty} W_n^*$$

Dimostrazione. Dalla stazionarietà W_n^* è non decrescente in n , quindi ammette un limite, anche infinito. Siccome

$$\frac{W^*(\mathbf{X}_0, \dots, W_{n-1})}{n} = \frac{1}{n} \sum_{t=0}^{n-1} W_t^*$$

segue dal teorema delle media di Cesàro che

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

□

Assumiamo che il vettore dei rendimenti $\mathbf{X}(\omega) \in \mathbb{R}_+^m$ sia definito su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$, il teorema citato in appendice ci assicura che \mathbf{X} ammette una distribuzione di probabilità condizionale e regolare data ogni sotto- σ -algebra di \mathcal{F} .

Teorema 2.7. *Sia $\{\bar{\mathcal{F}}_t\}_{0 \leq t < \infty}$ una successione crescente di sotto- σ -algebre di \mathcal{F} con σ -algebra terminale $\bar{\mathcal{F}}_\infty \subseteq \mathcal{F}$.*

- Se $\bar{\mathbf{P}}_t$ è una distribuzione di probabilità regolare di \mathbf{X} condizionata a $\bar{\mathcal{F}}_t$, allora

$$\bar{\mathbf{P}}_t \rightarrow \bar{\mathbf{P}}_\infty \quad \text{debolmente quasi certamente}$$

- Se $\mathbf{b}^*(\cdot)$ è un selettore misurabile di portafogli log-ottimali data una distribuzione, allora $\bar{\mathbf{b}}_t^* = \mathbf{b}^*(\bar{\mathbf{P}}_t)$ è un portafoglio $\bar{\mathcal{F}}_t$ misurabile che raggiunge il massimo log-return atteso dato $\bar{\mathcal{F}}_t$. Inoltre $\bar{\mathbf{b}}_t^{*t} \mathbf{X} \rightarrow \bar{\mathbf{b}}_\infty^{*t} \mathbf{X}$ quasi certamente e quindi

$$\log(\bar{\mathbf{b}}_t^{*t} \mathbf{X}) \rightarrow \log(\bar{\mathbf{b}}_\infty^{*t} \mathbf{X}) \quad \text{q.c.}$$

Se il portafoglio log-ottimale dato $\bar{\mathcal{F}}_\infty$ è unico allora anche $\bar{\mathbf{b}}_t^* \rightarrow \bar{\mathbf{b}}_\infty^*$ q.c..

- Se $w^*(\cdot)$ denota la funzione di massimo log-ritorno atteso, allora il massimo log-ritorno atteso condizionato a $\bar{\mathcal{F}}_t$ è dato da

$$\bar{w}_t^* = w^*(\bar{\mathbf{P}}_t) = \sup_{\mathbf{b} \in \bar{\mathcal{F}}_t} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X}) | \bar{\mathcal{F}}_t] = \mathbf{E}[\log(\bar{\mathbf{b}}_t^{*t} \mathbf{X}) | \bar{\mathcal{F}}_t]$$

In particolare, $\{\bar{w}_t^*, \bar{\mathcal{F}}_t\}_{0 \leq t < \infty}$ è una submartingala e

$$\bar{w}_t^* \rightarrow \bar{w}_\infty^* \quad \text{q.c. (e in } L^1 \text{ se } \bar{W}_\infty^* < \infty)$$

- Il massimo log-ritorno atteso dato $\bar{\mathcal{F}}_t$ è dato da

$$\bar{W}_t^* = \mathbf{E}[\bar{w}_t^*] = \sup_{\mathbf{b} \in \bar{\mathcal{F}}_t} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X})] = \mathbf{E}[\log(\bar{\mathbf{b}}_t^{*t} \mathbf{X})]$$

Inoltre,

$$\bar{W}_t^* \uparrow \bar{W}_\infty^* \quad \text{per } t \rightarrow \infty$$

Ci chiediamo ora se l'estremo superiore nella definizione di tasso di raddoppio ottimale viene effettivamente raggiunto nel semplice dei portafogli ammissibili. Enunciamo quindi due teoremi, dimostrati in [AC88] che mostrano risultati di compattezza e continuità per portafogli log-ottimali.

Dato un portafoglio fissato β tale che $\beta_j > 0$ per ogni j , definiamo il vettore dei rendimenti riscalato come

$$U = \mathbf{u}(\mathbf{X}), \quad \text{dove } \mathbf{u}(\mathbf{x}) = \frac{\mathbf{x}}{\beta^t \mathbf{x}}$$

Quindi U è la proiezione di \mathbf{X} sul semplice

$$\mathcal{U} = \{\mathbf{u} = (u_j)_{1 \leq j \leq m} \in \mathbb{R}_+^m : \beta^t \mathbf{u} = 1\}$$

Se $\mathbf{X} = 0$ poniamo $\mathbf{u}(0) = u_0$ per u_0 fissato. Su \mathcal{U} abbiamo una misura immagine di \mathbf{P} tramite $\mathbf{u}(\mathbf{x})$ in modo che

$$\mathbf{Q}(U \in A) = \mathbf{P}(\mathbf{u}(\mathbf{X}) \in A) = \mathbf{P}(\mathbf{u}^{-1}(A))$$

Dato che $\mathbf{X} = (\beta^t \mathbf{X})\mathbf{u}(\mathbf{X})$ possiamo scomporre il tasso di raddoppio grazie alle proprietà del logaritmo, ottenendo $\mathbf{E}_{\mathbf{P}}[\log \mathbf{b}^t \mathbf{X}] = \mathbf{E}_{\mathbf{P}}[\log \beta^t \mathbf{X}] + \mathbf{E}_{\mathbf{P}}[\log \mathbf{b}^t \mathbf{u}(\mathbf{X})]$ da cui $w(\mathbf{b}, \mathbf{P}) = r(\mathbf{P}) + w(\mathbf{b}, \mathbf{Q})$. Ponendo quindi $w^*(\mathbf{Q}) = \sup_{\mathbf{b}} \mathbf{E}_{\mathbf{Q}}[\log \mathbf{b}^t U]$ otteniamo che $w^*(\mathbf{P}) = r(\mathbf{P}) + w^*(\mathbf{Q})$. Abbiamo quindi diviso il nostro tasso di raddoppio in due addendi, uno fissato e l'altro che dipende dalla distribuzione immagine \mathbf{Q} .

Teorema 2.8. *Il tasso di raddoppio ottimale atteso $w^*(\mathbf{Q})$ è convesso, limitato tra 0 e $\max_j (-\log \beta_j)$ e uniformemente continuo quando lo spazio \mathcal{Q} delle misure di probabilità su \mathcal{U} è dotato della topologia di convergenza debole. L'insieme dei portafogli log-ottimali $\mathbf{B}^*(\mathbf{Q})$ è un sottoinsieme non vuoto, compatto e convesso di \mathcal{B} per ogni distribuzione \mathbf{Q} su \mathcal{U} , e un portafoglio log-ottimale $\mathbf{b}^*(\mathbf{Q}) \in \mathbf{B}^*(\mathbf{Q})$ può essere scelto per ogni $\mathbf{Q} \in \mathcal{Q}$ in modo che $\mathbf{b}^*(\mathbf{Q})$ sia misurabile in \mathbf{Q} .*

Teorema 2.9. *L'insieme $\text{Gr}(\mathbf{B}^*) = \{(\mathbf{Q}, \mathbf{b}^*) : \mathbf{b}^* \in \mathbf{B}^*(\mathbf{Q})\}$ è chiuso in $\mathcal{Q} \times \mathcal{B}$. Di conseguenza, una qualsiasi selezione di portafogli log-ottimali $\mathbf{Q} \mapsto \mathbf{b}^*(\mathbf{Q}) \in \mathbf{B}^*(\mathbf{Q})$ è continua in ogni $\mathbf{Q} \in \mathcal{Q}$ tale che $\mathbf{B}^*(\mathbf{Q}) = \{\mathbf{b}^*(\mathbf{Q})\}$ è un singolo.*

In generale $w^*(\mathbf{P})$ non è nè limitato nè continuo rispetto alla topologia della convergenza debole. Si ha però che se il supporto di \mathbf{P} è un sottoinsieme chiuso \mathcal{K} di \mathbb{R}_+^m valgono le seguenti implicazioni

- $w^*(\mathbf{P})$ è semicontinuo inferiormente e limitato inferiormente se e solo se \mathcal{K} è limitato e lontano da 0.
- $w^*(\mathbf{P})$ è semicontinuo superiormente e limitato superiormente se e solo se \mathcal{K} è limitato.
- $w^*(\mathbf{P})$ è uniformemente continuo e limitato se e solo se \mathcal{K} è limitato ed è lontano da 0.

Dove per lontano da 0 intendiamo che la distanza (nello spazio metrico \mathbb{R}_+^m) tra 0 e il sottoinsieme \mathcal{K} (che è chiuso) è maggiore di 0.

2.3.2 Principio di Ottimalità Asintotica

Normalizziamo ponendo il capitale iniziale $S_0 = 1$. Consideriamo quindi il mercato come una successione di vettori dei rendimenti $\{X_i\}_{0 \leq i < \infty}$ definiti su uno spazio di probabilità perfetto $(\Omega, \mathcal{F}, \mathbf{P})$. Ogni giorno investiamo il capitale con un portafoglio (variabile col tempo) \mathbf{b}_t .

L'obiettivo è massimizzare il valore atteso del tasso di raddoppio del capitale

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}[\log(S_n)]$$

Il portafoglio viene scelto ogni giorno sulla base dell'informazione fino a quel momento, cioè condizionando rispetto alle sigma algebre \mathcal{F}_t che rappresentano ciò che conosciamo fino al momento t . \mathbf{b}_t deve essere quindi \mathcal{F}_t -misurabile.

Sia ora \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

$$w_t^* = w^*(\mathbf{P}_t) = \mathbf{E}[\log(\mathbf{b}_t^{*t} \mathbf{X}) | \mathcal{F}_t] = \sup_{\mathbf{b} \in \mathcal{F}_t} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X}) | \mathcal{F}_t]$$

Per il teorema 2.7 il valore atteso di w_t^* coincide col tasso di raddoppio ottimale

$$W_t^* = \mathbf{E}[w_t^*] = \mathbf{E}[\log(\mathbf{b}_t^{*t} \mathbf{X})] = \sup_{\mathbf{b} \in \mathcal{F}_t} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X})]$$

Dimostriamo che $\{\mathbf{b}_t^*\}_{0 \leq t < \infty}$ è asintoticamente ottimale.

Teorema 2.10 (Principio di Ottimalità Asintotica). *Supponiamo che i $\{X_t\}_{0 \leq t < \infty}$ siano definiti su uno spazio di probabilità perfetto $(\Omega, \mathcal{F}, \mathbf{P})$, \mathcal{F}_t è una successione di sotto- σ -algebre di \mathcal{F} tali che $\sigma(X_0, \dots, X_{t-1}) \subseteq \mathcal{F}_t$ per ogni t . Siano quindi i capitali accumulati dopo n giorni ottenuti applicando il portafoglio \mathbf{b}_t^* e un qualsiasi altro \mathbf{b}_t rispettivamente*

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

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

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

Dimostrazione. 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}_t -misurabile e siccome \mathbf{b}_t^* è log-ottimale condizionato a \mathcal{F}_t possiamo scrivere la condizione di Kuhn-Tucker

$$\mathbf{E} \left[\frac{\mathbf{b}_n^t \mathbf{X}}{\mathbf{b}_n^{*t} \mathbf{X}} \middle| \mathcal{F}_n \right] \leq 1 \tag{2.3}$$

Si ottiene quindi che

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

Quindi $\left\{ \frac{S_n}{S_n^*}, \mathcal{F}_n \right\}_{0 \leq n < \infty}$ è una supermartingala non negativa e converge quindi quasi certamente a una variabile aleatoria Y (Teorema di convergenza di Lévy per martingale). Dal lemma di Fatou, inoltre

$$1 = \mathbf{E} \left[\frac{S_0}{S_0^*} \right] \geq \mathbf{E} \left[\frac{S_n}{S_n^*} \right] \geq \lim_{n \rightarrow \infty} \downarrow \mathbf{E} \left[\frac{S_n}{S_n^*} \right] \geq \mathbf{E}[Y]$$

Da $\mathbf{E} \left[\frac{S_n}{S_n^*} \right] \leq 1$ segue, usando la disuguaglianza di Markov, che per $r_n > 0$

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

Prendiamo r_n in modo che $\sum_n \frac{1}{r_n} < \infty$. Si ha così che

$$\mathbf{P} \left\{ \frac{S_n}{S_n^*} \geq r_n \right\} \leq \sum_{n=0}^{\infty} \frac{1}{r_n} < \infty$$

Dal lemma di Borel-Cantelli $\frac{S_n}{S_n^*} < r_n$ per n grande. In particolare, ponendo $r_n = e^{n\epsilon}$, $\epsilon > 0$ otteniamo che

$$\mathbf{P} \left\{ \frac{1}{n} \log \frac{S_n}{S_n^*} \geq \epsilon \text{ frequentemente} \right\} = 0$$

Siccome ϵ è arbitrariamente piccolo concludiamo infine che

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

□

2.3.3 Proprietà di Equipartizione Asintotica

In questa sezione estendiamo i primi risultati del capitolo riguardo a tassi di raddoppio ottimali, mostrati solo per mercati con outcomes i.i.d., a mercati ergodici e stazionari.

Sia quindi $\mathbf{X}(\omega) \in \mathbb{R}_+^m$ un vettore di rendimenti e sia T una trasformazione invertibile e metricamente transitiva che preserva la misura, definita su uno spazio di probabilità perfetto $(\Omega, \mathcal{F}, \mathbf{P})$. Sia quindi $\mathbf{X}_t(\omega) = \mathbf{X}(T^t(\omega))$. Data l'invertibilità, possiamo anche considerare la successione indicizzata in \mathbb{Z} : $\{\mathbf{X}_t\}_{-\infty < t < \infty}$.

Per la stazionarietà del processo, possiamo liberamente traslare temporalmente la nostra successione ad ogni passo: sia \mathbf{b}_t^* 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

$\bar{\mathcal{F}}_t = T^t \mathcal{F} = \sigma(\mathbf{X}_{-1}, \dots, \mathbf{X}_{-t})$. Questi portafogli raggiungono i rendimenti logaritmici condizionali ottimali, rispettivamente

$$w_t^* = \sup_{\mathbf{b} \in \mathcal{F}_t} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X}_t) | \mathcal{F}_t] \quad \text{e} \quad \bar{w}_t^* = \sup_{\mathbf{b} \in \bar{\mathcal{F}}_t} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X}_0) | \bar{\mathcal{F}}_t]$$

Denotiamo con $W_t^* = \mathbf{E}[\log(\mathbf{b}_t^{*t} \mathbf{X}_t)]$ e $\bar{W}_t^* = \mathbf{E}[\log(\bar{\mathbf{b}}_t^{*t} \mathbf{X}_0)]$ i rendimenti logaritmici ottimali. Allora, per stazionarietà (invarianza della distribuzione cumulata), vale

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

Se $\bar{\mathbf{b}}_\infty^*$ è il portafoglio log-ottimale per il periodo zero, basandoci sulle informazioni della σ 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 $\bar{W}_\infty^* = W^*(\mathbf{X}_0 | \mathbf{X}_{-1}, \mathbf{X}_{-2}, \dots)$.

Sia infine $S_n^* = \prod_{0 \leq t < n} \mathbf{b}_t^{*t} \mathbf{X}_t$ il capitale dopo n periodi di investimento.

Teorema 2.11 (Proprietà di Equipartizione Asintotica). *Se la successione dei vettori dei rendimenti degli asset $\{\mathbf{X}_t\}$ è un processo ergodico e stazionario, il capitale cresce in modo esponenziale (limitato dall'alto dalla strategia di investimento log-ottimale), 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. La σ -algebra di informazione $\mathcal{F}_t = \sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$ è approssimato 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 quindi $\mathbf{b}_t^{(k)}$ e $\mathbf{b}_t^{(\infty)}$ i corrispondenti portafogli log-ottimali e i rispettivi capitali dopo n periodi

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

Ricapitolando, $S_n^{(k)}, S_n^*, S_n^{(\infty)}$ rappresentano la crescita del capitale su n giorni di utilizzo dei portafogli log-ottimali quando l'investitore ha possibilità di guardare, ad ogni passo,

rispettivamente al k -passato (ma non prima del periodo 0), fino al periodo 0 e in tutto l'infinito passato.

