Global Sensitivity Analysis with Distance Correlation and Energy Statistics

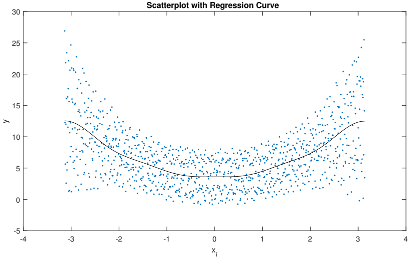
Elmar Plischke, joint work with Emanuele Borgonovo

Institut für Endlagerforschung
TU Clausthal
SAMO 2016, Université De La Réunion, December 1, 2015

Contents

Sensitivity Information from Scatterplots

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

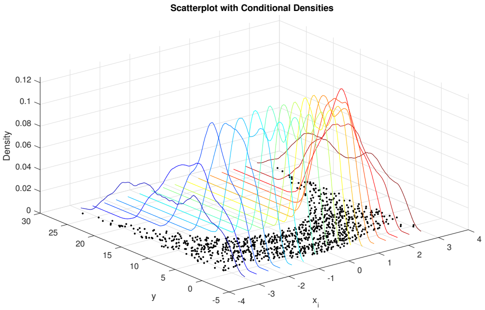


$$S_i = \frac{\mathbb{V}[\mathbb{E}[Y|X_i]]}{\mathbb{V}[Y]},\ \hat{S}_i = \frac{\sum_{j=1}^{n}(\mu_{Y|X_j=x_{ji}} - \mu_Y)^2}{\sum_{j=1}^{n}(y_j - \mu_Y)^2}$$

Conditioning on X_i: Local information from scatterplot

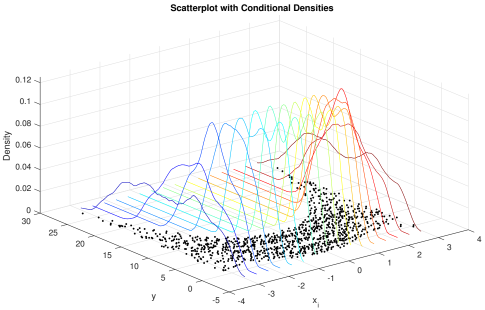
General Framework of Sensitivity Measures

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



General Framework of Sensitivity Measures

Idea: 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 $Y|X_i \in \mathcal{C}_m^i$ instead of point conditional $Y|X_i = x$)

General Framework II

A general sensitivity measure [Borgonovo et al., 2016]:

$$\zeta(X_i, Y) = \mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})]$$

Suitable distance measure $d(\cdot, \cdot)$: shift/separation/contrast function between total and conditional probability measures.

General Framework II

A general sensitivity measure [Borgonovo et al., 2016]:

$$\zeta(X_i, Y) = \mathbb{E}[d(\mathbb{P}_Y, \mathbb{P}_{Y|X_i})]$$

Suitable distance measure $d(\cdot, \cdot)$: shift/separation/contrast function between total and conditional probability measures.

Answers the question:
What is the (average) value of getting to know that $X_i = x$?

Examples: Shift/Separation/Contrast Functions

$d_{\mathrm{EI}}(\mu_Y, \mu_{Y\|X=x}) = \max\{\mu_{Y\|X=x}, 0\} - \max\{\mu_Y, 0\}$	EVPI, null alternative
$d_{\mathrm{SI}}(\mu_Y, \mu_{Y\|X=x}) = \sigma_Y^{-2}(\mu_Y - \mu_{Y\|X=x})^2$	Main Effect
$d_{\mathrm{KS}}(F_Y, F_{Y\|X=x}) = \sup \left\| F_Y - F_{Y\|X=x} \right\|$	Kolmogorov-Smirnov
$d_{\mathrm{Ku}}(F_Y, F_{Y\|X=x}) = \sup \left(F_Y - F_{Y\|X=x}\right) - \inf \left(F_Y - F_{Y\|X=x}\right)$	Kuiper
$d_{\mathrm{CvM}}(F_Y, F_{Y\|X=x}) = \frac{1}{2} \int \left(F_{Y\|X=x}(y) - F_Y(y)\right)^2 dy$	Cramér, L^2 (cdf)
$d_{\mathrm{Bo}}(f_Y, f_{Y\|X=x}) = \frac{1}{2} \int \left\| f_{Y\|X=x}(y) - f_Y(y) \right\| dy$	Borgonovo, L^1 (pdf)
$d_{\mathrm{KL}}(f_Y, f_{Y\|X=x}) = \int f_{Y\|X=x}(y) \log \frac{f_{Y\|X=x}(y)}{f_Y(y)} dy$	Kullback-Leibler
$d_{\mathrm{He}}(f_Y, f_{Y\|X=x}) = 1 - \int \sqrt{f_Y(y) \cdot f_{Y\|X=x}(y)} dy$	Hellinger

