GLOBAL SENSITIVITY WITH OPTIMAL TRANSPORT

Presenter: Emanuele Borgonovo

Bocconi University

Joint work with Alessio Figalli, Elmar Plischke and Giuseppe Savarè

Università Bocconi
MILANO

Joint work with

Alessio Figalli
ETH Zurich

Elmar Plischke
TU Clausthal

Giuseppe Savarè
Bocconi University

Motivation

- In several applications model output is multivariate in nature (e.g., vectorial, spatio-temporal dependent)
- We can look at one component at a time
- We also may wish to have an overall indicator of importance

Previous works

- Marrel, Iooss, et al (2011), Models with spatially dependent outputs, Environmetrics.
- Gamboa, Janon, Klein, Lagnoux (2014) Sensitivity Analysis for Multidimensional and Functional Outputs, *Generalized Sobol' Indices*
- Rahman (2016) – The f-sensitivity index (SIAM-JUQ)
- Marrel, Saint-Geours, De Lozzo (2017) in Handbook of Uncertainty Quantification
- Gamboa, Klein, Lagnoux (2018) – Cramer von Mises Distances (SIAM-JUQ)
- Fraiman, R., F. Gamboa, and L. Moreno. 2020. "Sensitivity Indices for Output on a Riemannian Manifold." *International Journal for Uncertainty Quantification* 10 (4): 297–314.

Università Bocconi MILANO

Previous work (cont.)

- Gamboa, Klein, Lagnoux, Moreno (2021), Sensitivity Analysis in General Metric Spaces: output belongs to a generic metric space
- Fort, Klein, Lagnoux (2021), sensitivity for stochastic outputs when the distance is the Wasserstein distance
- Lamboni (2019), (2020), generalized Sobol' indices and derivative based sensitivity measures
- Alexanderian, Gremaud and Smith (2020) variance-based indices are modified to take into account the temporal variation of the output process variance
- Zham, Constantine, Prieur Marzok, 2020, SIAM J Sc. Comp., with gradient based methods for multivariate vector valued functions
- Da Veiga (2021) (ArXiv) on sensitivity measures based on Kernel-embeddings
- Barr and Rabitz (2022), Siam-JUQ, also on a Kernel-based approach to global sensitivity analysis
- And many others

Optimal transport

- Originated by the work of Gaspard Monge (1781)
- Extended by Kantorovich to a more general formulation (1942)
- Has received increasing attention in mathematics and artificial intelligence
- Distance between images
- Cuturi (2013) breakthrough paper on the solution of the OT problem using the Sinkhorn approximation
- Recent review in SIAM-Review by Chen et al (2021): "Stochastic Control Liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger Bridge"
- Over time two Fields medal Cedric Villani and Alessio Figalli

OT Formulation

— $(\Omega, \mathcal{F}, \mathbb{P})$ reference probability space, let v and v' be two probability measures on this space.

Given $\nu, \nu' \in \mathscr{P}(\mathsf{Y})$ we denote by $\Pi(\nu, \nu')$ the set of plans or couplings $\pi \in \mathscr{P}(\mathsf{Y} \times \mathsf{Y})$ whose marginals are ν and ν' respectively. Posed a lower semicontinuous cost function $\mathsf{k} : \mathsf{Y} \times \mathsf{Y} \to [0, +\infty]$, the Kantorovich formulation of the optimal transport problem

$$K(\nu, \nu') = \inf_{\pi \in \Pi(\nu, \nu')} \mathcal{C}(\pi), \quad \mathcal{C}(\pi) := \int_{\mathsf{Y}^2} \mathsf{k}(y, z)\, \mathrm{d}\pi(y, z), \tag{1}$$

consists of finding a transfer plan $\pi \in \Pi(\nu, \nu')$ that minimizes the integrated cost $\mathcal{C}$.