Osserviamo che $\mathbf{b}_t^{(k)}(\omega) = \mathbf{b}_k^*(T^{t-k}\omega)$ se $t \geq k$ per stazionarietà. Espandiamo

$$\frac{1}{n} \log(S_n^{(k)}) = \frac{1}{n} \log(S_k^*) + \frac{1}{n} \sum_{k \leq t < n} \log(\mathbf{b}_t^{(k)t} \mathbf{X}_t)$$

dal teorema ergodico (con $g_t = \log(\mathbf{b}_t^{(k)t})$) segue che

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

La successione $\{\log(\mathbf{b}_t^{(\infty)t} \mathbf{X}_t)\}$ è stazionaria ergodica e $\mathbf{b}_0^{(\infty)} = \bar{\mathbf{b}}_\infty^*$. Dal teorema ergodico segue ancora che

$$\frac{1}{n} \log(S_n^{(\infty)}) = \frac{1}{n} \sum_{0 \leq t < n} \log(\mathbf{b}_t^{(\infty)t} \mathbf{X}_t) \rightarrow \bar{W}_\infty^* = \mathbf{E}[\log(\bar{\mathbf{b}}_\infty^{*t} \mathbf{X}_0)] \quad \text{q.c.}$$

$\mathcal{F}_t^{(k)} \subseteq \mathcal{F}_t$ 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. Dal principio di ottimalità asintotica segue che

$$\limsup_n \frac{1}{n} \log\left(\frac{S_n^{(k)}}{S_n^*}\right) \leq 0 \quad \text{e} \quad \limsup_n \frac{1}{n} \log\left(\frac{S_n^*}{S_n^{(\infty)}}\right) \leq 0 \quad \text{q.c.}$$

Quindi otteniamo la catena di disuguaglianze

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

La tesi segue da $W_k^* = \bar{W}_k^* \uparrow \bar{W}_\infty^*$ per $k \rightarrow \infty$. □

Per ultimo enunciamo un risultato di ottimalità competitiva, un argomento più avanzato che non tratteremo. Il portafoglio log-ottimale non garantisce di avere quasi certamente un rendimento migliore di ogni altro portafoglio. Tuttavia si può dimostrare che la probabilità di fare meglio di ogni altro investitore è maggiore di $\frac{1}{2}$

Teorema 2.12. *Sia S^* il capitale dopo un periodo di investimento usando il portafoglio log-ottimale e sia S quello dato da un qualsiasi altro portafoglio. Sia quindi U^* una variabile aleatoria indipendente da \mathbf{X} e uniforme su $[0, 2]$, e sia V una variabile aleatoria indipendente da U^* e \mathbf{X} con $V \geq 0$, $\mathbf{E}[V] = 1$. Allora*

$$\mathbf{P}(VS \geq U^* S^*) \leq \frac{1}{2}$$

Nota. In questo vaso V e U^* servono a introdurre una casualità nel capitale iniziale, in modo tale da ridurre gli effetti di piccole differenze iniziali, cosicché l'unico dato che influenza la probabilità di vincere contro l'altro investitore sia il rapporto $\frac{S}{S^*}$.

Capitolo 3

Portafogli Universali di Cover

Consideriamo come nei precedenti capitoli una successione di vettori di rendimento di m titolo $\{\mathbf{x}_i\}$, $\mathbf{x}_i \in \mathbb{R}_+^m$. In particolare x_{ij} è il rapporto tra i prezzi nel giorno i dell'asset j . Nella realtà è difficile trovare il portafoglio log-ottimale senza conoscere la distribuzione di probabilità legata al processo stocastico che modella il mercato. Questa distribuzione è per sua natura ignota e al più decidibile in maniera approssimativa con varie metodologie econometriche. Ci dobbiamo quindi affidare a strategie che asintoticamente raggiungono o si avvicinano il più possibile alla situazione ottimale.

Un approccio è quello di immaginare di avere a disposizione una quantità infinita di investitori, ognuno dei quali utilizzerà un portafoglio costante nel tempo (detto portafoglio costantemente ribilanciato), per poi mescolare le loro strategie. Siccome uno di loro farà esponenzialmente meglio degli altri mostreremo che possiamo raggiungere la performance del migliore a meno di un fattore $n^{\frac{m-1}{2}}$, dove n è il numero di giorni considerati.

Un secondo approccio è invece quello di un gioco contro un avversario che ad ogni turno sceglie il successivo vettore di stock. Definiremo quindi ad ogni periodo un portafoglio $\hat{\mathbf{b}}_i(\mathbf{x}_{i-1}, \dots, \mathbf{x}_1)$ dipendente solo dal passato. L'avversario, che conosce la nostra strategia, sceglie \mathbf{x}_i in modo tale da farci performare il peggio possibile. Indichiamo con $\hat{\mathbf{b}}_i(\mathbf{x}^{i-1})$ il portafoglio al giorno i e la successione di vettori fino al giorno n come \mathbf{x}^n .

Il miglior portafoglio costante $\mathbf{b}^*(\mathbf{x}^n)$ raggiunge un capitale di

$$S_n(\mathbf{x}^n) = \max_{\mathbf{b}} \prod_{i=1}^n \mathbf{b}^t \mathbf{x}_i$$

mentre il portafoglio variabile $\hat{\mathbf{b}}_n(\mathbf{x}^n)$ ci da

$$\hat{S}_n(\mathbf{x}^n) = \prod_{i=1}^n \hat{\mathbf{b}}_i^t(\mathbf{x}^{i-1}) \mathbf{x}_i$$

Mostreremo che è possibile trovare una strategia di questo tipo che nel caso peggiore si avvicina a S_n^* a meno di un fattore $V_n \approx n^{-\frac{m-1}{2}}$. Questa strategia dipende da n , l'orizzonte del gioco.

3.1 Portafogli universali a orizzonte finito

In questa sezione, dimostreremo i risultati relativi all'ultimo approccio mostrato. Diamo prima un lemma preliminare

Lemma 3.1. *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 $\frac{x}{0} = \infty$ per $x > 0$ e escludiamo per ipotesi che per qualche i $\frac{p_i}{q_i} = \frac{0}{0}$. In tal caso, infatti, possiamo eliminare tale indice dal termine a sinistra e quindi anche a destra, dove non sarebbe definito.

Dimostrazione. Sia I l'indice che minimizza il termine a destra. Assumiamo che $p_I > 0$ (altrimenti la tesi è ovvia). Inoltre, se $q_i = 0$ si ha che ogni altro q_i è nullo, altrimenti I non realizzerebbe il minimo, e quindi entrambi i lati sarebbero infiniti. Possiamo assumere che anche $q_I > 0$. Allora

$$\frac{\sum_{i=1}^m p_i}{\sum_{i=1}^m q_i} = \frac{p_I(1 + \sum_{i \neq I} \frac{p_i}{p_I})}{q_I(1 + \sum_{i \neq I} \frac{q_i}{q_I})} \geq \frac{p_I}{q_I}$$

dove l'ultimo passaggio segue da

$$\frac{p_i}{q_i} \geq \frac{p_I}{q_I} \longrightarrow \frac{p_i}{p_I} \geq \frac{q_i}{q_I}$$

per ogni i . □

Per dimostrare il prossimo teorema useremo una notazione particolare. Indicheremo infatti con $\hat{\mathbf{b}}, \mathbf{b}, \mathbf{x}' \in \mathbb{R}_+^{m^n}$ i portafogli e i vettori degli stock cumulativi di n giorni, in modo tale che le componenti dei portafogli siano tutti i possibili prodotti (m^n) della forma $b_1^{n_1} b_2^{n_2} \dots b_m^{n_m}$, cosicché

$$\frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} = \frac{\prod_{i=1}^n \hat{\mathbf{b}}_i^t \mathbf{x}_i}{\prod_{i=1}^n \mathbf{b}_i^t \mathbf{x}_i} = \frac{\hat{\mathbf{b}}^t \mathbf{x}}{\mathbf{b}^t \mathbf{x}'}$$

Teorema 3.2 (Portafoglio universale a orizzonte finito). *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. Dimostriamo il teorema nel caso $m = 2$, per m generico la soluzione è una generalizzazione di questa in modo completamente analogo.

Denotiamo gli stock con 1 e 2, il passaggio chiave è quello di esprimere

$$S_n(\mathbf{x}^n) = \prod_{i=1}^n \mathbf{b}_i^t \mathbf{x}_i$$

che è un prodotto di somme, come una somma di prodotti. Espandendo il lato a destra, ogni termine nella somma corrisponde a una successione di prezzi di stock relativi al primo o al secondo asset, moltiplicato per la proporzione b_{i1}, b_{i2} che la strategia pone rispettivamente per lo stock 1 e lo stock 2 al tempo i . Possiamo quindi vedere S_n come la somma su tutte le 2^n possibili n sequenze di 1 e 2 di un prodotto, costituito da prodotti di elementi di portafogli per prezzi relativi, ossia

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

Denotiamo con $w(j^n)$ il prodotto $\prod_{i=1}^n b_{ij_i}$, la frazione totale di capitale investita sulla sequenza j^n e sia

$$x(j^n) = \prod_{i=1}^n x_{ij_i}$$

il corrispondente ritorno per questa successione, allora possiamo scrivere

$$S_n(\mathbf{x}^n) = \sum_{j^n \in \{1,2\}^n} w(j^n) x(j^n)$$

Similmente applichiamo le stesse notazioni e considerazioni sul portafoglio casuale e su quello costantemente ribilanciato, cosicché

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

Applicando ora il lemma 3.1 abbiamo che

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

Il problema di massimizzare il rapporto tra i capitali si traduce quindi con l'assicurarsi che la frazione scommessa su una successione di stock dal portafoglio universale sia uniformemente vicina alla proporzione scommessa sulla stessa usando \mathbf{b}^* . L'obiettivo di tutto questo procedimento era ridurre un mercato su più periodi a un mercato su singolo periodo e successioni di stock.

Calcoliamo prima il peso $w^*(j^n)$ dato al miglior portafoglio costantemente ribilanciato \mathbf{b}^* . Osserviamo che per un portafoglio di questo tipo (che è costante nel tempo) abbiamo pesi della forma

$$w(j^n) = \prod_{i=1}^n b_{ij_i} = b^k (1-b)^{n-k}$$

dove k è il numero di volte che 1 compare in j^n .

Differenziando rispetto a b notiamo che

$$w^*(j^n) = \max_{0 \leq b \leq 1} b^k (1-b)^{n-k} = \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}$$

che è raggiunto da

$$\mathbf{b}^* = \left(\frac{k}{n}, \frac{n-k}{n}\right)$$

Notiamo che però in questo modo $\sum w^*(j^n) > 1$, mentre la nostra strategia in partenza deve avere $\sum \hat{w}(j^n) = 1$. Il modo più efficace per farlo, in modo da proteggersi da tutti i risultati peggiori, è quello di rinormalizzare e scegliere i $\hat{w}(j^n)$ proporzionali ai $w^*(j^n)$, cosicché il peggior rapporto $\frac{\hat{w}(j^n)}{w^*(j^n)}$ è massimizzato.

Definiamo quindi

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

e

$$\hat{w}(j^n) = V_n \left(\frac{k(j^n)}{n}\right)^{k(j^n)} \left(\frac{n-k(j^n)}{n}\right)^{n-k(j^n)} \quad (3.1)$$

In questo modo si ha $\hat{w}(j^n) \geq 0$ e $\sum_{j^n \in \{1,2\}^n} \hat{w}(j^n) = 1$. V_n è quindi la normalizzazione che rende $\hat{w}(j^n)$ una distribuzione di probabilità. Inoltre, da prima, per ogni successione \mathbf{x}^n

$$\begin{aligned} \frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} &\geq \min_{j^n} \frac{\hat{w}(j^n)}{w^*(j^n)} \\ &= \min_k \frac{V_n \left(\frac{k}{n}\right)^k \left(\frac{n-k}{n}\right)^{n-k}}{b^{*k} (1-b^*)^{n-k}} \\ &\geq V_n \end{aligned}$$

Abbiamo cioè che

$$\max_{\mathbf{b}} \min_{\mathbf{x}^n} \frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \geq V_n$$

Questo dimostra l'esistenza di un portafoglio che raggiunge asintoticamente il capitale del miglior portafoglio costantemente ribilanciato a meno di un fattore V_n . Per completare la dimostrazione vediamo che non si può fare di meglio, ossia che esiste una successione \mathbf{x}^n per cui ogni rapporto tra il capitale di una qualsiasi strategia e S_n^* è limitato esattamente da V_n .

Per ogni $j^n \in \{1, 2\}^n$ definiamo il corrispondente vettore di stock estremae $\mathbf{x}^n(j^n)$ come

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

Siano quindi $\mathbf{e}_1 = (1, 0)^t$ e $\mathbf{e}_2 = (0, 1)^t$ e sia

$$\mathcal{K} = \{\mathbf{x}(j^n) : j^n \in \{1, 2\}^n, \mathbf{x}_{ij_i} = \mathbf{e}_{j_i}\}$$

l'insieme delle successioni estremali. Queste sono 2^n e per ognuna di esse c'è solo un asset che ha un ritorno non nullo mentre il capitale investito nell'altro è perso. Quindi, dopo n periodi, il capitale ottenuto da una successione estrema è pari al prodotto delle quantità investite negli stock j_1, j_2, \dots, j_n , cioè $S_n(\mathbf{x}^n(j^n)) = \prod_i b_{j_i} = w(j^n)$. Il fatto che i ritorni sono del tipo 0 – 1 implica che per $\mathbf{x}^n \in \mathcal{K}$

$$\sum_{j^n \in \{1, 2\}^n} S_n(\mathbf{x}^n(j^n)) = \sum_{j^n \in \{1, 2\}^n} w(j^n) = 1$$

Per ogni sequenza estrema $\mathbf{x}^n(j^n) \in \mathcal{K}$, il miglior portafoglio costantemente ribilanciato è

$$\mathbf{b}^*(\mathbf{x}^n(j^n)) = \left(\frac{n_1(j^n)}{n}, \frac{n_2(j^n)}{n} \right)^t$$

dove $n_1(j^n)$ è il numero di occorrenze di 1 in j^n . Il corrispondente capitale al termine di n periodi è

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

dove l'ultimo passaggio segue dall'equazione 3.1. Quindi

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

Abbiamo così la seguente disuguaglianza, valida per ogni $\{\mathbf{b}_i\}_{i=1}^n$

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

dove la prima disuguaglianza segue dal fatto che il minimo di una successione è minore o uguale alla media pesata. Ne segue che

$$\max_{\mathbf{b}} \min_{\mathbf{x}^n \in \mathcal{K}} \frac{S_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} \leq V_n$$

□

Questa strategia è chiaramente dipendente da n . possiamo a questo punto riscriverla in termini incrementali. Consideriamo il peso $\hat{\mathbf{b}}_{i,1}$ assegnato dall'algoritmo sullo stock 1 al tempo i data la successione di stock precedenti \mathbf{x}^{i-1} . Possiamo calcolare tale peso sommando su tutte le sequenze j^n che hanno 1 in posizione i , ottenendo

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

dove

$$\hat{w}(j^i) = \sum_{j^n: j^i \subseteq j^n} w(j^n)$$

è il peso totale su tutte le sequenze che iniziano per j^i e

$$x(j^{i-1}) = \prod_{k=1}^{i-1} x_{kj_k}$$

è il ritorno di queste sequenze. Ciò che stiamo facendo è quindi ad ogni passo mediare su tutte le successioni che hanno 1 in quella posizione, in base ai pesi definiti secondo il processo sopra.

