Spotlights on the theory of elicitability

Habilitation Talk

Tobias Fissler

ETH Zurich

30 June 2023

Research Seminar

Institute for Statistics and Mathematics, WU Wien

Overview of Habilitation thesis

Publications included in Habilitation thesis

1. T. Fissler, S. M. Pesenti (2023).
   Sensitivity measures based on scoring functions.
   European Journal of Operational Research **307** (3), 1408–1423
2. T. Fissler, H. Holzmann (2022).
   Measurability of functionals and of ideal point forecasts.
   Electronic Journal of Statistics **16** (2), 5019–5034
3. C. Heinrich-Mertsching, T. Fissler (2022).
   Is the mode elicitable relative to unimodal distributions?
   Biometrika **109** (4), 1157–1164
4. T. Fissler, J. F. Ziegel (2021).
   On the elicitability of range value at risk.
   *Statistics & Risk Modeling **25** (1–2), 25–46*
5. T. Fissler, J. Hlavinová, B. Rudloff (2021).
   Elicitability and Identifiability of Systemic Risk Measures.
   Finance and Stochastics **25** (1), 133–165
6. T. Fissler, R. Frongillo, J. Hlavinová, B. Rudloff (2021).
   Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals.
   Electronic Journal of Statistics **15** (1), 1034–1084

Publications not included in the thesis

7. T. Fissler, Y. Hoga (2023).
   Backtesting Systemic Risk Forecasts using Multi-Objective Elicitability.
   Journal of Business & Economic Statistics.
8. T. Dimitriadis, T. Fissler, J. Ziegel (2023).
   Characterizing M-estimators.
   Biometrika (forthcoming).
9. T. Dimitriadis, T. Fissler, J. Ziegel (2023).
   Osband's Principle for Identification Functions.
   Statistical Papers
10. T. Fissler, M. Merz, M. V. Wüthrich (2023).
   Deep Quantile and Deep Composite Model Regression.
   Insurance: Mathematics and Economics **109**, 94–112
11. T. Fissler, J. F. Ziegel (2019).
   Order-Sensitivity and Equivariance of Scoring Functions.
   Electronic Journal of Statistics **13** (1), 1166–1211
12. T. Fissler, M. Podolskij (2017).
   Testing the maximal rank of the volatility process for continuous diffusions observed with noise.
   Bernoulli **23** (4B), 3021–3066

Publications not included in the thesis

13. T. Fissler, J. F. Ziegel (2016).
Higher order elicitability and Osband's principle.
Annals of Statistics **44** (4), 1680–1707.

14. T. Fissler, J. F. Ziegel, T. Gneiting (2016).
Expected Shortfall is jointly elicitable with Value at Risk – Implications for backtesting.
Risk Magazine, January 2016, 58–61.

15. T. Fissler, C. Thäle (2016).
A four moments theorem for Gamma limits on a Poisson chaos.
ALEA, Lat. Am. J. Probab. Math. Stat. **13** (1), 163–192.

Preprints

1. T. Fissler, C. Lorentzen, M. Mayer (2022).
Model Comparison and Calibration Assessment: User Guide for Consistent Scoring Functions in Machine Learning and Actuarial Practice.
`https://doi.org/10.48550/arXiv.2202.12780`

2. T. Dimitriadis, T. Fissler, J. F. Ziegel (2020).
The Efficiency Gap.
`https://doi.org/10.48550/arXiv.2010.14146`

Roadmap

1. Overview of Habilitation thesis ✓
2. Setup
3. Loss functions in statistical learning & forecast comparison
4. The elicitation problem
5. The wondrous tale about the mode
6. Range Value at Risk – a linear combination of Bayes risks
7. The dichotomy of the set-valued elicitation world
8. Summary and outlook

Setup

Setup

- Y: The quantity of interest or response:
  - Typically real-valued, but could also be multivariate, categorical etc.
  - **Examples:** Claim sizes, number of claims, temperature, precipitation, wind speed, demand for a product, GDP growth, inflation, loss of a company
- $\boldsymbol{X}$: Explanatory variables, regressors, features:
  - From a possibly high dimensional feature space $\mathcal{X}$.
  - Can contain metrical variables, categorical etc.
  - Can be exogenous variables (cross-sectional), but also past observations of Y (time series setup)

Learning: We want to exploit the information in $\boldsymbol{X}$ to describe Y as accurately as possible.
$\rightsquigarrow$ How to fit a model?

