On the moments and the distribution of the Choquet integral

Ivan Kojadinovic
LINA CNRS FRE 2729
Site école polytechnique de l'univ. de Nantes
Rue Christian Pauc
44306 Nantes, France
ivan.kojadinovic@univ-nantes.fr

Jean-Luc Marichal
Applied Mathematics Unit
Univ. of Luxembourg
162A, avenue de la Faïencerie
L-1511 Luxembourg, G.D. Luxembourg
jean-luc.marichal@uni.lu

Abstract

We investigate the distribution functions and the moments of the so-called *Choquet integral*, also known as the *Lovász extension*, when regarded as a real function of a random sample drawn from a continuous population. Since the Choquet integral includes weighted arithmetic means, ordered weighted averaging operators, and lattice polynomials as particular cases, our results encompass the corresponding results for these aggregation operators. After recalling the results obtained by the authors in the uniform case, we present approaches that can be used in the non-uniform case to obtain moment approximations.

Keywords: Discrete Choquet integral; Lovász extension; Order statistic; Distribution function; Moment; B-Spline; Divided difference; Moment approximation; Asymptotic distribution.

1 Introduction

Aggregation operators are of central importance in many fields such as statistics or decision theory. Among such commonly used operators, the most frequently employed is probably the weighted arithmetic mean because of its simplicity and its very intuitive interpretation.

Although very attractive in many fields, the weighted arithmetic mean is not suited for situations where the values to be aggregated display some interaction. Let us choose the framework of multi-criteria decision aid to elaborate this in more detail. We consider a set $N := \{1, \dots, n\}$ of *criteria* and a set $\mathcal{A}$ of *alternatives* evaluated according to these criteria. As classically done, we assume that with each alternative $a \in \mathcal{A}$ a vector $(a_1, \dots, a_n) \in \mathbb{R}^n$ is associated, where, for any $i \in N$, a_i represents the *partial score* of a related to criterion i. The partial scores are further assumed to be defined on the same interval scale.

From the vector of scores of any alternative, one can compute an overall evaluation by means of an aggregation operator. Once the overall evaluations are computed, they can be used to rank the alternatives. In such a context, it is very frequent in applications to have criteria that are *substitutive* or *complementary*. Substitutivity between two criteria arises when an alternative can be assigned a high overall score when only one of the two criteria has a high partial evaluation. Complementarity means that it is necessary that the two criteria have simultaneously a high evaluation for the alternative to receive an overall high score. A natural extension of the weighted arithmetic mean that is able to deal with such situations (and many others) is the so-called *Choquet integral* w.r.t. a *capacity* [1, 2, 3].

Also called *Lovász extension* [4] in the context of the extension of pseudo-Boolean functions, the Choquet integral includes weighted arithmetic means, ordered weighted averaging operators [5], and lattice polynomials as particular cases [6, 7].

In this paper, we investigate the distribution and the moments of the *Choquet integral* when considered as a real function of a random sample drawn from a continuous population. In the uniform case, we recall the results obtained by the authors in [7] and we provide algorithms for computing the probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of the Choquet integral. In the non-uniform case, we present approaches that can be used to obtain approximations of the moments of this functional.

In order to avoid a cumbersome notation, cardinality of subsets $S, T, \dots$ will be denoted whenever possible by the corresponding lower case letters $s, t, \dots$, otherwise by the standard notation $|S|, |T|, \dots$. Moreover, we will often omit braces for singletons, e.g., writing $\nu(i)$, $N \setminus i$ instead of $\nu(\{i\})$, $N \setminus \{i\}$. Finally, the set of permutations on N will be denoted by $\mathfrak{S}_n$.

2 The Choquet integral and its particular cases

A set function $\nu : 2^N \to [0,1]$ is a *capacity* [1] on $N := \{1,\ldots,n\}$ if it is monotone with respect to (w.r.t.) inclusion and satisfies $\nu(\emptyset) = 0$ and $\nu(N) = 1$. In the context of aggregation by the Choquet integral, for any $T \subseteq N$, the coefficient $\nu(T)$ is to be interpreted as the weight of importance of the combination T of criteria, or better, its importance or power to make the decision alone (without the remaining criteria).

