Lecturer: Rob Schapire
Scribe: Kevin Lin

Lecture #23
April 30, 2013

1 Portfolio Selection

Recall the setting from last time:

- N stocks
- Start on day 1 with \$1
- $p_t(i) = \frac{\text{price of stock i at the end of day t}}{\text{price of stock i at the start of day t}}$
- $S_t =$ wealth at the start of day t
- $w_t(i) =$ fraction of wealth in stock i at the start of day t

We call $p_t(i)$ the price relative. We choose $\mathbf{w}_t$ first and observe $\mathbf{p}_t$ at the end of the day. From these definitions, it is easy to show that

$$S_{t+1} = S_t(\mathbf{w}_t \cdot \mathbf{p}_t) \tag{1}$$

$$S_{T+1} = \prod_{t=1}^{T}(\mathbf{w}_t \cdot \mathbf{p}_t) \tag{2}$$

where S_{T+1} can be interpreted as total wealth. Also, we showed that our goal of maximizing wealth has an equivalent form

$$\max \prod_t (\mathbf{w}_t \cdot \mathbf{p}_t) \leftrightarrow \min \sum_t -\log(\mathbf{w}_t \cdot \mathbf{p}_t) \tag{3}$$

In this analysis, we are not going to make any statistical assumptions about the movement of the stock market. Instead, we would like to find a strategy that works no matter what the stock market does, so we will use online learning.

Last time, we used Bayes Algorithm for this. We had one expert for each stock, and its wealth was shown to be almost as good as the best stock **but**, as we noted, this was a trivial algorithm since we can achieve the same objective value by dividing the wealth among all N stocks equally and leaving it there. In this setting, we asked the question "How do we do as well as the best single stock?" We strive to do something more interesting.

Now, we might ask the question "How do we do as well as the best switching strategy?" A switching-strategy is defined to be a buy-and-hold strategy, but the stock that is held can vary depending on the time interval. We can do this, and this algorithm will simply redistribute the wealth among the N stocks. We could analyse this, but instead we will choose to analyse in terms of a *constant rebalanced portfolio*.

2 Constant Rebalanced Portfolio

A constant rebalanced portfolio (CRP) is exactly what it sounds like. We periodically rebalance our portfolio to maintain target ratios of wealth among the N stocks. That is, we decide on fixed proportions on each stock and we periodically rebalance our portfolio to achieve these proportions. For example, a very specific type of CRP is the *uniform constant rebalanced portfolio* where all N of our proportions are equal. Thus, each stock has an equal amount of wealth in it. Imagine the following two scenarios to get an intuition on why CRP is a good strategy:

Example 1: Imagine having 60% of your wealth in the stock fund, and 40% in the bond fund. If the value of stock goes up relative to the wealth of bond, then we naturally would want to sell our stock and use the proceeds to buy bond. We are buying high and selling low. A CRP strategy does this.

Example 2: Imagine having two stocks:

- Stock 1: Never does anything (i.e. you put your money under a mattress).
- Stock 2: Wildly volatile. On odd days, it halves in value. On even days, it doubles

Assuming both stocks start at \$1, we can easily see that the price of stock 1 will always remain at \$1, and the price of stock 2 will be \$1 on odd days and $\$\frac{1}{2}$ on even days. In terms of price relatives $p_t(i)$,

$$\text{price of stock } 1 = 1 \qquad \forall t \tag{4}$$

$$\text{price of stock } 2 = \begin{cases} 1 & \text{if } t \text{ is odd} \\ 1/2 & \text{if } t \text{ is even} \end{cases} \qquad \forall t \tag{5}$$

And in terms of price relatives:

$$p_t(1) = 1 \qquad \forall t \tag{6}$$

$$p_t(2) = \begin{cases} 1/2 & \text{if } t \text{ is odd} \\ 2 & \text{if } t \text{ is even} \end{cases} \qquad \forall t \tag{7}$$

We note that buy and hold will **never** earn money in this setting. What about UCRP?

$$S_1 = 1 \tag{8}$$

$$S_2 = S_1(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2}) = \frac{3}{4} S_1 \tag{9}$$

$$S_3 = S_2(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2) = \frac{3}{2} S_2 \tag{10}$$

so if we start with $\$S_t$ on day t, then $S_{t+2} = S_t \cdot \frac{3}{4} \cdot \frac{3}{2} = \frac{9}{8} S_t$. Thus, every 2 days, our money goes up by $\frac{9}{8}$. This is an exponential growth.

