Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems

Aurélien Alfonsi

CERMICS, Ecole des Ponts, Paris-Est University
Joint work with Jacopo Corbetta and Benjamin Jourdain

Jim Gatheral's conference, NY

October 15, 2017

Structure of the talk

The convex order

Let $X, Y : \Omega \to \mathbb{R}^d$ two random variables with respective law μ and ν. X is smaller than Y in the convex order if

$$\forall \phi : \mathbb{R}^d \to \mathbb{R},\ \mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)],$$

for any convex function ϕ such that both expectations exist. In this case, we write $X \leq_{cx} Y$ or $\mu \leq_{cx} \nu$.
Strassen's theorem : (1965) Assume $\int_{\mathbb{R}^d} |y|\nu(dy) < \infty$. $\mu \leq_{cx} \nu$ iff there exists a martingale kernel $Q(x, dy)$ such that $\mu Q = \nu$, i.e. $\int \mu(dx)Q(x, dy) = \nu(dy)$.
Notation : $\Pi^M(\mu, \nu) = \{\pi(dx, dy) = \mu(dx)Q(x, dy) : \forall x \in \mathbb{R}^d,\ \int_{\mathbb{R}^d} |y|Q(x, dy) < \infty \text{ and } \int_{\mathbb{R}^d} yQ(x, dy) = x\}$

Martingale Optimal Transport in Finance

We assume $r = 0$. $(S_t)_{t \geq 0}$: price process of d assets. Suppose that we know for $0 < T_1 < T_2$ the law of S_{T_1} and S_{T_2} (denoted μ_1 and μ_2), and that we want to price an option that pays $c(S_{T_1}, S_{T_2})$ at time T_2, with $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

Price bounds for the option :

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) \pi(dx, dy), \pi \in \Pi^M(\mu_1, \mu_2) \to \textit{minimize/maximize}.$$

Multi-marginal case : payoff $c(S_{T_1}, \dots, S_{T_n})$ with $c : (\mathbb{R}^d)^n \to \mathbb{R}$.
Beiglböck, Henry-Labordère, Penkner (2013) : Duality and connection with super/subhedging strategies.

Sampling in the Convex order

When approximating $\mu \leq_{\text{cx}} \nu$ by discrete measures $\mu_I = \sum_{i=1}^{I} p_i \delta_{x_i}$ and $\nu_J = \sum_{j=1}^{J} q_j \delta_{y_j}$ (typically with i.i.d. samples) we may not have $\mu_I \leq_{\text{cx}} \nu_J$.

Our goal : to construct $\tilde{\mu}_I$ (resp. $\tilde{\nu}_J$) "close" to μ_I (resp. ν_J) such that $\tilde{\mu}_I \leq_{\text{cx}} \tilde{\nu}_J$.

Motivation : Numerical methods for Martingale Optimal Transport (MOT) problems. We can use linear programming solvers to solve :

$$\sum_{i=1}^{I} \sum_{j=1}^{J} r_{ij} c(x_i, y_j)$$

under the constraints

$$r_{ij} \geq 0, \ \sum_{i=1}^{I} r_{ij} = q_j, \ \sum_{j=1}^{J} r_{ij} = p_i \text{ and } \sum_{j=1}^{J} r_{ij} y_j = p_i x_i.$$

Monte-Carlo : calculate together option prices and their bounds.

Existing methods to approximate measures in the convex order

Quantization :

- Dual quantization preserves the convex order in dimension one. In general, gives a measure $\hat{\nu}$ such that $\nu \leq_{cx} \hat{\nu}$. (Pagès Wilbertz 2012)
- Primal quantization gives a measure $\hat{\mu}$ such that $\hat{\mu} \leq_{cx} \mu$.
- Drawbacks : ν and thus μ must have a compact support. Only for 2 marginals. Computation time.

Dimension 1 (Baker's thesis, 2012) : Assume $\mu \leq_{cx} \nu$ and let $\hat{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{I\int_{\frac{i-1}{I}}^{\frac{i}{I}} F_\mu^{-1}(u)du}$ and $\hat{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{I\int_{\frac{i-1}{I}}^{\frac{i}{I}} F_\nu^{-1}(u)du}$.
Then, we have $\hat{\mu}_I \leq_{cx} \hat{\nu}_I$ for any $I \in \mathbb{N}^*$.

A first idea : to equalize the means

Suppose $\mu \leq_{cx} \nu$, $X_1, \dots, X_I$ i.i.d. $\sim \mu$ and $Y_1, \dots, Y_J$ i.i.d. $\sim \nu$. We set $\bar{X}_I = \frac{1}{I}\sum_{i=1}^I X_i$ and $\bar{Y}_J = \frac{1}{J}\sum_{j=1}^J Y_j$, and

$$\tilde{\mu}_I = \frac{1}{I}\sum_{i=1}^I \delta_{X_i+m-\bar{X}_I},\ \tilde{\nu}_J = \frac{1}{J}\sum_{j=1}^J \delta_{Y_j+m-\bar{Y}_J},$$

with $m = \int x\mu(dx)$ if it is known explicitly or $\bar{X}_I$ otherwise.