The Wasserstein Distance

When $\mathsf{k}(y,z) = \mathsf{d}^p(y,z)$ for a suitable continuous metric $\mathsf{d} : \mathsf{Y} \times \mathsf{Y} \to [0,+\infty)$, the Kantorovich problem

$$W_p^p(\nu,\nu') = \inf_{\pi \in \Pi(\nu,\nu')} \int \mathsf{d}^p(y,z)\,\mathrm{d}\pi(y,z) \tag{7}$$

defines the p-th power of the so-called p-Wasserstein distance W_p, which satisfies the axioms of a metric in the subset $\mathscr{P}_p(\mathsf{Y})$ whose measures have finite p-th moment

The Elliptical Distribution Case

In fact, ν and ν' are two normal distributions with mean values μ, μ' and covariance matrices Σ, Σ' respectively, then Givens and Shortt [1984] show that

$$W_2^2(\nu, \nu') = \|\mu - \mu'\|_2^2 + \mathrm{Tr}\left(\Sigma + \Sigma' - 2\left(\Sigma^{1/2}\Sigma'^{1/2}\right)^{1/2}\right), \tag{12}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix and $\Sigma^{1/2}$ is the square root of a symmetric and positive matrix. Gelbrich [1990] extends this result for the case in which ν and ν' are two elliptical distributions with the same characteristic generator.

The quantity in Eq. (12) is called the squared Wasserstein-Bures distance

OT-based Global Sensitivity Measures

We define a family of probabilistic sensitivity measures based on optimal transport. We follow the notation in Section 2, and consider an OT-problem with a lower semicontinuous cost function $\mathsf{k} : \mathsf{Y} \times \mathsf{Y} \to [0, +\infty]$ satisfying $\mathsf{k}(y, y') = 0 \;\Leftrightarrow\; y = y'$.

Lemma 2. *Let $\nu, \nu' \in \mathscr{P}(\mathsf{Y})$. The function $K(\cdot, \cdot) : \mathscr{P}(\mathsf{Y}) \times \mathscr{P}(\mathcal{Y}) \to [0, +\infty]$ defined by (1) is a separation measurement and $K(\nu, \nu') = 0 \iff \nu = \nu'$.*

Definition 3. *Let X, Y be random variables with laws μ, ν respectively and let $(\nu_x^{\mathcal{F}})_{x \in \mathsf{X}}$ be the conditional law of Y generated by $(X, \mathcal{F})$. We call*

$$\xi^K(Y, X) := \mathbb{E}[K(\mathbb{P}_Y, \mathbb{P}_{Y|X})] = \int_{\mathsf{X}} K(\nu, \nu_x)\, \mathrm{d}\mu(x) \tag{22}$$

the OT-based global sensitivity measure of X with respect to Y.

Università Bocconi
MILANO

Interpretation

- The most important input is the one for which we expect to spend more work (it is more costly on average) to pass from the distribution of Y to the conditional distribution of Y given X

- The class encompasses several global sensitivity measures. For instance it can be shown that the δ-importance measure (B. 2007) is a special case of an OT-based sensitivity measure if the space of the output is equipped with the discrete metric.

Properties 1: Zero-independence property

Proposition 5. $\xi^K(Y,X) \geq 0$ *and* $\xi^K(Y,X) = 0$ *if and only if* Y *and* X *are statistically independent.*

Properties 2: Functional Dependence

Let:

$$\mathbb{M}_{\mathsf{k}}[Y] := \int_{\mathsf{Y}^2} \mathsf{k}(y, y')\, \mathrm{d}\nu(y)\, \mathrm{d}\nu(y') = \mathbb{E}[\mathsf{k}(Y, Y')]$$

Lemma 6. *For every random variable X, Y*

$$\xi^K(Y, X) \leq \mathbb{M}_{\mathsf{k}}[Y]; \tag{26}$$

in particular ξ^K is finite if $\mathbb{M}_{\mathsf{k}}[Y] < \infty$. In this case, if Y is functionally dependent on X, i.e. $Y = f(X)$ $\mathbb{P}$-a.e. for some Borel map $f : \mathsf{X} \to \mathsf{Y}$, then $\xi^K(Y, X) = \mathbb{M}_{\mathsf{k}}[Y]$ so that the maximum value is attained in (26).

Notable Case

Remark 7. *When* $\mathsf{Y} = \mathbb{R}^m$ *and* $\mathsf{k}(y, y') := |y - y'|^2$ *we have*

$$\mathbb{M}_{\mathsf{k}}[Y] = \int_{\mathsf{Y}^2} |y - y'|^2 \, \mathrm{d}\nu(y) \, \mathrm{d}\nu(y') = 2 \int_{\mathsf{Y}} |y - \bar{\nu}|^2 \, \mathrm{d}\nu(y) = 2\mathrm{Var}[Y], \quad (27)$$

where $\bar{\nu} := \int y \, \mathrm{d}\nu(y) = \mathbb{E}[Y]$. *The previous lemma then yields*

$$\xi^{W_2^2} \leq 2\mathrm{Var}[Y]. \quad (28)$$

In particular ξ^K *is finite if* $\mathbb{E}\left[|Y|^2\right] < \infty$.

