IUKM 2019 - Nara, Japan

Choquet integral in decision making and metric learning

Vicenç Torra
Hamilton Institute, Maynooth University
Ireland

March 27, 2018

Overview

Basics and objectives:

- Using Choquet integral in two types of applications
  decision and metric learning (reidentification)
- Distances
- and distribution
  (for non-additive measures)

Outline

1. Preliminaries
   - Choquet integral: mathematical perspective
     - Non-additive measures
     - Now we need an integral
   - Choquet integral: Application perspective
     - Aggregation operators and CI in decision: MCDM
     - Aggregation operators and CI in reidentification: risk assessment
     - Zooming out
2. Distances in classification (filling the gaps)
3. Distributions

Choquet integral: a mathematical introduction

Non-additive measures

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:

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

Definitions: measures

Additive measures.

Example:

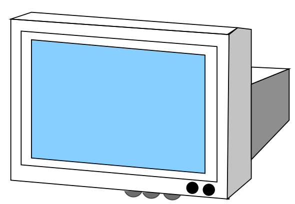
- Lebesgue measure. Unique measure λ s.t. $\lambda([a, b]) = b - a$ for every finite interval $[a, b]$
- Probability. When $\mu(X) = 1$.

Definitions: measures

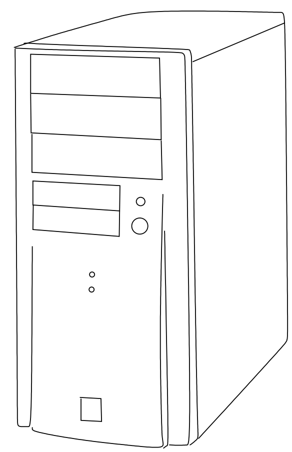
Additive measures.

Example:

- Lebesgue measure. Unique measure λ s.t. $\lambda([a, b]) = b - a$ for every finite interval $[a, b]$
- Probability. When $\mu(X) = 1$.
- Or just price ...



A



B

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**
  - $(X, \mathcal{A})$ 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)

- Naturally, additivity implies monotonicity
  - E.g., $B = A \cup C$ (with $A \cap C = \emptyset$) then $\mu(B) = \mu(A) + \mu(C) \geq \mu(A)$
  - But in non-additive measures, we allow

$$\mu(B = A \cup C) < \mu(A) + \mu(C)$$

$$\mu(B = A \cup C) > \mu(A) + \mu(C)$$

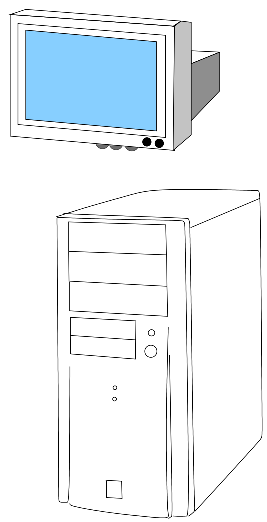
As e.g., $\mu(B) = 0.5 < \mu(A) + \mu(C) = 0.3 + 0.4 = 0.7$

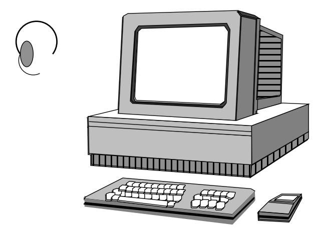
A way to represent interactions

Definitions: measures

- **Non-additive measures.** Price
  - When we have a discount, for disjoints A and B, we have

$$\mu(A \cup B) < \mu(A) + \mu(B) \quad \text{but} \quad \mu(A \cup B) \geq \mu(A)$$





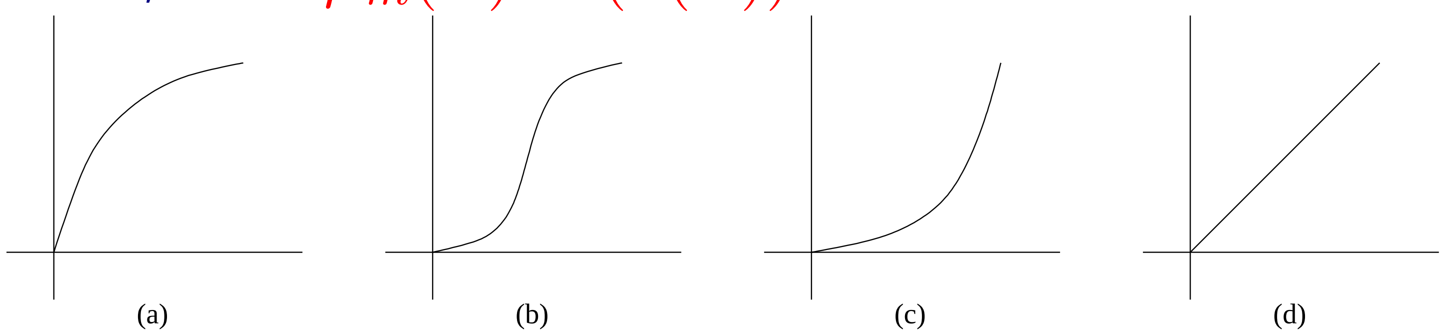
- There quite a large number of families of measures

Definitions: measures

- **Non-additive measures.** Distorted probabilities
  - $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function such that $m(0) = 0$; P be a probability.

$$\mu_{m,P}(A) = m(P(A)) \quad (1)$$

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



  - Used in economics: Prospect theory (Kahneman and Tversky, 1979). Small probabilities tend to be overestimated, while large ones, underestimated.

Definitions: measures

- **Non-additive measures.** Distorted Lebesgue
  - $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function such that $m(0) = 0$; λ be the Lebesgue measure.

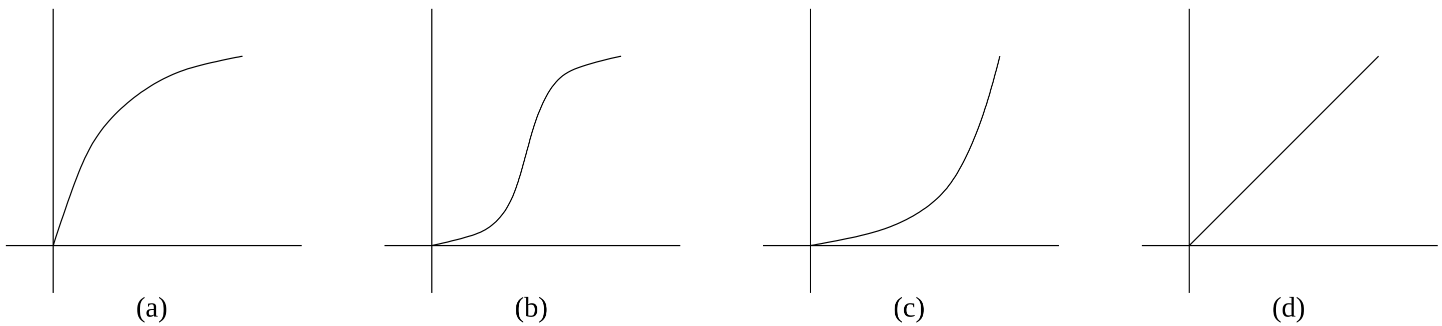
$$\mu_m(A) = m(\lambda(A)) \tag{2}$$

Definitions: measures

- **Non-additive measures.** Distorted Lebesgue
  - $m : \mathbb{R}^+ \to \mathbb{R}^+$ a continuous and increasing function such that $m(0) = 0$; λ be the Lebesgue measure.

$$\mu_m(A) = m(\lambda(A)) \quad (2)$$

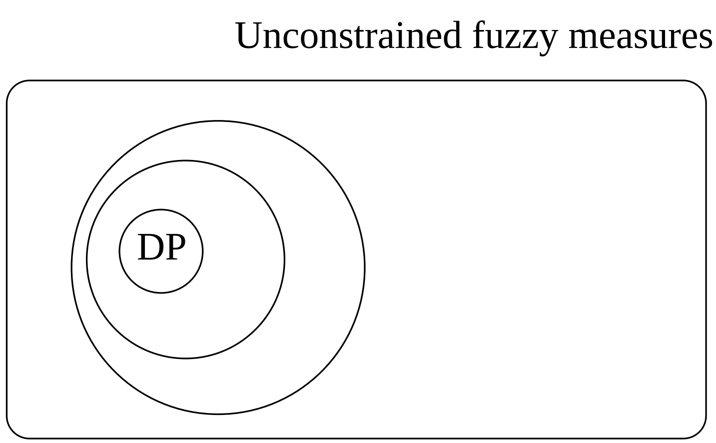
  - If $m(x) = x^2$, then $\mu_m(A) = (\lambda(A))^2$
  - If $m(x) = x^p$, then $\mu_m(A) = (\lambda(A))^p$



Definitions: measures

- **Non-additive measures.** A large number of families
  - Sugeno λ-measures: $\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B)$ ($\lambda > -1$)
  - For $\mathcal{P}$ a non empty set of probability measures, the upper and lower probabilities
    - $\bar{P}(A) = \sup_{P \in \mathcal{P}} P(A)$
    - $\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A)$