Which one is suitable?

Looking for sensitivity measures which are

- Simple to interpret
- Easy to estimate
- Invariant under monotonic transformations of inputs and outputs
- Detecting strong functional links: $Y = g(X) \implies \mathbb{E}[d(Y, Y|X)] = 1$
- Testing for independence:
  $\mathbb{E}[d(Y, Y|X)] = 0 \iff Y$ and X are independent

No “one–size–fits–all” sensitivity method

Which one is suitable?

Looking for sensitivity measures which are

- Simple to interpret
- Easy to estimate
- Invariant under monotonic transformations of inputs and outputs
- Detecting strong functional links: $Y = g(X) \implies \mathbb{E}[d(Y, Y|X)] = 1$
- Testing for independence:
  $\mathbb{E}[d(Y, Y|X)] = 0 \iff Y$ and X are independent

No “one–size–fits–all” sensitivity method

Numerical issues: pdf estimation, extreme values are not numerically stable

Quest for Distance Measures

Candidates should work on whole distribution, not just a point estimate: tests of statistical independence

Here: Closer look at two (and a half) candidates

- Distance Covariance
- Energy Statistics
- (Hilbert Schmidt independence criterion [Gretton et al., 2005])

Studied in SA context: [Da Veiga, 2015, Gamboa et al., 2015, Li et al., 2012]

Distance Covariance and Distance Correlation

At the SIAM UQ 2014 meeting, Grace Wahba (of spline fame) advised the use of Distance Correlation:

"Use Distance Correlation (DCOR) to select a subset of the variables that, taken together are related to the outcome of interest."

Distance Covariance and Distance Correlation

At the SIAM UQ 2014 meeting, Grace Wahba (of spline fame) advised the use of Distance Correlation:

"Use Distance Correlation (DCOR) to select a subset of the variables that, taken together are related to the outcome of interest."

Factor prioritization setting in sensitivity analysis

Distance Covariance

Let X and Y be p and q dimensional random vectors. [Székely et al., 2007] define

$$\mathrm{dCov}(X,Y)^2 = C_{p,q} \iint_{\mathbb{R}^{p+q}} \frac{|\varphi_{XY}(s,t) - \varphi_X(s)\varphi_Y(t)|^2}{\|s\|^{1+p}\,\|t\|^{1+q}} ds dt.$$

$\varphi_{XY}(s,t) = \mathbb{E}[e^{i\langle s,X\rangle + i\langle t,Y\rangle}]$, $\varphi_X(s) = \mathbb{E}[e^{i\langle s,X\rangle}]$, $\varphi_Y(t) = \mathbb{E}[e^{i\langle t,Y\rangle}]$ characteristic functions

$$\mathrm{dCorr}(X,Y) = \frac{\mathrm{dCov}(X,Y)}{\sqrt{\mathrm{dVar}(X)\,\mathrm{dVar}(Y)}}$$

Distance correlation with $\mathrm{dVar}(X) = \mathrm{dCov}(X,X)$

Distance Covariance II

In the scalar case with random variables X and Y:

$$\mathrm{dCov}(X,Y)^2 = \frac{1}{\pi^2}\iint_{\mathbb{R}^2} \frac{|\varphi_{XY}(s,t) - \varphi_X(s)\varphi_Y(t)|^2}{s^2t^2} dsdt.$$

