Weighted Total Variation distances for contextual robustification

David Nieto-Barba, Sébastien Destercke, Enrique Miranda

University of Oviedo, University of Technology of Compiègne

IPMU 2026. Rome, June 16, 2026

Summary

- A way to robustify an uncertainty model is to consider a neighbourhood around a probability measure.
- Usually, in these distortion models the neighbourhood is defined uniformly in all directions.
- We propose a non-uniform distortion model, based on an extension of the total variation distance.
- We determine the number of extreme points, and analyse the properties of the associated lower prevision.

Distortion models

Let $\mathcal{X}$ be a finite set, $\mathbb{P}(\mathcal{X})$ the family of probability measures on $\mathcal{P}(\mathcal{X})$ and $\mathbb{P}^*(\mathcal{X})$ those that are strictly positive on all non-empty events.

Given $P_0 \in \mathbb{P}(\mathcal{X})$, a distorting function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \to [0, +\infty)$ and a distortion parameter $\delta \geq 0$, the neighbourhood of P_0 through d with parameter δ is defined as:

$$B_d^\delta(P_0) := \{P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta\}.$$

As particular cases, we have the contamination, pari-mutuel, constant odds-ratio or the total variation models.

Distortion models

Let $\mathcal{X}$ be a finite set, $\mathbb{P}(\mathcal{X})$ the family of probability measures on $\mathcal{P}(\mathcal{X})$ and $\mathbb{P}^*(\mathcal{X})$ those that are strictly positive on all non-empty events.

Given $P_0 \in \mathbb{P}(\mathcal{X})$, a distorting function $d : \mathbb{P}(\mathcal{X}) \times \mathbb{P}(\mathcal{X}) \to [0, +\infty)$ and a distortion parameter $\delta \geq 0$, the neighbourhood of P_0 through d with parameter δ is defined as:

$$B_d^{\delta}(P_0) := \left\{ P \in \mathbb{P}(\mathcal{X}) \mid d(P, P_0) \leq \delta \right\}.$$

As particular cases, we have the contamination, pari-mutuel, constant odds-ratio or the total variation models.

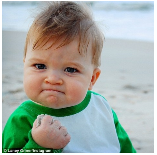

The total variation model

In the total variation model, the distorting function is the total variation distance:

$$d_{\mathrm{TV}}(P, P_0) := \max_{A \subseteq \mathcal{X}} \left| P(A) - P_0(A) \right| = \frac{1}{2} \sum_{x \in \mathcal{X}} \left| P(\{x\}) - P_0(\{x\}) \right|.$$

The resulting model has a number of nice properties:

- The neighbourhood has $n(n-1)$ extreme points, being computationally simpler than other distortion models.
- The associated lower prevision is 2-monotone.
- It behaves well under conditioning.



Motivation

However, in some cases we may aim for a non-uniform distortion:

- P_0 may represent a distribution of resources, and we may want to allow for transfers of resources that depend on the initial wealth.
- If P_0 is in the interior of the simplex, we may want the neighbourhood to stay in the interior so as to avoid zero lower probabilities, and this would require a smaller distortion in the directions that are closer to the boundary.
- P_0 may represent normalised utilities in a multicriteria decision making problem, and we may use weights to highlight the importance of each of the criteria.

The weighted total variation

Given $\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ such that $\alpha \succeq 0$, the α-Weighted Total Variation (α-WTV) is given by

$$d_{\mathrm{TV}}^{\alpha}(P, P_0) := \sum_{i=1}^{n} \alpha_i |P(\{x_i\}) - P_0(\{x_i\})| \quad \forall P, P_0 \in \mathbb{P}(\mathcal{X}).$$

If $\alpha_i = {}^1\!/_2$ for all i, we recover the TV distance.

