ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 19, Number 1, June 2015
Available online at http://acutm.math.ut.ee

.

Some properties of Choquet integral based probability functions

Vicenç Torra

Abstract. The Choquet integral permits us to integrate a function with respect to a non-additive measure. When the measure is additive it corresponds to the Lebesgue integral. This integral was used recently to define families of probability-density functions. They are the exponential family of Choquet integral (CI) based class-conditional probability-density functions, and the exponential family of Choquet–Mahalanobis integral (CMI) based class-conditional probability-density functions. The latter being a generalization of the former, and also a generalization of the normal distribution.

In this paper we study some properties of these distributions, and study the application of a few normality tests.

1. Introduction

New families of probability distributions based on the Choquet integral were recently introduced in [9], and further studied in [8].

The Choquet integral [1], introduced by Choquet in 1954, permits us to integrate a function with respect to a non-additive measure. Non-additive measures are also known by capacities and fuzzy measures. The Choquet integral has been applied successfully in decision making and in artificial intelligence. One of its advantages is that it permits to express interactions between variables (e.g., criteria in decision making or information sources in artificial intelligence) by means of the measure.

In short, the new probability distributions are defined replacing the Mahalanobis distance by *distances* based on the Choquet integral. Both the Mahalanobis distance and the Choquet integral based distance permit

Received November 24, 2014.

2010 *Mathematics Subject Classification.* 28E05, 28E10, 60A99.

Key words and phrases. Choquet integral, non-additive measures, Choquet integral based probability functions, normality tests.

http://dx.doi.org/10.12697/ACUTM.2015.19.04

us to take into account some interactions between the variables. However, the types of interactions expressed are different and expressed in a different way. In the Mahalanobis distance the interactions are represented in terms of a matrix (the covariance matrix) while in the Choquet integral distance the interactions are represented by means of a non-additive measure.

Another distribution was defined to encompass both types of interactions. That is, the one expressed in terms of the (covariance) matrix and the one expressed in terms of the non-additive measure.

In this paper we review these definitions, we study some of their properties and study whether data from these distributions pass some normality tests.

The structure of the paper is as follows. In Section 2, we review some preliminaries needed in the rest of the paper. In Section 3, we discuss the new distribitions, and in Section 4 we present our new results. The paper finishes with some conclusions.

2. Preliminaries

In this section we review the Choquet integral we need later on in this work. The section begins with the notation. From a formal point of view, the Choquet integral integrates a function with respect to a non-additive measure. Note that this is not the only integral for non-additive measures. The Sugeno integral [7] is another example. All these integrals are known as fuzzy integrals. See, e.g., [11] for details.

Let $Y = \{y_1, \dots, y_n\}$ be a set and let 2^Y represent the power set of Y. If $f : Y \to \mathbb{R}^+$ is a function, then $f(y_i) \in \mathbb{R}^+$. Fuzzy integrals aggregate the values $f(y_i)$ with respect to a fuzzy measure μ.

Definition 1. Let $Y = \{y_1, \dots, y_n\}$ be a set. Then, a set function $\mu : 2^Y \to [0, \infty)$ is a fuzzy measure (a non-additive measure) if it satisfies the following axioms:

(i) $\mu(\emptyset) = 0$ (boundary conditions),
(ii) $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ (monotonicity).

Definition 2. Let μ be a fuzzy measure on Y. Then, the *Choquet integral* of a function $f : Y \to \mathbb{R}^+$ with respect to the fuzzy measure μ is defined by

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

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

We also use the notation $CI_\mu(a_1, \dots, a_n)$ to express the Choquet integral of $a_i := f(y_i)$.

3. Choquet integral based distributions

We defined in [9] two distributions based on the Choquet integral. We review below the first one. The other generalizes this one as well as the normal distribution.

Definition 3. Let $Y = \{Y_1, \ldots, Y_n\}$ be a set of random variables describing data on a n-dimensional space. Let $\mu : 2^Y \to [0, 1]$ be a fuzzy measure and m a vector in $\mathbb{R}^n$.

