Statistical Sciences
UNIVERSITY OF TORONTO

Sensitivity Measures based on Scoring Functions

Silvana M. Pesenti, U. Toronto

joint work with

Tobias Fissler, Vienna University

available on ArXiv: https://doi.org/10.48550/arXiv.2203.00460

SAMO, Tallahassee – March 14-16, 2022

Sensitivity Measures

- Probability space: $(\Omega, \mathfrak{F}, \mathbb{P})$
- Model: $Y = g(\boldsymbol{X})$
- T: $\mathcal{M}' \to A$ law-invariant functional of interest.
- $\mathcal{M}'$ sets of probability measures on $\mathbb{R}$

Examples of T: Mean, mode, quantiles, Value-at-Risk, Expected Shortfall, law-invariant risk measures...

Sensitivity Measures

- Probability space: $(\Omega, \mathfrak{F}, \mathbb{P})$
- Model: $Y = g(\boldsymbol{X})$
- T: $\mathcal{M}' \to A$ law-invariant functional of interest.
- $\mathcal{M}'$ sets of probability measures on $\mathbb{R}$

Examples of T: Mean, mode, quantiles, Value-at-Risk, Expected Shortfall, law-invariant risk measures...

$\Rightarrow$ How to choose a sensitivity measure for T?

Sensitivity Measures

- ◇ variance-based [Saltelli & Tarantola, 2002, Saltelli et al., 2008];
- ◇ moment-independent [Borgonovo, 2007];
- ◇ quantile-based sensitivities [Browne et al., 2017];
- ◇ based on divergence measures
  [Gamboa et al., 2018, Pesenti et al., 2019, Fort et al., 2021]
- ◇ differential sensitivity measures [Tsanakas & Millossovich, 2016]
- ◇ Machine Learning & Interpretability
  [Mase et al., 2021, Bénard et al., 2021]
- ◇ Stochastic Models [Etoré et al., 2020]

Sensitivity Measures

- ◇ variance-based [Saltelli & Tarantola, 2002, Saltelli et al., 2008];
- ◇ moment-independent [Borgonovo, 2007];
- ◇ quantile-based sensitivities [Browne et al., 2017];
- ◇ based on divergence measures [Gamboa et al., 2018, Pesenti et al., 2019, Fort et al., 2021]
- ◇ differential sensitivity measures [Tsanakas & Millossovich, 2016]
- ◇ Machine Learning & Interpretability [Mase et al., 2021, Bénard et al., 2021]
- ◇ Stochastic Models [Etoré et al., 2020]

$\Rightarrow$ [Borgonovo et al., 2021] choice of a sensitivity measure should be tied to T via a **strictly consistent scoring functions** and thus reflect the **information value**.

Scoring Functions [Murphy & Daan, 1985, Engelberg et al., 2009]

A scoring function is a measurable map $S\colon \mathsf{A} \times \mathbb{R} \to [0, \infty]$.

Scoring Functions [Murphy & Daan, 1985, Engelberg et al., 2009]

A scoring function is a measurable map $S\colon \mathsf{A} \times \mathbb{R} \to [0, \infty]$.

For $T\colon \mathcal{M}' \to \mathsf{A}$ and a sub-class $\mathcal{M} \subseteq \mathcal{M}'$, we say

(i) S is $\mathcal{M}$-*consistent* for T, if for all $F \in \mathcal{M}$ and for all $z \in \mathsf{A}$

$$\int S(T(F), y)\, \mathrm{d}F(y) \leq \int S(z, y)\, \mathrm{d}F(y)\,. \qquad (1)$$

(ii) S is *strictly* $\mathcal{M}$-consistent for T, if it is $\mathcal{M}$-consistent for T and if (1) is strict for all $z \neq T(F)$.

Scoring Functions [Murphy & Daan, 1985, Engelberg et al., 2009]

A scoring function is a measurable map $S\colon \mathsf{A} \times \mathbb{R} \to [0, \infty]$.

For $T\colon \mathcal{M}' \to \mathsf{A}$ and a sub-class $\mathcal{M} \subseteq \mathcal{M}'$, we say

(i) S is $\mathcal{M}$-*consistent* for T, if for all $F \in \mathcal{M}$ and for all $z \in \mathsf{A}$

$$\int S(T(F), y)\, \mathrm{d}F(y) \leq \int S(z, y)\, \mathrm{d}F(y)\,. \tag{1}$$

(ii) S is *strictly* $\mathcal{M}$-consistent for T, if it is $\mathcal{M}$-consistent for T and if (1) is strict for all $z \neq T(F)$.

