数理解析研究所講究録
第 1906 巻 2014 年 136-143

Distributions based on the Choquet integral and non-additive measures

Vicenç Torra[1]

[1] IIIA - CSIC, Campus UAB s/n, 08193 Bellaterra, Catalonia, Spain
Email: vtorra@iiia.csic.es

Abstract. In recent papers we introduced multivariate probability distributions based on the Choquet integral. In this paper we review some previous results on multivariate distributions, and locate the Choquet based ones in this area.
Keywords: Multivariate distributions

1 Introduction

Non-additive measures [3, 19] are used to represent situations in which there are interactions between variables. They replace the axiom of additivity by the condition of monotonicity. Because of that, they generalize probabilities.

The Choquet integral [2] is one of the existing integrals to integrate a function with respect to a non-additive measure. This integral reduces to the Lebesgue integral when the measure is additive.

Nevertheless, non-additive measure is not the only construct that is used to represent situations in which there are interactions between variables. For example, the multivariate normal distribution also permits to consider interactions between variables. In multivariate normal distributions interactions are expressed in terms of the covariance matrix. In fact, the definition of the multivariate normal distribution can be seen as the definition of a distribution by means of the Mahalanobis distance. The Mahalanobis distance is a weighted distance where the covariance matrix plays the role of the weights.

In a recent paper [17] we introduced new multivariate distributions that generalize the multivariate normal distribution. They are based on the Choquet integral, through the use of a distance that is based on the Choquet integral. In short, the Mahalanobis distance that uses the covariance matrix to express weights (and interactions) of variables is replaced by a Choquet integral based distance that uses a non-additive measure to express weights (and interactions) of variables.

The literature on distributions presents other generalizations of the multivariate normal distribution. The spherical and the spherical distributions [5, 4, 7] are two of them.

In this paper we review some of these definitions, and some relationships between them. The structure of the paper is as follows. In Section 2 we review basic definitions needed later. In Section 3 we review the Choquet integral based distributions as well as some of the results on these distributions. The paper finishes with some conclusions and lines for future work

2 Preliminaries

This section reviews some definitions related to non-additive measures and probability distributions. We begin with some basic definitions that are needed later.

Definition 1. *The Hadamard or Schur product* $\circ$ *of vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *is defined as follows* $(\mathbf{v} \circ \mathbf{w}) = (v_1 w_1 \dots v_n w_n)$

Lemma 1. *Let* u, v *two arbitrary vectors in* $\mathbb{R}^n$, *and let* $\mathbf{a}_n$ *the vector in* $\mathbb{R}^n$ *defined by* $\mathbf{a}_n = (a, \dots, a)$. *When* $\mathbf{a}_n = (1, \dots, 1)$ *we will use* $\mathbf{1}_n$. *If* n *is clear from the context we will use only* $\mathbf{a}$. *Then,*

$$(\mathbf{u} \circ \mathbf{v})\mathbf{a} = a(\mathbf{uv}).$$

Proof. Note that,

$$(\mathbf{u} \circ \mathbf{v})\mathbf{a} = (u_1 v_1, \dots, u_n v_n)'(a, \dots, a) = a \sum_i u_i v_i = a(\mathbf{uv}).$$

□

Corollary 1. *For two arbitrary vectors in* $\mathbb{R}^n$ u, v *the following holds*

$$(\mathbf{u} \circ \mathbf{v})\mathbf{1} = (\mathbf{uv})$$

2.1 Non-additive measures and the Choquet integral

We begin defining non-additive measures, which generalize probabilities.

Definition 2. *Let* $(\Omega, \mathcal{F})$ *be a measurable space. A set function* μ *defined on* $\mathcal{F}$ *is called a non-additive measure if an only if*

- $0 \leq \mu(A) \leq \infty$ *for any* $A \in \mathcal{F}$;
- $\mu(\emptyset) = 0$;
- *If* $A_1 \subseteq A_2 \subseteq \mathcal{F}$ *then*
$$\mu(A_1) \leq \mu(A_2)$$

It is often required that $\mu(\Omega) = 1$. If the measure is additive and $\mu(\Omega) = 1$ then we have that μ is a probability distribution. In this work we do not presume this condition. In fact, some results, as Lemma 3, do not follow if μ is bounded by one.

We will use μ^1 to denote the additive measure $\mu(A) = |A|$ for all $A \subseteq \Omega$. That is, μ^1 is an additive measure where the *weight* of each element in A is one.

We will use $d(\mu)$ to denote the vector $(\mu(\{x_1\}), \dots, \mu(\{x_1\})$. Naturally, $d(\mu^1) = \mathbf{1}$.

Given a non-additive measure and a function f, the Choquet integral [2] of f with respect to μ is defined as follows.

