Sequential Investment, Universal Portfolio Algos and Log-loss

Chaitanya Ryali, ECE UCSD

March 3, 2014

Table of contents

Definitions and Notations

- A *market vector* $\underline{x} = \{x_1, x_2, \ldots, x_m\}$ for m assets is a vector of nonnegative real numbers representing price relatives for a given trading period.
- $x_i \geq 0$ denotes the ratio of closing to opening price of the ith asset for that period.
- An initial wealth invested in m assets according to the fractions $Q_1, Q_2, \ldots, Q_m$ multiplies by a factor of $\sum_{i=1}^{m} x_i Q_i$ at the end of the period.
- Market behavior during n trading periods is represented by a sequence of market vectors $\underline{x}^n = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$.

Definitions and Notations

- The probability simplex in $\mathbb{R}^m$ is denoted by $\triangle_{m-1}$.
- An investment strategy Q for n trading periods is a sequence $\underline{Q}_1, \ldots, \underline{Q}_n$ of vector valued functions $\underline{Q}_t : \mathbb{R}_+^{t-1} \to \triangle_{m-1}$
- ith component $Q_{i,t}(\underline{x}^{t-1})$of vector $\underline{Q}_t(\underline{x}^{t-1})$ denotes the fraction of the current wealth invested in the i th asset at the beginning of the tth period on the basis of the past market behavior $\underline{x}^{t-1}$

Wealth Factor

$$S_n(Q, \underline{x}^n) = \prod_{t=1}^{n} \left(\sum_{i=1}^{m} x_{i,t} Q_{i,t}(\underline{x}^{t-1}) \right) \qquad (1)$$

denotes the *wealth factor* of strategy Q after n trading periods.

- $\underline{Q}_t$ has nonnegative component summing to one expresses no short sales and no buying on margin.

Examples

- *Buy-and-Hold*:

$$S_n(Q, \underline{x}^n) = \sum_{j=1}^{m} Q_{j,1} \prod_{t=1}^{n} x_{j,t}$$
$$\leq \max_{j=1,\dots,m} \prod_{t=1}^{n} x_{j,t}$$

Examples

- *Constantly Rebalanced Portfolios*:
  - Parametrized by a probability vector $\underline{B} = (B_1, B_2, \ldots, B_m) \in \triangle_{m-1}$
  - $\underline{Q}_t(\underline{x}^{t-1}) = \underline{B}$ regardless of t and $\underline{x}^{t-1}$
- $$S_n(\underline{B}, \underline{x}^n) = \prod_{t=1}^{n} (\sum_{i=1}^{m} x_{i,t} B_i).$$
- Example: $(1, \frac{1}{2}), (1, 2), (1, \frac{1}{2}), (1, 2), \ldots$
  - Buy and Hold $\rightarrow$ No profit, No loss
  - CRP :$\underline{B} = (\frac{1}{2}, \frac{1}{2}) \rightarrow (\frac{9}{8})^{n/2}$, exponentially increasing wealth.

Minimax Wealth Ratio

- Given a class $\mathcal{Q}$ of investment strategies, the *worst case logarithmic wealth ratio* of a strategy P is given by

$$W_n(P,\mathcal{Q}) = \sup_{\underline{x}^n} \sup_{\underline{Q}\in\mathcal{Q}} ln \frac{S_n(\mathcal{Q},\underline{x}^n)}{S_n(P,\underline{x}^n)}.$$

- *Minimax logarithmic wealth ratio* is defined as:

$$W_n(\mathcal{Q}) = \inf_{P} W_n(P,\mathcal{Q}).$$

- $W_n(P,\mathcal{Q}) = o(n)$ means strategy P achieves the same exponent of growth as the best reference strategy in class $\mathcal{Q}$ for all market behaviors.

Prediction under log-loss and Investment

- Any investment strategy Q can be used define a forecaster that predicts elements $y_t \in \mathcal{Y}\{1, \ldots, m\}$ of a sequence $y^n \in \mathcal{Y}^n$ with probability vectors $\hat{p}_t \in \triangle_{m-1}$
- *Kelly Market Vectors*: Market vectors $\underline{x}$ with a single component equal to 1 and all other components equal to zero.
- If $\underline{x}_1, \ldots, \underline{x}_n$ are Kelly market vectors, we denote the index of the only non zero component of each vector $\underline{x}_t$ by y_t, we may define a forecaster f by

$$f_t(y|y^{t-1}) = Q_{y,t}(\underline{x}^{t-1}).$$

- f is *induced* by investment strategy Q.

