OT AND DATA DRIVEN METHODS: THEORY AND PRACTICE

(FROM MATHEMATICAL FINANCE AND STATISTICS)

Jan Obłój
Mathematical Institute
University of Oxford

joint works with
Daniel Bartl, Samuel Drapeau, Stephan Eckstein,
Gaoyue Guo, Tongseok Lim and Johannes Wiesel

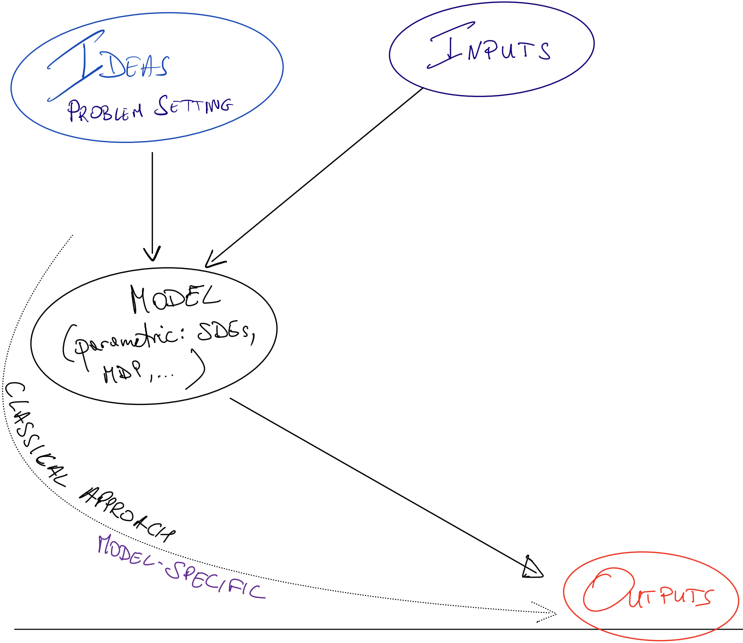
Kantorovich Initiative Seminar

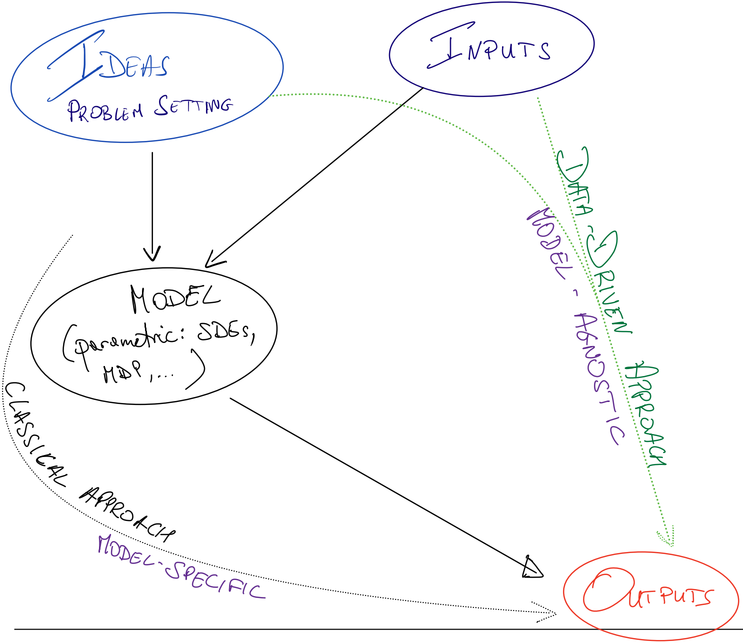
Oxford Mathematics

St John's College Oxford

University of Oxford Mathematical Institute

European Research Council Established by the European Commission


IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs,
MDP, ...)
CLASSICAL APPROACH
MODEL-SPECIFIC
OUTPUTS


IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs, MDP, ...)
CLASSICAL APPROACH
MODEL-SPECIFIC
DATA-DRIVEN APPROACH
MODEL-AGNOSTIC
OUTPUTS

Copulas vs OT

UNIVERSITY OF OXFORD
Mathematical Institute

Sklar's Theorem: d-dim df $=$ marginals $\oplus$ copula.

How to compute $\mathbb{E}[\xi(X, Y)]$?

Parametric Approach

Fix a copula C.
Estimate the marginals of X and Y.
Compute

$$\mathbb{E}[\xi(X, Y)] = \iint c(x, y) dF$$

where $F(x, y) = C(F_X(x), F_Y(y))$.

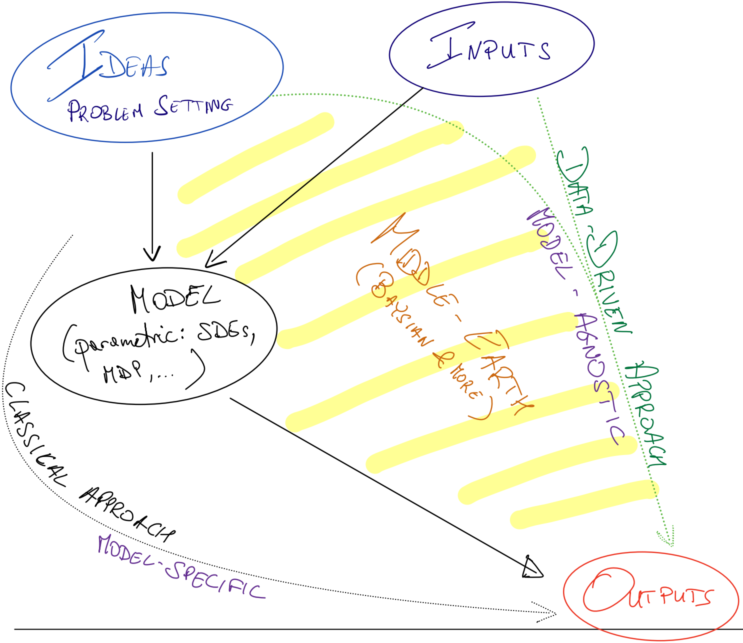
Non-parametric Approach

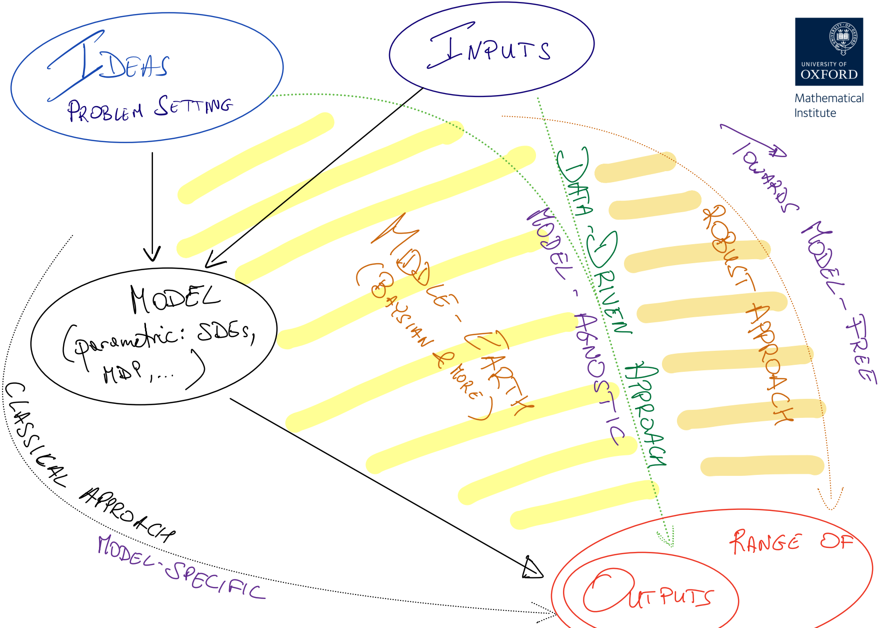
Estimate the marginals of X and Y.
Compute

$$\inf_{\pi \in \Pi(F_X, F_Y)} \iint c(x, y) d\pi$$

where $\Pi(F_X, F_Y)$...

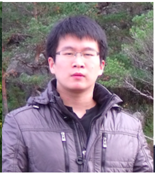
UNIVERSITY OF OXFORD
Mathematical Institute





IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs, MDP, ...)
CLASSICAL APPROACH
MODEL-SPECIFIC
MIDDLE-EARTH
(BAYESIAN & MORE)
MODEL - AGNOSTIC
DATA-DRIVEN APPROACH
ROBUST APPROACH
TOWARDS MODEL-FREE
RANGE OF
OUTPUTS

Data: Market Prices of Options


based on joint works with Stephan Eckstein, Gaoyue Guo, Tongseok Lim
see *SIAM J. Financial Math. (2021), Ann. App. Probab. (2019)*.

In the market I can expect to see prices of (many) European options.
What do I do with this information?

Mathematical Institute

In the market I can expect to see prices of (many) European options. What do I do with this information?

- Model specific: we typically consider $\{\mathbb{P}_\theta : \theta \in \Theta\}$ and use option prices to calibrate a particular $\mathbb{P}_{\theta^*}$.

In the market I can expect to see prices of (many) European options. What do I do with this information?

- Model specific: we typically consider $\{\mathbb{P}_\theta : \theta \in \Theta\}$ and use option prices to calibrate a particular $\mathbb{P}_{\theta^*}$.
- Robust approach: add these as inputs/trading instruments to lower the superhedging price
  via duality $\rightsquigarrow$ constraints on pricing measures

An (idealised) case study: the MOT problem

UNIVERSITY OF OXFORD
Mathematical Institute

- suppose you observe prices of call options:

$$\text{Price}((S_T - K)^+) = C(K), \quad K \in \mathbb{R}.$$

see Hobson '98, Breeden & Litzenberger '78.

An (idealised) case study: the MOT problem

UNIVERSITY OF OXFORD
Mathematical Institute

- suppose you observe prices of call options:

$$\mathrm{Price}((S_T - K)^+) = C(K), \quad K \in \mathbb{R}.$$

see Hobson '98, Breeden & Litzenberger '78.

- feasible pricing model $\leftrightsquigarrow$ probability measure $\mathbb{Q}$ s.t.