Prediction: We want to exploit the information in $\boldsymbol{X}$ to predict unseen Y as accurately as possible.
$\rightsquigarrow$ How to assess the accuracy?

Define your goal!

- Usually, $\boldsymbol{X}$ does not fully describe Y: There is no deterministic function g such that $Y = g(\boldsymbol{X})$.
- The remaining uncertainty of Y given $\boldsymbol{X}$ can be described in terms of the conditional distribution

$$F_{Y|\boldsymbol{X}}$$

Define your goal!

- Probabilistic predictions: Try to learn the full conditional distribution and come up with probabilistic forecasts $\widehat{F}_{Y|\boldsymbol{X}}$.
  - Very informative approach.
  - Often hard to implement.
  - Can be difficult to communicate.
- Point predictions: Summarise the conditional distribution with a functional of the conditional distribution

$$T(Y \mid \boldsymbol{X}) := T(F_{Y|\boldsymbol{X}})$$

  Note: The existence and $\sigma(\boldsymbol{X})$-measurability of $T(Y \mid \boldsymbol{X})$ has been established in F. & Holzmann (EJS, 2022).
- **Examples:**
  - mean, median, mode
  - quantiles, expectiles
  - Risk measures: Value at Risk, Expected Shortfall
- Come up with point forecasts $\widehat{T}(Y \mid \boldsymbol{X})$.
  - Loss of information
  - Easier to implement
  - Easier to communicate.

Loss functions in statistical learning & forecast comparison

Action domain and model choice

- Let $\mathcal{F}$ be a convex class of distributions such that $F_{Y|\boldsymbol{X}} \in \mathcal{F}$.
- Call the space where the chosen target functional maps to action domain $\mathcal{A}$. $T \colon \mathcal{F} \to \mathcal{A}$.
- **Examples:**
  - $\mathcal{A} = \mathbb{R}$ for the mean / quantile or $[0, \infty)$ for the mean / quantile of a positive Y
  - $\mathcal{A} = \mathbb{R}^k$ for mean of a multivariate observation, different quantiles of a real-valued observation
  - $\mathcal{A}$ finite for the mode of a categorical observation
  - $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^k)$ for prediction sets or systemic risk measures.
  - $\mathcal{A} = \mathcal{F}$, a class of probability distributions or densities for probabilistic forecasts (then T is the identity functional).
- Consider a model class $\mathcal{M}$ of models $m \colon \mathcal{X} \to \mathcal{A}$.
- **Examples:**
  - (Generalised) Linear Models
  - Neural nets
  - Isotonic regression functions
- Convexity of $\mathcal{F}$ ensures that $F_Y, F_{Y|m(\boldsymbol{X})} \in \mathcal{F}$ for all $m \in \mathcal{M}$.

Consistent loss functions and elicitability

Definition 1 (Consistency)

A loss function is a map

$$L\colon \mathcal{A} \times \mathbb{R} \to \mathbb{R}.$$

Sometimes, additional assumptions are imposed such as continuity (in the first argument), positivity etc.
It is $\mathcal{F}$-consistent for a functional T if

$$\mathrm{E}_{Y\sim F}\left[L(T(F), Y)\right] \leqslant \mathrm{E}_{Y\sim F}\left[L(a, Y)\right] \qquad \text{for all } a \in \mathcal{A},\ F \in \mathcal{F}.$$

L is strictly $\mathcal{F}$-consistent if equality arises only if $a = T(F)$.

Definition 2 (Elicitability)

A functional T is elicitable on $\mathcal{F}$ if there is a strictly $\mathcal{F}$-consistent loss function for it.

Alternative name for loss functions: Scoring functions

First examples of elicitable functionals

The mean is elicitable on the class of square integrable distributions. A strictly consistent loss function is given via the squared loss

$$L(a, y) = (a - y)^2.$$

The α-quantile is elicitable on the class of integrable distributions which are strictly increasing. A strictly consistent loss function is given via the pinball loss / asymmetric piecewise linear loss

$$L(a, y) = (\mathbb{1}\{y \leqslant a\} - \alpha)(a - y).$$

Learning via loss minimisation (M-estimation)

- Consider the statistical risk

$$R(m) = \mathrm{E}\left[L\big(m(\boldsymbol{X}), Y\big)\right] \\ = \mathrm{E}\Big[\mathrm{E}\left[L\big(m(\boldsymbol{X}), Y\big) \mid \boldsymbol{X}\right]\Big]$$

