Uncertainty functionals revisited:

Concavity and Jensen's inequality

Julien Bect & Xujia Zhu

Université Paris-Saclay, CNRS, CentraleSupélec

Laboratoire des Signaux et Systèmes (L2S), Gif-sur-Yvette, France

https://l2s.centralesupelec.fr/

mODa 14, Drongen Abbey

2026, June 15

Back to the early days of Bayesian experimental design

- Lindley [1956], DeGroot [1962] proposed to measure a priori the information brought by an experiment by the quantity

$$\Delta \;=\; H\left(\mathsf{P}^{\theta}\right) - \mathsf{E}\left(H\left(\mathsf{P}^{\theta|X}\right)\right), \tag{1}$$

Back to the early days of Bayesian experimental design

- Lindley [1956], DeGroot [1962] proposed to measure a priori the information brought by an experiment by the quantity

$$\Delta \;=\; H\left(\mathsf{P}^{\theta}\right) - \mathsf{E}\left(H\left(\mathsf{P}^{\theta|X}\right)\right), \tag{1}$$

where

  - P^{θ} is the prior distribution of θ,

Back to the early days of Bayesian experimental design

- Lindley [1956], DeGroot [1962] proposed to measure a priori the information brought by an experiment by the quantity

$$\Delta = H\left(\mathsf{P}^{\theta}\right) - \mathsf{E}\left(H\left(\mathsf{P}^{\theta|X}\right)\right), \tag{1}$$

where

  - P^{θ} is the prior distribution of θ,
  - $\mathsf{P}^{\theta|X}$ is the posterior distribution given the observation X,

Back to the early days of Bayesian experimental design

- Lindley [1956], DeGroot [1962] proposed to measure a priori the information brought by an experiment by the quantity

$$\Delta \;=\; H\left(\mathsf{P}^{\theta}\right) - \mathsf{E}\left(H\left(\mathsf{P}^{\theta|X}\right)\right), \tag{1}$$

  where
  - P^{θ} is the prior distribution of θ,
  - $\mathsf{P}^{\theta|X}$ is the posterior distribution given the observation X,
  - H is a suitable uncertainty functional.

Back to the early days of Bayesian experimental design

- Lindley [1956], DeGroot [1962] proposed to measure a priori the information brought by an experiment by the quantity

$$\Delta \;=\; H\left(\mathrm{P}^{\theta}\right) - \mathrm{E}\left(H\left(\mathrm{P}^{\theta|X}\right)\right), \tag{1}$$

  where
  - P^{θ} is the prior distribution of θ,
  - $\mathrm{P}^{\theta|X}$ is the posterior distribution given the observation X,
  - H is a suitable uncertainty functional.

- DeGroot [1962] proposed to focus on functionals with the following

Desirable property

$$H\left(\mathrm{P}^{\theta}\right) \;\geq\; \mathrm{E}\left(H\left(\mathrm{P}^{\theta|X}\right)\right), \text{ for all possible experiments}$$

Back to the early days of Bayesian experimental design

- Lindley [1956], DeGroot [1962] proposed to measure a priori the information brought by an experiment by the quantity

$$\Delta \; = \; H\left(\mathsf{P}^{\theta}\right) - \mathsf{E}\left(H\left(\mathsf{P}^{\theta|X}\right)\right), \tag{1}$$

  where
  - P^{θ} is the prior distribution of θ,
  - $\mathsf{P}^{\theta|X}$ is the posterior distribution given the observation X,
  - H is a suitable uncertainty functional.

- DeGroot [1962] proposed to focus on functionals with the following

Desirable property

$$H\left(\mathsf{P}^{\theta}\right) \; \geq \; \mathsf{E}\left(H\left(\mathsf{P}^{\theta|X}\right)\right), \text{ for all possible experiments}$$

- Remark: (1) is also used to construct sequential designs, using myopic (a.k.a. stepwise uncertainty reduction) approaches

Another use of uncertainty functionals

Global sensitivity analysis

Study of how uncertainties in a model's inputs propagate through the model and contribute to the variability of its output.

Another use of uncertainty functionals

Global sensitivity analysis

Study of how uncertainties in a model's inputs propagate through the model and contribute to the variability of its output.

Sobol' indices

Let $X = (X_1, \ldots, X_d)$ be the input vector, and let Y be a real-valued output with $\mathrm{var}(Y) < +\infty$.

Another use of uncertainty functionals

Global sensitivity analysis

Study of how uncertainties in a model's inputs propagate through the model and contribute to the variability of its output.

Sobol' indices

Let $X = (X_1, \ldots, X_d)$ be the input vector, and let Y be a real-valued output with $\mathrm{var}(Y) < +\infty$. For $u \subseteq [\![1, d]\!]$, the closed Sobol' index of the group of variables $X_u = (X_i)_{i \in u}$ is

$$S_u^{\mathrm{cl}} = \frac{\mathrm{var}\left(\mathrm{E}\left(Y \mid X_u\right)\right)}{\mathrm{var}(Y)} = \frac{\mathrm{var}(Y) - \mathrm{E}\left(\mathrm{var}\left(Y \mid X_u\right)\right)}{\mathrm{var}(Y)}.$$

Sobol' [1993], Homma and Saltelli [1996]

Another use of uncertainty functionals

Global sensitivity analysis

Study of how uncertainties in a model's inputs propagate through the model and contribute to the variability of its output.

