Optimal-Transport-based Sensitivity Measures and Their Computation

Elmar Plischke
Joint Work with E. Borgonovo, G. Savarè, A. Figalli

Institut für Endlagerforschung
TU Clausthal
UQSay, Oct 7, 2021

Background

- The pervasive use of scientific simulators requires approaches for providing transparent accounts of uncertainties [Saltelli et al., 2020]
- Most sensitivity analysis methods are built for univariate responses [Marrel et al., 2017]
- Recent research on multivariate responses based on variance-based indices [Gamboa et al., 2014]
- Optimal transport: Similarity measure based on a combinatorical optimization problem
- Our Goal: a comprehensive and theory-based approach

Contents

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

Wasserstein Sensitivity

Estimation

Examples

Discussion

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

Wasserstein Sensitivity

Estimation

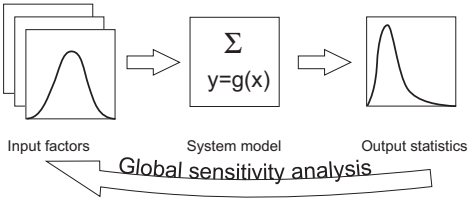
Examples

Discussion

Global Sensitivity Analysis

Saltelli in [Faivre et al., 2013]: "Sensitivity analysis is foremost about exploring the space of input assumptions in such a way as to be able to map the inference to the assumptions in a transparent way."

Iooss and Saltelli in [Ghanem et al., 2017]: "Sensitivity analysis provides users of mathematical and simulation models with tools to appreciate the dependency of the model output from model input and to investigate how important is each model input in determining its output."



Many Methods Available for Sensitivity Analysis

Different classes require finer detail of describing uncertainty in the input parameters [Borgonovo and Plischke, 2016]:

- Local Methods: Behaviour of the model about a nominal working/reference point
- Screening Methods: Behaviour of the system within given parameter ranges
- Global or Probabilistic Methods: Behaviour under given input parameter distributions

Many Methods Available for Sensitivity Analysis

Different classes require finer detail of describing uncertainty in the input parameters [Borgonovo and Plischke, 2016]:

- Local Methods: Behaviour of the model about a nominal working/reference point
- Screening Methods: Behaviour of the system within given parameter ranges
- Global or Probabilistic Methods: Behaviour under given input parameter distributions

Probabilistic Sensitivity Analysis: A General Framework

$(\Omega, \mathcal{A}, \mathbb{P})$	P-space
g	$\mathbb{R}^d \to \mathbb{R}^m$ simulation model (deterministic simulator, black-box)
X	d-dim. random vector, input uncertainty (known distribution)
$Y = g(X)$	m-dim. random vector, output uncertainty (unknown distribution)
α	index set in $[d] = \{1, \ldots, d\}$, factors of interest X_α
$\zeta(\cdot, \cdot)$	distance measure between P-measures on $\mathbb{R}^m$

General Probabilistic Sensitivity Measure [Borgonovo et al., 2016]

Given a shift/separation measure ζ,

$$\xi_\alpha(Y) = \mathbb{E}\left[\zeta\left(\mathbb{P}_Y, \mathbb{P}_{Y|X_\alpha}\right)\right]$$

Answers the question: What's the average value of learning that X_α is located at x_α?

Examples for shift/separation measures

For scalar output ($m = 1$):

$$\zeta_{\mathrm{CR}}(\mu_Y, \mu_{Y|X}) = \sigma_Y^{-2}(\mu_Y - \mu_{Y|X})^2$$ Sobol effects

$$\zeta_{\mathrm{KS}}(F_Y, F_{Y|X}) = \sup \left| F_Y - F_{Y|X} \right|$$ Kolmogorov–Smirnov

$$\zeta_{\mathrm{Gi}}(F_Y, F_{Y|X}) = \int \left(F_{Y|X}(y) - F_Y(y) \right)^2 dy$$ Gini Mean Distance, L^2 (cdf)

$$\zeta_{\mathrm{Bo}}(f_Y, f_{Y|X}) = \frac{1}{2} \int \left| f_{Y|X}(y) - f_Y(y) \right| dy$$ Borgonovo, L^1 (pdf)

$$\zeta_{\mathrm{KL}}(f_Y, f_{Y|X}) = \int f_{Y|X}(y) \log \frac{f_{Y|X}(y)}{f_Y(y)} dy$$ Kullback–Leibler

$$\zeta_{\mathrm{He}}(f_Y, f_{Y|X}) = 1 - \int \sqrt{f_Y(y) \cdot f_{Y|X}(y)} dy$$ Hellinger