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+] = C(K),\ K \in \mathbb{R},$$

which is equivalent to

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad S_T \sim_{\mathbb{Q}} \nu,\ \text{for } \nu(dK) = C''(dK).$$

An (idealised) case study: the MOT problem

University of Oxford Mathematical Institute

- suppose you observe prices of call options:

$$\text{Price}((S_T - K)^+) = C(K), \quad K \in \mathbb{R}.$$

see Hobson '98, Breeden & Litzenberger '78.

- feasible pricing model $\leftrightsquigarrow$ probability measure $\mathbb{Q}$ s.t.

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+] = C(K), \; K \in \mathbb{R},$$

which is equivalent to

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad S_T \sim_{\mathbb{Q}} \nu, \text{ for } \nu(dK) = C''(dK).$$

- Robust pricing of an exotic option with payoff ξ
$\rightsquigarrow$ sup $\mathbb{E}_{\mathbb{Q}}[\xi(S_t : t \leq T)]$ over such $\mathbb{Q}$s.
Robust hedging is its dual problem.

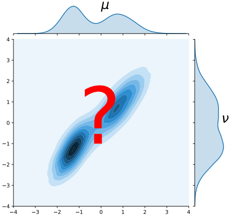
The MOT problem

Mathematical Institute

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\mathrm{P}(\mu,\nu) \quad := \quad \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_{\mathbb{Q}}\left[\xi(S_1,S_2)\right],$$

where
$\mathcal{M}(\mu,\nu) := \left\{\mathbb{Q}\in\mathcal{P}(\mathbb{R}^{2d}) : S_1\sim\mu,\ S_2\sim\nu \text{ and } \mathbb{E}_{\mathbb{Q}}\left[S_2\middle|S_1\right] = S_1\right\}.$

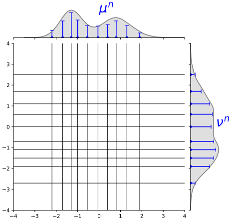


MOT Numerics: take I (WITH GAOYUE GUO)

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\mathrm{P}(\mu,\nu) \quad := \quad \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_{\mathbb{Q}}\big[\xi(S_1,S_2)\big],$$

where
$\mathcal{M}(\mu,\nu) := \left\{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^{2d}) : S_1 \sim \mu,\ S_2 \sim \nu \text{ and } \mathbb{E}_{\mathbb{Q}}\big[S_2 \big| S_1\big] = S_1\right\}$.



MOT Numerics: take I (WITH GAOYUE GUO)

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\mathrm{P}(\mu,\nu) \quad := \quad \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_{\mathbb{Q}}\big[\xi(S_1,S_2)\big],$$

where
$\mathcal{M}(\mu,\nu) := \big\{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^{2d}) : S_1 \sim \mu,\ S_2 \sim \nu \text{ and } \mathbb{E}_{\mathbb{Q}}\big[S_2 \big| S_1\big] = S_1\big\}$.

- If $\mu = \sum_{i=1}^m \alpha_i \delta_{x_i}(dx)$ and $\nu = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, then $\mathrm{P}(\mu,\nu)$ is an LP problem;
- Discretisation $(\mu,\nu) \rightsquigarrow (\mu^n,\nu^n)$ typically does NOT preserve the convex order, see Alfonsi et al. (2017).

MOT Numerics: take I (with Gaoyue Guo)

Mathematical Institute

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\mathrm{P}(\mu,\nu) \quad := \quad \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_{\mathbb{Q}}\big[\xi(S_1,S_2)\big],$$

where

$\mathcal{M}(\mu,\nu) := \big\{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^{2d}) : S_1 \sim \mu,\ \ S_2 \sim \nu$ and $\mathbb{E}_{\mathbb{Q}}\big[S_2\big|S_1\big] = S_1\big\}$.

- If $\mu = \sum_{i=1}^m \alpha_i \delta_{x_i}(dx)$ and $\nu = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, then $\mathrm{P}(\mu,\nu)$ is an LP problem;
- Discretisation $(\mu,\nu) \rightsquigarrow (\mu^n,\nu^n)$ typically does NOT preserve the convex order, see Alfonsi et al. (2017).
- Further, continuity of $(\mu,\nu) \to \mathrm{P}(\mu,\nu)$ is a hard problem.

MOT Numerics: take I (WITH GAOYUE GUO)

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\mathrm{P}(\mu,\nu) \quad := \quad \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \quad \mathbb{E}_{\mathbb{Q}}\left[\xi(S_1, S_2)\right],$$

where
$\mathcal{M}(\mu,\nu) := \left\{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^{2d}) : S_1 \sim \mu,\ \ S_2 \sim \nu \text{ and } \mathbb{E}_{\mathbb{Q}}\left[S_2 \middle| S_1\right] = S_1\right\}$.

- If $\mu = \sum_{i=1}^m \alpha_i \delta_{x_i}(dx)$ and $\nu = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, then $\mathrm{P}(\mu,\nu)$ is an LP problem;
- Discretisation $(\mu,\nu) \rightsquigarrow (\mu^n,\nu^n)$ typically does NOT preserve the convex order, see Alfonsi et al. (2017).
- Further, continuity of $(\mu,\nu) \to \mathrm{P}(\mu,\nu)$ is a hard problem.
- $\rightsquigarrow$ we propose to look at a suitable relaxation!

MOT Numerics: take I

Consider

$$
\begin{aligned}
\mathrm{P}_\varepsilon(\mu,\nu) &:= \sup_{\mathbb{Q}\in\mathcal{M}_\varepsilon(\mu,\nu)} \mathbb{E}_{\mathbb{Q}}\big[\xi(S_1,S_2)\big],\\
\mathcal{M}_\varepsilon(\mu,\nu) &:= \Big\{\mathbb{Q} : S_1\sim\mu,\ S_2\sim\nu \text{ and } \mathbb{E}_{\mathbb{Q}}\Big[\Big|\mathbb{E}_{\mathbb{Q}}\big[S_2\big|S_1\big]-S_1\Big|\Big]\le\varepsilon\Big\}.
\end{aligned}
$$

MOT Numerics: take I

UNIVERSITY OF OXFORD
Mathematical Institute

Consider

$$P_\varepsilon(\mu,\nu) \quad := \quad \sup_{\mathbb{Q}\in\mathcal{M}_\varepsilon(\mu,\nu)} \mathbb{E}_{\mathbb{Q}}\left[\xi(S_1,S_2)\right],$$

$$\mathcal{M}_\varepsilon(\mu,\nu) \quad := \quad \left\{\mathbb{Q} : S_1 \sim \mu,\ S_2 \sim \nu \text{ and } \mathbb{E}_{\mathbb{Q}}\left[\left|\mathbb{E}_{\mathbb{Q}}\left[S_2|S_1\right] - S_1\right|\right] \le \varepsilon\right\}.$$

Theorem

Assume $\mu \preceq \nu$ are in convex order and ξ is L-Lipschitz. Let $(\mu^n,\nu^n)_{n\geq 1}$ be a sequence converging to (μ,ν): $r_n := \mathcal{W}(\mu^n,\mu) + \mathcal{W}(\nu^n,\nu) \to 0$. Then,

$$\mathcal{M}_{r_n}(\mu^n,\nu^n) \neq \emptyset \quad \textit{and} \quad \lim_{n\to\infty} P_{r_n}(\mu^n,\nu^n) = P(\mu,\nu).$$

How do you actually discretise a measure μ?

If you can integrate against μ (or know the density)

- restrict to a ball of radius R,
- discretise on a lattice pulling mass on a cube to its corner,
- assuming $\theta > 1$ moment, gives $r_n \leq \frac{\theta}{\theta - 1} \frac{d}{n}$.
- In practice use point estimates of the density $\rightsquigarrow r_n \leq \text{const} \frac{L}{n^{1/(d+1)}}$.

How do you actually discretise a measure μ?

If you can integrate against μ (or know the density)

- restrict to a ball of radius R,
- discretise on a lattice pulling mass on a cube to its corner,
- assuming $\theta > 1$ moment, gives $r_n \leq \frac{\theta}{\theta-1}\frac{d}{n}$.
- In practice use point estimates of the density $\rightsquigarrow$ $r_n \leq \text{const}\frac{L}{n^{1/(d+1)}}$.

If you can simulate from μ

- let $\hat{\mu}_n = \frac{1}{n}\sum \delta_{X_i}$ be the empirical measure,
- take $\varepsilon_m \searrow 0$ with $\sum_{m\geq 1}\mathbb{E}\left[\mathcal{W}(\hat{\mu}_{n_m},\mu)+\mathcal{W}(\hat{\nu}_{n_m},\nu)\right]/\varepsilon_m < \infty$, then $\lim_{m\to\infty} \mathsf{P}_{\varepsilon_m}(\hat{\mu}_{n_m},\hat{\nu}_{n_m}) = \mathsf{P}(\mu,\nu)$ a.s.,
- use cnv rate in the Glivenko-Cantelli (Fournier & Guillin '15) + compute explicitly their constants.

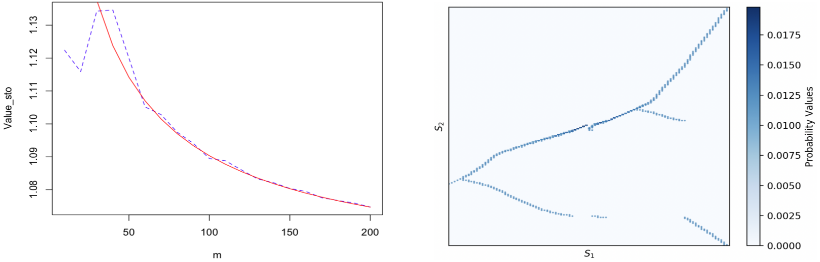


Figure: The first pane shows the convergence of $\mathrm{P}_{\varepsilon_m}(\hat{\mu}_n, \hat{\nu}_n)$ with respect to n for $m = 100$. The second pane draws the heat map of the optimiser for $n = 200$.

Further results

UNIVERSITY OF OXFORD
Mathematical Institute

- Results/methods extend to T-periods.
- For $T = 2$, $d = 1$:
  - bespoke discretisation
  - convergence rates
  - entropic regularisation + iterative Bregman projection method $\rightsquigarrow$ efficient numerics.
- BUT: quickly becomes infeasible: LP has n^{Td} parameters!
- *see also the works of Benjamin Jourdin and co-authors*.