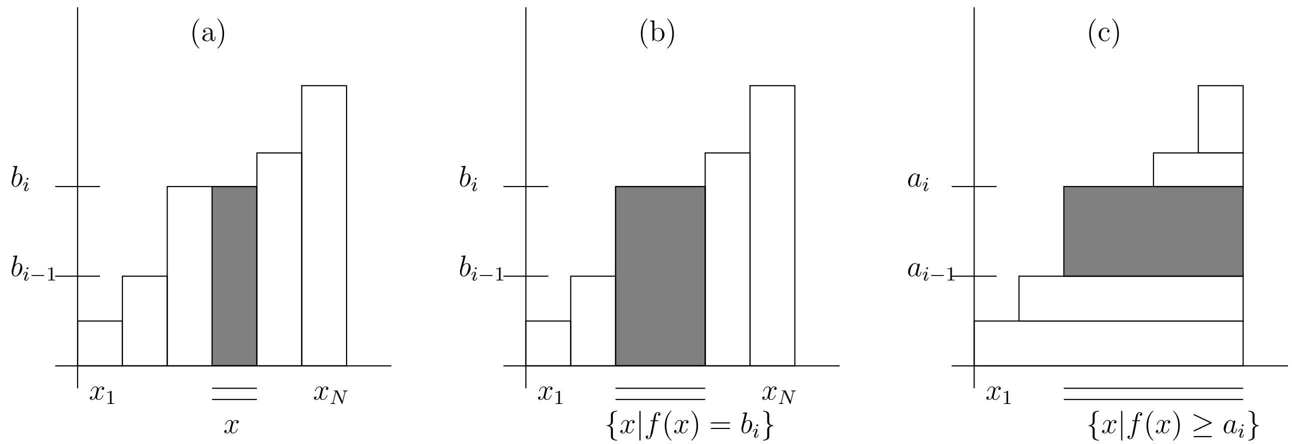
      (dual in the sense: $\bar{P}(A) = 1 - \underline{P}(A^c)$)
- m-dimensional distorted probabilities (NT/NT, 2005, 2011, 2012, 2018)



Now we need an integral

Definitions: integrals

- Additive measure: the way you add areas does not change[1] results



- Riemann integral (a) vs Lebesgue integral (c)
  - Riemann sum: $\sum_{I \in \mathcal{C}} f(x(I)) * \mu(I)$
    ($\mathcal{C}$ non-overlapping collection, $x(I)$ an element of I)
  - Lebesgue sum: $\sum_{a_i \in Range(f)} (a_i - a_{i-1})\mu(\Gamma(a_i))$
    where $\Gamma(a) := \{x|f(x) \geq a\}$

[1]Well, if it is calculable

Definitions: integrals

- Lebesgue integral

$$\int f d\mu := \int_0^\infty \mu_f(r) dr$$

where $\mu_f(r) = \mu(\{x | f(x) \geq r\})$

Definitions: integrals

- Choquet integral (Choquet, 1954):
  - μ a non-additive measure, f a measurable function. The Choquet integral of f w.r.t. μ, where $\mu_f(r) := \mu(\{x|f(x) > r\})$:

$$(C)\int f d\mu := \int_0^\infty \mu_f(r) dr.$$

Definitions: integrals

- Choquet integral (Choquet, 1954):
  - μ a non-additive measure, f a measurable function. The Choquet integral of f w.r.t. μ, where $\mu_f(r) := \mu(\{x|f(x) > r\})$:

$$(C)\int f d\mu := \int_0^\infty \mu_f(r) dr.$$

- Properties.
  - When the measure is additive, this is the Lebesgue integral (standard integral)

Definitions: integrals

Choquet integral. Discrete version

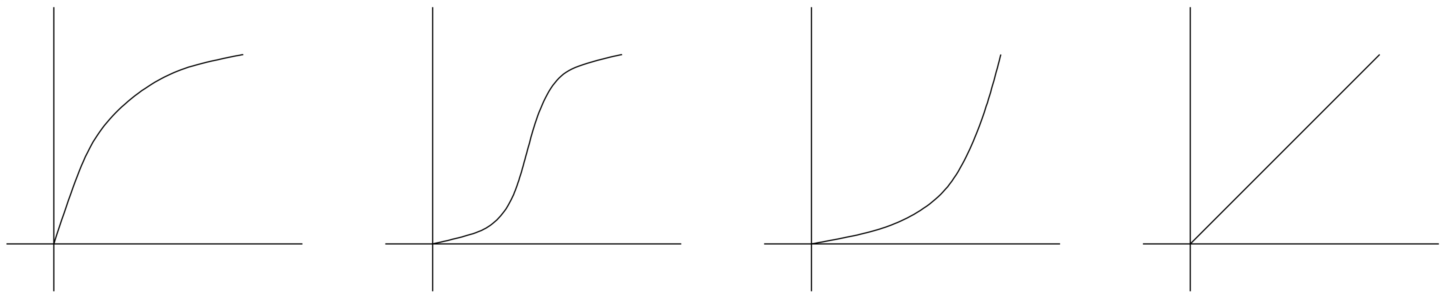
- μ a non-additive measure, f a measurable function. The Choquet integral of f w.r.t. μ,

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

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: integrals

- **Choquet integral.** Example:
  - Distorted probability $\mu_m(A) = m(P(A))$ (with $m(0) = 0$, $m(1) = 1$)
    $CI_{\mu_m}(f)$: $(a) \to$ *max*, $(b) \to$ *median*, $(c) \to$ *min*, $(d) \to$ *mean (expectation)*

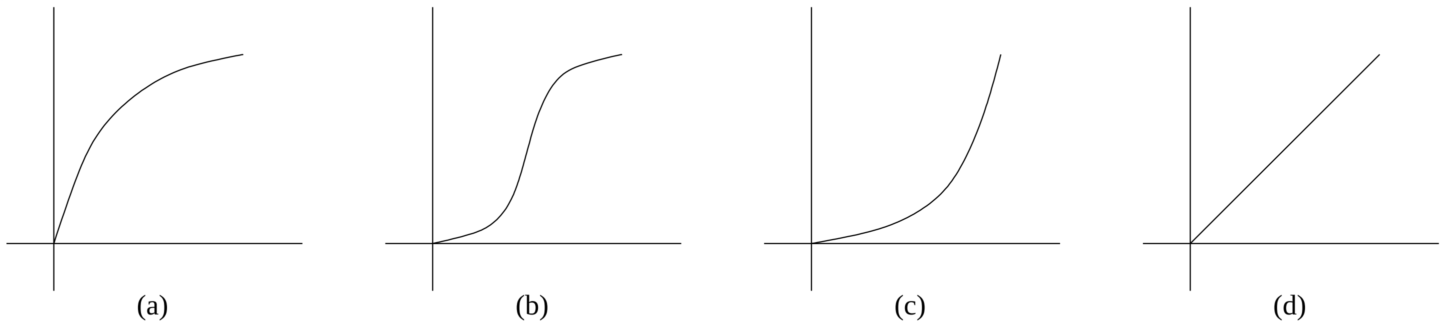


(a) (b) (c) (d)

  - Upper and lower probabilities: bounds for expectations
    $CI_{\underline{P}}(f) \leq \inf_P E_P(f) \leq \sup_P E_P(f) \leq CI_{\bar{P}}(f)$

Definitions: integrals

- **Choquet integral.** Example:
  - Distorted probability $\mu_m(A) = m(P(A))$ (with $m(0) = 0$, $m(1) = 1$)
    $CI_{\mu_m}(f)$: $(a) \to$ *max*, $(b) \to$ *median*, $(c) \to$ *min*, $(d) \to$ *mean (expectation)*



  - Upper and lower probabilities: bounds for expectations
    $CI_{\underline{P}}(f) \leq \inf_P E_P(f) \leq \sup_P E_P(f) \leq CI_{\bar{P}}(f)$
  - $(C) \int \chi_A d\mu = \mu(A)$

Application I
Aggregation operators & Choquet integral in Decision

MCDM: Aggregation for (numerical) utility functions

Aggregation and Choquet integral in MCDM

- Decision, utility functions

Aggregation and Choquet integral in MCDM

- Decision, utility functions

  Alternatives $=\{$ Ford T, Seat 600, Simca 1000, VW, Citr.Acadiane $\}$

Aggregation and Choquet integral in MCDM

- Decision, utility functions

  Alternatives = { Ford T, Seat 600, Simca 1000, VW, Citr.Acadiane }

  Criteria = { Seats, Security, Price, Comfort, trunk}

Aggregation and Choquet integral in MCDM

- Decision, utility functions

  Alternatives = { Ford T, Seat 600, Simca 1000, VW, Citr.Acadiane }

  Criteria = { Seats, Security, Price, Comfort, trunk}

  Decision making process:

Aggregation and Choquet integral in MCDM

- Decision, utility functions

  Alternatives = { Ford T, Seat 600, Simca 1000, VW, Citr.Acadiane }

  Criteria = { Seats, Security, Price, Comfort, trunk}

  Decision making process:

  Modelling=Criteria + Utilities, aggregation, selection

	Number of seats	Security	Price	Confort	trunk
Ford T	0	20	0	20	0
Seat 600	60	0	100	0	50
Simca 1000	100	30	100	50	70
VW Beetle	80	50	30	70	100
Citroën Acadiane	20	40	60	40	0