Then, the exponential family of Choquet integral based class-conditional probability-density functions is defined for $x \in \mathbb{R}^n$ by

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

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

Although the definition of the density function needs the constant K, the exact value of K is not relevant in classification problems, or for studying the shape of the distribution function. In any case, the K is the value such that

$$\int_{x \in \mathbb{R}^n} P(x) dx = 1.$$

So, K should be defined by

$$K = \int_{x \in \mathbb{R}^n} e^{-\frac{1}{2} CI_\mu((x-m)\circ(x-m))} dx.$$

In the experiments reported below we will use approximations of K computed through numerical integration.

Definition 4. Let $Y = \{Y_1, \ldots, Y_n\}$ be a set of random variables describing data on a n-dimensional space. Let $\mu : 2^Y \to [0, 1]$ be a fuzzy measure, m be a vector in $\mathbb{R}^n$, and Q $n \times n$ a positive-definite matrix.

Then, the exponential family of Choquet–Mahalanobis integral based class-conditional probability-density functions is defined for $x \in \mathbb{R}^n$ by

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

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

The Choquet–Mahalanobis integral (see [9]) of $(x - m)$, with respect to μ and Q, corresponds to $CI_\mu(v \circ w)$ as used in this definition.

This distribution generalizes the multivariate normal distribution. See [8] for a proof; [8] also compares these distributions with spherical and elliptical distributions.

4. Some basic properties of the Choquet integral based probability distributions

In this section we study some properties of the probability distributions based on the Choquet integral. Some of them generalize the ones presented whithout proof in [10].

Lemma 5. *Let $P(x)$ with $x \in \mathbb{R}^n$ be a Choquet integral based distribution according to Definition 3 defined in terms of a mean $m = (m_1, \ldots, m_n)$ and a fuzzy measure μ. Then, for all $x \in \mathbb{R}^n$ and all $i \in \{1, \ldots, n\}$,*

$$P(x_1, \ldots, x_{i-1}, x_i{+}m_i, x_{i+1}, \ldots, x_n){=}P(x_1, \ldots, x_{i-1}, -x_i{+}m_i, x_{i+1}, \ldots, x_n).$$

Proof. Without loss of generality, let us consider the case of $i = 1$. Then, it is easy to see that

$$\begin{aligned} P(m_1 + x_1, y_1, \ldots, y_{n-1}) &= \frac{1}{K} e^{-\frac{1}{2} CI_\mu((m_1+x-m_1)^2, y_1^2, \ldots, y_{n-1}^2)} \\ &= \frac{1}{K} e^{-\frac{1}{2} CI_\mu((m_1-x-m_1)^2, y_1^2, \ldots, y_{n-1}^2)} \\ &= P(m_1 - x_1, y_1, \ldots, y_{n-1}). \end{aligned}$$

□

Lemma 6. *Let A be a positive-definite diagonal matrix with diagonal elements $(a_1, \ldots, a_n)$. Then, its Cholesky decomposition is a diagonal matrix with diagonal elements $(\sqrt{a_1}, \ldots, \sqrt{a_n})$.*

This is easy to see from the definition of the Cholesky decomposition (see, e.g., [6], p. 335). Recall that the Cholesky decomposition of a positive-definite matrix A is the product LL^T.

Lemma 7. *Let $P(x)$ with $x \in \mathbb{R}^n$ be a Choquet–Mahalanobis integral based distribution according to Definition 4 defined in terms of a mean $m = (m_1, \ldots, m_n)$, a positive-definite diagonal matrix Q, and a fuzzy measure μ. Then, for all $x \in \mathbb{R}^n$ and all $i \in \{1, \ldots, n\}$,*

$$P(x_1, \ldots, x_{i-1}, x_i{+}m_i, x_{i+1}, \ldots, x_n){=}P(x_1, \ldots, x_{i-1}, -x_i{+}m_i, x_{i+1}, \ldots, x_n). \tag{1}$$

Proof. Let $Q = LL^T$ be the Cholesky decomposition of Q. Then, given $x \in \mathbb{R}^n$, we define $v = (x - m)^T L$ and $w = L^T (x - m)$. From Lemma 6, L is diagonal and, therefore, $L^T = L$ and $v = w^T$. Thus, $v \circ w = (v_1^2, \ldots, v_n^2)$.