MOT Numerics: take II (Eckstein & Kupper '17)

- Numerics on the dual (superhedging) problem

MOT Numerics: take II (Eckstein & Kupper '17)

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions

MOT Numerics: take II (Eckstein & Kupper '17)

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions
- $\rightsquigarrow$ Deep Neural Network implementation

MOT Numerics: take II (Eckstein & Kupper '17)

Mathematical Institute

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions
- $\rightsquigarrow$ Deep Neural Network implementation
  - hedging strategies $\in \mathcal{H}_n$ (a deep NN)

MOT Numerics: take II (Eckstein & Kupper '17)

Mathematical Institute

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions
- $\rightsquigarrow$ Deep Neural Network implementation
  - hedging strategies $\in \mathcal{H}_n$ (a deep NN)
  - superhedging "$\leq$" replaced by a smooth penalisation w.r.t. a reference measure allowing for gradient descent algorithms:

$$(D^m_{\theta,\gamma}) = \inf_{h \in \mathcal{H}^m} \varphi(h) + \int \beta_\gamma(\xi - h) d\theta$$

MOT Numerics: take II (Eckstein & Kupper '17)

Mathematical Institute

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions
- $\rightsquigarrow$ Deep Neural Network implementation
  - hedging strategies $\in \mathcal{H}_n$ (a deep NN)
  - superhedging "$\leq$" replaced by a smooth penalisation w.r.t. a reference measure allowing for gradient descent algorithms:

$$(D^m_{\theta,\gamma}) = \inf_{h \in \mathcal{H}^m} \varphi(h) + \int \beta_\gamma(\xi - h) d\theta$$

  - Dual optimiser $\hat{h}$ allows to recover the primal one $\hat{\mathbb{Q}}$ via

$$\frac{d\hat{\mathbb{Q}}}{d\theta} = \beta'_\gamma(\xi - \hat{h})$$

is an optimiser of $(P_{\theta,\gamma})$.

Market data: reality check

Mathematical Institute

- For $d > 1$ we do NOT have full marginals.
  Only marginals of marginals (the MMOT problem):

$$S_1^i \sim \mu_i, \quad S_2^i \sim \nu_i$$

Market data: reality check

Mathematical Institute

- For $d > 1$ we do NOT have full marginals.
  Only marginals of marginals (the MMOT problem):

$$S_1^i \sim \mu_i, \quad S_2^i \sim \nu_i$$

- Some interesting cases:
  - $d = 2$, $\xi(S) = (S_T^1 - \alpha S_T^2 - K)^+$ spread options
    $\rightsquigarrow$ *both LP and NN methods work*
  - $d = 30, 50, 100, \ldots, 500$ and $\xi(S) = \left(\sum_{i=1}^d \lambda_i S_T^i - K\right)^+$,
    i.e., calls/puts on an index
    $\rightsquigarrow$ *LP fails, NN work for* $dT \leq 30$ *and then harder, sampling the superhedging condition tricky!*

A case study: MMOT for $d = 2 = T$

Inputs:

- Two assets, two maturities.
- Option prices $\rightsquigarrow \mu_1, \mu_2$ and ν_1, ν_2 with $\mu_i \preceq \nu_i$
- Payoff: $\xi(S) = \xi(S_2^1, S_2^2)$ is a function of what happens at time $T = 2$, e.g., a spread option $\xi = (S_2^1 - S_2^2 - K)^+$.

A case study: MMOT for $d = 2 = T$

Inputs:

- Two assets, two maturities.
- Option prices $\rightsquigarrow \mu_1, \mu_2$ and ν_1, ν_2 with $\mu_i \preceq \nu_i$
- Payoff: $\xi(S) = \xi(S_2^1, S_2^2)$ is a function of what happens at time $T = 2$, e.g., a spread option $\xi = (S_2^1 - S_2^2 - K)^+$.

Beliefs: *minimal correlation between S^1 and S^2*

- PRIMAL: only consider $\mathbb{Q}$ s.t. $\mathrm{corr}(S_2^1, S_2^2) \geq \rho$
- DUAL: allow to sell $S_2^1 S_2^2$ at price $S_0^1 S_0^2 + \rho\sigma_2^1\sigma_2^2$.

A case study: MMOT for $d = 2 = T$

Inputs:

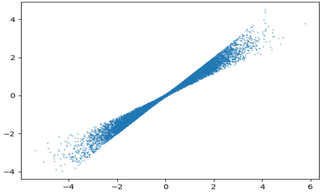
- Two assets, two maturities.
- Option prices $\rightsquigarrow \mu_1, \mu_2$ and ν_1, ν_2 with $\mu_i \preceq \nu_i$
- Payoff: $\xi(S) = \xi(S_2^1, S_2^2)$ is a function of what happens at time $T = 2$, e.g., a spread option $\xi = (S_2^1 - S_2^2 - K)^+$.

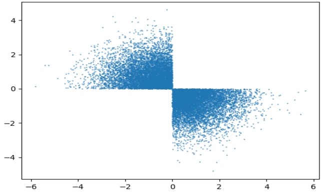
Beliefs: *minimal correlation between S^1 and S^2*

- $\textsc{Primal}$: only consider $\mathbb{Q}$ s.t. $\mathrm{corr}(S_2^1, S_2^2) \geq \rho$
- $\textsc{Dual}$: allow to sell $S_2^1 S_2^2$ at price $S_0^1 S_0^2 + \rho\sigma_2^1\sigma_2^2$.

Benchmarks:

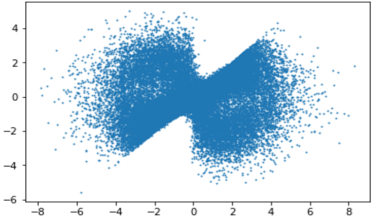
- $\mu_1 = \nu_1$ and $\mu_2 = \nu_2 \Rightarrow$ OT problem!
- Gaussian copula used to construct the join distribution

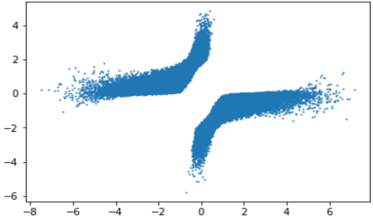




Minimisation: OT

Maximisation: OT





Minimisation: MMOT

Maximisation: MMOT

Problem: Maximise/Minimise $c = (S_2^1 - S_2^2)^+$

S.t.: $\mu_1 = \mathcal{N}(0, 1.8)$, $\mu_2 = \mathcal{N}(0, 0.2)$; $\nu_1 = \mathcal{N}(0, 1.9)$, $\nu_2 = \mathcal{N}(0, 1.3)$.

A Toy Example

INPUTS:

- Data recorded on 16/11/2018:
  - Spot prices $F_0 = 140$, $A_0 = 194$ for Facebook and Apple
  - Call/Puts prices for Facebook and Apple maturing $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$
- Beliefs: bounds on correlation between Facebook and Apple

A Toy Example

University of Oxford
Mathematical Institute

INPUTS:

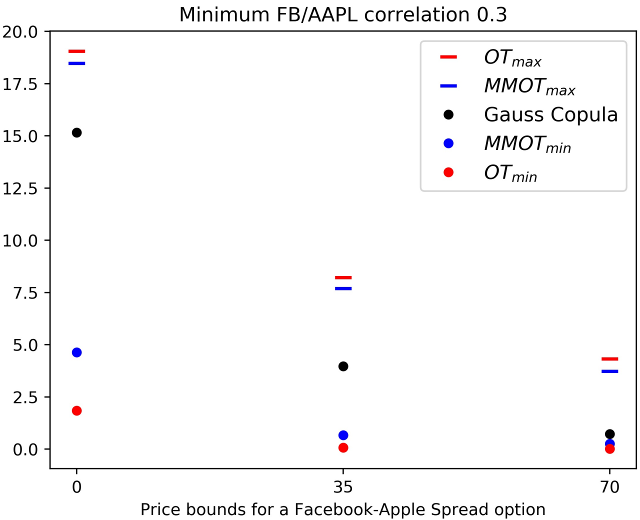
- Data recorded on 16/11/2018:
  - Spot prices $F_0 = 140$, $A_0 = 194$ for Facebook and Apple
  - Call/Puts prices for Facebook and Apple maturing $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$
- Beliefs: bounds on correlation between Facebook and Apple

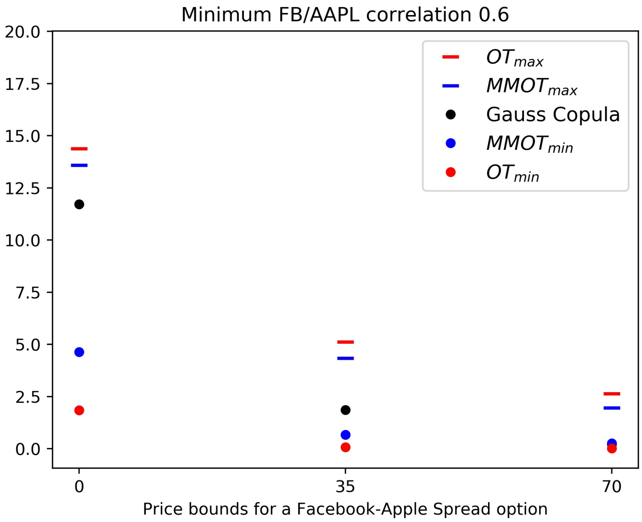
OUTPUTS:

- Range of no-arbitrage prices for a spread option:

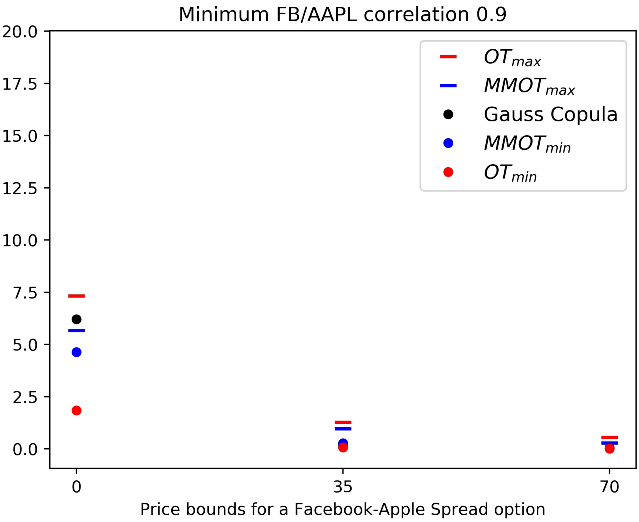
$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K\right)^+, \quad K = 0,\ 35,\ 70.$$