T is *elicitable* on $\mathcal{M}$, if there exists a strictly $\mathcal{M}$-consistent scoring function for T.

Elicitable Functionals [Osband 1985, Weber 2006, Gneiting 2011]

T	$S(z, y)$
mean	$(x - y)^2$
median	$\lvert x - y \rvert$
VaR_α	$(\mathbb{1}_{\{y \leq z\}} - \alpha)(z - y)$
variance	NO
Expected Shortfall (ES)	NO
(mean, variance)	YES!
$(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$	YES!

Information Value [Holzmann & Eulert, 2014]

Strictly consistent scoring functions respect increasing information sets:

If S is strictly consistent for T and $\mathfrak{A} \subseteq \mathfrak{F}$, then

$$\mathbb{E}[S(T(Y), Y)] - \mathbb{E}[S(T(Y|\mathfrak{A}), Y)] \geq 0.$$

- **resolution term** of the score decomposition [Pohle, 2020]
- **discrimination** [Gneiting & Resin, 2021]
- a special instance of a **score divergence** [Thorarinsdottir et al., 2013, Gneiting & Raftery, 2007]

Information Value [Holzmann & Eulert, 2014]

Strictly consistent scoring functions respect increasing information sets:

If S is strictly consistent for T and $\mathfrak{A} \subseteq \mathfrak{F}$, then

$$\mathbb{E}[S(T(Y), Y)] - \mathbb{E}[S(T(Y|\mathfrak{A}), Y)] \geq 0.$$

Sensitivity of Y to $\mathfrak{A}$ based on S is given by[1]

$$\xi_S(Y; \mathfrak{A}) = \frac{\mathbb{E}[S(T(Y), Y)] - \mathbb{E}[S(T(Y|\mathfrak{A}), Y)]}{\mathbb{E}[S(T(Y), Y)]} \in [0, 1] \,.$$

E.g. $T() = \mathbb{E}[\ ]$, $S(z, y) = (z - y)^2$, $\xi_S(Y; \mathfrak{A})$ are the Sobol indices.

[1]technical assumptions on S

Zero Information Gain

Zero information gain:

$T(Y|\mathfrak{A}) = T(Y)$ implies $\xi_S(Y;\mathfrak{A}) = 0$.
S is strictly consistent: $\xi_S(Y;\mathfrak{A}) = 0$ only if $T(Y|\mathfrak{A}) = T(Y)$.

- If Y and $\mathfrak{A}$ are independent: $\xi_S(Y;\mathfrak{A}) = 0$

Zero Information Gain

Zero information gain:

$T(Y|\mathfrak{A}) = T(Y)$ implies $\xi_S(Y; \mathfrak{A}) = 0$.
S is strictly consistent: $\xi_S(Y; \mathfrak{A}) = 0$ only if $T(Y|\mathfrak{A}) = T(Y)$.

- If Y and $\mathfrak{A}$ are independent: $\xi_S(Y; \mathfrak{A}) = 0$
- $Y = X_1 X_2 + X_3$; X_1, X_2, X_3 independent and $\mathbb{E}[X_2] = 0$.
  Then,
  $$T(Y|X_1) = \mathbb{E}[Y|X_1] = \mathbb{E}[X_3] = \mathbb{E}[Y] = T(Y)$$
  and $\xi_S(Y; X_1) = 0$ for any consistent scoring function S for the mean.

Zero Information Gain

Zero information gain:

$T(Y|\mathfrak{A}) = T(Y)$ implies $\xi_S(Y;\mathfrak{A}) = 0$.
S is strictly consistent: $\xi_S(Y;\mathfrak{A}) = 0$ only if $T(Y|\mathfrak{A}) = T(Y)$.

- ◇ If Y and $\mathfrak{A}$ are independent: $\xi_S(Y;\mathfrak{A}) = 0$
- ◇ $Y = X_1 X_2 + X_3$; X_1, X_2, X_3 independent and $\mathbb{E}[X_2] = 0$.
  Then,
  $$T(Y|X_1) = \mathbb{E}[Y|X_1] = \mathbb{E}[X_3] = \mathbb{E}[Y] = T(Y)$$
  and $\xi_S(Y;X_1) = 0$ for any consistent scoring function S for the mean.

$\Rightarrow$ *Nullity-implies-independence* is too strong.

Properties

Full information gain:

$T(Y|\mathfrak{A}) = T(\delta_Y)$ implies $\xi_S(Y;\mathfrak{A}) = 1$.
S strictly consistent: $\xi_S(Y;\mathfrak{A}) = 1$ only if $T(Y|\mathfrak{A}) = T(\delta_Y)$.