- Bayes rule is given by

$$m^* \in \arg\min_{m \in \mathcal{M}} R(m).$$

- If the true regression function $\boldsymbol{x} \mapsto T(Y \mid \boldsymbol{X} = \boldsymbol{x})$ is in $\mathcal{M}$ and if L is $\mathcal{F}$-consistent for T, we get

$$\mathrm{E}\left[L\big(T(Y \mid \boldsymbol{X}), Y\big) \mid \boldsymbol{X}\right] \leqslant \mathrm{E}\left[L\big(m(\boldsymbol{X}), Y\big) \mid \boldsymbol{X}\right].$$

- Therefore, $T(Y \mid \boldsymbol{X} = \cdot)$ is a Bayes rule.
- Due to Dimitriadis, F., Ziegel (Biometrika, 2023), the (strict) consistency of L is also necessary for $T(Y \mid \boldsymbol{X} = \cdot)$ to be the only Bayes act (under certain richness conditions).

Learning via loss minimisation (M-estimation)

- Let $D_{\text{train}} = \{(\boldsymbol{x}_i, y_i),\ i = 1, \ldots, n\}$ be a training sample. Define the empirical risk

$$\begin{aligned}\overline{R}(m; D_{\text{train}}) &= \frac{1}{n} \sum_{(\boldsymbol{x}_i, y_i) \in D_{\text{train}}} L(m(\boldsymbol{x}_i), y_i) \\ &\approx \mathrm{E}\left[L\big(m(\boldsymbol{X}), Y\big)\right] \\ &= R(m).\end{aligned}$$

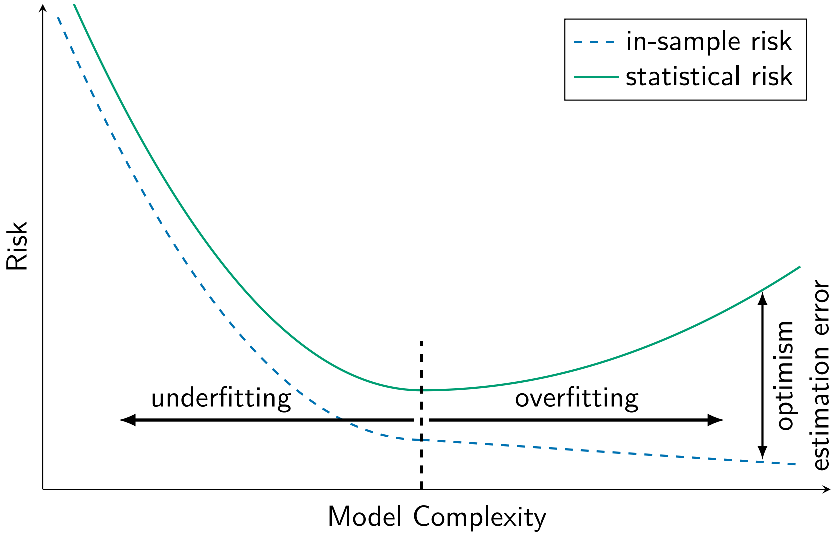
- M-estimator $\widehat{m}$ is an empirical risk minimiser

$$\widehat{m} \in \operatorname*{arg\,min}_{m \in \mathcal{M}} \overline{R}(m; D_{\text{train}})$$

Pitfall of overfitting

- Estimator $\widehat{m}$ depends on training sample D_{train}:
  - Prone to estimation error
  - Different training samples lead to different estimates.
  - Danger that $\widehat{m}$ learns the noisy pattern of the sample at hand and not the structure of the distribution.
  - In-sample performance $\overline{R}(\widehat{m}; D_{\text{train}})$ can be a bad estimate for the actual risk $R(\widehat{m})$.
  - $\rightsquigarrow$ pitfall of overfitting.
  - This problem gets bigger
    - ⋆ the more complex a model is;
    - ⋆ the smaller (less representative) the trainings sample is.

Pitfall of overfitting



Mitigating overfitting

There are two main strategies:

1. **Fitting:** Introduce a penalty term Ω, accounting for the model complexity:
$$\hat{m} = \arg\min_{m \in \mathcal{M}} \overline{R}(m; D_{\text{train}}) + \lambda\Omega(m).$$
Examples for Ω:
    - Number of parameters $\rightsquigarrow$ AIC and BIC
    - Norms of the parameter $\rightsquigarrow$ ridge and lasso regression
    - Number of optimisation steps when fitting a neural net
2. **Validation:** Monitor the out-of-sample risk on an (ideally) independent and identically distributed validation set $D_{\text{valid}} = \{(\boldsymbol{x}_i, y_i),\ i = 1, \dots, l\}$ via
$$\overline{R}(\hat{m}; D_{\text{valid}})$$
    - Better approximation of the statistical risk.
    - Can be made more efficient with cross-validation.