- Distribution of (F_{T_2}, A_{T_2}) for the minimiser/maximiser
- Robust hedging strategies

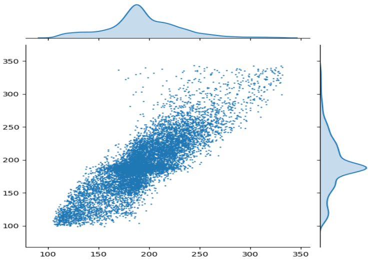

Minimum FB/AAPL correlation 0.3
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
35
70
Price bounds for a Facebook-Apple Spread option
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}

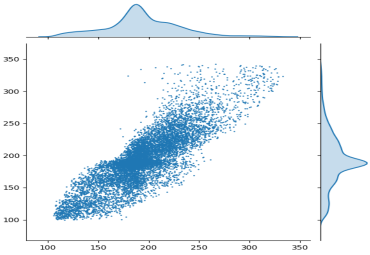


Minimum FB/AAPL correlation 0.6
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}
0
35
70
Price bounds for a Facebook-Apple Spread option


Minimum FB/AAPL correlation 0.9
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
35
70
Price bounds for a Facebook-Apple Spread option
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}

Temporal evolution under Extreme models

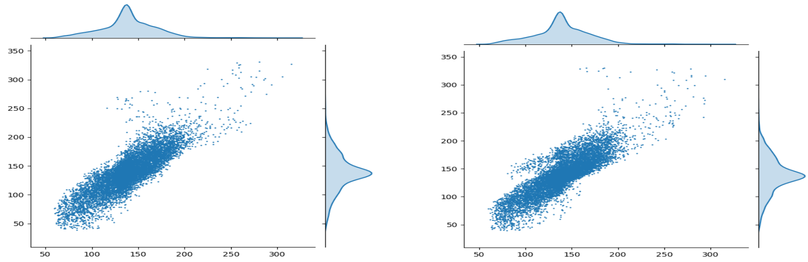




Joint distribution of (A_{T_1}, A_{T_2}), for the Minimiser and Maximiser $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$, $K = 35$ and $\rho \geq 0.6$ and

$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K\right)^+$$

Temporal evolution under Extreme models

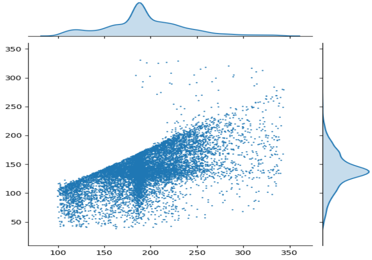


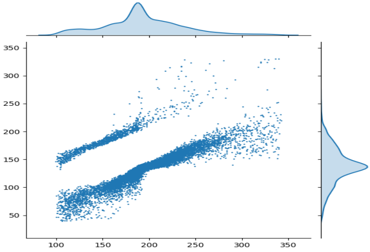
Joint distribution of (F_{T_1}, F_{T_2}), for the Minimiser and Maximiser $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$, $K = 35$ and $\rho \geq 0.6$ and

$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K \right)^+$$

Dependence Structure under Extreme models

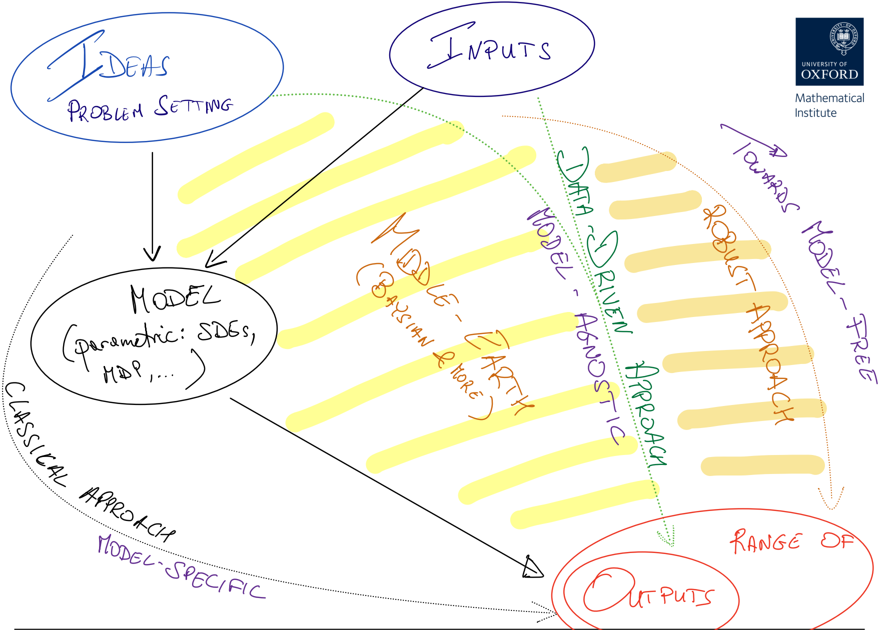
UNIVERSITY OF OXFORD
Mathematical Institute





Joint distribution of (A_{T_2}, F_{T_2}), $T_2 = 21/06/2019$, for the Minimiser and Maximiser for $K = 35$ and $\rho \geq 0.6$ and

$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K \right)^+$$



UNIVERSITY OF OXFORD
Mathematical Institute

Data: Historical Time Series

A Model's neighbourhood & Wasserstein distances

Model neighbourhood

UNIVERSITY OF OXFORD
Mathematical Institute

Measure μ (or $\mathbb{P}$) will denote a model, such as

- $\mu = \hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}$ is the empirical measure of the observations/test set.
- μ comes from a mathematical modelling effort, e.g., an SDE;

Model neighbourhood

Mathematical Institute

Measure μ (or $\mathbb{P}$) will denote a model, such as

- $\mu = \hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{x^i}$ is the empirical measure of the observations/test set.
- μ comes from a mathematical modelling effort, e.g., an SDE;

There are MANY ways to build a neighbourhood $B_\delta(\mu)$ of μ:

- data perturbation
- support estimates
- moments contraints
- density constraints
- Prokhorov distance
- Hellinger distance
- Kullback–Leibler divergence/entropy bounds
- and more...

Wasserstein distance

Mathematical Institute

For $p \geq 1$, $\mu, \nu \in \mathcal{P}(\mathcal{S})$ with p^{th} moments, set

$$W_p(\mu,\nu) = \inf \left\{ \int_{\mathcal{S}\times\mathcal{S}} d(x,y)^p \, \pi(dx,dy) \colon \pi \in \mathrm{Cpl}(\mu,\nu) \right\}^{1/p},$$

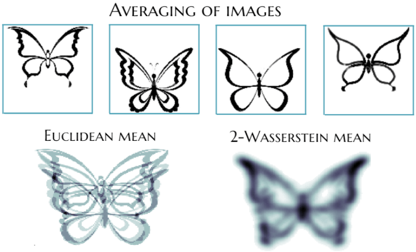
where $\mathrm{Cpl}(\mu,\nu) = \{\pi : \pi(\cdot \times \mathcal{S}) = \mu \text{ and } \pi(\mathcal{S} \times \cdot) = \nu\}$.

$$\text{metric } d \text{ on } \mathcal{S} \quad \Longrightarrow \quad \text{metric } W \text{ on } \mathcal{P}(\mathcal{S})$$

Observe historical returns $r^1, \ldots, r^N$ assumed to follow a time-homogeneous ergodic Markov chain on $\mathbb{R}^d$ with an invariant distribution μ. Should we work with

the data points $(r^i)_{i=1}^N$ or the empirical measure $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{r^i}$?

Source: J. Ebert, V. Spokoiny, A. Suvorikova arXiv:1703.03658



Wasserstein vs Euclidean mean (MNIST data)

University of Oxford
Mathematical Institute

Wasserstein vs Euclidean mean (MNIST data)

Small uncertainty limit

Key property: $\hat{\mu}_N \xrightarrow{W_p} \mu$ + cnv rates, see Fournier & Guillin '14

Esfahani & Kuhn '18 argue that using Wasserstein balls gives

- finite sample guarantees,
- asymptotic consistency,
- tractability (see also Eckstein & Kupper '19)

Large uncertainty limit

Mathematical Institute

Pflug, Pichler & Wozabal '12 use Wasserstein balls for robust portfolio selection:

$$\sup_{a:\langle a,1\rangle=1} \inf_{\nu\in B_\delta(\mu)} \Big(\mathbb{E}_\nu[\langle a, R\rangle] - \gamma \mathrm{Var}_\nu[\langle a, R\rangle]\Big)$$

and show that

$$a^*(\delta) \xrightarrow{\delta\to\infty} \left(\frac{1}{N},\dots,\frac{1}{N}\right)$$

which may not be true for weaker or stronger metrics.

UNIVERSITY OF OXFORD
Mathematical Institute

OT & Data-Driven Approach: risk estimation example

$$(r_1, \ldots, r_N) \in \mathbb{R}^{dN} \quad \text{v.s.} \quad \hat{\mathbb{P}}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{r_i} \in \mathcal{P}(\mathbb{R}^d)$$

based on O. and Wiesel, *Ann. Stat.* **49**(1): 508–530, 2021.

