Moment-Independent Sensitivity Measures

Elmar Plischke and Emanuele Borgonovo

Institut für Ressourcenökologie | Department of Decision Science
HZDR Dresden-Rossendorf | Uni Bocconi

SAMO Summer School, Parma, June 2024

Contents

Limits of Variance-Based Sensitivity Measures

Recall first order effect

$$S_i = \frac{\mathbb{V}[\mathbb{E}[Y|X_i]]}{\mathbb{V}[Y]} = \frac{\mathbb{E}\left[(\mathbb{E}[Y|X_i] - \mathbb{E}[Y])^2\right]}{\mathbb{E}[(Y - E[Y])^2]} = \frac{\mathbb{E}\left[\mathbb{E}[Y|X_i]^2\right] - \mathbb{E}[Y]^2}{\mathbb{E}[Y^2] - E[Y]^2}$$

What if ...

- ... difference between mean and conditional mean does not capture uncertainty (i.e. if constant or with bifurcations)?

Limits of Variance-Based Sensitivity Measures

Recall first order effect

$$S_i = \frac{\mathbb{V}[\mathbb{E}[Y|X_i]]}{\mathbb{V}[Y]} = \frac{\mathbb{E}\left[\left(\mathbb{E}[Y|X_i] - \mathbb{E}[Y]\right)^2\right]}{\mathbb{E}[(Y - E[Y])^2]} = \frac{\mathbb{E}\left[\mathbb{E}[Y|X_i]^2\right] - \mathbb{E}[Y]^2}{\mathbb{E}[Y^2] - E[Y]^2}$$

What if ...

- ... difference between mean and conditional mean does not capture uncertainty (i.e. if constant or with bifurcations)?
- ... variance (interpreted as squared Euclidean distance) is not the right measure (i.e. data spanning orders of magnitude)?

Limits of Variance-Based Sensitivity Measures

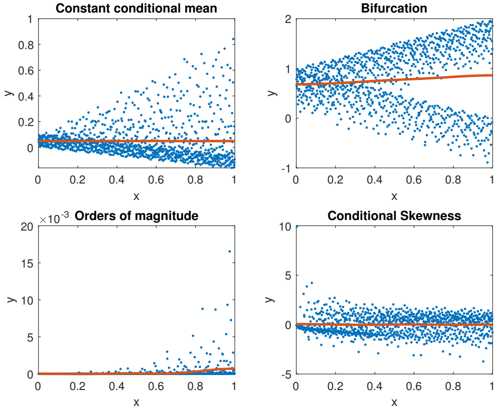
Recall first order effect

$$S_i = \frac{\mathbb{V}[\mathbb{E}[Y|X_i]]}{\mathbb{V}[Y]} = \frac{\mathbb{E}\left[\left(\mathbb{E}[Y|X_i] - \mathbb{E}[Y]\right)^2\right]}{\mathbb{E}[(Y - E[Y])^2]} = \frac{\mathbb{E}\left[\mathbb{E}[Y|X_i]^2\right] - \mathbb{E}[Y]^2}{\mathbb{E}[Y^2] - E[Y]^2}$$

What if ...

- ... difference between mean and conditional mean does not capture uncertainty (i.e. if constant or with bifurcations)?
- ... variance (interpreted as squared Euclidean distance) is not the right measure (i.e. data spanning orders of magnitude)?
- ... uncertainty is driven by skewness?

Examples: First order effects do not capture uncertainty



Common Rationale for Global Sensitivity Measures

Consider [Borgonovo et al., 2016]

$$\zeta_i = \mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i}) = \int_{\mathcal{X}_i} d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i=x_i}) p_i(x_i) dx_i$$

where d is a distance (dissimilarity measure) between probability measures, named separation / shift.

Answers the question

What is the average value of learning that X_α is located at x_α? (Expected value of perfect information)

The sensitivity measure ζ_i is

- a distance to independence if $d(\mathbb{P}, \mathbb{Q}) = 0 \iff \mathbb{P} = \mathbb{Q}$
- a dissimilarity measure to functional dependence if $d(\mathbb{P}, \mathbb{Q}) \leq d(\mathbb{P}, \delta)$ for measures $\mathbb{Q}$ and point-masses δ

Separation measurements

Shift/separation measures can be considered between

- cumulative distribution functions
- quantile functions
- probabilistic distribution functions
- characteristic functions
- point estimates (mean, median) – not moment-independent

Further statistics (median, mode, max, variance) on separations: [Pianosi and Wagener, 2015]

Examples for separation measures

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

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

$$d_{\mathsf{Ku}}(F_Y, F_{Y|X}) = \sup \left(F_Y - F_{Y|X} \right) - \inf \left(F_Y - F_{Y|X} \right)$$ Kuiper

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

$$d_{\mathsf{W}}(F_Y, F_{Y|X}) = \int_0^1 \left| F_Y^{-1}(v) - F_{Y|X}^{-1}(v) \right| \mathrm{d}v$$ Wasserstein, L^1 (qf)

$$d_{\mathsf{Bo}}(p_Y, p_{Y|X}) = \frac{1}{2} \int \left| p_{Y|X}(y) - p_Y(y) \right| dy$$ Borgonovo, L^1 (pdf)

$$d_{\mathsf{KL}}(p_Y, p_{Y|X}) = \int p_{Y|X}(y) \log \frac{p_{Y|X}(y)}{p_Y(y)} dy$$ Kullback–Leibler

$$d_{\mathsf{He}}(p_Y, p_{Y|X}) = 1 - \int \sqrt{p_Y(y) \cdot p_{Y|X}(y)} dy$$ Hellinger

Estimation of Measures in the Common Rationale

Formula $\mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})]$ suggests a double loop approach

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

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

Estimation of Measures in the Common Rationale

Formula $\mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})]$ suggests a double loop approach

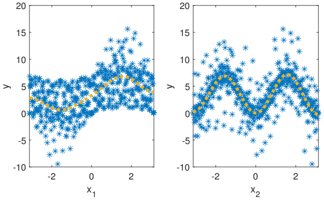
- Sample X_i, yielding realization x_i
  - Sample conditionally with respect to x_i being given
  - Compute the conditional distance, using e.g. kernel density estimators
- Average over the conditional distances

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

Replace point condition $X_i = x_i$ by an input binning approach: $X_i \in C_i^q$ where $\{C_i^q, q = 1, \ldots Q\}$ forms a partition of the support of X_i. Works also as a data-driven approach.