Aggregation and Choquet integral in MCDM

- Decision, utility functions
  Alternatives = { Ford T, Seat 600, Simca 1000, VW, Citr.Acadiane }
  Criteria = { Seats, Security, Price, Comfort, trunk}
  Decision making process:

Aggregation and Choquet integral in MCDM

- Decision, utility functions

  Alternatives = { Ford T, Seat 600, Simca 1000, VW, Citr.Acadiane }

  Criteria = { Seats, Security, Price, Comfort, trunk}

  Decision making process:

  Modelling, aggregation = $\mathbb{C}$, selection

	Seats	Security	Price	Comfort	trunk	$\mathbb{C} = AM$
Ford T	0	20	0	20	0	8
Seat 600	60	0	100	0	50	42
Simca 1000	100	30	100	50	70	70
VW	80	50	30	70	100	66
Citr. Acadiane	20	40	60	40	0	32

Aggregation and Choquet integral in MCDM

- MCDM: Aggregation to deal with contradictory criteria

Aggregation and Choquet integral in MCDM

- MCDM: Aggregation to deal with contradictory criteria
- But there are occasions in which ordering is clear

$$\text{when } a_i \leq b_i \text{ it is clear that } a \leq b$$

E.g.,

	Seats	Security	Price	Comfort	trunk	$\mathbb{C} = AM$
Seat 600	60	0	100	0	50	42
Simca 1000	100	30	100	50	70	70

Aggregation and Choquet integral in MCDM

- MCDM: Aggregation to deal with contradictory criteria
- But there are occasions in which ordering is clear

$$\text{when } a_i \leq b_i \text{ it is clear that } a \leq b$$

E.g.,

	Seats	Security	Price	Comfort	trunk	$\mathbb{C} = AM$
Seat 600	60	0	100	0	50	42
Simca 1000	100	30	100	50	70	70

Aggregation operators are appropriate because they satisfy monotonicity

Aggregation and Choquet integral in MCDM

- Decision making process:

Aggregation and Choquet integral in MCDM

- Decision making process:

  Modelling, aggregation, selection=order,first

Aggregation and Choquet integral in MCDM

- Decision making process:

  Modelling, aggregation, selection=order,first

- The function of aggregation functions
  - Different aggregations lead to different orders (in the PF)

Aggregation and Choquet integral in MCDM

- Decision making process:

  Modelling, aggregation, selection=order,first

- The function of aggregation functions
  - Different aggregations lead to different orders (in the PF)
  - Aggregation establishes which points are *equivalent*
  - Different aggregations, lead to different curves of points (level curves)

Criteria
Satisfaction on:

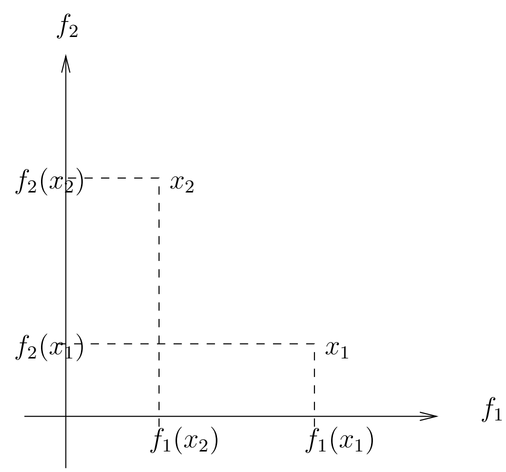
alt	Price	Quality	Comfort
FordT	0.2	0.8	0.3
206	0.7	0.7	0.8
...	...		

→

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

→

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



Aggregation and Choquet integral in MCDM

- Aggregation functions and different level curves
  - Arithmetic mean
  - Geometric mean, Harmonic mean, ...
  - Weighted mean
  - OWA, ...

Aggregation and Choquet integral in MCDM

- Aggregation functions and different level curves
  - Arithmetic mean
  - Geometric mean, Harmonic mean, ...
  - Weighted mean
  - OWA, ...
  - Choquet integral (generalization of the AM, WM, OWA)
    - ▷ to represent interactions between criteria
    - ▷ non-independent criteria allowed

Aggregation and Choquet integral in MCDM

- Aggregation functions and parameters
  - Arithmetic mean: no parameters
  - Geometric mean, Harmonic mean, ...: : no parameters
  - Weighted mean: weighting vector
  - OWA, ...: weighting vector
  - Choquet integral (generalization of the AM, WM, OWA) a measure
    - to represent interactions between criteria

      w(security,price,confort) $>$ *(or* $<$*) w(security)*$+$*w(price)*$+$*w(confort)*
    - non-independent criteria allowed

      $$\mu(\{c_1, c_2\}) \neq \mu(\{c_1\}) + \mu(\{c_2\})$$
    - $(C) \int \chi_A d\mu = \mu(A)$

MCDM: What fuzzy measures (and CI) can represent?

Aggregation and Choquet integral in MCDM

- Choquet integral can, and WM/Probability model cannot
  - An element/criteria is added into the set, and

    the preference is reversed

Aggregation and Choquet integral in MCDM

- Choquet integral can, and WM/Probability model cannot
  - An element/criteria is added into the set, and
    the preference is reversed
  - Example. Buying a house.
    When public transport is available, the preference changes[2]
    - ▷ If there is no bus I prefer a public library than a restaurant, but if there is a bus then I instead prefer the restaurant near.
    - ▷ Mathematically, with *B=Bus*, *R=Restaurant*, *L=Library* we have

$$\mu(\{R\}) \leq \mu(\{L\}) \text{ but } \mu(\{R, B\}) \geq \mu(\{L, B\})$$

[2] Ellesberg's paradox.

MCDM: Learn/identify the parameters (e.g. the measures)

Aggregation and Choquet integral in MCDM

- Available information?
  - Find measures from outcome: column vector with outcome

	Seats	Security	Price	Comfort	trunk	$\mathbb{C} = CI_\mu$
Seat 600	60	0	100	0	50	42
Simca 1000	100	30	100	50	70	70
...						

  - Find measures from preferences – (partial) order $<$: $S = \{(r_i, t_i)\}_i$

	Seats	Security	Price	Comfort	trunk	$\mathbb{C} = CI_\mu$
Seat 600	60	0	100	0	50	4th
Simca 1000	100	30	100	50	70	1st
...						

Aggregation and Choquet integral in MCDM

- Available information?
  - Measures from outcome: a column vector $\Rightarrow \min \sum (\mathbb{C}_P(a_r) - o_r)^2$
  - Measures from preferences – (partial) order $<$: $S = \{(r_i, t_i)\}_i$
    - ▷ Formulation: Find μ such that, for all $(r, t) \in S$, it follows that

$$\mathbb{C}_P(\text{evaluation-car r}) > \mathbb{C}_P(\text{evaluation-car t})$$

or, with a_r and a_s for rows r and s,

$$\mathbb{C}_P(a_{r1}, \ldots, a_{rn}) > \mathbb{C}_P(a_{t1}, \ldots, a_{tn})$$

Unfortunately, often, no solution: minimize failures $y_{(r,t)} \geq 0$

$$\mathbb{C}_P(a_{r1}, \ldots, a_{rn}) - \mathbb{C}_P(r_{t1}, \ldots, a_{tn}) + y_{(r,t)} > 0.$$

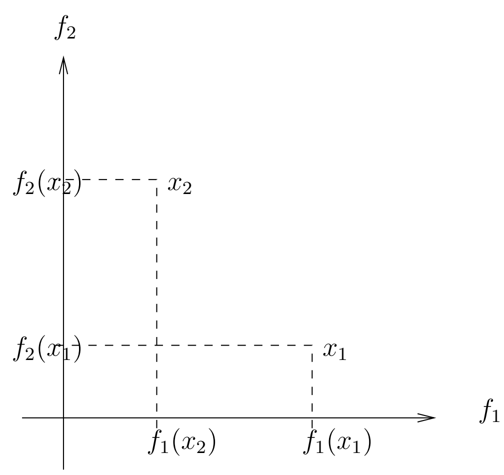
Aggregation and Choquet integral in MCDM

- Available information?
  - Measures from outcome: a column vector $\Rightarrow \min \sum (\mathbb{C}_P(a_r) - o_r)^2$
  - Measures from preferences – (partial) order $<$: $S = \{(r_i, t_i)\}_i$
    - ▷ Formulation: Find μ such that, for all $(r,t) \in S$, it follows that

$$
\begin{array}{ll}
\text{Minimize } \sum_{(r,t)\in S} y_{(r,t)} & \\
\text{Subject to} & \\
\quad \mathbb{C}_P(a_{r1}, \ldots, a_{rn}) - \mathbb{C}_P(a_{t1}, \ldots, a_{tn}) + & y_{(r,t)} > 0 \\
 & y_{(r,t)} \geq 0 \\
\quad \text{logical constraints on } P &
\end{array}
$$