Utilizzando i risultati in [OC98] si ottiene che

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

per m asset. In particolare per $m = 2$

$$V_n \sim \sqrt{\frac{2}{\pi n}}$$

e

$$\frac{1}{2\sqrt{n+1}} \leq V_n \leq \frac{2}{\sqrt{n+1}}$$

per ogni n . Per $m = 2$ il portafoglio appena mostrato ottiene quindi un capitale $\hat{S}_n(\mathbf{x}^n)$ tale che

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

per ogni successione \mathbf{x}^n .

3.2 Portafoglio Universale a orizzonte infinito

Vogliamo ora trovare una strategia a orizzonte infinito, cioè che non dipenda dal numero di giorni in cui verrà utilizzata. Come detto in precedenza, immaginiamo di avere infiniti manager, ognuno dei quali utilizzerà un portafoglio costantemente ribilanciato \mathbf{b} e successivamente di utilizzare un approccio che sia un misto di questi.

Sia $\mathcal{B} = \{\mathbf{b} \in \mathbb{R}^m : b_i \geq 0, \sum_i b_i = 1\}$ l'm simpleso, 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} .

Sia $S_n(\mathbf{b}, \mathbf{x}^n) = \prod_{i=1}^n \mathbf{b}^t \mathbf{x}_i$ il capitale dopo n periodi ottenuto dal portafoglio constantly rebalanced \mathbf{b} sulla sequenza \mathbf{x}^n . Considereremo il seguente portafoglio al periodo $i + 1$ dato da

$$\hat{\mathbf{b}}_{i+1}(\mathbf{x}^i) = \frac{\int_{\mathcal{B}} \mathbf{b} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})}{\int_{\mathcal{B}} S_i(\mathbf{b}, \mathbf{x}^i) d\mu(\mathbf{b})} \quad (3.2)$$

Notiamo che

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

Il prodotto $\prod_i \hat{\mathbf{b}}_i^t \mathbf{x}_i$ è telescopico, semplificando otteniamo che il capitale $\hat{S}_n(\mathbf{x}^n)$ dato da

$$\begin{aligned} \hat{S}_n(\mathbf{x}^n) &= \prod_{i=1}^n \hat{\mathbf{b}}_i^t(\mathbf{x}^{i-1}) \mathbf{x}_i \\ &= \int_{\mathbf{b} \in \mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n) d\mu(\mathbf{b}) \end{aligned} \quad (3.3)$$

Un altro modo di interpretare questo procedimento è che stiamo investendo una parte del capitale $d\mu(\mathbf{b})$ nel portafoglio \mathbf{b} che ci darà capitale $S_n(\mathbf{b}, \mathbf{x}^n)$, per un ritorno complessivo pari a $\int_{\mathbf{b} \in \mathcal{B}} S_n(\mathbf{b}, \mathbf{x}^n) d\mu(\mathbf{b})$. Quindi, $\hat{\mathbf{b}}_{i+1}(\mathbf{x}^i)$ è il portafoglio che media sulle performance dei vari manager.

Dimostriamo ora un lemma preliminare che assicura un limite su $\frac{\hat{S}_n}{S_n^*}$ in funzione della distribuzione $\mu(\mathbf{b})$.

Lemma 3.3. *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 $\hat{\mathbf{b}}(\cdot)$, dato da*

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

allora

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

Dimostrazione. In modo completamente analogo al teorema 3.2 (su successioni j^n in generale appartenenti a $\{1, \dots, m\}$) possiamo scrivere

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

dove $w^*(j^n) = \prod_{i=1}^n b_{j_i}^*$ è la quantità investita nella sequenza j^n e $x(j^n) = \prod_{i=1}^n x_{i j_i}$ è il corrispondente ritorno. Similmente, scriviamo

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

dove, in questo caso $\hat{w}(j^n) = \int_{\mathcal{B}} \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})$. Applicando ora il lemma 3.1 abbiamo

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

□

Applichiamo ora questo lemma quando $\mu(\mathbf{b})$ è la distribuzione Dirichlet($\frac{1}{2}$).

Teorema 3.4. *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 \mathbf{x}^n .

Dimostrazione. Come mostrato precedentemente nel teorema 3.2, possiamo dimostrare che il portafoglio costantemente ribilanciato ottimale \mathbf{b}^* sulla sequenza j^n pone pesi

$$\prod_{i=1}^n b_{j_i}^* = \binom{k}{n}^k \left(\frac{n-k}{n}\right)^{n-k} = 2^{-nH(\frac{k}{n})}$$

dove k è il numero di indici per cui $j_i = 1$. Nel caso della distribuzione Dirichlet($\frac{1}{2}$), definita per m variabili come

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

dove $\Gamma(x) = \int_0^\infty \exp^{-t} t^{x-1} dt$ indica la funzione Gamma, possiamo calcolare esattamente il numeratore. Per semplicità, nel caso $m = 2$

$$d\mu(b) = \frac{1}{\pi} \frac{1}{\sqrt{b(1-b)}} db, \quad 0 \leq b \leq 1$$

dove b è la frazione investita sul primo asset. Ora prendiamo una qualsiasi sequenza $j^n \in \{1, 2\}^n$ e consideriamo la quantità investita in quella successione,

$$b(j^n) = \prod_{i=1}^n b_{j_i} = b^l (1-b)^{n-l}$$

dove l è il numero di indici con $j_i = 1$. Quindi

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

dove $B(\lambda_1, \lambda_2)$ è la funzione beta, definita come

$$\begin{aligned} B(\lambda_1, \lambda_2) &= \int_0^1 x^{\lambda_1-1} (1-x)^{\lambda_2-1} dx \\ &= \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1 + \lambda_2)} \end{aligned}$$

Sappiamo che per ogni n intero, $\Gamma(n+1) = n!$ e $\Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}$. Possiamo calcolare $B\left(l + \frac{1}{2}, n - l + \frac{1}{2}\right)$ per semplice ricorsione usando l'integrazione per parti o usando la funzione Γ . Otteniamo in ogni caso

$$B\left(l + \frac{1}{2}, n - l + \frac{1}{2}\right) = \frac{\pi}{2^{2n}} \frac{\binom{2n}{n} \binom{n}{l}}{\binom{2n}{2l}}$$

Combinando questi risultati con il lemma precedente otteniamo che

$$\begin{aligned} \frac{\hat{S}_n(\mathbf{x}^n)}{S_n^*(\mathbf{x}^n)} &\geq \min_{j^n} \frac{\int_{\mathcal{B}} \prod_{i=1}^n b_{j_i} d\mu(\mathbf{b})}{\prod_{i=1}^n b_{j_i}^*} \\ &\geq \min_l \frac{\frac{1}{\pi} B\left(l + \frac{1}{2}, n - l + \frac{1}{2}\right)}{2^{-nH\left(\frac{l}{n}\right)}} \\ &\geq \frac{1}{2\sqrt{n+1}} \end{aligned}$$

usando i risultati in [CO96]. □

Concludiamo che per $m = 2$

$$\frac{\hat{S}_n}{S_n^*} \geq \frac{1}{\sqrt{2\pi}} V_n$$

per ogni n e ogni sequenza di prezzi di mercato. Quindi rimuovere l'orizzonte finito costa solo un fattore $\frac{1}{\sqrt{2\pi}}$. Asintoticamente, però,

$$\frac{1}{n} \ln(\hat{S}_n(\mathbf{x}^n)) - \frac{1}{n} \ln(S_n^*(\mathbf{x}^n)) \geq \frac{1}{n} \ln\left(\frac{V_n}{\sqrt{2\pi}}\right) \rightarrow 0.$$

e quindi il portafoglio universale ottiene lo stesso incremento asintotico del miglior portafoglio costantemente ribilanciato con conoscenza della distribuzione.

La scelta della densità Dirichlet $(\frac{1}{2}, \frac{1}{2})$ è dovuta al fatto che pone molto peso sui portafogli estremali, ossia quelli che distribuiscono quasi tutto il capitale attuale su un unico asset. Al primo ordine, scrivendo $\mathbf{x} = \mathbf{e} + \mathbf{r}$ con $\mathbf{e} = (1, \dots, 1)^t$, si ha che

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

Al primo ordine, quindi, conviene puntare tutto sull'asset che ci assicura $\mathbf{E}[x] = 1 + \mathbf{E}[r]$ massimo. L'approssimazione si giustifica col fatto che nel mondo reale x è generalmente vicino a 1 per un titolo.

Capitolo 4

Algoritmi di Predizione Universali

Nel seguente capitolo vogliamo ottenere risultati migliori rispetto alle strategie viste in precedenza, in cui ci avvicinavamo al miglior portafoglio costantemente ribilanciato.

Siamo interessati a trovare algoritmi che riescano col tempo a stimare sempre meglio la distribuzione di probabilità del processo stocastico che modella il mercato.

Dato quindi $\{\mathbf{X}_t\}$ per $t \in \mathbb{Z}$ processo stocastico ergodico e stazionario su uno spazio di probabilità perfetto $(\Omega, \mathcal{F}, \mathbf{P})$, ad ogni passo vogliamo costruire una stima della misura di probabilità per \mathbf{X}_t , sulla base della conoscenza fino a quel punto. Tale informazione a nostra disposizione sarà rappresentata da una sotto- σ -algebra $\mathcal{F}_t \supseteq \sigma(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$. Cerchiamo una stima che asintoticamente converga debolmente alla distribuzione effettiva, in modo tale che il tasso di raddoppio ottimale basato sull'informazione acquisita fino al tempo t converga a quello vero e proprio.

4.1 Raggiungimento del tasso di raddoppio ottimale

Vogliamo ora cercare sotto quali ipotesi sulle nostre stime di distribuzione possiamo trovare una strategia di portafoglio che ottenga il risultato ottimale.

Definizione 4.1. Diciamo che un vettore di rendimenti di mercato è sicuro se ogni asset ha valore atteso logaritmico maggiore di $-\infty$. In formule

$$\mathbf{E}[\log(X_{t,j})] > -\infty$$

Per ogni t e per ogni $1 \leq j \leq m$.

Teorema 4.1. Sia \mathbf{X}_t un processo ergodico e stazionario che rappresenti i rendimenti azionari di mercato, siano quindi \mathbf{b}_t^* e $\hat{\mathbf{b}}_t^*$ i portafogli log-ottimali derivanti rispettivamente dalla vera distribuzione di probabilità condizionata fino al tempo t $\mathbf{P}(d\mathbf{x}_t|\mathbf{X}^t)$ e da una stima $\hat{\mathbf{P}}(d\mathbf{x}_t|\mathbf{X}^t)$. Supponiamo inoltre che $\hat{\mathbf{P}}(d\mathbf{x}_t|\mathbf{X}^t)$ sia una versione traslata nel tempo di una distribuzione $\hat{\mathbf{P}}(d\mathbf{x}|\mathbf{X}^{-t})$ tale che

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

Se il mercato è sicuro, allora il portafoglio derivante dalla stima di probabilità ottiene lo stesso risultato asintotico di quello veritiero, cioè

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

Dimostrazione. L'AEP per il portafoglio log-ottimale \mathbf{b}_t^* e l'AOP applicata a $\hat{\mathbf{b}}_t^*$ ci danno le relazioni

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log S_n^* &= W(\mathbf{X}|\mathbf{X}^-) \quad \text{q.c.} \\ \limsup_{n \rightarrow \infty} \frac{1}{n} \log \left(\frac{\hat{S}_n^*}{S_n^*} \right) &\leq 0 \quad \text{q.c.} \end{aligned}$$

Da queste due otteniamo subito un limite superiore asintotico

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \hat{S}_n^* \leq W(\mathbf{X}|\mathbf{X}^-) = \lim_{n \rightarrow \infty} \frac{1}{n} \log S_n^* \quad \text{q.c.}$$

Per dimostrare l'altra parte, utilizzeremo il teorema ergodico generalizzato di Breiman. Siano $\bar{\mathbf{b}}_t^*$ e $\hat{\mathbf{b}}_t^*$ i portafogli log-ottimali rispettivamente per le distribuzioni $\mathbf{P}(d\mathbf{x}|\mathbf{X}^-)$ e $\hat{\mathbf{P}}(d\mathbf{x}|\mathbf{X}^{-t})$. Siccome quest'ultima converge debolmente quasi certamente alla prima, segue dal teorema 2.7 che il punto di accumulazione per $\hat{\mathbf{b}}_t^*$ è log-ottimale per $\mathbf{P}(d\mathbf{x}|\mathbf{X}^-)$. Quindi

$$\hat{\mathbf{b}}_t^{*t} \mathbf{X} \rightarrow \bar{\mathbf{b}}^{*t} \mathbf{X} \quad \text{q.c.}$$

Per dimostrare che $\frac{1}{n} \log \hat{S}_n^* \rightarrow W(\mathbf{X}|\mathbf{X}^-)$ dobbiamo mostrare che

$$\frac{1}{n} \sum_{0 \leq t < n} \hat{g}_t \circ T^t \rightarrow \mathbf{E}[g]$$

dove $\hat{g}_t = \log(\hat{\mathbf{b}}_t^{*t} \mathbf{X})$ e $g = \log(\bar{\mathbf{b}}_t^* \mathbf{X})$. Chiaramente $\hat{g}_t \rightarrow g$ quasi certamente e possiamo assumere che \mathbf{X} sia limitato, cosicché anche \hat{g}_t è limitato dall'alto da una costante. Per dimostrare che la successione $\{\hat{g}_t\}$ è dominata in L^1 è quindi sufficiente verificare la condizione $\mathbf{E}[\inf_t \hat{g}_t] > -\infty$. Si ha che $\inf_t \hat{g}_t \geq \min_j \log(X_{t,j}) > -\infty$ dato che $\mathbf{b}^t \mathbf{X}_t \geq \min_j X_{t,j}$ per ogni portafoglio \mathbf{b} e che il mercato è sicuro. Abbiamo quindi la tesi. \square

Possiamo, mediante un procedimento specifico, rimuovere l'ipotesi di mercato sicuro. Il fatto chiave è che per evitare di perdere tutto il capitale nel caso in cui esso venga suddiviso su asset che hanno rendimento 0 da un periodo all'altro, ci spostiamo nel simpleso dei portafogli ammissibili verso un portafoglio che pone massa su tutti i titoli. Sia β un portafoglio tale che $\beta_j > 0$ per ogni j e definiamo per ogni $0 \leq \lambda < 1$

$$\hat{S}_n^\lambda = \prod_{0 \leq t < n} \hat{\mathbf{b}}_t^{*\lambda t} \mathbf{X} \quad \text{dove } \hat{\mathbf{b}}_t^{*\lambda} = (1 - \lambda)\beta + \lambda \hat{\mathbf{b}}_t^*$$

La j -esima componente di $\hat{\mathbf{b}}_t^{*\lambda t}$ è limitata inferiormente dalla costante $(1 - \lambda)\beta_j$. Definiamo quindi, per ogni σ -algebra di informazione \mathcal{G}

$$W(\mathbf{X}|\mathcal{G}) = \mathbf{E}[\log(\mathbf{b}^{*\lambda t} \mathbf{X})]$$

dove $\mathbf{b}^{*\lambda} = (1 - \lambda)\boldsymbol{\beta} + \lambda\mathbf{b}^*$ e \mathbf{b}^* raggiunge l'estremo superiore in $\sup_{\mathbf{b} \in \mathcal{G}} \mathbf{E}[\log(\mathbf{b}^t \mathbf{X})|\mathcal{G}]$. Dal fatto che $\mathbf{b}_t^{*\lambda t} \mathbf{X} \geq \lambda\mathbf{b}^* \mathbf{X}$ vediamo che

$$W(\mathbf{X}|\mathcal{G}) + \log \lambda \leq W^\lambda(\mathbf{X}|\mathcal{G}) \leq W(\mathbf{X}|\mathcal{G})$$