Now, without loss of generality, we prove equation (1) for $i = 1$. Let $v' = (x_1 + m_1, x_2, \ldots, x_n) - m = (x_1, x_2 - m_2, \ldots, x_n - m_n)$, and let $v'' =$

$(-x_1+m_1, x_2, \ldots, x_n)-m = (-x_1, x_2-m_2, \ldots, x_n-m_n)$. As $v' \circ v' = v'' \circ v''$, $CMI_{\mu,Q}(v' \circ v') = CMI_{\mu,Q'}(v'' \circ v'')$, and the lemma is proved. □

Using these lemmas, we can prove the following propositions.

Proposition 8. *Let $P(x)$ with $x \in \mathbb{R}^n$ be an exponential Choquet integral probability-density function with mean $m = (m_1, \ldots, m_n)$. Then, for any fuzzy measure μ, the mean vector $\bar{X} = [E[X_1], E[X_2], \ldots, E[X_n]]$ is m and $\Sigma = [Cov[X_i, X_j]]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n$ is zero for all $i \neq j$ and, thus, diagonal.*

Proof. This distribution is

$$\bar{X} = \int_{x \in \mathbb{R}^n} xP(x)dx,$$

and for the ith component of the vector, we have

$$\bar{X}_i = \int_{x \in \mathbb{R}^n} x_i P(x)dx.$$

Without loss of generality, we consider the case of $i = 1$. Given any $y = (y_1, \ldots, y_{n-1}) \in \mathbb{R}^{n-1}$ and $x_1 \in \mathbb{R}$, we denote the vector $(x_1, y_1, \ldots, y_{n-1})$ by $(x_1|y)$. Let us apply Fubini's theorem to the expression above for $\bar{X}_1$:

$$\begin{aligned}
\bar{X}_1 &= \int_{x_1 \in \mathbb{R}} \int_{y \in \mathbb{R}^{n-1}} x_1 P((x_1|y))dydx_1 \\
&= \int_{x_1 \in \mathbb{R}^+} \int_{y \in \mathbb{R}^{n-1}} (x_1+m_1)P((x_1+m_1|y))dydx_1 \\
&\quad + \int_{x_1 \in \mathbb{R}^+} \int_{y \in \mathbb{R}^{n-1}} (-x_1+m_1)P((-x_1+m_1|y))dydx_1.
\end{aligned}$$

As Lemma 5 implies that $P((-x_1+m_1|y)) = P((x_1+m_1|y))$, we have

$$\begin{aligned}
\bar{X}_1 &= \int_{x_1 \in \mathbb{R}^+} \int_{y \in \mathbb{R}^{n-1}} (x_1 - x_1)P((m_1+x_1|y))dydx_1 \\
&\quad + \int_{x_1 \in \mathbb{R}} \int_{y \in \mathbb{R}^{n-1}} m_1 P((m_1+x_1|y))dydx_1 \\
&= 0 + m_1 \cdot 1 = m_1.
\end{aligned}$$

Without loss of generality, we consider the covariance of variables X_1 and X_2. Their covariance is

$$Cov(X_1, X_2) = \int_{x_1 \in \mathbb{R}} \int_{x_2 \in \mathbb{R}} P(x_1, x_2)(x_1 - m_1)(x_2 - m_2)dx_2dx_1.$$