Examples II

Shift/separation measures are considered between

- point estimators (mean, median)
- cumulative distribution functions [Gamboa et al., 2018]
- quantile functions
  [Fort et al., 2016, Elie-Dit-Cosaque and Maume-Deschamps, 2021, Kala, 2021]
- probabilistic distribution functions [Borgonovo, 2007]
- characteristic functions

Examples II

Shift/separation measures are considered between

- point estimators (mean, median)
- cumulative distribution functions [Gamboa et al., 2018]
- quantile functions
  [Fort et al., 2016, Elie-Dit-Cosaque and Maume-Deschamps, 2021, Kala, 2021]
- probabilistic distribution functions [Borgonovo, 2007]
- characteristic functions

Average Quantile Sensitivity

p-Wasserstein Sensitivity

For $m = 1$, $p \geq 1$ consider average L^p distance of quantiles

$$\xi_\alpha^{W_p}(Y) = \mathbb{E}\left[\sqrt[p]{\int_0^1 \left|Q_Y(u) - Q_{Y|X_\alpha}(u)\right|^p du}\right]$$

where $Q_Y(u) = \inf\{y : F_Y(y) \geq u\}$, $F_Y(y) = \mathbb{P}[Y \leq y]$

For $p = 1$,

$$\xi_\alpha^{W_1}(Y) = \mathbb{E}\left[\int_0^1 |Q_Y(u) - Q_{Y|X_\alpha}(u)| du\right] = \mathbb{E}\left[\int_{\mathbb{R}} |F_Y(y) - F_{Y|X_\alpha}(y)| dy\right]$$

Average Quantile Sensitivity

p-Wasserstein Sensitivity

For $m = 1$, $p \geq 1$ consider average L^p distance of quantiles

$$\xi_\alpha^{W_p}(Y) = \mathbb{E}\left[\sqrt[p]{\int_0^1 \left|Q_Y(u) - Q_{Y|X_\alpha}(u)\right|^p du}\right]$$

where $Q_Y(u) = \inf\{y : F_Y(y) \geq u\}$, $F_Y(y) = \mathbb{P}[Y \leq y]$

For $p = 1$,

$$\xi_\alpha^{W_1}(Y) = \mathbb{E}\left[\int_0^1 |Q_Y(u) - Q_{Y|X_\alpha}(u)| du\right] = \mathbb{E}\left[\int_{\mathbb{R}} |F_Y(y) - F_{Y|X_\alpha}(y)| dy\right]$$

For $m > 1$: generalization to the multivariate case is needed!

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

Wasserstein Sensitivity

Estimation

Examples

Discussion

Optimal Transport Problem

Kantorovich Formulation of Optimal Transport Problem

Let ν, ν' be two Borel P-measures on $\mathcal{Y}$ and $\mathcal{Z}$, respectively.

$$K(\nu, \nu') = \inf_{\pi \in \Pi(\nu, \nu')} \int c(y, z) d\pi(y, z)$$

where $\Pi(\nu, \nu')$ is the set of transport plans or couplings (P-measures on $\mathcal{Y} \times \mathcal{Z}$) whose marginals are ν and ν', respectively.

Wasserstein Distances

For $\mathcal{Y} = \mathcal{Z}$, let d be a metric and consider $c = d^p$. Then

$$W_p(\nu, \nu') = \sqrt[p]{K(\nu, \nu')} = \inf_{\pi \in \Pi(\nu, \nu')} \sqrt[p]{\int d(y, z)^p d\pi(y, x)}$$

is a distance called the p-Wasserstein distance [Villani, 2009].

