optimal transport research report

Zhang Jinrui*
jerryzhang40@gmail.com

20250714

Abstract

In this article, I tried to do some research about optimal transport.

1 recite of the problem

This link [5, Transportation-theory] is what I find the closest defination Prof. Pass gives on lecture.

1.1 one dimensional case

But the defination in the wiki still somehow different from what Prof. Pass gives on lecture.

So the one dimensional problem is we have two probabilistic distribution on $X = Y = [0, 1]$. The two distribution is $h(x) > 0$ and $g(y) > 0$ We have a cost function $c(x, y) = \|x - y\|^2$ of the transportation cost of the probabilistic density.

We want to find a Map $T : X \to Y$ satisfied $\forall y \in Y$ we have $g(y)dy = \int_{x \in T^{-1}(y)} h(x)dx$ which is $g(T(x))T'(x) = \int_{x \in T^{-1}(y)} h(x)$

If $T \in C^1$ then we can assert that $T' \neq 0$ for if we assume some $\xi \in X$ such that $T'(\xi) = 0$ then $0 = g(T(x))T'(x) = \int_{x \in T^{-1}(y)} h(x) > 0$ which can't be true. So $T'(\xi) \neq 0$

By the intermediate value theorem, T is a one to one map from X to Y then $T'(\xi) > 0$

then T is a monotonous function.

1.2 two dimensional case

So the two dimensional problem is we have two probabilistic distribution on $X = Y = [0, 1]^2$. The two distribution is $h(x) > 0$ and $g(y) > 0$ We have a cost function $c(x, y) = \|x - y\|^2$ of the transportation cost of the probabilistic density. same argument as one dimensional case we can get $\det(\frac{\partial y}{\partial x}) \neq 0$ which the $\frac{\partial y}{\partial x}$ is the jacobian matrix.

*alternative email:zhangjr1022@mails.jlu.edu.cn

1.3 one period binomial model

The model has some basic settings. For this model it only have discrete period. In this case it's one period.

Suppose the time involves tick 0 and tick 1, between them is the time period. the price at tick 0 is S^1 and at tick 1 the price is a random variable S^λ which shows by the following chart.

$variable$	up	$down$
S^λ	$uS^1 = S^u$	$dS^1 = S^d$
$P(S^\lambda)$	p	$1-p$

which satisfies $d < 1 + r < u$ the no-arbitage condition. which r is the bank account interest rate.

So the two assets during this period shows by the following chart.

$asset$	$tick0$	$tick1$
$stock$	S^1	S^λ
$bank$	1	$1+r$

We want to make a derivative product, call option. which gives the costomos back f_u or f_d amount of money at tick1, while charge x amount of money when tick0

$asset$	$tick0$	$tick1$
$stock$	S^1	S^λ
$bank$	1	$1+r$
$option$	f^1	f^λ

which f can be also view as a random variable which is shown the next chart.

$variable$	up	$down$
f^λ	f^u	f^d
$P(f^\lambda)$	p	$1-p$

we make a portfolio of Δ units of stock and Ψ units of the bank account.

then we want at tick1 we have

$\Delta S^\lambda + \Psi(1+r) = f_\lambda$

this is the equation in a vecotr matrix form.

where $f^\lambda = \begin{bmatrix} f_u \\ f_d \end{bmatrix}$ and we can denote $\Phi^\delta = \begin{bmatrix} \Delta \\ \Psi \end{bmatrix}$ and $S^\lambda = S^\delta = \begin{bmatrix} uS^1 \\ dS^1 \end{bmatrix}$ and $\Sigma^\lambda_\delta = \begin{bmatrix} S^u & r+1 \\ S^d & r+1 \end{bmatrix}$ and we can rewrite it as follow $\Sigma^\lambda_\delta \Phi^\delta = f^\lambda$

we then denote $\Sigma^k_\delta \Omega^\delta_\lambda = \delta^k_\lambda$ then $\Omega^k_\lambda \Sigma^\lambda_\delta \Phi^\delta = \Omega^k_\lambda f^\lambda$ then $\delta^k_\delta \Phi^\delta = \Omega^k_\lambda f^\lambda$ then $\Phi^k = \Omega^k_\lambda f^\lambda$