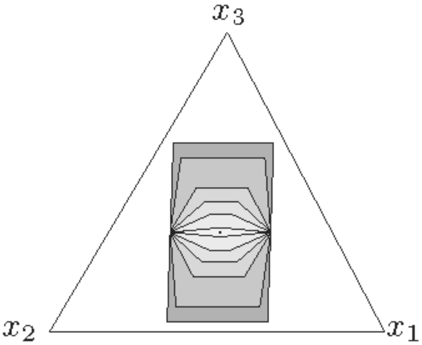
Basic properties

- For any P_0, $d^{\alpha}_{\mathrm{TV}}(\cdot, P_0)$ is continuous and convex.
- $d^{\alpha}_{\mathrm{TV}}(P, P) = 0$ for all $P \in \mathbb{P}(\mathcal{X})$.
- $d^{\alpha}_{\mathrm{TV}}(Q, P) = d^{\alpha}_{\mathrm{TV}}(P, Q)$.
- $d^{\alpha}_{\mathrm{TV}}(Q, P) \leq d^{\alpha}_{\mathrm{TV}}(Q, R) + d^{\alpha}_{\mathrm{TV}}(R, P)$ for every $P, Q, R \in \mathbb{P}(\mathcal{X})$.
- $d^{\alpha}_{\mathrm{TV}}(Q, P) = 0 \Rightarrow P = Q$ if and only if at most one $\alpha_i = 0$.

$\hookrightarrow$ Thus, d^{α}_{TV} is a distance if and only if at most one component of α is null.

Example

Let $P_0 = (1/3, 1/3, 1/3), \alpha = (0.5, 0.5, \alpha_3)$ and $\delta = \frac{3}{20}$. The neighbourhoods $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ for $\alpha_3 \in \{0, 0.1, 0.5, 1, 2, 10\}$ are given by:



Null weights

If more than one α_i is null, we have a number of undesirable properties:

- d^{α}_{TV} is no longer a distance.
- We do not have $B^{0}_{d^{\alpha}_{\mathrm{TV}}}(P_0) = \{P_0\}$.
- The neighbourhood $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ always touches the boundary of the simplex, even if P_0 is in the interior.

Null weights

If more than one α_i is null, we have a number of undesirable properties:

- d^{α}_{TV} is no longer a distance.
- We do not have $B^0_{d^{\alpha}_{\mathrm{TV}}}(P_0) = \{P_0\}$.
- The neighbourhood $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ always touches the boundary of the simplex, even if P_0 is in the interior.

$\hookrightarrow$ For this reason, we will focus on the case where at most one weight is null.

Case 1: $\alpha \succ 0$

Consider the case where all the weights α_i are strictly positive, and $P_0 \in \mathbb{P}^*(\mathcal{X})$. Let

$$\overline{\delta} := \min\left\{(\alpha_i + \alpha_j)P_0(\{x_j\}) \mid i, j \in \{1, \ldots, n\},\ i \neq j\right\}.$$

- $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ if and only if $\delta < \overline{\delta}$.
- If $\delta < \overline{\delta}$, then

$$\operatorname{ext}\left(B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)\right) = \{P_{ij} \mid i, j \in \{1, \ldots, n\},\ i \neq j\},$$

  where, for each $i, j \in \{1, \ldots, n\}$ with $i \neq j$:

$$P_{ij}(\{x\}) := \begin{cases} P_0(\{x\}) + (\alpha_i + \alpha_j)^{-1}\delta, & \text{if } x = x_i; \\ P_0(\{x\}) - (\alpha_i + \alpha_j)^{-1}\delta, & \text{if } x = x_j; \\ P_0(\{x\}), & \text{if } x \neq x_i, x_j. \end{cases}$$

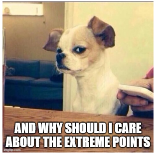
Case 1: $\alpha \succ 0$

Consider the case where all the weights α_i are strictly positive, and $P_0 \in \mathbb{P}^*(\mathcal{X})$. Let

$$\overline{\delta} := \min\left\{(\alpha_i + \alpha_j)P_0(\{x_j\}) \mid i, j \in \{1, \ldots, n\},\ i \neq j\right\}.$$

- $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ if and only if $\delta < \overline{\delta}$.
- If $\delta < \overline{\delta}$, then
$$\mathrm{ext}\left(B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)\right) = \{P_{ij} \mid i, j \in \{1, \ldots, n\},\ i \neq j\},$$
where, for each $i, j \in \{1, \ldots, n\}$ with $i \neq j$:
$$P_{ij}(\{x\}) := \begin{cases} P_0(\{x\}) + (\alpha_i + \alpha_j)^{-1}\delta, & \text{if } x = x_i; \\ P_0(\{x\}) - (\alpha_i + \alpha_j)^{-1}\delta, & \text{if } x = x_j; \\ P_0(\{x\}), & \text{if } x \neq x_i, x_j. \end{cases}$$

$\hookrightarrow$ Thus, $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ has exactly $n(n-1)$ extreme points!