Model agnostic forecast comparison

- Suppose the target functional T is fixed (could also be probabilistic).
- We have different methods of producing predictions, but we are agnostic about how they have been produced.

$\rightsquigarrow$ We adhere to the weak prequential principle (Dawid & Vovk, 1999).

- Example for two different forecasters: We have the prediction–observation sequence

$$(A_i^{(1)}, A_i^{(2)}, Y_i) \quad i = 1, \ldots, n$$

- Ranking in terms of the empirical loss difference

$$\frac{1}{n} \sum_{i=1}^{n} L\big(A_i^{(1)}, Y_i\big) - L\big(A_i^{(2)}, Y_i\big)$$

  - Forecast method 1 is deemed better than 2 if this is negative.
  - Tests for equal predictive accuracy $\mathrm{E}[L(A^{(1)}, Y)] = \mathrm{E}[L(A^{(2)}, Y)]$ and forecast dominance $\mathrm{E}[L(A^{(1)}, Y)] \geqslant \mathrm{E}[L(A^{(2)}, Y)]$ can be assessed via Diebold–Mariano tests (amounting to t-tests).
- To honour truthful forecasting, L should be (strictly) consistent for T!

The Elicitation Problem

The Elicitation Problem

Fix some functional $T: \mathcal{F} \to \mathcal{A}$.

(a) Is T elicitable?
(b) What is the class of (strictly) consistent loss functions for T?
(c) What is a particularly good choice of a loss function?
(d) What to do if T is not elicitable?

T	$L(x, y)$
mean	$(x-y)^2$
median	$\|x-y\|$
τ-expectile	$\|\mathbb{1}\{y \leqslant x\} - \tau\|(x-y)^2$
α-quantile	$\|\mathbb{1}\{y \leqslant x\} - \alpha\|\,\|x-y\|$
variance	×
Expected Shortfall	×
(mean, variance)	✓
(Value at Risk, Expected Shortfall)	✓
identity (probabilistic forecast)	$L(F, y) = -\log(f(y))$

(a) One-dimensional functionals

Theorem 3 (Convex level sets, Osband, 1985)

Let $T \colon \mathcal{F} \to \mathcal{A}$ be an elicitable functional. Let $F_0, F_1 \in \mathcal{F}$ and $F_\lambda = (1-\lambda)F_0 + \lambda F_1 \in \mathcal{F}$ for some $\lambda \in (0,1)$. Then

$$T(F_0) = T(F_1) \implies T(F_\lambda) = T(F_0)$$

Proof: Let $t = T(F_0) = T(F_1)$ and $x \neq t$. Then, due to the linearity of the expectation in the measure,

$$\begin{aligned} \mathrm{E}_{Y\sim F_\lambda}[L(t,Y)] &= (1-\lambda)\,\mathrm{E}_{Y\sim F_0}[L(t,Y)] + \lambda\,\mathrm{E}_{Y\sim F_1}[L(t,Y)] \\ &< (1-\lambda)\,\mathrm{E}_{Y\sim F_0}[L(x,Y)] + \lambda\,\mathrm{E}_{Y\sim F_1}[L(x,Y)] \\ &= \mathrm{E}_{Y\sim F_\lambda}[L(x,Y)]. \end{aligned}$$

□

(a) One-dimensional functionals

Theorem 4 (Convex level sets, Osband, 1985)

Let $T\colon \mathcal{F} \to \mathcal{A}$ be an elicitable functional. Let $F_0, F_1 \in \mathcal{F}$ and $F_\lambda = (1-\lambda)F_0 + \lambda F_1 \in \mathcal{F}$ for some $\lambda \in (0,1)$. Then

$$T(F_0) = T(F_1) \implies T(F_\lambda) = T(F_0)$$

Remarks:

- This shows that the variance or ES are generally not elicitable.

$$\mathrm{Var}(\delta_x) = \mathrm{Var}(\delta_y) = 0, \quad \mathrm{Var}\big(\lambda\delta_x + (1-\lambda)\delta_y\big) = \lambda(1-\lambda)(x-y)^2\,.$$

- Steinwart et al. (2014) showed that for $\mathcal{A} = \mathbb{R}$ and under some continuity assumptions on T, CxLS are also sufficient for elicitability.
- This argument is independent of the dimension of T.
- For $k > 1$, CxLS are generally not sufficient, e.g., $(\mathrm{VaR}_\alpha, \mathrm{CoVaR}_{\alpha|\beta})$.