Prediction under log-loss and Investment

- When $\underline{x}^n$ is a sequence of Kelly vectors determined by the indices y^n, we write $S_n(Q, y^n)$ for $S_n(Q, \underline{x}^n)$.
- Note that $S_n(Q, y^n) = f_n(y^n)$, where f is the forecaster induced by Q, where $f_n(y^n) = \prod_{t=1}^n f_t(y_t|y^{t-1})$, where $\sum_{y^n \in \mathcal{Y}^n} f_n(y^n) = 1$.
- Conversely, given a $f_n(y^n)$, we may define

$$f_t(y_t|y^{t-1}) = \frac{f_t(y^t)}{f_{t-1}(y^{t-1})},$$

where $f_t(y^t) = \sum_{y_{t+1}^n \in \mathcal{Y}^{n-t}} f_n(y^n)$.

Prediction under log-loss and Investment

- Log-loss: $l(f_t, y_t) = -\ln f_t(y_t|y^{t-1})$
- Regret against a reference forecaster f is

$$\hat{L}_n - L_{f,n} = \ln \frac{f_n(y^n)}{\hat{p}_n(y^n)} = \ln \frac{Q(y^n)}{P(y^n)},$$

where Q and P are the investment strategies induced by f and $\hat{p}$.

Lemma

Let $\mathcal{Q}$ be a class of investment strategies, and let $\mathcal{F}$ denote the class of forecasters induced by the strategies in $\mathcal{Q}$. Then, the minimax regret

$$V_n(\mathcal{F}) = \inf_{p_n} \sup_{y^n} \sup_{f \in \mathcal{F}} \ln \frac{f_n(y^n)}{p_n(y^n)}$$

satisfies $W_n(\mathcal{Q}) \geq V_n(\mathcal{F})$.

Proof.

Let P be any investment strategy and let p be it's induced forecaster. Then

$$\begin{aligned}\sup_{\underline{x}^n} \sup_{Q \in \mathcal{Q}} \ln \frac{S_n(Q, \underline{x}^n)}{S_n(P, \underline{x}^n)} &\geq \max_{y^n \in \mathcal{Y}^n} \sup_{Q \in \mathcal{Q}} \ln \frac{S_n(Q, y^n)}{S_n(P, y^n)} \\ &= \max_{y^n \in \mathcal{Y}^n} \sup_{f \in \mathcal{F}} \ln \frac{f_n(y^n)}{p_n(y^n)} \\ &= V_n(p, \mathcal{F}) \geq V_n(\mathcal{F}).\end{aligned}$$

$\square$

Given a prediction p, we define an investment strategy P as follows:

$$P_{j,t}(\underline{x}^{t-1}) = \frac{\sum_{y^{t-1}\in\mathcal{Y}^{t-1}} p_t(j|y^{t-1})p_{t-1}(y^{t-1})(\prod_{s=1}^{t-1} x_{y_s,s})}{\sum_{y^{t-1}\in\mathcal{Y}^{t-1}} p_{t-1}(y^{t-1})(\prod_{s=1}^{t-1} x_{y_s,s})}$$

- The obtained investment strategy induces p, and so we say p and P induce each other.
- $\prod_{s=1}^{t-1} x_{y_s,s}$ may be viewed as the return of the extremal investment strategy that, on each trading period t, invests everything on the y_t th asset.

Theorem

Let P be an investment strategy induced by a forecaster p, and let $\mathcal{Q}$ be an arbitrary class of investment strategies. Then for any market sequence $\underline{x}^n$,

$$\sup_{Q\in\mathcal{Q}} \ln \frac{S_n(Q,\underline{x}^n)}{S_n(P,\underline{x}^n)} \leq \max_{y^n\in\mathcal{Y}^n} \sup_{Q\in\mathcal{Q}} \ln \frac{\prod_{t=1}^{n} Q_{y,t}(\underline{x}^{t-1})}{p_n(y^n)}$$

Lemma

let $a_1, \dots, a_n$, $b_1, \dots, b_n$ be non negative numbers. Then,

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq \max_{j=1,\dots,n} \frac{a_j}{b_j},$$

where we define $0/0 = 0$.