- Y is $\mathfrak{A}$ measurable $\xi_S(Y;\mathfrak{A}) = 1$.
- E.g., $\mathfrak{A} = \sigma(\boldsymbol{X})$; Y and X_i are co-(counter)monotonic

Properties

Full information gain:

$T(Y|\mathfrak{A}) = T(\delta_Y)$ implies $\xi_S(Y;\mathfrak{A}) = 1$.
S strictly consistent: $\xi_S(Y;\mathfrak{A}) = 1$ only if $T(Y|\mathfrak{A}) = T(\delta_Y)$.

- Y is $\mathfrak{A}$ measurable $\xi_S(Y;\mathfrak{A}) = 1$.
- E.g., $\mathfrak{A} = \sigma(\boldsymbol{X})$; Y and X_i are co-(counter)monotonic

Monotonicity with respect to nested information:

If $\mathfrak{A} \subseteq \mathfrak{A}'$ then $\xi_S(Y;\mathfrak{A}) \leq \xi_S(Y;\mathfrak{A}')$.
S strictly consistent: $\xi_S(Y;\mathfrak{A}) = \xi_S(Y;\mathfrak{A}')$ only if $T(Y|\mathfrak{A}) = T(Y|\mathfrak{A}')$.

- $\xi_S\big(Y; h(\boldsymbol{X})\big) \leq \xi_S(Y;\boldsymbol{X})$, $h\colon \mathbb{R}^d \to \mathbb{R}^m$, $d \leq m$

Choice of Scoring Function

A score S is strictly consistent for $T(\cdot) = \mathbb{E}[\cdot]$, iff

$$S_\phi(z, y) = \phi(y) - \phi(z) + \phi'(z)(z - y)\,, \qquad z, y \in \mathbb{R}\,,$$

for a strictly convex function ϕ with subgradient ϕ'.

Choice of Scoring Function

A score S is strictly consistent for $T(\cdot) = \mathbb{E}[\cdot]$, iff

$$S_\phi(z, y) = \phi(y) - \phi(z) + \phi'(z)(z - y)\,, \qquad z, y \in \mathbb{R}\,,$$

for a strictly convex function ϕ with subgradient ϕ'.

Scale-invariant scores

$$\xi_S(cY; \mathfrak{A}) = \xi_S(Y; \mathfrak{A}) \quad \text{for all} \quad c > 0\,, \mathfrak{A} \subseteq \mathfrak{F}\,.$$

Proposition
If S is homogeneous, i.e., $S(cz, cy) = c^b S(z, y)$, for some $b \in \mathbb{R}$, then ξ_S is scale-invariant.

Murphy Diagrams

Homogeneous scores for $T(\cdot) = \mathbb{E}[\cdot]$ is the Patton family $(z, y > 0)$ [2]

$$S_b(z,y) = \begin{cases} \frac{y^b - z^b}{b(b-1)} - \frac{z^{b-1}}{b-1}(y-z), & b \in \mathbb{R} \setminus \{0,1\}, \\ \frac{y}{z} - \log\left(\frac{y}{z}\right) - 1, & b = 0, \\ y \log\left(\frac{y}{z}\right) - (y-z), & b = 1, \end{cases} \tag{2}$$

Unique up to a positive scaling.

[2] [Nolde & Ziegel, 2017]

Murphy Diagrams

Homogeneous scores for $T(\cdot) = \mathbb{E}[\cdot]$ is the Patton family $(z, y > 0)$ [2]

$$S_b(z, y) = \begin{cases} \frac{y^b - z^b}{b(b-1)} - \frac{z^{b-1}}{b-1}(y - z), & b \in \mathbb{R} \setminus \{0, 1\}, \\ \frac{y}{z} - \log\left(\frac{y}{z}\right) - 1, & b = 0, \\ y \log\left(\frac{y}{z}\right) - (y - z), & b = 1, \end{cases} \tag{2}$$

Unique up to a positive scaling.

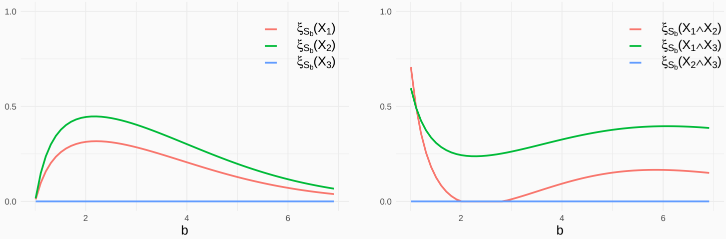
Murphy diagrams for sensitivity measures $b \mapsto \xi_{S_b}(Y; \mathfrak{A})$.