The wondrous tale about the mode

The wondrous tale about the mode

- The mode is the argmax of the counting / Lebesgue density.
- The mode functional has CxLS.
- On classes of discrete distributions only (say on $\mathbb{N}$), it is elicitable with the zero-one loss:

$$L(x, y) = \mathbb{1}\{x \neq y\}$$

- What about absolutely continuous distributions? Clearly, the zero-one loss is constant almost surely. But are there other candidates?

Theorem 5 (Heinrich-Mertsching and F. (Biometrika, 2022))

The mode is not elicitable on $\mathcal{F}_0$, the class of continuous and strongly unimodal densities on $\mathbb{R}$.
(And hence it fails to be elicitable on any superclass $\mathcal{F}$ of $\mathcal{F}_0$.)

This result substantially strengthens the result of Heinrich (2014) which establishes the non-elicitability on the class containing *all* absolutely continuous distributions on $\mathbb{R}$.

The wondrous tale about the mode (continued)

Proof:

- Key observation: The mode fails to be continuous:

  For any $a < b$ there are sequences of densities $f_n, g_n \in \mathcal{F}_0$
  - $\mathrm{mode}(f_n) =: x_1 \neq x_2 := \mathrm{mode}(g_n)$ for all $n \in \mathbb{N}$, $x_1, x_2 \in [a, b]$;
  - f_n and g_n converge pointwise to the uniform distribution on $[a, b]$ (which is not contained in $\mathcal{F}_0$).
- If a strictly $\mathcal{F}_0$-consistent loss function L existed, this would imply that

$$\int_a^b L(x_1, y)\mathrm{d}y = \int_a^b L(x_2, y)\mathrm{d}y \tag{1}$$

- Since (1) holds for all $a < b$, the Radon–Nikodym theorem implies that

$$L(x_1, y) = L(x_2, y) \text{ for almost all } y. \tag{2}$$

- (2) shows that L cannot be strictly $\mathcal{F}_0$-consistent. □

Range Value at Risk – a linear combination of Bayes risks

What to do if T is not elicitable?

A lot of (relevant) functionals are not elicitable: Variance, Expected Shortfall (ES), but also Range Value at Risk (RVaR).

$$\begin{aligned}
\mathrm{ES}_\alpha(Y) &= \frac{1}{\alpha}\int_\alpha^1 \mathrm{VaR}_\gamma(Y)\mathsf{d}\gamma \\
&\overset{(\star)}{=} \mathrm{E}\left[Y \,|\, \mathrm{VaR}_\alpha(Y) \leqslant Y\right] \\
\mathrm{RVaR}_{\alpha,\beta}(Y) &= \frac{1}{\beta-\alpha}\int_\alpha^\beta \mathrm{VaR}_\gamma(Y)\mathsf{d}\gamma \\
&\overset{(\star)}{=} \mathrm{E}\left[Y \,|\, \mathrm{VaR}_\alpha(Y) \leqslant Y \leqslant \mathrm{VaR}_\beta(Y)\right]
\end{aligned}$$

- $\mathrm{RVaR}_{\alpha,\beta}$ is an interpolation of ES_α and VaR_α.
- It is robust, but not coherent.
- Moreover, $\mathrm{RVaR}_{\alpha,1-\alpha}$ is a trimmed mean.

What to do if T is not elicitable?

Variance and ES can be written as the Bayes risk of a consistent loss function.

$$\mathrm{Var}(Y) = \min_{x \in \mathbb{R}} \mathrm{E}\left[(x - Y)^2\right]$$

$$\mathrm{ES}_\alpha(Y) = \min_{x \in \mathbb{R}} \mathrm{E}\left[\frac{1}{\alpha} S_\alpha(x, Y)\right], \quad S_\alpha(x, y) = (\mathbb{1}\{y \leqslant x\} - \alpha)x - \mathbb{1}\{y \leqslant x\}y$$

Theorem 6

Let T be elicitable with strictly consistent loss S. Then $(T, T^\star)$ is jointly elicitable where

$$T^\star(F) = \min_{x \in \mathcal{A}} \mathrm{E}_{Y \sim F}\left[S(x, Y)\right].$$