Extreme points

The extreme points of the set allow us to:

- Characterise the set using convex combinations.
- Determine the lower and upper probabilities and previsions that summarise it.
- Update the set by regular extension, which is an imprecise rule for conditioning.

$\hookrightarrow$ As a consequence, we would like them to be as few as possible and to have an explicit formula to compute them.

Lower prevision

If $\delta < \overline{\delta}$, the coherent lower prevision associated with $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ is given by

$$\underline{P}^{\alpha}_{\mathrm{TV}}(f) = P_0(f) - \delta \max_{i \neq j} \left\{ |f(x_i) - f(x_j)|(\alpha_i + \alpha_j)^{-1} \right\} \quad \forall f \in \mathcal{L}(\mathcal{X}),$$

and the coherent lower probability is

$$\underline{P}^{\alpha}_{\mathrm{TV}}(A) = P_0(A) - \delta \left(\min_{i|x_i \notin A} \alpha_i + \min_{j|x_j \in A} \alpha_j \right)^{-1} \quad \forall A \neq \emptyset, \mathcal{X}.$$

Case 2: $\alpha \succeq 0$

Consider the case where exactly one α_{k*} is zero, and $P_0 \in \mathbb{P}^*(\mathcal{X})$. Let

$$\overline{\delta} := \min \Big\{ \min \left\{ \alpha_i P_0(\{x_{k^*}\}) \mid i \in \{1, \dots, n\} \setminus \{k^*\} \right\}, \min \left\{ \alpha_j P_0(\{x_j\}) \mid j \in \{1, \dots, n\} \setminus \{k^*\} \right\} \Big\}$$

- $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0) \subseteq \mathbb{P}^*(\mathcal{X})$ if and only if $\delta < \overline{\delta}$.
- If $\delta < \overline{\delta}$, then

$$\mathrm{ext}\left(B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)\right) = \{P_{k^*j} \mid j \in \{1, \dots, n\} \setminus \{k^*\}\} \cup \{P_{ik^*} \mid i \in \{1, \dots, n\} \setminus \{k^*\}\}.$$

$\hookrightarrow$ As a consequence, $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ has exactly $2(n-1)$ extreme points.

Lower prevision

If $\delta < \overline{\delta}$, the coherent lower prevision associated with $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ is given by

$$\underline{P}^{\alpha}_{\mathrm{TV}}(f) = P_0(f) - \delta \max_{i \neq k^*} \left\{ |f(x_i) - f(x_{k^*})| \alpha_i^{-1} \right\} \quad \forall f \in \mathcal{L}(\mathcal{X}),$$

and the coherent lower probability is

$$\underline{P}^{\alpha}_{\mathrm{TV}}(A) = P_0(A) - \delta \left(\min_{i | x_i \notin A} \alpha_i + \min_{j | x_j \in A} \alpha_j \right)^{-1} \quad \forall A \neq \emptyset, \mathcal{X}.$$

2-monotonicity

The lower probability and the lower prevision do not give the same information, in the sense that $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ is a proper subset of

$$\{P \in \mathbb{P}^*(\mathcal{X}) : P(A) \geq \underline{P}^{\alpha}_{\mathrm{TV}}(A)\ \forall A\}.$$

A sufficient condition for the equality would be that $\underline{P}^{\alpha}_{\mathrm{TV}}$ is 2-monotone:

$$\underline{P}^{\alpha}_{\mathrm{TV}}(f_1 \vee f_2) + \underline{P}^{\alpha}_{\mathrm{TV}}(f_1 \wedge f_2) \geq \underline{P}^{\alpha}_{\mathrm{TV}}(f_1) + \underline{P}^{\alpha}_{\mathrm{TV}}(f_2) \quad \forall f_1, f_2 \in \mathcal{L}(\mathcal{X}).$$