A stronger result

Theorem 8. *Assuming* $Y \in L^2(\Omega, \mathbb{P})$, $\xi^{W_2^2}(Y, X) = \mathbb{M}_{\mathsf{k}}[Y] = 2\mathrm{Var}[Y]$ *if and only if there exists a Borel map* $f : \mathsf{X} \to \mathsf{Y}$ *such that* $Y = f(X)$ $\mathbb{P}$*-a.e.*

A New Sensitivity Index

Definition 12. *If* $\mathbb{M}_{\mathsf{k}}[Y] > 0$*, we let*

$$\iota^{\mathsf{k}}(Y,X) = \frac{\xi^{K}(Y,X)}{\mathbb{M}_{\mathsf{k}}[Y]} \tag{32}$$

Based on the results of Section 3.1, we immediately have that for any cost k, $0 \leq \iota^{\mathsf{k}}(Y,X) \leq 1$, with $\iota^{\mathsf{k}}(Y,X) = 0$ indicating statistical independence and $\iota^{\mathsf{k}}(Y,X) = 1$ in the case of functional dependence. In the remainder, we shall focus on the case in which the cost is associated with the squared 2-Wasserstein distance:

$$\iota(Y,X) = \frac{\xi^{W_2^2}(Y,X)}{2\mathbb{V}[Y]}. \tag{33}$$

With this choice, $\iota(Y,X) = 1$ is equivalent to functional dependence.

The Wasserstein-Bures Case

Proposition 13. *Assume that the second moment of Y is finite. If $\mathbb{P}_Y$ is elliptical with generating function h, and $\mathbb{P}_{Y|X}$ is elliptical with the identical generating function h for all values X, then*

$$\iota(Y,X) = \frac{\mathbb{E}[\|\mathbb{E}[Y] - \mathbb{E}[Y|X]\|_2^2] + \mathbb{E}[\mathrm{Tr}\left(\Sigma_Y + \Sigma_{Y|X} - 2\left(\Sigma_Y^{1/2}\Sigma_{Y|X}\Sigma_Y^{1/2}\right)^{1/2}\right)]}{2\mathbb{V}[Y]}, \qquad (34)$$

where the right hand side of (59) is the squared Wasserstein-Bures distance between $\mathbb{P}_Y$ and $\mathbb{P}_{Y|X}$.

Università Bocconi
MILANO

Advective and Diffusive Parts

- The previous index is the sum of two parts:

$$\mathrm{Adv}_X = \frac{\mathbb{E}[\|\mathbb{E}[Y] - \mathbb{E}[Y|X]\|_2^2]}{2\mathbb{V}[Y]},$$

- That accounts for the shift in mean values (centers of mass), and

$$\mathrm{Diff}_X = \frac{\mathbb{E}[\mathrm{Tr}\left(\Sigma_Y + \Sigma_{Y|X} - 2\left(\Sigma_Y^{1/2}\Sigma_{Y|X}\Sigma_Y^{1/2}\right)^{1/2}\right)}{2\mathbb{V}[Y]}.$$

- That accounts for multi-directional dispersion around the means

Link to Gamboa et al Generalized Indices

Proposition 14. *For Adv_X in (36), we have:*

$$Adv_X = \mathbb{E}\left[\|\mathbb{E}[Y] - \mathbb{E}[Y|X_i]\|_2^2\right] = \mathbb{E}\left[\sum_{t=1}^{n_Y}(\mathbb{E}[Y_t] - \mathbb{E}[Y_t|X_i])^2\right] = \sum_{t=1}^{n_Y}\xi_i^{V,t}, \qquad (38)$$

where $\xi_i^{V,t}$ is the univariate variance-based sensitivity measure (19) of X_i with respect to Y_t. Moreover, if we assume that the inputs are independent then we have $Adv_X = \sum_{t=1}^{n_Y}\mathbb{V}[Y^t]S_X^t$, where S_X^t is the Sobol' first order sensitivity measure of X with respect to the t^{th} component of the output, Y^t.