Definition 1. The Choquet integral of $x \in \mathbb{R}^n$ w.r.t. a capacity ν on N is defined by

$$C_\nu(x) := \sum_{i=1}^n p_i^{\nu,\sigma}\, x_{\sigma(i)}, \tag{1}$$

where σ is a permutation on N such that $x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}$, where

$$p_i^{\nu,\sigma} := \nu_i^\sigma - \nu_{i-1}^\sigma, \qquad \forall\, i \in N, \tag{2}$$

and where $\nu_i^\sigma := \nu\big(\{\sigma(1),\ldots,\sigma(i)\}\big)$ for any $i = 0,\ldots,n$. In particular, $\nu_0^\sigma := 0$.

The Choquet integral can therefore be regarded as a piecewise linear function that coincides with a weighted arithmetic mean on each n-dimensional region

$$R_\sigma := \{x \in \mathbb{R}^n \mid x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}\} \qquad (\sigma \in \mathfrak{S}_n), \tag{3}$$

whose union covers $\mathbb{R}^n$.

The Choquet integral satisfies very appealing properties for aggregation. For instance, it is continuous, non decreasing, comprised between min and max, stable under the same transformations of interval scales in the sense of the theory of measurement, and coincides with the weighted arithmetic mean whenever the capacity is additive. An axiomatic characterization is provided in [3].

We now present some subclasses of Choquet integrals. Any vector $\omega \in [0,1]^n$ such that $\sum_i \omega_i = 1$ will be called a *weight vector* as we continue.

2.1 The weighted arithmetic mean

Definition 2. For any weight vector $\omega \in [0,1]^n$, the weighted arithmetic mean operator WAM_ω associated to ω is defined by

$$\mathrm{WAM}_\omega(x) := \sum_{i=1}^n \omega_i\, x_i, \qquad \forall\, x \in \mathbb{R}^n.$$

We can easily see that WAM_ω is a Choquet integral C_ν with respect to the additive capacity defined by $\nu(T) := \sum_{i\in T} \omega_i$ for all $T \subseteq N$. Conversely, the weights associated to WAM_ω are defined by $\omega_i := \nu(i)$ for all $i \in N$.

The class of weighted arithmetic means WAM_ω includes two important special cases, namely:

- the arithmetic mean $\mathrm{AM}(x) := \frac{1}{n}\sum_{i=1}^n x_i$, when $\omega_i = 1/n$ for all $i \in N$. In this case, we have $\nu(T) := t/n$ for all $T \subseteq N$.
- the k-th projection $\mathrm{P}_k(x) := x_k$, when $\omega_k = 1$ for some $k \in N$. In this case, we have $\nu(T) := 1$ if $T \ni k$ and 0 otherwise.

2.2 The ordered weighted averaging operator

The concept of ordered weighted averaging operator was proposed in aggregation theory by Yager [5] and corresponds, in statistics, to that of linear combination of order statistics.

Definition 3. For any weight vector $\omega \in [0,1]^n$, the ordered weighted averaging operator OWA_ω associated to ω is defined by

$$\mathrm{OWA}_\omega(x) := \sum_{i=1}^n \omega_i\, x_{\sigma(i)}, \qquad \forall\, x \in \mathbb{R}^n,$$

where σ is a permutation on N such that $x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}$.

It is easy to verify that an OWA operator is a Choquet integral w.r.t. a capacity that depends only on the cardinality of subsets. The capacity ν associated to OWA_ω is defined by

$$\nu(T) := \sum_{i=1}^t \omega_i, \qquad T \subseteq N,\ T \neq \emptyset.$$

Conversely, the weights associated to OWA_ω are defined by $\omega_t := \nu(T) - \nu(T \setminus i)$ for all $T \subseteq N$ and all $i \in T$.

The class of ordered weighted averaging operators OWA_ω includes some important special cases, namely:

- the arithmetic mean when $\omega_i = 1/n$ for all $i \in N$.
- the k-th order statistic when $\omega_{n-k+1} = 1$ for some $k \in N$. In this case, we have
$$\nu(T) := \begin{cases} 1 & \text{if } t \geqslant n-k+1, \\ 0 & \text{otherwise,} \end{cases}$$
- the min operator
$$\min(x) = \min_{i\in N} x_i,$$

when $\omega_n = 1$. In this case, we have $\nu(T) := 1$ if $T = N$ and 0 otherwise.

- the max operator

$$\max(x) = \max_{i \in N} x_i,$$

when $\omega_1 = 1$. In this case, we have $\nu(T) := 1$ for all $T \neq \emptyset$.

