A Unifying Frame for Neighbourhood and Distortion Models

Enrique Miranda, Ignacio Montes, Sébastien Destercke

Univ. of Oviedo (Spain) and Univ. Téchnologique de Compiègne (France)

(mirandaenrique,imontes)@uniovi.es; sebastien.destercke@hds.utc.fr

ISIPTA'2019, Gent

What people expect to find when they come to Oviedo...

...and what they get (to their horror)

Summary

Summary of the paper

One family of imprecise probability models are those that arise from distorting *somehow* a fixed probability measure P_0 by some factor $\delta > 0$, representing:

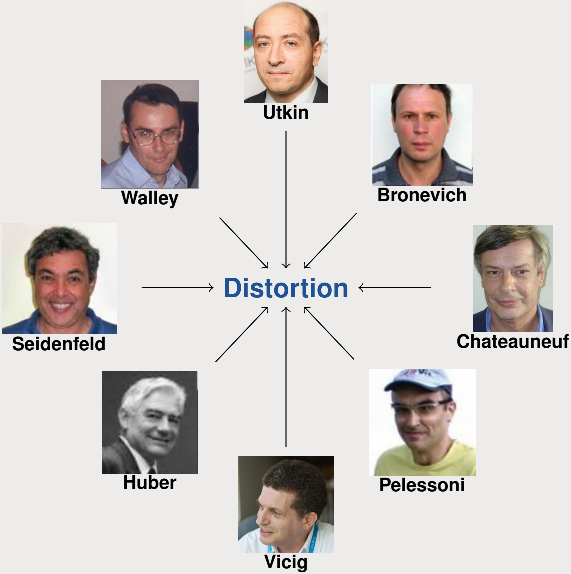
- the amount of contaminated data;
- a taxation from the house;
- the distance from the original model we are sensitive to;
- ...

- ▶ The goal of the paper is to compare a number of possible distortion models.

Here, we consider a finite space $\mathcal{X}$ and assume that $\forall x \in \mathcal{X}$ $P_0(\{x\}) > 0$ and that $\delta > 0$ is *small enough* (but this can be generalised).

Does anybody care?

Does anybody care?



Examples of distortion models (I)

1. Pari-mutuel model

$$\underline{P}_{PMM}(A) = \max\{0, (1+\delta)P_0(A) - \delta\}.$$

2. Linear-vacuous mixture

$$\underline{P}_{LV}(\mathcal{X}) = 1, \quad \underline{P}_{LV}(A) = (1-\delta)P_0(A) \quad \forall A \neq \mathcal{X}.$$

3. Constant odds ratio on gambles
$\underline{P}_{COR}(f)$ is the unique solution of

$$(1-\delta)P_0((f-\underline{P}_{COR}(f))^+) = P_0((f-\underline{P}_{COR}(f))^-).$$

4. Constant odds ratio on events

$$\underline{Q}_{COR}(A) = \frac{(1-\delta)P_0(A)}{1-\delta P_0(A)}.$$

Examples of distortion models (II)

Given a distance d, we can consider the credal set

$$\mathcal{M}(P_0, d, \delta) = \{P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta\}$$

and its lower envelope $\underline{P}_d$. In this way we can consider:

5. Total variation

$$d_{TV}(P, Q) = \sup_{A \subset \mathcal{X}} |P(A) - Q(A)|.$$

6. Kolmogorov

$$d_K(P, Q) = \sup_{x \in \mathcal{X}} |F_P(x) - F_Q(x)|,$$

assuming $\mathcal{X}$ totally ordered.

Distortion in terms of $\underline{P}$ or from a ball?

The first examples can also be obtained as envelopes of neighbourhoods induced by some d:

1. $\underline{P}_{PMM}$: $d_{PMM}(P,Q) = \max_{A \neq \mathcal{X}} \frac{Q(A)-P(A)}{1-Q(A)}$.
2. $\underline{P}_{LV}$: $d_{LV}(P,Q) = \max_{A \neq \emptyset} \frac{Q(A)-P(A)}{Q(A)}$.
3. $\underline{P}_{COR}$: $d_{COR}(P,Q) = \max_{A,B \neq \emptyset} \left\{1 - \frac{P(A) \cdot Q(B)}{P(B) \cdot Q(A)}\right\}$.
4. $\underline{Q}_{COR}$: $d'_{COR}(P,Q) = \max_{A \neq \mathcal{X}, \emptyset} \left\{1 - \frac{P(A)}{P(A^c)} \frac{Q(A^c)}{Q(A)}\right\}$.

In fact, something similar applies to what are called distortion models, given by $f(P_0)$ for some f.

Comparison between the models

We have compared the different examples according to the following criteria:

- How large is the credal set obtained when distorting P_0 by some fixed factor δ.
- The number of extreme points of this credal set.
- The properties of the associated coherent lower probability.
- The properties of the distorting function d.

Amount of imprecision

If we fix P_0 and δ, we can compare the amount of imprecision between the different neighbourhood models:

$$
\begin{tikzcd}
 & \mathcal{M}(\underline{P}_K) \arrow[d] & \\
 & \mathcal{M}(\underline{P}_{TV}) \arrow[dl] \arrow[dr] & \\
\mathcal{M}(\underline{P}_{PMM}) \arrow[dr] & & \mathcal{M}(\underline{P}_{LV}) \arrow[dl] \\
 & \mathcal{M}(\underline{Q}_{COR}) \arrow[d] & \\
 & \mathcal{M}(\underline{P}_{COR}) &
\end{tikzcd}
$$

Here, an arrow between two nodes means that parent includes the child.

Properties of the lower probabilities

Model	2-monotone	∞-monotone	Prob. interval
$\underline{P}_{PMM}$	YES	NO	YES
$\underline{P}_{LV}$	YES	YES	YES
$\underline{P}_{TV}$	YES	NO	NO
$\underline{P}_{COR}$	NO	NO	NO
$\underline{Q}_{COR}$	YES	YES	NO
$\underline{P}_{K}$	YES	YES	NO

Thus, the most precise model, that was the constant odds ratio (on gambles), is the one with worse properties, while the best is the linear vacuous.

Number of extreme points of $\mathcal{M}(\underline{P})$

In terms of the maximum number of extreme points of the neighbourhood model, we have the following:

Model	Maximal number of extreme points
$\underline{P}_{PMM}$	$\frac{n!}{\lfloor\frac{n}{2}\rfloor\left(\lfloor\frac{n}{2}\rfloor-1\right)!\left(n-\lfloor\frac{n}{2}\rfloor-1\right)!}$
$\underline{P}_{LV}$	n
$\underline{P}_{TV}$	$\frac{n!}{\left(\lfloor\frac{n}{2}\rfloor-1\right)!\left(n-\lfloor\frac{n}{2}\rfloor-1\right)!}$
$\underline{P}_{COR}$	2^n-2
$\underline{Q}_{COR}$	$n!$
$\underline{P}_{K}$	$\mathcal{P}_n$

where $\mathcal{P}_n$ denotes the n-th Pell number. The best is the linear vacuous and the worst is the constant odds ratio on events.

Conclusions and further work

- Distortion models can be seen as neighbourhood models induced by some d.
- The linear vacuous seems to be the best overall model, although this depends on the property of interest.
- The analysis of the imprecision can be done by means of other measures.

Additional results, to be reported elsewhere:

- Study of the model induced by the L_1 distance.
- Combination of distortion models.
- Study when the model is preserved by conditioning.

We eagerly look forward to your questions in the poster