[2] [Nolde & Ziegel, 2017]

Ishigami–Homma Test Function

$T = \mathbb{E}[\,]$, b-homogeneneous scores $S(cz, yz) = c^b S(z, y)$:

$$Y = \sin(X_1) + 7\sin(X_2)^2 + 0.1 X_3^4 \sin(X_1)\,,$$



Interaction information

$$\xi_S(Y; \mathfrak{A}_1 \wedge \mathfrak{A}_2) := \max\left\{\xi_S(Y; \sigma(\mathfrak{A}_1 \cup \mathfrak{A}_2)) - \xi_S(Y; \mathfrak{A}_1) - \xi_S(Y; \mathfrak{A}_2)\,,\ 0\right\},$$

Paper includes

Working paper includes

- multiple examples illustrating the properties
- Murphy diagrams for elementary scores
- non-linear insurance example
- Score-based sensitivity measures to $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$
- using neural nets to estimate $T(Y|\boldsymbol{X})$

https://doi.org/10.48550/arXiv.2203.00460

Thank you!

https://utstat.toronto.edu/pesenti/
silvana.pesenti@utoronto.ca

References

Bénard, C., Biau, G., Veiga, S., & Scornet, E. (2021).
Interpretable random forests via rule extraction.
In *International Conference on Artificial Intelligence and Statistics* (pp. 937–945).: PMLR.

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

Borgonovo, E., Hazen, G. B., Jose, V. R. R., & Plischke, E. (2021).
Probabilistic sensitivity measures as information value.
European Journal of Operational Research, 289(2), 595–610.

Browne, T., Fort, J.-C., Iooss, B., & Le Gratiet, L. (2017).
Estimate of quantile-oriented sensitivity indices.
Preprint.

Engelberg, J., Manski, C. F., & Williams, J. (2009).
Comparing the point predictions and subjective probability distributions of professional forecasters.
Journal of Business & Economic Statistics, 27(1), 30–41.

Etoré, P., Prieur, C., Pham, D. K., & Li, L. (2020).
Global sensitivity analysis for models described by stochastic differential equations.
Methodology and Computing in Applied Probability, 22(2), 803–831.

Fort, J.-C., Klein, T., & Lagnoux, A. (2021).
Global sensitivity analysis and wasserstein spaces.
SIAM/ASA Journal on Uncertainty Quantification, 9(2), 880–921.

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

Gneiting, T. & Raftery, A. (2007).
Strictly Proper Scoring Rules, Prediction, and Estimation.
Journal of the American Statistical Association, 102(477), 359–378.

Gneiting, T. & Resin, J. (2021).
Regression diagnostics meets forecast evaluation: Conditional calibration, reliability diagrams, and coefficient of determination.
Preprint.

Holzmann, H. & Eulert, M. (2014).
The role of the information set for forecasting – with applications to risk management.
Annals of Applied Statistics, 8(1), 79–83.

Mase, M., Owen, A. B., & Seiler, B. B. (2021).
Cohort shapley value for algorithmic fairness.
arXiv preprint arXiv:2105.07168.

Murphy, A. H. & Daan, H. (1985).
Forecast Evaluation.
In A. H. Murphy & R. W. Katz (Eds.), *Probability, Statistics and Decision Making in the Atmospheric Sciences* (pp. 379–437). Westview Press, Boulder, Colorado.

Nolde, N. & Ziegel, J. F. (2017).
Elicitability and backtesting: Perspectives for banking regulation.
Annals of Applied Statistics, 11(4), 1833–1874.

Pesenti, S. M., Millossovich, P., & Tsanakas, A. (2019).
Reverse sensitivity testing: What does it take to break the model?
European Journal of Operational Research, 274(2), 654–670.

Pohle, M.-O. (2020).
The Murphy decomposition and the calibration-resolution principle: A new perspective on forecast evaluation.
Preprint.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008).
Global sensitivity analysis: the primer.
John Wiley & Sons.

Saltelli, A. & Tarantola, S. (2002).
On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal.
Journal of the American Statistical Association, 97(459), 702–709.

Thorarinsdottir, T. L., Gneiting, T., & Gissibl, N. (2013).
Using proper divergence functions to evaluate climate models.
SIAM/ASA Journal on Uncertainty Quantification, 1(1), 522–534.

Tsanakas, A. & Millossovich, P. (2016).
Sensitivity analysis using risk measures.
Risk Analysis, 36(1), 30–48.