Aggregation and Choquet integral in MCDM

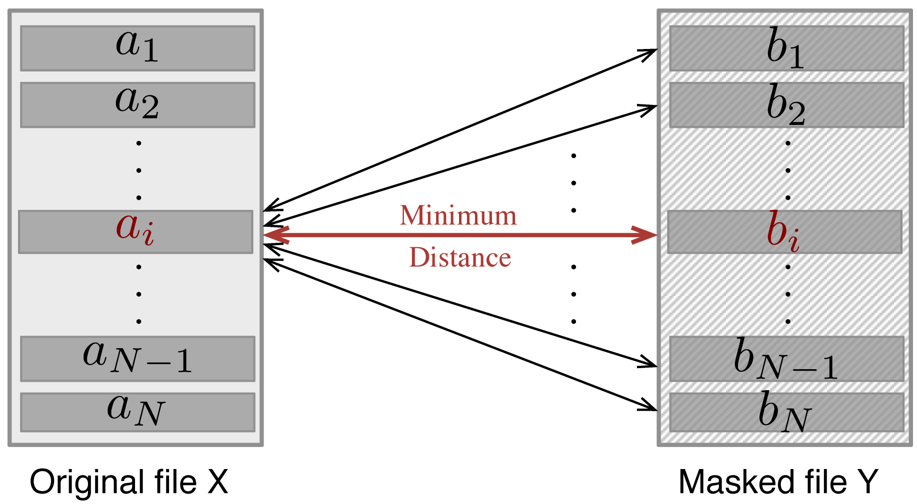
- Aggregation and selection
  - Selection of the one with maximum value of $\mathbb{C} = CI$ with μ (maximum distance to nadir – worst combination)
  $$d((a_1, \ldots, a_n), (0, \ldots, 0))$$
  - Selection of the one with minimum distance to ideal
  $$d((a_1, \ldots, a_n), (100, \ldots, 100))$$
  where d is computed as an aggregation



Application II
The Choquet integral in metric learning: reidentification

Aggregation operators and CI in reidentification

- **Re-identification**. Record linkage for databases, supervised approach
  - ML/Optimization for distance-based RL (A and B aligned).
    - ▷ Goal: as many correct reidentifications as possible: for each record i, we need $d(a_i, b_j) \geq d(a_i, b_i)$ for all j



$a_i = (a_{i1}, \ldots, a_{in})$ and $b_i = (b_{i1}, \ldots, b_{in})$

Aggregation operators and CI in reidentification

- **Re-identification**. Record linking for databases. Supervised approach
  - ML/Optimization for distance-based approach. (A and B aligned)
    - Goal: as many correct reidentifications as possible. But, if error for a_i: $K_i = 1$ and $d(a_i, b_j) + CK_i \geq d(a_i, b_i)$ for all j
    - or, expanding d,
    $\mathbb{C}_p(\mathit{diff}_1(a_{i1}, b_{j1}), \ldots, \mathit{diff}_n(a_{in}, b_{jn}) + CK_i \geq \mathbb{C}_p(\mathit{diff}_1(a_{i1}, b_{i1}), \ldots, \mathit{diff}_n(a_{in}, b_{in}))$
  - Formalization:

$$\text{Minimize} \quad \sum_{i=1}^{N} K_i$$

$$\text{Subject to:} \mathbb{C}_p(\mathit{diff}_1(a_{i1}, b_{j1}), \ldots, \mathit{diff}_n(a_{in}, b_{jn})) -$$

$$- \mathbb{C}_p(\mathit{diff}_1(a_{i1}, b_{i1}), \ldots, \mathit{diff}_n(a_{i1}, b_{i1})) + CK_i > 0$$

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

Additional constraints according to $\mathbb{C}$

Aggregation operators and CI in reidentification

- **Re-identification**. Record linking for databases. Supervised approach
  - ML/Optimization for distance-based approach. (A and B aligned)
  - Formalization for CI

$$\text{Minimize} \quad \sum_{i=1}^{N} K_i$$

$$\text{Subject to:} CI_\mu(\mathit{diff}_1(a_{i1}, b_{j1}), \ldots, \mathit{diff}_n(a_{in}, b_{jn})) -$$

$$- CI_\mu(\mathit{diff}_1(a_{i1}, b_{i1}), \ldots, \mathit{diff}_n(a_{i1}, b_{i1})) + CK_i > 0$$

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

Additional constraints for μ

Aggregation operators and CI in reidentification

- **Re-identification**. Record linking for databases. Supervised approach
  - ML/Optimization for distance-based approach. (A and B aligned)
  - Formalization for CI

$$\text{Minimize} \quad \sum_{i=1}^{N} K_i$$

$$\text{Subject to:} CI_\mu(\mathit{diff}_1(a_{i1}, b_{j1}), \ldots, \mathit{diff}_n(a_{in}, b_{jn})) -$$

$$- CI_\mu(\mathit{diff}_1(a_{i1}, b_{i1}), \ldots, \mathit{diff}_n(a_{i1}, b_{i1})) + CK_i > 0$$

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

Additional constraints for μ

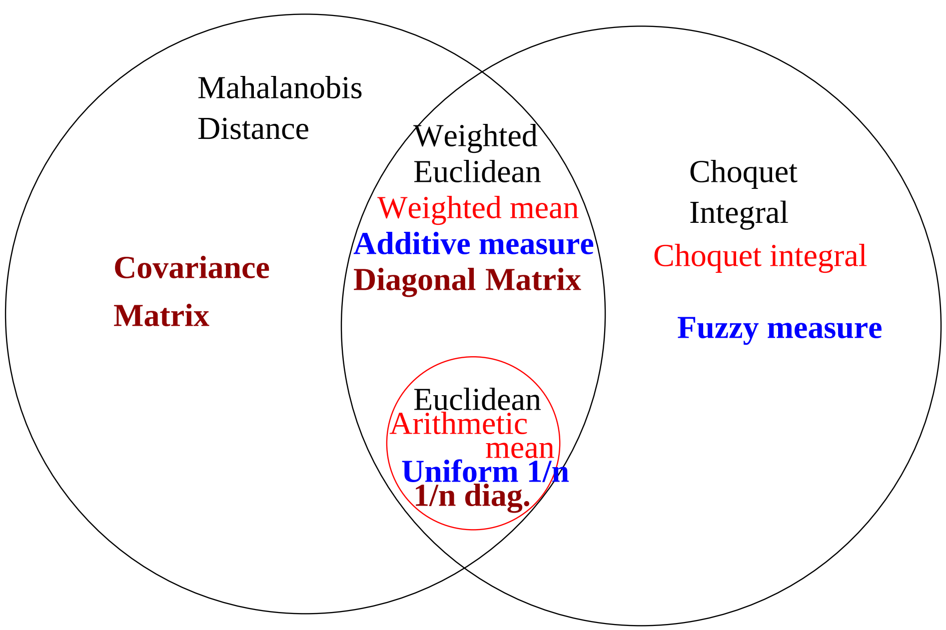
(but also WM, OWA, and Bilinear distance)

Zooming out: trying to understand

Aggregation, distances, and independence

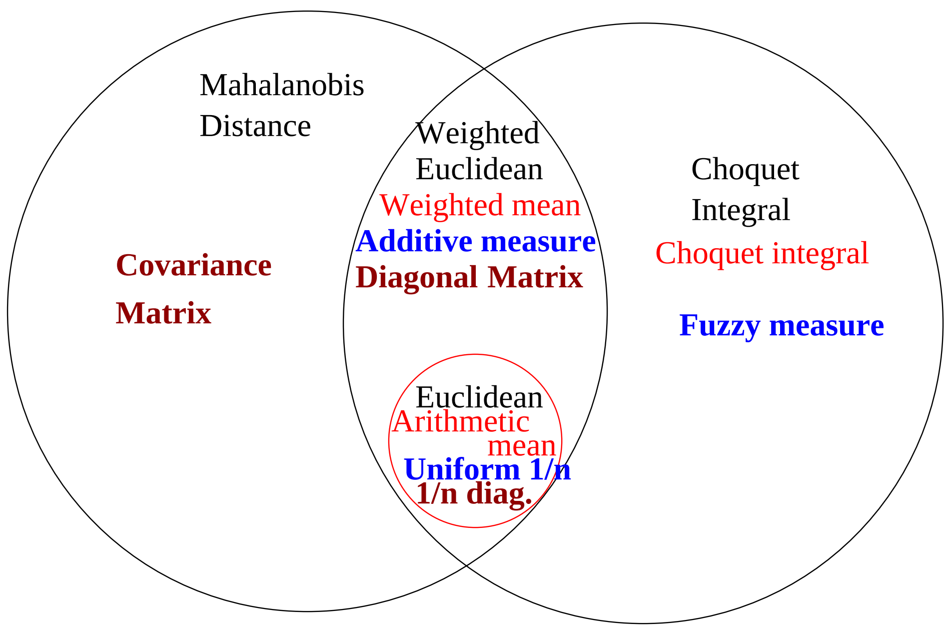
Aggregation, distance and independence

- **Aggregation and distance.**
  - Arithmetic mean (AM): Euclidean distance
  - Weighted mean (WM): Weighted euclidean
  - Choquet integral (CI): Choquet integral-based distance
  - ——— : Bilinear/Mahalanobis distance
- In a single picture: Mahalanobis and Choquet *distance*



Aggregation, distance and independence

- **Aggregation, distance and independence.**
  - Only with Choquet integral and Mahalanobis distances
    - Mahalanobis: covariance matrix
    - Choquet integral: fuzzy measure
  - In a single framework: Mahalanobis and Choquet *distance*



Filling gaps:

Aggregation, distances, and independence

Aggregation, distance and independence

- **Mahalanobis *distance*.**
  - between $\mathbf{x} \in \mathbb{R}^d$ and a vector $\mathbf{m} \in \mathbb{R}^d$
    with respect to the covariance matrix Σ

$$(\mathbf{x} - \mathbf{m})\Sigma^{-1}(\mathbf{x} - \mathbf{m}))$$

Aggregation, distance and independence

- **Choquet integral *distance*.**
  - between $\mathbf{x} \in \mathbb{R}^d$ and a vector $\mathbf{m} \in \mathbb{R}^d$
    with respect to a non-additive measure μ

$$CI_\mu((\mathbf{x} - \mathbf{m}) \circ (\mathbf{x} - \mathbf{m}))$$

$\mathbf{v} \circ \mathbf{w}$ is the Hadamard or Schur (elementwise) product of $\mathbf{v}$ and $\mathbf{w}$
(i.e., $(\mathbf{v} \circ \mathbf{w}) = (v_1 w_1 \ldots v_n w_n)$).