Optimal Transport as Linear Programming Problem

For discrete measures, $\nu = \sum_{i=1}^{n} \nu_i \delta_{y_i}$ and $\nu' = \sum_{j=1}^{n'} \nu'_j \delta_{z_j}$ we get

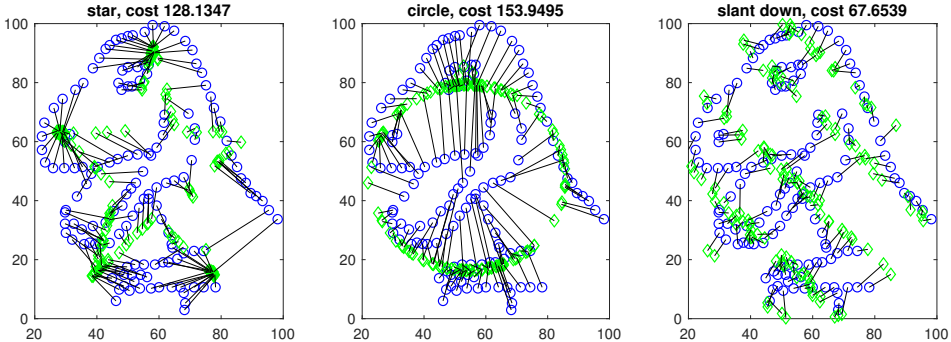
$$K(\nu, \nu') = \min \left\{ \sum_{i,j} c(y_i, z_j) p_{ij}, \qquad \sum_i p_{ij} = \nu'_j, \sum_j p_{ij} = \nu_i \right\}$$

This linear program can be solved via

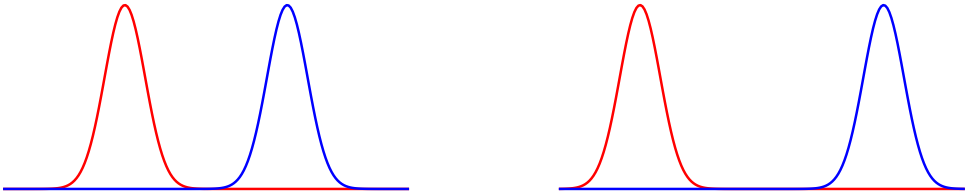
- Transport simplex algorithm
- Hungarian algorithm for the linear assignment problem (LAP) ($n = n'$, $\nu_i = \nu'_j = 1$)

For LAP, minimizing matrix P is a permutation

2D Example: Transporting the Datasaurus to different shapes



Revisiting the one dimensional case



- The intuition "red is further away from blue in the right panel compared to the left panel" works with Wasserstein, but not with cdf- or pdf-based distances
- Optimal transport plans to and from the $[0, 1]$ Lebesgue measure (standard uniform) are realized by quantile function and the cumulative distribution function: Wasserstein distances can be written compactly as seen on slide 10

Analytical solutions for the 2-Wasserstein distance

Consider $p = 2$. Let $\nu = \mathcal{N}(\mu, \Sigma)$ and $\nu' = \mathcal{N}(\mu', \Sigma')$ be two Gaussian distributions on $\mathbb{R}^m$ and let $d(y, z) = \|y - x\|_2$ be the Euclidean distance. Then the squared 2-Wasserstein distance is given by Wasserstein-Bures [Givens and Shortt, 1984]

$$W_2^2(\nu, \nu') = WB^2(\nu, \nu') = \|\mu - \mu'\|_2^2 + \text{trace}\left\{\Sigma + \Sigma' - 2\left(\Sigma^{1/2}\Sigma'\Sigma^{1/2}\right)^{1/2}\right\}.$$

Note the split into

- advective part, including the "centers of gravity" μ and μ'
- diffusive part, using the covariance information

The Entropic Problem

Regularize the Kantorovich problem through a penalty term [Cuturi, 2013]

$$\mathrm{KL}(\pi|\vartheta) = \int \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\vartheta}\right) \mathrm{d}\pi, \quad \text{if } \pi \ll \vartheta$$

using the Kullback-Leibler entropy of π w.r.t. a reference measure ϑ. For $\vartheta = \mu \otimes \nu$

$$S_\lambda(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \mathcal{C}(\pi) + \lambda \mathrm{KL}(\pi|\mu \otimes \nu).$$