Lemma

The wealth factor achieved by an investment strategy Q may be written as

$$S_n(Q, \underline{x}^n) = \sum_{y^n \in \mathcal{Y}^n} \left(\prod_{t=1}^{n} x_{y_t,t}\right)\left(\prod_{t=1}^{n} Q_{y_t,t}(\underline{x}^{t-1})\right).$$

If the investment strategy P is induced by a forecaster p_n, then

$$S_n(P, \underline{x}^n) = \sum_{y^n \in \mathcal{Y}^n} \left(\prod_{t=1}^{n} x_{y_t,t}\right) p_n(y^n).$$

Proof.

$$\begin{aligned} S_n(Q, \underline{x}^n) &= \prod_{t=1}^{n}(\sum_{j=1}^{m} x_{j,t} Q_{j,t}(\underline{x}^{t-1})) \\ &= \sum_{y^n \in \mathcal{Y}^n} (\prod_{t=1}^{n} x_{y_t,t} Q_{y_t,t}(\underline{x}^{t-1})) \\ &= \sum_{y^n \in \mathcal{Y}^n} (\prod_{t=1}^{n} x_{y_t,t})(\prod_{t=1}^{n} Q_{y_t,t}(\underline{x}^{t-1})). \end{aligned}$$

□

Proof.

$$
\begin{aligned}
S_n(P, \underline{x}^n) &= \prod_{t=1}^{n} \Big(\sum_{j=1}^{m} x_{j,t} P_{j,t}(\underline{x}^{t-1}) \Big) \\
&= \prod_{t=1}^{n} \frac{\sum_{j=1}^{m} \sum_{y^{t-1} \in \mathcal{Y}^{t-1}} p_t(y^{t-1} j) x_{j,t} (\prod_{s=1}^{t-1} x_{y_s,s})}{\sum_{y^{t-1} \in \mathcal{Y}^{t-1}} p_t(y^{t-1} j) x_{j,t} (\prod_{s=1}^{t-1} x_{y_s,s})} \\
&= \prod_{t=1}^{n} \frac{\sum_{y^t \in \mathcal{Y}^t} (\prod_{s=1}^{t} x_{y_s,s}) p_t(y^t)}{\sum_{y^{t-1} \in \mathcal{Y}^{t-1}} (\prod_{s=1}^{t-1} x_{y_s,s}) p_{t-1}(y^{t-1})} \\
&= \sum_{y^n \in \mathcal{Y}^n} \Big(\prod_{t=1}^{n} x_{y_t,t} \Big) p_n(y^n).
\end{aligned}
$$

$\square$

Proof.

Fix any market sequence $\underline{x}^n$ and choose any reference strategy $Q^{'} \in \mathcal{Q}$. Denote by $S_n(y^n, \underline{x}^n) = \prod_{t=1}^{n} x_{y_t,t}$, then

$$
\begin{aligned}
\frac{S_n(Q^{'}, \underline{x}^n)}{S_n(P, \underline{x}^n)} &= \frac{\sum_{y^n \in \mathcal{Y}^n} S_n(y^n, \underline{x}^n)(\prod_{t=1}^{n} Q^{'}_{y_t,t}(\underline{x}^{t-1}))}{\sum_{y^n \in \mathcal{Y}^n} S_n(y^n, \underline{x}^n) p_n(y^n)} \\
&\leq \max_{y^n : S_n(y^n, \underline{x}^n) > 0} \frac{S_n(y^n, \underline{x}^n)(\prod_{t=1}^{n} Q^{'}_{y_t,t}(\underline{x}^{t-1}))}{S_n(y^n, \underline{x}^n) p_n(y^n)} \\
&= \max_{y^n \in \mathcal{Y}^n} \frac{\prod_{t=1}^{n} Q^{'}_{y_t,t}(\underline{x}^{t-1})}{p_n(y^n)} \\
&\leq \max_{y^n \in \mathcal{Y}^n} \sup_{q \in \mathcal{Q}} \frac{\prod_{t=1}^{n} Q_{y_t,t}(\underline{x}^{t-1})}{p_n(y^n)}.
\end{aligned}
$$