Recall: Sensitivity Information from Scatterplots

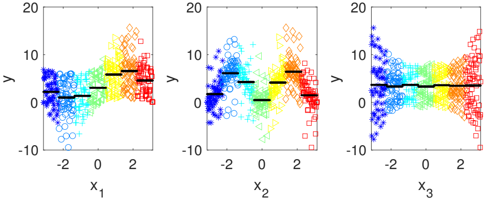
Variance-based first order effect as non-linear goodness-of-fit R^2:
Fraction of output variance explained by functional dependence on X_i



$$S_i = \frac{\mathbb{V}[\mathbb{E}[Y\,|\,X_i]]}{\mathbb{V}[Y]}$$

Recall: Sensitivity Information from Scatterplots

Variance-based first order effect as non-linear goodness-of-fit R^2:
Fraction of output variance explained by functional dependence on X_i

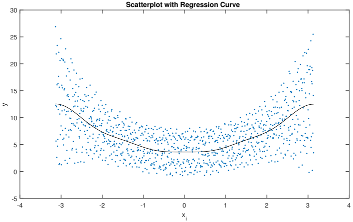


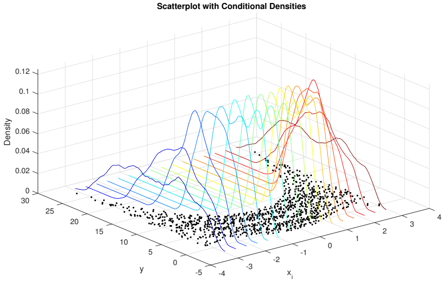
$$\widehat{S}_i = \frac{\sum_{j=1}^{n} (\mu_{Y|X_i=x_{ji}} - \mu_Y)^2}{\sum_{j=1}^{n} (y_j - \mu_Y)^2}$$

Conditioning on X_i: Local information from scatterplot (Pearson Correlation Ratio: Use piecewise constant approximation for $\mu_{Y|X_i}$)

Common Rationale revisited

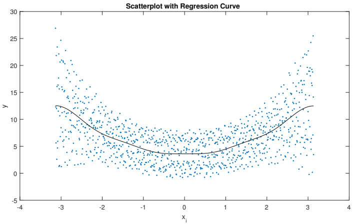
Replace $(\mu_{Y|X_i=x} - \mu_Y)^2$ by distance measures between Y and $Y|X_i = x$

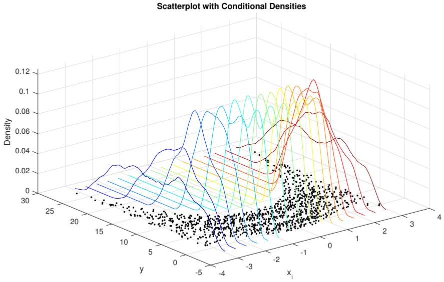




Common Rationale revisited

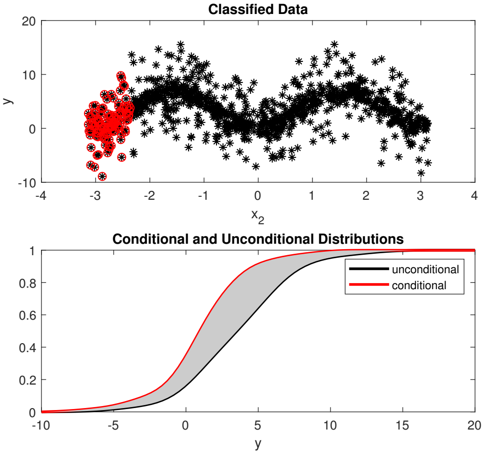
Replace $(\mu_{Y|X_i=x} - \mu_Y)^2$ by distance measures between Y and $Y|X_i = x$



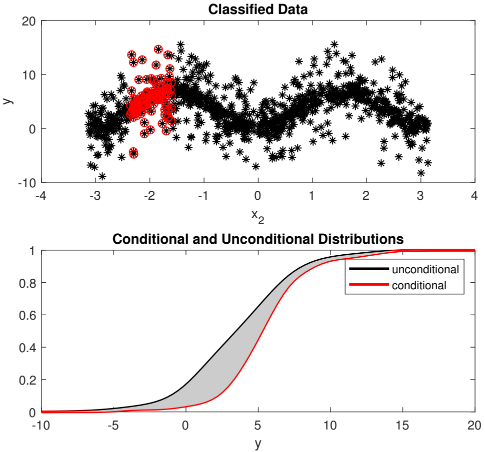


Estimation: Binning the scatterplot into vertical stripes
(interval conditional $Y|X_i \in \mathcal{C}_q^i$ instead of point conditional $Y|X_i = x$)

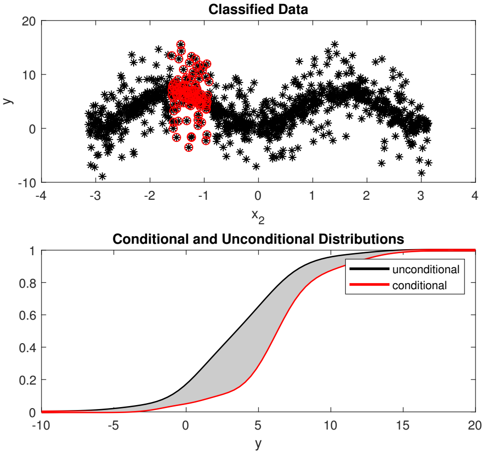
Example: cdf-based (area between curves)



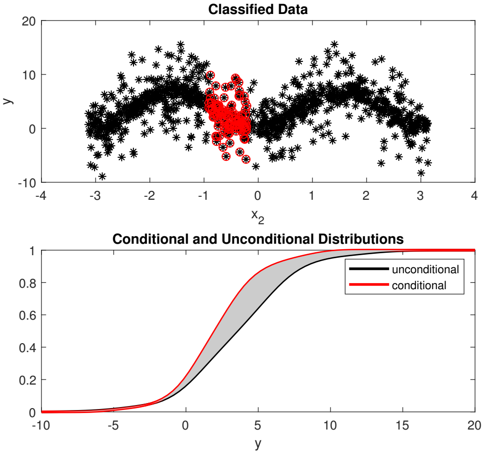
Example: cdf-based (area between curves)



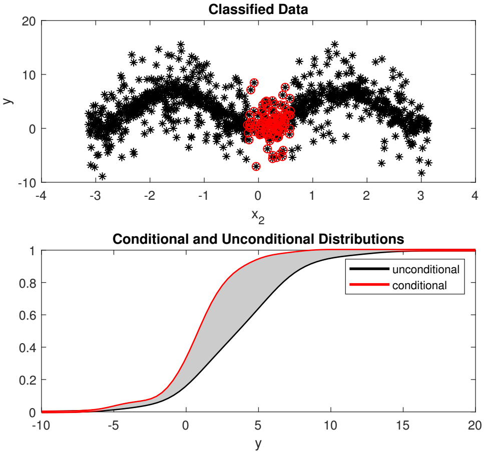
Example: cdf-based (area between curves)



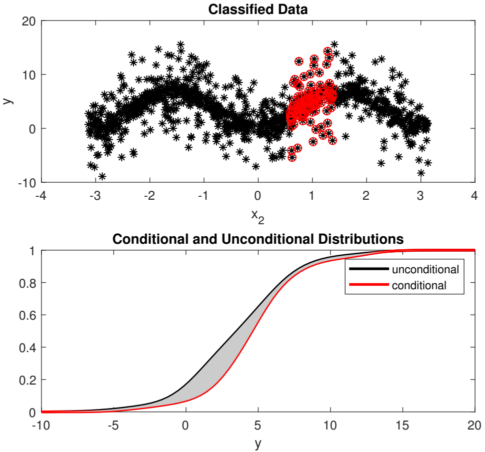
Example: cdf-based (area between curves)