In general, we have a lot of choices for CRP. We hope one of these choices for CRP will be decent, and we want an algorithm to find this. Let us move on to the general setting of CRP.

Let us define our CRP strategy by the proportions we allocate on the N stocks. We define b_i to be the fraction of wealth we put on stock i on each round (for example, 60% in stock, 40% in bond). We next define $\mathbf{b} = \langle b_1, ..., b_N \rangle$ (a valid distribution over the N stocks). There are uncountably many choices of $\mathbf{b}$. Wouldn't it be nice if someone told you which $\mathbf{b}$ is the best? Well, we can't hope for that, but we have a lot of algorithms that can do almost as well as the best $\mathbf{b}$ in hindsight. Let us analyse one particular algorithm tailored for this setting called the *Universal Portfolio Algorithm*.

3 Universal Portfolio Algorithm

From here on, we will freely interchange terminology among investment instrument, stock, and strategy. We previously talked about the buy&hold strategy. It generates almost as much wealth as the single best investment instrument (or stock). So if there are N stocks, we would divide evenly on all N stocks, and our total wealth after T rounds would be almost as much as the wealth generated if we had initially put all our wealth in the best single stock. We follow this setup to analyse CRP's, but instead we split our wealth among all possible CRP's. That is, we take our money and divide it evenly among *all* possible $\mathbf{b}$'s. But of course, there are infinite (uncountable) possibilities for $\mathbf{b}$, so we imagine this as putting an infinitesimal small amount of money on each possible $\mathbf{b}$. Thus, we are using buy&hold over uncountably many strategies. We can represent each $\mathbf{b}$ as a point in the simplex since $\|\mathbf{b}\|_1 = \sum_i b_i = 1$.

One question we seek to answer is, "On day t, how much money do we invest in each CRP strategy?" At the start of day t, we have:

$$\text{wealth invested in CRP } \mathbf{b} \text{ after } T \text{ time periods} = \prod_{s=1}^{t-1} (\mathbf{b} \cdot \mathbf{p}_s) d_\mu(\mathbf{b}) \tag{11}$$

where the term $\prod_{s=1}^{t-1} (\mathbf{b} \cdot \mathbf{p}_s)$ represents the amount of wealth generated had we started with \$1, and the product goes up to day $t-1$ since we are interested in the wealth generated up until the start of day t. On the other hand, $d_u(\mathbf{b})$ represents the scaling factor for the tiny amount of money we put in CRP $\mathbf{b}$.

So our total wealth (after T time periods) is simply

$$\text{total wealth after } T \text{ time periods} = \int_{\mathbf{b}} \prod_{s=1}^{t-1} (\mathbf{b} \cdot \mathbf{p}_s) d_\mu(\mathbf{b}) \tag{12}$$

where our integral is over the simplex of possible CRP's.

So how much money (wealth) was put in each stock i? We simply calculate another integral

$$\text{total wealth in stock } i \text{ after } T \text{ time periods} = \int_{\mathbf{b}} b_i \left[\prod_{s=1}^{t-1} (\mathbf{b} \cdot \mathbf{p}_s) d_\mu(\mathbf{b}) \right] \tag{13}$$

where the last term b_i represents the fraction of wealth in stock i.

Thus, finishing up our algorithm, we calculate the fraction of wealth $w_t(i)$ in stock i by simply taking the ratio

$$w_t(i) = \frac{\int_{\mathbf{b}} b_i \left[\prod_{s=1}^{t-1} (\mathbf{b} \cdot \mathbf{p_s}) d_u(\mathbf{b}) \right]}{\int_{\mathbf{b}} \prod_{s=1}^{t-1} (\mathbf{b} \cdot \mathbf{p_s}) d_\mu(\mathbf{b})} \tag{14}$$

This quantity represents how we rebalance our portfolio at each time step.