Hence

$$\mathrm{dCov}(X,Y)^2 = \frac{1}{\pi^2}\iint_{\mathbb{R}^2} \left| \frac{\iint e^{i(sx+ty)}\left(f_{XY}(x,y) - f_X(x)f_Y(y)\right)dxdy}{st} \right|^2 dsdt$$

$$= \frac{1}{\pi^2}\iint_{\mathbb{R}^2} \left| \mathbb{E}\left[\frac{e^{isX}}{s}\left(\mathbb{E}\left[\frac{e^{itY}}{t}\middle| X\right] - \mathbb{E}\left[\frac{e^{itY}}{t}\right]\right)\right] \right|^2 dsdt$$

Closely related, but not part of the framework suggested above

Empirical Distance Covariance

For a sample $(x_j, y_j)_{j=1,\dots,n}$ of (X, Y)

$$\widehat{\mathrm{dCov}}(X, Y) = \sqrt{\frac{1}{n(n-3)} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij}} = \sqrt{\frac{1}{n(n-3)} \operatorname{trace}(AB)}$$

with distance matrices computed via

$$a_{ij} = \left\| x_i - x_j \right\|, \qquad \bar{a}_{i\cdot} = \frac{1}{n-2} \sum_{j=1}^{n} a_{ij}, \qquad \bar{a}_{\cdot j} = \frac{1}{n-2} \sum_{i=1}^{n} a_{ij},$$

$$\bar{a} = \frac{1}{(n-1)(n-2)} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}, \qquad A_{ij} = a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}$$

Analogously for (y_j), forming B

Trace product is inner product for vectorized matrices

Hilbert–Schmidt Indepence Criterion

HSIC introduced by [Gretton et al., 2005].
Estimator of HSIC: Absolute values in dCov distances are replaced by kernel evaluations

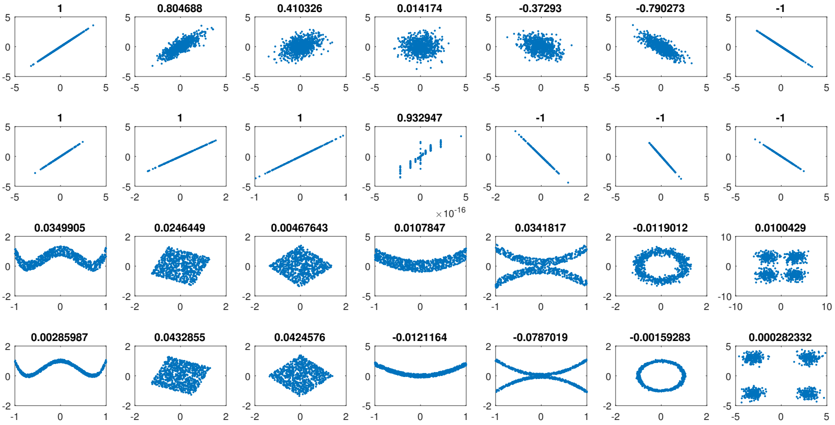
$$a_{ij} = k_X(x_i, x_j), \qquad b_{ij} = k_Y(y_i, y_j)$$

Mostly used with Gaussian kernels

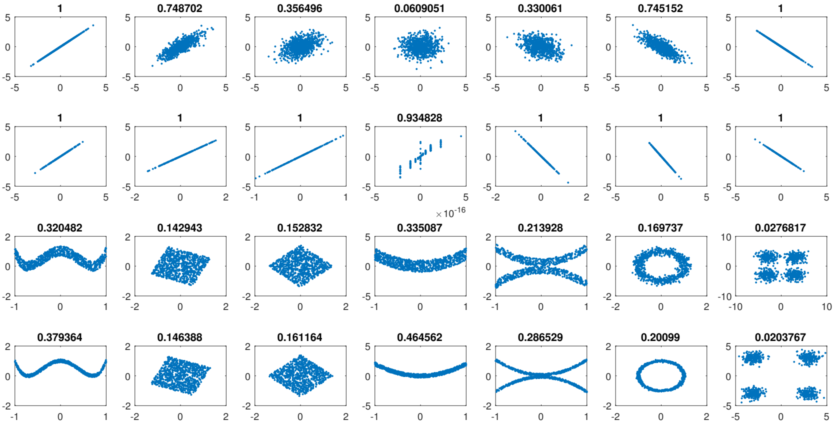
$$k(z_i, z_j) = \exp(-h^{-2}\|z_i - z_j\|^2)$$