2.3 Partial minimum and maximum

Definition 4. For any non-empty subset $A \subseteq N$, the partial minimum operator $\min_A$ and the partial maximum operator $\max_A$, associated to A, are respectively defined by

$$\begin{aligned}\min_A(x) &= \min_{i \in A} x_i,\\ \max_A(x) &= \max_{i \in A} x_i.\end{aligned}$$

For the operator $\min_A$ (resp. $\max_A$), for any $T \subseteq N$, we have

$$\nu(T) := \begin{cases} 1 & \text{if } T \supseteq A,\\ 0 & \text{otherwise.}\end{cases}$$

$$\left(\text{resp. } \nu(T) := \begin{cases} 1 & \text{if } T \cap A \neq \emptyset,\\ 0 & \text{otherwise.}\end{cases}\right)$$

These operators are particular cases of *lattice polynomials* that also correspond to special classes of Choquet integrals; see [6, 7] for more details.

3 Distributional relationships with linear combination of order statistics

From Definition 1, it is clear that the Choquet integral is a linear combination of order statistics whose coefficients depend on the order of the arguments. We state hereafter immediate relationships between the moments (resp. the c.d.f.) of the Choquet integral and the moments (resp. the c.d.f.) of linear combination of order statistics.

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with p.d.f. f and let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ denote the corresponding order statistics. Further, let $Y_\nu := C_\nu(X_1, \ldots, X_n)$ and let h be any function. By definition of the expectation, we have

$$\begin{aligned}\mathbf{E}[h(Y_\nu)] &= \int_{\mathbb{R}^n} h(C_\nu(x_1, \ldots, x_n)) \prod_{i=1}^n f(x_i)\mathrm{d}x_i\\ &= \sum_{\sigma \in \mathfrak{S}_n} \int_{R_\sigma} h\left(\sum_{i=1}^n p_i^{\nu,\sigma} x_{\sigma(i)}\right) \prod_{i=1}^n f(x_i)\mathrm{d}x_i\end{aligned}$$

Using the well-known fact that the joint p.d.f. of $X_{n:n} \geqslant \cdots \geqslant X_{1:n}$ is

$$n! \prod_{i=1}^n f(x_i), \qquad x_n \geqslant \cdots \geqslant x_1,$$

we obtain

$$\begin{aligned}\mathbf{E}[h(Y_\nu)] &= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbf{E}\left[h\left(\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n}\right)\right]\\ &= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbf{E}[h(Y_\nu^\sigma)]. \qquad (4)\end{aligned}$$

where $Y_\nu^\sigma := \sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n}$ are linear combinations of order statistics. Clearly, the special cases

$$h(x) = x^r, \ [x - \mathbf{E}(Y_\nu)]^r, \text{ and } e^{tx}$$

provide similar relationships, respectively, for raw moments, central moments, and moment-generating functions.

Now, consider the *minus* (resp. *plus*) *truncated power function* x_-^n (resp. x_+^n) defined to be x^n if $x < 0$ (resp. $x > 0$) and zero otherwise. Given $y \in \mathbb{R}$, taking $h(x) = (x - y)_-^0$ in (4) provides a relationship between the c.d.f. F_ν of Y_ν and those of the random variables $\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n}$. Indeed, we clearly have

$$F_\nu(y) := \mathbf{Pr}[Y_\nu \leqslant y] = \mathbf{E}[(Y_\nu - y)_-^0],$$

and, denoting by F_ν^σ the c.d.f. of Y_ν^σ,

$$\begin{aligned}F_\nu^\sigma(y) :=& \mathbf{Pr}\left[\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} \leqslant y\right]\\ =& \mathbf{E}\left[\left(\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} - y\right)_-^0\right],\end{aligned}$$

which immediately gives

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} F_\nu^\sigma(y) \qquad (5)$$

As one could have expected from the definition of the Choquet integral, the determination of the moments and the distribution functions of the Choquet integral is closely related to the determination of the moments and the distribution functions of linear combinations of order statistics.

4 The uniform case

In this section, we are interested in the moments and distribution functions of Y_ν when the random sample

$X_1, \ldots, X_n$ is drawn from the standard uniform distribution. In order to emphasize this last point, as classically done, we shall denote the random sample as $U_1, \ldots, U_n$ and the corresponding order statistics by $U_{1:n} \leqslant \cdots \leqslant U_{n:n}$.

Before yielding the main results obtained in [7], let us recall some basic material related to divided differences. See for instance [8, 9, 10] for further details.