And if we denote $\sigma^1_k = \begin{bmatrix} S^1 & 1 \end{bmatrix}$ then we get $f^1 = \sigma^1_k \Phi^k = \sigma_k \Omega^k_\lambda f^\lambda$

1.4 two periods binomial model

for one case if we already get to f^u at tick1, then we now have a price at uS^1

$variable$	up	$down$
$S^{u\lambda}$	$uS^u = S^{uu}$	$dS^u = S^{ud}$
$P(S^{u\lambda})$	p	$1-p$

$asset$	$tick0$	$tick1$	$tick2$
$stock$	S^1	S^u	$S^{u\lambda}$
$bank$	1	$1+r$	$(1+r)^2$
$option$	f^1	f^u	$f^{u\lambda}$

where

$variable$	up	$down$
$f^{u\lambda}$	f^{uu}	f^{ud}
$P(f^{u\lambda})$	p	$1-p$

$\Sigma_\delta^{u\lambda} = \begin{bmatrix} S^{uu} & r+1 \\ S^{ud} & r+1 \end{bmatrix}$ we then denote $\Sigma_\delta^{uk}\Omega_{\lambda\tau}^\delta = \delta_\lambda^k \delta_\tau^u$ then $\Sigma_\delta^{u\lambda}\Phi_\xi^\delta = f^{u\lambda}\delta_\xi^u$ then $\Omega_{\lambda\tau}^\delta \Sigma_\delta^{u\lambda}\Phi_\xi^\delta = f^\lambda \delta_\xi^u$

Then the same formula for the f^u holds true we get $\Phi_u^k = \Omega_{\lambda u}^k f^{u\lambda}$
$f^1 = \sigma_k^1 \Phi^k = \sigma_k \Omega_\lambda^k f^{u\lambda}$

2 Principal

The basic principal is to find the vulnerability of the market. To be more practical and more mathematically, we are longing for the breaking of the symmetries. According to the Efficient-market hypothesis [4, Hypothesis] we can give out the first and fundamental definition of this report.

Definition 2.1 (EMH means Symmetries)**.** We say a market is completely EMH if and only if for every condition A, the random variable of the logarithmic growth rate W agrees that $\mathbb{E}_{W|A}(W) \doteq 0$.

Analytically, if we want to prove that a market is EMH we need to prove for all the condition A $\mathbb{E}_{W|A}(W) \doteq 0$, this is obviously a very large amount of work to do, and this is also a very profitless effort(we do all the strive to find that we can't make money).

Base on this definition, all the experiments which have been taken or will be taken in the future are dedicated to find a condition X that $\mathbb{E}_{W|X}(W) > 0$ significantly.

3 Current Roadmap

3.1 statistics

3.1.1 X is Close Price

Choose $X = \mathbb{R}$ which is the close price.

3.1.2 X is the difference between MA and close price

By intuition, when the price is above the MA, It should some how comes back, and vice versa.

3.1.3 statistics for Total Variation and fractals

This method is about Hurst Exponent, Fractal Dimension and the Bounded Variation function model.

standard deviation with moving average. $S(t,n)$ means start from t include n data point, that is $S(t,n) = \sqrt{\frac{1}{n}\sum_{k=t}^{t+n-1}(P_k - mean(P))}$ for the Hurst Exponent the $S(n)$ shoud be stable over time

3.2 machine learning

3.2.1 X is prefix sequence

To find feature in the prefix sequence. The Transformer and 1D-CNN is now under the consideration. This topic derives experiments including EXP20250119

3.2.2 Data process

Currently method is only short period. The W in Definition 2.1 is Lr_t which is a very local property. The maximum and minimum is just in the range of 3 time unit. This is way to small period.

We need to some how ignore the local fluctuation, but get the long term trend and the maximum and minimum in a meaningful sense in a specific scale extend.

