From portfolio theory to optimal transport and Schrödinger bridge in-between

Soumik Pal
University of Washington, Seattle

McMaster University, Feb 14 2020

Based on joint work with T.-K. Leonard Wong
University of Toronto, formerly UW, Seattle.

Introduction: portfolio theory

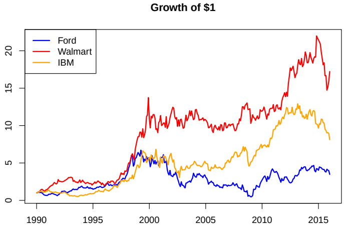
Stochastic portfolio theory

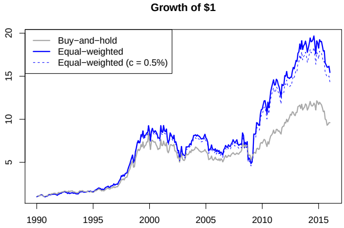
- Market weights for n stocks: $\mu = (\mu_1, \ldots, \mu_n)$ in Δ_n, **unit simplex**

$$\Delta_n = \left\{ (p_1, \ldots, p_n) : \; p_i > 0, \; \sum_i p_i = 1 \right\}.$$

- μ_i = Proportion of the total capital that belongs to ith stock.
- Process in time, $\mu(t), \; t = 0, 1, 2, \ldots$.
- Portfolio: $\pi = (\pi_1, \ldots, \pi_n) \in \Delta_n$.
- Portfolio weights: π_i=Proportion of the total value that belongs to ith stock.
- $\pi(t), \; t = 0, 1, 2, \ldots$ is another process in the unit simplex.

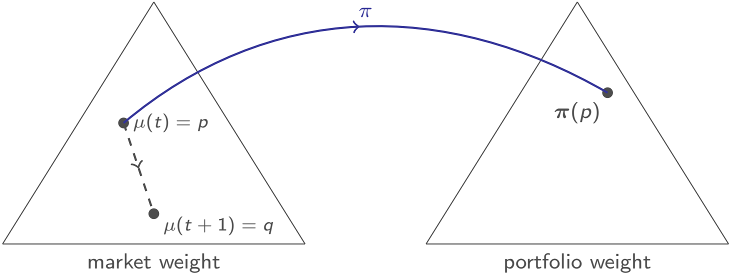
Actively managed portfolios vs. passive index portfolios





Portfolio map

- $\boldsymbol{\pi} : \Delta_n \to \Delta_n$. $\pi(t) \equiv \boldsymbol{\pi}(\mu(t))$.
- Start by investing \$1 in portfolio and compare with index.
- Relative value process: $V(\cdot)$ = ratio of growth of \$1.



$$\frac{V_\pi(t+1)}{V_\pi(t)} = \sum_{i=1}^{n} \boldsymbol{\pi}_i(p) \frac{q_i}{p_i}$$

Constant-weighted portfolio: $\boldsymbol{\pi}(p) \equiv \pi \in \Delta_n$

Relative value and MCM portfolios

$$\begin{tikzcd}
 & \mu_1 \arrow[dr] & \\
\mu_0 \arrow[ur] & & \mu_2 \arrow[d] \\
\mu_m \arrow[u] & & \mu_3 \arrow[dl] \\
 & \mu_4 \arrow[ul, dashed] &
\end{tikzcd}$$

Figure: A market cycle

- Suppose we make **no statistical assumptions**, but are confident on the support $S \subseteq \Delta_n$ of the future market weights.
- Given $\epsilon > 0$, want $\liminf_{t\to\infty} V(t) > \epsilon$, *irrespective* of market paths.
- Are there portfolio maps $\boldsymbol{\pi}$ that guarantee that. No transac cost.
- (Multiplicative cyclical monotonicity) Necessary that after any market cycle: $V(m+1) \geq 1$.

Definition

- $\varphi : \Delta_n \to \mathbb{R} \cup \{-\infty\}$ is exponentially concave if e^{φ} is concave.

$$\mathrm{Hess}(\varphi) + \nabla\varphi \, (\nabla\varphi)' \le 0.$$

- Examples: $p, \pi \in \Delta_n$, $0 < \lambda < 1$.

$$\begin{aligned} \varphi(p) &= \frac{1}{n} \sum_i \log p_i, \quad \varphi(p) = \sum_i \pi_i \log p_i, \\ \varphi(p) &= \log\left(\sum_i \pi_i p_i \right), \quad \varphi(p) = \frac{1}{\lambda} \log\left(\sum_i p_i^{\lambda} \right). \end{aligned}$$

- Also called (K, N) convexity by Erbar, Kuwada, and Sturm '15.
- Statistics, optimization, machine learning.
  Cesa-Bianchi and Lugosi '06, Mahdavi, Zhang, and Jin '15.
- Compare log-concave functions.