4.1 Divided differences

Let $\mathcal{A}^{(n)}$ be the set of $n-1$ times differentiable one-place functions g such that $g^{(n-1)}$ is absolutely continuous. The nth *divided difference* of a function $g \in \mathcal{A}^{(n)}$ is the symmetric function of $n+1$ arguments defined inductively by $\Delta[g : a_0] := g(a_0)$ and

$$\Delta[g : a_0, \ldots, a_n] :=$$
$$\begin{cases} \dfrac{\Delta[g : a_1, \ldots, a_n] - \Delta[g : a_0, \ldots, a_{n-1}]}{a_n - a_0}, & \text{if } a_0 \neq a_n, \\ \dfrac{\partial}{\partial a_0} \Delta[g : a_0, \ldots, a_{n-1}], & \text{if } a_0 = a_n. \end{cases}$$

The *Peano representation* of the divided differences, which can be obtained by a Taylor expansion of g, is given by

$$\Delta[g : a_0, \ldots, a_n] = \frac{1}{n!} \int_{\mathbb{R}} g^{(n)}(t)\, M(t \mid a_0, \ldots, a_n)\, \mathrm{d}t, \tag{6}$$

where $M(t \mid a_0, \ldots, a_n)$ is the *B-spline* of order n, with knots $\{a_0, \ldots, a_n\}$, defined as

$$M(t \mid a_0, \ldots, a_n) := n\, \Delta[(\cdot - t)_+^{n-1} : a_0, \ldots, a_n]. \tag{7}$$

We also recall the *Hermite-Genocchi formula*: For any function $g \in \mathcal{A}^{(n)}$, we have

$$\Delta[g : a_0, \ldots, a_n]$$
$$= \int_{R_{id} \cap [0,1]^n} g^{(n)} \Big[a_0 + \sum_{i=1}^{n} (a_i - a_{i-1}) x_i\Big]\, \mathrm{d}x, \tag{8}$$

where R_{id} is the region defined in (3) when σ is the identity permutation.

For distinct arguments $a_0, \ldots, a_n$, we also have the following formula, which can be verified by induction,

$$\Delta[g : a_0, \ldots, a_n] = \sum_{i=0}^{n} \frac{g(a_i)}{\prod_{j \neq i} (a_i - a_j)}. \tag{9}$$

4.2 Moments and distribution

Let $g \in \mathcal{A}^{(n)}$. From (8), we immediately have that

$$\mathbf{E}\left[g^{(n)}\left(\sum_{i=1}^{n} p_i^{\nu,\sigma}\, U_{n-i+1:n}\right)\right] = n!\Delta[g : \nu_0^\sigma, \ldots, \nu_n^\sigma]$$

since the joint p.d.f. of $U_{n:n} \geqslant \cdots \geqslant U_{1:n}$ is equal to $\frac{1}{n!}$ on $R_{id} \cap [0,1]^n$ and is zero elsewhere. Combining the previous expression with (4), we obtain

$$\mathbf{E}[g^{(n)}(Y_\nu)] = \sum_{\sigma \in \mathfrak{S}_n} \Delta[g : \nu_0^\sigma, \ldots, \nu_n^\sigma]. \tag{10}$$

Eq. (10) provides the expectation $\mathbf{E}[g^{(n)}(Y_\nu)]$ in terms of the divided differences of g with arguments $\nu_0^\sigma, \ldots, \nu_n^\sigma$ ($\sigma \in \mathfrak{S}_n$). An explicit formula can be obtained by (9) whenever the arguments are distinct for every $\sigma \in \mathfrak{S}_n$.

Clearly, the special cases

$$g(x) = \frac{r!}{(n+r)!}\, x^{n+r},\ \frac{r!}{(n+r)!}\, [x - \mathbf{E}(Y_\nu)]^{n+r}, \text{ and } \frac{e^{tx}}{t^n} \tag{11}$$

give, respectively, the raw moments, the central moments, and the moment-generating function of Y_ν. As far as the raw moments are concerned, the following result was obtained in [7].

