LinStat 2014

New results on the Choquet integral based distributions

Vicenç Torra
IIIA, Artificial Intelligence Research Institute
Bellaterra, Catalonia, Spain

August, 2014

From November: University of Skövde, Sweden

Overview

Basics and objectives:

- Distribution based on the Choquet integral (for non-additive measures)

Motivation:

- Theory: Mathematical properties
- Methodology: different ways to express interactions
- Application: statistical disclosure control (data privacy)

Outline

1. Preliminaries
2. Choquet integral based distribution
3. Choquet-Mahalanobis based distribution
4. Summary

Preliminaries
Non-additive measures and the Choquet integral

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.

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

- Probability: $\mu(X) = 1$

Definitions: measures

Non-additive measures.

- $(X, \mathcal{A})$ a measurable space, a non-additive 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 measure:

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

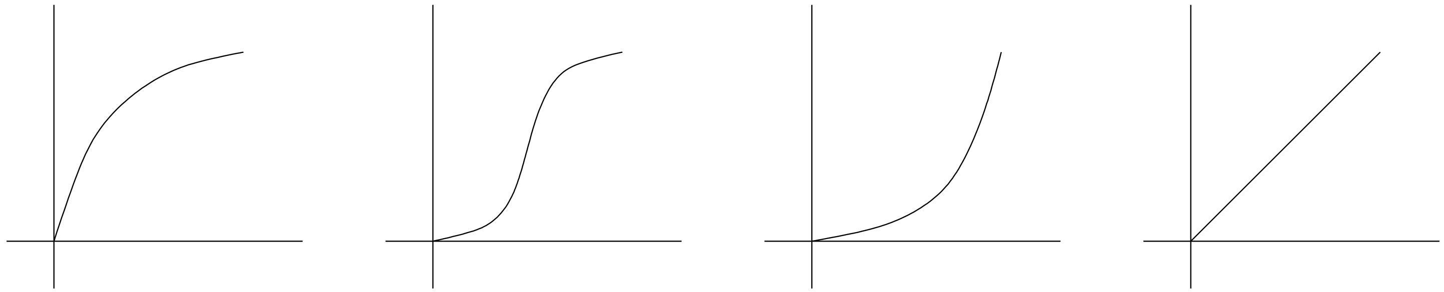
Definitions: measures

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 measure:

$$\mu_m(A) = m(\lambda(A)) \quad (1)$$

- 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 measure:

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

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 measure:

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

Applications.

- To represent interactions

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

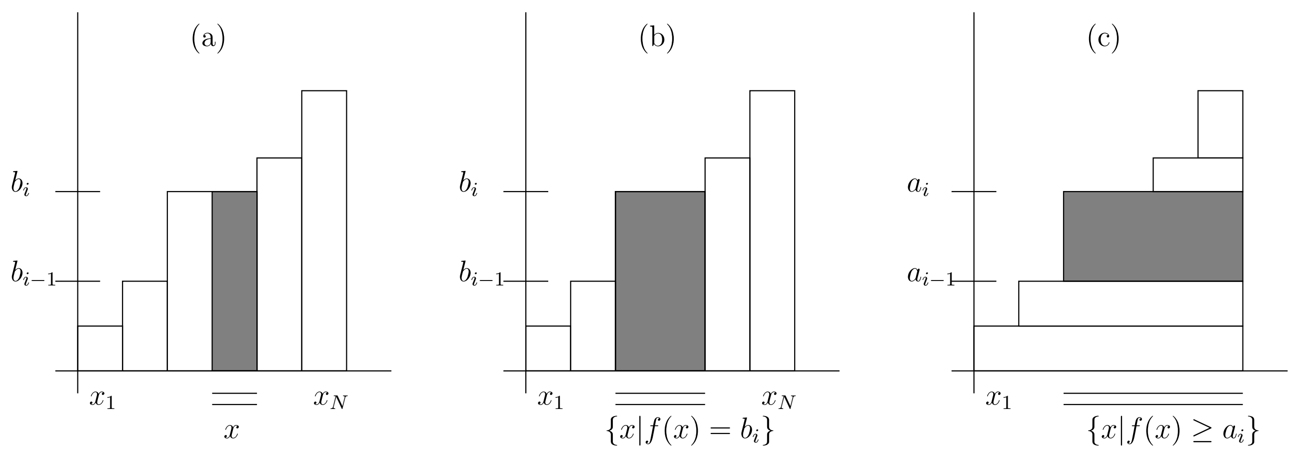
Definitions: integrals

Choquet integral (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 $A_{s(i)} = \{x_{s(i)}, \ldots, x_{s(N)}\}$.