Here

$$P(x_1, x_2) = \int_{x \in \mathbb{R}^{n-2}} P((x_1, x_2|x))dx,$$

and then,

$$
\begin{aligned}
&Cov(X_1, X_2) = \\
&= \int_{x_1\in\mathbb{R}^+}\int_{x_2\in\mathbb{R}^+} P(x_1+m_1, x_2+m_2)(x_1+m_1-m_1)(x_2+m_2-m_2)dx_2dx_1 \\
&+ \int_{x_1\in\mathbb{R}^+}\int_{x_2\in\mathbb{R}^+} P(-x_1+m_1, -x_2+m_2)(-x_1+m_1-m_1)(-x_2+m_2-m_2)dx_2dx_1 \\
&+ \int_{x_1\in\mathbb{R}^+}\int_{x_2\in\mathbb{R}^+} P(-x_1+m_1, x_2+m_2)(-x_1+m_1-m_1)(x_2+m_2-m_2)dx_2dx_1 \\
&+ \int_{x_1\in\mathbb{R}^+}\int_{x_2\in\mathbb{R}^+} P(x_1+m_1, -x_2+m_2)(x_1+m_1-m_1)(-x_2+m_2-m_2)dx_2dx_1.
\end{aligned}
$$

Using Lemma 5, we have

$$
\begin{aligned}
P(x_1+m_1, x_2+m_2) &= P(x_1+m_1, -x_2+m_2) = P(-x_1+m_1, x_2+m_2) \\
&= P(-x_1+m_1, -x_2+m_2).
\end{aligned}
$$

It follows that $Cov(X_1, X_2)$ equals to

$$
\int\limits_{x_1\in\mathbb{R}^+}\int\limits_{x_2\in\mathbb{R}^+} P(x_1+m_1, x_2+m_2)\,(x_1x_2+(-x_1)(-x_2)
$$

$$
+(-x_1)x_2+x_1(-x_2))\,dx_2dx_1 = 0.
$$

So, the proposition is proved. □

Proposition 9. *Let $P(x)$ with $x \in \mathbb{R}^n$ be an exponential Choquet–Mahalanobis integral probability-density function with mean $m = (m_1, \ldots, m_n)$. Then, for any fuzzy measure μ and any diagonal matrix Q, the mean vector*

$$[E[X_1], E[X_2], \ldots, E[X_n]]$$

is m (i.e., $E[X_i] = m_i$) and $\Sigma = [Cov[X_i, X_j]]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n$ is zero for all $i \neq j$ and thus, diagonal.

Proof. This proof is similar to the one of Proposition 8. The main changes is that we need now Lemma 7 instead of Lemma 5 used in Proposition 8. □

In fact, the property given above of the mean of the distribution follows also from the fact that odd-order moments of distributions symmetric with respect to zero are zero.

In the case when Σ is not diagonal and, thus, $\Sigma(X_i, X_j) \neq 0$ for $i \neq j$, we might have $Cov[X_i, X_j] \neq 0$. It is important to note that it is not at all required that $Cov[X_i, X_j] = \Sigma(X_i, X_j)$. The following example illustrates this fact.

Example 10. Let us consider the Choquet–Mahalanobis integral based distribution with a fuzzy measure $\mu(\emptyset) = 0$, $\mu(\{x\}) = 0.5$, $\mu(\{y\}) = 0.2$, $\mu(\{x, y\}) = 1$ and the matrix

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}.$$

The covariance matrix of this distribution is

$$\Sigma = \begin{pmatrix} 0.9548251 & 0.9262923 \\ 0.9262923 & 1.0293333 \end{pmatrix}.$$

The correlation coefficient between the two variables is 0.9343469.

4.1. Normality tests. There are several approaches [5] to check whether a distribution follows a multivariate normal distribution. One approach is to study the normality of its marginals. Another is to study directly the multivariate distribution. Mardia's test [4] is an example of the latter.

