LiU 2014

Non-additive measures and integrals

Vicenç Torra

March, 2014

* IIIA-CSIC (joint work with Yasuo Narukawa and Michio Sugeno)

A short motivation

Topic. Non-additive measures

- A generalization of additive measures (probabilities)
- Non-additive measures also known as
  - fuzzy measures (Sugeno, 1974),
  - capacities (Choquet, 1954),
  - monotone games (Aumann and Shapley, 1974),
  - premeasures (Šipoš, 1979)

A short motivation

Why are these measures studied?

- Mathematical interest
  - Properties
    - ⋆ Equalities and inequalities (e.g. Cauchy-Schwarz type inequalities)
    - ⋆ Measures and distances (e.g. entropy/Hellinger)
  - Constructions
    - ⋆ Integrals with respect to these measures (e.g. Choquet integral)

A short motivation

Why are these measures studied?

- Applications
  - Some problems that cannot be solved with additive measures can be solved with non-additive measures
    - ⋆ Decision making
    - ⋆ Subjective evaluation
    - ⋆ Data fusion (e.g. computer vision)

→ a common theme:
to take into account interactions

→ a common advantage:
more expressive power than with the additive models

Outline

1. Introduction
2. Some definitions
3. Distances (new definitions)
4. Properties
5. Applications
6. Summary

Some definitions

Definitions: measures

Additive measures.

- $(X, \mathcal{A})$ a measurable space; then, a set function μ is an additive measure if it satisfies
  (i) $\mu(A) \geq 0$ for all $A \in \mathcal{A}$,
  (ii) $\mu(X) \leq \infty$
  (iii) for every countable sequence A_i $(i \geq 1)$ of $\mathcal{A}$ that is pairwise disjoint (i.e,. $A_i \cap A_j = \emptyset$ when $i \neq j$)

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

Definitions: measures

Additive measures.

- $(X, \mathcal{A})$ a measurable space; then, a set function μ is an additive measure if it satisfies
  (i) $\mu(A) \geq 0$ for all $A \in \mathcal{A}$,
  (ii) $\mu(X) \leq \infty$
  (iii) for every countable sequence A_i $(i \geq 1)$ of $\mathcal{A}$ that is pairwise disjoint (i.e,. $A_i \cap A_j = \emptyset$ when $i \neq j$)

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$$

Finite case: $\mu(A \cup B) = \mu(A) + \mu(B)$ for disjoint A, B

Definitions: measures

Additive measures.

Example:

- Unique measure λ s.t. $\lambda([a,b]) = b - a$ for every finite interval $[a,b]$
  $\rightarrow$ the Lebesgue measure

Definitions: measures

Additive measures.

Example:

- Unique measure λ s.t. $\lambda([a,b]) = b - a$ for every finite interval $[a,b]$
  $\rightarrow$ the Lebesgue measure
- Probability, if $\mu(X) = 1$.

Definitions: measures

Non-additive measures.

- $(X, \mathcal{A})$ a measurable space, a non-additive (fuzzy) measure μ on $(X, \mathcal{A})$ is a set function $\mu : \mathcal{A} \to [0, 1]$ satisfying the following axioms:

(i) $\mu(\emptyset) = 0$, $\mu(X) = 1$ (boundary conditions)

(ii) $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ (monotonicity)

Definitions: measures

Non-additive measures. Examples. Distorted Lebesgue

- $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function such that $m(0) = 0$; λ be the Lebesgue measure.
  The following set function μ_m is a non-additive (fuzzy) measure:

$$\mu_m(A) = m(\lambda(A)) \tag{1}$$

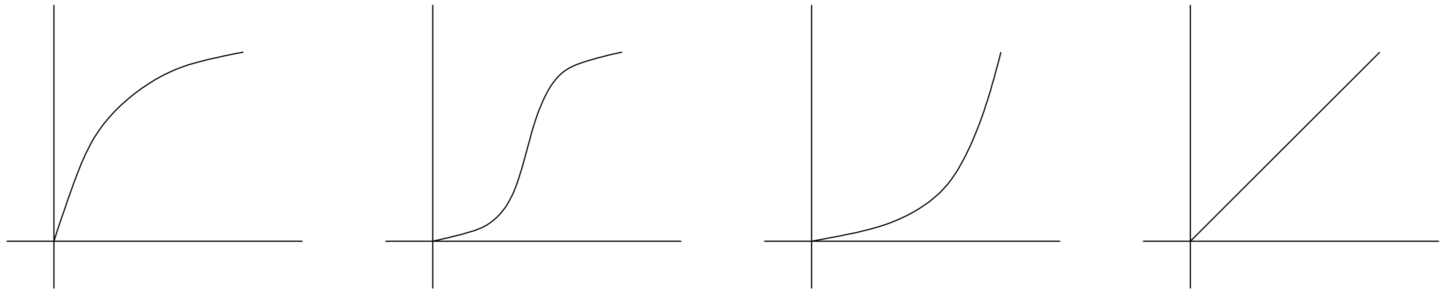
Definitions: measures

Non-additive measures. Examples. Distorted Lebesgue

- $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function such that $m(0) = 0$; λ be the Lebesgue measure.
  The following set function μ_m is a non-additive (fuzzy) measure:

$$\mu_m(A) = m(\lambda(A)) \tag{1}$$

- If $m(x) = x^2$, then $\mu_m(A) = (\lambda(A))^2$
- If $m(x) = x^p$, then $\mu_m(A) = (\lambda(A))^p$



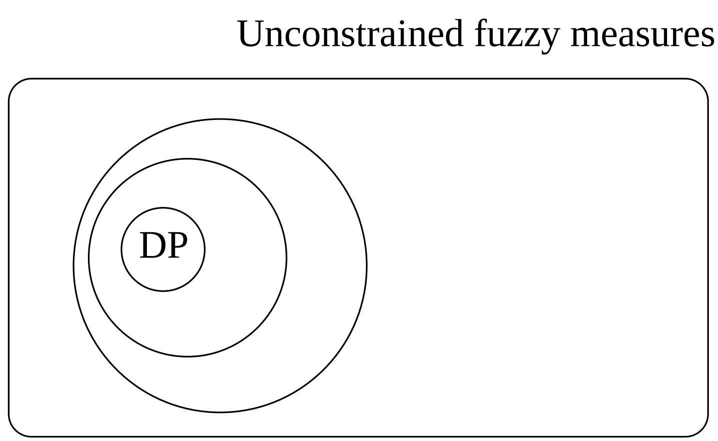
(a) (b) (c) (d)

Definitions: measures

Non-additive measures. Examples. Distorted probabilities

- $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function such that $m(0) = 0$; P be a probability.
  The following set function μ_m is a non-additive (fuzzy) measure:

$$\mu_{m,P}(A) = m(P(A)) \tag{2}$$



Definitions: integrals

Choquet integral (Choquet, 1954):