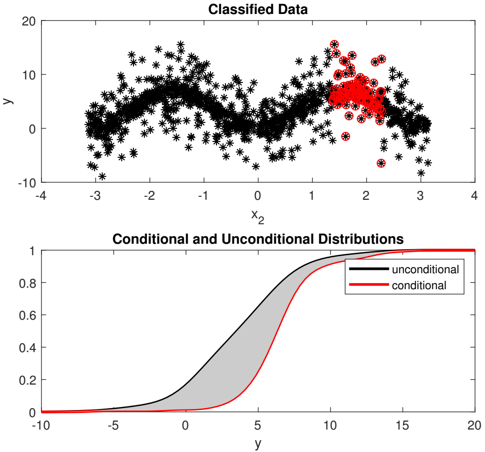
Example: cdf-based (area between curves)



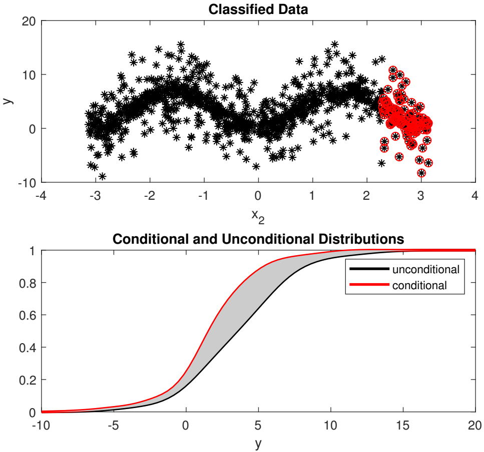
Example: cdf-based (area between curves)



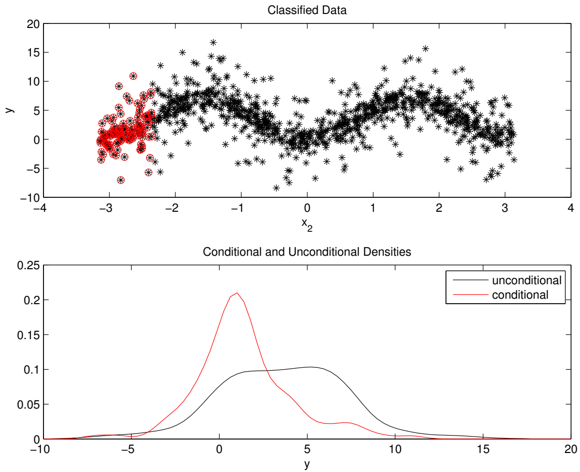
Example: cdf-based (area between curves)



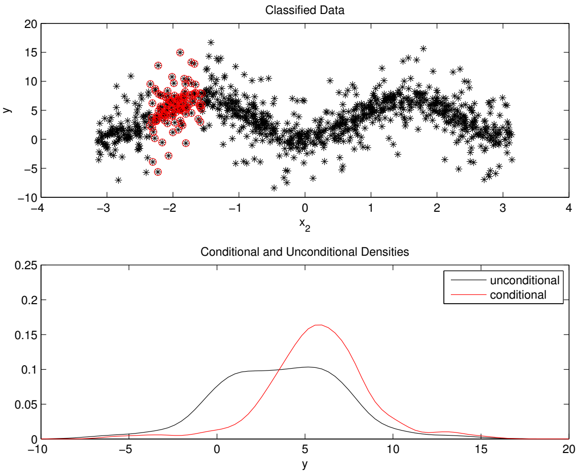
Example: cdf-based (area between curves)



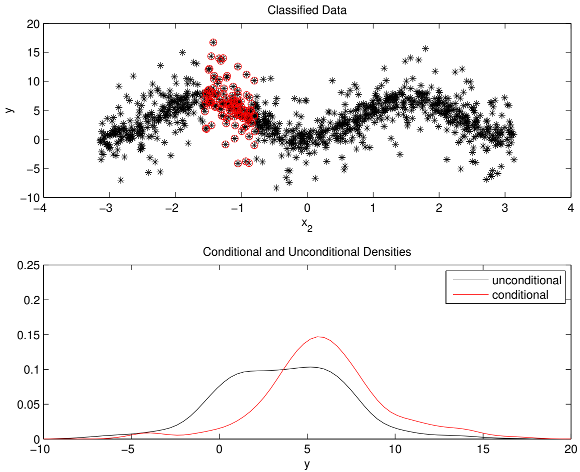
Example: pdf-based (area between curves)



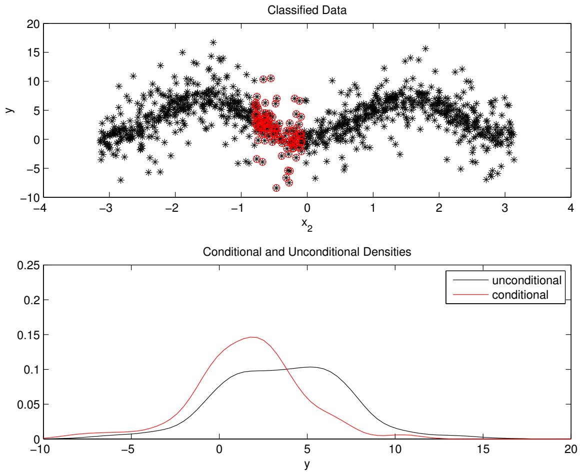
Example: pdf-based (area between curves)



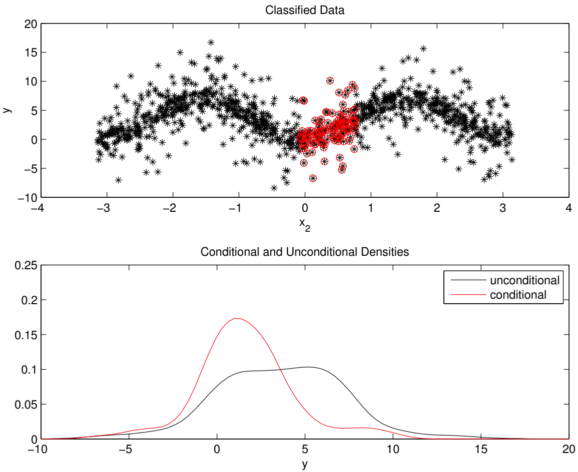
Example: pdf-based (area between curves)



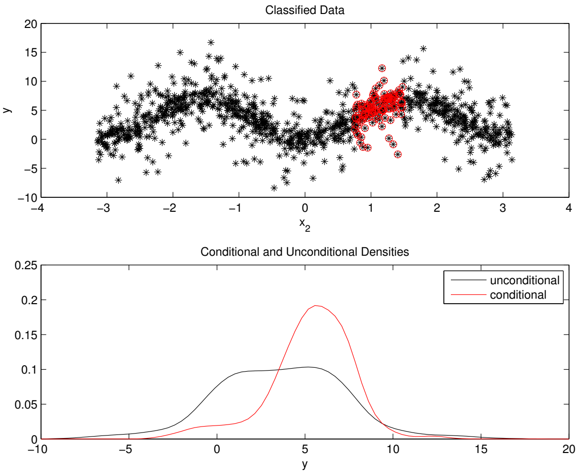
Example: pdf-based (area between curves)