Thus we have a variance contribution, plus other contributions in an OT-based sensitivity measure. If distributions are elliptical, we know the additional term, otherwise we have to solve the OT problem

Università Bocconi MILANO

Distortion

Theorem 10. *Let* $(\mathsf{Z}, \mathcal{G})$ *be a measurable space and let* $g : \mathsf{X} \to \mathsf{Z}$ *be a* $(\mu, \mathcal{G})$*-measurable map with* $Z := g \circ X$*. We have*

$$\xi^K(Y, X) \geq \xi^K(Y, Z) = \xi^K(Y, (X, \mathcal{F}')), \text{ where } \mathcal{F}'^{-1}(\mathcal{G}) \subset \mathcal{F}. \tag{30}$$

In particular, if the σ*-algebra generated by* g *coincides with* $\mathcal{F}$ *we have* $\xi^K(Y, X) = \xi^K(Y, Z)$.

Monotonicity

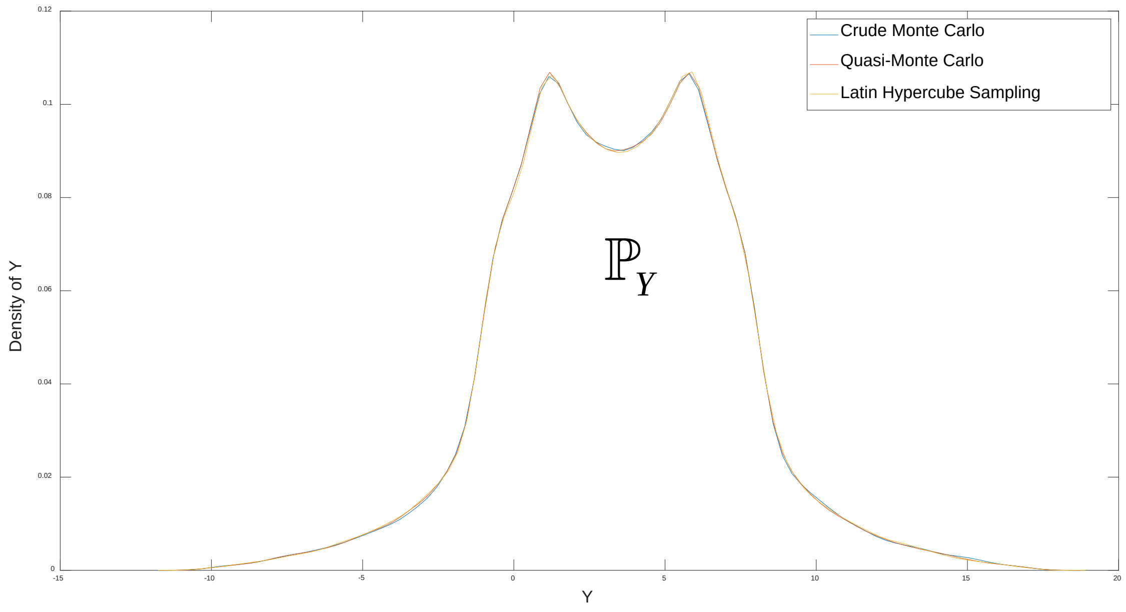
Theorem 11. *Let* $(\mathfrak{F}^n)_{n\in\mathbb{N}}$ *be an increasing family of sub-*σ*-algebras in* $\mathfrak{F}$ *with* $\mathfrak{F} = \bigvee_{n=1}^{\infty} \mathfrak{F}^n$*. We have*

$$\lim_{n\to\infty} \xi^K(Y,(X,\mathfrak{F}^n)) = \xi^K(Y,(X,\mathfrak{F})). \tag{31}$$