Superhedging with respect to risk measures

Returns $r \sim \mathbb{P}$. We want to build an estimator for

$$\pi^{\mathbb{P}}(\xi) = \inf \left\{ x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \; x + H(r-1) \geq \xi(r) \; \mathbb{P}\text{-a.s.} \right\}$$

Superhedging with respect to risk measures

Returns $r \sim \mathbb{P}$. We want to build an estimator for

$$\pi^{\mathbb{P}}(\xi) = \inf \left\{ x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \; x + H(r-1) \geq \xi(r) \; \mathbb{P}\text{-a.s.} \right\}$$

$$\pi^{\rho_{\mathbb{P}}}(\xi) = \inf \left\{ x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \rho_{\mathbb{P}}(\xi - x - H(r-1)) \leq 0 \right\},$$

where $\rho_{\mathbb{P}}$ is a law-invariant coherent risk measure:

$$\rho_{\mathbb{P}}(\xi) = \inf\{x \in \mathbb{R} \mid \rho_{\mathbb{P}}(\xi - x) \leq 0\}.$$

Superhedging with respect to risk measures

Returns $r \sim \mathbb{P}$. We want to build an estimator for

$$\pi^{\mathbb{P}}(\xi) = \inf\left\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \ x + H(r-1) \geq \xi(r)\ \mathbb{P}\text{-a.s.}\right\}$$
$$\pi^{\rho_{\mathbb{P}}}(\xi) = \inf\left\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \rho_{\mathbb{P}}(\xi - x - H(r-1)) \leq 0\right\},$$

where $\rho_{\mathbb{P}}$ is a law-invariant coherent risk measure:

$$\rho_{\mathbb{P}}(\xi) = \inf\{x \in \mathbb{R} \mid \rho_{\mathbb{P}}(\xi - x) \leq 0\}.$$

Under mild assumptions, the plug-in estimators are consistent:

$$\pi^{\hat{\mathbb{P}}_N}(\xi) \to \pi^{\mathbb{P}}(\xi) \quad \text{and} \quad \pi^{\rho_{\hat{\mathbb{P}}_N}}(\xi) \to \pi^{\rho_{\mathbb{P}}}(\xi) \quad \mathbb{P}^{\infty} - a.s.,$$

but are otherwise very poor and non-robust estimators!

Superhedging with respect to risk measures

Instead, we consider robust estimators. Consider $\beta_N \searrow 0$ and $\varepsilon_N \searrow 0$ s.t.

$$\mathbb{P}^N(W_p(\mathbb{P}, \hat{\mathbb{P}}_N) \geq \varepsilon_N) \leq \beta_N, \quad N > 1.$$

Define

$$\pi^{\rho}_{B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)}(\xi) := \inf\left\{x \in \mathbb{R}^d \;\middle|\; \exists H \in \mathbb{R}^d \text{ s.t. } \sup_{\nu \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \rho_\nu(\xi - x - H(r-1)) \leq 0\right\}.$$

Superhedging with respect to risk measures

University of Oxford
Mathematical Institute

Instead, we consider robust estimators. Consider $\beta_N \searrow 0$ and $\varepsilon_N \searrow 0$ s.t.

$$\mathbb{P}^N(W_p(\mathbb{P}, \hat{\mathbb{P}}_N) \geq \varepsilon_N) \leq \beta_N, \quad N > 1.$$

Define

$$\pi^{\rho}_{B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)}(\xi) := \inf\left\{x \in \mathbb{R}^d \;\middle|\; \exists H \in \mathbb{R}^d \text{ s.t. } \sup_{\nu \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \rho_\nu(\xi - x - H(r-1)) \leq 0\right\}.$$

Theorem

Assume g satisfies $|\xi(r) - \xi(\tilde{r})| \leq L_\gamma |r - \tilde{r}|^\gamma$ *for some* $\gamma \leq 1$, $L_\gamma \in \mathbb{R}$ *and that* $\sup_{\mu \in \mathfrak{P}} \int_0^1 \alpha^{-\gamma/p} m_\rho(d\alpha) < \infty$. *Then*

$$\lim_{n\to\infty} \pi^{\rho}_{B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)}(\xi) = \pi^{\rho_{\mathbb{P}}}(\xi) \qquad \mathbb{P}^\infty\text{-a.s.}$$

Robust Superhedging Price estimator

University of Oxford
Mathematical Institute

Take $k_N \to \infty$ and $k_N \varepsilon_N(\beta_N) \to 0$. Let

$$\pi_{\hat{\mathcal{Q}}_N}(\xi) = \sup_{\mathbb{P} \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\mathbb{Q} \in \mathcal{M} : \|d\mathbb{Q}/d\mathbb{P}\|_\infty \le k_N} \mathbb{E}_{\mathbb{Q}}[\xi]$$

Robust Superhedging Price estimator

Mathematical Institute

Take $k_N \to \infty$ and $k_N \varepsilon_N(\beta_N) \to 0$. Let

$$
\begin{aligned}
\pi_{\hat{\mathcal{Q}}_N}(\xi) &= \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\mathbb{Q}\in\mathcal{M}:\|d\mathbb{Q}/d\mathbb{P}\|_\infty\le k_N} \mathbb{E}_{\mathbb{Q}}[\xi] \\
&= \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\|d\mathbb{Q}/d\mathbb{P}\|_\infty\le k_N} \inf_{H\in\mathbb{R}^d} \mathbb{E}_{\mathbb{Q}}[\xi - H(r-1)] \\
&= \inf_{H\in\mathbb{R}^d} \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\|d\mathbb{Q}/d\mathbb{P}\|_\infty\le k_N} \mathbb{E}_{\mathbb{Q}}[\xi - H(r-1)] \\
&= \inf_{H\in\mathbb{R}^d} \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} AV@R^{\mathbb{P}}_{\frac{k_N-1}{k_N}}(\xi - H(r-1))
\end{aligned}
$$

$$
= \inf\left\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} AV@R^{\mathbb{P}}_{\frac{k_N-1}{k_N}}(\xi - H(r-1) - x) \le 0\right\}
$$

W_p-approach: Consistency & Robustness

Mathematical Institute

Theorem

Let g be Lipschitz continuous and bounded from below or continuous and bounded and $p \geq 1$. Then

$$\lim_{N\to\infty} \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{E}_{\mathbb{Q}}[\xi] = \pi^{\mathbb{P}}(\xi) \qquad \mathbb{P}^{\infty} - a.s.,$$

if NA($\mathbb{P}$) holds.

W_p-approach: Consistency & Robustness

Mathematical Institute

Theorem

Let g be Lipschitz continuous and bounded from below or continuous and bounded and $p \geq 1$. Then

$$\lim_{N \to \infty} \sup_{\mathbb{Q} \in \hat{\mathcal{Q}}_N} \mathbb{E}_{\mathbb{Q}}[\xi] = \pi^{\mathbb{P}}(\xi) \qquad \mathbb{P}^{\infty} - a.s.,$$

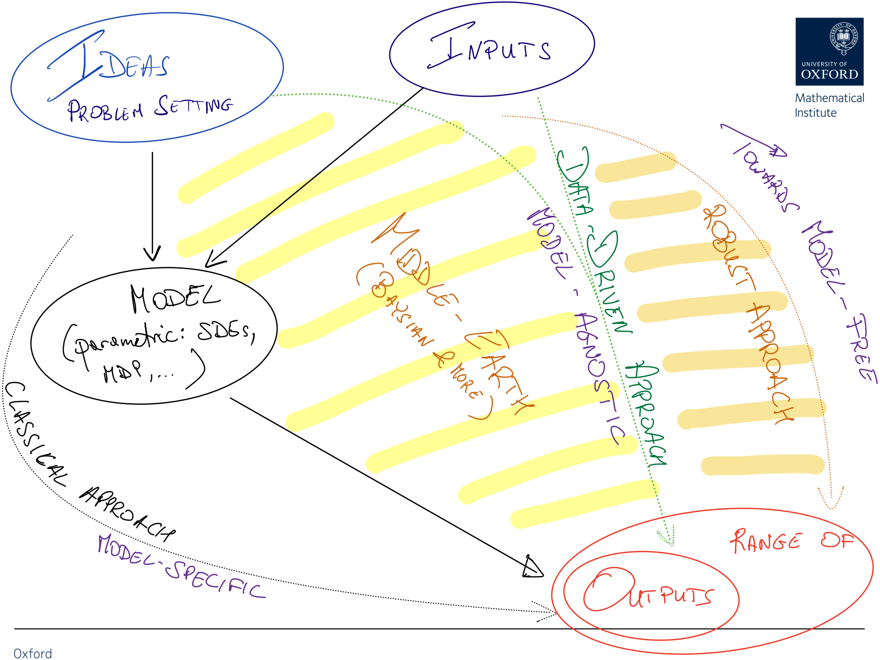
if $NA(\mathbb{P})$ holds. Further,

$$\sup_{\xi \in \mathcal{L}_1} \left| \sup_{\mathbb{Q} \in \hat{\mathcal{Q}}_N^1} \mathbb{E}_{\mathbb{Q}}[\xi] - \sup_{\mathbb{Q} \in \hat{\mathcal{Q}}_N^2} \mathbb{E}_{\mathbb{Q}}[\xi] \right|$$

$$\leq \max \left(\sup_{\mathbb{Q}^1 \in \hat{\mathcal{Q}}_N^1} \inf_{\mathbb{Q}^2 \in \hat{\mathcal{Q}}_N^2} W_p(\mathbb{Q}^1, \mathbb{Q}^2), \sup_{\mathbb{Q}^2 \in \hat{\mathcal{Q}}_N^2} \inf_{\mathbb{Q}^1 \in \hat{\mathcal{Q}}_N^1} W_p(\mathbb{Q}^2, \mathbb{Q}^1) \right).$$

where $\hat{\mathcal{Q}}_N^i$ are defined corresponding to some $\mathbb{P}^i \in \mathcal{P}(\mathbb{R}_+^d)$, $i = 1, 2$.

University of Oxford
Mathematical Institute



OT & Distributionally Robust Optimization

$\Upsilon =$ sensitivity w.r.t. the MODEL

based on Bartl, Drapeau, O. and Wiesel, *Proc. R. Soc. A* 477: 20210176, 2021
O. and Wiesel, *Math. Finance* 31(4): 1454–1493, 2021.

Consider the following optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a, x)\mu(dx),$$

where $\mathcal{A}$ is the set of controls, $\mathcal{S}$ is the state space and μ is the model.

Consider the following optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a,x)\mu(dx),$$

where $\mathcal{A}$ is the set of controls, $\mathcal{S}$ is the state space and μ is the model.

Examples:

- risk neutral pricing: $\mathbb{E}_{\mathbb{Q}}[f(S_T)]$,
- optimal investment: $\inf_{a \in \mathcal{A}} \mathbb{E}_{\mathbb{P}}[-U(x + \langle a, S_T - S_0 \rangle)]$,
- optimised certainty equivalents: $\inf_{a \in \mathbb{R}} \mathbb{E}_{\mathbb{P}}[a - U(X + a)]$
- marginal utility pricing (Davis' price)...