Sobol' indices ➙ "goal-oriented" generalization

Let $X = (X_1, \ldots, X_d)$ be the input vector, and let Y be a real-valued output with $\mathrm{var}(Y) < +\infty$. For $u \subseteq [\![1, d]\!]$, the closed Sobol' index of the group of variables $X_u = (X_i)_{i \in u}$ is

$$S_u^{\mathrm{cl}} = \frac{\mathrm{var}\left(\mathsf{E}\left(Y \mid X_u\right)\right)}{\mathrm{var}(Y)} = \frac{H(\mathsf{P}^Y) - \mathsf{E}(H(\mathsf{P}^{Y|X_u}))}{H(\mathsf{P}^Y)}.$$

Fort et al. [2016], Borgonovo et al. [2021], Fissler and Pesenti [2023], ...

Generalization. Replace the variance with other uncertainty measures.

Uncertainty functionals (our definition)

Let $(\mathbb{Y}, \mathcal{Y})$ denote a measurable space (e.g., $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition

inspired by DeGroot [1962], Rao [1982b, 1984], Hainy et al. [2014], Bect et al. [2019]

Uncertainty functionals (our definition)

Let $(\mathbb{Y}, \mathcal{Y})$ denote a measurable space (e.g., $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition

Let $\mathbb{P}$ be the space of all probability measures on $\mathbb{Y}$.

inspired by DeGroot [1962], Rao [1982b, 1984], Hainy et al. [2014], Bect et al. [2019]

Possible generalizations, not covered in this talk:

a) functionals defined on a convex subset of $\mathbb{P}$;

Uncertainty functionals (our definition)

Let $(\mathbb{Y}, \mathcal{Y})$ denote a measurable space (e.g., $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition

Let $\mathbb{P}$ be the space of all probability measures on $\mathbb{Y}$. A non-negative probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ is an *uncertainty functional* if it

inspired by DeGroot [1962], Rao [1982b, 1984], Hainy et al. [2014], Bect et al. [2019]

Possible generalizations, not covered in this talk:

a) functionals defined on a convex subset of $\mathbb{P}$;

b) other sets of uncertainty values ($\overline{\mathbb{R}}$, symmetric matrices, etc.).

Uncertainty functionals (our definition)

Let $(\mathbb{Y}, \mathcal{Y})$ denote a measurable space (e.g., $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition

Let $\mathbb{P}$ be the space of all probability measures on $\mathbb{Y}$. A non-negative probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ is an *uncertainty functional* if it

1. is measurable,[†]

inspired by DeGroot [1962], Rao [1982b, 1984], Hainy et al. [2014], Bect et al. [2019]

Possible generalizations, not covered in this talk:

a) functionals defined on a convex subset of $\mathbb{P}$;

b) other sets of uncertainty values ($\overline{\mathbb{R}}$, symmetric matrices, etc.).

[†] $\mathbb{P}$ is equipped with the σ-algebra generated by the evaluation maps.

Uncertainty functionals (our definition)

Let $(\mathbb{Y}, \mathcal{Y})$ denote a measurable space (e.g., $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition

Let $\mathbb{P}$ be the space of all probability measures on $\mathbb{Y}$. A non-negative probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ is an *uncertainty functional* if it

1. is measurable,
2. vanishes on Dirac measures: $H(\delta_y) = 0$ for all $y \in \mathbb{Y}$,

inspired by DeGroot [1962], Rao [1982b, 1984], Hainy et al. [2014], Bect et al. [2019]

Possible generalizations, not covered in this talk:

a) functionals defined on a convex subset of $\mathbb{P}$;

b) other sets of uncertainty values ($\overline{\mathbb{R}}$, symmetric matrices, etc.).

Uncertainty functionals (our definition)

Let $(\mathbb{Y}, \mathcal{Y})$ denote a measurable space (e.g., $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition

Let $\mathbb{P}$ be the space of all probability measures on $\mathbb{Y}$. A non-negative probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ is an *uncertainty functional* if it

1. is measurable,
2. vanishes on Dirac measures: $H(\delta_y) = 0$ for all $y \in \mathbb{Y}$,
3. decreases on average (DoA): $H(\mathrm{P}^Y) \geq \mathrm{E}\left(H(\mathrm{P}^{Y|X})\right)$ ‡,

inspired by DeGroot [1962], Rao [1982b, 1984], Hainy et al. [2014], Bect et al. [2019]

Possible generalizations, not covered in this talk:

a) functionals defined on a convex subset of $\mathbb{P}$;

b) other sets of uncertainty values ($\overline{\mathbb{R}}$, symmetric matrices, etc.).

‡ where Y is $\mathbb{Y}$-valued and $Y|X$ admits a regular conditional distribution

Examples

1. Trace-class generalized entropies (a.k.a. ϕ-entropies) on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$

Burbea and Rao [1982], Ben-Tal and Teboulle [1986], Scarfone [2013], Tempesta [2015], ...

$$H(\nu) \;=\; \sum_{y \in \mathbb{N}} h\Big(\nu\left(\{y\}\right)\Big)$$

with $h : [0, 1] \to \overline{\mathbb{R}}_+$ concave, $h(0) = h(1) = 0$.

Examples

1. Trace-class generalized entropies (a.k.a. ϕ-entropies) on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$

Burbea and Rao [1982], Ben-Tal and Teboulle [1986], Scarfone [2013], Tempesta [2015], ...

$$H(\nu) \;=\; \sum_{y \in \mathbb{N}} h\Bigl(\nu\left(\{y\}\right)\Bigr)$$

with $h : [0, 1] \to \overline{\mathbb{R}}_+$ concave, $h(0) = h(1) = 0$.