$\square$

Theorem

Let $\mathcal{Q}$ be a class of static investment strategies, and let $\mathcal{F}$ denote the class of forecasters induced by strategies in $\mathcal{Q}$. Then

$$W_n(\mathcal{Q}) = V_n(\mathcal{F}).$$

Furthermore, the minimax optimal investment strategy is defined by

$$P^*_{j,t}(\underline{x}^{t-1}) = \frac{\sum_{y^{t-1}\in\mathcal{Y}^{t-1}} p^*_t(j|y^{t-1})p^*_{t-1}(y^{t-1})(\prod_{s=1}^{t-1} x_{y_s,s})}{\sum_{y^{t-1}\in\mathcal{Y}^{t-1}} p^*_{t-1}(y^{t-1})(\prod_{s=1}^{t-1} x_{y_s,s})}$$

where p^* is the normalized maximum likelihood forecaster

$$p^*_n(y^n) = \frac{\sup_{Q\in\mathcal{Q}} \prod_{t=1}^{n} Q_{y_t,t}}{\sum_{y^n\in\mathcal{Y}^n} \sup_{Q\in\mathcal{Q}} \prod_{t=1}^{n} Q_{y_t,t}}.$$

Normalized Maximum Likelihood Forecaster

Definition

The normalized maximum likelihood forecaster is defined by the following:

$$p_n^*(y^n) = \frac{\sup_{f \in \mathcal{F}} f_n(y^n)}{\sum_{x^n \in \mathcal{Y}^n} \sup_{f \in \mathcal{F}} f_n(y^n)}$$

Normalized Maximum Likelihood Forecaster

Theorem

For any class $\mathcal{F}$ of experts and integer $n > 0$, the normalized maximum likelihood forecaster p^* is the unique forecaster such that

$$\sup_{y^n \in \mathcal{Y}^n} \left(\hat{L}(y^n) - \inf_{f \in \mathcal{F}} L_f(y^n)\right) = V_n(\mathcal{F}).$$

Moreover, p^* is an equalizer that is, for all $y^n \in \mathcal{Y}^n$,

$$\ln \frac{\sup_{f \in \mathcal{F}} f_n(y^n)}{p_n^*(y^n)} = \ln \sum_{x^n \in \mathcal{Y}^n} \sup_{f \in \mathcal{F}} f_n(x^n) = V_n(\mathcal{F}).$$

Proof.

The normalized maximum likelihood forecaster p^* is minimax optimal for the class $\mathcal{F}$; that is,

$$\max_{y^n \in \mathcal{Y}^n} \ln \sup_{Q \in \mathcal{Q}} \frac{\prod_{t=1}^{n} Q_{y_t,t}}{p_n^*(y^n)} = V_n(\mathcal{F}).$$

Now, let P^* be the investment strategy induced by minimax forecaster p^* for $\mathcal{Q}$. By theorem, we get

$$W_n(\mathcal{Q}) \leq \sup_{\underline{x}^n} \sup_{Q \in \mathcal{Q}} \ln \frac{S_n(Q, \underline{x}^n)}{S_n(P^*, \underline{x}^n)} \leq \max_{y^n \in \mathcal{Y}^n} \sup_{Q \in \mathcal{Q}} \ln \frac{\prod_{t=1}^{n} Q_{y_t,t}}{p_n^*(y^n)} = V_n(\mathcal{F})$$

$\square$

Constantly Rebalanced Portfolios

$$W_n(\mathcal{Q}) = \frac{m-1}{2} \ln n + \ln \frac{\Gamma(1/2)^m}{\Gamma(m/2)} + o(1)$$

Universal Portfolios

- We restrict our attention to class $\mathcal{Q}$ of all constantly rebalanced portfolios.
- Each strategy Q in this class is determined by a vector $\underline{B} = \{B_1, B_2, \ldots, B_m\} \in \triangle_{m-1}$
- The *Universal Portfolio* strategy P is given by

$$P_{j,t}(\underline{x}^{t-1}) = \frac{\int_{\triangle_{m-1}} B_j S_{t-1}(\underline{B}, \underline{x}^{t-1})\mu(\underline{B})d\underline{B}}{\int_{\triangle_{m-1}} S_{t-1}(\underline{B}, \underline{x}^{t-1})\mu(\underline{B})d\underline{B}},$$

$j = 1, 2 \ldots, m$, $t = 1, \ldots, n$, and μ is a density function on $\triangle_{m-1}$.