- Under suitable conditions, a.s., $\exists M, \forall I, J \geq M, \tilde{\mu}_I \leq_{cx} \tilde{\nu}_J$.
- For $X_1 \overset{law}{=} exp(\sigma_\mu G - \frac{\sigma_\mu^2}{2})$, $Y_1 \overset{law}{=} exp(\sigma_\nu G - \frac{\sigma_\nu^2}{2})$ with $\sigma_\mu = 0.24$, $\sigma_\nu = 0.28$, $I = 100$, $\mathbb{P}(\tilde{\mu}_I \leq_{cx} \tilde{\nu}_I) \approx 0.45$.
  $\Longrightarrow$ **need for a non asymptotic approach.**

Characterization of the convex order in dimension 1

We set $\mathcal{P}_1(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x|\mu(dx) < \infty\}$.
For $x \in \mathbb{R}$, $F_\mu(x) = \mu((-\infty, x])$, $\bar{F}_\mu(x) = \mu([x, +\infty))$. For $p \in (0,1)$,

$$F_\mu^{-1}(p) = \inf\{x \in \mathbb{R} : F_\mu(x) \geq p\}.$$

For $t \in \mathbb{R}$, we define $\varphi_\mu(t) = \int_{-\infty}^t F_\mu(x)dx = \int_{\mathbb{R}}(t-x)^+\mu(dx)$ and $\bar{\varphi}_\mu(t) = \int_t^{+\infty} \bar{F}_\mu(x)dx = \int_{\mathbb{R}}(x-t)^+\mu(dx)$.

Theorem 1

Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. The following conditions are equivalent :

(i) $\mu \leq_{cx} \nu$,

(ii) $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$ *and* $\forall t \in \mathbb{R}, \varphi_\mu(t) \leq \varphi_\nu(t)$,

(iii) $\int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$ *and* $\forall t \in \mathbb{R}, \bar{\varphi}_\mu(t) \leq \bar{\varphi}_\nu(t)$,

(iv) $\int_0^1 F_\mu^{-1}(p)dp = \int_0^1 F_\nu^{-1}(p)dp$ *and* $\forall q \in (0,1), \int_q^1 F_\mu^{-1}(p)dp \leq \int_q^1 F_\nu^{-1}(p)dp$.

Application to discrete random variables

Corollary 2

Let $\mu = \sum_{i=1}^{I} p_i \delta_{x_i}$ and $\nu = \sum_{j=1}^{J} q_j \delta_{y_j}$ be two probability measures on $\mathbb{R}$. Without loss of generality, we assume that $x_1 < \cdots < x_I$, $y_1 < \cdots < y_J$ and $p_1 p_I q_1 q_J > 0$. Then we have $\mu \leq_{\text{cx}} \nu$ if, and only if

(i) $y_1 \leq x_1$ and $y_J \geq x_I$,

(ii) for all j such that $x_1 \leq y_j \leq x_I$, $\varphi_\mu(y_j) \leq \varphi_\nu(y_j)$,

(iii) $\sum_{i=1}^{I} p_i x_i = \sum_{j=1}^{J} q_j y_j$.

We can replace (ii) by

(ii') for all j such that $x_1 \leq y_j \leq x_I$, $\bar{\varphi}_\mu(y_j) \leq \bar{\varphi}_\nu(y_j)$.

The increasing/decreasing convex orders

Definition/Proposition : The following statements are equivalent :

(i) $\mu \leq_{\text{icx}} \nu$ (resp. $\mu \leq_{\text{dcx}} \nu$),

(ii) $\forall t \in \mathbb{R}$, $\bar{\varphi}_\mu(t) \leq \bar{\varphi}_\nu(t)$ (resp. $\varphi_\mu(t) \leq \varphi_\nu(t)$),

(iii) $\forall q \in [0,1]$, $\int_q^1 F_\mu^{-1}(p)dp \leq \int_q^1 F_\nu^{-1}(p)dp$ (resp. $\int_0^q F_\mu^{-1}(p)dp \geq \int_0^q F_\nu^{-1}(p)dp$).

Rk 1 : $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, $\mu \leq_{\text{icx}} \nu \implies \int_{\mathbb{R}} x\mu(dx) \leq \int_{\mathbb{R}} x\nu(dx)$. Therefore,

$$\mu \leq_{\text{cx}} \nu \iff \mu \leq_{\text{icx}} \nu \text{ and } \int_{\mathbb{R}} x\mu(dx) = \int_{\mathbb{R}} x\nu(dx)$$
$$\iff \mu \leq_{\text{icx}} \nu \text{ and } \mu \leq_{\text{dcx}} \nu.$$

Rk 2 : $\mu \leq_{\text{dcx}} \nu \iff \mu^* \leq_{\text{icx}} \nu^*$, where $\mu^*((-\infty, x]) := \mu([-x, +\infty))$

The lattice structure (Kertz & Rösler 1992,2000)