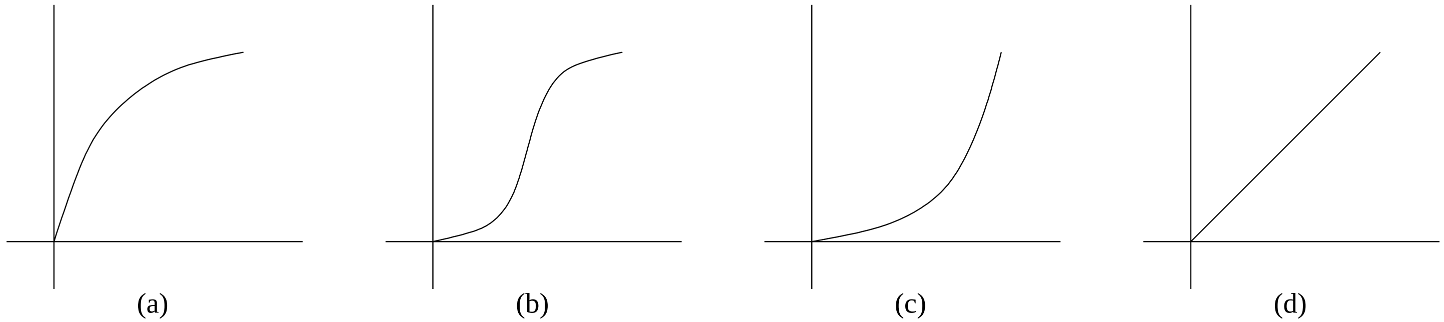
Definitions: measures

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 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*



Choquet integral based distribution

Choquet integral based distribution: Definition

Definition:

- $Y = \{Y_1, \ldots, Y_n\}$ random variables; $\mu : 2^Y \to [0,1]$ 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:

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

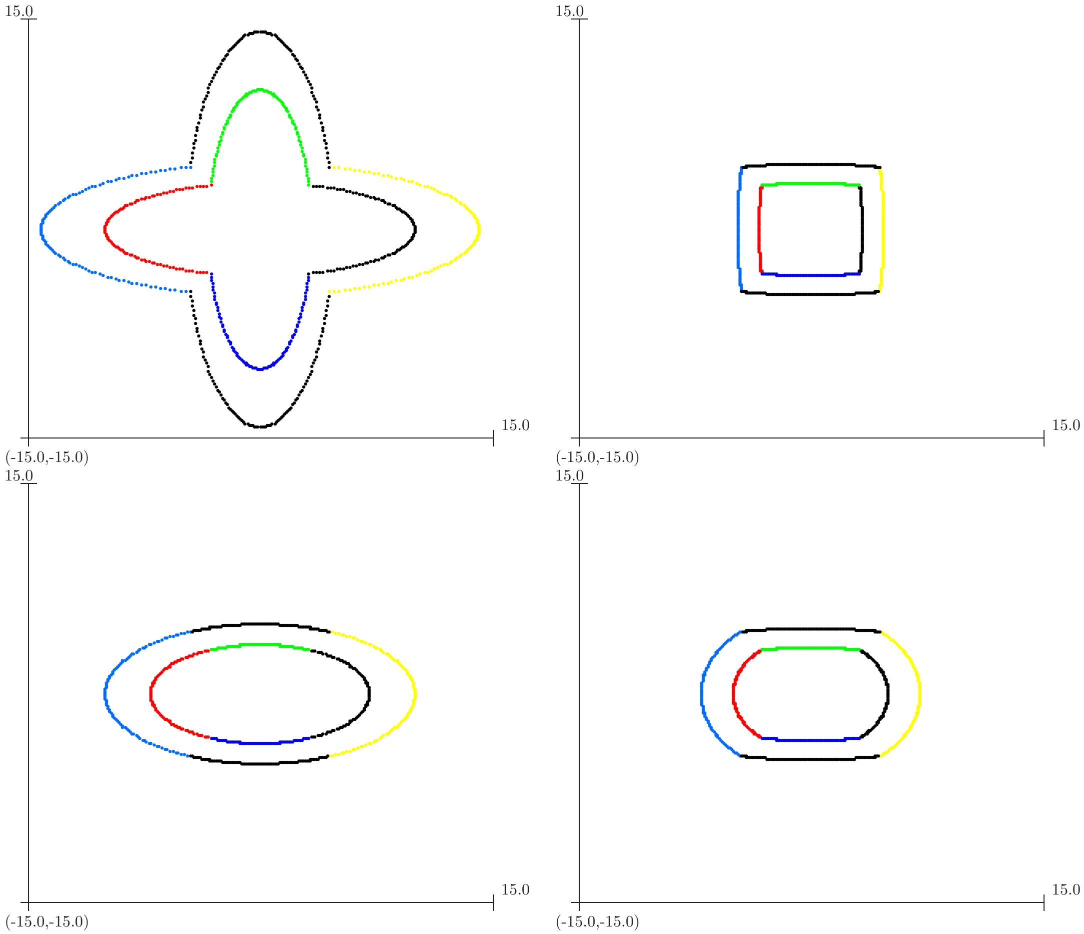
where K is a constant that is defined so that the function is a probability, and where $\mathbf{v} \circ \mathbf{w}$ denotes the Hadamard or Schur (elementwise) product of vectors $\mathbf{v}$ and $\mathbf{w}$ (i.e., $(\mathbf{v} \circ \mathbf{w}) = (v_1 w_1 \ldots v_n w_n)$).