- μ a non-additive measure, g a measurable function. The Choquet integral of g w.r.t. μ, where $\mu_g(r) := \mu(\{x|g(x) > r\})$:

$$(C)\int g d\mu := \int_0^\infty \mu_g(r) dr. \tag{3}$$

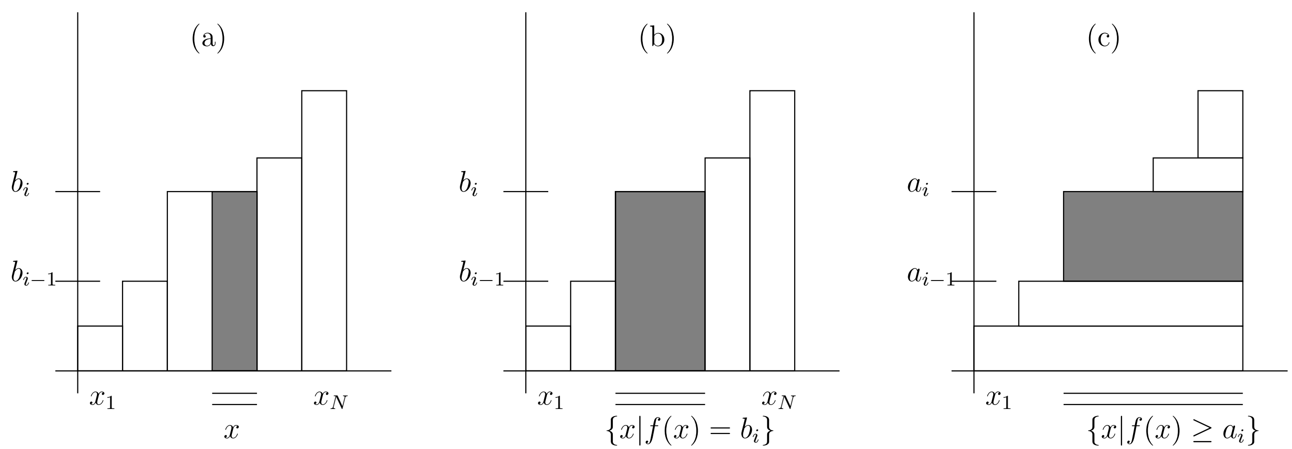
Definitions: integrals

Choquet integral (Choquet, 1954):

- μ a non-additive measure, g a measurable function. The Choquet integral of g w.r.t. μ, where $\mu_g(r) := \mu(\{x|g(x) > r\})$:

$$(C)\int g d\mu := \int_0^\infty \mu_g(r)dr. \tag{3}$$

- When the measure is additive, this is the Lebesgue integral



Definitions: integrals

Choquet integral. Discrete version

- μ a non-additive measure, f a measurable function. The Choquet integral of f w.r.t. μ,

$$(C)\int f d\mu = \sum_{i=1}^{N} [f(x_{s(i)}) - f(x_{s(i-1)})]\mu(A_{s(i)}),$$

  where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq \cdots \leq f(x_{s(N)}) \leq 1$, and where $f(x_{s(0)}) = 0$ and $A_{s(i)} = \{x_{s(i)}, \ldots, x_{s(N)}\}$.

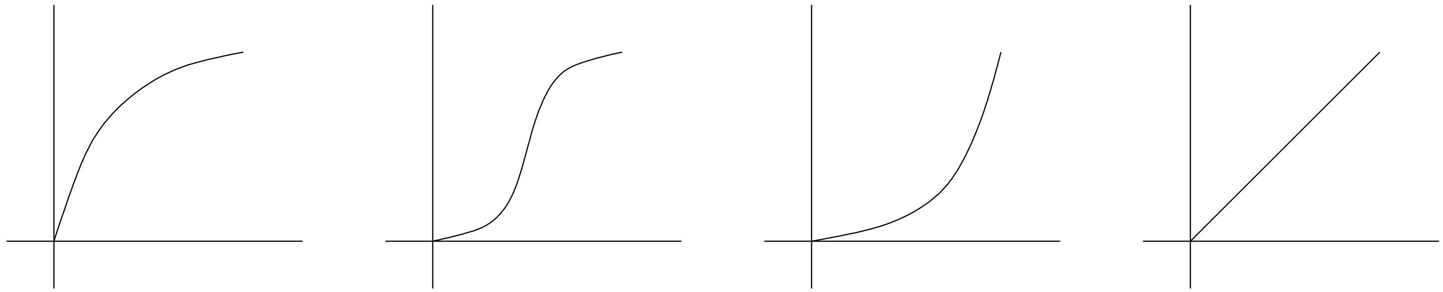
Definitions: measures

Choquet integral: Example:

- $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function s.t. $m(0) = 0$, $m(1) = 1$; P a probability distribution. μ_m, a non-additive (fuzzy) measure:

$$\mu_m(A) = m(P(A)) \quad (4)$$

- $CI_{\mu_m}(f)$
  $(a) \to$ *max*, $(b) \to$ *median*, $(c) \to$ *min*, $(d) \to$ *mean (expectation)*



(a) (b) (c) (d)

Definitions: properties

Properties: (X be a reference set)

- **Comonotonicity.** f and g are comonotonic if, for all $x_i, x_j \in X$,
$$f(x_i) < f(x_j) \text{ imply that } g(x_i) \leq g(x_j)$$
- $\mathcal{I}$ **is comonotonic monotone** if and only if, for comonotonic f and g,
$$f \leq g \text{ imply that } \mathcal{I}(f) \leq \mathcal{I}(g)$$
- $\mathcal{I}$ **is comonotonic additive** if and only if, for comonotonic f and g,
$$\mathcal{I}(f+g) = \mathcal{I}(f) + \mathcal{I}(g)$$

Definitions: properties

Choquet integral. Characterization

- Theorem (Schmeidler, 1986; Narukawa and Murofushi, 2003). Let $\mathcal{I} : [0,1]^n \to \mathbb{R}_+$ be a functional with the following properties
  - $\mathcal{I}$ is comonotonic monotone
  - $\mathcal{I}$ is comonotonic additive
  - $\mathcal{I}(1, \ldots, 1) = 1$

  Then, there exists a non-additive measure μ such that $\mathcal{I}(f)$ is the Choquet integral of f with respect to μ.

It is also true that a Choquet integral satisfies the conditions above.

Definitions: properties

Choquet integral. Properties

- Proposition 1. If μ is submodular, then

$$(C)\int (f+g)d\mu \leq (C)\int f d\mu + (C)\int g d\mu.$$

- Proposition 2. If μ is supermodular, then

$$(C)\int (f+g)d\mu \geq (C)\int f d\mu + (C)\int g d\mu.$$

where

- submodular $\mu(A)+\mu(B) \geq \mu(A\cup B)+\mu(A\cap B)$

  When adding an element, the smaller the set, the larger the increase
- supermodular $\mu(A)+\mu(B) \leq \mu(A\cup B)+\mu(A\cap B)$

Definitions: properties

Choquet integral. Properties

- Cauchy-Schwarz inequality: If μ is a submodular non-additive measure; then

$$((C)\int fgd\mu)^2 \leq (C)\int f^2 d\mu (C)\int g^2 d\mu.$$

- Another inequality: If μ is a submodular non-additive measure; then

$$((C)\int (f+g)^2 d\mu)^{\frac{1}{2}} \leq ((C)\int f^2 d)^{\frac{1}{2}} + ((C)\int g^2 d\mu)^{\frac{1}{2}}$$

Definitions: Radon-Nikodym derivative

Radon-Nikodym derivative: (additive measures)

- Concept: ν absolutely continuous w.r.t. μ (if $\mu(A) = 0$ then $\nu(A) = 0$)
- **Theorem.** μ and ν two additive measures on $(\Omega, \mathcal{F})$ and μ be σ-finite. If $\nu << \mu$, then there exists a nonnegative measurable function f on Ω such that

$$\nu(A) = \int_A f d\mu$$

Definitions: Radon-Nikodym derivative

Radon-Nikodym derivative: (additive measures)

- Concept: ν absolutely continuous w.r.t. μ (if $\mu(A) = 0$ then $\nu(A) = 0$)
- **Theorem.** μ and ν two additive measures on $(\Omega, \mathcal{F})$ and μ be σ-finite. If $\nu << \mu$, then there exists a nonnegative measurable function f on Ω such that

$$\nu(A) = \int_A f d\mu$$

$\rightarrow$ this permits to define the Radon-Nikodym derivative:

Definitions: Radon-Nikodym derivative

Radon-Nikodym derivative: (additive measures)