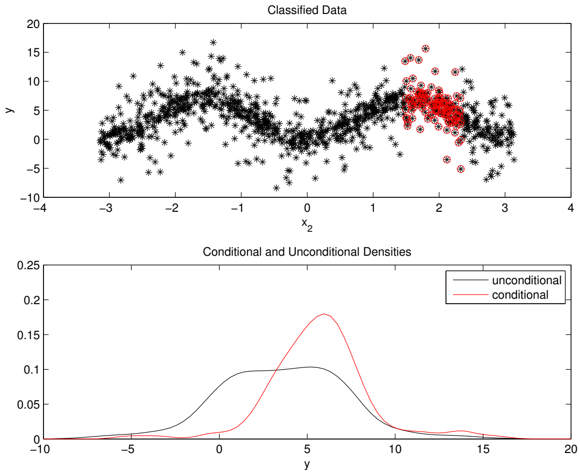
Example: pdf-based (area between curves)



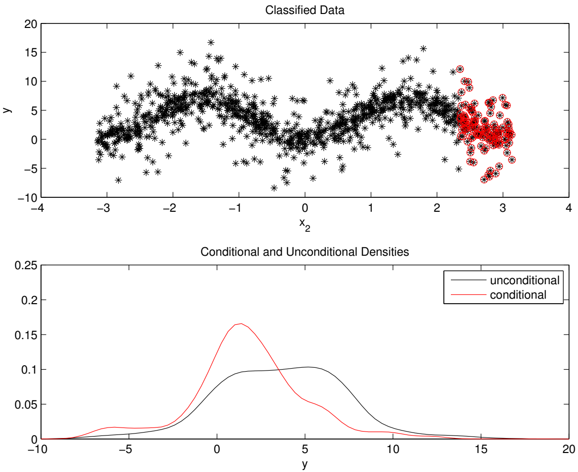
Example: pdf-based (area between curves)



Example: pdf-based (area between curves)



Example: pdf-based (area between curves)

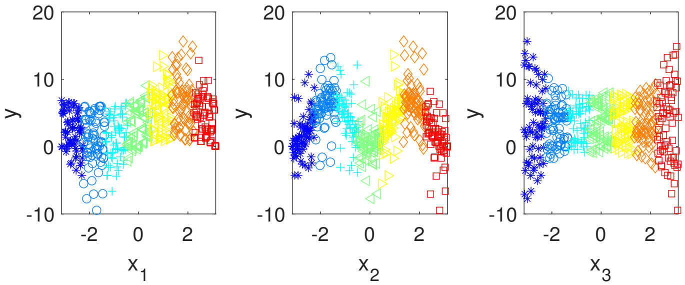


MatLab Implementation for $\delta = \mathbb{E}[d_{Bo}(Y, Y|X_i)]$ estimation

```
Kernel=@(x)3/(4*sqrt(5))*max(1-(x.^2/5),0);[n,k]=size(x);
mady=median(abs(median(y)-y));
% bandwidth estimate (rule of thumb)
stdy=min(std(y),mady/(2*0.675)); h=stdy*((4/(3*n))^(1/5));
z=linspace(min(y),max(y),100); % quadrature points
W=Kernel(bsxfun(@minus,z,y)/h)/h; densy=mean(W); %KDE
[xr,indxx]=sort(x);
for i=1:k;   xr(indxx(:,i),i)=1:n; end % ranks
for j=1:M
   indx=((j-1)*n/M<xr) & (xr ≤ j*n/M);   nm(:,j)=sum(indx);
   for i=1:k
   densc=mean(W(indx(:,i),:)); % conditional density
   Sm(i,j)=trapz(z,max(densy-densc,0)); %only positive part
end
Sm(Sm<Cutoff.*sqrt(1/n+1./nm))=0; d=sum(Sm.*nm,2)'/n;
```

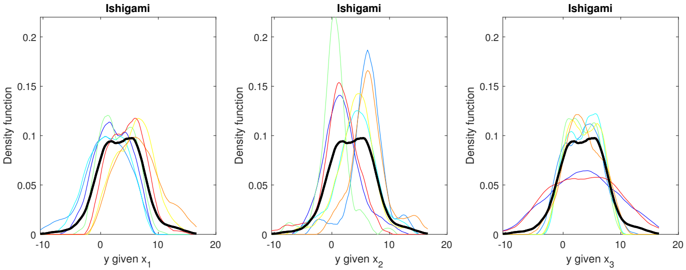
Conditional separation: Localized Information

Visualizing $x_i \mapsto d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i=x_i})$



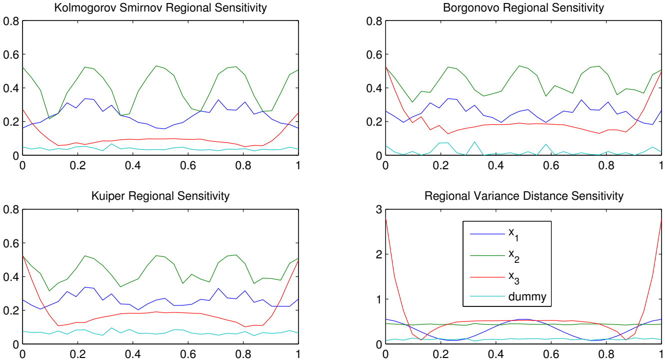
Conditional separation: Localized Information

Visualizing $x_i \mapsto d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i=x_i})$



Conditional separation: Localized Information

Visualizing $x_i \mapsto d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i=x_i})$



Sensitivity Measures from the Common Rationale

Still a large class. Some more structure?

- Transformation invariance

Sensitivity Measures from the Common Rationale

Still a large class. Some more structure?

- Transformation invariance
- Information value under a proper scoring rule

Sensitivity Measures from the Common Rationale

Still a large class. Some more structure?

- Transformation invariance
- Information value under a proper scoring rule
- Optimal transport based

Spearman Rank Correlation

Replace each value with its rank before computing the regression line

$$Rk(z) = \sum_{i=1}^{n} \mathbf{1}\{z_i \leq z\}$$

Almost the same as the empirical cdf

$$\hat{F}_Z(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{z_i \leq z\}$$

Counting the number of realisations smaller or equal to a given value

Generalization

Replacing regression line by regression curve: Transformation-invariant first order effect (equivalent to Kruskal-Wallis test)

Transformation Invariance

Copula Theory

Study of properties invariant under strictly monotonic transformations

Transformation-invariant moment-independent sensitivity measures capture copula properties.