Consider the following optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a,x)\mu(dx),$$

where $\mathcal{A}$ is the set of controls, $\mathcal{S}$ is the state space and μ is the model.

Examples:

- risk neutral pricing: $\mathbb{E}_{\mathbb{Q}}[f(S_T)]$,
- optimal investment: $\inf_{a\in\mathcal{A}} \mathbb{E}_{\mathbb{P}}[-U(x + \langle a, S_T - S_0\rangle)]$,
- optimised certainty equivalents: $\inf_{a\in\mathbb{R}} \mathbb{E}_{\mathbb{P}}[a - U(X+a)]$
- marginal utility pricing (Davis' price)...
- OLS regression: $\inf_{a\in\mathbb{R}^d} \frac{1}{N}\sum_{i=1}^N (y^i - \langle a, x^i\rangle)^2$,
- ML/NN: $\inf \frac{1}{N}\sum_{i=1}^N |y^i - ((A_2(\cdot)+b_2)\circ\sigma\circ(A_1(\cdot)+b_1))(x^i)|^p$ over $a = (A_1, A_2, b_1, b_2) \in \mathcal{A} = \mathbb{R}^{k\times d}\times\mathbb{R}^{d\times k}\times\mathbb{R}^k\times\mathbb{R}^d$, where $(x^i, y^i)_{i=1}^N$ is the training set.
-

Given our optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a,x)\mu(dx),$$

we want to understand its dependence on the "model" μ.

We are interested in computing

$$\frac{\partial V}{\partial \mu} \quad \text{– the uncertainty sensitivity of the problem}$$

- parametric programming and statistical inference
  see Armacost & Fiacco '76 ... Bonnans & Shapiro '13;
- qualitative/quantitative stability in μ
  see Dupačová '90, Römisch '03
- robust optimisation
  see Bertsimas, Gupta & Kallus '18

Distributionally Robust Optimisation (DRO) considers

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} f(a,x)\nu(dx),$$

see Scarf '58, … ,Rahimian & Mehrotra '19, where

$B_\delta(\mu)$ is a δ–neighbourhood of the model μ.

Distributionally Robust Optimisation (DRO) considers

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} f(a,x)\nu(dx),$$

see Scarf '58, ... ,Rahimian & Mehrotra '19, where

$$B_\delta(\mu) \quad \text{is a } \delta\text{–neighbourhood of the model } \mu.$$

We propose to compute

$$\Upsilon := V'(0) = \lim_{\delta \searrow 0} \frac{V(\delta) - V(0)}{\delta} \quad \text{and} \quad \beth := \lim_{\delta \searrow 0} \frac{a^*(\delta) - a^*(0)}{\delta},$$

with $B_\delta(\mu)$ being Wasserstein balls around μ.

Υ *the sensitivity of the value w.r.t.* $\Upsilon\pi o\delta\varepsilon\gamma\mu\alpha$, *the Model.*

$\beth$ *the sensitivity of* בקרה, the control, w.r.t. the Model.

Uncertainty Sensitivity of DRO problems

UNIVERSITY OF OXFORD
Mathematical Institute

Recall our DRO problem (for simplicity $\mathcal{A} = \mathbb{R}^k$, $\mathcal{S} = \mathbb{R}^d$)

$$V(\delta) = \inf_{a \in \mathbb{R}^k} \sup_{\nu \in B_\delta(\mu)} \int_{\mathbb{R}^d} f(x, a)\, \nu(dx).$$

Uncertainty Sensitivity of DRO problems

Mathematical Institute

Recall our DRO problem (for simplicity $\mathcal{A} = \mathbb{R}^k$, $\mathcal{S} = \mathbb{R}^d$)

$$V(\delta) = \inf_{a \in \mathbb{R}^k} \sup_{\nu \in B_\delta(\mu)} \int_{\mathbb{R}^d} f(x, a)\, \nu(dx).$$

Theorem

For $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, and under suitable assumptions, we have

$$\Upsilon := V'(0) = \lim_{\delta \to 0} \frac{V(\delta) - V(0)}{\delta} = \inf_{a^* \in A^{\mathrm{opt}}(0)} \left(\int_{\mathbb{R}^d} |\nabla_x f(x, a^*)|^q \, \mu(dx) \right)^{1/q},$$

where $A^{\mathrm{opt}}(\delta)$ denotes the set of optimisers for $V(\delta)$.

Υ: uncertainty sensitivity of the value function

We can restate the result as

$$\inf_{a\in\mathbb{R}^k} \sup_{\nu\in B_\delta(\mu)} \int_{\mathbb{R}^d} f(x,a)\,\nu(dx) \approx \inf_{a\in\mathbb{R}^k} \int_{\mathbb{R}^d} f(x,a)\,\mu(dx) + \Upsilon\delta + o(\delta)$$

where

$$\Upsilon = \inf_{a^*\in A^{\mathrm{opt}}(0)} \left(\int_{\mathbb{R}^d} |\nabla_x f(x,a^*)|^q\,\mu(dx)\right)^{1/q}.$$

Υ: uncertainty sensitivity of the value function

We can restate the result as

$$\inf_{a\in\mathbb{R}^k} \sup_{\nu\in B_\delta(\mu)} \int_{\mathbb{R}^d} f(x,a)\,\nu(dx) \approx \inf_{a\in\mathbb{R}^k} \int_{\mathbb{R}^d} f(x,a)\,\mu(dx) + \Upsilon\delta + o(\delta)$$

where

$$\Upsilon = \inf_{a^*\in A^{\mathrm{opt}}(0)} \left(\int_{\mathbb{R}^d} |\nabla_x f(x,a^*)|^q\,\mu(dx)\right)^{1/q}.$$

- extends to general semi-norms;
- extends to sensitivity at a fixed $\delta > 0$: $V'(\delta+)$;
- extends to DRO problems with linear constraints, e.g., martingale;
- no first order loss from using $a^*(0)$ instead of $a^*(\delta)$.

Sketch of the proof (1)

Sensitivity of the value function: "$\leq$"

$$V(\delta) - V(0) \leq \sup_{\pi \in C_\delta(\mu)} \int f(y, a^*) - f(x, a^*)\, \pi(dx, dy)$$

$$= \sup_{\pi \in C_\delta(\mu)} \int \int_0^1 \langle \nabla_x f(x + t(y - x), a^*), (y - x) \rangle \, dt \, \pi(dx, dy)$$

$$\leq \delta \sup_{\pi \in C_\delta(\mu)} \int_0^1 \left(\int |\nabla_x f(x + t(y - x), a^*)|^q \pi(dx, dy) \right)^{1/q} dt.$$

+ growth conditions + DCT.

Sketch of the proof (2)

Sensitivity of the value function: "$\geq$"

$$T(x) := \frac{\nabla_x f(x, a^*)}{|\nabla_x f(x, a^*)|^{2-q}} \left(\int |\nabla_x f(z, a^*)|^q \, \mu(dz) \right)^{1/q-1}$$

$$\pi^\delta := [x \mapsto (x, x + \delta T(x))]_\# \mu \in C_\delta(\mu)$$

We can use π^δ to get a lower bound:

$$\begin{aligned}
\frac{V(\delta) - V(0)}{\delta} &\geq \frac{1}{\delta} \int f(x + \delta T(x), a^\delta) - f(x, a^\delta) \, \mu(dx) \\
&= \int \int_0^1 \langle \nabla_x f(x + t\delta T(x), a^\delta), T(x) \rangle \, dt \, \mu(dx) \\
&\xrightarrow{\delta \to 0} \int \langle \nabla_x f(x, a^*), T(x) \rangle \, \mu(dx) = \left(\int |\nabla_x f(x, a^*)|^q \, \mu(dx) \right)^{1/q}.
\end{aligned}$$

Ex 1: Call Price Sensitivity, classical vs robust

UNIVERSITY OF OXFORD
Mathematical Institute

Take $r = q = 0$, $T = 1$, $S_0 = 1$ and $\mu =$BS(σ) log-normal.

$$BS(\sigma) = \int_{\mathcal{S}} (s - K)^+ \mu(ds).$$

Ex 1: Call Price Sensitivity, classical vs robust

Take $r = q = 0$, $T = 1$, $S_0 = 1$ and $\mu =$BS(σ) log-normal.

$$\mathcal{R}BS(\delta) = \sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} (s-K)^+ \nu(ds).$$

Parametric Approach

$$B_\delta(\mu) = \{\mathrm{BS}(\tilde{\sigma}) : |\tilde{\sigma} - \sigma| \leq \delta\}$$

Then

$$\mathcal{R}BS'(0) = \mathcal{V} = S_0 \phi(d_+).$$

Non-parametric Approach

$$B_\delta(\mu) = \{\nu : W_2(\mu, \nu) \leq \delta\}$$

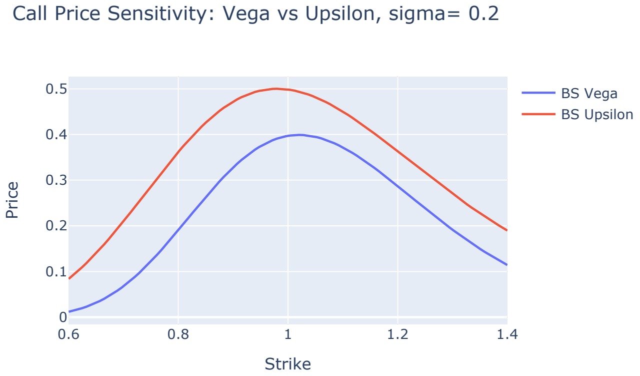
Then

$$\mathcal{R}BS'(0) = \Upsilon = S_0 \sqrt{\Phi(d_-)(1 - \Phi(d_-))}$$

BS Call: Vega($\mathcal{V}$) vs Upsilon(Υ)

Consider the simple example of a call option pricing.

Take $r = q = 0,\ T = 1,\ S_0 = 1$ and $\mu =$BS(σ) model.