Proposition 1. *For any integer $r \geqslant 1$, setting $A_0 := N$, we have,*

$$\mathbf{E}[Y_\nu^r] = \frac{1}{\binom{n+r}{r}} \sum_{\substack{A_1 \subseteq N \\ A_2 \subseteq A_1 \\ \cdots \\ A_r \subseteq A_{r-1}}} \prod_{i=1}^{r} \frac{1}{\binom{|A_{i-1}|}{|A_i|}}\, \nu(A_i).$$

Proposition 1 provides an explicit expression for the rth raw moment of Y_ν as a sum of $(r+1)^n$ terms. For instance, the first two moments are

$$\mathbf{E}[Y_\nu] = \frac{1}{n+1} \sum_{A \subseteq N} \frac{1}{\binom{n}{|A|}}\, \nu(A),$$
$$\mathbf{E}[Y_\nu^2] = \frac{2}{(n+1)(n+2)} \times$$
$$\sum_{A_1 \subseteq N} \frac{1}{\binom{n}{|A_1|}}\, \nu(A_1) \sum_{A_2 \subseteq A_1} \frac{1}{\binom{|A_1|}{|A_2|}}\, \nu(A_2).$$

As far as the distribution function $F_\nu(y) := \mathbf{Pr}[Y_\nu \leqslant y]$ of Y_ν is concerned, using (10) with $g(x) = \frac{1}{n!}(x-y)_-^n$, the following result was obtained in [7].

Theorem 1. *There holds*

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_-^n : \nu_0^\sigma, \ldots, \nu_n^\sigma] \tag{12}$$
$$= 1 - \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_+^n : \nu_0^\sigma, \ldots, \nu_n^\sigma].$$

It follows from (12) that the distribution function of Y_ν is absolutely continuous and hence its probability density function is simply given by

$$f_\nu(y) = \frac{1}{(n-1)!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_+^{n-1} : \nu_0^\sigma, \ldots, \nu_n^\sigma], \tag{13}$$

or, using the B-spline notation (7),

$$f_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} M(y \mid \nu_0^\sigma, \ldots, \nu_n^\sigma).$$

Remark 1. (i) When the arguments $\nu_0^\sigma, \ldots, \nu_n^\sigma$ are distinct for every $\sigma \in \mathfrak{S}_n$, then combining (9) with (12) immediately yields the following explicit expression

$$F_\nu(y) = 1 - \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=0}^{n} \frac{(\nu_i^\sigma - y)_+^n}{\prod_{j \neq i}(\nu_i^\sigma - \nu_j^\sigma)},$$

or, using the minus truncated power function,

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=0}^{n} \frac{(\nu_i^\sigma - y)_-^n}{\prod_{j \neq i}(\nu_i^\sigma - \nu_j^\sigma)}.$$

(ii) The case of linear combinations of order statistics, called *ordered weighted averaging operators* in aggregation theory (see § 2.2), is of particular interest. In this case, each ν_i^σ is independent of σ, so that we can write $\nu_i := \nu_i^\sigma$. The main formulas then reduce to (see for instance [11] and [12])

$$\begin{aligned} \mathbf{E}[g^{(n)}(Y_\nu)] &= n!\,\Delta[g : \nu_0, \ldots, \nu_n], \\ F_\nu(y) &= \Delta[(\cdot - y)_-^n : \nu_0, \ldots, \nu_n], \\ f_\nu(y) &= M(y \mid \nu_0, \ldots, \nu_n). \end{aligned}$$

We also note that the Hermite-Genocchi formula (8) provides nice geometric interpretations of $F_\nu(y)$ and $f_\nu(y)$ in terms of volumes of slices and sections of canonical simplices (see also [13] and [14]).

4.3 Algorithms for computing divided differences

Both functions F_ν and f_ν require the computation of divided differences of truncated power functions. On this issue, we recall a recurrence equation, due to de Boor [15] and rediscovered independently by Varsi [16] (see also [13]), which allows to compute $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n]$ in $O(n^2)$ operations.

Rename as $b_1, \ldots, b_r$ the elements a_i such that $a_i < y$ and as $c_1, \ldots, c_s$ the elements a_i such that $a_i \geqslant y$ so that $r + s = n + 1$. Then, the unique solution of the recurrence equation

$$\alpha_{k,l} = \frac{(c_l - y)\alpha_{k-1,l} + (y - b_k)\alpha_{k,l-1}}{c_l - b_k} \qquad (k \leqslant r,\, l \leqslant s),$$

with initial values $\alpha_{1,1} = (c_1 - b_1)^{-1}$ and $\alpha_{0,l} = \alpha_{k,0} = 0$ for all $l, k \geqslant 2$, is given by

$$\alpha_{k,l} := \Delta[(\cdot - y)_+^{k+l-2} : b_1, \ldots, b_k, c_1, \ldots, c_l],\ (k+l \geqslant 2).$$

Algorithm 1 Algorithm for the computation of $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n]$.