Teorema 4.2. *Consideriamo un mercato ergodico e stazionario con distribuzione ignota. Se il capitale è frazionato ad ogni periodo di investimento t secondo il portafoglio $\hat{\mathbf{b}}_t^{*\lambda}$, allora il capitale risultante \hat{S}_n^λ cresce in modo esponenziale quasi certamente con un tasso di crescita limite ben definito.*

$$\frac{1}{n} \log \hat{S}_n^\lambda \rightarrow W^\lambda(\mathbf{X}|\mathbf{X}^-) \quad \text{q.c.}$$

dove

$$W^\lambda(\mathbf{X}|\mathbf{X}^-) = \mathbf{E}[\log(\bar{\mathbf{b}}_t^{*\lambda t} \mathbf{X})]$$

Dimostrazione. Usiamo anche questa volta il teorema ergodico generalizzato sulle variabili aleatorie

$$\begin{aligned} \hat{g}_t^\lambda &= \log \left(\frac{((1 - \lambda)\boldsymbol{\beta} + \lambda\hat{\mathbf{b}}_t^{*\lambda})^t \mathbf{X}_t}{\boldsymbol{\beta}^t \mathbf{X}} \right) \\ g^\lambda &= \log \left(\frac{((1 - \lambda)\boldsymbol{\beta} + \lambda\bar{\mathbf{b}}_t^{*\lambda})^t \mathbf{X}_t}{\boldsymbol{\beta}^t \mathbf{X}} \right) \end{aligned}$$

Siccome $\hat{g}_t^\lambda \rightarrow g^\lambda$ quasi certamente e $\{\hat{g}_t^\lambda\}$ è limitato tra $\log(1 - \lambda)$ e $\max_j \log\left(\frac{1}{\beta_j}\right)$, abbiamo

$$\frac{1}{n} \sum_{0 \leq t < n} \hat{g}_t^\lambda \circ T^t \rightarrow \mathbf{E}[g^\lambda] \quad \text{q.c.}$$

La tesi segue da

$$\begin{aligned} \frac{1}{n} \log \hat{S}_n^\lambda &= \frac{1}{n} \sum_{0 \leq t < n} \hat{g}_t^\lambda \circ T^t + \frac{1}{n} \sum_{0 \leq t < n} \log(\boldsymbol{\beta}^t \mathbf{X}) \\ &\rightarrow \mathbf{E}[g^\lambda] + \mathbf{E}[\log(\boldsymbol{\beta}^t \mathbf{X})] = W^\lambda(\mathbf{X}|\mathbf{X}^-) \quad \text{q.c.} \end{aligned}$$

□

Nella realtà, purtroppo, tali stime di probabilità non sono facili da trovare. Nei prossimi paragrafi, però, analizzeremo strategie simili che raggiungono gli stessi risultati.

4.2 Metodo delle Frequenze Empiriche

Ragioniamo inizialmente per semplicità su variabili aleatorie discrete J_t a valori in un alfabeto finito $\mathcal{J} = \{1, \dots, m\}$. Il metodo più immediato per stimare la distribuzione effettiva è quello di approssimare il processo ergodico e stazionario con precisione k calcolando le frequenze empiriche di ogni simbolo dati i precedenti k . Definiamo quindi la nostra stima di probabilità come

$$\hat{\mathbf{P}}_t(j|J^{-k}) = \frac{1 + c_t(j|J^{-k})}{m + c_t(J^{-k})} \quad (4.1)$$

dove

$$\begin{aligned} c_t(j|J^{-k}) &= \#\{\tau : 1 \leq \tau \leq t, (J_{-\tau-k}, \dots, J_{-\tau-1}, J_{-\tau}) = (J_{-k}, \dots, J_{-1}, j)\} \\ c_t(J^{-k}) &= \#\{\tau : 1 \leq \tau \leq t, (J_{-\tau-k}, \dots, J_{-\tau-1}) = (J_{-k}, \dots, J_{-1})\} \\ &= \sum_{1 \leq j \leq m} c_t(j|J^{-k}) \end{aligned}$$

Stiamo quindi facendo il rapporto tra il numero di volte in cui nel passato il simbolo j è comparso dopo il blocco J^{-k} e tutte le occorrenze di quest'ultimo. In ogni caso, per evitare di escludere la possibilità che compaia un simbolo mai visto, abbiamo dato un peso 1 a priori per tutti.

Questa strategia può essere usata per le corse dei cavalli ma presto vedremo come estenderla al caso del mercato azionario.

Notiamo anzitutto che data la stazionarietà del processo possiamo traslare temporalmente queste stime di distribuzione ottenendo sempre lo stesso risultato. Inoltre, vale il seguente risultato:

Teorema 4.3.

$$\hat{\mathbf{P}}_t(j|J^{-k}) \rightarrow \mathbf{P}(j|J^{-k}) \quad \text{q.c. per } t \rightarrow \infty$$

Dimostrazione. Utilizziamo anche in questo caso il teorema ergodico generalizzato di Breiman.

Fissata una successione di k termini $j^k = (j_{-k}, \dots, j_{-1})$ e $j \in \mathcal{J}$ consideriamo la funzione

$$g = \chi_{\{(J_{-k}, \dots, J_{-1}, J_0) = (j^k, j)\}}$$

Per stazionarietà è evidente che $\hat{g}_t \rightarrow g$ quasi certamente, dato anche che posso usare lo shift temporale liberamente.

Dal teorema ergodico otteniamo che

$$\frac{1}{n} \sum_{0 \leq t < n} g \circ T^t = \frac{c_t(j|j^k)}{n} \rightarrow \mathbf{P}(J_{-k} = j_{-k}, \dots, J_0 = j) \quad \text{q.c.}$$

Analogamente, considerando la funzione

$$g = \chi_{\{(J_{-k}, \dots, J_{-1}) = j^k\}}$$

otteniamo che

$$\frac{1}{n} \sum_{0 \leq t < n} g \circ T^t = \frac{c_t(j^k)}{n} \rightarrow \mathbf{P}(J_{-k} = j_{-k}, \dots, J_{-1} = j_{-1}) \quad \text{q.c.}$$

Concludiamo quindi che

$$\frac{c_t(j|j^k)}{c_t(j^k)} \rightarrow \frac{\mathbf{P}(J_{-k} = j_{-k}, \dots, J_0 = j)}{\mathbf{P}(J_{-k} = j_{-k}, \dots, J_{-1} = j_{-1})} = P(J_0 = j | J^{-k} = j^k) \quad \text{q.c.}$$

Togliendo il caso di successioni di misura nulla, inoltre, $c_t(j|j^k)$ e $c_t(j^k)$ divergono quasi certamente, quindi anche

$$\frac{c_t(j|j^k) + 1}{c_t(j^k) + m} = \frac{c_t(j|j^k) \left(1 + \frac{1}{c_t(j|j^k)}\right)}{c_t(j^k) \left(1 + \frac{m}{c_t(j^k)}\right)} \rightarrow P(J_0 = j | J^{-k} = j^k) \quad \text{q.c.}$$

□

Un procedimento completamente analogo può essere applicato, usando determinati accorgimenti, al caso dei vettori di rendimenti del mercato per m asset. In particolare per ogni $k \geq 0$ consideriamo una sotto- σ -algebra finita \mathcal{F}^{-k} su $\sigma(\mathbf{X}_{-1}, \mathbf{X}_{-2}, \dots, \mathbf{X}_{-k})$. Definiamo in questo caso la frequenza empirica come

$$\hat{\mathbf{P}}_t(d\mathbf{x} | \mathcal{F}^{-k}) = \frac{1 + \sum_{\tau \in I_t^{-k}} \delta_{\mathbf{X}_{-\tau}}(d\mathbf{x})}{\|\mathcal{F}^{-k}\| + \|I_t^{-k}\|}$$

dove

$$I_t^{-k} = \{\tau : 1 \leq \tau \leq t, (\mathbf{X}_{-\tau-k}, \dots, \mathbf{X}_{\tau-1}) \text{ e } (\mathbf{X}_{-k}, \dots, \mathbf{X}_{-1}) \\ \text{appartengono allo stesso atomo di } \mathcal{F}^{-k}\}$$

In modo analogo a prima si dimostra che queste stime convergono debolmente quasi certamente alla distribuzione vera e propria. Siccome si tratta di una sotto- σ -algebra finita basta verificare che la misura degli atomi di \mathcal{F}^{-k} secondo le stime converga a quella effettiva. Il procedimento è a questo punto identico al teorema precedente. (FORSE SCRIVERLO BENE?).

A questo punto ci possiamo accontentare di un determinato k (nella sperimentazione useremo $k = 1$, unico valore efficiente per i dati a disposizione) oppure possiamo applicare varie strategie, ognuna con il proprio k e mescolarle tra loro. In particolare supponiamo che le σ -algebre \mathcal{F}^{-k} siano tutte finite e che per $k \rightarrow \infty$ queste crescano con σ -algebra terminale \mathcal{F} .

Siano $\hat{\mathbf{b}}_t^k$ e $\bar{\mathbf{b}}^k$ i portafogli log-ottimali rispettivamente per $\hat{\mathbf{P}}_t(d\mathbf{x}_t | \mathcal{F}_t^{-k})$ e $\mathbf{P}(d\mathbf{x} | \mathcal{F}^{-k})$. Sia quindi $\boldsymbol{\beta}$ un portafoglio con $\beta_j > 0$ per ogni j e $0 \leq \lambda < 1$

$$\hat{S}_n^{k,\lambda} = \prod_{0 \leq t < n} \hat{\mathbf{b}}_t^{(k,\lambda)t} \mathbf{X}_t \quad \text{dove } \hat{\mathbf{b}}_t^{k,\lambda} = (1 - \lambda)\boldsymbol{\beta} + \lambda \hat{\mathbf{b}}_t^k$$

In modo completamente analogo a quanto fatto nella dimostrazione del teorema 4.2, usando il teorema ergodico sappiamo che

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \hat{S}_n^{k,\lambda} = W^\lambda(\mathbf{X}|\mathcal{F}^{-k}) = \mathbf{E}[\log(\bar{\mathbf{b}}^{(k,\lambda)t} \mathbf{X})] \quad \text{q.c.}$$

dove $\bar{\mathbf{b}}^{(k,\lambda)t} = (1 - \lambda)\beta + \lambda\bar{\mathbf{b}}^k$. Notiamo che $W^\lambda(\mathbf{X}|\mathcal{F}^{-k})$ cresce a $W(\mathbf{X}|\mathcal{F}^{-k})$ per $\lambda \uparrow 1$ visto che (come visto in

$$W(\mathbf{X}|\mathcal{F}^{-k}) + \log \lambda \leq W^\lambda(\mathbf{X}|\mathcal{F}^{-k}) \leq W(\mathbf{X}|\mathcal{F}^{-k})$$

Per ottenere il massimo tasso di raddoppio $W(\mathbf{X}|\mathbf{X}^-)$ dividiamo il capitale in numerabili manager, indicizzati da $k \geq 0$. Sia $\mu_k > 0$ la quantità iniziale affidata al k -esimo manager ($\sum_k \mu_k = 1$) e supponiamo che il suo capitale a disposizione sia investito secondo la strategia $\{\hat{\mathbf{b}}_t^{k,\lambda_j}\}$, dove $\lambda_k \uparrow 1$ e \mathcal{F}^{-k} è finito e cresce a $\sigma(\mathbf{X}^-)$ per $k \rightarrow \infty$. Il capitale totale, ottenuto sommando quelli dei vari manager, cresce esponenzialmente con tasso di rassoppio limite

$$W(\mathbf{X}|\mathbf{X}^-) = \sup_k W^{\lambda_k}(\mathbf{X}|\mathcal{F}^k)$$

4.3 Algoritmo di Lempel-Ziv

L'algoritmo di parsing incrementale di Lempel-Ziv [LZ78] è un algoritmo di compressione. Questo suddivide una successione j^n inserendo virgole tra i simboli. La prima è posta all'inizio della successione. Da qui in poi, ogni frase comincia con una virgola e 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.

Denotiamo con ν_n il numero di frasi complete in cui viene suddivisa la sequenza j^n . Per esempio 0101000100 che ha lunghezza $n = 10$ genera , 0, 1, 01, 00, 010, 0 e contiene $\nu_n = 5$ frasi complete.

Questo procedimento può essere realizzato in modo efficace costruendo dinamicamente una struttura ad albero in cui ogni nodo ha come figli possibili gli m simboli dell'alfabeto \mathcal{J} . Inizialmente, l'albero è costituito da un solo nodo, la radice. Iniziamo la ricerca per una nuova frase partendo dalla radice e scendendo nell'albero seguendo i simboli che vengono letti dalla successione. Questo procedimento finisce quando da un nodo ci si deve muovere verso un figlio mai visitato prima. Viene quindi aggiunto il nodo e il ramo corrispondente a questo, si inserisce la virgola nella sequenza e si ricomincia dalla radice. Consideriamo quindi un processo $\{J_t\}$, $t \in \mathbb{N}$ e denotiamo con T_n l'albero costruito dopo aver letto j^n . I rami uscenti da ogni nodo sono segnati con i simboli dell'alfabeto $\{1, \dots, m\}$, ma alcuni nodi possono essere incompleti nel senso che non tutti i rami uscenti sono presenti a un dato tempo e sono quindi già stati visitati. Definiamo di conseguenza \bar{T}_n come il completamento dell'albero T_n , a cui aggiungiamo tutti i rami possibili ai nodi di T_n . Ogni nodo di T_n è quindi un nodo interno di \bar{T}_n .

Lemma 4.4. T_n ha $\nu_n + 1$ nodi, \bar{T}_n ha $1 + m(\nu_n + 1)$ nodi di cui $\nu_n + 1$ interni.

Dimostrazione. Lo dimostriamo per induzione:

- $n = 1$. In questo caso abbiamo una sola frase completa costituita da un unico carattere x , $\nu_n = 1$. L'albero T_n contiene il nodo radice e il nodo corrispondente a x , mentre \bar{T}_n contiene il nodo radice con m figli, tra cui il nodo x con i suoi m figli.
- Passo induttivo. Se l'ultimo carattere letto non comporta l'aggiunta di una virgola, il numero di nodi all'interno degli alberi rimane invariato, così come $\nu_n = \nu_{n+1}$ e la tesi è verificata. Supponiamo ora l' $n + 1$ -esimo carattere comporti la formazione di una nuova parola, da cui $\nu_{n+1} = \nu_n + 1$. In questo caso viene aggiunto un nodo all'albero T_n , corrispondente a x e quindi T_{n+1} ha $|T_n| + 1 = 1 + \nu_n + 1 = \nu_{n+1} + 1$ nodi. Per quanto riguarda \bar{T}_{n+1} , esso è costituito da \bar{T}_n , con m nuovi figli aggiunti al nodo corrispondente all'ultimo carattere letto. In totale abbiamo $|\bar{T}_{n+1}| = \bar{T}_n + m = 1 + m(\nu_n + 1) + m = 1 + m(\nu_{n+1} + 1)$ nodi, di cui $\nu_{n+1} + 1$ interni.

□

Si nota che la ricerca di una nuova frase finisce quando ci troviamo in una foglia dell'albero completato \bar{T}_n . Questa foglia diventa quindi un nodo interno e i suoi m figli aggiunti a \bar{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

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

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

$c_n(j|z)$ è quindi 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 .

Inoltre, $1 + c_n(z)$ 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.

Intuitivamente e come suggerito in [Alg92] la scelta ottimale per una stima di probabilità $\hat{\mathbf{P}}(j|z)$ dato il nodo dell'albero z è

$$\hat{\mathbf{P}}(j|z) = \frac{c_n(j|z)}{c_n(z)}$$

Questa strategia però non tiene conto di frasi che non sono mai comparse nel passato e può quindi erroneamente suggerire che tali frasi compaiano con probabilità 0. Usiamo quindi una versione modificata della precedente stima, guardando le dimensioni dei sotto-alberi di \bar{T}_n , in cui ci sono anche i nodi non ancora visitati adiacenti a quelli di T_n . Definiamo quindi

$$\gamma_n(z_n) = 1 + (m - 1)(c_n(z_n) + 1)$$

$$\gamma_n(j|z_n) = 1 + (m - 1)(c_n(j|z_n) + 1)$$

e

$$\hat{\mathbf{P}}(j|j^n) = \frac{\gamma_n(j|z_n)}{\gamma_n(z_n)} = \frac{m + (m-1)c_n(j|z_n)}{m + (m-1)c_n(z_n)} \quad (4.2)$$

Come si vede in [Alg92]([citare articolo di Algoet o dimostrare, a questo punto diventa rilevante la dualità codici-gambling](#)) e in [CT06] questa stima di probabilità è universale nel senso che asintoticamente ha la stessa entropia della distribuzione effettiva e quindi anche il medesimo tasso di raddoppio ottimale per investimenti su corse di cavalli. Questa strategia verrà adottata anche per il mercato azionario con qualche modificazione in una successiva sperimentazione, in modo analogo all'adattamento del metodo delle frequenze empiriche.