A strictly consistent loss for $(T, T^\star)$ is given by

$$L(x_1, x_2; y) = \phi'(x_2)\big(x_2 - S(x_1, y)\big) - \phi(x_2) + L_T(x_1, y),$$

where ϕ is strictly convex, $\phi' < 0$, and L_T is a consistent loss for T.

$$L(x_1, x_2; y) = \phi'(x_2)\big(x_2 - S(x_1, y)\big) - \phi(x_2) + L_T(x_1, y),$$

Idea:

- For fixed x_2, the map

$$(x_1, y) \mapsto L(x_1, x_2; y) = -\phi'(x_2) S(x_1, y) + L_T(x_1, y) + \kappa(x_2)$$

  is strictly consistent for T, since $\phi' < 0$.

- For fixed x_1, the map

$$(x_2, y) \mapsto L(x_1, x_2; y) = \phi'(x_2)\big(x_2 - S(x_1, y)\big) - \phi(x_2) + \dots$$

  is strictly consistent for $F \mapsto \mathrm{E}_{Y \sim F} S(x_1, Y)$, since ϕ is convex.

Corollary 7 (F and Ziegel (AoS, 2016))

The pairs (mean, variance) and $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ *are elicitable!*

Elicitability of RVaR

RVaR is the scaled difference of Bayes risks!

$$\mathrm{ES}_\alpha(Y) = \frac{1}{\beta-\alpha}\left(\min_{x\in\mathbb{R}} \mathrm{E}\left[S_\alpha(x,Y)\right] - \min_{x\in\mathbb{R}} \mathrm{E}\left[S_\beta(x,Y)\right]\right)$$

Theorem 8 (F and Ziegel (Stat. Risk Model., 2021))

$$\begin{aligned} L(x_1,x_2,x_3;y) &= (\mathbb{1}\{y\leqslant x_1\}-\alpha)(g_1(x_1)-g_1(y)) \\ &\quad + (\mathbb{1}\{y\leqslant x_2\}-\beta)(g_2(x_2)-g_2(y)) \\ &\quad + \phi'(x_3)\left(x_3 - \frac{1}{\beta-\alpha}(S_\alpha(x_1,y)-S_\beta(x_2,y))\right) - \phi(x_3) \end{aligned} \tag{3}$$

is strictly consistent for $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ *if*

- ϕ *is strictly convex;*
- *for all* x_3*:* $x_1 \mapsto g_1(x_1) - x_1\phi'(x_3)/(\beta-\alpha)$ *is strictly increasing;*
- *for all* x_3*:* $x_2 \mapsto g_2(x_2) + x_2\phi'(x_3)/(\beta-\alpha)$ *is strictly increasing.*

Any strictly consistent loss is essentially of the form (3).

The dichotomy of the set-valued elicitation world

Set-valued functionals

The functional T maps to a subset of $\mathcal{P}(\mathbb{R}^k)$.

- **Mode:** $\mathrm{mode}(F) = \mathrm{argmax}_x f(x)$
- **Quantiles:** $q_\alpha(F) := \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\}$
- **Prediction intervals:** Any $[a, b]$ s.t. $F([a, b]) := F(b) - F(a-) \geqslant \alpha$
- **Systemic risk measures:** $R(F_Y) = \{k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) \leqslant 0\}$, see Feinstein, Rudloff and Weber (2017)
- **Functionals of random sets:** Climatology, reliability engineering, medicine, econometrics; see Molchanov (2017); Molchanov and Molinari (2018).

Selective vs. exhaustive forecasts

- Example of the α-quantile $q_\alpha \colon \mathbb{R} \to \mathcal{P}(\mathbb{R})$

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- Choice of the action domain $\mathcal{A}$:

  $\mathcal{A}_{\mathrm{sel}} \subseteq \mathbb{R}$: The forecasts are points in $\mathbb{R}$. There are multiple best actions, namely each selection $x \in q_\alpha(F)$.
  $\rightsquigarrow$ Selective forecasts

  $\mathcal{A}_{\mathrm{exh}} \subseteq \mathcal{P}(\mathbb{R})$: The forecasts are subsets of $\mathbb{R}$. There is a unique best action namely to report the entire set $B = q_\alpha(F)$.
  $\rightsquigarrow$ Exhaustive forecasts

Two modes of elicitability

Definition 9 (Elicitability)

(a) A functional $T\colon \mathcal{F} \to \mathcal{P}(\mathcal{A}_{\mathrm{sel}})$ is selectively elicitable if there is a selective loss function $L_{\mathrm{sel}}\colon \mathcal{A}_{\mathrm{sel}} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathrm{E}_F[L_{\mathrm{sel}}(t, Y)] \leqslant \mathrm{E}_F[L_{\mathrm{sel}}(x, Y)]$$

for all $F \in \mathcal{F}$, for all $t \in T(F)$, for all $x \in \mathcal{A}_{\mathrm{sel}}$ and where equality implies that $x \in T(F)$.