$(\mathcal{P}_1(\mathbb{R}), \leq_{\text{icx}})$ and $(\mathcal{P}_1(\mathbb{R}), \leq_{\text{dcx}})$ are complete lattices. From $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, we can define $\mu \vee_{\text{icx}} \nu$ and $\mu \wedge_{\text{icx}} \nu$ such that

$$\mu \wedge_{\text{icx}} \nu \leq_{\text{icx}} \eta \leq_{\text{icx}} \mu \vee_{\text{icx}} \nu \text{ for } \eta \in \{\mu, \nu\},$$
$$\mu \wedge_{\text{dcx}} \nu \leq_{\text{dcx}} \eta \leq_{\text{icx}} \mu \vee_{\text{dcx}} \nu \text{ for } \eta \in \{\mu, \nu\}.$$

$$\int_{-\infty}^t F_{\mu \vee_{\text{dcx}} \nu}(x)dx = \max\left(\int_{-\infty}^t F_\mu(x)dx, \int_{-\infty}^t F_\nu(x)dx\right)$$
$$\int_{-\infty}^t F_{\mu \wedge_{\text{dcx}} \nu}(x)dx = \text{convex hull of } \min\left(\int_{-\infty}^t F_\mu(x)dx, \int_{-\infty}^t F_\nu(x)dx\right)$$
$$\int_t^{+\infty} \bar{F}_{\mu \vee_{\text{icx}} \nu}(x)dx = \max\left(\int_t^{+\infty} \bar{F}_\mu(x)dx, \int_t^{+\infty} \bar{F}_\nu(x)dx\right)$$
$$\int_t^{+\infty} \bar{F}_{\mu \vee_{\text{icx}} \nu}(x)dx = \text{convex hull of } \min\left(\int_t^{+\infty} \bar{F}_\mu(x)dx, \int_t^{+\infty} \bar{F}_\nu(x)dx\right).$$

Rk : We also have $\int_0^q F^{-1}_{\mu \wedge_{\text{dcx}} \nu}(p)dp = \min(\int_0^q F_\mu^{-1}(p)dp, \int_0^q F_\nu^{-1}(p)dp)$.

Application to sampling in the convex order

We define for $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$:

$$\mu \wedge \nu = 1_{\{\int_{\mathbb{R}} x\mu(dx) \leq \int_{\mathbb{R}} x\nu(dx)\}} \mu \wedge_{\text{dcx}} \nu + 1_{\{\int_{\mathbb{R}} x\mu(dx) > \int_{\mathbb{R}} x\nu(dx)\}} \mu \wedge_{\text{icx}} \nu,$$
$$\mu \vee \nu = 1_{\{\int_{\mathbb{R}} x\mu(dx) \leq \int_{\mathbb{R}} x\nu(dx)\}} \mu \vee_{\text{dcx}} \nu + 1_{\{\int_{\mathbb{R}} x\mu(dx) > \int_{\mathbb{R}} x\nu(dx)\}} \mu \vee_{\text{icx}} \nu,$$

so that $\mu \leq_{\text{cx}} \mu \vee \nu$ and $\mu \wedge \nu \leq_{\text{cx}} \nu$. When μ and ν are discrete with finite support, these measures can be calculated explicitly (Andrew's monotone chain convex hull algorithm).

Proposition 3

$(X_i)_{i\geq 1}$ i.i.d. $\sim \mu$. $(Y_j)_{j\geq 1}$ i.i.d. $\sim \nu$. $\mu_I = \frac{1}{I}\delta_{X_i}$, $\nu_J = \frac{1}{J}\delta_{Y_j}$ As $I, J \to \infty$, μ_I and $\mu_I \vee \nu_J$ (resp. $\mu_I \wedge \nu_J$ and ν_J) converges a.s. weakly to μ and ν.

Rk : Easy extension to the multi-marginal case.

The algorithm

Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. $(X_i)_{i\geq 1}$ i.i.d. $\sim \mu$. $(Y_j)_{j\geq 1}$ i.i.d. $\sim \nu$. $\mu_I = \frac{1}{I}\delta_{X_i}$, $\nu_J = \frac{1}{J}\delta_{Y_j}$. We consider the following minimization problem :

$$\begin{cases} \text{minimize } \frac{1}{I}\sum_{i=1}^{I} \left| X_i - \sum_{j=1}^{J} q_{ij} Y_j \right|^2 \\ \text{under the constraints } \forall i,j,\ q_{ij} \geq 0,\ \forall i,\ \sum_{j=1}^{J} q_{ij} = 1 \text{ and } \forall j,\ \sum_{i=1}^{I} q_{ij} = \frac{I}{J}. \end{cases}$$

This is a quadratic minimization under linear constraints :

- There exists a minimizer $q_\star$.
- $\mu_I \curlywedge_2 \nu_J = \frac{1}{I}\sum_{i=1}^{I} \delta_{\tilde{X}_i}$, with $\tilde{X}_i = \sum_{j=1}^{J} (q_\star)_{ij} Y_j$ is uniquely defined, and satisfy
$$\mu_I \curlywedge_2 \nu_J \leq_{cx} \nu_J.$$
- Efficient solvers already exist.