- Concept: ν absolutely continuous w.r.t. μ (if $\mu(A) = 0$ then $\nu(A) = 0$)
- **Theorem.** μ and ν two additive measures on $(\Omega, \mathcal{F})$ and μ be σ-finite. If $\nu << \mu$, then there exists a nonnegative measurable function f on Ω such that

$$\nu(A) = \int_A f d\mu$$

$\rightarrow$ this permits to define the Radon-Nikodym derivative:

$\rightarrow$ The function f is called the Radon-Nikodym derivative of ν w.r.t. μ, denoted

$$f = \frac{d\nu}{d\mu}.$$

Definitions: Radon-Nikodym derivative

Radon-Nikodym derivative: (additive measures)

- Concept: ν absolutely continuous w.r.t. μ (if $\mu(A) = 0$ then $\nu(A) = 0$)
- **Theorem.** μ and ν two additive measures on $(\Omega, \mathcal{F})$ and μ be σ-finite. If $\nu << \mu$, then there exists a nonnegative measurable function f on Ω such that

$$\nu(A) = \int_A f d\mu$$

$\rightarrow$ this permits to define the Radon-Nikodym derivative:
$\rightarrow$ The function f is called the Radon-Nikodym derivative of ν w.r.t. μ, denoted

$$f = \frac{d\nu}{d\mu}.$$

- f may not be unique, but if f_0 and f_1 are both Radon-Nikodym derivatives of ν, then $f_0 = f_1$ almost everywhere μ

Derivatives w.r.t. non-additive measures

Derivative (Choquet integral): (non-additive measures)

- $(\Omega, \mathcal{F})$ a measurable space, $\nu, \mu : \mathcal{F} \to \mathbb{R}^+$ non-additive measures.
  $\to$ ν is a Choquet integral of μ if there exists a measurable function $g : \Omega \to \mathbb{R}^+$ s.t. for all $A \in \mathcal{F}$

$$\nu(A) = (C) \int_A g d\mu \tag{5}$$

Derivatives w.r.t. non-additive measures

Derivative (Choquet integral): (non-additive measures)

- $(\Omega, \mathcal{F})$ a measurable space, $\nu, \mu : \mathcal{F} \to \mathbb{R}^+$ non-additive measures.
  $\to$ ν is a Choquet integral of μ if there exists a measurable function $g : \Omega \to \mathbb{R}^+$ s.t. for all $A \in \mathcal{F}$

$$\nu(A) = (C) \int_A g d\mu \tag{5}$$

- μ, ν two non-additive measures. If μ is a Choquet integral of ν, and g is a function such that Equation 5 is satisfied, then

$$d\nu/d\mu = g,$$

$\to$ g is a derivative of ν with respect to μ.
$\to$ Graf and Sugeno studied conditions of when this derivative exists.

Derivatives w.r.t. non-additive measures

Derivative (Choquet integral): (Proposition 4 in (Sugeno, 2013))

- Let $f(t)$ be a continuous and increasing function with $f(0) = 0$, let μ_m be a distorted Lebesgue measure, then there exists an increasing (non-decreasing) function g so that $f(t) = (C)\int_{[0,t]} g(\tau) d\mu_m$ and the following holds:

$$G(s) = F(s)/sM(s) \quad (6)$$

$$g(t) = L^{-1}[F(s)/sM(s)]. \quad (7)$$

Here, $F(s)$ is the Laplace transformation of f, M the Laplace transformation of m, and G the Laplace transformation of g.

Derivatives w.r.t. non-additive measures

Computation:

- It is possible to compute the Radon-Nikodym derivative (for some examples)
- Computations use the Laplace transformation

Derivatives w.r.t. non-additive measures

Computation (Example): Applying Proposition 4 (Sugeno, 2013), we have

$$L[\frac{d\nu^p}{d\mu_m}] = \frac{N^p(s)}{sM(s)} = \frac{\Gamma(p+1)}{2s^{p-1}}.$$

Then using the inverse Laplace transform on this expression we obtain:

$$\begin{aligned} \frac{d\nu^p}{d\mu_m} &= L^{-1}[\frac{\Gamma(p+1)}{2s^{p-1}}] = \frac{\Gamma(p+1)}{2\Gamma(p-1)} t^{p-2} \\ &= \frac{p(p-1)}{2} t^{p-2}. \end{aligned}$$

f-divergence for non-additive measures

f-Divergence

Given: P, Q two probabilities a.c. w.r.t. a prob. ν.

- f-divergence between P and Q w.r.t. ν

$$D_{f,\nu}(P,Q) = \int \frac{dQ}{d\nu} f\left(\frac{dP/d\nu}{dQ/d\nu}\right) d\nu$$

f-Divergence and distances

Examples of f-divergence between P and Q w.r.t. ν

$$D_{f,\nu}(P,Q) = \int \frac{dQ}{d\nu} f\left(\frac{dP/d\nu}{dQ/d\nu}\right) d\nu$$

Some particular distances

- Hellinger distance when $f(x) = (1-\sqrt{x})^2$,

$$H(P,Q) = \sqrt{\frac{1}{2}\int \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}}\right)^2 d\nu}$$

  Here $dP/d\nu$ and $dQ/d\nu$ are the Radon-Nikodym derivatives
- Variation distance when $f(x) = |x-1|$,

$$\delta(P,Q) = \frac{1}{2}\int \left|\frac{dP}{d\nu} - \frac{dP}{d\nu}\right| d\nu$$

- Kullback-Leibler, Rényi distance, χ^2-distance

f-Divergence: non-additive measures

Definition:

- μ_1, μ_2 two non-additive measures that are Choquet integrals of ν. The f-divergence between μ_1 and μ_2 with respect to ν is defined as

$$D_{f,\nu}(\mu_1, \mu_2) = (C) \int \frac{d\mu_2}{d\nu} f\left(\frac{d\mu_1/d\nu}{d\mu_2/d\nu}\right) d\nu$$

  Here $d\mu_1/d\nu$ and $d\mu_2/d\nu$ are the derivatives of μ_1 and μ_2.

f-Divergence and Hellinger distance: non-additive measures

Definition:

- μ_1, μ_2 two non-additive measures that are Choquet integrals of ν. The Hellinger distance between μ_1 and μ_2 with respect to ν is defined as

$$H_\nu(\mu_1, \mu_2) = \sqrt{\frac{1}{2}(C)\int \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}}\right)^2 d\nu}$$

  Here $d\mu_1/d\nu$ and $d\mu_2/d\nu$ are the derivatives of μ_1 and μ_2.

f-Divergence and variation distance: non-additive measures

Definition:

- μ_1, μ_2 two non-additive measures that are Choquet integrals of ν. The Variation distance between μ_1 and μ_2 with respect to ν is defined as

$$\delta_\nu(\mu_1, \mu_2) = \frac{1}{2}(C)\int \left| \frac{d\mu_1}{d\nu} - \frac{d\mu_2}{d\nu} \right| d\nu$$

  Here $d\mu_1/d\nu$ and $d\mu_2/d\nu$ are the derivatives of μ_1 and μ_2.

Properties

Distances: properties

Properties:

- Proper generalization?

Distances: properties

Properties:

- Proper generalization?
- Yes: Let ν, μ_1, μ_2 be three additive measures such that μ_1 and μ_2 are absolutely continuous with respect to ν. Then, $D_{f,\nu}(\mu_1, \mu_2)$ is the standard f-divergence.