Notation:

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

Choquet integral based distribution: Examples

- 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 $\mu_D(\{y\}) = 0.9$.

Choquet integral based distribution: Properties

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)

Choquet integral based distribution: Properties

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)

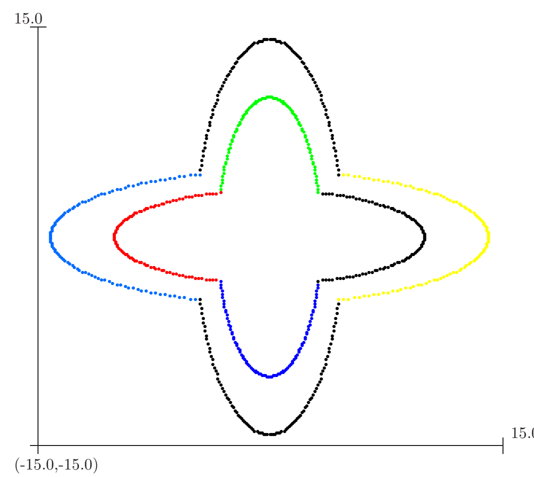
Corollary:

- 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 $A \subseteq X$.

Choquet integral based distribution: N vs. C

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.



- A generalization of both: Choquet-Mahalanobis based distribution.
  - Mahalanobis: Σ represents some interactions
  - Choquet (measure): μ represents some interactions

Choquet-Mahalanobis based distribution

Choquet integral based distribution: Definition

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 $CMI(\mathbf{m}, \mu, \mathbf{Q})$.

Choquet integral based distribution: Properties

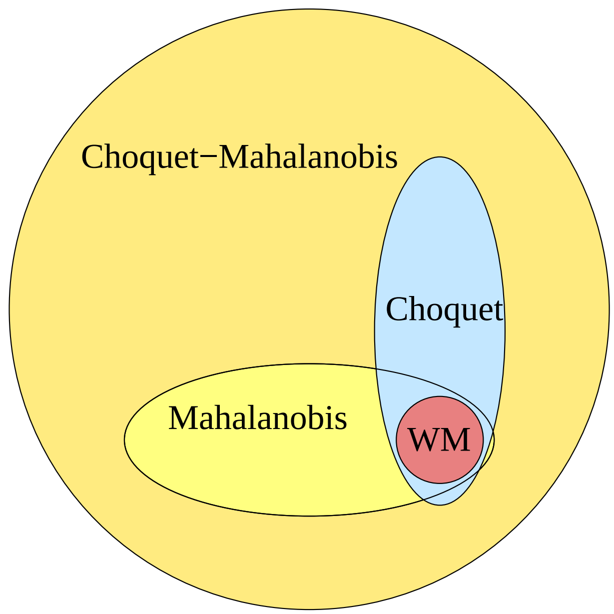
Property:

- 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 $CI(\mathbf{m}, \mu)$.