4.4 Codici e Scommesse

4.4.1 I codici

In questa sezione analizziamo la dualità tra le codifiche e le scommesse, argomenti che si mostrano essere tra di loro molto collegati. Seguiremo i libri [MM08] e [CT06], dove si trovano analisi più approfondite sui codici.

Dato un processo stocastico di variabili aleatorie discrete $(X_n)_{n \geq 0}$ vogliamo codificare i dati X_n per comprimerli o trasmetterli. Supponiamo quindi di avere un alfabeto $\mathcal{L} = \{l_1, l_2, \dots, l_m\}$ di m lettere e indichiamo con \mathcal{L}^* l'insieme delle parole w ottenute concatenando in successione un numero finito di elementi dell'alfabeto, in formule $w = l_{i_1}l_{i_2} \dots l_{i_k}$. Indichiamo invece con \mathcal{L}^k l'insieme delle parole di esattamente k simboli. Denotiamo quindi con $l(w)$ la lunghezza della parola w in simboli.

Supponiamo quindi che $X_i : \Omega \rightarrow C$ dove C è l'insieme dei dati che X_i può assumere: una codifica è una funzione $c_i : C \rightarrow \mathcal{L}^*$ che associa ad ogni dato una parola. Chiamiamo quindi L l'insieme immagine di un codice c , detto anche vocabolario.

La richiesta più naturale è che il codice possa essere decodificato, questo si traduce nel fatto che la funzione c_i deve essere almeno iniettiva. In realtà dato che il codice finale, composto dalla codifica di più dati, avrà una forma del tipo

$$c_1(X_1)c_2(X_2) \dots$$

vogliamo poter suddividere il codice in arrivo in parole che vengono poi decodificate singolarmente per ottenere la stringa iniziale.

Esistono vari tipi di codici, e si suddividono inizialmente in

- *Codici univocamente decodificabili*: per ogni stringa in arrivo esiste al più un modo di suddividerlo in parole e ottenere i dati iniziali.
- *Codici prefissi*: per ogni coppia di parole $w, w' \in L$ non avviene mai che una è parte iniziale dell'altra (il suo "prefisso").

Nei testi si dimostra che scegliere un codice univocamente decodificabile rispetto a un prefisso, nonostante ogni l'insieme di questi ultimi è incluso negli altri e sembrano quindi più complessi, non porta ad alcun vantaggio computazionale o teorico, per entrambi vale infatti il seguente risultato:

Lemma 4.5 (Disuguaglianza di Kraft-Mc Millian). *Se L è un vocabolario di un codice decodificabile basato su un alfabeto \mathcal{L} di m lettere,*

$$\sum_{w \in L} \frac{1}{m^{l(w)}} \leq 1$$

Vale anche l'inverso

Teorema 4.6 (di McMillan). *Siano $\{a_1, a_2, \dots\}$ numeri interi tale che*

$$\sum_{i \geq 1} \frac{1}{m^{a_i}} \leq 1$$

allora esiste un vocabolario prefisso $L = \{w_1, w_2, \dots\}$ di parole fatte con l'alfabeto \mathcal{L} tale che $l(w_i) = a_i$.

Il nostro obiettivo è a questo punto comprimere i dati, ossia prendere le stringhe e codificarle in modo che quella in uscita sia più corta, pur mantenendo l'invertibilità del processo. Consideriamo solo il caso in cui l'insieme dei dati C è finito e di un alfabeto in arrivo con $m = 2$ lettere.

Chiamiamo un codice ottimale se il valore atteso della lunghezza delle parole in uscita $\mathbf{E}[l(w)] = \sum_{w \in L} l(w)p(w)$ è minimo possibile (in questo caso $p(w)$ è la distribuzione immagine sul vocabolario).

Consideriamo a questo punto l'insieme dei dati sorgente finito e una codifica che ad ogni parola di n dati $x_1 x_2 \dots x_n$ associa una parola lunga $l(x_1, x_2, \dots, x_n)$. Il valore atteso della lunghezza del codice sarà dato da

$$\mathbf{E}[l(X_1, X_2, \dots, X_n)] = \sum_{x^n} p(x_1, x_2, \dots, x_n) l(x_1, x_2, \dots, x_n)$$

Il primo risultato che enunciamo mostra come l'entropia del processo stocastico (o di X se sono i.i.d) sia un limite inferiore sulla lunghezza media per simbolo del codice, ma anche il valore ottimale del fattore di compressione.

Massimizzare questa quantità vuol dire scegliere nel modo migliore possibile data una distribuzione p la lunghezza del codice associato a una successione di n dati x^n .

Teorema 4.7. *Il minimo valore atteso della lunghezza per simbolo di una codifica a n simboli Len_n^* , con un alfabeto di m lettere è tale che*

$$\frac{H(X_1, X_2, \dots, X_n)}{n} \leq Len_n^* < \frac{H(X_1, X_2, \dots, X_n)}{n} + \frac{1}{n}$$

Inoltre, se $(X_n)_{n \geq 0} = \mathcal{X}$ è un processo stazionario

$$Len_n^* \rightarrow H(\mathcal{X})$$

dove $H(\mathcal{X})$ è il tasso di entropia del processo.

Nota. Nel caso il processo sia costituito da variabili i.i.d. e per $n = 1$ si ha

$$H(X) \leq \mathbf{E}[l] < H(X) + 1$$

Questo risultato suggerisce il fatto che la lunghezza ottimale di una stringa codificata $l^*(x_1, x_2, \dots, x_n)$ sia uguale a $\log \frac{1}{p(x_1, x_2, \dots, x_n)}$ (eventualmente arrotondato a un intero). Viceversa, trovare un codice ottimale ci suggerisce quale sia la distribuzione del processo: proprio per questo le due strategie mostrate in precedenza provengono da algoritmi di compressione ottimali.

Nel caso in cui la distribuzione è sconosciuta, un algoritmo adattivo di codifica che suddivide la stringa sorgente in parole di lunghezza variabile e le codifica, in modo che la lunghezza media delle parole del codice in uscita sia ottimale, è detto codifica universale. Ne è un esempio l'algoritmo di Lempel-Ziv [LZ78] sopra citato.

Questi sono quindi gli algoritmi di fondamentale importanza per la stima di una distribuzione di probabilità e il raggiungimento del tasso di raddoppio ottimale

4.4.2 Dualità Codici-Scommesse

Nel capitolo 1 abbiamo mostrato come il tasso di raddoppio ottimale e l'entropia di un processo stocastico e ergodico che modella una successione di corse a m cavalli siano legati dalla relazione

$$H(\mathcal{X}) + W_\infty^* = \log(m)$$

Intuitivamente, più l'entropia è bassa e il processo prevedibile, maggiore è il tasso di raddoppio ottimale e quindi il guadagno atteso e viceversa.

Il seguente algoritmo di compressione, descritto in [CT06], mostra come da una strategia di investimento si possa ottenere direttamente un algoritmo di compressione e viceversa. Supponiamo di avere due gemelli identici, uno dei quali codificherà la successione degli esiti delle corse e l'altro la decodificherà applicando poi la stessa strategia di scommesse. Mappiamo una successione $\mathbf{x} = x_1 x_2 \dots x_n \in \{0, 1\}^n$ (gli esiti delle gare, supponendo $m = 2$) in una sequenza codificata $c_1 c_2 \dots c_k \in \{0, 1\}^k$. Entrambi i gemelli conoscono n . Ordiniamo quindi in modo lessicografico le 2^n sequenze possibili. Il codificatore osserva la successione \mathbf{x} e calcola quindi quale sarebbe stato il capitale risultante $S_n(\mathbf{x}')$ per ogni successione $\mathbf{x}' \leq \mathbf{x}$ e $F(\mathbf{x}) = \sum_{\mathbf{x}' \leq \mathbf{x}} 2^{-n} S_n(\mathbf{x}')$. Chiaramente $F(\mathbf{x}) \in [0, 1]$ ed è una funzione di ripartizione per una distribuzione particolare. Sia quindi $k = \lceil n - \log S_n(\mathbf{x}) \rceil$ e esprimiamo $F(\mathbf{x})$ fino alla k -esima cifra decimale $\lfloor F(\mathbf{x}) \rfloor = 0.c_1 c_2 \dots c_k$. Il gemello codificatore trasmette quindi questa sequenza $c_1 c_2 \dots c_k$.

Il decodificatore calcola esattamente il valore $S_n(\mathbf{x}')$ per ognuna delle 2^n successioni possibili ed è a conoscenza quindi delle somme cumulate di $2^{-n} S_n(\mathbf{x}')$. Trova quindi il primo \mathbf{x} per cui viene superata la quantità $0.c_1 \dots c_k$, esso è tale che la somma cumulata è nell'intervallo $[0.c_1 \dots c_k, 0.c_1 \dots c_k + \frac{1}{2^k}]$ e che il codice sia univocamente decodificabile, dato che sommare $S_n(\mathbf{x}) 2^{-n}$ porterebbe fuori dall'intervallo. Il secondo gemello decodifica quindi con successo \mathbf{x} . Il numero di bit richiesti è $k = \lceil n - \log S_n(\mathbf{x}) \rceil$ e quello di bit risparmiati $n - k = \lfloor \log S_n(\mathbf{x}) \rfloor$. Per scommesse proporzionali, si è dimostrato che $S_n(\mathbf{x}) = 2^n p(\mathbf{x})$, dato che il pagamento è 2. Il numero atteso di bit è quindi

$$\mathbf{E}[k] = \sum p(\mathbf{x}) \lceil -\log(p\mathbf{x}) \rceil \leq H(X_1, \dots, X_n) + 1$$

Si ha quindi che se la strategia di scommessa è deterministica e nota a entrambi i gemelli, il numero di bits necessari a codificare $x_1x_2\dots x_n$ è meno di $n - \log S_n + 1$. Inoltre, se la distribuzione è nota e viene usata la strategia del Kelly gambling, la lunghezza attesa di descrizione è $\mathbf{E}[n - \log S_n] \leq H(X_1, X_2, \dots, X_n) + 1$. Quindi, il capitale risultante è strettamente collegato al fattore di compressione raggiunto. Migliore la stima della distribuzione di probabilità, maggiore è il capitale e migliore è la compressione dei dati.

Capitolo 5

Sperimentazioni

In quest'ultimo capitolo verranno mostrati risultati di alcune sperimentazioni da me eseguite di alcuni dei metodi citati, appositamente adattati, su dati reali di mercato.

In particolare il mercato sarà sempre modellizzato come un processo stocastico stazionario ed ergodico (X_n) , scegliendo in vari modi la frequenza di investimento (giornaliera, settimanale, mensile, ...). Tranne che per il metodo Nearest Neighbour [GUW08], dove consideriamo 3 asset, le altre sperimentazioni saranno effettuate su $m = 2$ titoli.

5.1 Metodi di scelta dei portafogli

Come prima cosa vediamo come scegliere il portafoglio log-ottimale data una distribuzione di probabilità \mathbf{P} , la quale può essere sia stimata che reali. Data la complessità del problema utilizzeremo due metodi, rispettivamente ottenuti da un'approssimazione al primo ordine e al secondo ordine.

5.1.1 Primo ordine

Il metodo più semplice corrisponde ad una approssimazione di $\mathbf{E}[\log \mathbf{b}^t \mathbf{X}]$ al primo ordine. Siccome nella maggior parte dei casi $X_i \approx 1$ per ogni i (ossia il rendimento è vicino a 1), possiamo scrivere $\mathbf{X} \approx \mathbf{e} + \mathbf{r}$ dove $r_i \approx 0$ e $\mathbf{e} = (1, \dots, 1)^t$ per ogni i .

Si ha quindi che

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

Evidentemente, questo metodo ci suggerisce di investire tutto il capitale nel titolo che assicura valore atteso maggiore, senza tenere conto di correlazioni o incertezze su \mathbf{X} , contribuendo a un rischio maggiore di perdita.

5.1.2 Secondo Ordine

Il metodo al secondo ordine, che tiene conto di correlazioni e incertezze sugli asset, è detto anche criterio di Kelly ed è ampiamente descritto in [Tho06].

In generale il modello prevede la presenza di un asset detto risk-free, ossia con rendimento r generalmente basso ma quasi certo. Questo titolo può essere però tolto dal modello,

semplicemente ponendo $r = 0$. Consideriamo quindi un vettore dei rendimenti per m asset, con vettore dei drift attesi $\mathbf{E}[\mathbf{X} - \mathbf{e}] = \mathbf{M}$ e matrice delle covarianze $C = (c_{i,j})$ con $c_{i,j}$ covarianza tra i titoli i e j . Sia infine $\mathbf{R} = (r, r, \dots, r)^t$.

Il criterio di Kelly ci suggerisce di scegliere il portafoglio dato da

$$\mathbf{b} = C^{-1}(\mathbf{M} - \mathbf{R})$$

dove si è supposta l'invertibilità della matrice C . Nel caso in cui non vi sia un titolo risk-free si pone quindi $\mathbf{b} = C^{-1}\mathbf{M}$.

Il criterio di Kelly può in molti casi, restituire un portafoglio non normalizzato a 1, oppure con pesi negativi. Questi rappresentano situazioni in cui il modello ci suggerisce di prendere prestiti e acquistare o vendere più del possibile.

In generale eviteremo queste situazioni e ci comporteremo in questo modo:

- Se il modello ci propone pesi negativi su uno o più asset non investiremo alcuna parte del capitale in questi.
- Se la somma dei pesi del portafoglio suggerito è maggiore di 1, normalizzeremo a questa quantità mantenendo le proporzioni.
- Se la somma dei pesi è minore di 1 investiremo solo una parte del capitale, corrispondente a tale somma e terremo il resto senza investirlo.

5.2 Metodi e risultati sperimentali

In questa sezione vengono descritti i metodi utilizzati per le sperimentazioni e analizzati i risultati ottenuti. I codici utilizzano dati di asset con una storia ampia: nel caso di 2 titoli sono stati scelti Coca Cola (KO) e IBM, mentre nel caso di 3 titoli a questi viene affiancato General Electric (GE). Per ogni strategia riporteremo risultati più o meno approfonditi per i tre periodi di investimento tipici: giornaliero, settimanale (≈ 5 giorni) e mensile (≈ 21 giorni). Questi ultimi due sono dati più affidabili, in quanto in nessun caso terremo conto dei costi di transazione e ribilanciamento del capitale che nel caso giornaliero possono incidere drasticamente sul risultato finale.

5.2.1 Quantità di analisi

Per la parte di analisi dei risultati considereremo le seguenti quantità, basate su n periodi di investimento, w_t il capitale al t -esimo periodo (con $w_0 = 1$):

- *Crescita media annua* μ_{ann} : è la crescita percentuale media per ogni anno calcolata come

$$\mu_{\text{ann}} = k \frac{1}{n} \sum_{i=1}^n \left(\frac{w_t}{w_{t-1}} - 1 \right)$$

dove k è il numero di periodi di investimento in un anno che dipendono dalla frequenza di investimento

- *Deviazione standard annua* σ_{ann} : è la deviazione standard dei drift di ogni periodo $r_t = \left(\frac{w_t}{w_{t-1}} - 1\right)$, moltiplicata per un fattore \sqrt{k} .