Distances: properties

Properties:

- Proper generalization?
- Yes: Let ν, μ_1, μ_2 be three additive measures such that μ_1 and μ_2 are absolutely continuous with respect to ν. Then, $D_{f,\nu}(\mu_1, \mu_2)$ is the standard f-divergence.
- Also, $H_\nu(\mu_1, \mu_2)$ and $\delta_\nu(\mu_1, \mu_2)$ are the Hellinger distance and the variation distance

Distances: properties

Properties:

- Proper generalization?
- Yes: Let ν, μ_1, μ_2 be three additive measures such that μ_1 and μ_2 are absolutely continuous with respect to ν. Then, $D_{f,\nu}(\mu_1, \mu_2)$ is the standard f-divergence.
- Also, $H_\nu(\mu_1, \mu_2)$ and $\delta_\nu(\mu_1, \mu_2)$ are the Hellinger distance and the variation distance
- $D_{f,\nu}(\mu_1, \mu_2)$ with appropriate f (i.e., $f(x) = (1 - \sqrt{x})^2$ and $f(x) = |x - 1|$) correspond to Hellinger and variation distance. I.e.,

Distances: properties

Properties:

- Proper generalization?
- Yes: Let ν, μ_1, μ_2 be three additive measures such that μ_1 and μ_2 are absolutely continuous with respect to ν. Then, $D_{f,\nu}(\mu_1, \mu_2)$ is the standard f-divergence.
- Also, $H_\nu(\mu_1, \mu_2)$ and $\delta_\nu(\mu_1, \mu_2)$ are the Hellinger distance and the variation distance
- $D_{f,\nu}(\mu_1, \mu_2)$ with appropriate f (i.e., $f(x) = (1 - \sqrt{x})^2$ and $f(x) = |x - 1|$) correspond to Hellinger and variation distance. I.e.,

$$\sqrt{\frac{1}{2} D_{f,\nu}(\mu_1, \mu_2)} = H_\nu(\mu_1, \mu_2).$$

$$\frac{1}{2} D_{f,\nu}(\mu_1, \mu_2) = \delta_\nu(\mu_1, \mu_2).$$

Distances: properties

Properties:

- Distance?

Distances: properties

Properties:

- Distance?
  - f-divergence is not a distance for additive measures
    (it is not symmetric, it does not no satisfy triangle inequality)

Distances: properties

Properties:

- Distance?
  - f-divergence is not a distance for additive measures (it is not symmetric, it does not no satisfy triangle inequality)
  - Hellinger distance, variation distance are distances. (satisfy positiveness, symmetry, and triangular inequality)

Distances: properties

Properties:

- Distance?
  - f-divergence is not a distance for additive measures (it is not symmetric, it does not no satisfy triangle inequality)
  - Hellinger distance, variation distance are distances. (satisfy positiveness, symmetry, and triangular inequality)
- So, we only consider distance for Hellinger and variation distance

Distances: properties

Properties:

- Distance?
  - Positiveness: $D_{f,\nu}(\mu_1, \mu_2) = 0$ if $\mu_1 = \mu_2$.
    So, Hellinger and variation distance are positive

Distances: properties

Properties:

- Distance?
  - Positiveness: $D_{f,\nu}(\mu_1, \mu_2) = 0$ if $\mu_1 = \mu_2$.
    So, Hellinger and variation distance are positive
  - Symmetry: Hellinger and variation symmetric by definition

Distances: properties

Properties:

- Distance?
  - Positiveness: $D_{f,\nu}(\mu_1, \mu_2) = 0$ if $\mu_1 = \mu_2$.
    So, Hellinger and variation distance are positive
  - Symmetry: Hellinger and variation symmetric by definition
  - Triangular inequality:
    - ⋆ If ν is submodular, then we have

$$H_\nu(\mu_1, \mu_2) + H_\nu(\mu_2, \mu_3) \geq H_\nu(\mu_1, \mu_3).$$

Distances: properties

Properties:

- Distance?
  - Positiveness: $D_{f,\nu}(\mu_1, \mu_2) = 0$ if $\mu_1 = \mu_2$.
    So, Hellinger and variation distance are positive
  - Symmetry: Hellinger and variation symmetric by definition
  - Triangular inequality:
    - ⋆ If ν is submodular, then we have

$$H_\nu(\mu_1, \mu_2) + H_\nu(\mu_2, \mu_3) \geq H_\nu(\mu_1, \mu_3).$$

    - ⋆ Also, if ν is submodular, then we have

$$\delta_\nu(\mu_1, \mu_2) + \delta_\nu(\mu_2, \mu_3) \geq \delta_\nu(\mu_1, \mu_3).$$

Distances: properties

Properties:

- Triangular inequality. Proof
  - Proof of triangular inequality for Hellinger distance comes from (seen above)

$$((C)\int(f+g)^2 d\mu)^{\frac{1}{2}} \leq ((C)\int f^2 d)^{\frac{1}{2}} + ((C)\int g^2 d\mu)^{\frac{1}{2}}$$

  - Proof of triangular inequality for variation distance comes from (seen above)

$$(C)\int(f+g)d\mu \leq (C)\int f d\mu + (C)\int g d\mu.$$

Distances: properties

Properties:

- Triangular inequality Hellinger distance. Proof

$$
\begin{aligned}
H_\nu(\mu_1,\mu_2) + H_\nu(\mu_2,\mu_3) &= \{\frac{1}{2}(C)\int \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}}\right)^2 d\nu\}^{1/2} + \{\frac{1}{2}(C)\int \left(\sqrt{\frac{d\mu_2}{d\nu}} - \sqrt{\frac{d\mu_3}{d\nu}}\right)^2 d\nu\}^{1/2} \\
&\geq \{\frac{1}{2}(C)\int \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}}\right)^2 + \left(\sqrt{\frac{d\mu_2}{d\nu}} - \sqrt{\frac{d\mu_3}{d\nu}}\right)^2 d\nu\}^{1/2} \\
&= \{\frac{1}{2}(C)\int \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}}\right)^2 + \left(\sqrt{\frac{d\mu_3}{d\nu}} - \sqrt{\frac{d\mu_2}{d\nu}}\right)^2 d\nu\}^{1/2} \\
&\geq \{\frac{1}{2}(C)\int \left(\sqrt{\frac{d\mu_1}{d\nu}} - \sqrt{\frac{d\mu_3}{d\nu}}\right)^2 d\nu\}^{1/2} \\
&= H_\nu(\mu_1,\mu_3)
\end{aligned}
$$

Distances: properties

Properties:

- Triangular inequality variation distance. Proof

$$
\begin{aligned}
\delta_\nu(\mu_1,\mu_2)+\delta_\nu(\mu_2,\mu_3) &= \frac{1}{2}(C)\int\left|\frac{d\mu_1}{d\nu}-\frac{d\mu_2}{d\nu}\right|d\nu+\frac{1}{2}(C)\int\left|\frac{d\mu_2}{d\nu}-\frac{d\mu_3}{d\nu}\right|d\nu \\
&\geq \frac{1}{2}(C)\int\left(\left|\frac{d\mu_1}{d\nu}-\frac{d\mu_2}{d\nu}\right|+\left|\frac{d\mu_2}{d\nu}-\frac{d\mu_3}{d\nu}\right|\right)d\nu \\
&\geq \frac{1}{2}(C)\int\left|\frac{d\mu_1}{d\nu}-\frac{d\mu_3}{d\nu}\right|d\nu \\
&= \delta_\nu(\mu_1,\mu_3)
\end{aligned}
$$

Distances: properties

Properties:

- Distance?

Distances: properties

Properties:

- Distance?
  - If ν is submodular, Hellinger distance is a distance.
  - If ν is submodular, Variation distance is a distance.