Aggregation, distance and independence

- **Choquet-Mahalanobis integral *distance*.**
  - between $\mathbf{x} \in \mathbb{R}^d$ and a vector $\mathbf{m} \in \mathbb{R}^d$
    with respect to μ and a positive-definite matrix Q

$$CMI(\mathbf{m}, \mu, \mathbf{Q}) = CI_{\mu}(\mathbf{v} \circ \mathbf{w})$$

where
- ▷ $\mathbf{L}\mathbf{L}^T = \mathbf{Q}$ is the Cholesky decomposition of the matrix $\mathbf{Q}$,
- ▷ $\mathbf{v} = (\mathbf{x} - \mathbf{m})^T \mathbf{L}$,
- ▷ $w = \mathbf{L}^T (\mathbf{x} - \mathbf{m})$, and where
- ▷ $\mathbf{v} \circ \mathbf{w}$ is the Hadamard (elementwise) product of $\mathbf{v}$ and $\mathbf{w}$.

Choquet integral based distribution: generalized *distance*

Well defined when Σ is a covariance matrix.

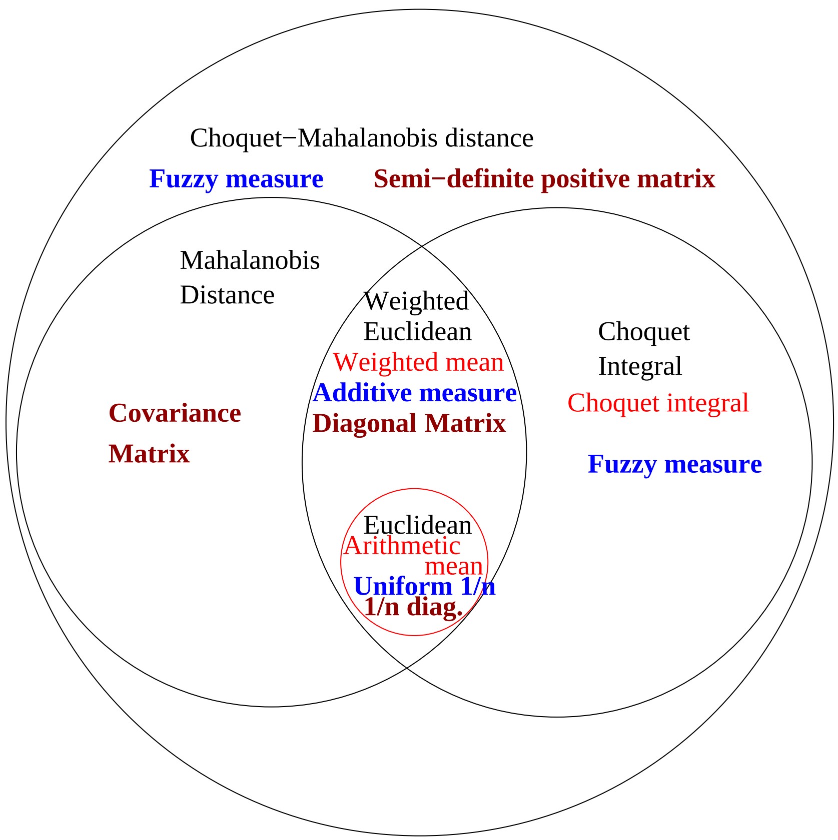
- When Σ^{-1} is a definite-positive matrix, the Cholesky descomposition is unique. This is the case when Σ is a covariance matrix valid for generating a probability-density function.

Proper generalization:

- Generalization of both the Mahalanobis and the Choquet integral based distance.
  - The definition with Σ equal to the identity results into the Choquet integral of $(x - \bar{x}) \otimes (x - \bar{x})$ with respect to μ.
  - The definition with μ corresponding to an additive probability $\mu(A) = 1/|A|$ results into $1/n$ of the Mahalanobis distance with respect to Σ.

Aggregation, distance and independence

- **Aggregation and distance.**
  - Arithmetic mean (AM): Euclidean distance
  - Weighted mean (WM): Weighted euclidean
  - Choquet integral (CI): Choquet integral-based distance
  - —— : Bilinear/Mahalanobis distance
  - Choquet-Mahalanobis *integral*: CMI-distance



A natural construction:

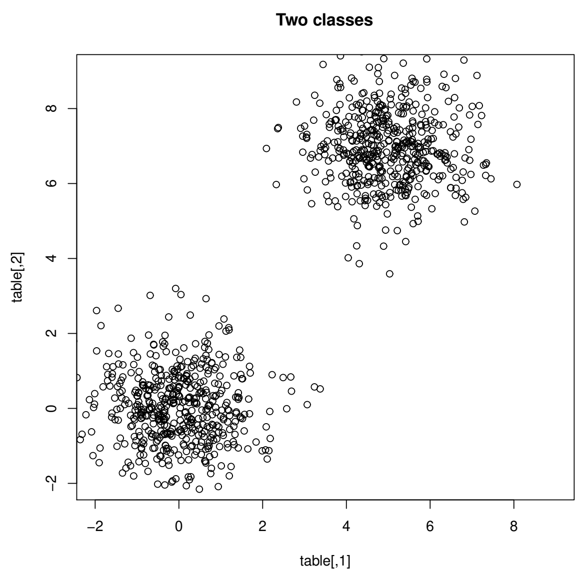
Distributions

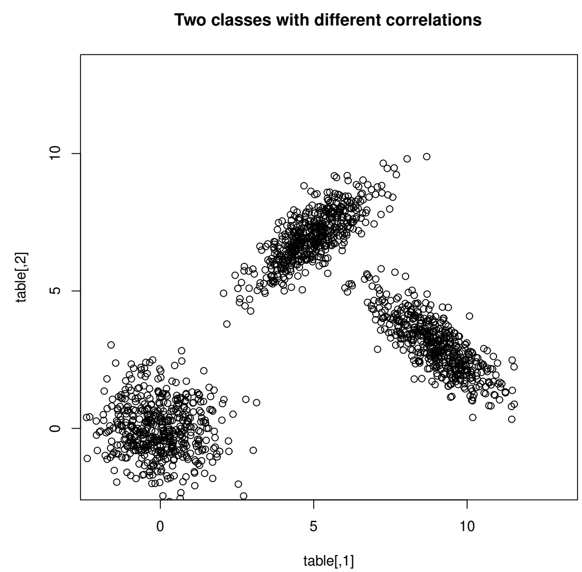
Distributions

- E.g. in Classification data drawn from normal Gaussian distributions.
  - Parameters $N(\mu, \Sigma)$ determined from real data or *known*
  - Set of k classes $\Omega = \{\omega_1, \dots, \omega_k\}$
  - covariance matrices Σ_i
  - means $\bar{x}_i$

  class-conditional probability-density function Gaussian distribution

$$P(x|\omega_i) = \frac{1}{(2\pi)^{m/2}|\Sigma_i|^{1/2}} e^{-\frac{1}{2}(x-\bar{x}_i)^T \Sigma_i^{-1}(x-\bar{x}_i)}$$





Distributions

- Define distributions based on the Choquet integral. Why?
  - Non-additive measures on a set X permit us to represent interactions between objects in X !!
    ... similar to covariances but different types of interactions !!