Generalization of the problem : Wasserstein projection

First generalization : For a Markov kernel $Q(x,dy)$, we set $m_Q(x) = \int_{\mathbb{R}^d} yQ(x,dy)$. For $\rho \geq 1$ and $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$,

$$\begin{cases} \text{Minimize } \mathcal{J}_\rho(Q) := \int_{\mathbb{R}^d} |x - m_Q(x)|^\rho \mu(dx) \\ \text{under the constraint that } Q \text{ is a Markov kernel such that } \mu Q = \nu \end{cases}.$$

Second generalization : Let $\Pi(\mu,\eta) = \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \int_{y\in\mathbb{R}^d} \pi(dx,dy) = \mu(dx), \int_{x\in\mathbb{R}^d} \pi(dx,dy) = \eta(dy)$.

$$W_\rho(\mu,\eta) = \left(\inf_{\pi\in\Pi(\mu,\eta)} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^\rho \pi(dx,dy) \right)^{1/\rho}.$$

Under suitable conditions (μ absolutely continuous), $\exists T : \mathbb{R}^d \to \mathbb{R}^d$, $T\#\mu = \eta$ and $W_\rho^\rho(\mu,\eta) = \int_{\mathbb{R}^d} |x - T(x)|^\rho \mu(dx)$.

Minimize $W_\rho^\rho(\mu,\eta)$ under the constraint $\eta \leq_{\mathrm{cx}} \nu \to \mu \curlywedge_\rho \nu := \eta^\star$.

Main result & definition of $\mu \curlywedge_\rho \nu$

Proposition 4

Let $\rho \geq 1$, $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$. One has

$$\inf_{Q:\mu Q=\nu} \mathcal{J}_\rho(Q) = \inf_{\eta:\eta \leq_{cx} \nu} W_\rho^\rho(\mu, \eta),$$

where both infima are attained. If $\rho > 1$, then the functions $\{m_{Q_\star} : \mu Q_\star = \nu$ and $\mathcal{J}_\rho(Q_\star) = \inf_{Q:\mu Q=\nu} \mathcal{J}_\rho(Q)\}$ are $\mu(dx)$ a.e. equal, $\mu \curlywedge_\rho \nu := m_{Q_\star} \# \mu = \eta^\star$ is the unique $\eta \leq_{cx} \nu$ minimizing $W_\rho^\rho(\mu, \eta)$ and $\mu(dx)\delta_{m_{Q_\star}(x)}(dy)$ the unique optimal transport plan $\pi \in \Pi(\mu, \mu \curlywedge_\rho \nu)$ such that $W_\rho^\rho(\mu, \mu \curlywedge_\rho \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^\rho \pi(dx, dy)$.

Estimate on the approximation

Proposition 5

Let $\rho \geq 1$, $\mu, \nu, \mu_I, \nu_J \in \mathcal{P}_\rho(\mathbb{R}^d)$ such that $\mu \leq_{\text{cx}} \nu$. Then, we have

$$W_\rho(\mu, \mu_I \curlywedge_\rho \nu_J) \leq 2W_\rho(\mu, \mu_I) + W_\rho(\nu, \nu_J).$$

- For $(X_i)_{i\geq 1}$ i.i.d. $\sim \mu$, $(Y_j)_{j\geq 1}$ i.i.d. $\sim \nu$, $W_\rho(\mu, \mu_I) + W_\rho(\nu, \nu_I) \underset{I\to 0}{\to} 0$.
- Fournier Guillin (2015) : if $\int_{\mathbb{R}^d} e^{\gamma|x|^\alpha}\nu(dx) < \infty$ for some $\alpha > \rho$ and $\gamma > 0$, $W_\rho(\mu, \mu_I \curlywedge_\rho \nu_J) \underset{I,J\to+\infty}{=} \mathcal{O}\left(\left(\frac{\log(I\wedge J)}{I\wedge J}\right)^{\frac{1}{d\vee(2\rho)}}\right)$, *a.s.*.
- Multi-marginal case : for $\mu^1_{I_1}, \ldots, \mu^\ell_{I_\ell}$ approximations of $\mu^1 \leq_{\text{cx}} \cdots \leq_{\text{cx}} \mu^\ell$: $\mu^1_{I_1} \curlywedge_\rho (\ldots(\mu^{\ell-1}_{I_{\ell-1}} \curlywedge \mu^\ell_{I_\ell})) \leq_{\text{cx}} \cdots \leq_{\text{cx}} \mu^\ell_{I_\ell}$ and

$$W_\rho(\mu^k, \mu^k_{I_k}\curlywedge_\rho(\cdots\curlywedge_\rho(\mu^{\ell-1}_{I_{\ell-1}}\curlywedge_\rho\mu^\ell_{I_\ell}))) \leq 2\sum_{k'=k}^{\ell-1} W_\rho(\mu^{k'}, \mu^{k'}_{I_{k'}})+W_\rho(\mu^\ell, \mu^\ell_{I_\ell}).$$

Proof of Proposition 5.