4 Theoretical Guarantees for UP

We now prove the following about the Universal Portfolio (UP) algorithm.

Theorem: Wealth of UP $\geq$ (Wealth of best CRP in hindsight)$\left(\frac{1}{(T+1)^{N-1}}\right)$ where T is the number of time steps, and N is the number of stocks.
We interpret the best CRP as the best sequence of proportions. We first note that this bound may not look satisfactory at all since $\left(\frac{1}{(T+1)^{N-1}}\right)$, a scaling factor, is a very small fraction. In terms of actual wealth, this bound is extremely weak. However, our goal is to show that UP matches the exponential rate as before in Example 2, that is, matches a rate of the form 2^{cT} where c is the doubling rate. This rate is what we wish to match.

Proof Idea: We note that if we completed all T rounds using UP, looking back, we would notice one particular $\mathbf{b}^*$ that did extremely well. Fortunately, our algorithm placed some weight (albeit infinitesimal at first) on $\mathbf{b}^*$, so in some sense, we captured this best performing $\mathbf{b}^*$. We aim to show two things:

- If we look at a neighborhood of $\mathbf{b}^*$, we would've generated a large amount of money from this neighborhood.
- This neighborhood is actually quite large compared to the size of the simplex.

We will actually prove a slightly weaker bound (but not by much).

Proof: Let $\triangle$ denote the simplex of all CRP's. Formally:

$$\triangle = \{\mathbf{b} \in [0,1]^N \ \mid \ \sum_i (b_i) = 1\} \tag{15}$$

and let $\mathbf{b}^*$ be the best CRP in hindsight. We now define the neighborhood η around $\mathbf{b}^*$ to be

$$\eta(\mathbf{b}^*) = \{(1-\alpha)\mathbf{b}^* + \alpha\mathbf{z} \ \ |\mathbf{z} \in \triangle\} \tag{16}$$

where α will be some small number that we will determine later on in this proof. We interpret $\mathbf{z}$ to be alternative CRP's that we use to "jiggle" the best CRP with.

Step 1: If $\mathbf{b} = (1-\alpha)\mathbf{b}^* + \alpha\mathbf{z}$, then the wealth obtained by $\mathbf{b}$ will be at least $(1-\alpha)^T$(wealth obtained by $\mathbf{b}^*$)
Proof of Step 1: If we give at least $(1-\alpha)$ of wealth to $\mathbf{b}^*$, then as long as $\alpha\mathbf{z}$ doesn't go negative, we are done. This is trivially true since $\mathbf{z}$ is a discrete distribution, and $\mathbf{p_t}$ is a vector of price relatives. Price relatives (as we recall from the definition) are nonnegative. Formally, on each time step t,

$$\begin{aligned}
\text{wealth obtained by } \mathbf{b} \text{ on time step } t &= \mathbf{b} \cdot \mathbf{p_t} && (17)\\
&= (1-\alpha)\mathbf{b}^* \cdot \mathbf{p_t} + \alpha\mathbf{z} \cdot \mathbf{p_t} && (18)\\
&\geq (1-\alpha)\mathbf{b}^* \cdot \mathbf{p_t} && (19)
\end{aligned}$$

Applying this over all time steps will give us the proof. $\square$.

Step 2: $\frac{\mathrm{Vol}(\eta(\mathbf{b}^*))}{\mathrm{Vol}(\triangle)} = \alpha^{N-1}$ where Vol stands for the volume function.