Bandwidth h chosen in such way that half of the off-diagonal entries are below e^{-1}
Normalized version: Hilbert–Schmidt Independence Quotient (new)

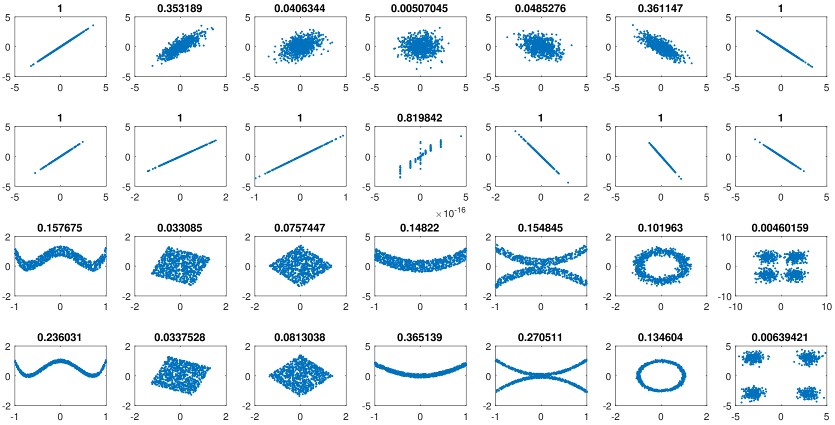
Example: Correlation



Example: Distance Correlation



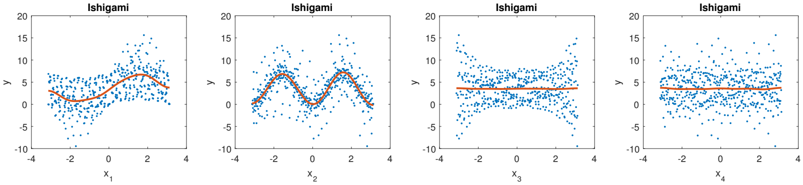
Example: Hilbert Schmidt Independence Quotient



dCov as Sensitivity Measure

Replace correlation with new measure!

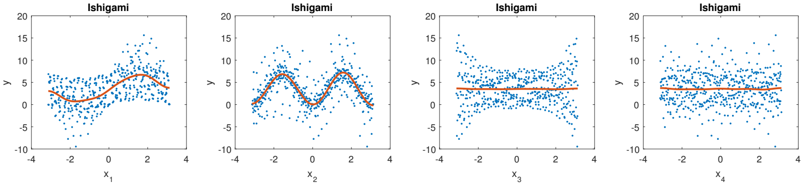
Example: Ishigami function with dummy parameter, QMC 512



dCov as Sensitivity Measure

Replace correlation with new measure!

Example: Ishigami function with dummy parameter, QMC 512



Parameter	Correlation	DCorrelation	First Order Effect
1	0.4304	0.4441	0.3141
2	-0.0043	0.1808	0.4443
3	-0.0018	0.1179	0.0001
4	-0.0013	0.0393	0.0005

dCov as Sensitivity Measure II

- Generalizes linear correlation
- Can detect heteroscedasticity (changing conditional variance)
- Different ranking than from first order effects
- Independent dummy parameter is identified

dCov as Sensitivity Measure II

- Generalizes linear correlation
- Can detect heteroscedasticity (changing conditional variance)
- Different ranking than from first order effects
- Independent dummy parameter is identified

Interaction effects: Consider dCorr($\{X_i, X_j\}, Y$) of Ishigami

$\{i, j\}$	1	2	3	4
1	0.4441	0.3762	0.3612	0.3644
2	0.3762	0.1808	0.1311	0.1023
3	0.3612	0.1311	0.1179	0.0918
4	0.3644	0.1023	0.0918	0.0393

dCov as Sensitivity Measure II

- Generalizes linear correlation
- Can detect heteroscedasticity (changing conditional variance)
- Different ranking than from first order effects
- Independent dummy parameter is identified

Interaction effects: Consider dCorr($\{X_i, X_j\}, Y$) of Ishigami

$\{i,j\}$	1	2	3	4
1	0.4441	0.3762	0.3612	0.3644
2	0.3762	0.1808	0.1311	0.1023
3	0.3612	0.1311	0.1179	0.0918
4	0.3644	0.1023	0.0918	0.0393