Distributions

Definition:

- $Y = \{Y_1, \ldots, Y_n\}$ random variables; $\mu : 2^Y \rightarrow [0,1]$ a non-additive measure and $\mathbf{m}$ a vector in $\mathbb{R}^n$.
- The exponential family of Choquet integral based class-conditional probability-density functions is defined by:

$$PC_{\mathbf{m},\mu}(\mathbf{x}) = \frac{1}{K} e^{-\frac{1}{2} CI_\mu((\mathbf{x}-\mathbf{m}) \circ (\mathbf{x}-\mathbf{m}))}$$

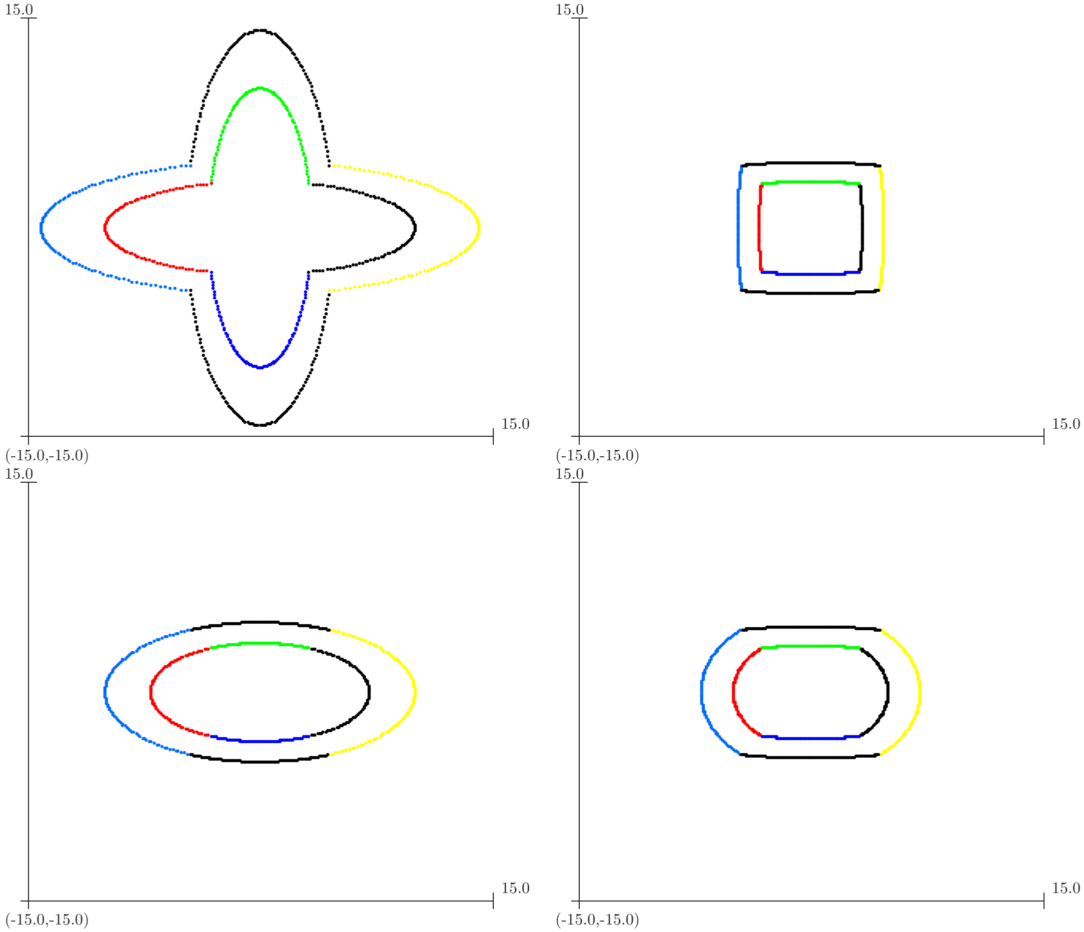
where K is a constant that is defined so that the function is a probability, and where $\mathbf{v} \circ \mathbf{w}$ denotes the Hadamard or Schur (elementwise) product of vectors $\mathbf{v}$ and $\mathbf{w}$ (i.e., $(\mathbf{v} \circ \mathbf{w}) = (v_1 w_1 \ldots v_n w_n)$).

Notation:

- We denote it by $C(\mathbf{m}, \mu)$.

Distributions

- Shapes (level curves)



(a) $\mu_A(\{x\}) = 0.1$ and $\mu_A(\{y\}) = 0.1$, (b) $\mu_B(\{x\}) = 0.9$ and $\mu_B(\{y\}) = 0.9$, (c) $\mu_C(\{x\}) = 0.2$ and $\mu_C(\{y\}) = 0.8$, and (d) $\mu_D(\{x\}) = 0.4$ and $\mu_D(\{y\}) = 0.9$.

Distributions

Property:

- The family of distributions $N(\mathbf{m}, \mathbf{\Sigma})$ in $\mathbb{R}^n$ with a diagonal matrix Σ of rank n, and the family of distributions $C(\mathbf{m}, \mu)$ with an additive measure μ with all $\mu(\{x_i\}) \neq 0$ are equivalent.
  ($\mu(X)$ is not necessarily here 1)

Follows from additivity in μ = probability = diagonal Σ

Distributions

Property:

- The family of distributions $N(\mathbf{m}, \mathbf{\Sigma})$ in $\mathbb{R}^n$ with a diagonal matrix Σ of rank n, and the family of distributions $C(\mathbf{m}, \mu)$ with an additive measure μ with all $\mu(\{x_i\}) \neq 0$ are equivalent.
  ($\mu(X)$ is not necessarily here 1)

Follows from additivity in μ = probability = diagonal Σ

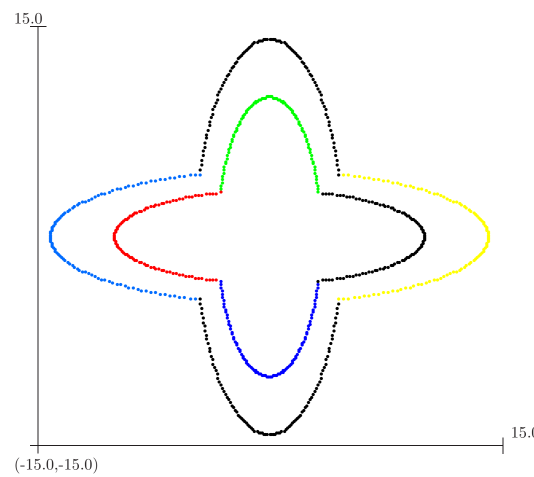
Corollary:

- The distribution $N(\mathbf{0}, \mathbb{I})$ corresponds to $C(\mathbf{0}, \mu^1)$ where μ^1 is the additive measure defined as $\mu^1(A) = |A|$ for all $A \subseteq X$.

Distributions

Properties:

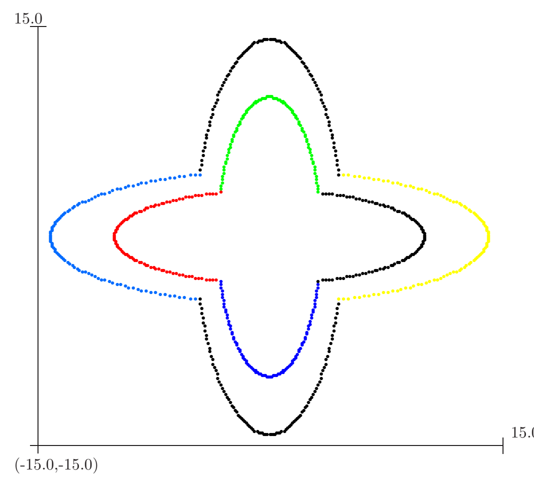
- In general, the two families of distributions $N(\mathbf{m}, \mathbf{\Sigma})$ and $C(\mathbf{m}, \mu)$ are different.
- $C(\mathbf{m}, \mu)$ always symmetric w.r.t. Y_1 and Y_2 axis.



Distributions

Properties:

- In general, the two families of distributions $N(\mathbf{m}, \mathbf{\Sigma})$ and $C(\mathbf{m}, \mu)$ are different.
- $C(\mathbf{m}, \mu)$ always symmetric w.r.t. Y_1 and Y_2 axis.



- Using the CMI distance, we consider both types of interactions
  - Mahalanobis: Σ
  - Choquet (measure): μ

Distributions

Definition:

- $Y = \{Y_1, \ldots, Y_n\}$ random variables, $\mu : 2^Y \to [0,1]$ a measure, $\mathbf{m}$ a vector in $\mathbb{R}^n$, and Q a positive-definite matrix.
- The exponential family of Choquet-Mahalanobis integral based class-conditional probability-density functions is defined by:

$$PCM_{\mathbf{m},\mu,\mathbf{Q}}(x) = \frac{1}{K} e^{-\frac{1}{2} CI_\mu(\mathbf{v} \circ \mathbf{w})}$$

where K is a constant that is defined so that the function is a probability, where $\mathbf{L}\mathbf{L}^T = \mathbf{Q}$ is the Cholesky decomposition of the matrix $\mathbf{Q}$, $\mathbf{v} = (\mathbf{x} - \mathbf{m})^T \mathbf{L}$, $w = \mathbf{L}^T(\mathbf{x} - \mathbf{m})$, and where $\mathbf{v} \circ \mathbf{w}$ denotes the elementwise product of vectors $\mathbf{v}$ and $\mathbf{w}$.

Notation:

- We denote it by $CMI(\mathbf{m}, \mu, \mathbf{Q})$.