Require: n, $a_0, \ldots, a_n$, y
  $S \leftarrow 0$, $R \leftarrow 0$
  for $i = 0, 1, \ldots, n$ **do**
    if $x_i - y \geqslant 0$ **then**
      $S \leftarrow S + 1$
      $C_S \leftarrow x_i - y$
    else
      $R \leftarrow R + 1$
      $B_R \leftarrow x_i - y$
    end if
  end for
  $A_0 \leftarrow 0$, $A_1 \leftarrow 1/(C_1 - B_1)$ {Initialization of the unidimensional temporary array of size $S+1$ necessary for the computation of the divided difference}
  for $j = 2, \ldots, S$ **do**
    $A_j \leftarrow -B_1 A_{j-1}/(C_j - B_1)$
  end for
  for $i = 2, \ldots, R$ **do**
    for $j = 1, \ldots, S$ **do**
      $A_j \leftarrow (C_j A_j - B_i A_{j-1})/(C_j - B_i)$
    end for
  end for
  return A_R {Contains the value of $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n]$.}

In order to compute $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n] = \alpha_{r,s}$, it suffices therefore to compute the sequence $\alpha_{k,l}$ for $k + l \geqslant 2$, $k \leqslant r$, $l \leqslant s$, by means of 2 nested loops, one on k, the other on l. We detail this computation in Algorithm 1; see also [13, 16].

We can compute $\Delta[(\cdot - y)_-^n : a_0, \ldots, a_n]$ similarly. Indeed, the same recurrence equation applied to the initial values $\alpha_{0,l} = 0$ for all $l \geqslant 1$ and $\alpha_{k,0} = 1$ for all $k \geqslant 1$, produces the solution

$$\alpha_{k,l} := \Delta[(\cdot - y)_-^{k+l-1} : b_1, \ldots, b_k, c_1, \ldots, c_l] \qquad (k+l \geqslant 1).$$

Example. As we have already mentioned, the Choquet integral is widely used in non-additive expected utility theory, cooperative game theory, complexity analysis, measure theory, etc. (see [17] for an overview.) For instance, when a discrete Choquet integral is used as an aggregation tool in a given decision making problem, it is then very informative for the decision maker to know its distribution. In that context, the most natural *a priori* density on $[0,1]^n$ is the uniform one, which makes the results presented in this section of particular interest.

Let ν be the capacity on $\{1,2,3\}$ defined by $\nu(\{1\}) = 0.1$, $\nu(\{2\}) = 0.6$, $\nu(\{3\}) = \nu(\{1,2\}) = \nu(\{1,3\}) = \nu(\{2,3\}) = 0.9$, and $\nu(\{1,2,3\}) = 1$. The density of the Choquet integral w.r.t. ν, which can be computed through (13) and by means of Algorithm 1, is

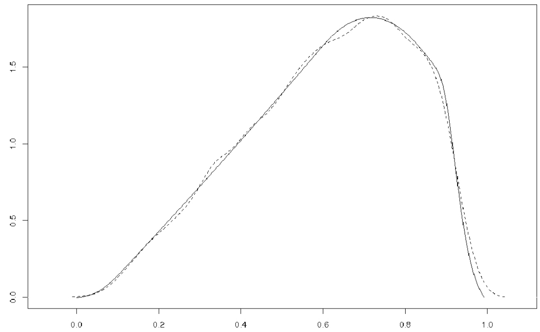

Figure 1: Density of a discrete Choquet integral (solid line).

represented in Figure 1 by the solid line. The dotted line represents the density estimated by the kernel method from 10 000 randomly generated realizations. The typical value and standard deviation can also be calculated through the raw moments: we have

$$\mathbf{E}[Y_\nu] \approx 0.608 \quad \text{and} \quad \sqrt{\mathbf{E}[Y_\nu^2] - \mathbf{E}[Y_\nu]^2} \approx 0.204.$$

From a practical perspective, routines for computing the p.d.f. and the c.d.f. of the Choquet integral in the uniform case have been implemented in the Kappalab package for GNU R [18].

5 The non-uniform case

We now turn to the non-uniform case. Let $X_1, \dots, X_n$ be a random sample from a continuous distribution with c.d.f. F. Unlike in the uniform case, in this section we will only be able to present results allowing to compute approximations of moments of $Y_\nu := C_\nu(X_1, \dots, X_n)$.