The Entropic Problem

Regularize the Kantorovich problem through a penalty term [Cuturi, 2013]

$$\mathrm{KL}(\pi|\vartheta) = \int \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\vartheta}\right) \mathrm{d}\pi, \quad \text{if } \pi \ll \vartheta$$

using the Kullback-Leibler entropy of π w.r.t. a reference measure ϑ. For $\vartheta = \mu \otimes \nu$

$$S_\lambda(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \mathcal{C}(\pi) + \lambda \mathrm{KL}(\pi|\mu \otimes \nu).$$

Smooth convex problem that now can be solved with gradient-based methods / Sinkhorn-Knopp algorithm

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

Wasserstein Sensitivity

Estimation

Examples

Discussion

OT-based Probabilistic Sensitivity Measures

Definition

We call

$$\xi_\alpha^K = \mathbb{E}[\zeta^K(\mathbb{P}_Y, \mathbb{P}_{Y|X_\alpha})] = \mathbb{E}\left[\inf_{\pi \in \Pi(\mathbb{P}_Y, \mathbb{P}_{Y|X_\alpha})} \int c d\pi\right]$$

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

- ξ_α^K: expected amount of work needed to move from the current state of information about X to the state in which we learn the value of X_α.
- Most important input (group) X_α is associated with the highest expected amount of work to move the unconditional distribution of Y to the conditional distribution of Y once X_α is fixed.

Properties

[Borgonovo et al., 2021] prove the following:

Theorem

- *If* ξ_α^K *is finite then* $\xi_\alpha^K \geq 0$ *and* $\xi_\alpha^K = 0$ *if and only if* Y *and* X_α *are statistically independent.*
- *If* $\xi_\alpha^{W_p}$ *and* $\xi_\alpha^{W_q}$ *are finite,* $\xi_\alpha^{W_p} \leq \xi_\alpha^{W_q}$ *if* $p < q$*;*
- *Let* $\mathcal{Y}$ *be equipped with the discrete metric, i.e., a metric such that for all* $y, y' \in \mathcal{Y}$*,* $\mathrm{d}(y, y') = 0$ *if* $y = y'$ *and* $\mathrm{d} = 1$ *if* $y \neq y'$*, then* $\xi_\alpha^K = \xi_\alpha^{L^1}$*.*

Wasserstein-Bures Sensitivity

We may use the squared Wasserstein-Bures distance in a sensitivity measure, using means and empirical variance-covariance matrices as estimator inputs (taking the symmetric square root from a singular value decomposition avoids issues with non-pos. definiteness)

$$\xi^{\mathrm{WB}^2} = \mathbb{E}\left(\|\hat{\mu}_Y - \hat{\mu}_{Y|X_\alpha}\|^2 + \mathrm{trace}\left(\widehat{\Sigma}_Y + \widehat{\Sigma}_{Y|X_\alpha} - 2\left(\widehat{\Sigma}_Y^{1/2}\widehat{\Sigma}_{Y|X_\alpha}\widehat{\Sigma}_Y^{1/2}\right)^{1/2}\right)\right)$$

The advective part from the squared Wasserstein-Bures sensitivity coincides with point-wise variance-based Sobol' indices (unnormalized), a multivariate sensitivity measure sugested by [Gamboa et al., 2014].

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

Wasserstein Sensitivity

Estimation

Examples

Discussion

Estimators for Sensitivity Measures in the Common Rationale

$$\xi_\alpha(Y) = \mathbb{E}\left[\zeta\left(\mathbb{P}_Y, \mathbb{P}_{Y|X_\alpha}\right)\right]$$

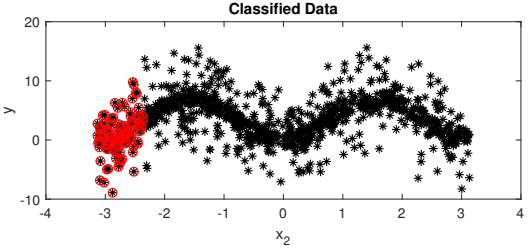
Formula suggests a double loop approach

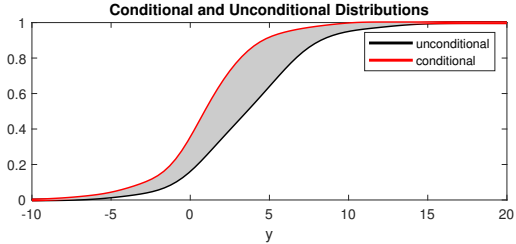
- Sample X_α, yielding realization x_α
  - Sample conditionally with respect to x_α being given
  - Compute the conditional distance, using e.g. kernel density estimators
- Average over the conditional distances