$Q^\rho_{\mu_I}$ (resp. Q^ρ_ν) : Markov kernel such that $\mu_I Q^\rho_{\mu_I} = \mu$ (resp. $\nu Q^\rho_\nu = \nu_J$) is optimal for $W_\rho(\mu_I, \mu)$ (resp. $W_\rho(\nu, \nu_J)$). Let $R(x, dy)$ be a martingale kernel such that $\nu = \mu R$. Since $Q^\rho_{\mu_I} R Q^\rho_\nu$ is a Markov kernel s.t. $\mu_I Q^\rho_{\mu_I} R Q^\rho_\nu = \mu R Q^\rho_\nu = \nu Q^\rho_\nu = \nu_J$ we get

$$
\begin{aligned}
W_\rho(\mu_I, \mu_I \curlywedge_\rho \nu_J) &\leq \mathcal{J}_\rho^{\frac{1}{\rho}}(Q^\rho_{\mu_I} R Q^\rho_\nu) = \left(\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} (x - w + z - y) Q^\rho_{\mu_I}(x, dw) R(w, dz) Q^\rho_\nu(z, dy) \right|^\rho \mu_I(dx) \right)^{\frac{1}{\rho}} \text{ (mg.)} \\
&\leq \left(\int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d} |x - w + z - y|^\rho \, Q^\rho_{\mu_I}(x, dw) R(w, dz) Q^\rho_\nu(z, dy) \mu_I(dx) \right)^{1/\rho} \text{ (Jensen)} \\
&\leq \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - w|^\rho Q^\rho_{\mu_I}(x, dw) \mu_I(dx) \right)^{1/\rho} + \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |z - y|^\rho \nu(dz) Q^\rho_\nu(z, dy) \right)^{1/\rho} \text{ (Minkowski)} \\
&= W_\rho(\mu_I, \mu) + W_\rho(\nu_J, \nu).
\end{aligned}
$$

The claim follows since $W_\rho(\mu, \mu_I \curlywedge_\rho \nu_J) \leq W_\rho(\mu, \mu_I) + W_\rho(\mu_I, \mu_I \curlywedge_\rho \nu_J)$.

Definition of $\mu \curlyvee_\rho \nu$

"Dual" problem :
find $\eta \in \mathcal{P}_\rho(\mathbb{R}^d)$ such that $\mu \leq_{cx} \eta$ that minimizes $W_\rho(\eta, \nu)$.

Proposition 6

For $\rho > 1$, if $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$, then $\inf_{\eta : \mu \leq_{cx} \eta} W_\rho^\rho(\nu, \eta)$ is attained by some probability measure $\mu \curlyvee_\rho \eta$ which is unique when ν is absolutely continuous with respect to the Lebesgue measure or $d = 1$. If $\rho > 1$ and $\mu, \nu, \mu_I, \nu_J \in \mathcal{P}_\rho(\mathbb{R}^d)$, then $\mu \leq_{cx} \nu \Rightarrow W_\rho(\mu_I \curlyvee_\rho \nu_J, \nu) \leq W_\rho(\mu, \mu_I) + 2W_\rho(\nu, \nu_J)$.

Property : $W_\rho(\mu \curlyvee_\rho \nu, \nu) = W_\rho(\mu, \mu \curlywedge_\rho \nu, \nu)$ and there exists an optimal transport map between $\mu \curlyvee_\rho \nu$ and ν.

- $\mu \curlyvee_2 \nu$ a priori less easy to calculate numerically than $\mu \curlywedge_2 \nu$.

Back in dimension 1

Proposition 7

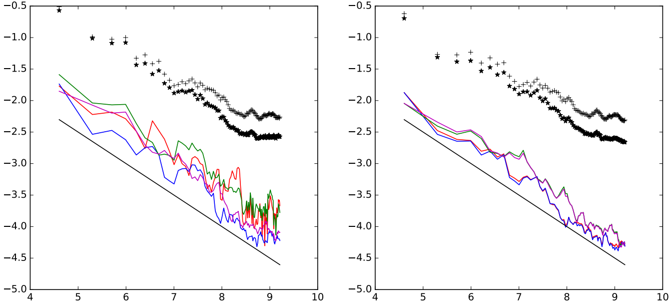
Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. Let ψ denote the convex hull (largest convex function bounded from above by) of the function $[0,1] \ni q \mapsto \int_0^q F_\mu^{-1}(p) - F_\nu^{-1}(p)dp$. Then, there exists probability measures $\mu \curlywedge \nu$ and $\mu \curlyvee \nu$ such that for all $q \in [0,1]$,

$$\int_0^q F_{\mu \curlywedge \nu}^{-1}(p)dp = \int_0^q F_\mu^{-1}(p)dp - \psi(q),$$
$$\int_0^q F_{\mu \curlyvee \nu}^{-1}(p)dp = \int_0^q F_\nu^{-1}(p)dp + \psi(q).$$

Moreover, $\mu \curlyvee_\rho \nu = \mu \curlyvee \nu$ for each $\rho > 1$ such that $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$.

Rk : For discrete probability measures μ and ν, ψ (and thus $\mu \curlywedge \nu$ and $\mu \curlyvee \nu$) can be calculated explicitly.