Examples:

- $h(p) = -p\,\log(p)$ ➙ Shannon (or Boltzmann-Gibbs) entropy,
- $h(p) = p\,(1 - p)$ ➙ Gini-Simpson index, Simpson [1949]
- $h(p) = (\alpha - 1)^{-1}\,(p - p^{\alpha})$, $\alpha > 1$ ➙ α-entropy. Havrda and Charvát [1967] Tsallis [1988]

Examples (cont'd)

2. Regular quadratic uncertainty functionals (QUFs)

Rao [1982a, 1984], Rao and Nayak [1985], Lau [1985], Rao [2010], Hainy et al. [2014], ...

a.k.a. quadratic entropies

$$H(\nu) \;=\; \iint \rho(y, y')\, \nu(\mathrm{d}y)\, \nu(\mathrm{d}y')$$

with $\rho : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$ symmetric, measurable, conditionally negative definite (CND), and vanishing on the diagonal.

Examples (cont'd)

2. Regular quadratic uncertainty functionals (QUFs)

Rao [1982a, 1984], Rao and Nayak [1985], Lau [1985], Rao [2010], Hainy et al. [2014], ...

a.k.a. quadratic entropies

$$H(\nu) \;=\; \iint \rho(y, y')\, \nu(\mathrm{d}y)\, \nu(\mathrm{d}y')$$

with $\rho : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$ symmetric, measurable, conditionally negative definite (CND), and vanishing on the diagonal.

Examples:

- $\mathbb{Y} = \mathbb{R}$, $\rho(y, y') = \frac{1}{2}\,(y - y')^2$ ➙ variance,
- $\mathbb{Y} = \mathbb{R}$, $\rho(y, y') = |y - y'|$ ➙ Gini's mean difference, Yitzhaki [2003]
- $\mathbb{Y} = L^2(\mu)$, μ σ-finite, $\rho(y, y') = \|y - y'\|^2$ ➙ integrated variance,
- ... (more examples in appendix)

A foundational question

How can we characterize the DoA (decreasing on average) property?

How is this related to concavity?

A foundational question

How can we characterize the DoA (decreasing on average) property?

How is this related to concavity?

These questions have been explored by DeGroot [1962] a long time ago, but only in the case of a *finite* set $\mathbb{Y}$...

From DoA to Jensen

- Recall the DoA property:

$$H\left(\mathrm{P}^{Y}\right) \;\geq\; \mathrm{E}\left(H\left(\mathrm{P}^{Y|X}\right)\right)$$

for all (X, Y) where Y is $\mathbb{Y}$-valued and $Y|X$ admits a regular conditional distrib.

From DoA to Jensen

- Recall the DoA property:

$$H\left(\mathsf{P}^{Y}\right) \;\geq\; \mathsf{E}\left(H\left(\mathsf{P}^{Y|X}\right)\right)$$

for all (X, Y) where Y is $\mathbb{Y}$-valued and $Y|X$ admits a regular conditional distrib.

- $\boldsymbol{\nu} = \mathsf{P}^{Y|X}$ can be seen as a random measure on $\mathbb{Y}$, with expectation

$$\overline{\boldsymbol{\nu}} \;=\; \mathsf{E}\left(\boldsymbol{\nu}(\cdot)\right) \;=\; \mathsf{E}\left(\mathsf{P}\left(Y \in \cdot \mid X\right)\right) \;=\; \mathsf{P}^{Y}.$$

From DoA to Jensen (and back)

- Recall the DoA property:

$$H\left(\mathsf{P}^{Y}\right) \;\geq\; \mathsf{E}\left(H\left(\mathsf{P}^{Y|X}\right)\right)$$

for all (X, Y) where Y is $\mathbb{Y}$-valued and $Y|X$ admits a regular conditional distrib.

- $\boldsymbol{\nu} = \mathsf{P}^{Y|X}$ can be seen as a random measure on $\mathbb{Y}$, with expectation

$$\overline{\boldsymbol{\nu}} \;=\; \mathsf{E}\left(\boldsymbol{\nu}(\cdot)\right) \;=\; \mathsf{E}\left(\mathsf{P}\left(Y \in \cdot \mid X\right)\right) \;=\; \mathsf{P}^{Y}.$$

Proposition

H is DoA iff it satisfies Jensen's inequality (concave version):

$$H\left(\overline{\boldsymbol{\nu}}\right) \;\geq\; \mathsf{E}\left(H(\boldsymbol{\nu})\right)$$

for all random probability measures $\boldsymbol{\nu}$ on $(\mathbb{Y}, \mathcal{Y})$.

From Jensen's inequality to concavity

Let $H : \mathbb{P} \to \overline{\mathbb{R}}$ denote a measurable (signed) probability functional.

Theorem (part 1)

If H satisfies Jensen's inequality on $\mathbb{P}$ (concave version), then H is concave: for all $\lambda \in (0, 1)$ and all ν_1, $\nu_2 \in \mathrm{dom}(H) = \{\nu \in \mathbb{P} \mid H(\nu) > -\infty\}$,

$$H\left(\lambda\nu_1 + (1-\lambda)\nu_2\right) \;\geq\; \lambda H(\nu_1) + (1-\lambda)H(\nu_2).$$

DeGroot [1962]

From Jensen's inequality to concavity

Let $H : \mathbb{P} \to \overline{\mathbb{R}}$ denote a measurable (signed) probability functional.