Requires the computational model g to be available (and quickly evaluated)

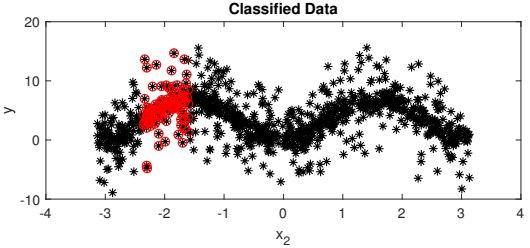
Replace point condition: $X_\alpha = x_\alpha$ by an input binning approach: $X_\alpha \in C_\alpha^q$ where $\{C_\alpha^q, q = 1, \dots Q\}$ forms a partition of the support of X_α

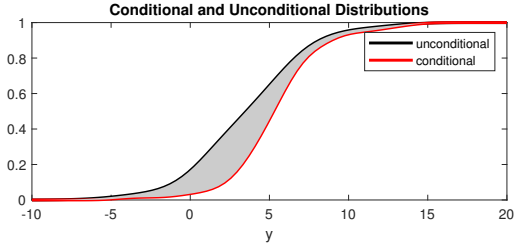
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)



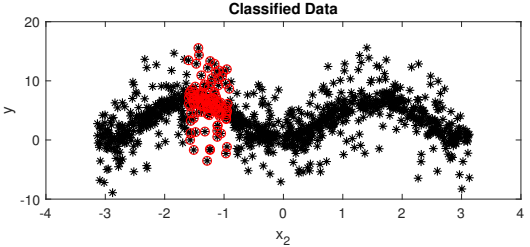


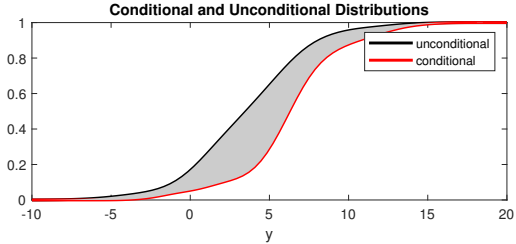
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)



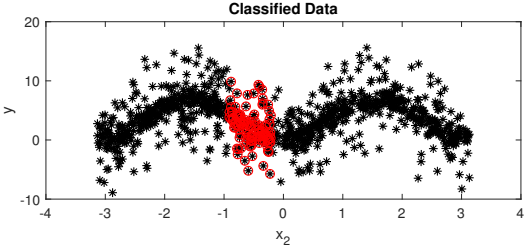


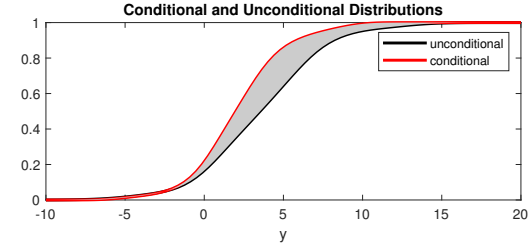
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)



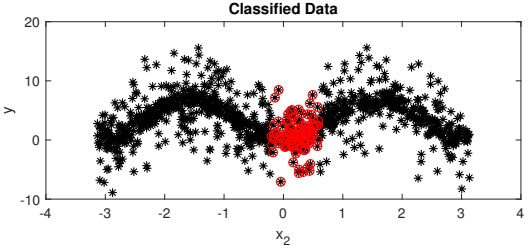


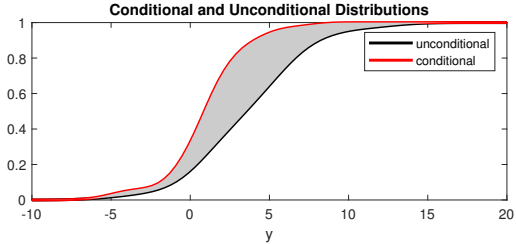
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)