Convergence in Wasserstein distance W_2



$\log(W_2)$ in function of $\log(I)$. Right : $W_2(\mu, \mu_I \wedge \nu_I)$, $W_2(\mu, \mu_I \curlywedge \nu_I)$, $W_2(\nu, \mu_I \vee \nu_I)$ and $W_2(\nu, \mu_I \curlyvee \nu_I)$, Left : same with “tilde” measures.
$\mu = \mathcal{N}(0,1)$, $\nu = \mathcal{N}(0,1.1)$, $\mu_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i}$, $\nu_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{Y_i}$, $\bar{X}_I = \frac{1}{I}\sum_{i=1}^{I} X_i$, $\bar{Y}_I = \frac{1}{I}\sum_{i=1}^{I} Y_i$, $\tilde{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i - \bar{X}_I}$ and $\tilde{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{Y_i - \bar{Y}_I}$

An example with explicit MOT

Let $\rho > 2$, $\mu(dx) = \frac{1}{2}1_{[-1,1]}(x)dx$ and $\nu(dx) = \frac{1}{4}1_{[-2,2]}(x)dx$. We consider the following MOT problem :

$$\min_{\pi \in \Pi^M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |y-x|^\rho \pi(dx,dy).$$

- For any $\pi \in \Pi^M(\mu,\nu)$, we have $\int_{\mathbb{R}\times\mathbb{R}} |y-x|^2\pi(dx,dy) = \int_{\mathbb{R}} y^2\nu(dy) - \int_{\mathbb{R}} x^2\mu(dx) = 1$. For $\rho > 2$, Jensen's inequality gives

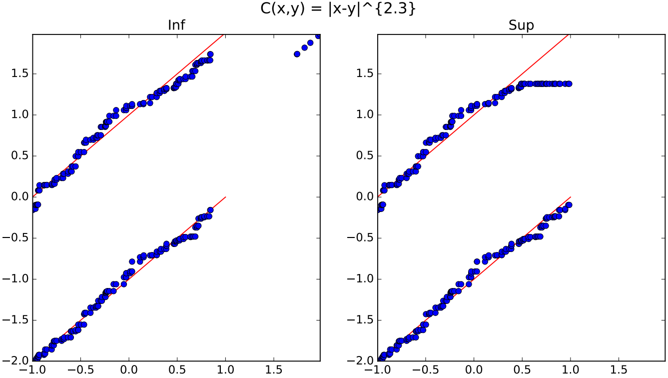
$$\int_{\mathbb{R}\times\mathbb{R}} |y-x|^\rho \pi(dx,dy) \geq \left(\int_{\mathbb{R}\times\mathbb{R}} |y-x|^2\pi(dx,dy)\right)^{\frac{\rho}{2}} = 1.$$

- We observe that

$$\pi^\star(dx,dy) = \frac{1}{2}1_{[-1,1]}(x)\frac{\delta_{x+1}(dy)+\delta_{x-1}(dy)}{2}dx$$

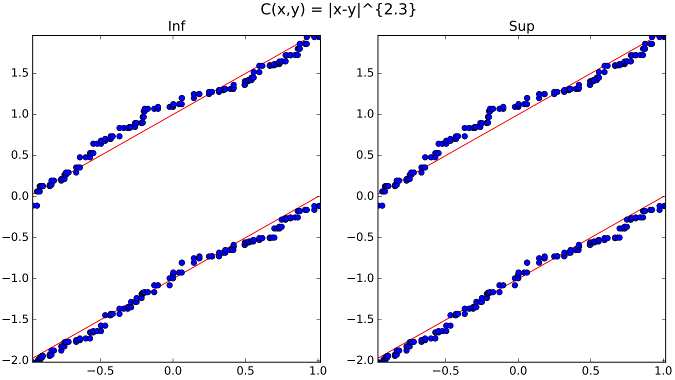
achieves this lower bound.

MOT for $(\mu_I \wedge \nu_I, \nu_I)$ (left) and $(\mu_I, \mu_I \vee \nu_I)$ (right)



$I = 100$. $\mu_I = \frac{1}{I}\sum_{i=1}^{I}\delta_{X_i}$, $\nu_I = \frac{1}{I}\sum_{i=1}^{I}\delta_{Y_i}$. Points with positive probability in the optimal coupling.

MOT for $(\tilde{\mu}_I \wedge \tilde{\nu}_I, \tilde{\nu}_I)$ (left) and $(\tilde{\mu}_I, \tilde{\mu}_I \vee \tilde{\nu}_I)$ (right)



$I = 100$. $\bar{X}_I = \frac{1}{I}\sum_{i=1}^{I} X_i$, $\bar{Y}_I = \frac{1}{I}\sum_{i=1}^{I} Y_i$, $\tilde{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i - \bar{X}_I}$ and $\tilde{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{Y_i - \bar{Y}_I}$

Comparison on the costs

$I = 100$. We have run 100 times the discrete MOT and calculated means and standard deviations of the cost (Optimal cost 1 for the continous MOT) :

	$(\mu_I \wedge \nu_I, \nu_I)$	$(\mu_I, \mu_I \vee \nu_I)$	$(\tilde{\mu}_I \wedge \tilde{\nu}_I, \tilde{\nu}_I)$	$(\tilde{\mu}_I, \tilde{\mu}_I \vee \tilde{\nu}_I)$
mean	0.7506	0.7319	1.002	1.002
std. dev.	0.2148	0.2148	0.14	0.14

- Few differences between $(\mu_I \wedge \nu_I, \nu_I)$ and $(\mu_I, \mu_I \vee \nu_I)$.
- Equalizing the means really improves the approximation.