We have considered both approaches for bivariate Choquet integral based distributions. In particular, we have considered distributions on $X = \{x_1, x_2\}$ based on the Choquet integral with measures defined so that $\mu(\{x_1\}) = i/10$ and $\mu(\{x_2\}) = j/10$ for $i, j \in \{1, 2, 3, \ldots, 9\}$.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.973	0.987	0.969	0.990	0.989	0.987	0.988	0.996	0.987
0.2	0.985	0.975	0.985	0.982	0.994	0.987	0.995	0.994	0.994
0.3	0.990	0.977	0.992	0.984	0.980	0.988	0.983	0.986	0.981
0.4	0.968	0.988	0.988	0.991	0.993	0.991	0.985	0.987	0.985
0.5	0.988	0.991	0.978	0.979	0.961	0.980	0.994	0.980	0.982
0.6	0.979	0.992	0.992	0.991	0.988	0.985	0.993	0.994	0.988
0.7	0.988	0.989	0.987	0.995	0.995	0.990	0.986	0.988	0.984
0.8	0.984	0.996	0.948	0.990	0.986	0.986	0.989	0.990	0.982
0.9	0.990	0.976	0.978	0.988	0.985	0.988	0.993	0.982	0.978

TABLE 1. Values W of Shapiro–Wilk statistic for Choquet integral based distributions with $\mu(\{x\}) = i/10$ and $\mu(\{y\}) = j/10$ for $i, j = 1, 2, \ldots, 9$. Position (a, b) in the table corresponds to the case $\mu(\{x\}) = a$ and $\mu(\{y\}) = b$.

Without loss of generality we have considered the marginals of these distributions on the variable x_2. For each of the marginals we have generated a sample with $n = 100$ data, and studied its normality using the Shapiro–Wilk test. The W value for each sample is included in Table 1. The p-values obtained for these samples are given in Table 2. The null hypothesis is rejected when the p-value is ≤ 0.05. That is, only for p-values lower than 0.05 we conclude that the sample is not normal.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.037	0.420	0.018	0.673	0.567	0.435	0.479	0.991	0.460
0.2	0.320	0.059	0.319	0.177	0.932	0.444	0.976	0.918	0.898
0.3	0.680	0.071	0.836	0.251	0.150	0.512	0.236	0.417	0.178
0.4	0.016	0.576	0.493	0.778	0.901	0.764	0.316	0.460	0.296
0.5	0.482	0.744	0.087	0.111	0.005	0.127	0.928	0.122	0.209
0.6	0.104	0.808	0.774	0.751	0.490	0.294	0.894	0.948	0.532
0.7	0.536	0.602	0.446	0.977	0.964	0.650	0.350	0.516	0.288
0.8	0.227	0.012	0.212	0.481	0.405	0.056	0.134	0.478	0.014
0.8	0.247	0.993	0.00061	0.694	0.360	0.358	0.595	0.655	0.181
0.9	0.653	0.064	0.082	0.496	0.329	0.483	0.877	0.186	0.091

TABLE 2. p-Values of Shapiro–Wilk statistic for Choquet integral based distributions with $\mu(\{x\}) = i/10$ and $\mu(\{y\}) = j/10$ for $i, j = 1, 2, \ldots, 9$. Position (a, b) in the table corresponds to the case $\mu(\{x\}) = a$ and $\mu(\{y\}) = b$.

In order to compute the test, we need (i) to determine the marginal and (ii) to construct the sample from the marginal. All calculations have been done using the statistical software R by means of numerical approximations. More particularly, the marginal has been computed using the function `integrate` integrating the density function for x_1 in the interval $[-20, 20]$. We selected this interval because results had enough accuracy. Then, given the marginal $f(x_2)$ the sampling has been constructed using the inverse of the marginal. That is, given the marginal density function f, we consider its cumulative function $F(x) = \int_{-\infty}^{x} f(y)dy$ and its inverse, the quantile function, which given a value r in $[0, 1]$ returns x_r such that $r = F(x_r)$. Note that given a value r from a uniform distribution in $[0, 1]$, x_r follows the distribution described by f. Function F was computed using `integrate` and the value x_r using the function `uniroot` (package *stats*). The construction of each sample with $n = 100$ required 15–25 minutes in a standard PC.