IMAGINE

2-monotonicity

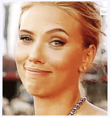
The lower probability and the lower prevision do not give the same information, in the sense that $B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0)$ is a proper subset of

$$\{P \in \mathbb{P}^*(\mathcal{X}) : P(A) \geq \underline{P}^{\alpha}_{\mathrm{TV}}(A)\ \forall A\}.$$

A sufficient condition for the equality would be that $\underline{P}^{\alpha}_{\mathrm{TV}}$ is 2-monotone:

$$\underline{P}^{\alpha}_{\mathrm{TV}}(f_1 \vee f_2) + \underline{P}^{\alpha}_{\mathrm{TV}}(f_1 \wedge f_2) \geq \underline{P}^{\alpha}_{\mathrm{TV}}(f_1) + \underline{P}^{\alpha}_{\mathrm{TV}}(f_2) \quad \forall f_1, f_2 \in \mathcal{L}(\mathcal{X}).$$

Lower previsions, 2-monotone ones, neighbourhood models...I wish there was a place to discuss this calmly!



12th SIPTA School on Imprecise Probabilities
Munich, July 27-31, 2026

SIPTA

https://www.sipta.org/

So when do we have the equality?

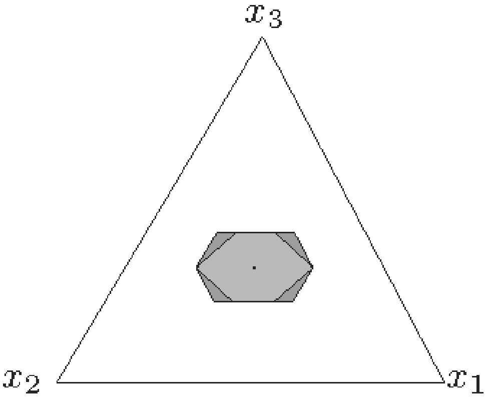
If $P_0 \in \mathbb{P}^*(\mathcal{X})$ and $0 < \delta < \overline{\delta}$, then

$$B^{\delta}_{d^{\alpha}_{\mathrm{TV}}}(P_0) = \mathcal{M}(\underline{P}^{\alpha}_{\mathrm{TV}}) \Leftrightarrow \left[n = 2 \text{ or } \alpha_1 = \cdots = \alpha_n > 0\right].$$

$\hookrightarrow$ Consequently, $\underline{P}^{\alpha}_{\mathrm{TV}}$ is 2-monotone if and only if the previous condition holds.

Example

Let $P_0 = (1/3, 1/3, 1/3), \alpha = (0.5, 0.5, 1)$ and $\delta = 3/20$.



Conclusions

Summary:

- The complexity of the model is similar to the TV case, and reduces if one weight is null.
- 2-monotonicity only holds for the TV case or if $n = 2$.

Further work:

- Study of the behaviour under conditioning.
- Probabilistic dependent weights.
- Extension to the imprecise case.

References

T. Herron, T. Seidenfeld, L. Wasserman, *Divisive conditioning: further results on dilation*. Philosophy of Science 64, 411–444 (1997)

I. Montes, E. Miranda, S. Destercke, *Unifying neighbourhood and distortion models: Part II- New models and synthesis*. International Journal of General Systems 49(6), 636–674 (2020).

D. Nieto-Barba, I. Montes, E. Miranda, *The imprecise total variation model and its connections with game theory*. Fuzzy Sets and Systems 517, 109448 (2025)

S. Destercke, *A new contextual discounting rule for lower probabilities*. Proceedings of IPMU'10, Part II. pp. 198–207 (2010).

Funding

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

UNIÓN EUROPEA
Fondo Social Europeo
El FSE invierte en tu futuro

AGENCIA ESTATAL DE INVESTIGACIÓN

This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/AEI/10.13039/501100011033 and “FEDER/UE”.

Thank you for the attention...

...and for the questions!