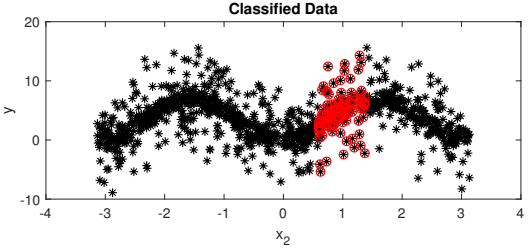


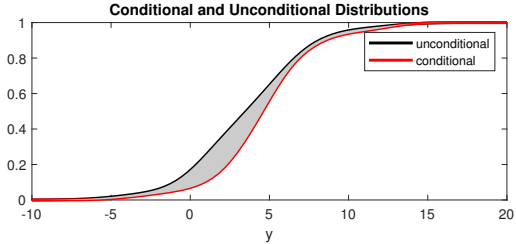
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)



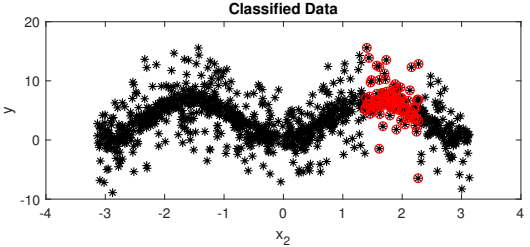


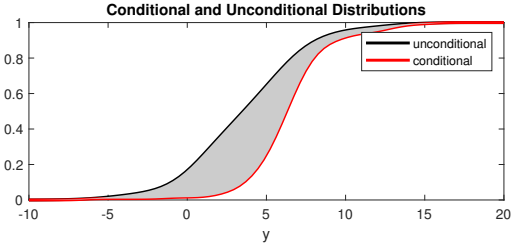
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)



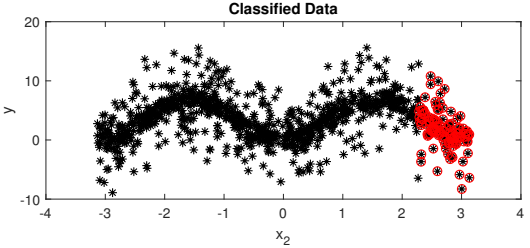


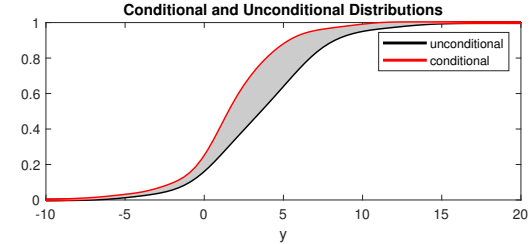
Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)





Example: $m = 1$, $p = 1$ (1-Wasserstein is area between cdfs)





Multivariate Output

For each input $i = 1, \ldots, d$ and each partition interval $q = 1, \ldots, Q$ we need to solve an OT problem with $n \gg n'$.

Simplex algorithm is of non-polynomial running time (worst case visits all corners)

Multivariate Output

For each input $i = 1, \ldots, d$ and each partition interval $q = 1, \ldots, Q$ we need to solve an OT problem with $n \gg n'$.

Trade in precision for speed:

- Use methods for LAP ($n = n'$) by
  - upscaling: reusing the conditional output multiple times, until unconditional sample size is reached
  - downscaling: randomly picking a subset from the unconditional output (same size as the conditional sample) to compare with the conditional output

  ([Puccetti, 2017] lifts the idea of reordering to the multidim. case)

Multivariate Output

For each input $i = 1, \ldots, d$ and each partition interval $q = 1, \ldots, Q$ we need to solve an OT problem with $n \gg n'$.

Trade in precision for speed:

- Use methods for LAP ($n = n'$) by
  - upscaling: reusing the conditional output multiple times, until unconditional sample size is reached
  - downscaling: randomly picking a subset from the unconditional output (same size as the conditional sample) to compare with the conditional output

  ([Puccetti, 2017] lifts the idea of reordering to the multidim. case)
- Solution of entropic problem: $\lambda \to 0$ regains OT (numerical instability!)

Multivariate Output

For each input $i = 1, \ldots, d$ and each partition interval $q = 1, \ldots, Q$ we need to solve an OT problem with $n \gg n'$.