- *Crescita media geometrica* μ_{geom} : indica il rendimento medio geometrico ed è calcolato come

$$\mu_{\text{geom}} = \left(\frac{w_n}{w_0}\right)^{\frac{1}{n}} - 1$$

- *drowdown* $dd(t)$: è il rapporto percentuale tra la perdita totale fino al periodo t rispetto al capitale massimo raggiunto precedentemente e si calcola come

$$dd(t) = 1 - \frac{w_t}{\max_{0 \leq s \leq t} w_s}$$

- *Massimo drowdown* $\max dd$: è il massimo drowdown tra tutti i periodi di investimento e indica quanto il metodo ha perso rispetto a un suo massimo locale.
- *Calmar ratio* CR : indica la performance di un metodo di investimento rispetto alla massima perdita registrata, o massimo drowdown. In generale, considerando anche un titolo risk-free con rendimento r_f che poniamo a zero per il nostro caso, il calmar ratio si calcola come

$$CR = \frac{\mu_{\text{ann}} - r_f}{\max dd}$$

- *sharpe ratio* SR : indica la performance di un portafoglio tenendo conto dell'incertezza dei risultati e viene calcolato come

$$SR = \frac{\mu_{\text{ann}} - r_f}{\sigma_{\text{ann}}}$$

5.2.2 Metodo empirico sui rendimenti

Il primo metodo che presentiamo si basa sull'analisi del capitolo 4.2, applicato a 2 asset. In particolare, per questione dell'ampiezza temporale dei dati disponibili (al massimo poco meno di 15000 giorni di investimento), si è scelto di fermarsi a $k = 1$. Le stime di probabilità si fermano quindi alle transizioni tra singoli stati, come per una catena di markov.

Partizioniamo la retta reale estesa $[-\infty, +\infty]$ in un numero finito di intervalli $\{A_1, \dots, A_k\}$, ognuno dei quali sarà il simbolo una parte dei possibili rendimenti percentuali di un asset. Gli atomi o stati che andiamo a considerare sono tutti i possibili prodotti $A_i \times A_j$, corrispondenti rispettivamente alla fascia di rendimento del primo e secondo asset. Per esempio uno di questi potrà essere l'insieme dei vettori di rendimento in cui il primo titolo da un periodo all'altro ha tra l'1% e il 2% di rendimento, mentre il secondo titolo ha un rendimento maggiore del 10%. Le stime di probabilità di transizione saranno calcolate quindi come nell'equazione 4.1 con una piccola modifica

$$\hat{\mathbf{P}}_t(j|J^{-k}) = \frac{1 + p \cdot c_t(j|J^{-k})}{m + p \cdot c_t(J^{-k})} \quad (5.1)$$

dove p è il peso che scegliamo di dare alle transizioni osservate rispetto a quelle assegnate a priori, J^{-k} è lo stato attuale (rappresentato da una sequenza di k simboli, nel nostro caso $k = 1$, riferito al k -passato) e j è lo stato di cui andiamo a studiare la probabilità di transizione dallo stato attuale. Scegliere $p > 1$ nella maggior parte dei casi porta a risultati migliori, in quanto la quantità degli stati e il limitato numero di dati storici disponibili possono portare al fatto che i pesi dati a priori per non escludere transizioni non ancora osservate possano influire troppo.

Ad ogni periodo di investimento t , quindi, si possono presentare due casi:

- il vettore dei rendimenti \mathbf{X}_{t-1} appartiene a uno stato già osservato in precedenza, possiamo quindi basarci sulla stima della probabilità di transizione data dalla formula 5.1 e calcolare il valore atteso del vettore \mathbf{X}_t .
- il vettore dei rendimenti \mathbf{X}_{t-1} appartiene a uno stato che non è mai stato osservato prima. In questo caso usiamo la stessa formula ma avremo che la stima di probabilità è data solo dai pesi a priori assegnati a tutte le transizioni possibili da quello stato.

A questo punto per ogni insieme A_j sopra citato viene scelto un valore di riferimento, che può essere la media degli estremi se è un intervallo limitato o l'estremo inferiore/superiore nel caso sia illimitato superiormente/inferiormente. Si calcola così la media pesata e si ottiene un rendimento stimato per ogni asset.

In questo caso ci affidiamo solo alla approssimazione al primo ordine: sceglieremo quindi un portafoglio che pone tutto il capitale sull'asset che assicura valore atteso maggiore. Nel caso essi siano uguali, poniamo $\frac{1}{2}$ del capitale su entrambi i titoli. A causa dell'incertezza molto grande, il metodo di Kelly dà come risultato portafogli a pesi negativi, nel qual caso abbiamo scelto di non investire nulla.

[Il codice occupa davvero tanto spazio.](#)

Qui di seguito riportiamo le quattro partizioni dell'intervallo reale che sono state considerate per costruire gli stati $A_{i_1} \times A_{i_2}$, descritte dagli estremi di tali intervalli in successione, da intendersi in punti per mille di rendimento.

1. $\{-\infty, -100, -50, -20, -10, -5, -2, -1, 0, 1, 2, 5, 10, 20, 50, 100, +\infty\}$
2. $\{-\infty, -100, -50, -10, -5, 0, 5, 10, 50, 100, +\infty\}$.
3. $\{-\infty, -150, -100, -75, -50, -30, -10, -5, 0, 5, 10, 30, 50, 75, 100, 150, +\infty\}$
4. $\{-\infty, -1000, -700, -500, -300, -100, -50, -10, 0, 10, 50, 100, 300, 500, 700, 1000, +\infty\}$

In tutti i casi il numero totale di periodi di investimento nel caso giornaliero è $n = 14640$, nel caso settimanale $n = 2927$ e nel caso mensile $n = 696$. Il capitale iniziale $S_0 = 1$. Per ogni periodo di investimento riportiamo i risultati generali, focalizzandoci su quelli ottenuti da partizione e peso p ottimale.

Risultati

- Periodo di investimento giornaliero: I risultati migliori si ottengono per $p = 5$. La partizione ottimale è la numero 4, probabilmente perché gli stati da essa derivanti corrispondono a incrementi simili a quelli del mercato reale, ossia dell'ordine di qualche punto percentuale.
- Periodo di investimento settimanale: in questo caso i risultati sono molto più simili tra loro. Il peso migliore si ha per $p = 10$, essendoci meno periodi di investimento su cui costruire le stime di distribuzione. La partizione migliore è la numero 3, tuttavia per confronto utilizziamo i risultati per la numero 4, i cui valori si discostano comunque poco dalla precedente.
- Periodo di investimento mensile: la partizione migliore si rivela ancora essere la 4. I risultati ottimali si ottengono aumentando il peso, ossia ponendo $p \approx 20$, confermando la tendenza.

Riportiamo qui di seguito i risultati e il grafico in scala semilogaritmica (per apprezzare meglio l'andamento esponenziale) per i tre periodi di investimento.

| Quantità | Giornaliero | Settimanale | Mensile |
|-----------------------|-------------|-------------|-----------|
| S_n | 43025.29 | 7584.782 | 18071.66 |
| μ_{ann} | 21.70% | 18.56% | 19.89% |
| σ_{ann} | 0.2460 | 0.2394 | 0.2243 |
| μ_{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 |

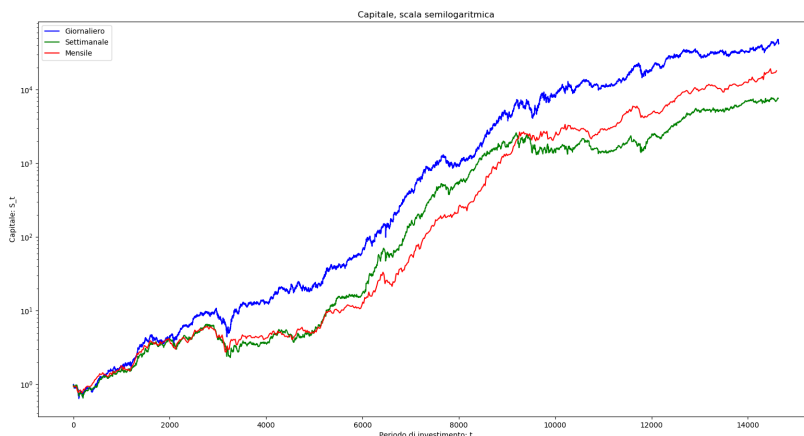


Figura 5.1

In generale, i risultati al variare di tutti i parametri sono comunque allineati a questi, per tutte le quantità considerate. Essendo il drowdown molto elevato, questi metodi sono in realtà poco efficaci: l'investitore in certi periodi ha perso più di metà del capitale massimo che aveva accumulato.

5.2.3 Metodo empirico sui rapporti rendimento-incertezza

Questo secondo metodo è simile al precedente, ma non si basa sulle stime di probabilità costruite a partire dalle osservazioni. Per ogni periodo di investimento t , guardiamo gli ultimi τ vettori di log-rendimento $\log(\mathbf{X}_{t-\tau}), \dots, \log(\mathbf{X}_{t-1})$ e calcoliamo media e deviazione standard empirica, rispettivamente $\hat{\boldsymbol{\mu}}$ e $\hat{\boldsymbol{\sigma}}$.

Partizioniamo come prima la retta reale estesa $[-\infty, +\infty]$ in un numero finito di intervalli $\{A_1, \dots, A_k\}$. Per ogni asset $j = 1, 2$ andiamo a calcolare $\frac{\log X_{t-1,j} - \hat{\mu}_j}{\hat{\sigma}_j}$, e troviamo a quale intervallo A_{i_j} appartiene. Lo stato attuale sarà simboleggiato dal prodotto cartesiano dei due intervalli $A_{i_1} \times A_{i_2}$.

In pratica calcoliamo per ogni asset quanto l'ultimo rendimento si discosta dalla media degli ultimi τ giorni, in rapporto alla deviazione standard, e vediamo in quale fascia di valori si trova questa quantità (ricordando che, generalmente, tale valore supera raramente 5 in modulo).

Per ognuno di questi stati $A_{i_1} \times A_{i_2}$ consideriamo le occorrenze di essi $B_{i_1, i_2, t} = \{\tau \leq s \leq t : \text{al tempo } s \text{ ero nello stato } A_{i_1} \times A_{i_2}\}$. Siano quindi $R_{i_1, i_2, t} = \{\mathbf{X}_s : s \in B_{i_1, i_2, t}\}$ i rendimenti del periodo corrispondente a ogni occorrenza dello stato nel passato. Da questi possiamo ottenere una stima del valore atteso e dell'incertezza di \mathbf{X}_t e utilizzando entrambe le approssimazioni al primo e secondo ordine, semplicemente facendo medie e covarianze dei vettori in $R_{i_1, i_2, t}$.

Questo metodo viene utilizzato effettivamente solo dal τ -esimo periodo di investimento in poi, prima utilizziamo un portafoglio che suddivide il capitale in parti uguali tra i titoli.

Le partizioni della retta reale utilizzate per questa sperimentazione, indicate dagli estremi come nel capitolo precedente, sono:

1. $\{-\infty, -3, -1, -0.5, 0, 0.5, 1, 3, +\infty\}$
2. $\{-\infty, -2, -1, -0.5, -0.25, 0, 0.25, 0.5, 1, 2, +\infty\}$.

Daremo una analisi generale dei risultati generali in funzione della variabile τ e della partizione, con le stesse condizioni iniziali del capitolo precedente.

Risultati

I risultati mostrano chiaramente che in tutti i casi l'approssimazione al primo ordine ottiene risultati migliori del metodo di Kelly, mantenendo però un drowdown molto alto. Questo metodo è migliore del precedente, probabilmente perché utilizza dati storici per calcolare una stima del valore atteso del rendimento futuro. Le partizioni, essendo simili, forniscono valori finali molto ravvicinati, favorendo leggermente la seconda.

- Periodo di investimento giornaliero: il valore di τ ottimale è in tutti i casi 63, corrispondente a un periodo di circa 3 mesi.
- Periodo di investimento settimanale: il valore di τ ottimale è in tutti i casi 52 settimane, corrispondente a un periodo di 1 anno.
- Periodo di investimento mensile: il valore di τ ottimale tra quelli testati ($\tau = 36, 60$) è $\tau = 36$, corrispondente a un periodo di 3 anni.

Riportiamo qui di seguito la tabella e il grafico dei risultati migliori, riguardanti quindi l'approssimazione al primo ordine.

| Quantità | Giornaliero | Settimanale | Mensile |
|-----------------------|-------------|-------------|-----------|
| S_n | 64102.81 | 4166.649 | 2384.919 |
| μ_{ann} | 22.37% | 17.44% | 16.59% |
| σ_{ann} | 0.2453 | 0.2371 | 0.2357 |
| μ_{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 |



Figura 5.2

5.2.4 Metodo ad albero (Lempel-Ziv)

Questo metodo prende spunto direttamente da quello descritto nella sezione 4.3 e segue passo per passo l'algoritmo di codifica lì descritto.

Definiamo, esattamente come nel metodo empirico sui rendimenti (paragrafo 5.2.2) gli

stati simboleggiati dai prodotti cartesiani degli intervalli reali di rendimenti $A_{i_1} \times A_{i_2}$. A questo punto procede esattamente come nell'algoritmo di Lempel-Ziv descritto in precedenza, come se la stringa da codificare in tempo reale fosse la successione delle fasce di rendimento (gli $A_{i_1} \times A_{i_2}$).

La formula per ottenere la stima di probabilità è una piccola modificazione della formula 4.2 a cui abbiamo dato un peso differente p alle osservazioni rispetto alle frequenze date a priori.

$$\hat{\mathbf{P}}(j|j^n) = \frac{\gamma_n(j|z_n)}{\gamma_n(z_n)} = \frac{m + p \cdot (m - 1)c_n(j|z_n)}{m + p \cdot (m - 1)c_n(z_n)}$$

Il calcolo dei valori attesi procede come descritto nel paragrafo 5.2.2, ossia assegnando un valore particolare agli asset per ogni stato e facendo una media pesata con le probabilità di transizione appena ottenute.

Anche in questo caso si hanno incertezze molto elevate che rendono il metodo di Kelly inefficace: ci baseremo quindi solo su una approssimazione al primo ordine.

Le partizioni della retta reale utilizzate per costruire gli stati sono le stesse del metodo empirico sui rendimenti.

Risultati

La strategia descritta per funzionare in modo ottimale richiede molti dati storici e pochi stati, data la complessa struttura ad albero che viene costruita passo dopo passo.

- Periodo di investimento giornaliero: la partizione migliore è la 4 e maggiore il peso p , migliori sono i risultati (almeno per valori contenuti). In questo caso è stato utilizzato $p = 10$.
- Periodo di investimento settimanale: alcune partizioni (2 e 3), mostrano come pochi dati disponibili incidano drasticamente sui risultati finali. La partizione ottimale si conferma essere la 4, combinata con $p = 10$.
- Periodo di investimento mensile: in questo caso, più la partizione è semplice più i risultati migliorano, tanto che la 3 e la 4 hanno una performance nettamente inferiore rispetto alla 1 e soprattutto alla 2. Quest'ultima, con $p = 10$ si è rivelata essere la migliore.