(b) A functional $T\colon \mathcal{F} \to \mathcal{A}_{\mathrm{exh}}$ is exhaustively elicitable if there is an exhaustive loss function $L_{\mathrm{exh}}\colon \mathcal{A}_{\mathrm{exh}} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathrm{E}_F[L_{\mathrm{exh}}(T(F), Y)] \leqslant \mathrm{E}_F[L_{\mathrm{exh}}(B, Y)]$$

for all $F \in \mathcal{F}$, for all $B \in \mathcal{A}_{\mathrm{exh}}$ and where equality implies that $B = T(F)$.

Mutual exclusivity results

Theorem 10 (F., Frongillo, Hlavinová, Rudloff (EJS, 2021))

If there exist $F, G \in \mathcal{F}$ *such that* $\emptyset \neq T(F_0) \subsetneq T(F_1)$ *and* $(1-\lambda)F_0 + \lambda F_1 \in \mathcal{F}$ *for all* $\lambda \in (0,1)$ *then:*

(i) *If* T *is selectively elicitable, it is not exhaustively elicitable.*

(ii) *If* T *is exhaustively elicitable, it is not selectively elicitable.*

Main idea for proofs:

Exploit "linearity" of the expected loss $\bar{L}(x, F)$ in its second argument and use a refinement of the fact that convex level sets are necessary for elicitability.

Examples: Mode and Quantiles

- If all $F \in \mathcal{F}_{count}$ have countable support, the mode is selectively elicitable on $\mathcal{F}_{count}$ with the loss

$$L_{mode}(x, y) = \mathbb{1}\{x \neq y\}.$$

$\rightsquigarrow$ The mode is generally not exhaustively elicitable.

- The α-quantile is selectively elicitable with a strictly consistent loss

$$L_\alpha(x, y) = |\mathbb{1}\{y \leqslant x\} - \alpha||x - y|.$$

$\rightsquigarrow$ The α-quantile is generally not exhaustively elicitable.

- One can also show that the lower quantile (or any other selection of it) is in general not elicitable!

Examples: Prediction intervals and systemic risk measures

- The class of α-prediction intervals is exhaustively elicitable.
- The class of shortest α-prediction interval is neither selectively nor exhaustively elicitable.
- Systemic risk measures of the form $R(F_Y) = \{k \in \mathbb{R}^d \mid \rho(\Lambda(Y + k)) \leqslant 0\}$ are exhaustively elicitable, if ρ is elicitable. (F, Hlavinová, Rudloff (Fin. Stoch., 2021)).

Closed Random Sets

- Let $(\Omega, \mathfrak{F}, \mathbf{P})$ be a non-atomic probability space.
- A closed random set $\boldsymbol{Y}$ is a map from Ω into the collection $\mathfrak{U}$ of closed sets in $\mathbb{R}^d$ (or some general separable Banach space).
- It is measurable if for all compact sets $K \subseteq \mathbb{R}^d$

$$\{\omega \,|\, \boldsymbol{Y}(\omega) \cap K \neq \emptyset\} \in \mathfrak{F}.$$

  See Molchanov (2017) for details.
- Examples:
  - region of a blackout in a country
  - region affected by a flood, avalanche, disease
  - tumorous tissue in the human body
- There are interesting set-valued functionals of random sets:
  - Vorob'ev quantiles
  - Vorob'ev expectation
  - Selection expectation ($\approx$ Minkowski average)

Vorob'ev Quantiles of Closed Random Sets

Definition 11

The upper excursion set of the coverage function $u \mapsto \mathbf{P}(u \in \boldsymbol{Y})$ at level $\alpha \in [0, 1]$,

$$Q_\alpha(\boldsymbol{Y}) := \{u \in \mathbb{R}^d \,|\, \mathbf{P}(u \in \boldsymbol{Y}) \geqslant \alpha\},$$

is called the *Vorob'ev α-quantile* of $\boldsymbol{Y}$.

$Q_\alpha(\boldsymbol{Y})$ is always a closed set.