Table 2 shows that for the most of the cases, the samples pass the Shapiro–Wilk test with p-value larger than 0.05. In fact, it results that the sample with the second lowest p-value is precisely the one with $\mu(\{x_1\}) = \mu(\{x_2\}) = 0.5$, and another that fails is $\mu(\{x_1\}) = 0.8$ and $\mu(\{x_2\}) = 0.2$, and precisely both correspond to normal distributions. Recall that all measures with $\mu(\{x_1\}) + \mu(\{x_2\}) = 1$ lead to normal bivariate distributions.

For the case $\mu(\{x_1\}) = 0.8$ we have computed two pairs of samples for each $\mu(\{x_2\})$. The p-values of the two samples are included in the table, and results show that there is a large variation in the value.

The samples with a p-value lower than 0.05 are: (i) $\mu(\{x_1\}) = 0.1$ and $\mu(\{x_2\}) = 0.3$, (ii) $\mu(\{x_1\}) = 0.4$ and $\mu(\{x_2\}) = 0.1$, (iii) $\mu(\{x_1\}) = 0.5$ and

$\mu(\{x_2\}) = 0.5$, (iv) $\mu(\{x_1\}) = 0.8$ and $\mu(\{x_2\}) = 0.2$, (v) $\mu(\{x_1\}) = 0.8$ and $\mu(\{x_2\}) = 0.9$, and (vi) $\mu(\{x_1\}) = 0.8$ and $\mu(\{x_2\}) = 0.3$.

The results show that this test in the most of the cases fail to detect that the distribution is not normal.

To further illustrate the marginal distributions of the bivariate Choquet integral based distributions, we display in Figure 1 some of these marginals together with the normal distribution with the same variance. The marginal is on the left hand side of the figure and the normal distribution on the right hand side. We can see that the two distributions are slightly different and the marginal is more peaked than the normal distribution. This is because the distributions displayed use low values of μ. In the case of large μ the reversal is true. Figure 2 illustrates this case with the marginal of the distribution with $\mu(\{x_1\}) = \mu(\{x_2\}) = 0.9$ and the normal distribution with the same variance. The variances of the distributions have been computed numerically from the marginal (defined in terms of the numerical integration of the bivariate distribution) using the `integrate` function in R.

We have also considered Mardia's test for the bivariate distributions. Mardia's test is based on multivariate extensions of skewness and kurtosis. In particular, for the multivariate skewness of a sample in a k-dimensional space Mardia obtained the expression

$$b_{1,k} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} [(x_i - \bar{x})' \hat{\Sigma} (x_j - \bar{x})]^3.$$

In the case of the multivariate kurtosis, the expression obtained is

$$b_{2,k} = \frac{1}{n} \sum_{i=1}^{n} [(x_i - \bar{x})' \hat{\Sigma} (x_i - \bar{x})]^2.$$

Here, $\bar{x}$ is the sample mean vector and $\hat{\Sigma}$ is the covariance matrix. They correspond to

$$\bar{x} = (1/n) \sum_{i=1}^{n} x_i, \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T.$$

Then, when the distribution is a multivariate normal distribution (i.e., when the null hypothesis holds), the expression

$$A = n \cdot b_{1,k}/6$$

follows a chi-squared distribution with $k(k+1)(k+2)/6$ degrees of freedom, and the expression

$$B = \sqrt{\frac{n}{8k(k+2)}}\,(b_{2,k} - k(k+2))$$

follows a standard normal random variable $N(0, 1)$.