Choquet integral based distribution: Properties

Graphically:

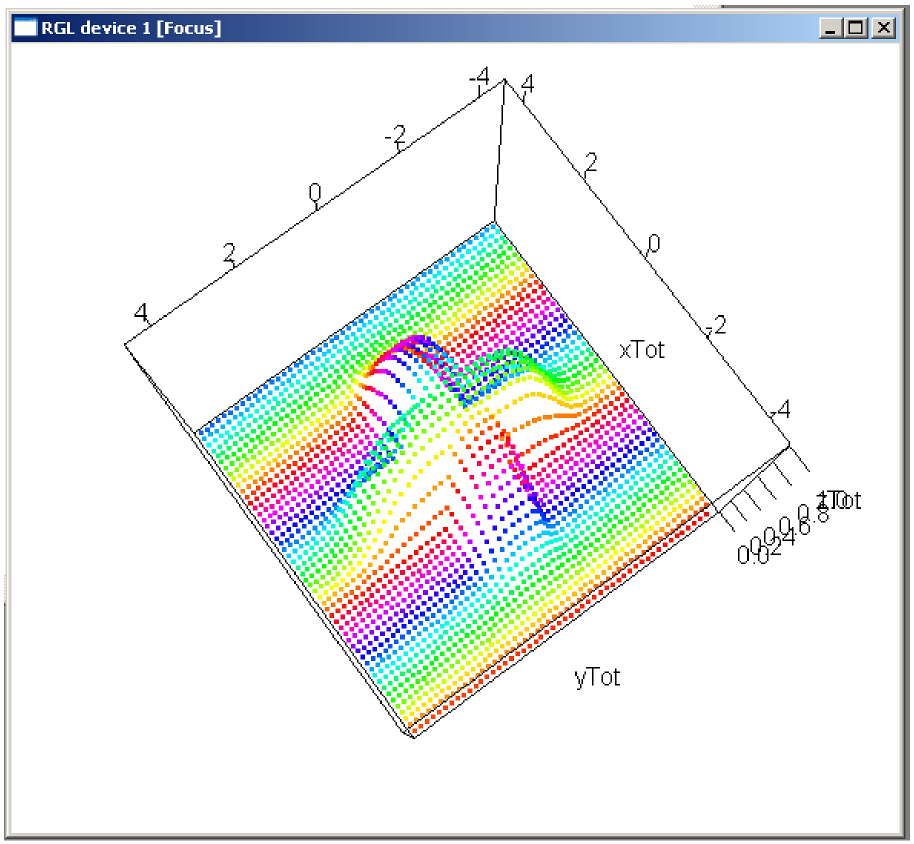
- Choquet-integral (CI distribution) and Mahalobis distance (multivariate normal distribution) and a generalization