Definition 3. *[2] Let μ be a non-additive measure on X; then, the* Choquet integral *of a function $f : X \to \mathbb{R}^+$ with respect to the non-additive measure μ is defined by*

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

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

We will also use the notation $CI_\mu(\mathbf{y})$ for a given vector $\mathbf{y}$. In this case, we understand that $f(x_i) = y_i$.

The Choquet integral corresponds to the Lebesgue integral when the measure is additive. Note also that the following holds.

Lemma 2. *When the measure is additive, $CI_\mu(\mathbf{x}) = d(\mu)\mathbf{x}$.*

In this paper we will use the following distance, which is based on the Choquet integral. We will denote by dCI^2 the square of the distance. Only when the measure μ is submodular (i.e., $\mu(A)+\mu(B) \geq \mu(A\cup B)+\mu(A\cap B)$), we have that dCI is, properly speaking, a distance (i.e., it satisfies the triangular inequality). This definition is used in [12, 1].

Definition 4. *Given two vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ and a non-additive measure on the set $\{1, \ldots, n\}$ we define*

$$dCI_\mu^2(\mathbf{a}, \mathbf{b}) = CI_\mu((a_1 - b_1)^2, \ldots, (a_n - b_n)^2).$$

2.2 Multivariate probability distributions

Multivariate normal distribution are defined in terms of the probability density function as follows:

$$PM_{\mathbf{m},\Sigma}(\mathbf{x}) = \frac{1}{(2\pi)^{m/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})'\Sigma^{-1}(\mathbf{x}-\mathbf{m})}.$$

However, as pointed out in [4], there are different ways to extend the normal univariate distribution to the multivariate case. Doing so through the definition of the density function is only one of such approaches. [4] considers the following four extensions.

- Extension by means of density functions.
- Extension by characteristic functions.
- Extension by means of a linear combination.
- Extension by means of stochastic decomposition. If $x \sim N(m, \sigma)$, x can be expressed by $x = m + \sigma y$ where $y \sim N(0, 1)$. So, if $\mathbf{y} = (y_1, \ldots, y_p)'$ have independent identically distributed components and $y_i \sim N(0, 1)$, an extension is to consider $\mathbf{x} = \mathbf{m} + \mathbf{A}\mathbf{y}$ where $\mathbf{x}, \mathbf{m} \in \mathbb{R}^n$ and $\mathbf{A}$ is a $n \times p$ matrix.

Real data does not always satisfy the assumption of normality. One of the existing lines of research introduces new distributions that extend and diverge from the normality model. This line studies elliptical distributions. This type of distributions have two advantages [7]. First, the elliptical distribution generalizes the multivariate normal distribution (as well as other well known distributions). Second, many results for multivariate normal distributions also hold for elliptical distributions.

Following, [7] and [4] we review elliptical distributions starting from spherical distributions.

Definition 5. *[4] A n-random vector* $\mathbf{x}$ *is said to have a spherical distribution if for every orthogonal* $n \times n$ *matrix* Γ, $\mathbf{\Gamma x} \stackrel{d}{=} \mathbf{x}$

Here $\stackrel{d}{=}$ is the operator that denotes that the two variables have the same distribution.

A spherical distributions is an extension of the multivariate standard normal distribution $N(\mathbf{0}, \mathbf{I}_n)$.

Definition 6. *[4] A n-random vector* $\mathbf{x}$ *is said to have an elliptical distribution with parameter* $\mathbf{m}$ *(an n vector) and* Σ *(an* $n \times n$ *matrix) if*

$$x \stackrel{d}{=} \mathbf{A}'\mathbf{y}$$

where $\mathbf{y}$ *has a spherical distribution,* $\mathbf{A}$ *is a* $p \times k$ *matrix and the matrix* $\mathbf{V}$ *defined by* $\mathbf{V} = \mathbf{A}\mathbf{A}'$ *has rank* k *(i.e.,* $rank(\mathbf{V}) = k$*).*

An elliptical distribution is an extension of the multivariate normal distribution $N(\mathbf{m}, \mathbf{\Sigma})$.

3 Extensions of probability distributions using the Choquet integral

In this section we review different families of distributions we have introduced based on the Choquet integral. Using the classification in [4] reviewed in Section 2.2, these new extensions were introduced by means of density functions.

Definition 7. *[17] Let* $Y = \{Y_1, \ldots, Y_n\}$ *be a set of random variables describing data on a* $\mathbb{R}^n$ *dimensional space. Let* $\mu : 2^Y \to [0,1]$ *be a non-additive measure and* $\mathbf{m}$ *a vector in* $\mathbb{R}^n$.

Then, 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., $(\mathbf{v} \circ \mathbf{w}) = (v_1 w_1 \dots v_n w_n)$).