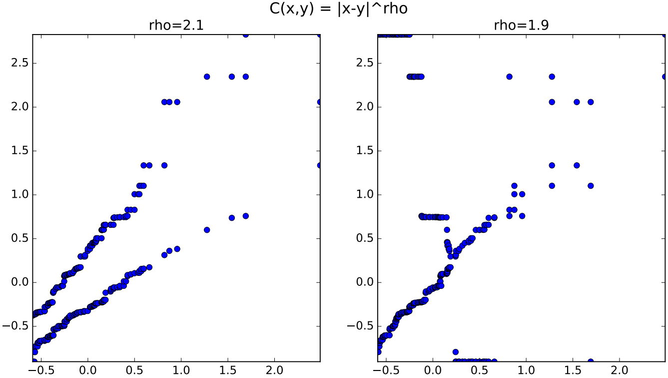
Another numerical example

We consider again the following MOT problem :

$$\min_{\pi \in \Pi^M(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |y-x|^\rho \pi(dx,dy),$$

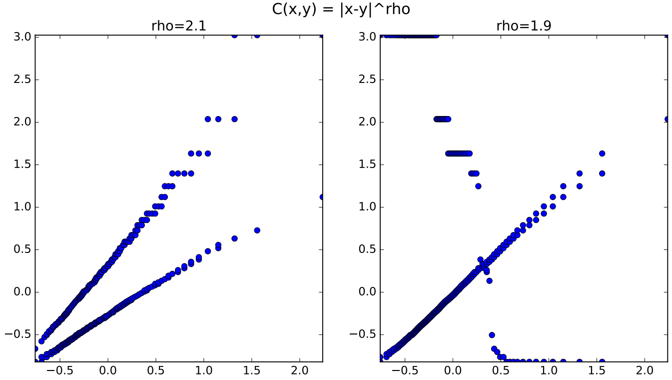
with μ (resp. ν) being the law of $\exp\left(\sigma_X G - \frac{1}{2}\sigma_X^2\right) - 1$ (resp. $\exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$), with $G \sim \mathcal{N}(0,1)$, $\sigma_X = 0.24$ and $\sigma_Y = 0.28$.

MOT for $(\tilde{\mu}_I \wedge \tilde{\nu}_I, \tilde{\nu}_I)$ (left $\rho = 2.1$, right $\rho = 1.9$)



$I = 100$. $\bar{X}_I = \frac{1}{I}\sum_{i=1}^{I} X_i$, $\bar{Y}_I = \frac{1}{I}\sum_{i=1}^{I} Y_i$, $\tilde{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{X_i - \bar{X}_I}$ and $\tilde{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{Y_i - \bar{Y}_I}$

MOT for $(\hat{\mu}_{I'} \wedge \hat{\nu}_{I'}, \hat{\nu}_{I'})$ (left $\rho = 2.1$, right $\rho = 1.9$)



Baker : $\hat{\mu}_{I'} = \frac{1}{I'} \sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\tilde{\mu}_I \wedge \tilde{\nu}_I}(u)du}$, $\hat{\nu}_{I'} = \frac{1}{I'} \sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\tilde{\nu}_I}(u)du}$.

$I' = 100$, $I = 10000$.

Comparison of the std deviation on the first example

On 100 independent runs.

- $I = 100$ $(\tilde{\mu}_I \wedge \tilde{\nu}_I, \tilde{\nu}_I)$. Mean : 1.002, Std. dev. : 0.14.
- $\hat{\mu}_{I'} = \frac{1}{I'} \sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\tilde{\mu}_I \wedge \tilde{\nu}_I}(u)du}$, $\hat{\nu}_{I'} = \frac{1}{I'} \sum_{i=1}^{I'} \delta_{I' \int_{\frac{i-1}{I'}}^{\frac{i}{I'}} F^{-1}_{\tilde{\nu}_I}(u)du}$, with $I' = 100$, $I = 10000$. Mean : 0.9981, Std dev. : 0.0148.

An example with three marginals

- Laws : $X_i \overset{(d)}{=} \exp\left(\sigma_X G - \frac{1}{2}\sigma_X^2\right) - 1$, $Y_i \overset{(d)}{=} \exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$ and $Z_i \overset{(d)}{=} \exp\left(\sigma_Y G - \frac{1}{2}\sigma_Y^2\right) - 1$, with $G \sim \mathcal{N}(0,1)$, $\sigma_X = 0.24$, $\sigma_Y = 0.28$, $\sigma_Z = 0.32$.
- Payoff/cost function : $c(x,y,z) = (z - \frac{x+y}{2})^+$.
- BS price$\approx$ 0.0681, lower bound : 0.0303, upper bound 0.0856 obtained with $(\hat{\mu}_{I'}, \hat{\nu}_{I'}, \hat{\eta}_{I'})$ ($I = 2500$ and $I' = 25$).
- Minimize/maximize