5.1 Expectation and variance of the Choquet integral

We first focus on the expectation and the variance of Y_ν. Starting from (4) with $h(x) = x$, we immediately obtain

$$\mathbf{E}[Y_\nu] = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{n} p_i^{\nu,\sigma}\, \mathbf{E}[X_{n-i+1:n}]. \tag{14}$$

Similarly, for $h(x) = x^2$, we get

$$\mathbf{E}[Y_\nu^2] = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i^{\nu,\sigma} p_j^{\nu,\sigma}\, \mathbf{E}[X_{n-i+1:n} X_{n-j+1:n}]. \tag{15}$$

It immediatelty follows that the expectation and the variance of the Choquet integral can be computed in the non-uniform case only if the first product moments of order statistics from the same underlying distribution can be computed. As we shall see in the next subsection, it is possible to obtain approximations of these product moments provided the inverse of F and the derivatives of the inverse can be computed.

5.2 Moments of order statistics and their approximation

Let $U_1, \dots, U_n$ be a random sample from the standard uniform distribution. The product moments of the corresponding order statistics are then given by the following formula (see e.g. [19, Chap. 3] and the references therein):

$$\mathbf{E}\left(\prod_{j=1}^{l} U_{i_j:n}^{m_j}\right) = \frac{n!}{\left(n + \sum_{j=1}^{l} m_j\right)!} \times \prod_{j=1}^{l} \frac{(i_j + m_1 + \cdots + m_j - 1)!}{(i_j + m_1 + \cdots + m_{j-1} - 1)!}. \tag{16}$$

Now, it is well-known that the c.d.f. of $X_{i:n}$ is given by

$$\mathbf{Pr}(X_{i:n} \leqslant x) = \sum_{j=i}^{n} \binom{n}{j} F^j(x)[1 - F(x)]^{n-j}.$$

It immediately follows that

$$\mathbf{Pr}(F^{-1}(U_{i:n}) \leqslant x) = \mathbf{Pr}(U_{i:n} \leqslant F(x)) = \mathbf{Pr}(X_{i:n} \leqslant x),$$

i.e. that $F^{-1}(U_{i:n})$ and $X_{i:n}$ are equal in distribution.

Starting from this distributional equality, David and Johnson [20] expanded $F^{-1}(U_{i:n})$ in a Taylor series around the point $\mathbf{E}(U_{i:n}) = i/(n+1)$ in order to obtain approximations of product moments of non-uniform order statistcs; see also [19, §4.6]. Setting $p_i := i/(n+1)$ and $G := F^{-1}$, we have

$$X_{i:n} = G(p_i) + (U_{i:n} - p_i)G'(p_i) + \frac{1}{2}(U_{i:n} - p_i)^2 G''(p_i) + \frac{1}{6}(U_{i:n} - p_i)^3 G'''(p_i) + \dots \tag{17}$$

Taking the expectation of the previous expression and using (16), the following approximation for the expectation of $X_{i:n}$ can be obtained to order $(n+2)^{-2}$ [19, §4.6]:

$$\mathbf{E}[X_{i:n}] \approx G_i + \frac{p_i q_i}{2(n+2)} G_i'' + \frac{p_i q_i}{(n+2)^2}\left[\frac{1}{3}(q_i - p_i)G_i''' + \frac{1}{8} p_i q_i G_i''''\right], \tag{18}$$

where $q_i := 1 - p_i$ and $G_i := G(p_i)$, $G_i' := G'(p_i)$, etc. Similarly, for the first product moment, we have, to order $(n+2)^{-2}$,

$$\begin{aligned}\mathbf{E}[X_{i:n}X_{j:n}] \approx\ & G_iG_j + \frac{p_iq_j}{n+2}G_i'G_j' + \frac{p_iq_i}{2(n+2)}G_jG_i'' \\ & + \frac{p_jq_j}{2(n+2)}G_iG_j'' + \frac{p_iq_j}{(n+2)^2}\Big[(q_i-p_i)G_i''G_j' \\ & + (q_j-p_j)G_i'G_j'' + \frac{1}{2}p_iq_iG_i'''G_j' + \frac{1}{2}p_jq_jG_i'G_j''' \\ & + \frac{1}{2}p_iq_jG_i''G_j''\Big] + \frac{p_ip_jq_iq_j}{4(n+2)^2}G_i''G_j'' \\ & + \frac{p_iq_iG_j}{(n+2)^2}\Big[\frac{1}{8}p_iq_iG_i'''' + \frac{1}{3}(q_i-p_i)G_i''\Big] \\ & + \frac{p_jq_jG_i}{(n+2)^2}\Big[\frac{1}{8}p_jq_jG_j'''' + \frac{1}{3}(q_j-p_j)G_j''\Big]. \qquad (19)\end{aligned}$$

The accuracy of the above approximations is discussed in [19, §4.6]. Note that Childs and Balakrishnan [21] have recently proposed MAPLE routines facilitating the computations and permitting the inclusion of higher order terms.