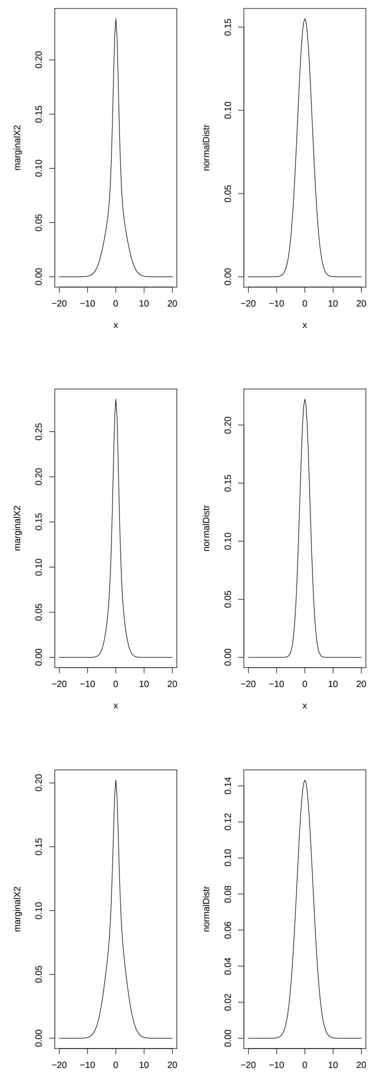

FIGURE 1. Marginals of the bivariate Choquet integral based distributions, and the normal distribution with the same variance. Cases (i) $\mu(\{x_1\}) = 0.1$ and $\mu(\{x_2\}) = 0.1$; (ii) $\mu(\{x_1\}) = 0.1$ and $\mu(\{x_2\}) = 0.2$; and (iii) $\mu(\{x_1\}) = 0.2$ and $\mu(\{x_2\}) = 0.1$.

Our experiments show that most of the distributions pass the normality test. The statistic A based on $b_{1,k}$ (skewness) is about zero for all these

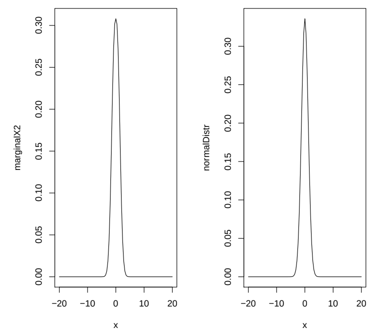

FIGURE 2. Marginal of the bivariate Choquet integral based distributions with $\mu(\{x_1\}) = 0.9$ and $\mu(\{x_2\}) = 0.9$ (left) and the normal distribution with the same variance (right).

distributions and, thus, passes the skewness test. This is natural, as skewness is a measure of asymmetry. Being all almost zero, we do not include the exact values in this work. According to what has been stated above, the statistic A follows a chi-squared distribution with $k(k+1)(k+2)/6 = 4$ degrees of freedom which means that at a confidence level of 0.95 should be less than 9.487729.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	3.54	2.61	1.55	7.71e-01	2.44e-01	-8.23e-02	-0.26	-0.35	-0.37
0.2	2.61	1.86	1.14	6.44e-01	3.16e-01	1.20e-01	0.02	-0.01	0.02
0.3	1.55	1.14	0.65	3.27e-01	1.29e-01	2.82e-02	-0.00	0.03	0.09
0.4	0.77	0.64	0.33	1.31e-01	2.88e-02	-3.00e-07	0.03	0.09	0.20
0.5	0.24	0.32	0.13	2.88e-02	-3.45e-09	2.85e-02	0.10	0.21	0.34
0.6	-0.08	0.12	0.03	-3.00e-07	2.85e-02	1.03e-01	0.21	0.35	0.51
0.7	-0.26	0.02	-0.00	2.78e-02	1.02e-01	2.14e-01	0.36	0.52	0.71
0.8	-0.35	-0.01	0.03	9.83e-02	2.10e-01	3.53e-01	0.52	0.71	0.92
0.9	-0.37	0.02	0.09	2.00e-01	3.42e-01	5.13e-01	0.71	0.92	1.14

TABLE 3. Values of Mardia's B test for Choquet-integral based distributions with $\mu(\{x\}) = i/10$ and $\mu(\{y\}) = i/10$ for $i = 1, 2, \ldots, 9$. Position (a, b) in the table corresponds to the case $\mu(\{x\}) = a$ and $\mu(\{y\}) = b$. The matrix is symmetric.