The constant K is defined so that P is a density function. Therefore, the value of K is such that,

$$\int_{\mathbf{x} \in X} PC_{\mathbf{m},\mu}(\mathbf{x}) = 1,$$

so,

$$K = \int_{\mathbf{x} \in X} e^{-\frac{1}{2} CI_{\mu}((\mathbf{x}-\mathbf{m})\circ(\mathbf{x}-\mathbf{m}))}.$$

We will denote by $C(\mathbf{m}, \mu)$ a distribution of this form.

In [17], this distribution was introduced for a classification problem. In such type of problems, the exact value of K is not needed.

Lemma 3. *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.*

Proof. To proof this, let us first consider $N(\mathbf{m}, \mathbf{\Sigma})$ with a diagonal matrix Σ. Then, if we define the additive measure $\mu(\{x_i\}) = 1/(\sigma_i^2)$ (and, therefore, $\mu(A) = \sum_{a \in A} \mu(\{a\})$ for all $A \subseteq X$ such that $|A| > 1$), we have that $C(\mathbf{m}, \mu)(\mathbf{x}) = N(\mathbf{m}, \mathbf{\Sigma})(\mathbf{x})$ for all vector $\mathbf{x}$.

Similarly, if we consider a distribution $C(\mathbf{m}, \mu)$ with an additive measure μ, then we define Σ as the diagonal matrix with $\sigma_i^2 = 1/\mu(\{x_i\})$. □

Corollary 2. *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 $A \subseteq X$.*

In general, the two families of distributions $N(\mathbf{m}, \mathbf{\Sigma})$ and $C(\mathbf{m}, \mu)$ are different. The following family of distributions was introduced to make a generalization of both $N(\mathbf{m}, \mathbf{\Sigma})$ and $C(\mathbf{m}, \mu)$.

Definition 8. *[17] Let $Y = \{Y_1, \dots, Y_n\}$ be a set of random variables describing data on a $\mathbb{R}^n$ dimensional space. Let $\mu : 2^Y \to [0, 1]$ be a fuzzy measure, $\mathbf{m}$ be a vector in $\mathbb{R}^n$, and Q a positive-definite matrix.*

Then, 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 $\mathbf{w}$.

We will denote by $CMI(\mathbf{m}, \mu, \mathbf{Q})$ a distribution of this form.

Proposition 1. *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 a multivariate normal distributions,*
- *A $CMI(\mathbf{m}, \mu, \mathbf{Q})$ with $Q = \mathbb{I}$ corresponds to a $CI(\mathbf{m}, \mu)$.*

Proof. We first proof the equality between $CMI(\mathbf{m}, \mu^1, \mathbf{Q})$ and multivariate normal distributions.

First note that when $\mu = \mu^1$ we have

$$(\mathbf{x} - \mathbf{m})^T \mathbf{Q}(\mathbf{x} - \mathbf{m}) = (\mathbf{x} - \mathbf{m})\mathbf{T}\mathbf{L}\mathbf{L}^T(\mathbf{x} - \mathbf{m}).$$

Now, as matrix multiplication is associative we have that this expression is equivalent to $((\mathbf{x} - \mathbf{m})^T\mathbf{L})(\mathbf{L}^T(\mathbf{x} - \mathbf{m}))$ which using Lemma 1 corresponds to

$$(((\mathbf{x} - \mathbf{m})^T\mathbf{L}) \circ (\mathbf{L}^T(\mathbf{x} - \mathbf{m})))\mathbf{1}.$$

Using Lemma 2 in this expression, we have that the last expression is equivalent to $CI_{\mu^1}(((\mathbf{x} - \mathbf{m})^T\mathbf{L}) \circ (\mathbf{L}^T(\mathbf{x} - \mathbf{m})))$. From this follows the equality of the two distributions $(PM_{\mathbf{m},\mathbf{Q}}(\mathbf{x}) = PCM_{\mathbf{m},\mu,\mathbf{Q}}(\mathbf{x}))$ and also that for any $N(\mathbf{m}, \mathbf{Q})$ we have $CMI(\mathbf{m}, \mu^1, \mathbf{Q})$ that has the same distribution.

Now we prove the equality between $CMI(\mathbf{m}, \mu, \mathbb{I})$ and the Choquet integral based distributions.

If $\mathbf{Q} = \mathbb{I}$, then $\mathbf{L} = \mathbf{L}^T = \mathbb{I}_n$, therefore, $\mathbf{v} = (\mathbf{x} - \mathbf{m})^T\mathbf{L} = (\mathbf{x} - \mathbf{m})$ and $\mathbf{w} = \mathbf{L}^T(\mathbf{x} - \mathbf{m}) = (\mathbf{x} - \mathbf{m})$ and

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

From this equations we have that for any distribution $C(m, \mu)$ we have $CMI(\mathbf{m}, \mu, \mathbb{I})$ with the same distribution. □

In general, the family $CMI(\mathbf{m}, \mu, \mathbf{Q})$ is different than the spherical and the elliptical distributions. For non-additive measures, $CMI(\mathbf{m}, \mu, \mathbf{Q})$ cannot be expressed as spherical or elliptical distributions. On the contrary, a spherical distribution does not need to have its maximum in the mean vector. Consider for example the 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. As distributions $CMI(\mathbf{m}, \mu, \mathbf{Q})$ have its maximum in the mean vector $\mathbf{m}$, they cannot represent this type of spherical distribution.