Hellinger distance: properties

Example: Computation of a Hellinger distance between two distorted Lebesgue measures w.r.t. a third one

Hellinger distance: properties

Example: Computation of a Hellinger distance between two distorted Lebesgue measures w.r.t. a third one

- Measures:

Hellinger distance: properties

Example: Computation of a Hellinger distance between two distorted Lebesgue measures w.r.t. a third one

- Measures:
  - μ_m be the distorted Lebesgue measure with $m(t) = t^2$,
  - ν^p be the distorted Lebesgue measure with $n^p(t) = t^p$ (i.e., $\nu^p(A) = (\lambda(A))^p$ for $p \geq 2$, and $\nu^p([0, t]) = t^p$)

Hellinger distance: properties

Example: Computation of a Hellinger distance between two distorted Lebesgue measures w.r.t. a third one

- Measures:
  - μ_m be the distorted Lebesgue measure with $m(t) = t^2$,
  - ν^p be the distorted Lebesgue measure with $n^p(t) = t^p$ (i.e., $\nu^p(A) = (\lambda(A))^p$ for $p \geq 2$, and $\nu^p([0,t]) = t^p$)
- Computation: Hellinger distance between ν^2 and ν^p w.r.t. μ_m.

Hellinger distance: properties

Example: Computation of a Hellinger distance between two distorted Lebesgue measures w.r.t. a third one

- Measures:
  - μ_m be the distorted Lebesgue measure with $m(t) = t^2$,
  - ν^p be the distorted Lebesgue measure with $n^p(t) = t^p$ (i.e., $\nu^p(A) = (\lambda(A))^p$ for $p \geq 2$, and $\nu^p([0,t]) = t^p$)
- Computation: Hellinger distance between ν^2 and ν^p w.r.t. μ_m.

$$H_{\mu_m}(\nu^2, \nu^p) = \sqrt{\frac{1}{2}(C)\int_0^1 \left(\sqrt{\frac{d\nu^2}{\mu_m}} - \sqrt{\frac{d\nu^p}{\mu_m}}\right)^2 d\mu_m} \qquad (8)$$

Hellinger distance: properties

Example (II): Hellinger distance between ν^2 and ν^p w.r.t. μ_m where distortions are $n^p(t) = t^p$ and $m(t) = t^2$.

- Recall (from a previous example) that

$$\frac{d\nu^p}{d\mu_m} = \frac{p(p-1)}{2} t^{p-2}$$

Hellinger distance: properties

Example (II): Hellinger distance between ν^2 and ν^p w.r.t. μ_m where distortions are $n^p(t) = t^p$ and $m(t) = t^2$.

- Recall (from a previous example) that

$$\frac{d\nu^p}{d\mu_m} = \frac{p(p-1)}{2} t^{p-2}$$

- Computation (with more Choquet integral – and Laplace transforms):

$$H_{\mu_m}(\nu^2, \nu^p) = \sqrt{\frac{1}{2}(C)\int_0^1 \left(\sqrt{\frac{d\nu^2}{\mu_m}} - \sqrt{\frac{d\nu^p}{\mu_m}}\right)^2 d\mu_m}$$

$$= \sqrt{\frac{1}{2}(C)\int_0^1 \left(1 - \sqrt{\frac{p(p-1)}{2}t^{(p-2)/2}}\right)^2 d\mu_m} \quad (9)$$

$$= \sqrt{1 - \frac{4\sqrt{2p(p-1)}}{(p+2)p}} \quad (10)$$

Hellinger distance: properties

Example 2:

- $\mu_{m'}$ be the distorted Lebesgue measure with $m'(t) = t$.
- ν^p be the distorted Lebesgue measure with $n(t) = t^p$ (i.e., $\nu^p(A) = (\lambda(A))^p$ for $p \geq 2$, and $\nu^p([0, t]) = t^p$)
- Compute the Hellinger distance between ν^2 and ν^p w.r.t. $\mu_{m'}$. Only difference from Example 1 is $\mu_{m'}$

Hellinger distance: properties

Example 2: Hellinger distance between ν^2 and ν^p w.r.t. μ_m where distortions are $n^p(t) = t^p$ and $m(t) = t$.

- First,

$$\frac{d\nu^p}{d\mu_{m'}} = pt^{p-1}$$

Hellinger distance: properties

Example 2: Hellinger distance between ν^2 and ν^p w.r.t. μ_m where distortions are $n^p(t) = t^p$ and $m(t) = t$.

- First,

$$\frac{d\nu^p}{d\mu_{m'}} = pt^{p-1}$$

- Computation (with more Choquet integral – and Laplace transforms):

$$\begin{aligned} H_{\mu_{m'}}(\nu^2, \nu^p) &= \sqrt{\frac{1}{2}(C)\int_0^1 \left(\sqrt{\frac{d\nu^2}{\mu_m}} - \sqrt{\frac{d\nu^p}{\mu_m}}\right)^2 d\mu_m} \\ &= \sqrt{\frac{1}{2}(C)\int_0^1 \left(\sqrt{2t} - \sqrt{pt^{p-1}}\right)^2 d\mu_m} \quad (11) \\ &= \sqrt{1 - \frac{2\sqrt{2p}}{p+2}} \quad (12) \end{aligned}$$

Hellinger distance: properties

Properties:

- Compare:

$$H_{\mu_m}(\nu^2, \nu^p) = \sqrt{1 - \frac{4\sqrt{2p(p-1)}}{(p+2)p}}$$
$$H_{\mu_{m'}}(\nu^2, \nu^p) = \sqrt{1 - \frac{2\sqrt{2p}}{p+2}} \quad (13)$$

- The Hellinger distance depends on μ_m

Hellinger distance: properties

Properties:

- Compare:

$$
\begin{aligned}
H_{\mu_m}(\nu^2, \nu^p) &= \sqrt{1 - \frac{4\sqrt{2p(p-1)}}{(p+2)p}} \\
H_{\mu_{m'}}(\nu^2, \nu^p) &= \sqrt{1 - \frac{2\sqrt{2p}}{p+2}}
\end{aligned}
\tag{13}
$$

- The Hellinger distance depends on μ_m
  Different for additive measures: $H_\nu(\mu_1, \mu_2)$ is independent of ν.

Hellinger distance: properties

Properties related to the previous example:

When $p \to \infty$,
$H_{\mu_m}(\nu^2, \nu^p) = 1$ and $H_{\mu_{m'}}(\nu^2, \nu^p) = 1$.

Both $H_{\mu_m}(\nu^2, \nu^p)$ and $H_{\mu_{m'}}(\nu^2, \nu^p)$ are increasing w.r.t. $p > 2$, and the following holds

- $H_{\mu_m}(\nu^2, \nu^p) \in [0, 1]$ for all $p \geq 2$,
- $H_{\mu_{m'}}(\nu^2, \nu^p) \in [0, 1]$ for all $p \geq 2$.

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - Recall that conjugate of a measure: $\mu^c(A) = 1 - \mu_m(X \setminus A)$

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_{\nu^c}(\mu_1^c, \mu_2^c)$

    where $\mu^c(A) = 1 - \mu(X \setminus A)$

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_{\nu^c}(\mu_1^c, \mu_2^c)$

    where $\mu^c(A) = 1 - \mu(X \setminus A)$
- Partial answers:

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_{\nu^c}(\mu_1^c, \mu_2^c)$

    where $\mu^c(A) = 1 - \mu(X \setminus A)$
- Partial answers:
  - Dual of Distorted Lebesgue is Distorted Lebesgue

    $\mu^c(A) = 1 - \mu_m(X \setminus A) = 1 - m(1 - x)$

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_{\nu^c}(\mu_1^c, \mu_2^c)$

    where $\mu^c(A) = 1 - \mu(X \setminus A)$
- Partial answers:
  - Dual of Distorted Lebesgue is Distorted Lebesgue
    $\mu^c(A) = 1 - \mu_m(X \setminus A) = 1 - m(1 - x)$
  - If ν is submodular, ν^c is supermodular
    So, $H_\nu(\mu_1, \mu_2)$ is a distance but $H_{\nu^c}(\mu_1^c, \mu_2^c)$ is not