Theorem (part 1)

If H satisfies Jensen's inequality on $\mathbb{P}$ (concave version), then H is concave: for all $\lambda \in (0,1)$ and all ν_1, $\nu_2 \in \mathrm{dom}(H) = \{\nu \in \mathbb{P} \mid H(\nu) > -\infty\}$,

$$H\left(\lambda\nu_1 + (1-\lambda)\nu_2\right) \;\geq\; \lambda H(\nu_1) + (1-\lambda)H(\nu_2).$$

DeGroot [1962]

Proof. Let $X \sim \mathrm{Ber}(\lambda)$. Consider the random measure

$$\boldsymbol{\nu} \;=\; \nu_1\, \mathbb{1}_{X=1} + \nu_2\, \mathbb{1}_{X=0}.$$

Then we have

$$\begin{aligned} \overline{\boldsymbol{\nu}} \;&=\; \lambda\, \nu_1 + (1-\lambda)\, \nu_2, \\ \mathsf{E}(H(\boldsymbol{\nu})) \;&=\; \lambda\, H(\nu_1) + (1-\lambda)\, H(\nu_2), \end{aligned}$$

which proves the claim. □

The case of finite spaces

DeGroot [1962] also observed that the converse holds when $\mathbb{Y}$ is finite.

Theorem (part 2)

If $\mathbb{Y}$ is finite and H is concave, then H satisfies Jensen's inequality on $\mathbb{P}$.

The case of finite spaces

DeGroot [1962] also observed that the converse holds when $\mathbb{Y}$ is finite.

Theorem (part 2)

If $\mathbb{Y}$ is finite and H is concave, then H satisfies Jensen's inequality on $\mathbb{P}$.

Proof. Assume that $\mathbb{Y} = \{1, \ldots m\}$. Then $\mathbb{P}$ can be identified with the $(m-1)$-dimensional probability simplex in $\mathbb{R}^m$:

$$\Sigma_{m-1} = \left\{ \nu \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^{m} \nu_i = 1,\ \nu_i \geq 0,\ \text{for } i = 1, \ldots, m \right\},$$

and the conclusion follows from Jensen's inequality on $\mathbb{R}^m$. $\square$

The case of finite spaces

DeGroot [1962] also observed that the converse holds when $\mathbb{Y}$ is finite.

Theorem (part 2)

If $\mathbb{Y}$ is finite and H is concave, then H satisfies Jensen's inequality on $\mathbb{P}$.

Proof. Assume that $\mathbb{Y} = \{1, \ldots m\}$. Then $\mathbb{P}$ can be identified with the $(m-1)$-dimensional probability simplex in $\mathbb{R}^m$:

$$\Sigma_{m-1} = \left\{ \nu \in \mathbb{R}^m \;\middle|\; \sum_{i=1}^{m} \nu_i = 1,\ \nu_i \geq 0,\ \text{for } i = 1, \ldots, m \right\},$$

and the conclusion follows from Jensen's inequality on $\mathbb{R}^m$. □

What if $\mathbb{Y}$ is not finite?

Main result

Theorem

Assume that $(\mathbb{Y}, \mathcal{Y}) \cong (\mathbb{R}, \mathcal{B}(\mathbb{R})) \otimes (\mathbb{Y}', \mathcal{Y}')$, for some measurable space $(\mathbb{Y}', \mathcal{Y}')$.

Remarks

- $(\mathbb{Y}, \mathcal{Y})$ must be "rich enough".

Main result: Concavity is not enough

Theorem

Assume that $(\mathbb{Y}, \mathcal{Y}) \cong (\mathbb{R}, \mathcal{B}(\mathbb{R})) \otimes (\mathbb{Y}', \mathcal{Y}')$, for some measurable space $(\mathbb{Y}', \mathcal{Y}')$. Then, there exists a concave, non-negative, measurable probability functional on $(\mathbb{Y}, \mathcal{Y})$ that vanishes on Dirac measures but fails to satisfy Jensen's inequality.

Remarks

- $(\mathbb{Y}, \mathcal{Y})$ must be "rich enough".

Main result: Concavity is not enough

Theorem

Assume that $(\mathbb{Y}, \mathcal{Y}) \cong (\mathbb{R}, \mathcal{B}(\mathbb{R})) \otimes (\mathbb{Y}', \mathcal{Y}')$, for some measurable space $(\mathbb{Y}', \mathcal{Y}')$. Then, there exists a concave, non-negative, measurable probability functional on $(\mathbb{Y}, \mathcal{Y})$ that vanishes on Dirac measures but fails to satisfy Jensen's inequality.

Remarks

- $(\mathbb{Y}, \mathcal{Y})$ must be "rich enough".
- It has been known since Perlman [1974] that Jensen's inequality does not hold in general in finite-dimensional spaces.

Main result: Concavity is not enough

Theorem

Assume that $(\mathbb{Y}, \mathcal{Y}) \cong (\mathbb{R}, \mathcal{B}(\mathbb{R})) \otimes (\mathbb{Y}', \mathcal{Y}')$, for some measurable space $(\mathbb{Y}', \mathcal{Y}')$. Then, there exists a concave, non-negative, measurable probability functional on $(\mathbb{Y}, \mathcal{Y})$ that vanishes on Dirac measures but fails to satisfy Jensen's inequality.

Remarks

- $(\mathbb{Y}, \mathcal{Y})$ must be "rich enough".
- It has been known since Perlman [1974] that Jensen's inequality does not hold in general in finite-dimensional spaces.
- The theorem asserts the existence of a counter-example on $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ that fullfills all the requirements of an uncertainty functional, with "concave" instead of "DoA".

Main result: Concavity is not enough

Theorem

Assume that $(\mathbb{Y}, \mathcal{Y}) \cong (\mathbb{R}, \mathcal{B}(\mathbb{R})) \otimes (\mathbb{Y}', \mathcal{Y}')$, for some measurable space $(\mathbb{Y}', \mathcal{Y}')$. Then, there exists a concave, non-negative, measurable probability functional on $(\mathbb{Y}, \mathcal{Y})$ that vanishes on Dirac measures but fails to satisfy Jensen's inequality.