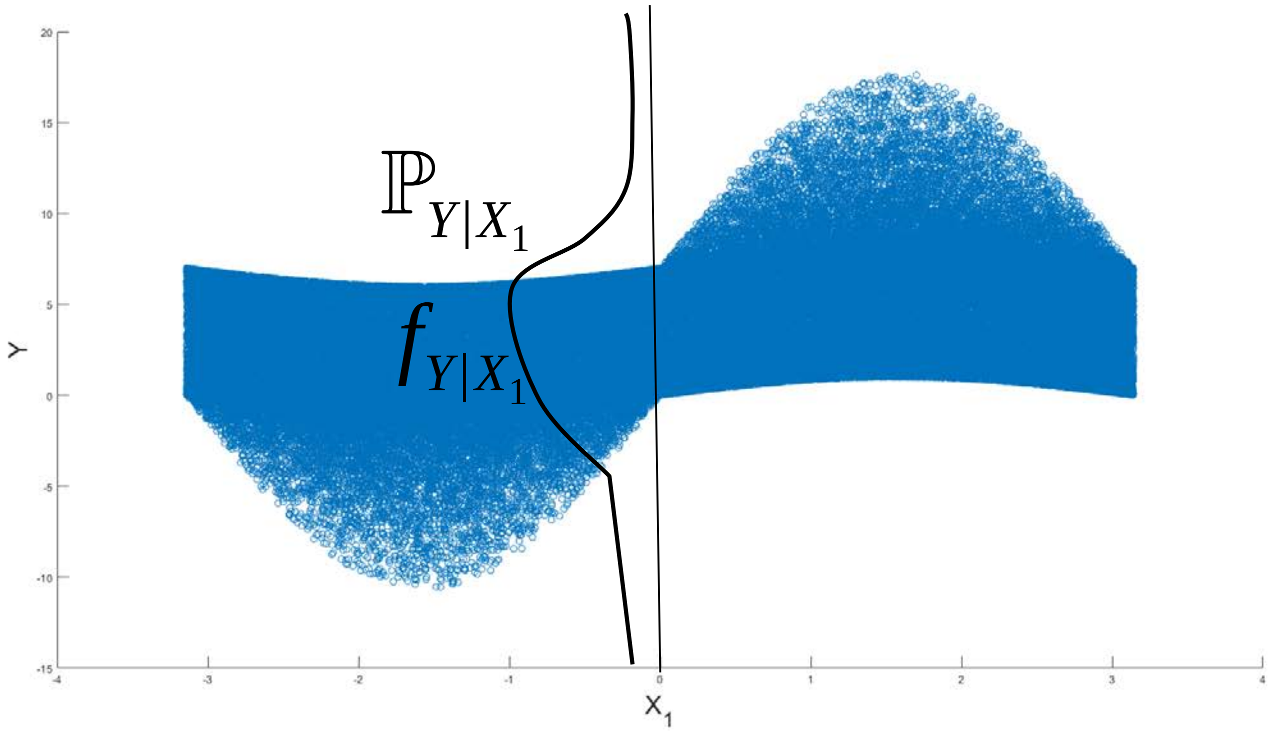
ESTIMATION

Università Bocconi
MILANO

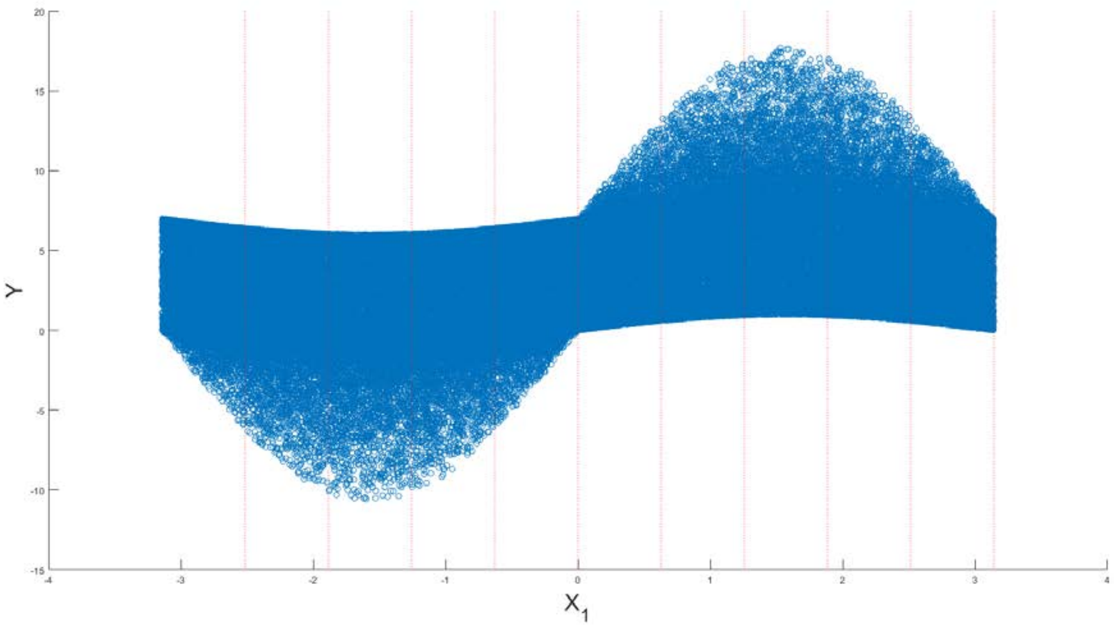
Ishigami Function Distribution



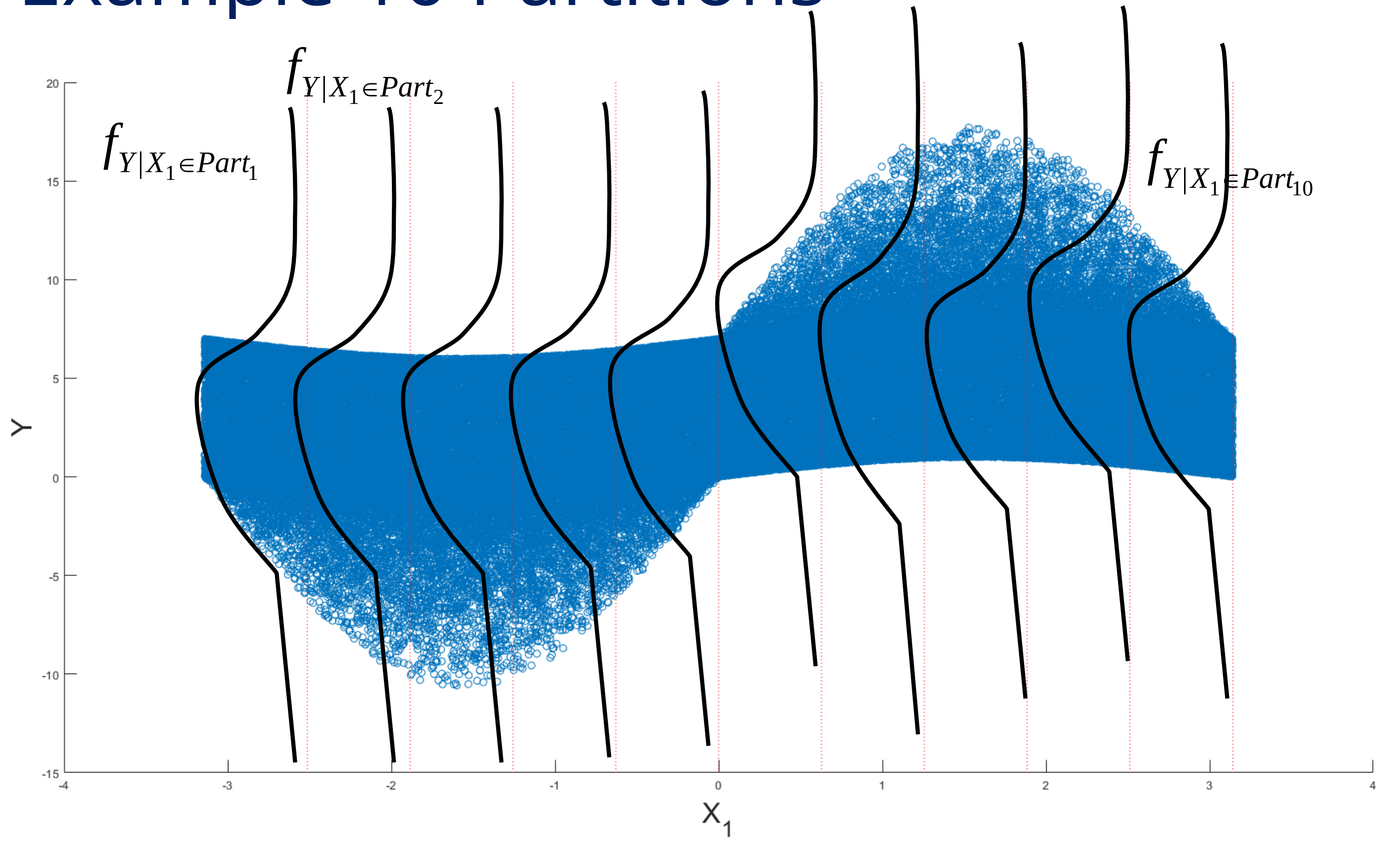
2) Form the scatterplot

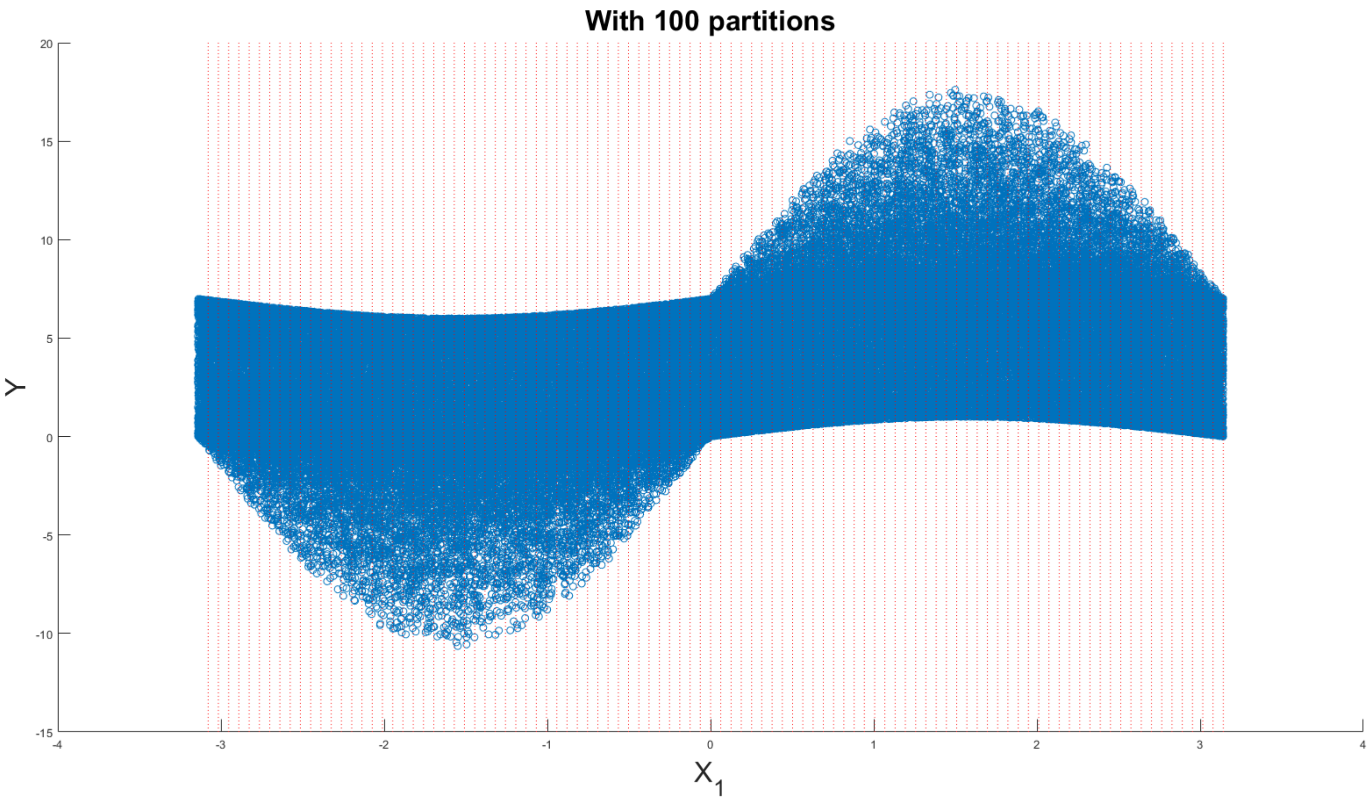


Partition the horizontal axis