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_{\nu^c}(\mu_1^c, \mu_2^c)$

    where $\mu^c(A) = 1 - \mu(X \setminus A)$
- Partial answers:
  - Dual of Distorted Lebesgue is Distorted Lebesgue
    $\mu^c(A) = 1 - \mu_m(X \setminus A) = 1 - m(1 - x)$
  - If ν is submodular, ν^c is supermodular
    So, $H_\nu(\mu_1, \mu_2)$ is a distance but $H_{\nu^c}(\mu_1^c, \mu_2^c)$ is not
    Therefore, only $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$ makes sense

Hellinger distance: properties

Properties:

- Conjugate of the measure, same distance ?
  - First question, which conjugate in $H_\nu(\mu_1, \mu_2)$?
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$
    - ⋆ $H_\nu(\mu_1, \mu_2) = (?) H_{\nu^c}(\mu_1^c, \mu_2^c)$

    where $\mu^c(A) = 1 - \mu(X \setminus A)$
- Partial answers:
  - Dual of Distorted Lebesgue is Distorted Lebesgue
    $\mu^c(A) = 1 - \mu_m(X \setminus A) = 1 - m(1 - x)$
  - If ν is submodular, ν^c is supermodular
    So, $H_\nu(\mu_1, \mu_2)$ is a distance but $H_{\nu^c}(\mu_1^c, \mu_2^c)$ is not
    Therefore, only $H_\nu(\mu_1, \mu_2) = (?) H_\nu(\mu_1^c, \mu_2^c)$ makes sense
  - This case, difficult (work in progress)
    E.g., if $m(x) = x^2$, then $m^c(x) = 2x - x^2$.

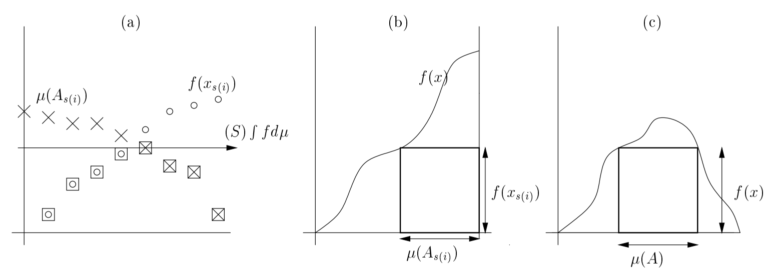
Some definitions (II): The Sugeno integral

Definitions: integrals

Sugeno integral (Sugeno, 1974):

- μ a non-additive measure, g a measurable function. The Sugeno integral of g w.r.t. μ, where $\mu_g(r) := \mu(\{x|g(x) > r\})$:

$$(S)\int g d\mu := \sup_{r\in[0,1]} [r \wedge \mu_g(r)]. \quad (14)$$



Definitions: integrals

Sugeno integral. Discrete version

- μ a non-additive measure, f a measurable function. The Sugeno integral of f w.r.t. μ,

$$(S)\int f d\mu = \max_{i=1,N} \min(f(x_{s(i)}), \mu(A_{s(i)})),$$

where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq ... \leq f(x_{s(N)}) \leq 1$ and $A_{s(i)} = \{x_{s(i)}, ..., x_{s(N)}\}$.

Definitions: properties

Properties: (X a reference set, a a value in $[0,1]$)

- f, g functions $f, g : X \to [0,1]$. Then,
  - $\mathcal{I}$ is minimum homogeneous if and only if, for comonotonic f and g,
  $$\mathcal{I}(a \wedge f) = a \wedge \mathcal{I}(f)$$
  - $\mathcal{I}$ is comonotonic maxitive if and only if, for comonotonic f and g,
  $$\mathcal{I}(f \vee g) = \mathcal{I}(f) \vee \mathcal{I}(g)$$

Definitions: properties

Characterization of the Sugeno integral

- Theorem (Ralescu and Sugeno, 1996; Marichal, 2000; Benvenuti and Mesiar, 2000). Let $\mathcal{I} : [0,1]^n \to \mathbb{R}_+$ be a functional with the following properties
  - $\mathcal{I}$ is comonotonic monotone
  - $\mathcal{I}$ is comonotonic maxitive
  - $\mathcal{I}$ is minimum homogeneous
  - $\mathcal{I}(1, \ldots, 1) = 1$

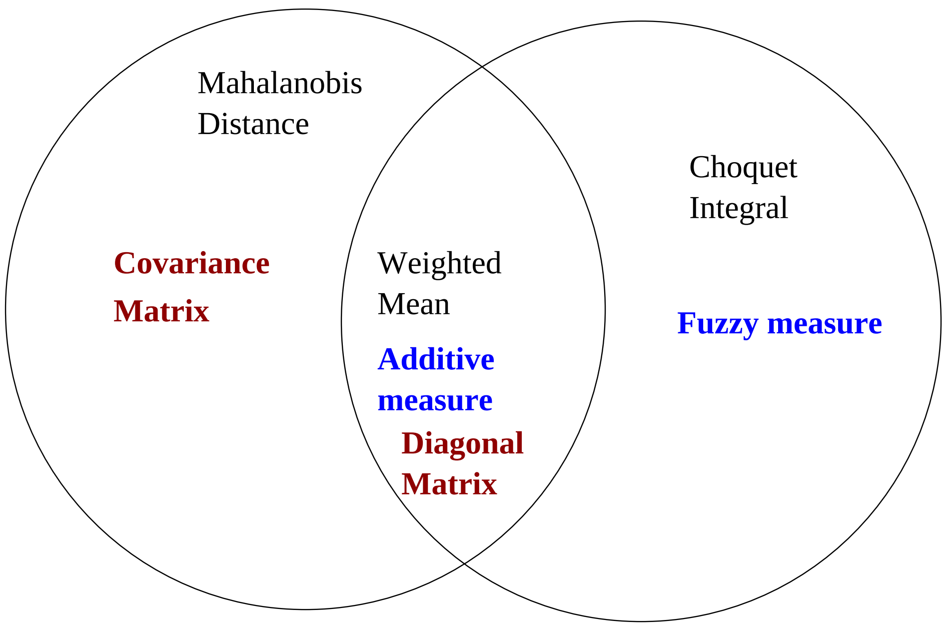
  Then, there exists a fuzzy measure μ such that $\mathcal{I}(f)$ is the Sugeno integral of f with respect to μ.

Applications

Aggregation operators

Independence.

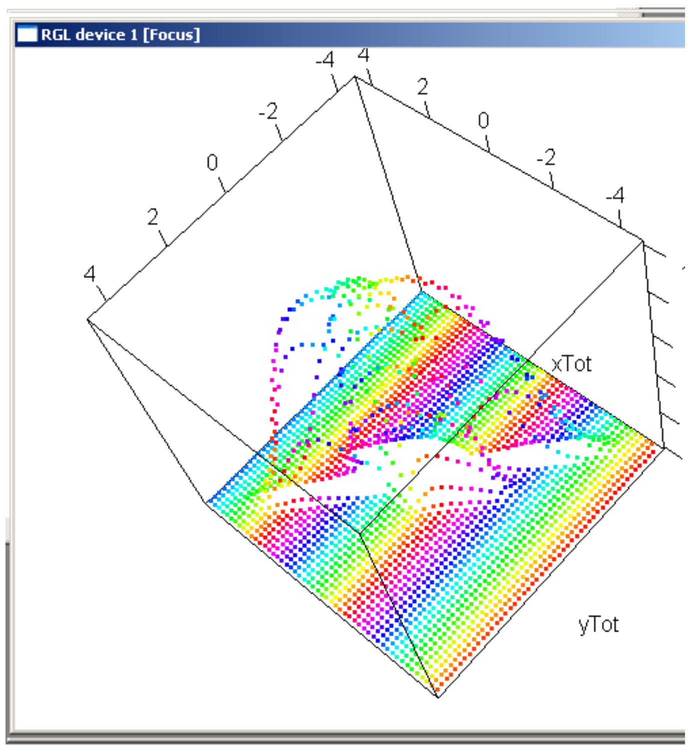
- Choquet integral and Mahalanobis distance
  - Mahalanobis: covariance matrix
  - Choquet integral: fuzzy measure
- In a single framework: Mahalanobis and Choquet *distance*