Here I want to introduce a new method based on the gradient descent.

The method in the history include the linear regression by Gauss. And the fellow mathematician developed some high degree regression method, include polynomial regression, logistic regression which is even transfinite.

But I want a more simple and more interpolation like method.

We use a simplest interpolation, which is the linear interpolation.

By the classical interpolation theory, the interpolation point in dimension t should also be non-uniform, and it will be determined by the Gauss' method.

Now we can use the power of the gradient descent.

Suppose we want to interpolate N points to a period of time series $P_t, t \in [T_s, T_e]$. $\Delta = \{\delta_n | \delta_n = (\tau_n, \pi_n)\}_{n=1}^{N}$. Then for this regression problem the parameter set is $\Delta^* = \Delta \setminus \{\tau_1, \tau_N\}$ which $|\Delta^*| = 2\,(N-1)$ The boundary must be fixed, that is $\tau_1 = T_s, \tau_N = T_e$

Initialize the set $\{\tau_n\}_{n=1}^N$ to be uniformly space from T_s to T_e

We assume $\forall 1 \le n \le N-1, \tau_n < \tau_{n+1}$ and for each segment $[\tau_n, \tau_{n+1}]$ we do the oridary linear regression for the points $\{\psi_t = (t, P_t)\,|\tau_n \le t \le \tau_{n+1}\}$

Then we can get the coefficient of determination $\left\{R_n^2\right\}_{n=1}^{N-1}$ for each line segment $[\tau_n, \tau_{n+1}]$

Then set the target function $f(\psi_{t t\in[T_s,T_e]}, \Delta^*) = \sum_{n=1}^{N-1} R_n^2$

We want to optimize the Δ^* and make the $f(\psi_{t t\in[T_s,T_e]}, \Delta^*)$ optimum.

To be more practical, we explicitly define each variable name.

There is one hyper-parameter N, which defined the number of segments.

The input date to fit is a (K,2)-tensor $\psi_k = (t_k, P_k)$, $k \in \mathbb{Z}, 1 \le k \le K$

Then define a (N)-tensor χ_n, $n \in \mathbb{Z}, 1 \le n \le N$ Use this χ_n we can calculate the length L_n of each segment. use the softmax $\sigma(\cdot)$ to make the sum of L_n is a const; Set the data length $\Lambda = t_K - t_1$ Then $L_n = \Lambda\sigma(\chi_n)$ This ensure $\sum_{n=1}^N L_n = \Lambda$

Then $\tau_n = (\sum_{i=1}^n L_i) + t_1$ and $n \in \mathbb{Z}, 0 \le n \le N$ $\tau_0 = t_1$

Then define a new (N+1)-tensor π_n $n \in \mathbb{Z}, 0 \le n \le N$

The implementation is in [1, computeGraph].

The result image is as follow.

Figure 1: Interpolation N=5 N=10

3.2.3 evolution strategy

evolution strategy can be viewed as a zeroth-order gradient descent method. Now let's define some notations.

Definition 3.1 (notations)**.** τ represents the discrete time of the stock.

$\Delta\tau$ denotes the discretized time interval.

$\pi(\tau)$ is the stock price at time τ.

Suppose at (t, p) an automlized amount of money ΔM is invested. We need to have a strategy when to sell. so $P_{S_\theta}(\ln(\frac{\tau}{t}), \ln(\frac{\pi}{p}))$ is the probability to sell at the state $(\ln(\frac{\tau}{t}), \ln(\frac{\pi}{p}))$

Suppose at (t, p) an automlized amount of money ΔM is selled. We need to have a strategy when to buy. so $P_{B_\theta}(\ln(\frac{\tau}{t}), \ln(\frac{\pi}{p}))$ is the probability to buy at the state $(\ln(\frac{\tau}{t}), \ln(\frac{\pi}{p}))$

all the probability is to make a decision whether to buy or sell. but if the balance is not enough then we can't buy or sell. this situation is the survival conditon that also i want the model to learn.