Interaction sensitivity from dCorr smaller than sum of individual contributions!?
Interactions of factors 1 and 3 not visible! Group sensitivity questionable

Energy Statistics as Sensitivity Measure

For d-dimensional RVecs Y, Z (and iid copies Y', Z') [Székely and Rizzo, 2013] define

$$\mathcal{E}(Y,Z) = 2\,\mathbb{E}\,\|Z - Y\| - \mathbb{E}\,\|Z - Z'\| - \mathbb{E}\,\|Y - Y'\|$$

1D: Gini mean difference $\mathcal{G}(Y) = 2\int F_Y(t)(1 - F_Y(t))dt = 4\,\mathrm{Cov}(Y, F_Y) = \mathbb{E}\,|Y - Y'|$

Energy Statistics as Sensitivity Measure

For d-dimensional RVecs Y, Z (and iid copies Y', Z') [Székely and Rizzo, 2013] define

$$\mathcal{E}(Y,Z) = 2\,\mathbb{E}\,\|Z-Y\| - \mathbb{E}\,\|Z-Z'\| - \mathbb{E}\,\|Y-Y'\|$$

1D: Gini mean difference $\mathcal{G}(Y) = 2\int F_Y(t)(1-F_Y(t))dt = 4\,\mathrm{Cov}(Y, F_Y) = \mathbb{E}\,|Y-Y'|$
Sensitivity framework: $Z = (Y|X=x)$ with separation

$$d(Y, Y|X=x) = \int \Big(F_{Y|X=x}(y) - F_Y(y)\Big)^2 dy = \tfrac{1}{2}\mathcal{E}(Y, Y|X=x)$$
$$= \mathbb{E}\,|(Y|X=x) - Y| - \tfrac{1}{2}(\mathcal{G}(Y) + \mathcal{G}(Y|X=x))$$

Gini (or Cramér–von Mises) sensitivity: $\gamma = \mathbb{E}[d(Y, Y|X)]$

Estimating Energy Statistics

Hoeffding U-statistics (as for dCov): Sample is full of iid copies.
But for large samples: First you run out of patience then you run out of memory.

For 1D: Increasingly sorted output sample $(y_{(j)})$ without ties

$$\widehat{\mathcal{E}}(Y, Y|X_i \in \mathcal{C}_m) = \sum_{j=1}^{n-1} \left(\hat{F}_Y(y_{(j)}) - \hat{F}_{Y|X_i\in\mathcal{C}_m}(y_{(j)})\right)^2 \left(y_{(j+1)} - y_{(j)}\right).$$

By construction $\hat{F}_Y(y_{(j)}) = \frac{j}{n}$, $\hat{F}_{Y|X_i\in\mathcal{C}_m}$ from cumulative sum of the indicator function for $\mathcal{C}_m$

Weighted Energy Statistics

Consider

$$\mathcal{E}^*(Y, Y|X_i = x) = \int \Big(F_Y(y) - F_{Y|X_i=x}(y)\Big)^2 dF_Y(y)$$

$$\widehat{\mathcal{E}}^*(Y, Y|X_i \in \mathcal{C}_m) = \frac{1}{n}\sum_{j=1}^{n-1}\Big(\hat{F}_Y(y_{(j)}) - \hat{F}_{Y|X_i\in\mathcal{C}_m}(y_{(j)})\Big)^2$$

yields transformation-invariant version of Gini measure

$$\gamma^* = 6\iint \Big(F_Y(y) - F_{Y|X_i=x}(y)\Big)^2 dF_Y(y)dF_{X_i}(x)$$

Scale factor: Monotonic dependence yields $\gamma^* = 1$

Weighted Energy Statistics: The Flavour Method of [Gamboa et al., 2015]

- Replace the output Y with quantile indicator functions $1\!\!1\{Y \leq F_Y^{-1}(\theta)\}$
- Compute the variance of the conditional output/first order effects times output variance (scaled by 6) of the indicators
- Average over all quantiles

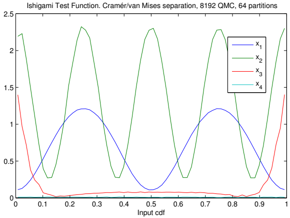
Example: Ishigami 512 QMC, using COSI 10 DCT coefficients vs. partition size 16