Riportiamo quindi la tabella e il grafico dei risultati.

| Quantità | Giornaliero | Settimanale | Mensile |
|-----------------------|-------------|-------------|-----------|
| S_n | 17806.67 | 13149.88 | 8605.029 |
| μ_{ann} | 20.12% | 19.32% | 18.71% |
| σ_{ann} | 0.2446 | 0.2303 | 0.2292 |
| μ_{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 |

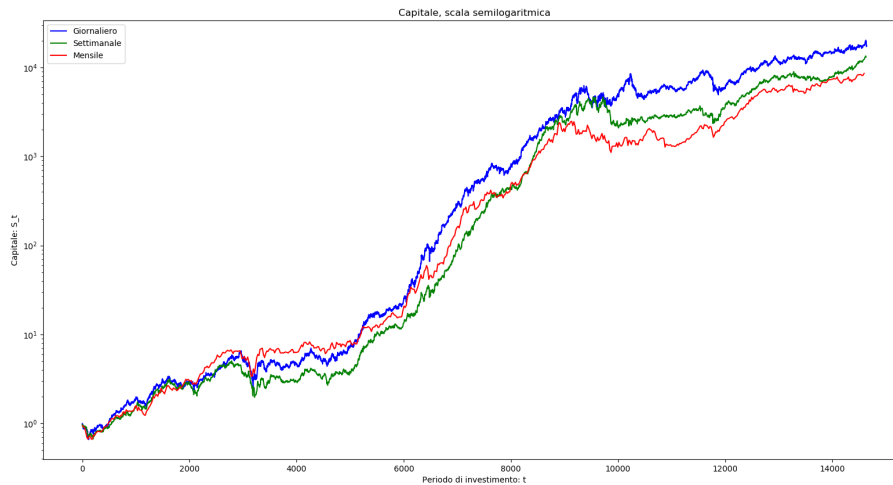


Figura 5.3

5.2.5 Metodo del portafoglio universale di Cover

La seguente sperimentazione è un'applicazione diretta del portafoglio universale a orizzonte infinito di Cover e Thomas descritto in [CT06] e nel paragrafo 3.2. L'algoritmo viene applicato sia su $m = 2$ asset e gli integrali nelle equazioni 3.2 e 3.3 vengono calcolati grazie a una libreria di python.

Dalla definizione (il prodotto nell'equazione 3.3 è telescopico) risulta evidente che i tre periodi di investimento hanno lo stesso andamento, tanto che nei giorni in cui i tre portafogli vengono aggiornati simultaneamente, essi ottengono lo stesso risultato.

| Quantità | Giornaliero | Settimanale | Mensile |
|-----------------------|-------------|-------------|-----------|
| S_n | 4764.502 | 5324.182 | 4777.275 |
| μ_{ann} | 16.93% | 17.04% | 16.84% |
| σ_{ann} | 0.2058 | 0.2003 | 0.1942 |
| μ_{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 |

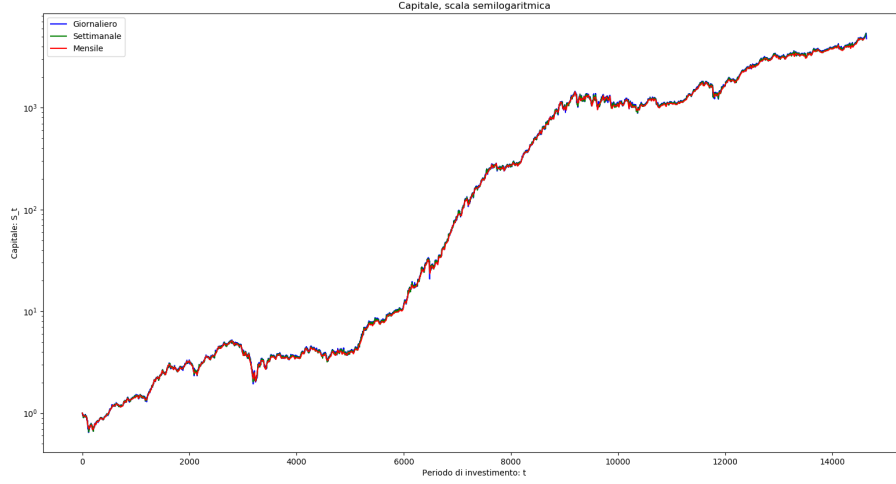


Figura 5.4

5.2.6 Metodo Nearest-Neighbour

Il seguente metodo è descritto in [GUW08] sia nella forma generale che in quella approssimata utilizzata in questa sperimentazione, ed è un esempio di metodo a finestra scorrevole.

Scegliamo inizialmente due parametri: k , la dimensione della finestra che teniamo in considerazione e s_l ossia la lunghezza di ogni segmento nel passato in cui cerchiamo il vicino a distanza minore dalla finestra attuale.

Al periodo di investimento $n > s_l + 1$, per ogni $1 \leq i \leq \lfloor \frac{n-1}{s_l} \rfloor$, definiamo quindi

$$N_i = \arg \min_{(i-1)s_l + k \leq j < i s_l} \|\mathbf{X}_{j-k}^{j-1} - \mathbf{X}_{n-k}^{n-1}\|$$

dove $\|\cdot\|$ è una qualsiasi norma sullo spazio $\mathbb{R}^{m \times k}$ (m è il numero di asset considerati), e \mathbf{X}_{j-k}^{j-1} indica il vettore $(\mathbf{X}_{j-k}, \dots, \mathbf{X}_{j-1})$. A questo punto scegliamo il portafoglio definito da

$$\hat{\mathbf{b}} = \arg \max_{\mathbf{b} \in \mathcal{B}} \prod_{1 \leq i \leq \lfloor \frac{n-1}{s_l} \rfloor} \mathbf{b}^t \mathbf{X}_{N_i}$$

Nell'articolo citato l'algoritmo combina insieme capitali ottenuti utilizzando diversi valori k , s_l , calcolando una media pesata di essi. Nel nostro caso, utilizzeremo diversi valori dei parametri che però non combineremo ma confronteremo tra loro.

Data l'elevata complessità faremo una ulteriore approssimazione: la finestra scorrevole viene mossa solo nei 10 anni precedenti al periodo di investimento, ossia circa 2560 giorni. Il portafoglio viene poi scelto in tre modi diversi, ossia con quello descritto precedentemente, sia usando una approssimazione al primo ordine e al secondo ordine sfruttando i dati storici degli \mathbf{X}_{N_i} .

I principali parametri di cui tenere conto sono quindi s_l e k . Riportiamo qui sotto le analisi per periodo di investimento.

Risultati

Si osserva subito che questa strategia, pur mantenendo un drowdown molto alto, ha una performance decisamente migliore rispetto alle precedenti e l'approssimazione al primo ordine ha risultati peggiori del kelly gambling e il Nearest Neighbour standard. Quest'ultimo fornisce in tutti i casi i risultati migliori, nonostante il Kelly Gambling si mantenga sempre sullo stesso ordine di grandezza.

- Periodo di investimento giornaliero: Sono stati sperimentati i parametri corrispondenti a segmenti di lunghezza $s_l = 21, 42, 63$, ossia 1, 2, 3 mesi reali, con finestre lunghe $k = 1, 5, 10$ giorni. Dai risultati ottenuti è apparso chiaramente che il valore di s_l ottimale per qualsiasi finestra è $s_l = 21$, inoltre segmenti più lunghi danno risultati peggiori. In generale, la lunghezza ottimale di una finestra è intorno a $k = 3$ o $k = 5$.
- Periodo di investimento settimanale: i parametri sperimentati in questo caso corrispondono a $s_l = 12, 24$ in combinazione con $k = 2, 4, 6$. Mentre $s_l = 12$ si è rivelato essere meno efficiente, a causa dei troppi segmenti e poco spazio dove muovere la finestra, per la lunghezza di quest'ultima i risultati migliori si hanno con $k = 4$, all'incirca lo stesso valore del caso giornaliero.
- Periodo di investimento mensile: in quest'ultimo caso sono state provate le due combinazioni di parametri $k = 1, s_l = 6$ e $k = 2, s_l = 6$. La migliore è risultata essere la prima.

Riportiamo dati e grafici per le combinazioni migliori, tutte ottenute utilizzando il metodo standard.

| Quantità | Giornaliero | Settimanale | Mensile |
|-----------------------|-------------|-------------|-----------|
| \bar{S}_n | 1485001 | 77065.69 | 25047.91 |
| μ_{ann} | 28.13% | 22.97% | 20.65% |
| σ_{ann} | 0.2564 | 0.2541 | 0.2327 |
| μ_{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 |

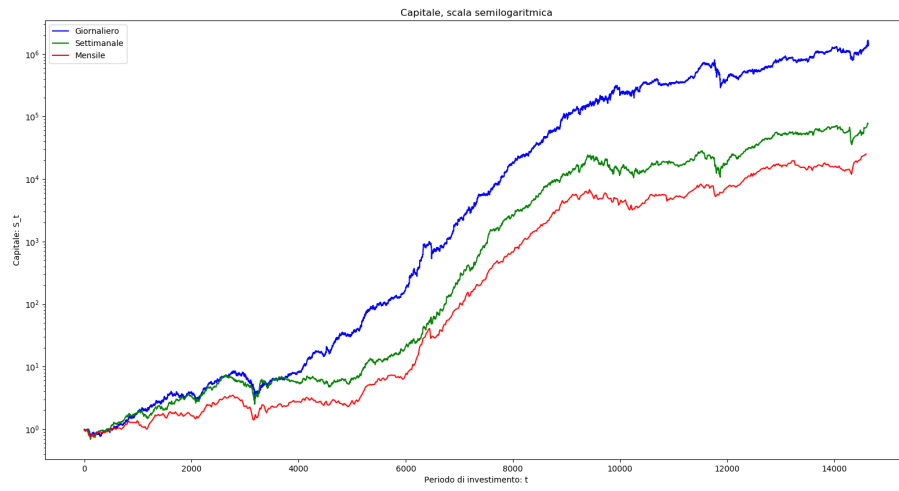


Figura 5.5

Appendice A

Definizioni

Diamo qui una serie di definizioni e risultati che vengono utilizzati nei vari capitoli e stanno alla base di molti argomenti trattati.

A.1 Speranza condizionale e distribuzioni condizionate

Qui di seguito definiamo invece alcuni concetti di teoria della probabilità. In ogni caso lavoreremo su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbf{P})$ e seguiremo il testo [Kle13]:

Definizione A.1. Sia E uno spazio metrico dotato della sigma algebra di Borel \mathcal{E} . Una successione limitata di misure positive \mathbf{P}_n , $n \in \mathbb{N}$ su (E, \mathcal{E}) converge debolmente alla misura finita e positiva \mathbf{P} se $\forall f$ funzione limitata e continua si ha

$$\mathbf{E}_{\mathbf{P}_n}[f] \rightarrow \mathbf{E}_{\mathbf{P}}[f]$$

Sia quindi $\mathcal{G} \subset \mathcal{F}$ e $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Per quanto riguarda la speranza condizionale e le distribuzioni condizionali e regolari usiamo le seguenti definizioni.

Definizione A.2. Una variabile aleatoria Y è detta speranza condizionale di X data \mathcal{G} , in formule $\mathbf{E}[X|\mathcal{G}] = Y$ se

- Y è \mathcal{G} -misurabile
- Per ogni $A \in \mathcal{G}$ vale $\mathbf{E}[X\mathbb{1}_A] = \mathbf{E}[Y\mathbb{1}_A]$. Inoltre, per $B \in \mathcal{F}$, $\mathbf{P}(B|\mathcal{G}) = \mathbf{E}[\mathbb{1}_B|\mathcal{G}]$ è detta la probabilità condizionale di B rispetto alla σ -algebra \mathcal{G} .

La speranza condizionale è unica a meno di equivalenza quasi ovunque.

Nel caso in cui stiamo condizionando rispetto a una σ algebra generata $\sigma(Y)$ scriveremo anche $\mathbf{E}[X|\sigma(Y)] = \mathbf{E}[X|Y]$.

Definizione A.3. Sia $Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ e $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$. Definiamo la speranza condizionale di Y dato $X = x$ come $\mathbf{E}[Y|X = x] = \phi(x)$ dove $\phi(X) = Z = \mathbf{E}[Y|X]$ è la funzione che fattorizza $\mathbf{E}[Y|X]$ con X (rif. [Kle13]).

Definizione A.4. Siano dati due spazi di misura $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$. Una mappa $\kappa : \Omega_1 \times \mathcal{F}_2 \rightarrow [0, \infty]$ è detta nucleo di transizione (σ -)finito se

- $\omega_1 \mapsto \kappa(\omega_1, A_2)$ è misurabile per ogni $A_2 \in \mathcal{F}_2$.
- $A_2 \mapsto \kappa(\omega_1, A_2)$ è una misura (σ -)finita su $(\Omega_2, \mathcal{F}_2)$ per ogni $\omega_1 \in \Omega_1$.

Se la misura nel secondo punto è di probabilità allora la funzione è anche detta nucleo stocastico o di Markov.

Definizione A.5. Sia Y una variabile aleatoria a valori in (E, \mathcal{E}) e sia $\mathcal{G} \subset \mathcal{F}$. Un nucleo stocastico $\kappa_{Y, \mathcal{G}}$ da (Ω, \mathcal{G}) a (E, \mathcal{E}) è detto una distribuzione condizionale regolare di Y data \mathcal{G} se

$$\kappa_{Y, \mathcal{G}}(\omega, B) = \mathbf{P}[\{Y \in B\} | \mathcal{G}](\omega)$$

per ogni $B \in \mathcal{E}$ e per quasi ogni ω (rispetto a \mathbf{P}) o equivalentemente se

$$\int \mathbb{1}_B(Y) \mathbb{1}_A d\mathbf{P} = \int \kappa_{Y, \mathcal{G}}(\cdot, B) \mathbb{1}_A d\mathbf{P} \quad \text{per ogni } A \in \mathcal{G}, \text{ per ogni } B \in \mathcal{E}$$

Un caso speciale si ha quando $\mathcal{G} = \sigma(X)$ in cui si ha che

$$(x, A) \mapsto \kappa_{Y, X}(x, A) = \mathbf{P}(\{Y \in A\} | X = x) = \kappa_{Y, \sigma(X)}(X^{-1}(x), A)$$

Nel caso in cui lo spazio in arrivo sia \mathbb{R} con σ -algebra dei boreliani, esiste sempre una versione della distribuzione condizionale regolare, qualsiasi sia la σ -algebra di condizionamento (Teorema 8.29 in [Kle13]).

Nota. Siano X e Y variabili aleatorie reali con densità congiunta $f(x, y)$ rispetto alla misura di Lebesgue λ . Per $x \in \mathbb{R}$, consideriamo

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \lambda(dy)$$

Questa la densità marginale di X , ossia la distribuzione \mathbf{P}_X su \mathbb{R} . La distribuzione condizionata regolare di Y rispetto a X ha densità

$$\frac{\mathbf{P}(Y \in dy | X = x)}{dy} = f_{Y|X}(x, y) = \frac{f(x, y)}{f_X(x)}$$

Per quasi ogni x rispetto a \mathbf{P}_X . Questo si verifica facilmente notando che dati $A, B \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} \int_A \mathbf{P}(X \in dx) \int_B f_{Y|X}(x, y) \lambda(dy) &= \int_A \mathbf{P}(X \in dx) \frac{1}{f_X(x)} \int_B f(x, y) \lambda(dy) \\ &= \int_A \lambda(dx) \int_B f(x, y) \lambda(dy) \\ &= \int_{A \times B} f d\lambda^2 = \mathbf{P}(X \in A, Y \in B) \end{aligned}$$

A.2 Martingale

Introduciamo ora alcuni concetti riguardo a filtrazioni e martingale. In generale considereremo processi stocastici $X = \{X_t, t \in I\}$.

Definizione A.6. Sia $\mathbb{F} = \{\mathcal{F}_t, t \in I\}$ una famiglia di σ -algebre con $\mathcal{F}_t \subseteq \mathcal{F}$ per ogni $t \in I$. \mathbb{F} è detta filtrazione se $\mathcal{F}_s \subseteq \mathcal{F}_t$ per ogni $s, t \in I$ con $s \leq t$.

Definizione A.7. Un processo stocastico è detto adattato alla filtrazione \mathbb{F} se per ogni $t \in I$ X_t è \mathcal{F}_t -misurabile. Se $\mathcal{F}_t = \sigma(X_s, s \leq t)$ per ogni t , allora denotiamo $\mathbb{F} = \sigma(X)$ la filtrazione generata dal processo X .

Definizione A.8. Sia $(\Omega, \mathcal{F}, \mathbf{P})$ uno spazio di probabilità, $I \subseteq \mathbb{R}$ e \mathcal{F} una filtrazione. Sia $X = (X_t)_{t \in I}$ un processo stocastico a valori reali adattato a \mathbb{F} con $\mathbf{E}[|X_t|] < \infty$ per ogni t . X è chiamato una

- *Martingala* se $\mathbf{E}[X_t | \mathcal{F}_s] = X_s$ per ogni $s, t \in I$ con $t > s$.
- *Submartingala* se $\mathbf{E}[X_t | \mathcal{F}_s] \geq X_s$ per ogni $s, t \in I$ con $t > s$.
- *Supermartingala* se $\mathbf{E}[X_t | \mathcal{F}_s] \leq X_s$ per ogni $s, t \in I$ con $t > s$.