The values obtained for the statistic B based on $b_{2,k}$ (kurtosis) are presented in Table 3. According to the results of Mardia, the statistic should follow a $N(0,1)$, and with a confidence level of 0.95 values should be lower than 1.959964. The table is symmetric as the distribution is probability distribution is symmetric with respect to the axis x and y. From the data in the table we can see that the test fails only on the pairs (i) $\mu(\{x\}) = 0.1$ and $\mu(\{y\}) = 0.1$, (ii) $\mu(\{x\}) = 0.2$ and $\mu(\{y\}) = 0.1$.

5. Conclusions

In this paper we have reviewed the probability distribution based on the Choquet integral, and we have studied some of their properties. We have presented some results about the means and covariances of these distributions. Then, we have studied some normality tests for bivariate distributions. We have seen that the Shapiro–Wilt for samples from marginal distributions does not permit us to detect that the distribution is not normal. We have also considered Mardia's test. In this case, only the most extreme Choquet integral based distributions have been detected. In particular, two distributions do not pass the normality test. They are the distribution built with $\mu(\{x_1\}) = 0.1$ and $\mu(\{x_2\}) = 0.1$ and the distribution built with $\mu(\{x_1\}) = 0.2$ and $\mu(\{x_2\}) = 0.1$.

As future work we plan to consider more exhaustive analysis of the application of normality tests to these distributions. We need to consider other tests and also the application of these tests to distributions of larger dimensions.

6. Acknowledgments

Partial support by the Spanish MEC project COPRIVACY (TIN2011-27076-C03-03) is acknowledged.

The comments of the referees are gratefully acknowledge.

References

[1] G. Choquet, *Theory of capacities*, Ann. Inst. Fourier **5** (1953/54), 131–295.

[2] S. Greco, B. Matarazzo, and S. Giove, *The Choquet integral with respect to a level dependent capacity*, Fuzzy Sets and Systems **175** (2011), 1–35.

[3] A. Kankainen, S. Taskinen, and H. Oja, *On Mardia's tests of multinormality*, in: Theory and Applications of Recent Robust Methods (Edited by Hubert, M., Pison, G., Struyf, A., and Van Aelst, S.), Birkhäuser, Basel, 2004, pp. 153–164.

[4] K. V. Mardia, *Measures of multivariate skewness and kurtosis with applications*, Biometrika **57**(3) (1970), 519–530.

[5] C. J. Mecklin and D. J. Mundfrom *An appraisal and bibliography of tests for multivariate normality*, International Statistical Review **72**(1) (2004), 123–138.

[6] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, Wiley-Interscience, Hoboken, NJ, 2003.

[7] M. Sugeno, *Theory of Fuzzy Integrals and its Applications*, Ph. D. Dissertation, Tokyo Institute of Technology, Tokyo, 1974.

[8] V. Torra, *Distributions based on the Choquet integral and non-additive measures*, RIMS Kokyuroku **1906** (2014), 136–143.

[9] V. Torra and Y. Narukawa, *On a comparison between Mahalanobis distance and Choquet integral: The Choquet-Mahalanobis operator*, Inform. Sci. **190** (2012), 56–63.

[10] V. Torra and Y. Narukawa, *Exponential family of level dependent Choquet integral based class-conditional probability functions*, in: Proc. AGOP 2013, Advances in Intelligent Systems and Computing **228**, Springer, 2013, pp. 477–484.

[11] V. Torra, Y. Narukawa, and M. Sugeno (eds.), *Non-additive Measures: Theory and Applications*, Springer, 2013.

[12] R. Wicklin, *What is Mahalanobis distance?* Entry February 15, 2012. http://blogs.sas.com/content/iml/2012/02/15/what-is-mahalanobis-distance/

University of Skövde, Skövde, 541 28 Sweden
E-mail address: vtorra@his.se