Ex 2: Decision making & prefs representation

Mathematical Institute

Let X be agent's wealth/consumption. Savage '51, von Neuman & Morgenstern '53 give

$$\mathbb{P} \succeq \check{\mathbb{P}} \quad \Leftrightarrow \quad \mathbb{E}_{\mathbb{P}}[u(X)] \geq \mathbb{E}_{\check{\mathbb{P}}}[u(X)].$$

Ex 2: Decision making & prefs representation

Let X be agent's wealth/consumption. Savage '51, von Neuman & Morgenstern '53 give

$$\mathbb{P} \succeq \check{\mathbb{P}} \quad \Leftrightarrow \quad \mathbb{E}_{\mathbb{P}}[u(X)] \geq \mathbb{E}_{\check{\mathbb{P}}}[u(X)].$$

An ambiguity averse agent of Gilboa & Schmeidler '89, might instead consider

$$\mathbb{P} \succeq_{\rho} \check{\mathbb{P}} \quad \Leftrightarrow \quad \min_{\tilde{\mathbb{P}} \in B_\delta(\mathbb{P})} \mathbb{E}_{\tilde{\mathbb{P}}}[u(X)] \geq \min_{\tilde{\mathbb{P}} \in B_\delta(\check{\mathbb{P}})} \mathbb{E}_{\tilde{\mathbb{P}}}[u(X)].$$

for $B_\delta(\mathbb{P})$ a δ-ball around $\mathbb{P}$ in some metric ρ,

(also called *constraint preferences* by Hansen & Sargent '01).

Variational prefs: relative entropy vs Wasserstein

UNIVERSITY OF OXFORD
Mathematical Institute

The variational/constraint preferences with ρ-ball $B_\delta(\mathbb{P})$

$$\mathcal{U}(X) := \min_{\tilde{\mathbb{P}} \in B_\delta(\mathbb{P})} \mathbb{E}_{\tilde{\mathbb{P}}}[u(X)]$$

up to $o(\delta)$ are equivalent to:

$\rho =$ Rel. entropy

$$\mathcal{U}(X) \approx \mathbb{E}_{\mathbb{P}}[u(X))] - \delta\sqrt{2\mathrm{Var}_{\mathbb{P}}(u(X))}$$

(cf. Lam '16)

$\rho = W_2$ Wasserstein

$$\mathcal{U}(X) \approx \mathbb{E}_{\mathbb{P}}[u(X))] - \delta\sqrt{\mathbb{E}_{\mathbb{P}}\left[|u'(X)|^2\right]}$$

(cf. our Υ-sensitivity)

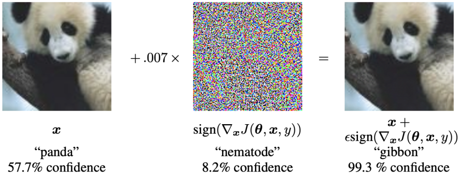
Example 3: NN & adversarial data

Most works focus on explaining the effects and creating algorithms to build adversarial examples.

Consider data (x, y) from μ and a 1-layer NN: $(A_1^*, A_2^*, b_1^*, b_2^*)$ solve

$$\inf \int \underbrace{|y - ((A_2(\cdot) + b_2) \circ \sigma \circ (A_1(\cdot) + b_1))(x)|^p}_{=:f(x,y;A,b)} \mu(dx, dy),$$

where the inf is taken over $(A_1, A_2, b_1, b_2) \in \mathbb{R}^{k\times d} \times \mathbb{R}^{d\times k} \times \mathbb{R}^k \times \mathbb{R}^d$.



Source: Goodfellow, Shlens & Szegedy ICLR 2015

Example 3: NN & adversarial data

Most works focus on explaining the effects and creating algorithms to build adversarial examples.
Consider data (x, y) from μ and a 1-layer NN: $(A_1^*, A_2^*, b_1^*, b_2^*)$ solve

$$\inf \int \underbrace{|y - ((A_2(\cdot) + b_2) \circ \sigma \circ (A_1(\cdot) + b_1))\,(x)|^p}_{=:f(x,y;A,b)} \mu(dx, dy),$$

where the inf is taken over $(A_1, A_2, b_1, b_2) \in \mathbb{R}^{k\times d} \times \mathbb{R}^{d\times k} \times \mathbb{R}^k \times \mathbb{R}^d$.

Then, sensitivity to adversarial data examples from $\hat{\mu} \in B_\delta(\mu)$ given by:

$$\left(\int |\nabla_{(x,y)} f(x, y; A^*, b^*)|^q \,\mu(dx, dy) \right)^{1/q}.$$

Sensitivity of optimisers

Mathematical
Institute

Theorem

For $p = q = 2$, under suitable regularity and growth assumptions,

$$\lim_{\delta \to 0} \frac{a^*(\delta) - a^*}{\delta} = -\frac{1}{\Upsilon}(\nabla_a^2 V(0, a^*))^{-1} \int \nabla_x \nabla_a f(x, a^*) \nabla_x f(x, a^*) \, \mu(dx),$$

where $a^* := a^*(0)$.

The results extends to general $p > 1$ and semi-norms.

Example 1: Square-root LASSO

Consider $\|(x,y)\|_* = |x|_r 1_{\{y=0\}} + \infty 1_{\{y\neq 0\}}$, $r>1$, $(x,y)\in\mathbb{R}^k\times\mathbb{R}$.
Then (see Blanchet, Kang & Murthy '19)

$$\inf_{a\in\mathbb{R}^k}\sup_{\nu\in B_\delta(\hat\mu_N)}\int (y-\langle x,a\rangle)^2\,d\nu = \inf_{a\in\mathbb{R}^k}\left(\sqrt{\int (y-\langle a,x\rangle)^2\,d\hat\mu_N} + \delta|a|_s\right)^2,$$

where $1/r+1/s=1$. $\hat\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{(x^i,y^i)}$ encodes the observations.
System is overdetermined so that $D=\int xx^T\,\mu(dx)$ is invertible.
$\delta=0$ case is the ordinary least squares regression: $a^* = \frac{1}{N}D^{-1}\int yx d\mu$.

Example 1: Square-root LASSO

Consider $\|(x,y)\|_* = |x|_r 1_{\{y=0\}} + \infty 1_{\{y\neq 0\}}$, $r>1$, $(x,y)\in\mathbb{R}^k\times\mathbb{R}$.
Then (see Blanchet, Kang & Murthy '19)

$$\inf_{a\in\mathbb{R}^k}\sup_{\nu\in B_\delta(\hat{\mu}_N)}\int (y-\langle x,a\rangle)^2\,d\nu = \inf_{a\in\mathbb{R}^k}\left(\sqrt{\int (y-\langle a,x\rangle)^2\,d\hat{\mu}_N} + \delta|a|_s\right)^2,$$

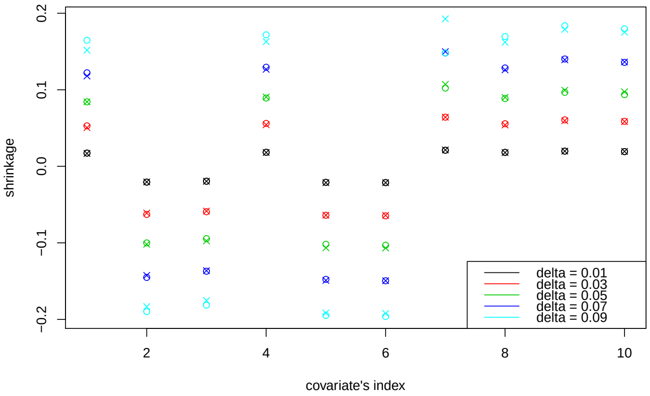
where $1/r+1/s=1$. $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{(x^i,y^i)}$ encodes the observations.
System is overdetermined so that $D=\int xx^T\,\mu(dx)$ is invertible.
$\delta=0$ case is the ordinary least squares regression: $a^* = \frac{1}{N}D^{-1}\int yx d\mu$.
$\delta>0$, $s=1 \rightsquigarrow$ RHS = square-root LASSO regression Belloni et al. '11
$\delta>0$, $s=2 \rightsquigarrow$ RHS $\approx$ Ridge regression
Then $a^*(\delta)$ is approximately, for $s=1$ and $s=2$ (cf. Tibshirani '96):

$$a^* - \sqrt{V(0)}D^{-1}\mathrm{sgn}(a^*)\delta \quad \text{and} \quad a^*\left(1-\frac{\sqrt{V(0)}}{|a^*|_2}D^{-1}\delta\right)$$

Square-root LASSO: numerics

Comparison of exact (o) and first-order (x) approximation of square-root LASSO coefficients for 2000 data generated from: (with all X_i, ε i.i.d. $\mathcal{N}(0,1)$)

$$Y = 1.5X_1 - 3X_2 - 2X_3 + 0.3X_4 - 0.5X_5 - 0.7X_6 + 0.2X_7 + 0.5X_8 + 1.2X_9 + 0.8X_{10} + \varepsilon.$$



Example 2: a CLT of Blanchet, Murphy and Si '19

UNIVERSITY OF OXFORD
Mathematical Institute

Consider the empirical measure $\hat{\mu}_N$ of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\text{arg}} \sup_{\nu \in B_\delta(\hat{\mu}_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \underset{min}{\text{arg}} \int f(x,a)\,\hat{\mu}_N(dx), \;\; a^* = \underset{min}{\text{arg}} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a^{*,N}_{1/\sqrt{N}} \to a^*$.

Example 2: a CLT of Blanchet, Murphy and Si '19

Consider the empirical measure $\hat{\mu}_N$ of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\arg} \sup_{\nu \in B_\delta(\hat{\mu}_N)} \int f(x,a)\,\nu(dx),\ \ a^{*,N} = \underset{min}{\arg} \int f(x,a)\,\hat{\mu}_N(dx),\ \ a^* = \underset{min}{\arg} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a^{*,N}_{1/\sqrt{N}} \to a^*$.

Let $\sigma^2 := \int \nabla_a f(x,a^*)^T \nabla_a f(x,a^*)\,\mu(dx)$. Classical results give

$$\sqrt{N}\left(a^{*,N} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1} H, \quad \text{where } H = \mathcal{N}(0,\sigma^2).$$

Example 2: a CLT of Blanchet, Murphy and Si '19

Consider the empirical measure $\hat{\mu}_N$ of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\arg} \sup_{\nu \in B_\delta(\hat{\mu}_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \underset{min}{\arg} \int f(x,a)\,\hat{\mu}_N(dx), \;\; a^* = \underset{min}{\arg} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a^{*,N}_{1/\sqrt{N}} \to a^*$.