Transformation Invariance

Copula Theory

Study of properties invariant under strictly monotonic transformations

Transformation-invariant moment-independent sensitivity measures capture copula properties.
Empirical CDFs

$$(x, y) \mapsto (u, v) \in [0,1]^2, u = \widehat{F}_X(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{x_j \leq x\}, v = \dots$$

Transformation of marginal distributions to uniform

Transformation Invariance

Copula Theory

Study of properties invariant under strictly monotonic transformations

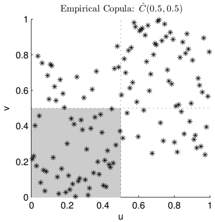
Transformation-invariant moment-independent sensitivity measures capture copula properties.
Empirical CDFs

$$(x, y) \mapsto (u, v) \in [0, 1]^2, u = \widehat{F}_X(x) = \tfrac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{x_j \le x\}, v = \ldots$$

Transformation of marginal distributions to uniform
Same idea in 2D: Empirical bivariate copula

$$\widehat{C}(u, v) = \tfrac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{u_j \le u\}\mathbf{1}\{v_j \le v\}$$

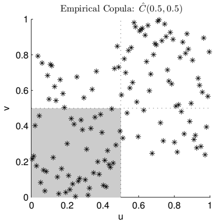
Sensitivity from the Empirical Copula



Count number of realizations inside the box spanned by (u, v) and the origin

Under Independence: $\widehat{C}(u, v)$ is a MC integral (relative frequency of hits) of the box area $u \cdot v$

Sensitivity from the Empirical Copula



Count number of realizations inside the box spanned by (u, v) and the origin

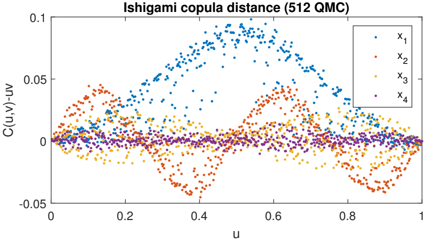
Under Independence: $\widehat{C}(u, v)$ is a MC integral (relative frequency of hits) of the box area $u \cdot v$
Hence: Compare the independent (product) copula with the empirical copula

$$\varphi : u \mapsto uv - \widehat{C}(u, v)$$

Visual tool

Copula distance plots: u vs. $\varphi(u)$ [Plischke and Borgonovo, 2019]

Ishigami Copula Distance: Quasi MC, 512 runs



Spearman ϱ^*	$12\int_u \varphi(u)du$	$(0.44, 0.00, 0.00, -0.00)$
Schweizer Wolff	$12\int_u \lvert\varphi(u)\rvert\, du$	$(0.44, 0.22, 0.09, 0.03)$
Discrepancy	$4\max_u \lvert\varphi(u)\rvert$	$(0.38, 0.18, 0.10, 0.04)$

Not in the Common Rationale

Transformation invariant sensitivity measure

Note that $u_i = F_i(x_i)$ and $v = F_Y(y)$, hence $du_i = p_i(x_i)dx_i$ and $dv = p_Y(y)dy$:

$$\begin{aligned}
\delta(Y, X_i) &= \frac{1}{2} \iint \left| p_{Y|X_i = x_i}(y) - p_Y(y) \right| p_i(x_i) dx_i dy \\
&= \frac{1}{2} \iint \left| p_{X_i,Y}(x_i, y) - p_i(x_i) p_Y(y) \right| dx_i dy \\
&= \frac{1}{2} \iint \left| \frac{p_{X_i,Y}\left(F_i^{-1}(u_i), F_Y^{-1}(v)\right)}{p_i(F_i^{-1}(u_i)) p_Y(F_Y^{-1}(y))} - 1 \right| du_i dv
\end{aligned}$$

But quotient is bivariate copula density: $c(u_i, v) = \frac{\partial^2 C(u_i, v)}{\partial u_i \partial v}$: $\delta(Y, X_i) = \frac{1}{2} \iint \left| c(u_i, v) - 1 \right| du_i dv$

Further transformation invariant sensitivity measures

If the separation is invariant with respect to monotonic transformation, so is its sensitivity measure in the common rationale.
Check sensitivity measures:

- Kullback–Leibler ($\iint c(u_i, v) \log c(u_i, v) du_i dv$)
- Hellinger ($\iint \sqrt{c(u_i, v)} du_i dv$)
- modified Gini measure (needs conditional copula)

$$\iint \left(F_{Y|X_i=x_i}(y) - F_Y(y)\right)^2 p_y(y) dy p_i(x_i) dx_i = \iint \left(v - F_{V|U_i=u_i}(v)\right)^2 du_i dv$$

- Kolmogorov–Smirnov and Kuiper

Estimating transformation-invariant sensitivity measures

An empirical (checkerboard) copula can be approximated by Bernstein polynomials (B-splines).
This family of polynomials is closed under differentiation, hence approximate conditional copula and copula density can be obtained analytically.
Bernstein polynomials offer a partition of unity.

Estimating Conditional Copula/Copula Density

Given a copula C, the Bernstein copula approximation of order m is given by [Sancetta and Satchell, 2004]

$$C_m(u,v) = \sum_{i=1}^{m}\sum_{j=1}^{m} C(\tfrac{i}{m},\tfrac{j}{m}) P_{mi}(u) P_{mj}(v)$$

with Bernstein polynomials $P_{mk}(t) = \binom{m}{k} t^k (1-t)^{m-k}$, $t \in [0,1]$.
Then for $C = \widehat{C}$

$$\widehat{C}_m(v|u) = \sum_i \sum_j \widehat{C}(\tfrac{i}{m},\tfrac{j}{m}) P'_{mi}(u) P_{mj}(v)$$

$$\widehat{c}_m(u,v) = \sum_i \sum_j \widehat{C}(\tfrac{i}{m},\tfrac{j}{m}) P'_{mi}(u) P'_{mj}(v)$$

Coditional copula, copula density

MatLab Implementation for Copula Estimation

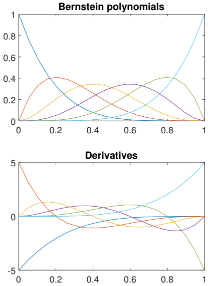
```
[n,d]=size(x); uv=empcdf([x,y]); m=100; % Bernstein polynomial order
u=linspace(1/(m+1),1,m); N=max(2*m,20); t=linspace(1/(2*N),1-1/(2*N),N)';%Grids
[F,F0]=bern(t,m);  Z=zeros(m+1,m+1);
for k=1:d
 for i=1:m,for j=1:m, Z(j+1,i+1)=sum( uv(:,k)<u(i) & uv(:,end)<u(j))/n;end;end
    L=F*Z*F'-t*t'; Q=F0*Z*F0'; R=F*Z*F0'; %c. minus indep., c.density, cond.c.
    S=R-t*ones(1,N); % cdf difference
    discrepancy(k)=4*max(max(abs(L))); rho(k)=12*mean(mean(L));
    eta(k)=12*mean( mean(R).^2)-3;  Δ(k)=mean(mean(abs(Q-1)))/2;
    smirnov(k)=mean(max(abs(S)));   kuiper(k)=mean(max(S)-min(S));
    gini2(k)=6*mean(mean(S.^2));    hellinger(k)=1-mean(mean(sqrt(max(Q,0))));
end
```