Distributions

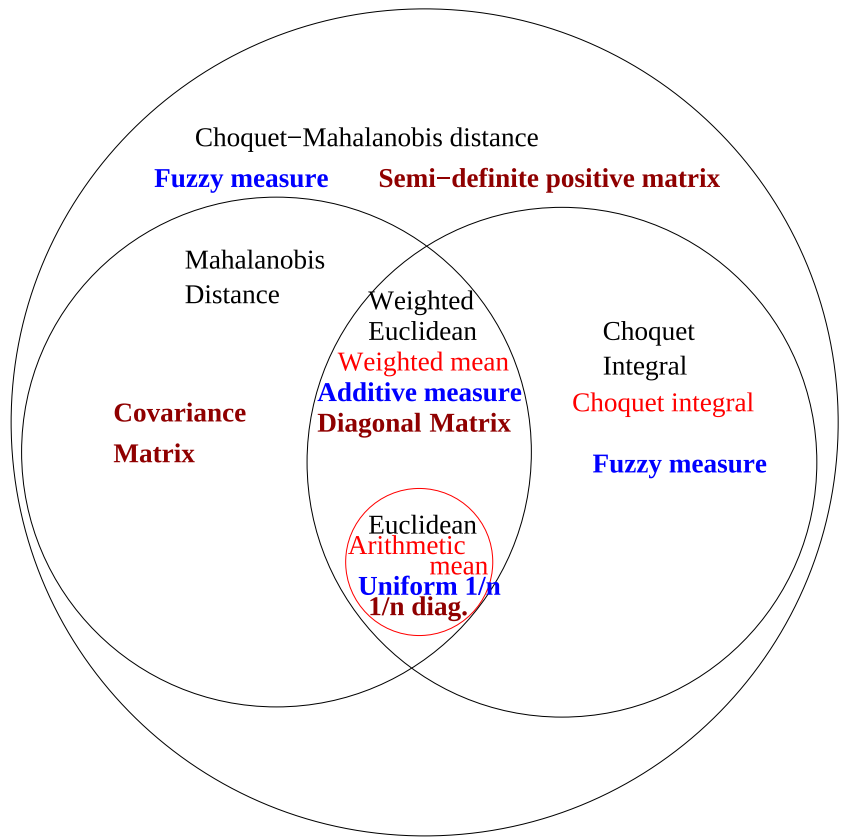
Property:

- The distribution $CMI(\mathbf{m}, \mu, \mathbf{Q})$ generalizes the multivariate normal distributions and the Choquet integral based distribution. In addition
  - A $CMI(\mathbf{m}, \mu, \mathbf{Q})$ with $\mu = \mu^1$ corresponds to multivariate normal distributions,
  - A $CMI(\mathbf{m}, \mu, \mathbf{Q})$ with $Q = \mathbb{I}$ corresponds to a $CI(\mathbf{m}, \mu)$.

Distributions

Graphically:

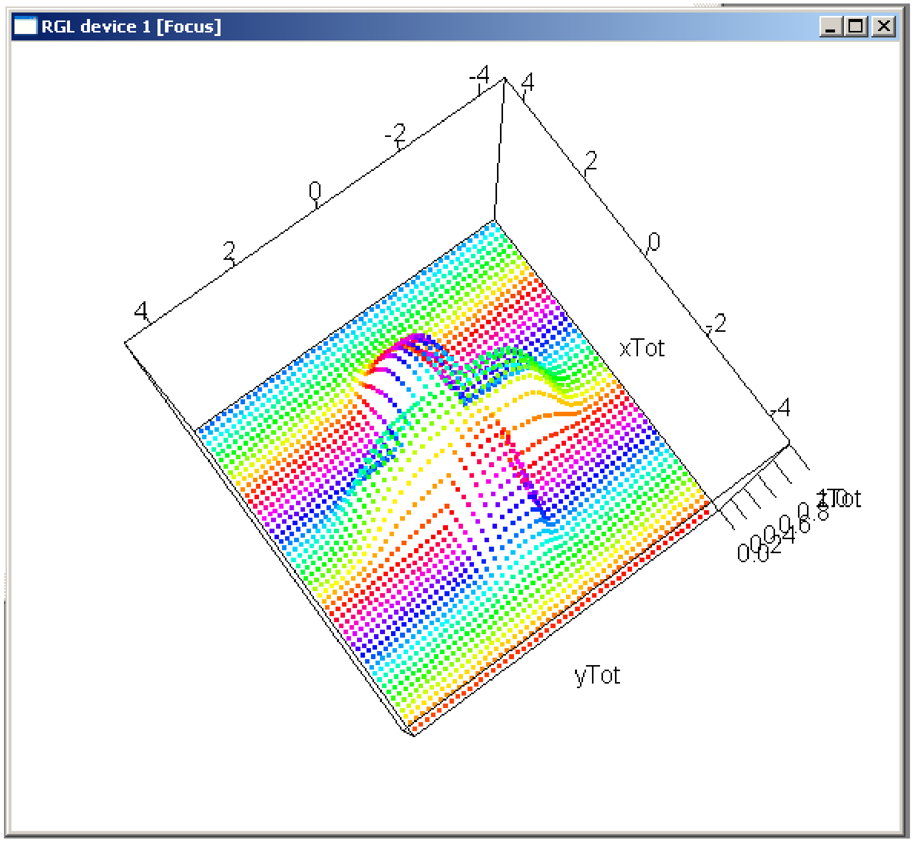
- Choquet integral (CI distribution), Mahalobis distance (multivariate normal distribution), generalization (CMI distribution)

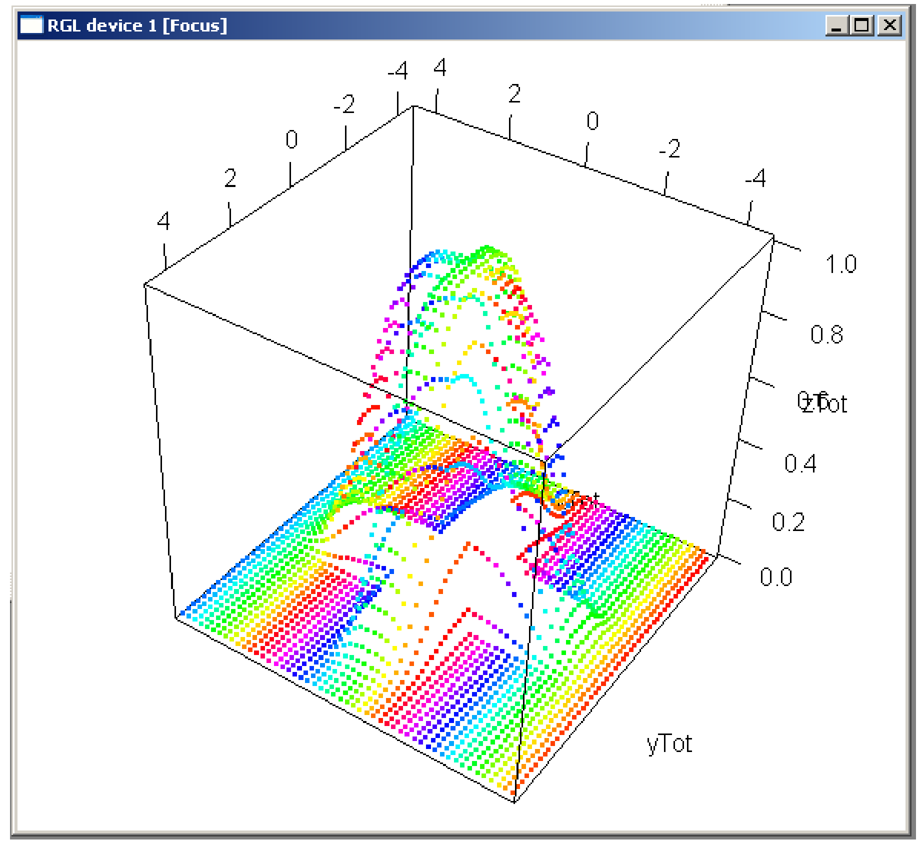


Distributions

1st Example: Interactions only expressed in terms of a measure.

- No correlation exists between the variables.
- CMI with $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho_{12} = 0.0$, $\mu_x = 0.01$, $\mu_y = 0.01$.