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk}\, c(x_i, y_j, z_k)$$

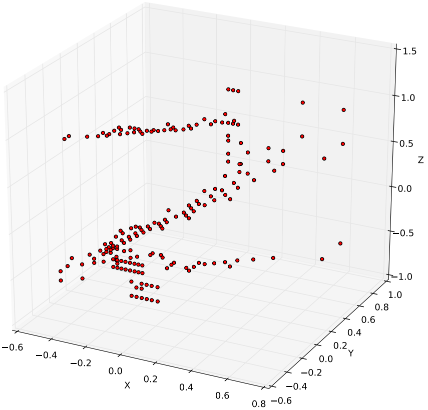
under the constraints

$$\forall i,j,k,\ r_{ijk} \geq 0,\ \forall i, \sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk} = p_i,\ \forall j, \sum_{i=1}^{I}\sum_{k=1}^{K} r_{ijk} = q_j,\ \forall k, \sum_{i=1}^{I}\sum_{j=1}^{J} r_{ijk} = s_k,$$

$$\forall i, \sum_{j=1}^{J}\sum_{k=1}^{K} r_{ijk}(y_j - x_i) = 0,\ \forall i,j, \sum_{k=1}^{K} r_{ijk}(z_k - y_j) = 0.$$

with $\mu = \sum_{i=1}^{I} p_i \delta_{x_i}$, $\nu = \sum_{j=1}^{J} q_j \delta_{y_j}$ and $\eta = \sum_{k=1}^{K} s_k \delta_{z_k}$ satisfying $\mu \leq_{\text{cx}} \nu \leq_{\text{cx}} \eta$.

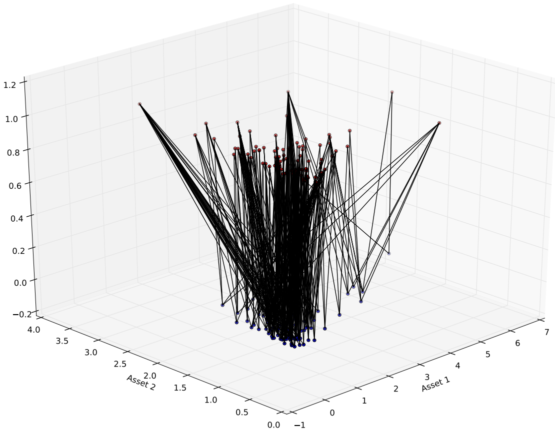
Martingale Optimal transport for $(\hat{\mu}_{I'}, \hat{\nu}_{I'}, \hat{\eta}_{I'})$ (min)



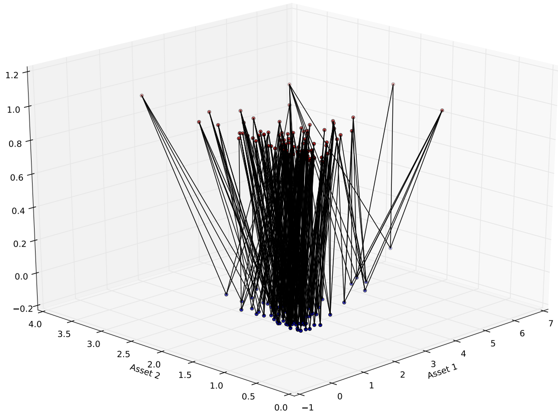
An example in dimension 2

- (G^1, G^2) : centered Gaussian vector with covariance matrix $\Sigma = \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}$.
- μ : law of (X^1, X^2) with $X^\ell = \exp(G^\ell - \Sigma_{\ell\ell}/2)$, $\ell \in \{1, 2\}$.
- ν : law of (Y^1, Y^2) with $Y^\ell = \exp(\sqrt{2}G^\ell - \Sigma_{\ell\ell})$, $\ell \in \{1, 2\}$.
- Payoff/cost function : $\max(Y^1 - X^1, Y^2 - X^2, 0)$ (best performance if positive).
- $\tilde{\mu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{(X_i^1+1-\bar{X}_I^1, X_i^2+1-\bar{X}_I^2)}$, $\tilde{\nu}_I = \frac{1}{I}\sum_{i=1}^{I} \delta_{(Y_i^1+1-\bar{Y}_I^1, Y_i^2+1-\bar{Y}_I^2)}$
- BS price$\approx$ 0.345, lower bound (on 100 indep runs) : mean 0.2293 (std. dev 0.0848), upper bound mean 0.4111 (std. dev 0.1422), obtained with $(\tilde{\mu}_I \curlywedge_2 \tilde{\nu}_I, \tilde{\nu}_I)$, $I = 100$.

Martingale Optimal transport for $(\tilde{\mu}_I \curlywedge_2 \tilde{\nu}_I, \tilde{\nu}_I)$ (min)



Martingale Optimal transport for $(\tilde{\mu}_I \curlywedge_2 \tilde{\nu}_I, \tilde{\nu}_I)$ (max)



Conclusion

- The methods that we have presented, enable to calculate with a MC method at the same time option prices, and their bounds on all other models sharing the same marginal laws.
- The accuracy of the price bounds (maybe not so important in practice) is limited by the dimension of the linear programming problem.
- A possible direction is to develop approximated linear programming solvers (Benamou, Carlier, Cuturi, Nenna, 2015 in the OT case).