MatLab Implementation for the Bernstein polynomial basis

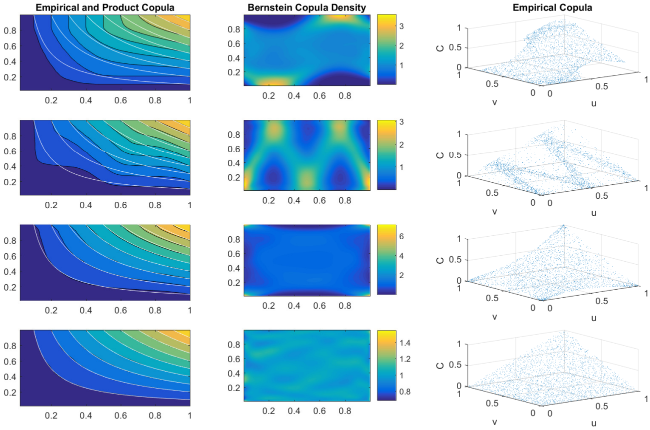
Numerical stable algorithm which allows for high polynomial degree.
Note that

$$1 = ((1-t)+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = \sum_{k=0}^{n} B_{k,n}(t)$$

```
function [b,B]=bern(t,n)
%BERN Bernstein base polynomials
 O=zeros(length(t),1); b=[1-t,t];
 for k=2:n
    T=repmat(t,1,k+1);
    B=b; b=(1-T).*[b,O]+T.*[O,b];
 end
 B=n*([O,B]-[B,O]); % derivative
end
```



Bernstein Copula, $m = 48$, 4096 QMC sample



Bregman Scoring Rules: Replacing the variance/square in variance-based sensitivity measures

$$\epsilon_i^{\psi} = \frac{\mathbb{E}[\psi(\mathbb{E}[Y|X_i])] - \psi(\mathbb{E}[Y])}{\mathbb{E}[\psi(Y)] - \psi(\mathbb{E}[Y])}$$

Normalized value-of-information based sensitivity measure where ψ is any convex function

Bregman Scoring Rules: Replacing the variance/square in variance-based sensitivity measures

$$\epsilon_i^{\psi} = \frac{\mathbb{E}[\psi(\mathbb{E}[Y|X_i])] - \psi(\mathbb{E}[Y])}{\mathbb{E}[\psi(Y)] - \psi(\mathbb{E}[Y])}$$

Normalized value-of-information based sensitivity measure where ψ is any convex function

Bregman function $\psi(y)$	Range	Scoring Name
y^2	$\mathbb{R}$	Quadratic / Brier
$-\log(y)$	$\mathbb{R}_{>0}$	Burg entropy
$y\log(y)$	$\mathbb{R}_{\geq 0}$	Logarithmic
$(y-a)\log(y-a)+(b-y)\log(b-y)$	$[a,b]$	Interval
$\log(1+\exp(y))$	$\mathbb{R}$	Dual logistic
$\sqrt{1-y^2}$	$[-1,1]$	Hellinger
$\|y\|^{\alpha}$, $\alpha > 1$	$\mathbb{R}$	Absolute Power

Bregman scores: Theory

Bregman score

$$S^{\psi}(y,a) = \psi(a) + \psi'(a)(y-a) - \psi(y)$$

y observation, a forecast, ψ convex function

Bregman scores: Theory

Bregman score

$$S^{\psi}(y,a) = \psi(a) + \psi'(a)(y-a) - \psi(y)$$

y observation, a forecast, ψ convex function

Theorem

The condition $S^{\psi}(\lambda y, \lambda a) = \lambda^{\alpha} S^{\psi}(y,a)$ is satisfied for

- $\alpha = 0$ *by the logarithmic score,*
- $\alpha = 1$ *by the Burg entropy,*
- $\alpha \neq 0, 1$ *by the Power score.*

Bregman scores: Theory

Bregman score

$$S^{\psi}(y,a) = \psi(a) + \psi'(a)(y-a) - \psi(y)$$

y observation, a forecast, ψ convex function

Theorem

The condition $S^{\psi}(\lambda y, \lambda a) = \lambda^{\alpha} S^{\psi}(y,a)$ is satisfied for

- $\alpha = 0$ *by the logarithmic score,*
- $\alpha = 1$ *by the Burg entropy,*
- $\alpha \neq 0, 1$ *by the Power score.*

Still working with the first conditional moment! But penalty term is not restricted to quadratic loss.

From Scores to Sensitivity Measures

For proper scores, value a^* to be forecasted (mean, quantile, cdf, pdf, ...) (uniquely) maximizes the expected score:

$$\max_{a \in A} \mathbb{E}[S(Y,a)] = \mathbb{E}[S(Y,a^*)], \qquad a^* = \operatorname*{argmax}_{a \in A} \mathbb{E}[S(Y,a)]$$

From Scores to Sensitivity Measures

For proper scores, value a^* to be forecasted (mean, quantile, cdf, pdf, ...) (uniquely) maximizes the expected score:

$$\max_{a \in A} \mathbb{E}[S(Y,a)] = \mathbb{E}[S(Y,a^*)], \qquad a^* = \operatorname*{argmax}_{a \in A} \mathbb{E}[S(Y,a)]$$

Average value-of-information of Y given X_i under scoring rule S

$$\xi_i^S = \mathbb{E}\left[\max_{a \in A} \mathbb{E}\left[S(Y,a)|X_i\right]\right] - \max_{a \in A} \mathbb{E}\left[S(Y,a)\right]$$

Upper limit: Total value-of-information of Y under scoring rule S

$$\tau^S = \mathbb{E}\left[\max_{a \in A} S(Y,a)\right] - \max_{a \in A} \mathbb{E}\left[S(Y,a)\right]$$

Normalized sensitivity measure under S: $\epsilon_i^S = \frac{\xi_i^S}{\tau^S} \in [0,1]$

Generalized Scoring Rules

There exist proper scoring rules for

- the mean (Bregman sores)
- quantiles (tick/newsvendor scores)
- densities (generalized Bregman scores)
- distribution functions (continuously ranked probability score – CRPS)

Scoring functions for quantiles

Generalized piece-wise linear loss functions associated with the p-quantile of Y (check function / newsvendor function)

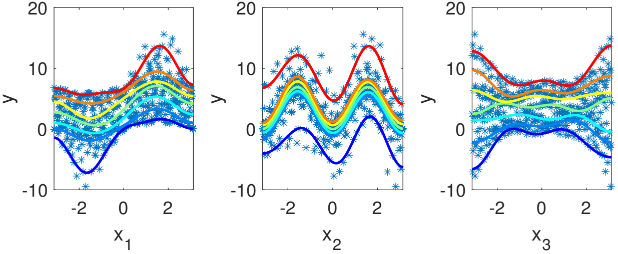
$$\mathcal{S}_p^Q(y,a) = k - h \cdot \left[p(t(y)-t(a))^+ + (1-p)(t(a)-t(y))^+\right], \tag{1}$$

where $p \in (0,1)$, $t : \mathbb{R} \mapsto \mathbb{R}$ is a non-decreasing function, $k \in \mathbb{R}$ and $h \in \mathbb{R}^+$ are constants.
$\mathbb{E}^{a,b}[Z] = \int_a^b z dF_Z(z)$: partial expectation of a random variable Z over interval (a,b). Then

$$\mathbb{E}[\mathcal{S}_p^Q(Y,a)] = \mathbb{E}^{-\infty,a}[Y] - a(F_Y(a)-p) - p\mathbb{E}[Y]$$

Optimal report a^* under $\mathcal{S}_p^Q(y,a)$ is p-quantile of Y, denoted by $Q_Y(p)$.