Trade in precision for speed:

- Use methods for LAP ($n = n'$) by
  - upscaling: reusing the conditional output multiple times, until unconditional sample size is reached
  - downscaling: randomly picking a subset from the unconditional output (same size as the conditional sample) to compare with the conditional output

  ([Puccetti, 2017] lifts the idea of reordering to the multidim. case)
- Solution of entropic problem: $\lambda \to 0$ regains OT (numerical instability!)
- Use Wasserstein-Bures formula, ignoring non-normality

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

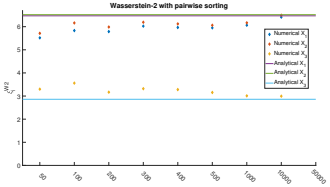
Wasserstein Sensitivity

Estimation

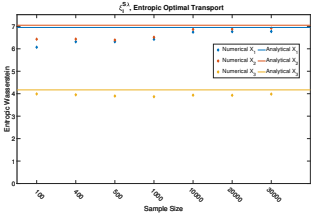
Examples

Discussion

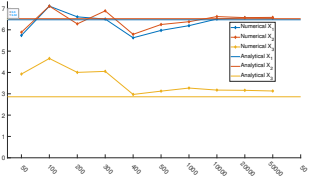
Multivariate Gaussian linear: analytical solution known from Wasserstein-Bures



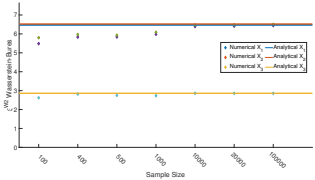
sorting algorithm $\hat{\xi}_i^{W2}$



entropic OT with stabilization $\hat{\xi}_i^{S,\lambda}$



entropic OT $\hat{\xi}_i^{S,\lambda}$



Wasserstein-Bures $\hat{\xi}_i^{WB}$

Computational Aspects

The estimation implies the sequential solution of optimal transport problems: computationally demanding

- Reordering takes 11 hours at $N = 10,000$
- Sinkhorn without numerical stabilization takes about 33 seconds, while solving 66 optimization problems
- Sinkhorn with numerical stabilization takes about 30 minutes
- Wasserstein-Bures takes a few seconds

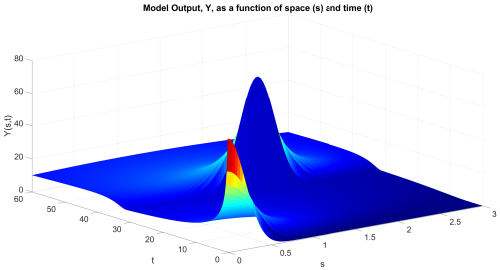
Computational Aspects

The estimation implies the sequential solution of optimal transport problems: computationally demanding

- Reordering takes 11 hours at $N = 10,000$
- Sinkhorn without numerical stabilization takes about 33 seconds, while solving 66 optimization problems
- Sinkhorn with numerical stabilization takes about 30 minutes
- Wasserstein-Bures takes a few seconds

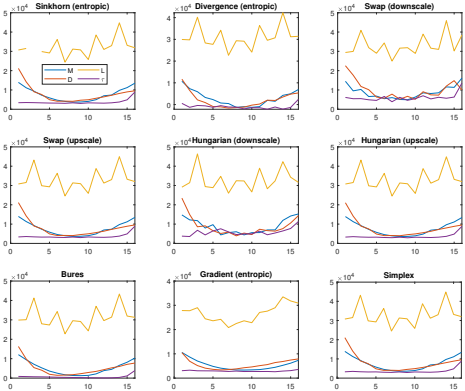
Results from June 2021. But: OT is fast evolving with new algorithms by the hour

Environmental example: pollutant transport, with response in space and time



$d = 4$ inputs, $m = 5 \cdot 200$ outputs, 2048 Quasi-Monte Carlo realizations, $Q = 16$ partitions

Results: Regional sensitivities $x \mapsto \zeta^K(\mathbb{P}_Y, \mathbb{P}_{Y|X_i=x})$



All methods perform comparably well (Sinkhorn with convergence holes)