S_θ, B_θ is the parameter of the model. we can define a FCN network with sigmoid function to represent the probability.

4 Principal of Unbiased Model

The long last dream of finding the asymmetry in market then make money from EMH failure seems to be too hard to achieve. So in the other hand, given the EMH hypothesis, hold true for the market, we can dedicate ourself to find a Unbiased model.

Definition 4.1 (UnbiasedModel)**.** We say a model is Unbiased if and only if it work under given condition $\mathbb{E}_{W|A}(W) \doteq 0$ 2.1.

This is hard to manage to do, but not all totally impossible.

4.1 sochastic differential equation

This section is a background theory which not a specific method.

Definition 4.2 (autonomous sochastic differential equation)**.** $x_0 = X(t_0)$ is the initial value. $dx_t = V(x_t)dt$ where $V(x)$ is a random variable which determined only by the position vector x_t Then the solution of this equation $X(t) = \int_{t_0}^{t} V(X(s))ds$ then for each t we want to get the random distribution of $X(t)$ this random distribution is the solution to the original equation.

4.2 +1s-2s method

The theory is as follow. We begin with the random variable in 2.1 as a normal distribution $N(W, \sigma)$ which $\mathbb{E}(W) \doteq 0$

A simple derivation would give us result as follow. set a lower sell bar $Ls = -2\sigma$, and a higher sell bar $Hs = +1\sigma$ for the first time step dt we have a distribution $D_1(x) = N(W, \sigma)$ By the way, D_0 is a single point distribution. and the probability of the first step can be simplify as a discrete variable W_1 which

$action$	Ls	$Hold$	Hs
W_1	-2σ	0	σ
$P(W_1)$	0.0228	0.8185	0.1587

Which the expect contribution is $\mathbb{E}_1 = -2\sigma 0.0228 + \sigma 0.1587 = 0.1131\sigma$

Then we need to calculate the second step $D_2(x) = \frac{\int_{-2\sigma}^{\sigma} D_1(t)D_1(x-t)dt}{\int_{-2\sigma}^{\sigma} D_1(t)dt}$ simplify a little bit we can get $\overline{D_2(x)} = 0.8185 D_2(x) = \int_{-2\sigma}^{\sigma} D_1(t)D_1(x-t)dt$ The discrete variable W_2 which

$action$	Ls	$Hold$	Hs
W_2	-2σ	0	σ
$P(W_2)$	0.0713	0.7544	0.1743

Which the expect contribution is $\mathbb{E}_2 = (-2\sigma 0.0713 + \sigma 0.1743)0.8185 = 0.02596\sigma$

Repeat this process we can get $D_3(x) = \frac{\int_{-2\sigma}^{\sigma} D_2(t)D_2(x-t)dt}{\int_{-2\sigma}^{\sigma} D_2(t)dt}$ simplify a little bit we can get $\overline{D_3(x)} = 0.7544 * 0.8185 D_2(x) = \int_{-2\sigma}^{\sigma} D_2(t)D_2(x-t)dt$ The discrete variable W_3 which

$action$	Ls	$Hold$	Hs
W_3	-2σ	0	σ
$P(W_3)$	0.1619	0.6827	0.1554

Which the expect contribution is $\mathbb{E}_3 = (-2\sigma 0.1619 + \sigma 0.1554)0.8185 * 0.7544 = -0.1040\sigma$

So we should end sell it within two day.

For further experiments we need to statistically get the distribution $D_1(W)$ in stead of assuming it is a normal distribution.

The action strategy is to buy and wait one day. check if it exceed the Hs or Ls. and then wait for the next day. and sell any way. to get a total expectation of $\mathbb{E} = \mathbb{E}_1 + \mathbb{E}_2 = 0.1391\sigma$

And notice when in practice $\mathbb{E} - fee$ should greater than zero, or it won't beat the cost fee.