Theorem 12 (F., Frongillo, Hlavinová, Rudloff (EJS, 2021))

(i)

$$L\colon \mathfrak{U} \times \mathfrak{U} \to [0, \infty], \quad L(X, Y) = \alpha\mu(X \backslash Y) + (1 - \alpha)\mu(Y \backslash X),$$

is a non-negative $\mathcal{F}$-consistent loss function for $Q_\alpha/$

(ii) *If $Q_\alpha(F) = \mathrm{cl}(Q_\alpha^{>}(F))$ and $Q_\alpha(F) = \mathrm{cl}(\mathrm{int}(Q_\alpha(F))$ for all $F \in \mathcal{F}$, then Q_α is exhaustively elicitable on $\mathcal{F}$.*
Moreover, for any σ-finite positive measure μ on $\mathbb{R}^d$ such that $\mathrm{E}_F[\mu(\boldsymbol{Y})] < \infty$ and $\pi(Q_\alpha(F)) < \infty$ for all $F \in \mathcal{F}$, the restriction of L to the family $\mathfrak{U}'\colon = \{U \in \mathfrak{U} \,|\, U = \mathrm{cl}(\mathrm{int}(U))\}$ is a strictly $\mathcal{F}$-consistent exhaustive loss function for Q_α.

Interpretation of loss

$$L(X, Y) = \alpha\mu(X\setminus Y) + (1-\alpha)\mu(Y\setminus X)$$

Decomposition into

- false positive $X\setminus Y$
- false negative $Y\setminus X$

Applications:

- Evaluation of warnings (in spacetime) where asymmetric costs for false positives and false negatives are present.
- Pattern recognition in learning and diagnostics.
- Mathematical statistics: A confidence set is actually a random set.

Summary

- Loss functions play a crucial role learning and in forecast assessment and comparison.
- They should be chosen in line with the target functional of interest. $\rightsquigarrow$ They should be consistent.
- Strict consistency ensures that the oracle regression function is eventually learned. It ensures incentive compatible forecast comparison.
- We have revisited the elicitation problem.
  - CxLS are necessary for elicitability. The mode shows that they are not generally sufficient.
  - Linear combinations of Bayes risks are elicitable.
  - A refinement of the CxLS property establishes that set-valued functionals can either be selectively elicitable, exhaustively elicitable or not elicitable at all.

Omitted achievements, open questions, outlook

- **Omitted achievements:**
  - Discussion of calibration assessment with identification functions.
  - Loss functions as measures of information (generalising the coefficient of determination) (F & Pesenti, EJOR, 2023)
  - Multivariate loss functions (F & Hoga, JBES, 2023)
  - Loss functions in modern statistical learning (= machine learning) (F, Merz, & Wüthrich, IME, 2023)
- **Open questions & outlook:**
  - Replace strict consistency (ranking of expectations) by requirement that average scores rank with high probability.
  - Better understanding of generative AI such as ChatGPT. ("Small" input vector / prompt associated with very complex response)

Further Reading

- **Scoring rules for probabilistic forecasts**:

  T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation.

  Journal of the American Statistical Association, 102:359–378, 2007

- **Good introduction to elicitability**:

  T. Gneiting. Making and evaluating point forecasts.

  Journal of the American Statistical Association, 106(494):746–762, 2011

- **Traditional and Comparative backtests**:

  N. Nolde and J. F. Ziegel. Elicitability and backtesting: Perspectives for banking regulation.

  The Annals of Applied Statistics, 11(4):1833–1874, 2017

Additional References

- A. Philip Dawid and Vladimir G. Vovk. Prequential probability: principles and properties.
  Bernoulli, 5(1):125–162, 2 1999
- Z. Feinstein, B. Rudloff, and S. Weber. Measures of systemic risk.
  SIAM Journal on Financial Mathematics, 8:672–708, 2017
- C. Heinrich. The mode functional is not elicitable.
  Biometrika, 101(1):245–251, 2014
- I. Molchanov. *Theory of Random Sets.*
  Probability Theory and Stochastic Modelling. Springer-Verlag London, London, 2 edition, 2017
- I. Molchanov and F. Molinary. *Random Sets in Econometrics.*
  Cambridge University Press, 2018
- I. Steinwart, C. Pasin, R. Williamson, and S. Zhang. Elicitation and Identification of Properties.
  JMLR Workshop Conf. Proc., 35:1–45, 2014

Thank you for your attention!

Looking forward to our discussion!