Aggregation operators

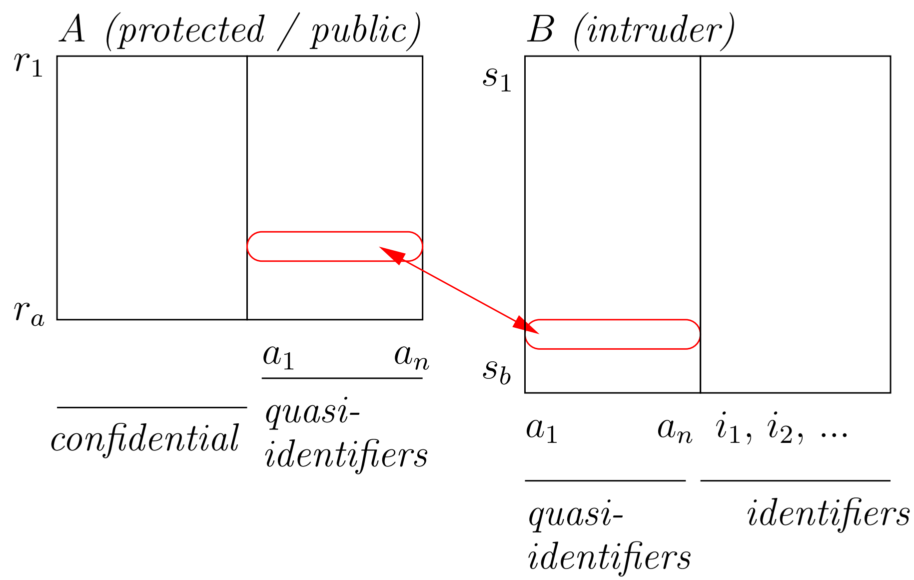
Independence.

- Choquet integral and Mahalanobis distance
  - Mahalanobis: covariance matrix
  - Choquet integral: fuzzy measure
- A generalization: Choquet-Mahalanobis distance/distribution



Record Linkage

Record Linkage:



Record Linkage

Record Linkage:

$$Minimize \sum_{i=1}^{N} K_i \tag{15}$$

$$Subject\ to:$$

$$CI_\mu(d(V_1(a_i), V_1(b_j)), \ldots, d(V_n(a_i), V_n(b_j)))- \\ - CI_\mu(d(V_1(a_i), V_1(b_i)), \ldots, d(V_n(a_i), V_n(b_i))) + CK_i > 0 \quad \forall i \forall j \tag{16}$$

$$K_i \in \{0, 1\} \tag{17}$$

$$\mu(A) \in [0, 1] \tag{18}$$

$$\mu(A) \leq \mu(B) \quad \forall A, B\ s.t.\ \ A \subseteq B \subseteq X \tag{19}$$

Decision

Decision:

- Different alternatives
- Users have preferences (an order on the alternatives $\prec$)
- GOAL: We want to model these preferences (to model $\prec$)

Decision under certainty

Decision under certainty. Multicriteria decision making

- Alternatives expressed in terms of utility functions
- Select best alternative by:

  Step 1. Aggregate utilities: Choquet integral for non-independence

  Step 2. Rank according to aggregated utilities

Criteria Satisfaction on:

alt	Price	Quality	Comfort
FordT	0.2	0.8	0.3
206	0.7	0.7	0.8
...	...		

→

alt	Consensus
FordT	0.35
206	0.72
...	...

→

alt	Ranking
206	0.72
FordT	0.35
...	...

Decision under uncertainty

Decision under uncertainty.

- Decision theory based on probability and utility functions to model lack of knowledge (Savage, 1954; Ramsey and von Neumann):
  - classical/subjective expected utility

Decision under uncertainty

Decision under uncertainty.

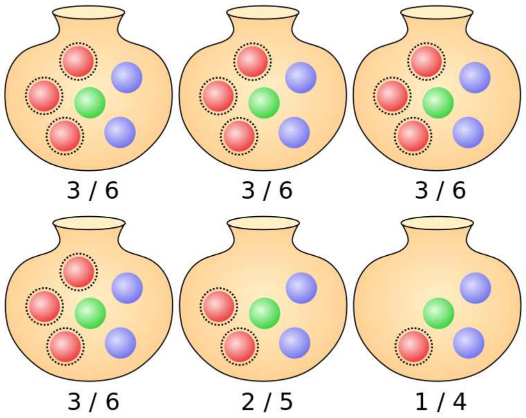
- Decision theory based on probability and utility functions to model lack of knowledge (Savage, 1954; Ramsey and von Neumann):
  - classical/subjective expected utility
- Ellsberg paradox: people behave differently than the model!!
  - Ellsberg paradox violates the postulates of the theory
  - Alternative model based on non-additive (fuzzy) measures

Decision under uncertainty: Ellsberg paradox

Decision making: (Ellsberg, 1961) 90 balls in an urn

- A player and different games, which prefer? $(f_R, f_B, ...)$

Color of balls	Red	Black	Yellow
Number of balls	30	60	
f_R	\$ 100	0	0
f_B	\$ 0	\$ 100	0
f_{RY}	\$ 100	0	\$ 100
f_{BY}	\$ 0	\$ 100	\$ 100



Decision under uncertainty: Ellsberg paradox

- How we model $\prec$ with classical expected utility ?
  - a (finite) state space S (*options* $=$ the balls)
  - a (finite) set of outcomes X (*benefits* $=$ the money)

Decision under uncertainty: Ellsberg paradox

- How we model $\prec$ with classical expected utility ?
  - a (finite) state space S (*options* = the balls)
  - a (finite) set of outcomes X (*benefits* = the money)
  - P be a probability measure on $(X, \mathcal{A}, P)$ (P on the balls)
  - $u : X \to \mathbb{R}^+$ be a utility function (utility of the money)

Decision under uncertainty: Ellsberg paradox

- How we model $\prec$ with classical expected utility ?
  - a (finite) state space S (*options* = the balls)
  - a (finite) set of outcomes X (*benefits* = the money)
  - P be a probability measure on $(X, \mathcal{A}, P)$ (P on the balls)
  - $u : X \to \mathbb{R}^+$ be a utility function (utility of the money)
  - a function from S to X (an act), $\mathcal{F}$ the set of acts (the alternatives).
  - User preferences on $\mathcal{F} = \{f | f : S \to X\}$ denoted by $\prec$

Decision under uncertainty: Ellsberg paradox

- How we model $\prec$ with classical expected utility ?
  - a (finite) state space S (*options* $=$ the balls)
  - a (finite) set of outcomes X (*benefits* $=$ the money)
  - P be a probability measure on $(X, \mathcal{A}, P)$ (P on the balls)
  - $u : X \to \mathbb{R}^+$ be a utility function (utility of the money)
  - a function from S to X (an act), $\mathcal{F}$ the set of acts (the alternatives).
  - User preferences on $\mathcal{F} = \{f | f : S \to X\}$ denoted by $\prec$
  - $\prec$ is represented by P and u when (user preference model)

$$E(u(f)) < E(u(g)) \text{ if and only if } f \prec g$$

where

$$E(u(f)) = \sum_{s \in S} u(f(s)) P(\{s\}) = \sum_{x \in X} u(x) P(f^{-1}(x)).$$

Decision under uncertainty: Ellsberg paradox

- Computation of the expected utility for a particular act (alternative)
  - $S = \{Red, Black, Yellow\}$
  - $f_{RY} =$ (0 for a Black, \$ 100 for Red, and \$ for Yellow)

$$
\begin{aligned}
E(u(f_{RY})) &= u(0)P(f^{-1}(0)) + u(100)P(f^{-1}(100)) \\
&= u(0)P(\{B\}) + u(100)P(\{Y, R\}) \\
&= u(0)P(\{B\}) + u(100)P(\{Y\}) + u(100)P(\{R\})
\end{aligned}
$$

  - Problem. Given a player, and preferences $\prec$, determine P and u
  - E.g., $P(x) = 1/3$ and $u(x) = x$.