Remarks

- $(\mathbb{Y}, \mathcal{Y})$ must be "rich enough". We don't know if the result holds for $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.
- It has been known since Perlman [1974] that Jensen's inequality does not hold in general in finite-dimensional spaces.
- The theorem asserts the existence of a counter-example on $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ that fullfills all the requirements of an uncertainty functional, with "concave" instead of "DoA".

Proof

- Let us construct a counter-example on $\mathbb{Y} = \mathbb{R}$.

Proof

- Let us construct a counter-example on $\mathbb{Y} = \mathbb{R}$.
- Consider the integrated variance functional defined by

$$H(\nu) \;=\; \int_{\mathbb{R}} \mathrm{var}_\nu \left(\mathbb{1}_{Y=s} \right) \, \mu(\mathrm{d}s),$$

where μ denotes the counting measure on $\mathbb{R}$.

Proof

- Let us construct a counter-example on $\mathbb{Y} = \mathbb{R}$.
- Consider the integrated variance functional defined by

$$H(\nu) \;=\; \int_{\mathbb{R}} \mathsf{var}_\nu \left(\mathbb{1}_{Y=s} \right) \, \mu(\mathrm{d}s),$$

where μ denotes the counting measure on $\mathbb{R}$.

- Remark: Let $\nu_y = \nu\left(\{y\}\right)$. Let $D_\nu = \{y \in \mathbb{R} \mid \nu_y > 0\}$. Then

$$H(\nu) \;=\; \sum_{y \in D_\nu} \nu_y \left(1 - \nu_y\right).$$

Proof

- Let us construct a counter-example on $\mathbb{Y} = \mathbb{R}$.
- Consider the integrated variance functional defined by

$$H(\nu) \;=\; \int_{\mathbb{R}} \mathsf{var}_\nu \left(\mathbb{1}_{Y=s} \right) \, \mu(\mathrm{d}s),$$

where μ denotes the counting measure on $\mathbb{R}$.

- Remark: Let $\nu_y = \nu\left(\{y\}\right)$. Let $D_\nu = \{y \in \mathbb{R} \mid \nu_y > 0\}$. Then

$$H(\nu) \;=\; \sum_{y \in D_\nu} \nu_y \left(1 - \nu_y\right).$$

- H is nonnegative, measurable, concave, and vanishes on Dirac measures. But we will show that it violates Jensen's inequality.

Proof (cont'd)

- Consider the random measure $\boldsymbol{\nu} = \frac{1}{2}\left(\delta_{-X} + \delta_X\right)$, where $X \sim \mathrm{Exp}(1)$.

Proof (cont'd)

- Consider the random measure $\boldsymbol{\nu} = \frac{1}{2}\left(\delta_{-X} + \delta_X\right)$, where $X \sim \mathrm{Exp}(1)$.
- $H\left(\boldsymbol{\nu}\right) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$ almost surely, and therefore

$$\mathsf{E}\left(H(\boldsymbol{\nu})\right) = \frac{1}{2}.$$

Proof (cont'd)

- Consider the random measure $\boldsymbol{\nu} = \frac{1}{2}\left(\delta_{-X} + \delta_X\right)$, where $X \sim \mathrm{Exp}(1)$.
- $H(\boldsymbol{\nu}) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$ almost surely, and therefore

$$\mathsf{E}\left(H(\boldsymbol{\nu})\right) = \frac{1}{2}.$$

- $\overline{\boldsymbol{\nu}}$ is a Laplace distribution, which has no atom, therefore

$$H\left(\overline{\boldsymbol{\nu}}\right) = 0.$$

Proof (cont'd)

- Consider the random measure $\boldsymbol{\nu} = \frac{1}{2}\left(\delta_{-X} + \delta_X\right)$, where $X \sim \mathrm{Exp}(1)$.
- $H(\boldsymbol{\nu}) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$ almost surely, and therefore

$$\mathsf{E}\left(H(\boldsymbol{\nu})\right) \;=\; \frac{1}{2}.$$

- $\overline{\boldsymbol{\nu}}$ is a Laplace distribution, which has no atom, therefore

$$H(\overline{\boldsymbol{\nu}}) \;=\; 0.$$

- Thus Jensen's inequality does not hold:

$$H(\overline{\boldsymbol{\nu}}) \;\not\geq\; \mathsf{E}\left(H(\boldsymbol{\nu})\right). \qquad \square$$

A sufficient condition

Proposition

Let $\mathbb{D}$ be a non-empty set, and $L : \mathbb{Y} \times \mathbb{D} \to \overline{\mathbb{R}}_+$ be a function such that, for all $d \in \mathbb{D}$, $L(\cdot, d)$ is measurable.

A sufficient condition: Uncertainty as Bayes risk

Proposition

Let $\mathbb{D}$ be a non-empty set, and $L : \mathbb{Y} \times \mathbb{D} \to \overline{\mathbb{R}}_+$ be a function such that, for all $d \in \mathbb{D}$, $L(\cdot, d)$ is measurable. Then,

$$H(\nu) \;=\; \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y), \quad \nu \in \mathbb{P}, \qquad \rightarrow \text{ min ?}$$

defines a non-negative measurable probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ that satisfies Jensen's inequality.

(well known; see, e.g., Bect et al. [2019])

A sufficient condition: Uncertainty as Bayes risk

Proposition

Let $\mathbb{D}$ be a non-empty set, and $L : \mathbb{Y} \times \mathbb{D} \to \overline{\mathbb{R}}_+$ be a function such that, for all $d \in \mathbb{D}$, $L(\cdot, d)$ is measurable. Then,

$$H(\nu) \;=\; \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y), \quad \nu \in \mathbb{P}, \qquad \to \text{ min ?}$$

defines a non-negative measurable probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ that satisfies Jensen's inequality.