Results: Timings (1024 QMC, $Q = 16$) in s

Sinkhorn 1.78	Divergence 51.78	Swap (downscale) 2.09
Swap (upscale) 450.1	Hungarian (downscale) 3.36	Hungarian (upscale) 270.8
Bures 26.51	Gradient 6.33	Simplex 104.8

A Common Rationale for Sensitivity Measures

Optimal Transport and the Wasserstein Distance

Wasserstein Sensitivity

Estimation

Examples

Discussion

Some Open Questions

- Balance between sample size and partition size
- Downsampling and bootstrapping
- Choice of output norm underlying the Wasserstein distance
- Benefits of Quasi Monte Carlo sampling compared to crude Monte Carlo sampling
- Pick'n'freeze design for computing OT-based sensitivities
- Error from the Wasserstein-Bures approximation
- ...

Conclusions

- New framework for the global sensitivity analysis of multivariate output
- Family of sensitivity measures: Wasserstein, Sinkhorn Divergence, Wasserstein-Bures
- Exact solution challenges current best practices
- Approximate solutions of OT actively researched in Machine Learning and Artificial Intelligence
- This work opens opportunities for cross fertilization and future research

Thank You!

Questions, Comments

mailto:elmar.plischke@tu-clausthal.de

Preprints, Scripts, Stuff

https://artefakte.rz-housing.tu-clausthal.de/epl/

References I

Borgonovo, E. (2007).
A new uncertainty importance measure.
Reliability Engineering&System Safety, 92(6):771–784.

Borgonovo, E., Figalli, A., Plischke, E., and Savarè, G. (2021).
Probabilistic sensitivity with optimal transport.
Under Review.

Borgonovo, E., Hazen, G. B., and Plischke, E. (2016).
A common rationale for global sensitivity measures and their estimation.
Risk Analysis, 36(10):1871–1895.

Borgonovo, E. and Plischke, E. (2016).
Sensitivity analysis: A review of recent advances.
European Journal of Operational Research, 248(3):869–887.

Cuturi, M. (2013).
Sinkhorn distances: Lightspeed computation of optimal transportation distances.
Advances in Neural Information Processing Systems, 26:2292–2300.

Elie-Dit-Cosaque, K. and Maume-Deschamps, V. (2021).
On quantile oriented sensitivity analysis.
preprint, Hyper articles en ligne.
hal-03137402.

References II

Faivre, R., Iooss, B., Mahévas, S., Makowski, D., and Monod, H., editors (2013).
Analyse de sensibilité et exploration de modèles.
Éditions Quæ, Versailles, France.

Fort, J.-C., Klein, T., and Rachdi, N. (2016).
New sensitivity analysis subordinated to a contrast.
Communications in Statistics - Theory and Methods, 45(15):4349–4364.

Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014).
Sensitivity analysis for multidimensional and functional outputs.
Electronic Journal of Statistics, 8:573–603.

Gamboa, F., Klein, T., and Lagnoux, A. (2018).
Sensitivity analysis based on Cramér–von Mises distance.
SIAM/ASA J. Uncertainty Quantification, 6(2):522–548.

Ghanem, R., Higdon, D., and Owhadi, H., editors (2017).
Handbook of Uncertainty Quantification.
Springer Verlag.

Givens, C. R. and Shortt, R. M. (1984).
A class of Wasserstein metrics for probability distributions.
Michigan Math. J., 31:231–240.

References III

Kala, Z. (2021).
Global sensitivity analysis of quantiles: New importance measure based on superquantiles and subquantiles.
Symmetry, 13(2):263.

Marrel, A., Saint-Geours, N., and De Lozzo, M. (2017).
Sensitivity analysis of spatial and/or temporal phenomena.
In *Handbook of Uncertainty Quantification*, pages 1327–1357. Springer Verlag, Cham.

Puccetti, G. (2017).
An algorithm to approximate the optimal expected inner product of two vectors with given marginals.
Journal of Mathematical Analysis and Applications, 451(1):132–145.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Wendy Nelson Espeland, J. K., Lo Piano, S., Mayo, D., Pielke, Jr, R., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., Stirling, A., van der Sluijs, J., and Vineis, P. (2020).
Five ways to ensure that models serve society: a manifesto.
Nature, 582:482–484.

Villani, C. (2009).
Optimal Transport: Old and New.
Springer Verlag, Berlin.