Gradients of e-concave functions

- Fact 1: Gradients of exp-concave functions are probabilities.
- (Fernholz '02, P. and Wong '15). φ, exp-concave on Δ_n. Define π by

$$\pi_i = p_i \left(1 + D_{e(i)-p}\varphi(p)\right).$$

  Then $\pi \in \Delta_n$. $e(i)$ is ith standard basis vector.
- Portfolio map: $\pi : \Delta_n \to \Delta_n$.
- Example: $\varphi(p) = \frac{1}{n}\sum_i \log p_i$. Then $\pi(p) \equiv (1/n, \ldots, 1/n)$.

Theorem (P.-Wong '15, Fernholz '02)

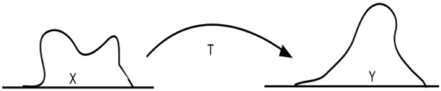
Assume $S \subseteq \Delta_n$ convex. π is MCM portfolio map on S if and only if $\exists\, \varphi : \Delta \to (0, \infty)$, exponentially concave:

1. $\exists \epsilon > 0$ *s.t.* $\inf_{p \in S} \varphi(p) \geq \log \epsilon$.
2. *And*
$$\frac{\pi_i(p)}{p_i} = 1 + D_{e(i)-p}\varphi(p).$$

- The 'if' part was essentially shown by Fernholz. Functionally generated portfolios.
- We show the 'only if' part.

Optimal Transportation

The Monge problem 1781



- P, Q - probabilities on $\mathcal{X} = \mathbb{R}^d = \mathcal{Y}$.
- $c(x, y)$ - cost of transport. E.g., $c(x, y) = \|x - y\|$ or $c(x, y) = \frac{1}{2}\|x - y\|^2$.
- Monge problem: minimize among $T : \mathbb{R}^d \to \mathbb{R}^d$, $T_{\#}P = Q$,

$$\int c(x, T(x))\, dP.$$

Kantorovich relaxation 1939

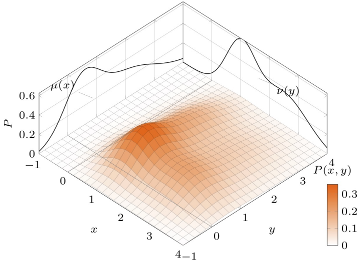


Figure: by M. Cuturi

- $\Pi(P, Q)$ - *couplings* of (P, Q) (joint dist. with given marginals).
- (Monge-) Kantorovich relaxation: minimize among $\nu \in \Pi(P, Q)$

$$\inf_{\nu \in \Pi(P,Q)} \left[\int c\,(x, y)\, d\nu \right].$$

- Linear optimization in ν over convex $\Pi(P, Q)$.

Example: quadratic Wasserstein

- Consider $c(x, y) = \frac{1}{2} \|x - y\|^2$.
- Assume P, Q has densities ρ_0, ρ_1.

$$\mathbb{W}_2^2(P, Q) = \mathbb{W}_2^2(\rho_0, \rho_1) = \inf_{\nu \in \Pi(\boldsymbol{\rho_0}, \boldsymbol{\rho_1})} \left[\int \|x - y\|^2 \, d\nu \right].$$

Theorem (Y. Brenier '87)

There exists convex ϕ such that $T(x) = \nabla \phi(x)$ solves both Monge and Kantorovich OT problems for (ρ_0, ρ_1) uniquely.

Idea: Rockafellar's cyclical monotonicity.

A MK optimal transport problem

- Unit simplex is an abelian group. If $p, q \in \Delta_n$, then

$$(p \odot q)_i = \frac{p_i q_i}{\sum_{j=1}^n p_j q_j}, \quad \left(p^{-1}\right)_i = \frac{1/p_i}{\sum_{j=1}^n 1/p_j}.$$

- $e = (1/n, \ldots, 1/n)$.
- K-L divergence or relative entropy as "distance":

$$H(q \mid p) = \sum_{i=1}^n q_i \log(q_i/p_i).$$

- Take $\mathcal{X} = \mathcal{Y} = \Delta_n$.

$$c(p, q) = H\left(e \mid p^{-1} \odot q\right) = \log\left(\frac{1}{n}\sum_{i=1}^n \frac{q_i}{p_i}\right) - \frac{1}{n}\sum_{i=1}^n \log\frac{q_i}{p_i} \geq 0.$$

An optimal transport description of mcm portfolios

Theorem (P.-Wong '15, '18)

Given density (ρ_0, ρ_1) on Δ_n, there exists an exp concave function φ such that the map

$$q = T(p) \propto 1 + D_{e(\cdot)-p}\varphi(p) \in \Delta_n$$

solves the Monge and MK transport problem uniquely.