Unfortunately, $\mathbb{E}_1 >= 0$ could be a result that $\int_{-2\sigma}^{\sigma} D_2(t)dt <= 0$ and exactly $\mathbb{E}_1 + \int_{-2\sigma}^{\sigma} D_2(t)dt = 0$

But this method still have one hope to be verify by the pure mathematics is that the expectation sequence add up to zero which means $\delta_i^i \mathbb{E}_i = 0$ but we just need to cut off when $\mathbb{E}_i < 0$ and sell out. it will be fine?

Short gacha guess would reveal that it will have a expectation of $\mathbb{E}_{total} = -0.5\sigma < 0$

which we need the $\mathbb{E}_{cut} > 0.5\sigma$

4.3 max draw back control bsPair method

Figure 2: illustration

Sometimes, the "keep holding" strategy is to cost in one side, is that, the draw back is very high, basically, the same as the local variation rate. So we need to control the maximum draw back.

This time I want to log a method, bsPair method. The basic idea is to sell when there is a 1% draw back. We sell it, and then, when the price hit this again, we buy it again. As Figures2 shows, 'b' is the buy point, and 's' is the sell point. suppose the trade fee is ϕ. And the random variable N_ϕ is the total pair number of buy and sell, after the case is over. I say a case is over is the automata fufill the sell and finished bar such as reach the level P_H. Then the expectation of the random variable is $\mathbb{E}(N_\phi)$.

Suppose the price is P_{t_0} at time t_0, we buy μ amount of stock at time t_0, The two line is P_L and P_H.

Normally in automata we have $C_H = \ln(\frac{P_H}{P_{t_0}})$ and $C_L = \ln(\frac{P_L}{P_{t_0}})$ is fixed, to inversely get P_L and P_H.

μ amount of stock is the principal. And the total gas fee is $\phi\mathbb{E}(N_\phi)$, By the time we sell at P_H.

So all the fee is $\phi\mathbb{E}(N_\phi)$ the total profit is $\mu\frac{P_H}{P_{t_0}} = \mu\exp(C_H)$.

And the fee rate is C_ϕ which $\phi = \mu\exp(C_\phi)$

We need $\mu\exp(C_\phi)\mathbb{E}(N_\phi) \ll \mu\exp(C_H)$

Which is equivalent to $C_\phi + \ln(\mathbb{E}(N_\phi)) \ll C_H$

When $C_\phi \doteq 5 \times 10^{-3}$ and $C_H \doteq 10 \times 10^{-3}$ we can have total logarithmic growth rate at $C = \ln(\exp(C_H) - \exp(C_\phi)\mathbb{E}(N_\phi))$.

When in classic automata, we have $\mathbb{E}(N_\phi) = 2$

But if we have this method, we can control the maximum draw back, which is estimated as $\chi = \mu(\exp(C_\phi)\mathbb{E}(N_\phi) + \exp(C_L))$

So we can take back $1 - \chi$ of the principal.

If the first condition $\mu\exp(C_\phi)\mathbb{E}(N_\phi) \ll \mu\exp(C_H)$ is satisfied, then we can estimated $\exp(C_\phi)\mathbb{E}(N_\phi) + \exp(C_L) < \exp(C_L) + \exp(C_H)$

Definition 4.3 (max draw back)**.** The max draw back is the max draw back because of the variation of the price.

5 Current Roadmap of UnbiasedModel

5.1 Grids

Grids are a very old and simple method and there is nothing special to say about this method.

But there are several experiments need to be carried out to get a better optimization of this ancient method.

First we need to get the distribution of the P_t We always refered to the distribution of this $Lr_t = \ln(\frac{P_{t+1}}{P_t})$ But now we want the distribution of P_t itself.

P_t is the historical data which is definite set the random variable ρ is the random variable of P_t The meaning is you get a time you can say the possibility of get P is $P(\rho = P)$

6 EXP20250119 prefix sequence

This experiment use the prefix sequence to be the condition space X.