Proof of Step 2: We first note that the volume of a simplex does not change if we shift the entire simplex. Thus

$$\begin{aligned}
\text{Vol}(\eta(\mathbf{b}^*)) &= \text{Vol}(\{(1-\alpha)\mathbf{b}^* + \alpha\mathbf{z} \;\; |\mathbf{z} \in \triangle\}) && (20)\\
&= \text{Vol}(\{\alpha\mathbf{z} \;\; |\mathbf{z} \in \triangle\}) && (21)\\
&= \alpha^{N-1}\text{Vol}(\triangle) && (22)
\end{aligned}$$

where the last equality uses the fact that the simplex is an $N-1$ dimensional object, so when we shrink $\triangle$ down to $\alpha\triangle$, each dimension is scaled down by α, and there are $N-1$ dimensions. □

All that remains is to choose an appropriate value of α and to tie these two steps together. From both steps, we can conclude that

$$\text{Wealth generated by the entire neighborhood } \eta(\mathbf{b}^*) \geq \alpha^{N-1}(1-\alpha)^T(\text{wealth of } \mathbf{b}^*) \tag{23}$$

So we choose $\alpha = \frac{1}{T+1}$, which will give a right hand side of

$$\frac{1}{e(T+1)^{N-1}}\text{Wealth of best CRP in hindsight} \tag{24}$$

which is the weaker bound we sought to prove. ■

We make some closing remarks that this analysis did not take into account transaction costs, and this strategy works better if the stocks are uncorrelated and volatile.

5 Game Theory Introduction

We now start the last topic for the course, game theory. Game theory is the study of not just ordinary games, but any interaction of any kind. We can intuitively get the feeling that it is connected to online learning, and as we will see, game theory will tie many topics together.

So what is a game? We start with the most basic example, Rock-Paper-Scissor. If we put this game into a matrix, we get:

	R	P	S
R	1/2	1	0
P	0	1/2	1
S	1	0	1/2

and we imagine the rows as the action of the row player Mindy, and the columns as the action of the column player Max. The objective for Mindy is to minimize the resulting objective value, and Max to maximize. We look at this game from Mindy's perspective.

In principle, we can put any game into this type of matrix notation. For example, we can do this in chess as long as we enumerate all possible strategies, where a strategy says what move you would make in any possible board position. Once both sides have decided on a strategy, the outcome of the game can be determined without actually playing the game. This (of course) is unreasonable, but theoretically possible. To play such a game, Mindy picks row i and Max picks column j simultaneously. These are *pure strategies* (deterministic). The loss for Mindy is denoted by the (i, j)-th coordinate of the game matrix M, or $M(i, j) \in [0, 1]$.

We change things slightly so now Mindy picks an action from a distribution P over the rows, and

Max picks a distribution Q over the columns. These are now *mixed strategies*. We still have both pick their actions simultaneously. We can calculate the expected loss as follows

$$\text{Expected loss} = \sum_{i,j} P(i)M(i,j)Q(j) \tag{25}$$
$$= P^\intercal M Q \tag{26}$$
$$=: M(P,Q) \tag{27}$$

where the last equality is a definition for a function M that takes two distributions P and Q as input. This is a slight overload of notation. We note that if P is fixed, then $P^\intercal M$ is fixed, and this expected loss becomes linear in Q. Thus, Max would pick a pure strategy to maximize his objective value given a fixed P.

So what happens if Mindy must pick first and Max picks afterwards? If Mindy picks a strategy P, then with this knowledge, Max will pick Q to maximize $M(P,Q)$, giving a loss to Mindy of

$$\max_Q M(P,Q).$$

Knowing this, Mindy will choose P to minimize this expression. Thus, the outcome of this game would be

$$\min_P \max_Q M(P,Q) = \min_P \max_j M(P,j) \tag{28}$$

where the minimum and maximum on the left hand side are taken over all possible mixed strategies for Mindy and Max respectively. The right hand side of this equation uses our observation above of Max reducing his choices to only pure strategies. Similarly, if Max were to pick first, the outcome of the game would be

$$\max_Q \min_P M(P,Q) = \max_Q \min_i M(i,Q) \tag{29}$$

From our observations from Lecture #14 with Lagrange duality for SVM's, we know that playing second must be better (or at least not worse) since the second player gets more information. However, as we will show in next class, thanks to Von Neumann's minmax theorem, we amazingly get

$$\min_P \max_Q M(P,Q) = \max_Q \min_P M(P,Q) \tag{30}$$

which means there is no advantage to playing second.