Universal Portfolios

The wealth achieved by the universal portfolio is just the average of the wealths achieved by the individual strategies in the class.

$$
\begin{aligned}
S_n(P, \underline{x}^n) &= \prod_{t=1}^{n} \sum_{j=1}^{m} P_{j,t}(\underline{x}^{t-1}) x_{j,t} \\
&= \prod_{t=1}^{n} \frac{\int_{\triangle_{m-1}} \sum_{j=1}^{m} x_{j,t} B_j S_{t-1}(\underline{B}, \underline{x}^{t-1}) \mu(\underline{B}) d\underline{B}}{\int_{\triangle_{m-1}} S_{t-1}(\underline{B}, \underline{x}^{t-1}) \mu(\underline{B}) d\underline{B}} \\
&= \prod_{t=1}^{n} \frac{\int_{\triangle_{m-1}} S_t(\underline{B}, \underline{x}^{t}) \mu(\underline{B}) d\underline{B}}{\int_{\triangle_{m-1}} S_{t-1}(\underline{B}, \underline{x}^{t-1}) \mu(\underline{B}) d\underline{B}} \\
&= \int_{\triangle_{m-1}} S_n(\underline{B}, \underline{x}^n) \mu(\underline{B}) d\underline{B}
\end{aligned}
$$

Universal Portfolios

It's like a Buy-and-Hold on all Constantly Rebalanced Portfolios (CRP)

$$S_n(P, \underline{x}^n) = \int_{\triangle_{m-1}} S_n(\underline{B}, \underline{x}^n)\mu(\underline{B})d\underline{B}$$

If it helps, think of it as: (Riemann sum approximation)

$$S_n(P, \underline{x}^n) = \sum_i Q_i S_n(\underline{B}_i, \underline{x}^n),$$

where, given the elements $\triangle_i$ of a fine partition of the simplex $\triangle_{m-1}$, we assume that $\underline{B}_i \in \triangle_{m-1}$ and $Q_i = \int_{\triangle_i} \mu(\underline{B})d\underline{B}$.

Universal Portfolios

Theorem

If μ *is the uniform density on the probability density simplex* $\triangle_{m-1} \in \mathbb{R}^m$*, then the wealth achieved by the universal portfolio satisfies*

$$\sup_{\underline{x}^n} \sup_{\underline{B} \in \triangle_{m-1}} \ln \frac{S_n(\underline{B}, \underline{x}^n)}{S_n(P, \underline{x}^n)} \leq (m-1) \ln(n+1).$$

If the universal portfolio is defined using the $Dirichlet(1/2, \ldots, 1/2)$ *density* μ*, then*

$$\sup_{\underline{x}^n} \sup_{\underline{B} \in \triangle_{m-1}} \ln \frac{S_n(\underline{B}, \underline{x}^n)}{S_n(P, \underline{x}^n)} \leq \frac{m-1}{2} \ln n + \ln \frac{\Gamma(1/2)^m}{\Gamma(m/2)} + \frac{m-1}{2} \ln 2 + o(1).$$

EG investment strategy

- Universal Portfolio involves integration over m-dimensional simplex.
- EG's computational cost is linear in m
- EG investment strategy invests at a time t using the vector $\underline{P}_t = (P_{1,t}, \ldots, P_{m,t})$ where $\underline{P}_t = (1/m, \ldots, 1/m)$ and

$$P_{i,t} = \frac{P_{i,t-1} \exp(\eta(x_{i,t-1}/\underline{P}_{t-1} \cdot \underline{x}_{t-1}))}{\sum_{j=1}^{m} P_{j,t-1} \exp(\eta(x_{j,t-1}/\underline{P}_{t-1} \cdot \underline{x}_{t-1}))}$$

where $i = 1, 2, \ldots, m$ and $t = 2, 3, \ldots$.