Definition 6.1 (prefix space)**.** All possible prefix sequence form a space X

If the market is EMH. By Definition 2.1, this means for all subset of X, $\mathbb{E}_{W|A}(W) \doteq 0$. That is $\forall A \subseteq X, \mathbb{E}_{W|A}(W) \doteq 0$

So if we want to find the vulnerability of the market we want to find a X such that there exist a $A \subseteq X$ agrees that $\mathbb{E}_{W|A}(W) > 0$

Consider a simplest senario, that is, we only treat the sequence as descend and ascend. Then for a space of length L then $|X| = 2^L$, which is a nightmare for statistics, so I'm forced to use machine learning to try to find something.

6.1 Data regularization

Current data regularization is as follow. Suppose we have a time series $P_t \in \mathbb{R}_+$ We want it range from $(0, 1)$ and have a symmetry about descend and ascend around zero. So let $Lr_t = \ln(\frac{P_{t+1}}{P_t})$. $Lr_t \in \mathbb{R}$ and then $Sg_t = \tanh(Lr_t)$

Then $X = \{(Sg_k)|t_{start} \leq k \leq t_{end}\}$

6.2 Conclusion

In [2, PrefixSequence] I implemented the function to get Sg_t

The first Stage experiment result is very bad in [3, Cov1d]

7 EXP20250120 MA

This experiment use the difference between MA and close price as the space X.

Definition 7.1 (MACP difference). $D_t^N := A_t^N - P_t$ which $A_t^N = \sum_{k=t-N+1}^{t} P_k$ which N is a hyper-parameter when N is a single number then $D_t := D_t^N$ is a time series of number when N can be multiple different numbers then $D_t^N$is a vector time series.

7.1 Fix N

Suppose N is a fixed single number then D_t is a time serie. To simplify the problem we define a classification variable V

Definition 7.2 (Ternary Classification). $V_t := \begin{cases} 1, & P_t \in (0.1, +\infty) \\ 0, & P_t \in (-0.1, 0.1) \\ -1, & P_t \in (-\infty, -0.1) \end{cases}$

Use the bin size of 0.01 which is the smallest increment of the price. Do the statistics over all $D_t \in \mathbb{R}$ with the bin size 0.01 to obtain the conditional probability distribution of the random variable V_t

7.2 Result

Figure 3: RandomWalk(left),realdata(right)

The result shows for the Random RandomWalk, the curve is like sigmoid but exchude a normal distribution for $V_t = 0$ While the distribution for realdata is more intricate the blue curve is more like a t-distribution.

Figure 4: Fixed:RandomWalk(left),realdata(right)

The original code that GPT write, use D_{t+1} to predict V_t which in the RandomWalk shows significantly distribution bias.This fixed version has test on the RandomWalk don't have any bias on the RandomWalk.

8 EXP20250615 +1s-2s method

This experiment would practice the +1s-2s method. Basically it needs to statistically get the distribution of the random variable D_1 and to compute the convolution cross verify the result of D_2 and get the expectation.

8.1 statistic of D_1

9 EXP20250702 ρ

Suppose we have the historical data P_t We want to get ρ. We need a function to get the distribution of the ρ, Like a box-line graph, of any given time period.

References

[1] Zhang Jinrui. Data process. `https://github.com/unjerry/bnbCAC/blob/report/gptTestField/computeGraph.py`. Accessed: 2025-01-27.

[2] Zhang Jinrui. Exp20250119prefixsequence. `https://github.com/unjerry/bnbCAC/blob/report/gptTestField/EXP20250119PrefixSequence.py`. Accessed: 2025-01-20.

[3] Zhang Jinrui. Exp20250119prefixsequencegptcov1d. `https://github.com/unjerry/bnbCAC/blob/report/gptTestField/EXP20250119PrefixSequenceGPTCov1d.py`. Accessed: 2025-01-20.

[4] wikipedia. Efficient-market hypothesis. `https://www.wikiwand.com/en/articles/Efficient-market_hypothesis`. Accessed: 2025-01-19.

[5] wikipedia. Transportation theory. `https://www.wikiwand.com/en/articles/Transportation_theory_(mathematics)`. Accessed: 2025-07-14.