(well known; see, e.g., Bect et al. [2019])

Proof. For any random probability measure $\boldsymbol{\nu}$ and any $d \in \mathbb{D}$,

$$\mathsf{E}\left(\int_{\mathbb{Y}} L(y, d)\, \boldsymbol{\nu}(\mathrm{d}y) \right) \;=\; \int_{\mathbb{Y}} L(y, d)\, \overline{\boldsymbol{\nu}}(\mathrm{d}y).$$

The result then follows from the property $\inf_d \left(\mathsf{E}\left(\cdot\right) \right) \geq \mathsf{E}\left(\inf_d \left(\cdot\right) \right)$. $\square$

A sufficient condition: Uncertainty as Bayes risk

Proposition

Let $\mathbb{D}$ be a non-empty set, and $L : \mathbb{Y} \times \mathbb{D} \to \overline{\mathbb{R}}_+$ be a function such that, for all $d \in \mathbb{D}$, $L(\cdot, d)$ is measurable. Then,

$$H(\nu) \;=\; \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y), \quad \nu \in \mathbb{P}, \qquad \rightarrow \text{ min ?}$$

defines a non-negative measurable probability functional $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ that satisfies Jensen's inequality.

(well known; see, e.g., Bect et al. [2019])

Proof. For any random probability measure $\boldsymbol{\nu}$ and any $d \in \mathbb{D}$,

$$\mathsf{E}\left(\int_{\mathbb{Y}} L(y, d)\, \boldsymbol{\nu}(\mathrm{d}y) \right) \;=\; \int_{\mathbb{Y}} L(y, d)\, \overline{\boldsymbol{\nu}}(\mathrm{d}y).$$

The result then follows from the property $\inf_d \left(\mathsf{E}\left(\cdot\right)\right) \geq \mathsf{E}\left(\inf_d \left(\cdot\right)\right)$. $\square$

Connection with . . .

When the infimum is attained, we obtain an interesting class of functionals.

Proposition

A probability functional H admits the previous representation, in which the infimum over $\mathbb{D}$ is attained for every $\nu \in \mathbb{P}$:

$$H(\nu) \;=\; \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$$

Connection with proper scoring rules

When the infimum is attained, we obtain an interesting class of functionals.

Proposition

A probability functional H admits the previous representation, in which the infimum over $\mathbb{D}$ is attained for every $\nu \in \mathbb{P}$:

$$H(\nu) \;=\; \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$$

if and only if it is the generalized entropy function associated with a proper, non-negative, negatively oriented scoring rule $S : \mathbb{Y} \times \mathbb{P} \to \overline{\mathbb{R}}_+$:

$$H(\nu) \;=\; \min_{\nu' \in \mathbb{P}} \int S\left(y, \nu'\right)\; \nu(\mathrm{d}y).$$

Grünwald and Dawid [2004]

Connection with proper scoring rules

When the infimum is attained, we obtain an interesting class of functionals.

Proposition

A probability functional H admits the previous representation, in which the infimum over $\mathbb{D}$ is attained for every $\nu \in \mathbb{P}$:

$$H(\nu) \;=\; \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$$

if and only if it is the generalized entropy function associated with a proper, non-negative, negatively oriented scoring rule $S : \mathbb{Y} \times \mathbb{P} \to \overline{\mathbb{R}}_+$:

$$H(\nu) \;=\; \min_{\nu' \in \mathbb{P}} \int S\left(y, \nu'\right)\; \nu(\mathrm{d}y).$$

Grünwald and Dawid [2004]

Proof. Take $S(y, \nu) = L(y, d^*(\nu))$ with $d^*(\nu) \in \mathrm{argmin}_d \int L(y, d)\, \nu(\mathrm{d}y)$.

Example: Gini's mean difference

- Recall Gini's mean difference on $\mathbb{Y} = \mathbb{R}$:

$$H(\nu) \;=\; \mathsf{E}_\nu\left(|Y - Y'|\right), \qquad Y, Y' \overset{\text{iid}}{\sim} \nu.$$

This is the regular QUF with CND kernel $\rho(y, y') = |y - y'|$.

Example: Gini's mean difference

- Recall Gini's mean difference on $\mathbb{Y} = \mathbb{R}$:

$$H(\nu) \;=\; \mathsf{E}_\nu\left(|Y - Y'|\right), \qquad Y, Y' \overset{\text{iid}}{\sim} \nu.$$

This is the regular QUF with CND kernel $\rho(y, y') = |y - y'|$.

- Representation as a "generalized variance": if $H(\nu) < +\infty$,

$$H(\nu) \;=\; \min_{t \in \mathbb{H}} \mathsf{E}_\nu\left(\|\psi(Y) - t\|^2_{\mathbb{H}},\right)$$

where $\mathbb{H}$ is a Hilbert space and $\psi : \mathbb{Y} \to \mathbb{H}$ is weakly measurable.

Example: Gini's mean difference

- Recall Gini's mean difference on $\mathbb{Y} = \mathbb{R}$:

$$H(\nu) \;=\; \mathsf{E}_\nu\left(|Y - Y'|\right), \qquad Y, Y' \overset{\text{iid}}{\sim} \nu.$$

This is the regular QUF with CND kernel $\rho(y, y') = |y - y'|$.