EG investment strategy

Special case of gradient-based forecaster:

$$P_{i,t} = \frac{P_{i,t-1} \exp(\eta \nabla \ell_{t-1}(\underline{P}_{t-1})_i)}{\sum_{j=1}^{m} P_{j,t-1} \exp(\eta \nabla \ell_{t-1}(\underline{P}_{t-1})_j)}$$

when the loss function is set as $\ell_{t-1}(\underline{P}_{t-1}) = -\ln \underline{P}_{t-1} \cdot \underline{x}_{t-1}$.

EG investment strategy

Theorem

Assume that the price relatives $x_{i,t}$ all fall between two positive constants $c < C$. Then the worst-case logarithmic wealth ratio of the EG investment strategy with $\eta = (c/C)\sqrt{(8 \ln m)/n}$ is bounded by

$$\frac{\ln m}{\eta} + \frac{n\eta}{8}\frac{C^2}{c^2} = \frac{C}{c}\sqrt{\frac{n}{2}\ln m}.$$

Simple proof without transaction costs

- *Main Idea*: Portfolios that are "near" each other perform similarly, and there is a large fraction of portfolios "near" the optimal one.
- Suppose in hindsight $\underline{B}^*$ is the optimal CRP. Let $\underline{B} = (1-\alpha)\underline{B}^* + \alpha\underline{z}$, for some $\underline{z} \in \triangle_{m-1}$.(Meaning, $\underline{B}$ is close to $\underline{B}^*$).
- For a single period

$$\text{gain of } CRP_{\underline{B}} \geq (1-\alpha)(\text{gain of} CRP_{\underline{B}^*}).$$

- Over n periods,

$$\text{wealth of CRP}_{\underline{B}} \geq (1-\alpha)^n(\text{wealth of CRP}_{\underline{B}^*}).$$

Simple proof without transaction costs

$$
\begin{aligned}
\frac{\text{wealth of UNIVERSAL}}{\text{wealth of best CRP}} &\geq E_{\underline{B} \in \triangle_{m-1}}[(1-\alpha)^n] \\
&= \int_0^1 Prob_{\underline{B} \in \triangle_{m-1}}[(1-\alpha)^n \geq x]dx \\
&= \int_0^1 (1-x^{1/n})^{m-1} dx \\
&= n \int_0^1 y^{n-1}(1-y)^{m-1} dy \\
&= \ldots \\
&= n(\frac{(m-1)!(n-1)!}{n+m-2}!) \\
&= \frac{1}{\binom{n+m-1}{m-1}}
\end{aligned}
$$

Commission

The following assumptions are made:

- The costs paid changing from distribution $\underline{B}_1$ to $\underline{B}_3$ is no more than the costs paid changing from $\underline{B}_1$ to $\underline{B}_2$ and then from $\underline{B}_2$ to $\underline{B}_3$.
- The cost, per dollars, of changing from a distribution $\underline{B}$ to a distribution $(1-\alpha)\underline{B}_1 + \alpha\underline{B}^{'}$ is no more than αc, because at most an α fraction of the money is being moved.
- An investment strategy I which invests an initial fraction α of it's money according to investment strategy I_1 and an initial $1-\alpha$ of it's money according to I_2, will achieve at least α times the wealth of I_1 plus $1-\alpha$ times the wealth of I_2.

Result with Commission

Theorem

In the presence of commission $0 \leq c \leq 1$,

$$\frac{\text{wealth of } UNIVERSAL_c}{\text{wealth of best } CRP} \geq \binom{(1+c)n+m-1}{m-1}^{-1} \geq \frac{1}{((1+c)n+1)^{m-1}}.$$

Result with Commission

Proof.

Based on the properties we assumed, if $B_j \geq (1-\alpha)B_j^*$, then

$$\frac{\text{single-period profit of } \mathrm{CRP}_{\underline{B}}}{\text{single-period profit of } \mathrm{CRP}_{\underline{B}^*}} \geq (1-\alpha)(1-c\alpha).$$

Over n periods, this gives

$$\text{wealth of } \mathrm{CRP}_{\underline{B}} \geq (1-\alpha)^{(1+c)n}(\text{wealth of } \mathrm{CRP}_{\underline{B}^*}).$$

The previous proof can be applied and we can replace n by $(1+c)n$in the final guarantee. □