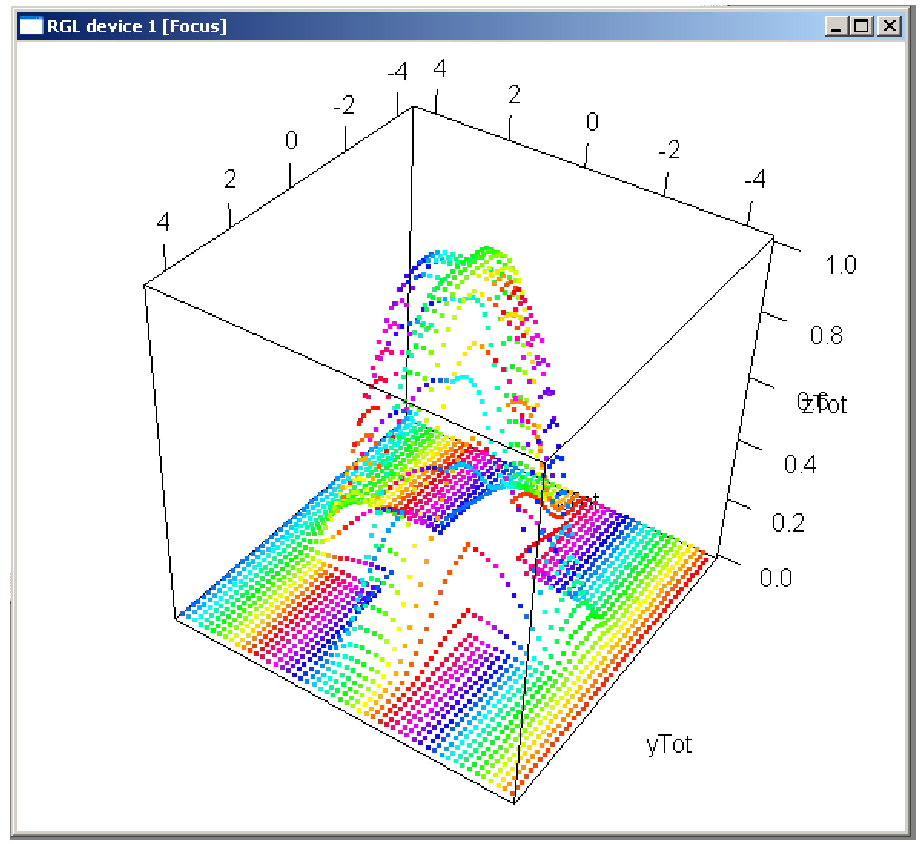


Choquet integral based distribution: Examples

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

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

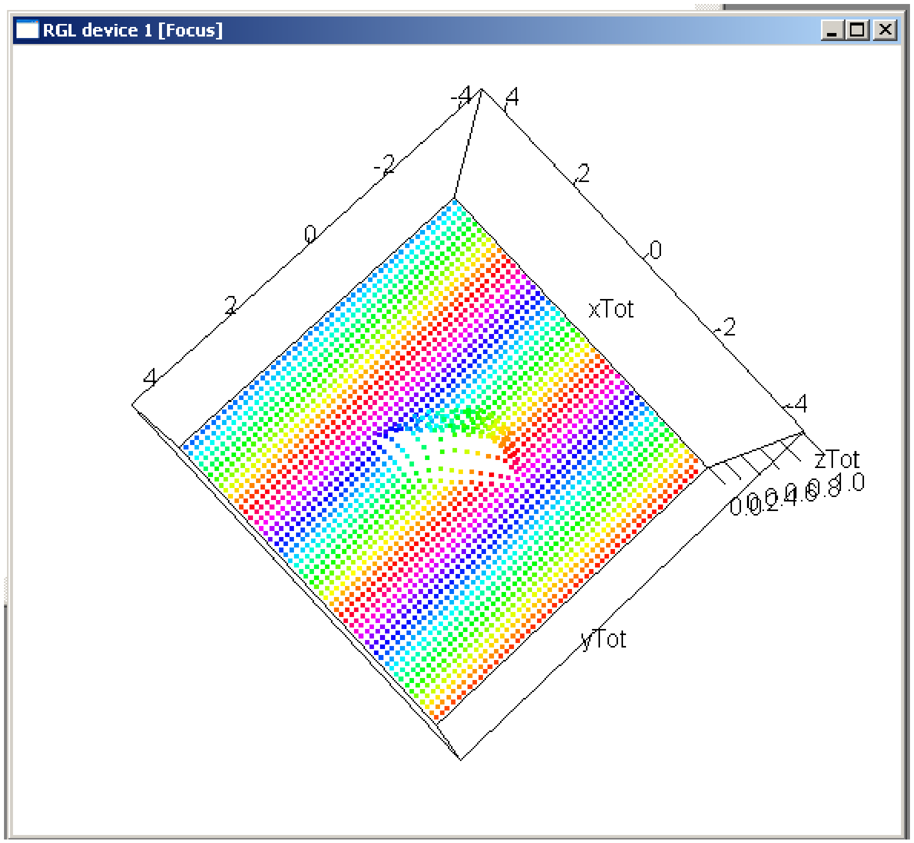


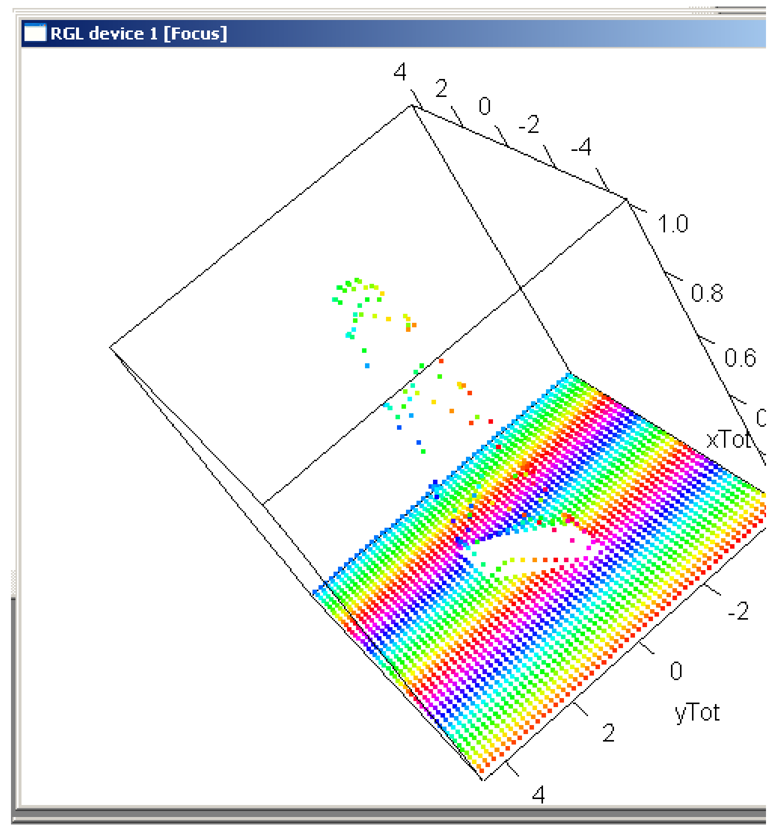


Choquet integral based distribution: Examples

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

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

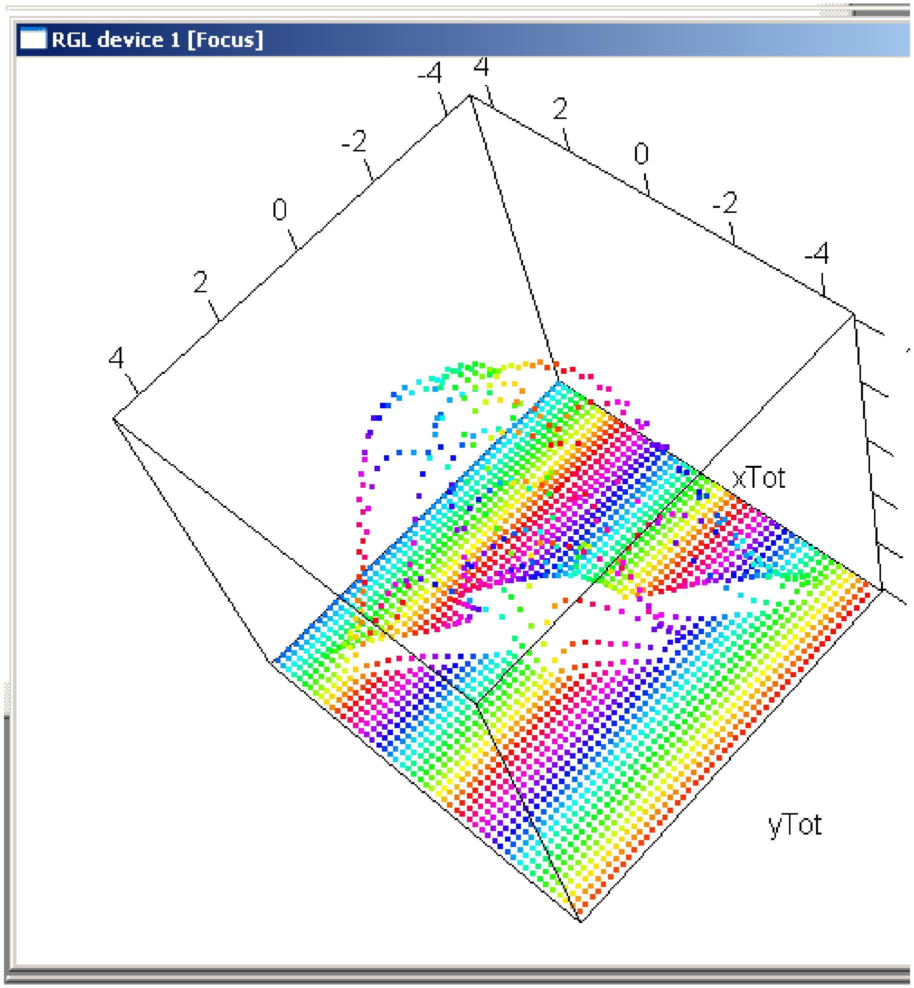


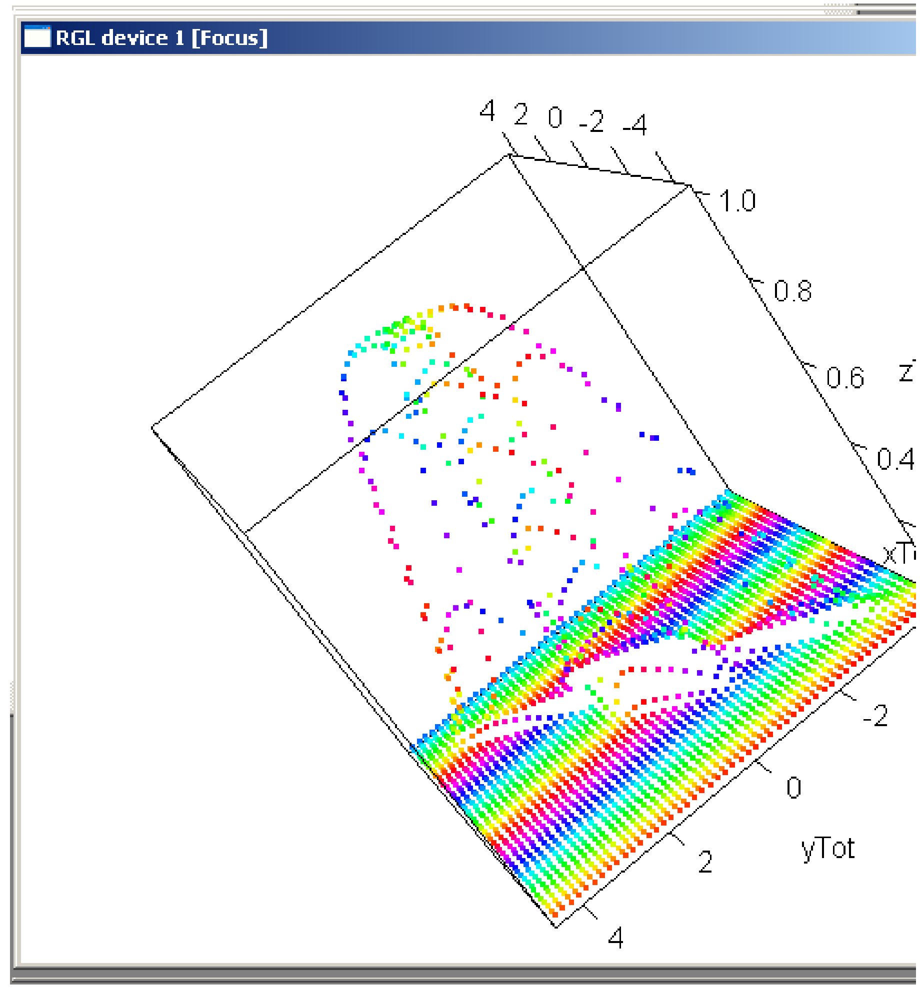


Choquet integral based distribution: Examples

3rd Example: Interactions expressed in both terms: covariance matrix and measure.

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