Decision under uncertainty: Ellsberg paradox

Decision making: (Ellsberg, 1961) 90 balls in an urn

- A player and different games, which prefer? $(f_R, f_B, \ldots)$

Color of balls	Red	Black	Yellow
Number of balls	30	60	
f_R	\$ 100	0	0
f_B	\$ 0	\$ 100	0
f_{RY}	\$ 100	0	\$ 100
f_{BY}	\$ 0	\$ 100	\$ 100

- Most people prefer
  - $f_B \prec f_R$

Decision under uncertainty: Ellsberg paradox

Decision making: (Ellsberg, 1961) 90 balls in an urn

- A player and different games, which prefer? $(f_R, f_B, \ldots)$

Color of balls	Red	Black	Yellow
Number of balls	30	60	
f_R	\$ 100	0	0
f_B	\$ 0	\$ 100	0
f_{RY}	\$ 100	0	\$ 100
f_{BY}	\$ 0	\$ 100	\$ 100

- Most people prefer
  - $f_B \prec f_R$
  - $f_{RY} \prec f_{BY}$

Decision under uncertainty: Ellsberg paradox

Decision making: (Ellsberg, 1961) 90 balls in an urn

- A player and different games, which prefer? $(f_R, f_B, ...)$

Color of balls	Red	Black	Yellow
Number of balls	30	60	
f_R	\$ 100	0	0
f_B	\$ 0	\$ 100	0
f_{RY}	\$ 100	0	\$ 100
f_{BY}	\$ 0	\$ 100	\$ 100

- Most people prefer
  - $f_B \prec f_R$
  - $f_{RY} \prec f_{BY}$
- No solution exist with probabilities (additive measures), but can be solved with non-additive (fuzzy) measures

Decision under uncertainty: Ellsberg paradox

- Choquet expected utility model (Schmeidler, 1989)
  - Choquet integral (CI), utility u, non-additive (fuzzy) measure μ

$$
\begin{aligned}
E(u(f_{RY})) &= u(0)\mu(\{B\}) + u(100)\mu(\{Y,R\}) \\
&\neq u(0)\mu(\{B\}) + u(100)\mu(\{Y\}) + u(100)\mu(\{R\})
\end{aligned}
$$

  - User preferences on $\mathcal{F}$ denoted by $\prec$
  - $\prec$ is represented by P and u when (user preference model)

$$E(u(f)) = CI_{u,\mu}(f) < E(u(g)) = CI_{u,\mu}(g) \text{ if and only if } f \prec g$$

where

$$E(u(f)) = CI_{u,\mu}(f) = \sum_{x_{\sigma(i)} \in X} (u(x_{\sigma(i)}) - u(x_{\sigma(i-1)}))\mu(f^{-1}(x)).$$

Summary

Hellinger distance: properties

Summary:

- Review of non-additive measures
- Extension of the Hellinger distance to non-additive measures
- Some properties
- Some applications

Thank you

References

References.

- Denneberg, D. (1994) Non Additive Measure and Integral, Kluwer Academic Publishers.
- Graf, S. (1980) A Radon-Nikodym theorem for capacities, Journal für die reine und angewandte Mathematik, 1980 192-214.
- Narukawa, Y., Murofushi, T. (2008) Choquet Stieltjes integral as a tool for decision modeling, Int. J. of Intel. Syst. 23 115-127.
- Nguyen, H. T. (2006) An Introduction to Random Sets, Chapman and Hall, CRC Press.
- Schmeidler, D. (1986) Integral representation without additivity, Proceedings of the American Mathematical Society, 97 253-261.
- Sugeno, M. (2013) A note on derivatives of functions with respect to fuzzy measures, Fuzzy Sets and Systems 222 1-17.
- Torra, V., Narukawa, Y. (2007) Modeling Decisions: Information Fusion and Aggregation Operators, Springer.
- Torra, V., Narukawa, Y., Sugeno, M., Carlson, M. (2014) On the f-divergence for non-additive measures, submitted.
- Torra, V., Narukawa, Y., Sugeno, M. (eds) (2014) Non-additive measures: Theory and applications, Springer.

Choquet expected utility model

- Why classical expected utility cannot represent Ellsberg paradox ?
  - to representation $\prec$ in terms of u and P, we need

$$E(u(f)) \leq E(u(g)) \quad \text{for all} \quad f \prec g.$$

Choquet expected utility model

- Why classical expected utility cannot represent Ellsberg paradox ?
  - From $f_{RY} \prec f_{BY}$,

$$\begin{aligned} E(u(f_{RY})) &= u(0)P(B) + u(100)P(Y) + u(100)P(R) \\ &< u(100)P(B) + u(100)P(Y) + u(0)P(R) = E(u(f_{BY})) \end{aligned}$$

    so, $u(0)P(B) + u(100)P(R) < u(100)P(B) + u(0)P(R)$
  - From $f_B \prec f_R$,

$$\begin{aligned} E(u(f_B)) &= u(100)P(B) + u(0)P(Y) + u(0)P(R) \\ &< u(0)P(B) + u(0)P(Y) + u(100)P(R) = E(u(f_R)) \end{aligned}$$

    so, $u(100)P(B) + u(0)P(R) < u(0)P(B) + u(100)P(R)$.

    Inequalities 1 and 2 are in contradiction: no u and P exist

Choquet expected utility model

- How Choquet expected utility represents Ellsberg paradox ?
  Using:
  - $\mu(\emptyset) = 0$
  - $\mu(\{R\}) = 1/3$, $\mu(\{B\}) = \mu(\{Y\}) = 2/9$
  - $\mu(\{R, Y\}) = 5/9$, $\mu(\{B, Y\}) = \mu(\{R, B\}) = 2/3$
  - $\mu(\{R, B, Y\}) = 1$

Choquet expected utility model

- How Choquet expected utility represents Ellsberg paradox ?
  - From $f_{RY} \prec f_{BY}$ we have

$$\begin{aligned} CI_\mu(u(f_{RY})) &= u(0)\mu(\{B\}) + u(100)\mu(\{Y,R\}) \\ &< u(100)\mu(\{B,Y\}) + u(0)\mu(\{R\}) = CI_\mu(u(f_{BY})) \end{aligned}$$

    so, $0 \cdot 2/9 + 100 \cdot 5/9 < 100 \cdot 2/3 + 0 \cdot 1/3.$
  - From $f_B \prec f_R$,

$$\begin{aligned} CI_\mu(u(f_B)) &= u(100)\mu(\{B\}) + u(0)\mu(\{Y,R\}) \\ &< CI_\mu(u(f_R)) = u(0)\mu(\{B,Y\}) + u(100)\mu(\{R\}) \end{aligned}$$

    so, $100 \cdot 2/9 + 0 \cdot 5/9 < 0 \cdot 2/3 + 100 \cdot 1/3.$