Lemma 4. *In general, neither $CMI(\mathbf{m}, \mu, \mathbf{Q})$ is more general than spherical / elliptical distributions, nor spherical / elliptical distributions are more general than $CMI(\mathbf{m}, \mu, \mathbf{Q})$.*

In [18] a more general definition of Choquet integral based distributions was introduced. It is based on the generalized mesures defined by Greco et al. in [6]. That is, in the level dependent measures.

Definition 9. *Let $\{Y_1, \ldots, Y_n\}$ be a set of random variables describing data on a $\mathbb{R}^n$ dimensional space. Let $(\alpha, \beta) \subseteq \mathbb{R}$, and $\mu^G : 2^Y \times (\alpha, \beta) \to [0, 1]$ be a generalized non-additive measure.*

Let $\mathbf{m} \in \mathbb{R}^n$ be the mean of the random variables; then, the exponential family of level dependent Choquet integral based class-conditional probability-density functions is defined by:

$$P(\mathbf{x}) = \frac{1}{K} e^{-\frac{1}{2} CI^G_{\mu^G}((\mathbf{x}-\mathbf{m}) \circ (\mathbf{x}-\mathbf{m}))}$$

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

4 Summary and future work

In this paper we have reviewed our definitions and results on distributions based on the Choquet integral. We have shown some relationship between these distributions and some other multivariate distributions. Future work on this topic includes the identification of the parameters of the Choquet integral based distributions and the study of tests of hypothesis.

Acknowledgements

Partial support by the Spanish MEC (projects ARES – CONSOLIDER INGENIO 2010 CSD2007-00004, and co-privacy – TIN2011-27076-C03-03) is acknowledged.

References

1. Abril, D., Navarro-Arribas, G., Torra, V. (2012) Choquet Integral for Record Linkage, Annals of Operations Research 195 97-110.
2. Choquet, G. (1953/54) Theory of capacities, Ann. Inst. Fourier, 5, 131-295.
3. Denneberg, D. (1994) Non Additive Measure and Integral, Kluwer Academic Publishers.
4. Fang, K.-T., Kotz, S., Ng, K.-W. (1990) Symmetric multivariate and related distributions, Chapman and Hall.

5. Fang, K.-T., Zhang, Y.-T. (1980) Generalized multivariate analysis, Springer-Verlag, Science Press.
6. Greco, S., Matarazzo, B., Giove, S. (2011) The Choquet integral with respect to a level dependent capacity, Fuzzy Sets and Systems 175 1-35.
7. Kollo, T., von Rosen, D. (2005) Advanced multivariate statistics with matrices, Springer.
8. Keener, R. W. (2010) Theoretical Statistics, Springer.
9. Keller, J. M., Gader, P. D., Hocaoglu, A. K. (2000) Fuzzy Integrals in Image Processing and Recognition, in M. Grabisch, T. Murofushi, M. Sugeno, Fuzzy Measures and Integrals, Springer-Verlag, 435-466.
10. Narukawa, Y., Murofushi, T. (2008) Choquet Stieltjes integral as a tool for decision modeling, Int. J. of Intel. Syst. 23 115-127.
11. Ozaki, H. (2013) Integral with respect to Non-additive Measure in Economics, in V. Torra, Y. Narukawa, M. Sugeno (eds.), Non-additive measures: theory and applications, Springer.
12. Pham, T. D., Yan, H. (1999) Color image segmentation using fuzzy integral and mountain clustering, Fuzzy Sets and Systems 107 121-130.
13. Schmeidler, D. (1989) Subjective probability and expected utility without additivity, Econometrica 57 517-587.
14. Shao, J. (2010) Mathematical Statistics, Springer.
15. Sugeno, M. (2013) A note on derivatives of functions with respect to fuzzy measures, Fuzzy Sets and Systems 222 1-17.
16. Torra, V. (2011) Expressing interactions: Mahalanobis and Choquet operators, Proc. AGOP 2011.
17. Torra, V., Narukawa, Y. (2012) On a comparison between Mahalanobis distance and Choquet integral: The Choquet-Mahalanobis operator, Information Sciences 190 56-63.
18. Torra, V., Narukawa, Y. (2013) Exponential family of Level dependent Choquet integral based class-conditional probability functions, Proc. AGOP 2013.
19. Torra, V., Narukawa, Y., Sugeno, M. (eds.) (2013) Non-additive measures: theory and applications, Springer.