Enunciamo quindi un importante teorema

Teorema A.1 (Teorema di convergenza per martingale). *Sia $(X_n)_{n \in \mathbb{N}}$ una submartingala con $\sup\{\mathbf{E}[X_n^+] : n \geq 0\} < \infty$, allora esiste una variabile aleatoria X_∞ che è \mathcal{F}_∞ misurabile con $\mathbf{E}[|X_\infty|] < \infty$ e $X_n \rightarrow X_\infty$ quasi certamente (dove \mathcal{F}_∞ è la sigma algebra terminale della filtrazione).*

A.3 Sistemi dinamici misurabili

Definiamo come prima cosa il concetto di processo stazionario

Definizione A.9. Sia $I \subseteq \mathbb{R}$ un insieme chiuso per addizione. Un processo stocastico $X = (X_t)_{t \in I}$ è detto stazionario se

$$\mathcal{L}[(X_{t+s})_{t \in I}] = \mathcal{L}[(X_t)_{t \in I}]$$

dove con \mathcal{L} indichiamo la legge del processo.

Come prima cosa definiamo un sistema dinamico misurabile e alcuni concetti ad esso collegato:

Definizione A.10. Un sistema dinamico misurabile è una quadrupla $(\Omega, \mathcal{F}, \mathbf{P}, T)$ dove:

- Ω è un insieme.
- \mathcal{F} è una σ -algebra su Ω .
- $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$ è una misura di probabilità.

- $T : \Omega \rightarrow \Omega$ è una trasformazione misurabile che preserva la misura, cioè $\forall A \in \mathcal{F}$ $\mathbf{P}(T^{-1}(A)) = \mathbf{P}(A)$

Strettamente collegato a questo è il concetto di ergodicità

Definizione A.11. Un sistema dinamico misurabile è detto ergodico se rispetta una delle seguenti definizioni equivalenti:

- Per ogni $A \in \mathcal{F}$

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

- Per ogni $A \in \mathcal{F}$ T -invariante ($T^{-1}(A) = A$) si ha che $\mathbf{P}(A) = 0$ o $\mathbf{P}(A) = 1$.
- Ogni integrale primo ($\phi \in L^1(\Omega)$ t.c $\phi = \phi \circ T$ q.c.) è costante q.c.
- Per ogni $\phi \in L^1(\Omega)$ vale

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq t < n} \phi \circ T^t = \int_{\Omega} \phi \, d\mathbf{P} \quad \text{q.c.}$$

- Per ogni $A, B \in \mathcal{F}$ si ha

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq t < n} \mathbf{P}(T^{-t}(A) \cap B) = \mathbf{P}(A)\mathbf{P}(B)$$

Inoltre, data una variabile aleatoria X su $(\Omega, \mathcal{F}, \mathbf{P}, T)$ sistema dinamico ergodico, considerando il processo $\{X_t\}$ dove $X_t(\omega) = X(T^t(\omega))$ per $t \in \mathbb{N}$ (o in \mathbb{Z} se T è invertibile e l'inversa conserva la misura), si ha che tale processo stocastico è ergodico e stazionario. Enunciamo un importante teorema, più volte utilizzato:

Teorema A.2 (Teorema ergodico generalizzato di Breiman). *Sia $\{g_t\}$ una successione di variabili aleatorie reali definite su un sistema dinamico ergodico $(\Omega, \mathcal{F}, \mathbf{P}, T)$. Se $g_t \rightarrow g$ q.c. e $\{g_t\}$ è dominata in L^1 , cioè $\mathbf{E}[\sup_t |g_t|] < \infty$, allora*

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

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

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

Appendice B

Codici

In questa sezione riportiamo i codici utilizzati per le sperimentazioni sui dati reali di mercato.

B.1 Metodo delle frequenze empiriche

```
1 # Empirical.py
2 # Metodo basato su frequenze empiriche dei rendimenti
3
4 # importing modules
5 import csv
6 import numpy as np
7 from matplotlib import pyplot as plt
8 import statistics
9 import math
10 from scipy.special import gamma
11 from scipy.integrate import dblquad
12 from scipy import integrate
13
14 # Questo algoritmo utilizza un procedimento suggerito dall'articolo di
15 # Algoet,
16 # Studia l'andamento di due titoli come una catena di Markov a k stati
17 # approssimati
18 # In questo caso ogni stato e rappresentato da un intervallo di punti
19 # percentuali del
20 # Rendimento di ogni asset
21 # L'algoritmo trascura le incertezze e le covarianze, quindi il
22 # portafoglio ad ogni passo
23 # e' distribuito tutto su un solo titolo
24
25 # Qui vengono settate i principali valori
26
27 # Finestra di tempo che vado a considerare, si usa solo 1 in realta
28 # Gli altri casi richiederebbero una finestra temporale troppo ampia per
29 # diventare efficienti
30
31 k = 1
```

```

27 # Portafoglio iniziale
28 portfolio= 1.
29 portfolios = []
30 portfolio2 = 1.
31
32 # Questi sono gli indicatori degli stati che andiamo a considerare
33 # Per come e scritto l'algoritmo ci interessano in realta gli intervalli
   tra questi valori
34 # Il primo e l'ultimo servono solo per considerare valori estremali
35 # Questi valori sono da intendersi in punti per mille di rendimento
36 s = [-10000, -1000, -700, -500, -300, -100, -50, -10, 0, 10, 50, 100,
   300, 500, 700, 1000, 10000]
37
38 # Inizializzo il dizionario e variabili utili
39 di = {}
40 stot = len(s)
41 ma = s[stot-2]
42
43 # L'algoritmo da un peso base di uno per stati gia visitati, qua si puo
44 # settare un peso diverso per le transizioni di cui so il numero di
   occorrenze
45 peso = 20.
46
47 # Qui si puo settare la frequenza, ossia ogni quanti giorni andiamo a
   operare
48 freq = 21
49
50 # Importiamo i dati del primo asset dal file CSV
51
52 # csv file name
53 filename = "K0.csv"
54
55 # initializing the titles and rows list
56 titles1 = []
57 rows1 = []
58
59 with open(filename, 'r') as csvfile:
60     csvreader = csv.reader(csvfile)
61     titles1 = csvreader.next()
62     for row in csvreader:
63         rows1.append(row)
64
65 # csv file name
66 filename = "IBM.csv"
67
68 # initializing the titles and rows list
69 titles2 = []
70 rows2 = []
71
72 with open(filename, 'r') as csvfile:
73     csvreader = csv.reader(csvfile)
74     titles2 = csvreader.next()
75     for row in csvreader:
76         rows2.append(row)

```

```
77
78 # Qua troviamo l'asset con minore periodo e facciamo partire gli altri al
    punto giusto
79
80 l = min(len(rows1), len(rows2))
81 print(l)
82
83 if(len(rows1)!=1):
84     start1 = len(rows1)-1
85 else:
86     start1 = 0
87
88 if(len(rows2)!=1):
89     start2 = len(rows2)-1
90 else:
91     start2 = 0
92
93 # Qui settiamo i vettori con i closing degli asset e quello coi rapporti
94
95 asset1 = []
96 asset2 = []
97
98 for i in range(0,(int)(l/freq)):
99     asset1.append((float)(rows1[start1+freq*i][5]))
100    asset2.append((float)(rows2[start2+freq*i][5]))
101
102 l = len(asset1)
103
104 # ~ asset1start = []
105 # ~ asset2start = []
106
107 # ~ for i in range(0,l):
108 # ~ asset1start.append((float)(rows1[start1+i][1]))
109 # ~ asset2start.append((float)(rows2[start2+i][1]))
110
111 assets = []
112
113 for i in range(1,l):
114     day = []
115     day.append(asset1[i]/asset1[i-1])
116     day.append(asset2[i]/asset2[i-1])
117     assets.append(day)
118
119
120 # ~ for i in range(1,l):
121 # ~ day = []
122 # ~ day.append(asset1[i]/asset1start[i])
123 # ~ day.append(asset2[i]/asset2start[i])
124 # ~ assets.append(day)
125
126 l = l-1
127
128 # Funzione che ottiene lo stato della catena in cui mi trovo
129 # Restituisce la stringa con cui viene salvato uno stato, semplicemente
```

```

130 # Ogni stato e rappresentato dal rendimento in punti percentuali dei due
      asset
131 # separati da una x
132 # Ogni singolo asset, invece, e rappresentato da un intervallo
133
134 def obtainword(st, en):
135     w = ''
136     for t in range(st, en+1):
137         if t == st:
138             approx = (int)(assets[t][0]*1000.-1000.)
139             for a in range(1, stot):
140                 if approx <= s[a]:
141                     if a == 1:
142                         w = '<' + str(s[1]) + '%'
143                     else:
144                         if a == stot-1:
145                             w = '>' + str(s[a-1]) + '%'
146                         else:
147                             w = str(s[a-1]) + 'to' + str(s[a]) + '%'
148                     break
149             approx = (int)(assets[t][1]*1000.-1000.)
150             for a in range(1, stot):
151                 if approx <= s[a]:
152                     if a == 1:
153                         w = w + 'x' + '<' + str(s[1]) + '%'
154                     else:
155                         if a == stot-1:
156                             w = w + 'x' + '>' + str(s[a-1]) + '%'
157                         else:
158                             w = w + 'x' + str(s[a-1]) + 'to' + str(s[a]) + '%'
159                     break
160             else:
161                 approx = (int)(assets[t][0]*1000.-1000.)
162                 for a in range(1, stot):
163                     if approx <= s[a]:
164                         if a == 1:
165                             w = w + ' ' + '<' + str(s[1]) + '%'
166                         else:
167                             if a == stot-1:
168                                 w = w + ' ' + '>' + str(s[a-1]) + '%'
169                             else:
170                                 w = w + ' ' + str(s[a-1]) + 'to' + str(s[a]) + '%'
171                         break
172                 approx = (int)(assets[t][1]*1000.-1000.)
173                 for a in range(1, stot):
174                     if approx <= s[a]:
175                         if a == 1:
176                             w = w + 'x' + '<' + str(s[1]) + '%'
177                         else:
178                             if a == stot-1:
179                                 w = w + 'x' + '>' + str(s[a-1]) + '%'
180                             else:
181                                 w = w + 'x' + str(s[a-1]) + 'to' + str(s[a]) + '%'
182                     break

```

```

183     return w
184
185
186 # Algoritmo vero e proprio
187
188 # Inizializzo dei contatori per l'analisi finale
189 indecisioni = 0
190 primoasset = 0
191 secondoasset = 0
192
193 for i in range(0,1):
194     if i > k:
195         # Vediamo in che stato mi trovo ora, cioe l'ultima finestra lunga k
196         state = obtainword(i-k, i-1)
197
198         # Prendo dal dizionario le frequenze passate e stimo il valore atteso
199         v1 = 0.
200         v2 = 0.
201
202         if state in di:
203             # Se sono gia stato in questo stato, semplicemente trovo
204             empiricamente
205             # Facendo un rapporto di frequenze le probabilita di transizione in
206             ogni altro
207             # Stato e calcolo il valore atteso del rendimento
208             for a in range(0,stot):
209                 for b in range(0,stot):
210                     if a == 0:
211                         increment1 = (1. + s[1]/1000.)
212                         newstate = state + ' ' + '<' + str(s[1]) + '%'
213                     else:
214                         if a == stot-1:
215                             increment1 = (1. + s[a-1]/1000.)
216                             w = '>' + str(s[a-1]) + '%'
217                         else:
218                             increment1 = (1. + (s[a-1]+s[a])/2000.)
219                             newstate = state + ' ' + str(s[a-1]) + 'to' + str(s[a]) + '
220                             %'
221                     if b == 0:
222                         increment2 = (1. + s[1]/1000.)
223                         newstate = newstate + 'x' + '<' + str(s[1]) + '%'
224                     else:
225                         if b == stot-1:
226                             increment2 = (1. + s[b-1]/1000.)
227                             w = '>' + str(s[b-1]) + '%'
228                         else:
229                             increment2 = (1. + (s[b-1]+s[b])/2000.)
230                             newstate = newstate + 'x' + str(s[b-1]) + 'to' + str(s[b])
231                             + '%'
232                     if newstate in di:
233                         # Se ho gia fatto questo passaggio prima
234                         v1 = v1 + increment1*(1. + peso*(float)(di[newstate]))/((stot
235                         )**2 + peso*(float)(di[state]))

```

```

231         v2 = v2 + increment2*(1. + peso*(float)(di[newstate]))/((stot
) **2 + peso*(float)(di[state]))
232     else:
233         # Se questa transizione di stato non e mai accaduta
234         v1 = v1 + increment1*(1.)/((stot)**2 + peso*(float)(di[state
]))
235         v2 = v2 + increment2*(1.)/((stot)**2 + peso*(float)(di[state
]))
236     # Scelgo ora di puntare tutto sull'asset che mi da valore atteso
maggiore, sto trascurando le covarianze
237     if v1 > v2:
238         portfolio = portfolio*assets[i][0]
239         primoasset = primoasset + 1
240     else:
241         if v2 > v1:
242             portfolio = portfolio*assets[i][1]
243             secondoasset = secondoasset + 1
244         else:
245             portfolio = portfolio * (0.5*assets[i][0] + 0.5*assets[i][1])
246             indecisioni = indecisioni + 1
247     else:
248         # Se non sono mai stato in questo stato
249         portfolio = portfolio * (0.5*assets[i][0] + 0.5*assets[i][1])
250         indecisioni = indecisioni + 1
251
252     # A questo punto aggiorno il dizionario degli stati
253     state = obtainword(i-k+1, i)
254     if state in di:
255         di[state] = di[state] + 1
256     else:
257         di[state] = 1
258
259     state = obtainword(i-k,i)
260     if state in di:
261         di[state] = di[state] + 1
262     else:
263         di[state] = 1
264
265     else:
266         # Se e troppo presto
267         portfolio = portfolio * (0.5*assets[i][0] + 0.5*assets[i][1])
268         indecisioni = indecisioni + 1
269
270     # Aggiornamenti vari
271     portfolios.append(portfolio)
272
273
274 # Analisi dei dati
275
276 wmax = 1.
277 dd = []
278 ddmax = 0.
279 for i in range(0, len(portfolios)):
280     wmax = max(wmax, portfolios[i])

```

```
281 dd.append(1.-portfolios[i]/wmax)
282 if dd[i] > ddmx:
283     ddmx = dd[i]
284
285 cal = []
286 for i in range(1,len(portfolios)):
287     cal.append(portfolios[i]/portfolios[i-1]-1)
288 mum = np.mean(cal)
289 sdev = np.std(cal)
290
291 sharpe = 16*mum/(sdev*(freq**0.5))
292 calmar = 16*mum/(ddmx*freq)
293
294 mug = portfolios[len(portfolios)-1]**(1./(len(portfolios)))-1
```


Bibliografia

- [AC88] 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.
- [Alg92] Paul Algoet. Universal schemes for prediction, gambling and portfolio selection. *Ann. Probab.*, 20(2):901–941, 04 1992.
- [CO96] T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348–363, March 1996.
- [CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.
- [GUW08] László Györfi, Frederic Udina, and Harro Walk. Experiments on universal portfolio selection using data from real markets. 02 2008.
- [Kle13] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2013.
- [LZ78] A. Lempel and J. Ziv. Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, 24(5):530–536, Sep. 1978.
- [MM08] A. C. G. Mennucci and S. K. Mitter. *Probabilità e informazione*. Edizioni della Normale, Pisa, Italia, 2008.
- [OC98] Erik Ordentlich and Thomas M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4):960–982, 1998.
- [Tho06] E. O. Thorp. The kelly criterion in blackjack, sports betting and the stock market. *Handbook of Asset and Liability Management*, 1:385–428, 2006.