Quantile Sensitivity for Ishigami Function



Sensitivity value is obtained from the mean of output values beneath the quantile curve that is compared with the mean of output values beneath the unconditional output quantile (horizontal line).

Scoring functions for densities

Let the set of alternatives be given by probability densities on $\mathcal{Y}$. Define the generalized Bregman score as

$$S(y,p) = \psi'(p(y)) + \int_{\mathbb{R}} |\psi(p(z)) - p(z)\psi(p(z))| \, dz$$

When $\psi(p) = p - p\log p$ then we regain Kullback-Leibler sensitivity measure,

$$\zeta_i^{\mathsf{KL}} = \mathbb{E}\left[\int p_{Y|X_i}(y) \left(\log p_{Y|X_i}(y) - \log p_Y(y)\right) dy\right].$$

Scoring functions for distributions

Let the set of alternatives be given by cumulative distribution functions on $\mathcal{Y}$. The cdf generated by a single observation is a Heaviside step function, so we compare it to the forecasted cdf:

$$S^{\mathsf{CRPS}}(y, F) = -\int_{\mathbb{R}} \left(F(z) - \mathbb{1}\{z \geq y\}\right)^2 dz$$

The information-value sensitivity measure derived from this approach is

$$\zeta^{\mathsf{CRPS}}(Y, X_i) = \mathbb{E}\left[\int_{\mathbb{R}} \left(F_Y(y) - F_{Y|X_i}(y)\right)^2 dy\right]$$

which is the same as the one generated by the Gini mean difference / Cramér-von Mises divergence / energy distance.

Optimal Transport based Sensitivity Measures

Reframe the question:
How much does it cost on average to transport the conditional output measures $\mathbb{P}_{Y|X_i=x_i}$ to the unconditional output measure $\mathbb{P}_Y$?
For two probability measures ν, ν' define

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

where $\Pi(\nu,\nu')$ set of transport plans with marginals ν and ν', respectively, k cost function.

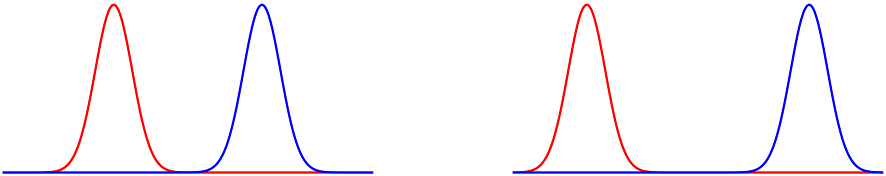
Optimal Transport based Sensitivity Measures

Reframe the question:
How much does it cost on average to transport the conditional output measures $\mathbb{P}_{Y|X_i=x_i}$ to the unconditional output measure $\mathbb{P}_Y$?
For two probability measures ν, ν' define

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

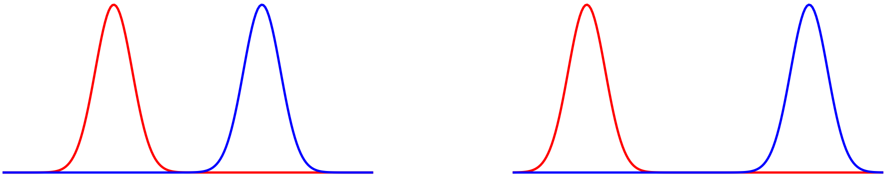
where $\Pi(\nu,\nu')$ set of transport plans with marginals ν and ν', respectively, k cost function.

The one-dimensional case



- The intuition “red density is further away from blue in the right panel compared to the left panel” works with OT 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 in 1D can be written compactly

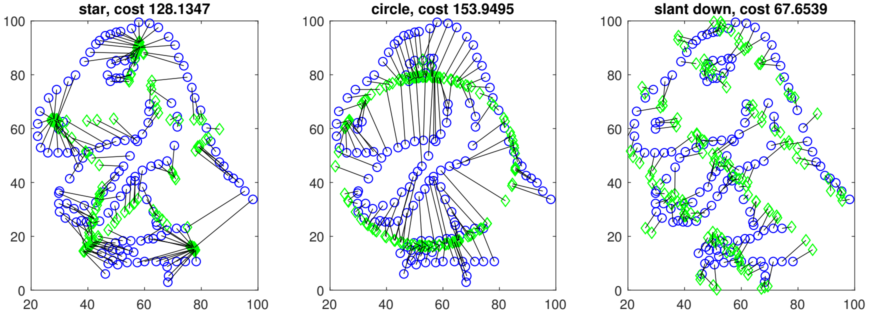
The one-dimensional case



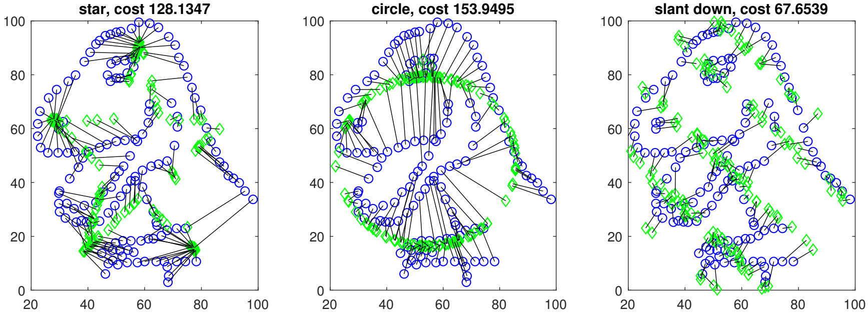
- The intuition “red density is further away from blue in the right panel compared to the left panel” works with OT Wasserstein, but not with cdf- or pdf-based distances
- Optimal transport plans to and from the $[0, 1]$ Lebesgue measure (standard uniform) are realized by quantile function and the cumulative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

Wasserstein sensitivity in 1D: Average distance of quantile functions

The two-dimensional case: Transporting the Datasaurus to different shapes



The two-dimensional case: Transporting the Datasaurus to different shapes



The longer the dark lines, the more costly the transport

Plugging Wasserstein into The Common Rationale

Using the squared Euclidean distance for k and plugging $\mathcal{K}$ into the common rationale yields the Wasserstein sensitivity measure

$$\zeta_i^{W_2^2} = \mathbb{E}\left[\inf_{\pi \in \Pi(\mathbb{P}_{Y|X_i}, \mathbb{P}_Y)} \int \|y - y'\|_2^2 \, d\pi(y, y')\right]$$

which also is defined if Y is multivariate.