Let $\sigma^2 := \int \nabla_a f(x,a^*)^T \nabla_a f(x,a^*)\,\mu(dx)$. Classical results give

$$\sqrt{N}\left(a^{*,N} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1} H, \quad \text{where } H = \mathcal{N}(0,\sigma^2).$$

Our results show that

$$\sqrt{N}\left(a^{*,N}_{1/\sqrt{N}} - a^{*,N}\right) \approx (\nabla_a^2 V(0,a^*))^{-1} \cdot \nabla_a \sqrt{\int |\nabla_x f(x,a^{*,N})|_s^2\,\hat{\mu}_N(dx)}.$$

Putting the two together yields the CLT of Blanchet, Murphy and Si '19

$$\sqrt{N}\left(a^{*,N}_{1/\sqrt{N}} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1}\left(H - \nabla_a \sqrt{\int |\nabla_x f(x,a^*)|_s^2\,\mu(dx)}\right).$$

$\rightsquigarrow$ out-of-sample error estimates.

Example 3: EUM & Optimal investment

$X = S_T - S_0 \sim \mu$ vector of returns in $\mathcal{S} \subset \mathbb{R}^d$ and $\mathcal{A} \subseteq \mathbb{R}^d$ admissible strategies; wlog $r = 0$, initial capital $x = 0$.
$u : \mathbb{R} \to \mathbb{R}$ strictly concave, continuously differentiable, bounded from above. Consider

$$V(\delta) = \sup_{a \in \mathcal{A}} \inf_{\nu \in B_\delta(\mu)} \mathbb{E}_\nu \left[u\left(\langle X, a \rangle\right)\right]$$

Then, under mild technical assumptions,

$$a^{\star\prime}(0) = \| u'(\langle X, a^\star \rangle) \|_{L^q(\mu)}^{1-q} \cdot \left(\nabla_\pi^2 V(0)\right)^{-1} \cdot \frac{a^\star}{|a^\star|} \cdot \left(\mathbb{E}_\mu \left[\frac{\langle X, a^\star \rangle u''(\langle X, a^\star \rangle) + u'(\langle X, a^\star \rangle)}{|u'(\langle X, a^\star \rangle)|^{1-q}} \right] \right).$$

Binomial model with a log utility

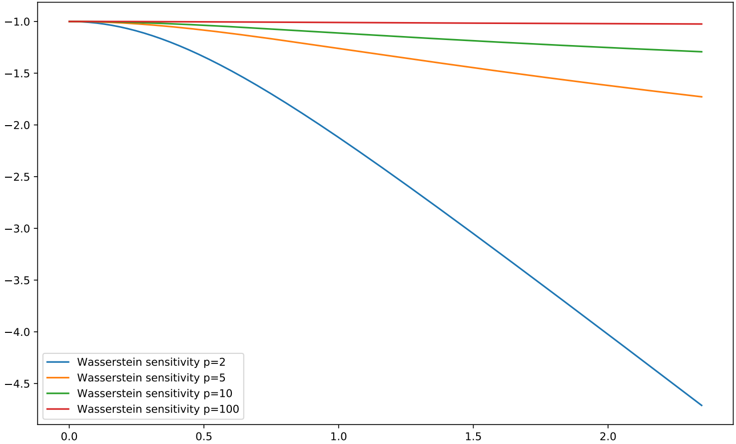


Figure: Wasserstein sensitivity ${a^{\star}}'(0)$ in a Binomial-model

Ex 4: Marginal utility (Davis') price

Recall the EUM setup. For a continuous payoff $g \geq 0$ consider

$$V(\varepsilon, p_d) := \sup_{a \in \mathcal{A}} \mathbb{E}_\mu \left[u \left(-\varepsilon + \langle X, a \rangle + \frac{\varepsilon}{p_d} g(X) \right) \right],$$

Definition

Suppose that for each $p_d > 0$, the function $\varepsilon \mapsto V(\varepsilon, p_d)$ is differentiable at $\varepsilon = 0$ and $\hat{p}_d$ is a solution to

$$\partial_\varepsilon V(0, p_d) = 0.$$

Then $\hat{p}_d$ is called a marginal utility price of the option g.

Characterisation of the marginal utility price

Theorem (Davis (1997))

Under mild technical assumptions $\hat{p}_d$ is unique and satisfies

$$\hat{p}_d = \frac{\mathbb{E}_\mu \left[u'(\langle X, a^\star \rangle) g(X)\right]}{\mathbb{E}_\mu \left[u'(\langle X, a^\star \rangle)\right]}.$$

In this way $\hat{p}_d$ is the price under a subjective martingale measure:

$$X = S_T - S_0 \quad \text{and} \quad \mathbb{E}_\mu \left[u'(\langle X, a^\star \rangle) X\right] = 0.$$

Robust marginal utility price

UNIVERSITY OF OXFORD
Mathematical Institute

Definition

Let us define

$$V(\delta, \varepsilon, p_d) = \sup_{a \in \mathcal{A}} \inf_{\nu \in B_\delta(\mu)} \mathbb{E}_\nu \left[u \left(-\varepsilon + \langle X, a \rangle + \frac{\varepsilon}{p_d} g(X) \right) \right].$$

Suppose that for each $p_d > 0$ the function $\varepsilon \mapsto V(\delta, \varepsilon, p_d)$ is differentiable. A number $\hat{p}_d(\delta)$, which satisfies

$$\partial_\varepsilon V(\delta, 0, \hat{p}_d(\delta)) = 0.$$

is called a robust marginal utility price of g at the uncertainty level δ.

Characterisation of DR marginal utility price

UNIVERSITY OF OXFORD
Mathematical Institute

Theorem

Fix $\delta \geq 0, p_d > 0$. Under mild technical assumptions the robust marginal utility price $\hat{p}_d(\delta)$ is given by

$$\hat{p}_d(\delta) = \frac{\mathbb{E}_{\mu^\star}\left[u'(\langle X - X_0, a^\star_\delta\rangle)\, g(X)\right]}{\mathbb{E}_{\mu^\star}\left[u'(\langle X - X_0, a^\star_\delta\rangle)\right]}$$

for any pair of optimisers $a^\star_\delta \in \mathcal{A}$ and $\mu^\star \in B_\delta(\mu)$.

As before, $\hat{p}_d(\delta)$ is the price under a subjective martingale measure but which also depends on δ.

Characterisation of DR marginal utility price

UNIVERSITY OF OXFORD
Mathematical Institute

Theorem

Fix $\delta \geq 0, p_d > 0$. Under mild technical assumptions the robust marginal utility price $\hat{p}_d(\delta)$ is given by

$$\hat{p}_d(\delta) = \frac{\mathbb{E}_{\mu^\star}\left[u'(\langle X - X_0, a^\star_\delta \rangle)\, g(X)\right]}{\mathbb{E}_{\mu^\star}\left[u'(\langle X - X_0, a^\star_\delta \rangle)\right]}$$

for any pair of optimisers $a^\star_\delta \in \mathcal{A}$ and $\mu^\star \in B_\delta(\mu)$.

As before, $\hat{p}_d(\delta)$ is the price under a subjective martingale measure but which also depends on δ.

Special cases: $\hat{p}_d = \hat{p}_d(\delta)$ for all $\delta > 0$, e.g., for $\mu = \mathcal{N}(m, \sigma^2)$, $p = \infty$ and an agent with an exponential utility.

Sensitivity of the marginal utility price

Mathematical Institute

Theorem

Under mild technical assumptions the following holds:

(i) *If $a^\star = 0$, then the Davis price $\hat{p}_d(\delta)$ satisfies*

$$\hat{p}_d'(0) = -\left(\mathbb{E}_\mu\left[|\nabla g(x)|^q\right]\right)^{1/q}.$$

(ii) *If $a^\star \neq 0$ then*

$$\hat{p}_d'(0) = \frac{1}{\mathbb{E}_\mu\left[u'(\langle X, a^\star\rangle)\right]}\Bigg(\mathbb{E}_\mu\Big[u''(\langle X, a^\star\rangle)\cdot\Big(\langle T(X), a^\star\rangle - \langle X, a'(0)\rangle\Big)$$

$$\cdot\left(\mathbb{E}_{\hat{\mu}}\left[g(X)\right] - g(X)\right)\Big]\Bigg) - \mathbb{E}_{\hat{\mu}}\left[\langle\nabla g(X), T(X)\rangle\right],$$

where $\frac{d\hat{\mu}}{d\mu} \propto u'(\langle X, a^\star\rangle)$ *and* $T(x) \propto \frac{a^\star}{|a^\star|}|u'(\langle x, a^\star\rangle)|^{q-1}$.

Conclusion & Outlook

Mathematical Institute

- Constrained (martingale, covariance) variants of OT appear naturally in applications
- Numerics pose interesting new challenges.
- OT allows to conceptualise and quantify the impact of model uncertainty
- Useful in data-driven and classical modelling approaches alike
- Wasserstein balls capture model uncertainty well, small and large uncertainty alike
- First-order approximations for DRO available analytically
- Applications in finance, statistics, UQ, ML and more!

Thank You

papers available at `www.maths.ox.ac.uk/people/jan.obloj`

The robust optimisation problem rewritten

UNIVERSITY OF OXFORD
Mathematical Institute

Consider the simplified problem

$$\sup_{\nu \in B^p_{\delta^{1/p}}(\mu)} \int f(x)\, \nu(dx).$$

Theorem (Bartl, Drapeau & Tangpi '19; Blanchet, Kang & Murthy '19)

For $f : \mathbb{R} \to \mathbb{R}$ *bounded below*

$$\sup_{\nu \in B^p_{\delta^{1/p}}(\nu)} \int f(x)\, \nu(dx) = \inf_{\lambda \geq 0} \left(\int f^{\lambda|\cdot|^p}(x)\, \mu(dx) + \delta\lambda \right),$$

where

$$f^{\lambda|\cdot|^p}(x) := \sup\left\{ f(y) - \lambda|x-y|^p \; : \; y \in \mathbb{R}^d \text{ s.t. } f(y) < \infty \right\}.$$