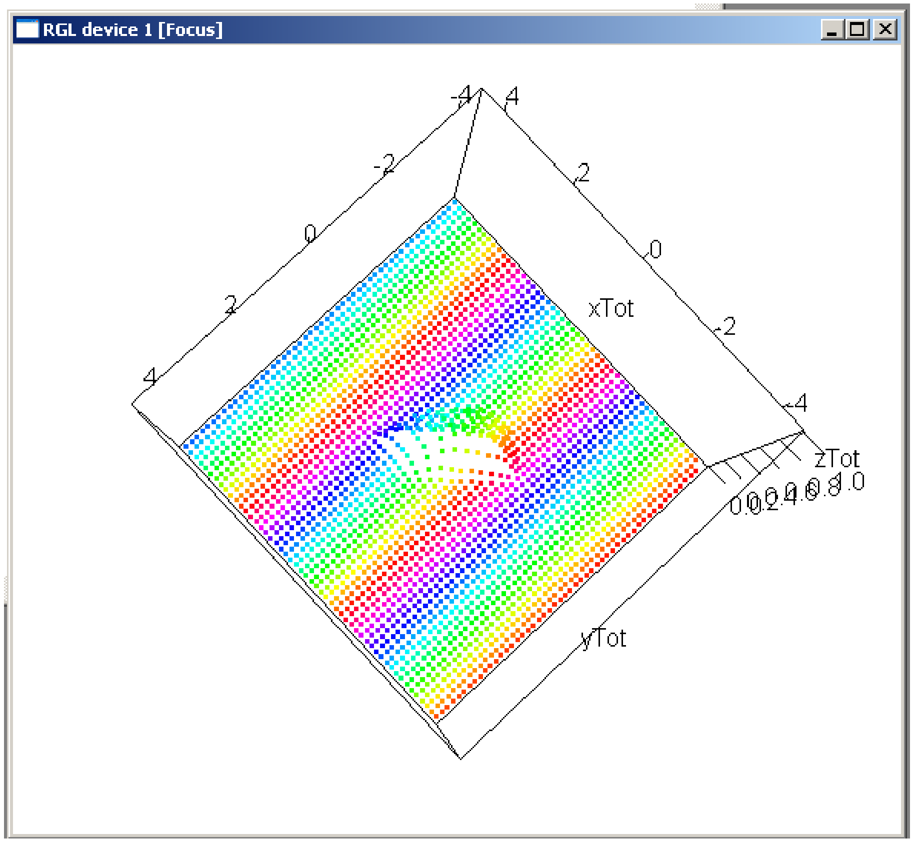


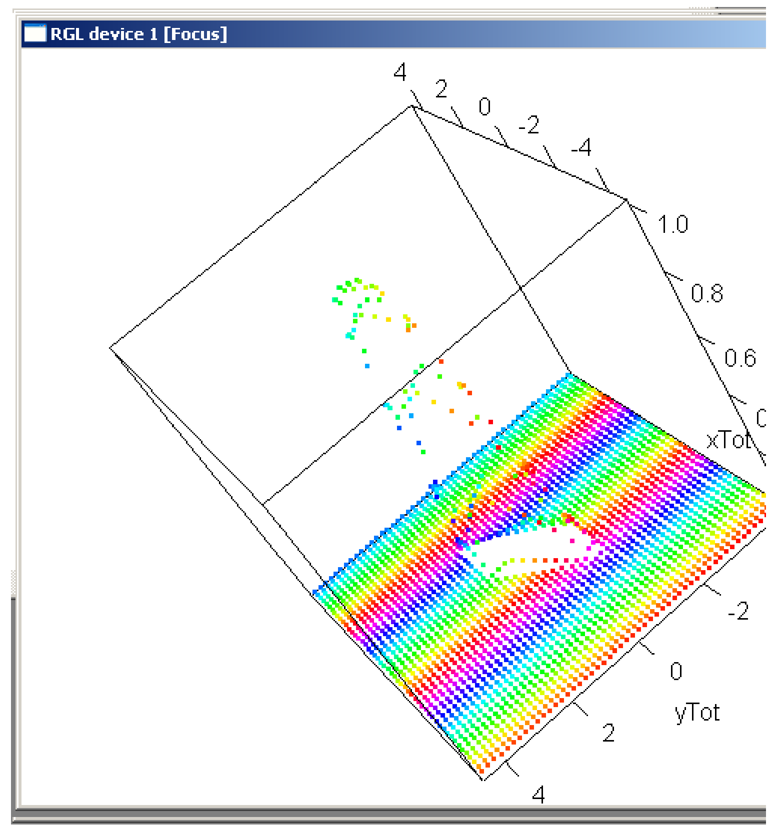


Distributions

2nd Example: Interactions only in terms of a covariance matrix.

- CMI with $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho_{12} = 0.9$, $\mu_x = 0.10$, $\mu_y = 0.90$.

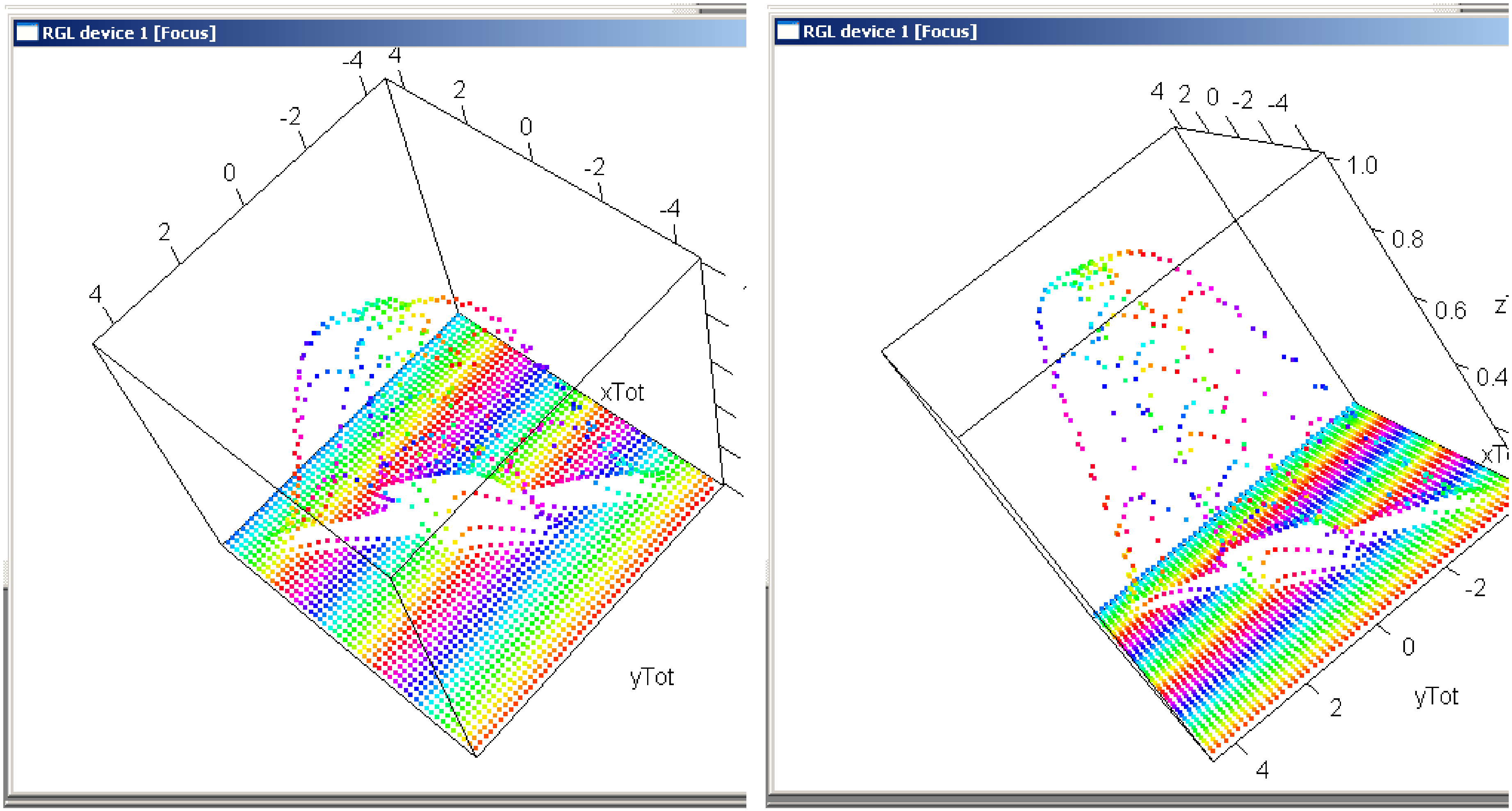




Distributions

3rd Example: Interactions both: covariance matrix and measure.

- CMI with $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho_{12} = 0.9$, $\mu_x = 0.01$, $\mu_y = 0.01$.



Distributions

More properties: Data not always acc. normality assumption

- spherical, elliptical distributions
- They generalize, respectively, $N(\mathbf{0}, \mathbb{I})$ and $N(\mathbf{m}, \Sigma)$

Distributions

More properties: Data not always acc. normality assumption

- ◦ spherical, elliptical distributions
- ◦ They generalize, respectively, $N(\mathbf{0}, \mathbb{I})$ and $N(\mathbf{m}, \Sigma)$
- Neither $CMI(\mathbf{m}, \mu, \mathbf{Q}) \subseteq / \supseteq$ spherical / elliptical distributions.

Distributions

More properties: Data not always acc. normality assumption

- ◦ spherical, elliptical distributions
- ◦ They generalize, respectively, $N(\mathbf{0}, \mathbb{I})$ and $N(\mathbf{m}, \Sigma)$
- Neither $CMI(\mathbf{m}, \mu, \mathbf{Q}) \subseteq / \supseteq$ spherical / elliptical distributions.

Example:

- Non-additive μ: $CMI(\mathbf{m}, \mu, \mathbf{Q})$ not repr. spherical/elliptical

Distributions

More properties: Data not always acc. normality assumption

- ◦ spherical, elliptical distributions
- ◦ They generalize, respectively, $N(\mathbf{0}, \mathbb{I})$ and $N(\mathbf{m}, \Sigma)$
- Neither $CMI(\mathbf{m}, \mu, \mathbf{Q}) \subseteq / \supseteq$ spherical / elliptical distributions.

Example:

- Non-additive μ: $CMI(\mathbf{m}, \mu, \mathbf{Q})$ not repr. spherical/elliptical
- No CMI for the following spherical distribution: Spherical distribution with density

$$f(r) = (1/K)e^{-\left(\frac{r-r_0}{\sigma}\right)^2},$$

where r_0 is a radius over which the density is maximum, σ is a variance, and K is the normalization constant.

Summary

Summary

Summary:

- Choquet integral and non-additive measures for decision and reidentification
- Definition of distances based on the Choquet integral
- Comparison with the Mahalanobis distance
- Construction of distributions
- Relationship with multivariate normal and spherical distributions

Thank you