3) Example 10 Partitions





With 100 partitions
Y
X1
20
15
10
5
0
-5
-10
-15
-4
-3
-2
-1
0
1
2
3
4

Given Data Estimation Theorem

1. *If* X *is finite then*

$$\lim_{N\to\infty} \xi_N^K = \xi^K(Y, X). \tag{44}$$

2. *In the general case, if* (42) *holds true,*

$$\lim_{M\to\infty} \lim_{N\to\infty} \xi_{M,N}^K = \xi^K(Y, X). \tag{45}$$

Experiments

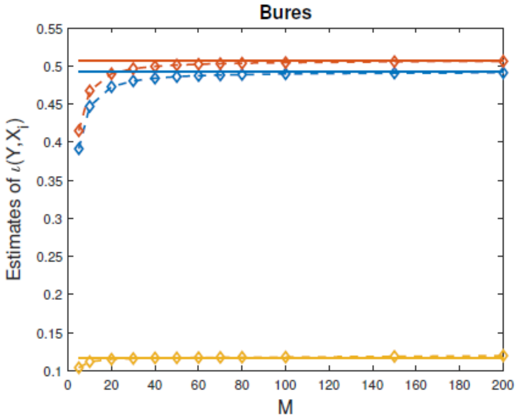


Figure 2: $\hat{\iota}(Y, X_i)$ for the multivariate-output analytical test case. The sample size is fixed at $N = 50000$. On the horizontal axis, partition cardinalities vary from $M = 5$ to $M = 200$.