Choquet integral based distribution: Properties

More properties: (comparison with spherical and elliptical distributions)

- In general, neither $CMI(\mathbf{m}, \mu, \mathbf{Q})$ is more general than spherical / elliptical distributions, nor spherical / elliptical distributions are more general than $CMI(\mathbf{m}, \mu, \mathbf{Q})$.

Example:

- For non-additive measures, $CMI(\mathbf{m}, \mu, \mathbf{Q})$ cannot be expressed as spherical or elliptical distributions.
- The following spherical distribution cannot be represented with CMI: 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.

Choquet integral based distribution: Properties

More properties:

- When $\mathbf{Q}$ is not diagonal, we may have

$$Cov[X_i, X_j] \neq Q(X_i, X_j).$$

Normality test CI-based distribution:

Mardia's test based on skewness and kurtosis

- Skewness test is passed.
- Almost all distributions (in $\mathbb{R}^2$) pass kurtosis test in experiments:
  - Choquet-integral distributions with $\mu(\{x\}) = i/10$ and $\mu(\{y\}) = i/10$ for $i = 1, 2, \ldots, 9$.
    Test only fails in (i) $\mu(\{x\}) = 0.1$ and $\mu(\{y\}) = 0.1$, (ii) $\mu(\{x\}) = 0.2$ and $\mu(\{y\}) = 0.1$.

Summary

Summary

Summary:

- Definition of distributions based on the Choquet integral
  Integral for non-additive measures
- Relationship with multivariate normal and spherical distributions

Thank you