- Representation as a "generalized variance": if $H(\nu) < +\infty$,

$$H(\nu) \;=\; \min_{t \in \mathbb{H}} \mathsf{E}_\nu\left(\|\psi(Y) - t\|^2_{\mathbb{H}},\right) \qquad \rightarrow \; \text{argmin} = \int \psi \mathrm{d}\nu$$

where $\mathbb{H}$ is a Hilbert space and $\psi : \mathbb{Y} \to \mathbb{H}$ is weakly measurable.

Example: Gini's mean difference

- Recall Gini's mean difference on $\mathbb{Y} = \mathbb{R}$:

$$H(\nu) \;=\; \mathsf{E}_\nu\left(|Y - Y'|\right), \qquad Y, Y' \overset{\text{iid}}{\sim} \nu.$$

This is the regular QUF with CND kernel $\rho(y, y') = |y - y'|$.

- Representation as a "generalized variance": if $H(\nu) < +\infty$,

$$H(\nu) \;=\; \min_{t \in \mathbb{H}} \mathsf{E}_\nu\left(\|\psi(Y) - t\|_{\mathbb{H}}^2,\right) \qquad \rightarrow \; \text{argmin} = \int \psi \mathrm{d}\nu$$

where $\mathbb{H}$ is a Hilbert space and $\psi : \mathbb{Y} \to \mathbb{H}$ is weakly measurable.

- Alternative representation using a proper scoring rule:

$$H(\nu) \;=\; \min_{\nu' \in \mathbb{P}} \mathsf{E}_\nu\left(S(Y, \nu')\right)$$

where $S(y, \nu) = 2\,\mathrm{CRPS}(y, \nu) = 2 \int \left(F_\nu(y') - \mathbb{1}_{y' \le y}\right)^2 \mathrm{d}y'$.

Example: Gini's mean difference

- Recall Gini's mean difference on $\mathbb{Y} = \mathbb{R}$:

$$H(\nu) \;=\; \mathsf{E}_\nu\left(|Y - Y'|\right), \qquad Y, Y' \overset{\text{iid}}{\sim} \nu.$$

  This is the regular QUF with CND kernel $\rho(y, y') = |y - y'|$.

- Representation as a "generalized variance": if $H(\nu) < +\infty$,

$$H(\nu) \;=\; \min_{t \in \mathbb{H}} \mathsf{E}_\nu\left(\|\psi(Y) - t\|_{\mathbb{H}}^2,\right) \qquad \rightarrow \; \text{argmin} = \int \psi \mathrm{d}\nu$$

  where $\mathbb{H}$ is a Hilbert space and $\psi : \mathbb{Y} \to \mathbb{H}$ is weakly measurable.

- Alternative representation using a proper scoring rule:

$$H(\nu) \;=\; \min_{\nu' \in \mathbb{P}} \mathsf{E}_\nu\left(S(Y, \nu')\right) \qquad \rightarrow \; \text{argmin} = \nu$$

  where $S(y, \nu) = 2\,\mathrm{CRPS}(y, \nu) = 2 \int \left(F_\nu(y') - \mathbb{1}_{y' \leq y}\right)^2 \mathrm{d}y'$.

One problem (almost) solved, two more to go...

Let $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ be measurable and vanish on the diagonal.

H is concave

$$\Uparrow$$

H is decreasing on average ($\Leftrightarrow$ satisfies Jensen's inequality)

$$\Uparrow$$

$H(\nu) = \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

$$\Uparrow$$

$H(\nu) = \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

One problem (almost) solved, two more to go...

Let $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ be measurable and vanish on the diagonal.

H is concave

$\Uparrow$ $\quad \not\Downarrow$ if $\mathbb{Y}$ is "rich enough"

H is decreasing on average ($\Leftrightarrow$ satisfies Jensen's inequality)

$\Uparrow$

$H(\nu) = \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

$\Uparrow$

$H(\nu) = \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

One problem (almost) solved, two more to go...

Let $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ be measurable and vanish on the diagonal.

H is concave

$\Uparrow \quad \not\Downarrow$ if $\mathbb{Y}$ is "rich enough"

H is decreasing on average ($\Leftrightarrow$ satisfies Jensen's inequality)

$\Uparrow \quad \Downarrow ?$ ➙ open problem

$H(\nu) = \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

$\Uparrow$

$H(\nu) = \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

One problem (almost) solved, two more to go...

Let $H : \mathbb{P} \to \overline{\mathbb{R}}_+$ be measurable and vanish on the diagonal.

H is concave

$\Uparrow$ $\not\Downarrow$ if $\mathbb{Y}$ is "rich enough"

H is decreasing on average ($\Leftrightarrow$ satisfies Jensen's inequality)

$\Uparrow$ $\Downarrow$? ➙ open problem

$H(\nu) = \inf_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

$\Uparrow$ $\Downarrow$? ➙ another open problem

$H(\nu) = \min_{d \in \mathbb{D}} \int L(y, d)\, \nu(\mathrm{d}y)$ for some $\mathbb{D}$, L

TTBOOK all the existing examples of uncertainty functionals are in the last class!

References I

J. Bect, F. Bachoc, and D. Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli*, 25(4A):2883–2919, 2019.

A. Ben-Tal and M. Teboulle. Rate distortion theory with generalized information measures via convex programming duality. *IEEE Transactions on Information Theory*, 32(5):630–641, 1986.

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

J. Burbea and C. R. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28(6):489–495, 1982.

M. H. DeGroot. Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33(2):404–419, 1962.

T. Fissler and S. M. Pesenti. Sensitivity measures based on scoring functions. *European Journal of Operational Research*, 307:1408–1423, 2023.