Parameter	10%	30%	50%	70%	90%	mean	γ^*	γ
1	.111	.193	.206	.234	.101	.167	0.1789	0.3452
2	.045	.538	.728	.498	.052	.372	0.3399	0.6082
3	.063	.025	.022	.025	.062	.038	0.0347	0.1005
4	.002	.018	.001	.017	.004	.005	0.0147	0.0298

Example: Ishigami

Separation measure for Ishigami (8192 QMC)

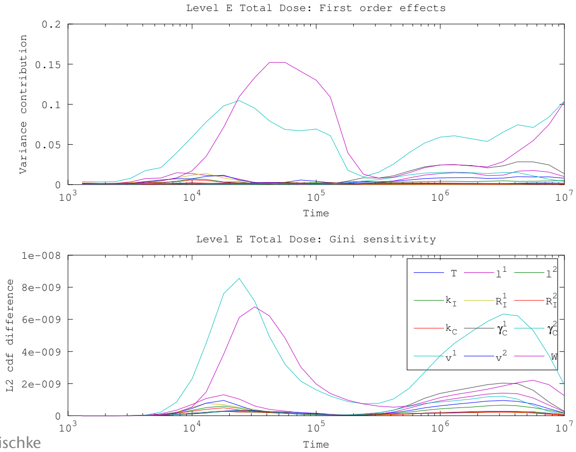
$$\theta \mapsto \widehat{\mathcal{E}}\left((y_j), \left(y_j \middle| \widehat{F}_{X_i}^{-1}(x_{ji}) \in \left[\theta - \frac{1}{M}, \theta\right)\right)\right), \quad \theta = \frac{1}{M}, \ldots, \frac{M}{M}$$



Parameter	1	2	3	4
Gini, 8192 QMC	0.3373	0.6347	0.0822	0.0020
Gini, 512 QMC	0.3807	0.6574	0.1234	0.0526

Example: Geosphere Transport Model Level-E, $k = 12$ inputs

Time-dependent analysis



First order effects: Are all individual contributions accounted for? γ provides additional hints that all indivial factors influencing the conditional distribution (instead of the conditional mean) are being identified.

Conclusions

- Attractive computational schemes
- dCorr: Group interaction may score less than sum of individual contributions?
- Gini / Cramér / von Mises fits into sensitivity framework
- Multidimensional version: Energy statistics
- Transformation-invariant version of Gini / Cramér / von Mises sensitivity
- Prototype software implementations available in MATLAB

Conclusions

- Attractive computational schemes
- dCorr: Group interaction may score less than sum of individual contributions?
- Gini / Cramér / von Mises fits into sensitivity framework
- Multidimensional version: Energy statistics
- Transformation-invariant version of Gini / Cramér / von Mises sensitivity
- Prototype software implementations available in MATLAB

Outlook

- Robust estimation
- Interactions and Group sensitivity

Thank You!

Questions, Comments

`mailto:elmar.plischke@tu-clausthal.de`

Preprints, Scripts, Stuff

`http://www.immr.tu-clausthal.de/~epl/`

References I

Borgonovo, E., Hazen, G. B., and Plischke, E. (2016).
A common rationale for global sensitivity measures and their estimation.
Risk Analysis.
DOI: 10.1111/risa.12555.

Da Veiga, S. (2015).
Global sensitivity analysis with dependence measures.
Journal of Statistical Computation and Simulation, 85(7):1283–1305.

Gamboa, F., Klein, T., and Lagnoux, A. (2015).
Sensitivity analysis based on Cramér von Mises distance.
arXiv:1506.04133v1.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005).
Measuring statistical dependence with Hilbert-Schmidt norms.
In *Algorithmic Learning Theory. 16th International Conference, ALT 2005*, pages 63–77. Springer Verlag, Berlin.

Li, L., Lu, Z., Feng, J., and Wang, B. (2012).
Moment-independent importance measure of basic variable and its state dependent parameter solution.
Structural Safety, 38:40–47.

References II

Székely, G. J. and Rizzo, M. L. (2013).
Energy statistics: A class of statistics based on distances.
Journal of Statistical Planning and Inference, 143:1249–1272.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007).
Measuring and testing dependence by correlation of distances.
The Annals of Statistics, 35(6):2769–2794.