- The portfolio map is

$$\boldsymbol{\pi}(p) = T(p) \odot p^{-1}, \quad T(p) = p \odot \boldsymbol{\pi}(p).$$

- Conversely all MCM portfolios are given this way.
- Transport maps are smooth MTW (Khan & Zhang '19).

Models parametrized by probabilities

- What do ρ_0, ρ_1 signify in portfolio theory?
- Roughly ρ_0 is the distribution of the market weights.
- ρ_1 is the distribution of the proportions of shares held in portfolio.
- They affect solely by their supports.
- Can be used from data to fit portfolios.

A tabular comparison

Group	$(\mathbb{R}^n, +)$	$(\Delta_n, \odot)$
Id	0	$e = (1/n, \ldots, 1/n)$
Cost	$\|y - x\|^2$	$H(e \mid q \odot p^{-1})$
Potential	convex	exp-concave
Monge solution	$y = \nabla \phi(x)$	$q = \widetilde{\nabla} \varphi(p)$
Displacement	$y - x$	$\pi(p) = q \circ p^{-1}$.

Computations from discrete data

Big interest in statistics

- Transport of discrete probabilities. Atoms $(x_1, x_2, \ldots, x_N), (y_1, y_2 \ldots, y_N)$.
- $p = (p_1, \ldots, p_N) \mapsto q = (q_1, \ldots, q_N)$.
- OT is a linear program. $O(N^3)$ steps.
- (Cuturi '13) "Entropic regularization" can be computed in about $O(N^2 \log N)$ steps.
- Sinkhorn algorithm - discrete IPFP.
- What about explicit approximate solutions?

Stochastic processes and OT

- Define transition kernel of Brownian motion with diffusion h:

$$p_h(x, y) = (2\pi h)^{-d/2} \exp\left(-\frac{1}{2h}\|x - y\|^2\right),$$

  and joint distribution $\mu_h(x, y) = \rho_0(x) p_h(x, y)$ of a particle initially sampled from ρ_0 and evolving as BM.
- Imagine large N many Brownian particles - temperature $h \approx 0$.

Schrödinger's problem

- Condition on initial configuration $\approx \rho_0$ and terminal configuration $\approx \rho_1$.
- Exponentially rare. On this rare event what do particles do?
- Schrödinger '31, Föllmer '88, Léonard '12.
- There is a coupling between initial and terminal configurations.
- Given $X_0 = x_0$ and $X_1 = x_1$, the path is a Brownian bridge with diffusion h.
- As $h \to 0+$, straight lines joining MK optimal coupling (ρ_0, ρ_1).
- *Schrödinger's bridge*.

Explicit solution

- Suppose distinct data.

$$L_0 = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \quad L_1 = \frac{1}{N}\sum_{j=1}^{N} \delta_{y_j}.$$

- Conditional coupling is explicit. $\mathcal{S}_N$ - set of permutations.
- Then

$$\nu_N^* = \sum_{\sigma\in\mathcal{S}_N} q(\sigma)\frac{1}{N}\sum_{i=1}^{N} \delta_{(x_i, y_{\sigma_i})}.$$

- Gibbs measure on $\mathcal{S}_N$:

$$q(\sigma) = \frac{\exp\left(-\frac{1}{2h}\sum_i \|x_i - y_{\sigma_i}\|^2\right)}{\sum_{\rho\in\mathcal{S}_N}\exp\left(-\frac{1}{h}\sum_i \|x_i - y_{\rho_i}\|^2\right)}.$$

Back to the Dirichlet transport

- If $p, q \in \Delta_n$, then

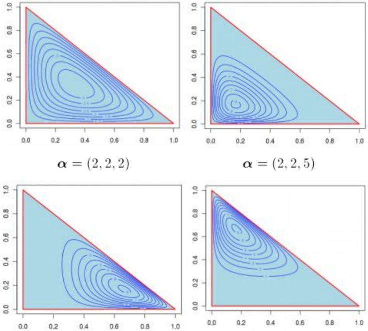
$$(p \odot q)_i = \frac{p_i q_i}{\sum_{j=1}^n p_j q_j}, \quad \left(p^{-1}\right)_i = \frac{1/p_i}{\sum_{j=1}^n 1/p_j}.$$