J.-C. Fort, T. Klein, and N. Rachdi. New sensitivity analysis subordinated to a contrast. *Communications in Statistics - Theory and Methods*, 45(15):4349–4364, 2016.

P. D. Grünwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *The Annals of Statistics*, 32(4):1367 – 1433, 2004.

References II

M. Hainy, W. G. Müller, and H. P. Wynn. Learning functions and approximate Bayesian computation design: ABCD. *Entropy*, 16(8):4353–4374, 2014.

J. Havrda and F. Charvát. Quantification method of classification processes. concept of structural a-entropy. *Kybernetika*, 03(1):30–35, 1967.

T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

K. S. Lau. Characterization of Rao's quadratic entropies. *The Indian Journal of Statistics*, 47 (3):295–309, 1985.

D. V. Lindley. On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

M. D. Perlman. Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *Journal of Multivariate Analysis*, 4(1):52–65, 1974.

C. R. Rao. Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology*, 21(1):24–43, 1982a.

C. R. Rao. Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 1–22, 1982b.

References III

C. R. Rao. Convexity properties of entropy functions and analysis of diversity. In *Inequalities in Statistics and Probability*, volume 5 of *IMS Lecture Notes-Monograph Series*, pages 68–77. Institute of Mathematical Statistics, 1984.

C. R. Rao. Quadratic entropy and analysis of diversity. *Sankhya A*, 72(1):70–80, 2010.

C. R. Rao and T. Nayak. Cross entropy, dissimilarity measures, and characterizations of quadratic entropy. *IEEE Transactions on Information Theory*, 31(5):589–593, 1985.

A. M. Scarfone. Entropic forms and related algebras. *Entropy*, 15(2):624–649, 2013.

E. H. Simpson. Measurement of diversity. *Nature*, 163(4148):688–688, 1949.

I. M. Sobol'. Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling & Computational Experiment*, 1:407–414, 1993.

G. J. Székely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of statistical planning and inference*, 143(8):1249–1272, 2013.

P. Tempesta. A theorem on the existence of trace-form generalized entropies. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2183):20150165, 2015.

References IV

C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of statistical physics*, 52(1):479–487, 1988.

S. Yitzhaki. Gini's mean difference: A superior measure of variability for non-normal distributions. *Metron*, 61(2):285–316, 2003.

Appendices

Conditionally negative definite kernel

Definition

A measurable symmetric function $\rho : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$,

Conditionally negative definite kernel

Definition

A measurable symmetric function $\rho : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$, is called conditionally negative definite (CND) if,

Conditionally negative definite kernel

Definition

A measurable symmetric function $\rho : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$, is called conditionally negative definite (CND) if, for all $m \geq 1$, all points $y_1, \dots, y_m \in \mathbb{Y}$, and all $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ with $\sum_{k=1}^{m} \lambda_k = 0$, it satisfies

$$\sum_{k,l}^{m} \lambda_k \lambda_l \, \rho(y_k, y_l) \ \leq \ 0.$$

Conditionally negative definite kernel

Definition

A measurable symmetric function $\rho : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_+$, is called conditionally negative definite (CND) if, for all $m \geq 1$, all points $y_1, \ldots, y_m \in \mathbb{Y}$, and all $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ with $\sum_{k=1}^{m} \lambda_k = 0$, it satisfies

$$\sum_{k,l}^{m} \lambda_k \lambda_l \, \rho(y_k, y_l) \; \leq \; 0.$$

Characterization of ρ

Any CND kernel ρ vanishing on the diagonal admits a Hilbertian representation: there exist a Hilbert space $\mathbb{H}$ and a map $\psi : \mathbb{Y} \to \mathbb{H}$ such that

$$\rho(y, y') \; = \; \|\psi(y) - \psi(y')\|_{\mathbb{H}}^2 , \qquad \forall y, y' \in \mathbb{Y}.$$

Conversely, kernels defined this way are CND and vanish on the diagonal.

Some examples of regular QUFs

$\mathbb{Y}$	**CND kernel** ρ	**Regular QUF** $H(\nu)$	**Name**
$\{0,1\}$	$\frac{1}{2}\mathbb{1}_{y\neq y'}$	$\nu_1\,(1-\nu_1)$	Bernoulli variance
$\mathbb{N}$	$\mathbb{1}_{y\neq y'}$	$1-\sum_{y\in\mathbb{N}}\nu_y^2$	Gini-Simpson index[(1)]
$\mathbb{R}$	$\frac{1}{2}\,(y-y')^2$	$\mathrm{var}_\nu\,(Y)$	variance
$\mathbb{R}$	$\lvert y-y'\rvert$	$\mathsf{E}_\nu\,(\lvert Y-Y'\rvert)$	Gini's mean diff.[(2)]
$\mathbb{R}$	$\lvert y-y'\rvert^\beta$ (†)	$\mathsf{E}_\nu\left(\lvert Y-Y'\rvert^\beta\right)$	Fract. Brownian var.[(3)]
$\mathbb{R}$	$1-e^{-(y-y')^2}$	$\mathsf{E}_\nu\left(1-e^{-(Y-Y')^2}\right)$	(unnamed ?)
$L^2(\mu)$	$\lVert y-y'\rVert^2$	$\int \mathrm{var}_\nu\,(Y(s))\;\mu(\mathrm{d}s)$	integrated variance

Notations: $\nu\in\mathbb{P}$; $\nu_y=\nu(\{y\})$ for all $y\in\mathbb{Y}$; (†) $0<\beta<2$; $Y, Y' \overset{\text{iid}}{\sim} \nu$; μ σ-finite. Refs: [(1)] Simpson [1949], [(2)] Yitzhaki [2003], [(3)] Székely and Rizzo [2013]