From a practical perspective, the previous expressions are useful only if $G := F^{-1}$ and its derivates can be easily computed. This is the case for instance when F is the c.d.f. of the standard normal distribution. Indeed, there exists algorithms that enable an accurate computation of F^{-1} (see [22] and the references therein) and it can be verified (see [19, p 85]) that $G' = (f\circ G)^{-1}$,

$$G'' = \frac{G}{f^2\circ G}, G''' = \frac{1+2G^2}{f^3\circ G} \text{ and } G'''' = \frac{G(7+6G^2)}{f^4\circ G},$$

where $f = F'$.

5.3 Back to the two first moments of the Choquet integral

Combining (18) and (19) with (14) and (15), it is therefore possible to obtain approximate values for the expectation and the variance of the Choquet integral provided F^{-1} and its derivates can be easily computed.

6 Future work

Using the expresssions given in Section 3 and distributional results on linear combinations of order statistics [19], it is possible to obtain the exact distribution of the Choquet integral for certain non-uniform distributions and also conditions under which the Choquet integral is asymptotically normal. These aspects will be studied in a forthcoming paper.

References

[1] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] M. Grabisch. Fuzzy integral in multicriteria decision making. *Fuzzy Sets and Systems*, 69:279–298, 1995.

[3] J.-L. Marichal. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8(6):800–807, 2000.

[4] L. Lovász. Submodular function and convexity. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical programming: The state of the art*, pages 235–257. Springer-Verlag, 1983.

[5] R. D. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, 18:183–190, 1988.

[6] J.-L. Marichal. Cumulative distribution functions and moments of lattice polynomials. *Statistics and Probability Letters*, page in press, 2005.

[7] J.-L. Marichal and I. Kojadinovic. The distribution of linear combinations of lattice polynomials from the uniform distribution. submitted, 2006.

[8] Philip J. Davis. *Interpolation and approximation. 2nd ed.* Dover Books on Advanced Mathematics. New York: Dover Publications, 1975.

[9] R.A. DeVore and G.G. Lorentz. *Constructive approximation.* Berlin: Springer-Verlag, 1993.

[10] M.J.D. Powell. *Approximation theory and methods.* Cambridge University Press, 1981.

[11] J. A. Adell and C. Sangüesa. Error bounds in divided difference expansions. A probabilistic perspective. *J. Math. Anal. Appl.*, 318(1):352–364, 2006.

[12] Girdhar G. Agarwal, Rohan J. Dalpatadu, and Ashok K. Singh. Linear functions of uniform order statistics and B-splines. *Commun. Stat., Theory Methods*, 31(2):181–192, 2002.

[13] M. M. Ali. Content of the frustum of a simplex. *Pacific Journal of Mathematics*, 48(2):313–322, 1973.

[14] L. Gerber. The volume cut off a simplex by a half space. *Pacific Journal of Mathematics*, 94(2):311–314, 1981.

[15] C. de Boor. On calculating with B-splines. *Journal of Approximation Theory*, 6:50–62, 1972.

[16] G. Varsi. The multidimensional content of the frustrum of the simplex. *Pacific Journal of Mathematics*, 46(1):303–314, 1973.

[17] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy measures and integrals*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000. Theory and applications.

[18] M. Grabisch, I. Kojadinovic, and P. Meyer. `kappalab`*: Non additive measure and integral manipulation functions*, 2006. R package version 0.3.

[19] H.A. David and H.N. Nagaraja. *Order statistics. 3rd ed.* Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons., 2003.

[20] F.N. David and N.L. Johnson. Statistical treatment of censored data, I. Fundamental formulae. *Biometrika*, 41:228–240, 1954.

[21] A. Childs and N. Balakrishnan. Series approximation for moments of order statistics using maple. *Computational Statistics and Data Analysis*, 38:331–347, 2002.

[22] P. J. Acklam. An algorithm for computing the inverse normal cumulative distribution function. `http://home.online.no/~pjacklam/notes/invnorm/`.