Università Bocconi MILANO

The Given Data OT-problem

Equation (46) implies the solution of an OT-problem for each of the M partitions. If the cost function is the squared Wasserstein metric, we need to solve:

$$\begin{gathered}\inf_{s} \sum_{k=1}^{N} \sum_{j: x_{j,i} \in \mathbb{X}_{\alpha}^{m}(N)} s_{k,j} \sum_{t=1}^{n_Y} (y_{k,t} - y_{j,t})^2 \\ \text{subject to} \\ \sum_{k=1}^{N} s_{k,j} = \frac{1}{N}, \sum_{j: x_{j,i} \in \mathbb{X}^{m}(N)} s_{k,j} = \frac{1}{N_m}, \quad N_m = \#\{j : x_{j,i} \in \mathbb{X}^{m}(N)\},\end{gathered} \tag{47}$$

Results for Alternative Solvers

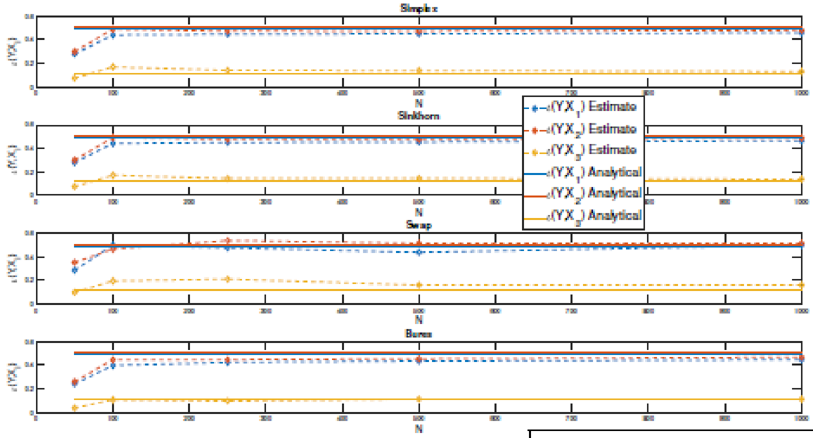
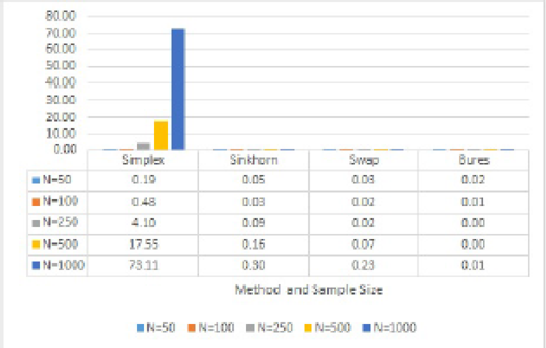


fig:MultivGaussTime

(a) $\hat{\iota}(Y, X_i)$



	Simplex	Sinkhorn	Swap	Bures
N=50	0.19	0.05	0.08	0.02
N=100	0.48	0.03	0.02	0.01
N=250	4.10	0.09	0.02	0.00
N=500	17.55	0.16	0.07	0.00
N=1000	73.11	0.30	0.23	0.01

(b) Times in seconds.

Figure 1: Estimates and times for experiments with sample sizes from $N = 50$ to $N = 1,000$ for four different algorithmic estimators of $\iota(Y, X_i)$ for the multivariate-output analytical test case used in this section.

Conclusions

- New Family of Indices for Multivariate Responses based on the theory of optimal transport
- Convenient Properties
- Generalize the previous multivariate indices of Lamboni (2011) and Gamboa et al (2014)
- Experiments Carried out thus far show promising results
- OT is a topical subject in AI, recent works propose breakthroughs in solution algorithms with interesting application perspectives