- $H(q \mid p) = \sum_{i=1}^n q_i \log(q_i/p_i)$.
- MK OT with cost

$$c(p,q) = H\left(e \mid p^{-1} \odot q\right) = \log\left(\frac{1}{n}\sum_{i=1}^n \frac{q_i}{p_i}\right) - \frac{1}{n}\sum_{i=1}^n \log\frac{q_i}{p_i} \geq 0.$$

- What is the corresponding picture for the Schrödinger bridge?

Dirichlet distribution



$\boldsymbol{\alpha} = (2,2,2)$ $\boldsymbol{\alpha} = (2,2,5)$

$\boldsymbol{\alpha} = (5,2,2)$ $\boldsymbol{\alpha} = (2,5,2)$

- Symmetric Dirichlet distribution $\mathrm{Diri}(\lambda)$, density $\propto \prod_{j=1}^{n} p_j^{\lambda/n-1}$.
- Probability distribution on the unit simplex. If $U \sim \mathrm{Diri}(\cdot)$,

$$\mathrm{E}\,(U) = e = (1/n, \ldots, 1/n), \quad \mathrm{Var}(U_i) = O\left(\frac{1}{\lambda}\right).$$

Dirichlet transition

- Haar measure on $(\Delta_n, \odot)$ is Diri (0), $\nu(p) = \prod_{i=1}^n p_i^{-1}$.
- Consider transition probability: $p \in \Delta_n$, $U \sim \mathrm{Diri}(\lambda)$, $Q = p \odot U$.

$$f_\lambda(p, q) = c\nu(q) \exp\left(-\lambda c(p, q)\right), \quad \text{(P.-Wong '18)}.$$

- Compare with Brownian transition. Temperature: $h = \frac{1}{\lambda}$.
- As $\lambda \to \infty$, $f_\lambda \to \delta_p$. As $\lambda \to 0+$, $f_\lambda \to \mathrm{Diri}(0)$.

Multiplicative Schrödinger problem

- Given discrete i.i.d. samples $p_1, \ldots, p_N \sim \rho_0$
- $q_1, \ldots, q_N \sim \rho_1$.
- $\mathcal{S}_N$ - set of permutations.
- Define "Schrödinger bridge":

$$\nu_N^* = \sum_{\sigma \in \mathcal{S}_n} q(\sigma) \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_i, y_{\sigma_i})}.$$

- Gibbs measure on $\mathcal{S}_N$:

$$q(\sigma) = \frac{\prod_{i=1}^N f_\lambda(x_i, y_{\sigma_i})}{\sum_{\rho \in \mathcal{S}_N} \prod_{i=1}^N f_\lambda(x_i, y_{\rho_i})}.$$

Pointwise convergence

Theorem (P.-Wong '18)

Let $\lambda = \lambda_N = N^{2/n}$. *Then, almost surely,*

$$\mathbb{W}_2^2(\nu_N^*, Monge) = O\left(N^{-1/n} \log N\right),$$

where Monge is the optimal Monge coupling between ρ_0, ρ_1.

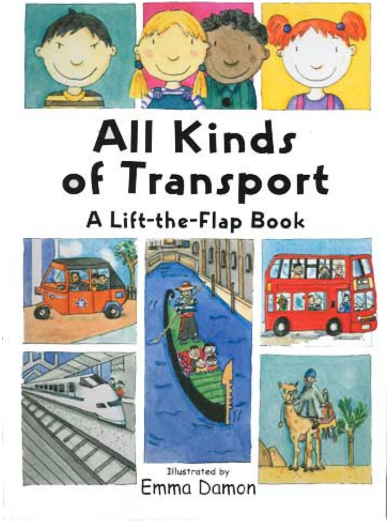
The explicit Schrödinger coupling is an approximate solution to the OT for discrete large data.

On the difference between entropic cost and the optimal transport cost
arxiv math.PR:1905.12206

Multiplicative Schrödinger problem and the Dirichlet transport
(With Leonard Wong) 1806.05649. To appear in PTRF.

Exponentially concave functions and a new information geometry
(With Leonard Wong) AOP '18.

The geometry of relative arbitrage
(With Leonard Wong) Mathematics and Financial Economics '15



Merci beaucoup et Thank you very much