Estimation: With a partitioning approach, need to solve Q OT problems (may be recast as linear programming problem) – fast (approximate) solvers

Normalization: $\zeta_i^{W_2^2}$ is bounded by twice the sum of the output variances, $2\,\mathrm{trace}\,\Sigma_Y$ (The expected squared Euclidean distance between two iid. random vectors is twice the variance).

Wasserstein-Bures Sensitivity

In the Gaussian case, a closed form OT solution named Wasserstein-Bures is known. As sensitivity measure, using means and empirical variance-covariance matrices as estimator inputs, we have

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

The advective part from the squared Wasserstein-Bures sensitivity coincides with point-wise variance-based Sobol' indices (unnormalized).

In 1D, WB sensitivity simplifies to $\frac{\xi^{\mathsf{WB}^2}}{2\mathbb{V}[Y]} = 1 - \mathbb{E}\left[\sqrt{\frac{\mathbb{V}[Y|X_\alpha]}{\mathbb{V}[Y]}}\right]$.

OT Solvers

Trade in precision for speed?

- Simplex algorithm is of non-polynomial running time

OT Solvers

Trade in precision for speed?

- Simplex algorithm is of non-polynomial running time
- Linear Assignment Problem (#sources = #sinks) via downscaling: pick a random subset from the unconditional output

OT Solvers

Trade in precision for speed?

- Simplex algorithm is of non-polynomial running time
- Linear Assignment Problem (#sources $=$ #sinks) via downscaling: pick a random subset from the unconditional output
- [Puccetti, 2017]: Bubble-sort inspired algorithm also for dim> 1

OT Solvers

Trade in precision for speed?

- Simplex algorithm is of non-polynomial running time
- Linear Assignment Problem (#sources $=$ #sinks) via downscaling: pick a random subset from the unconditional output
- [Puccetti, 2017]: Bubble-sort inspired algorithm also for dim> 1
- [Cuturi, 2013]: Entropic OT, with fast Sinkhorn-Knopp algorithm using matrix-vector operations

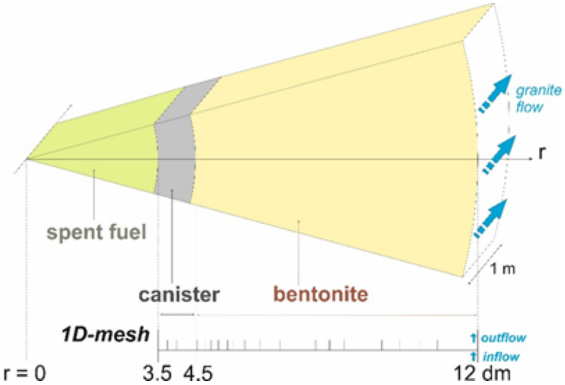
OT Solvers

Trade in precision for speed?

- Simplex algorithm is of non-polynomial running time
- Linear Assignment Problem ($\#$sources $= \#$sinks) via downscaling: pick a random subset from the unconditional output
- [Puccetti, 2017]: Bubble-sort inspired algorithm also for dim> 1
- [Cuturi, 2013]: Entropic OT, with fast Sinkhorn-Knopp algorithm using matrix-vector operations
- Ignore nonnormality and use Wasserstein-Bures

1D axi-symmetric reactive transport model

Long-term geochemical evolution of a High Level Waste repository in granite



(Data provided by Javier Sampre Calvete, Universidade de Coruña)

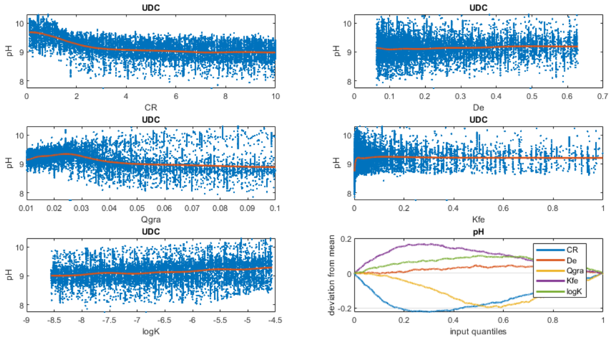
Inputs

- Corrosion rate (CR),
- Effective solute diffusion in the bentonite (De),
- Groundwater flow in granite (Qgra),
- Fe exchange selectivity (KFe),
- Log magnetite solubility constant (log K)

Outputs

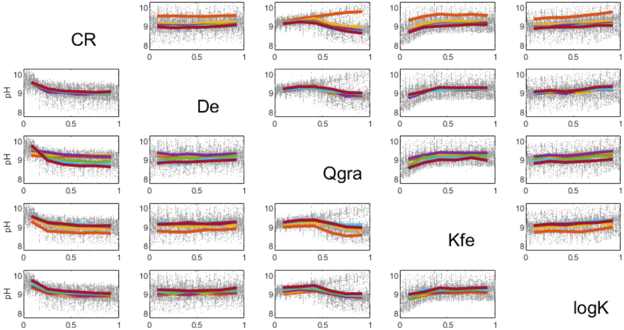
Quantity of interest: pH at end of simulation time (Eh, Corrosion products; different timesteps also available)

Variance-based analysis



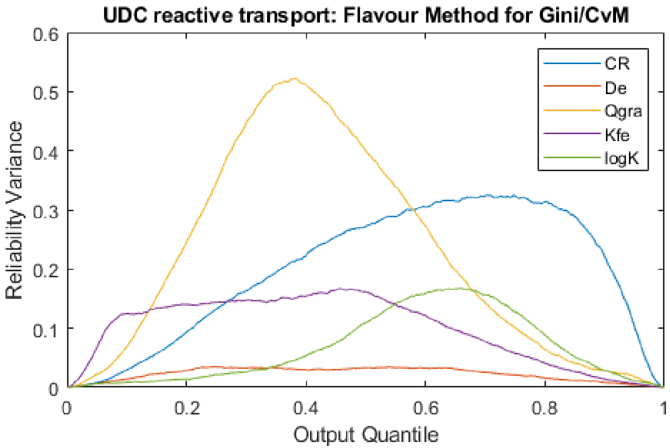
CR	DE	Qgra	KFe	logK
0.3269	0.0090	0.1785	0.1753	0.0502

Search for Interactions



Interaction plots: Watch out for non-parallel curves, intersections

Localized Sensitivity



Thank You!

Questions, Comments

mailto:e.plischke@hzdr.de

Preprints, Scripts, Stuff

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

GitLab Repository

https://gitlab.gwdg.de/elmar.plischke/global-sensitivity-analysis-collection

References (I)

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.

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

Pianosi, F. and Wagener, T. (2015).
A simple and efficient method for global sensitivity analysis based on cumulative distribution functions.
Environmental Modelling&Software, 67:1–11.

Plischke, E. and Borgonovo, E. (2019).
Copula theory and probabilistic sensitivity analysis: Is there a connection?
European Journal of Operational Research, 277(3):1046–1059.

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.

Sancetta, A. and Satchell, S. (2004).
The Bernstein copula and its applications to modeling and approximations of multivariate distributions.
Econometric Theory, 20(3):535–562.