An empirical study of statistical properties of Choquet and Sugeno integrals

Michel Grabisch, Éric Raufaste

An empirical study of statistical properties of the Choquet and Sugeno integrals

Michel Grabisch
Université Paris I Panthéon-Sorbonne
LIP6, 8 rue du Capitaine Scott, 75015 Paris, France
email: michel.grabisch@lip6.fr

Eric Raufaste
Université Toulouse II Le Mirail
LTC, 5 Allées Machado 31058 Toulouse Cedex, France
email: raufaste@univ-tlse2.fr

Abstract

This paper investigates the statistical properties of the Choquet and Sugeno integrals, used as multiattribute models. The investigation is done on an empirical basis, and focuses on two topics: the distribution of the output of these integrals when the input is corrupted with noise, and the robustness of these models, when they are identified using some set of learning data through some learning procedure.

I. Introduction

The Choquet integral [3] and the Sugeno integral [31], also known under the generic name of fuzzy integral, have become widely used aggregation functions, especially in multicriteria decision making [9], [18], subjective evaluation [2], [15], [24], [32], pattern classification [12], [23], image processing [16], [22], [23], information fusion [1], [5], [33], regression analysis [34], etc. (see also [20] for a detailed study and many references).

Their mathematical properties as aggregation functions have been studied extensively [4], [6], [25], [29], and it is known that many classical aggregation functions are particular cases of these so-called fuzzy integrals, e.g., the weighted arithmetic mean, ordered weigthed averages (OWA), weighted minimum and maximum, etc.

It is surprising that almost no study (at least to the knowledge of the authors) has been done concerning the statistical properties of the Choquet and Sugeno integrals, since this question is of primary importance in any application, where the robustness of models against noise has to be evaluated. The answer may lie in the mathematical difficulty to analyze the statistical behavior of these integrals, due to their nonlinear character. However, a very recent theoretical work has been done in this direction by Marichal, who obtained the mathematical expression of the distribution of the Sugeno integral [26]. We will present this result in Section III.

This paper aims to fill this gap, by providing an empirical analysis of statistical properties of the Choquet and Sugeno integrals, based on synthetic and real data. Two questions are addressed: for an input vector corrupted with Gaussian noise, what is the distribution, mean and variance of the output of Choquet and Sugeno integrals? Second, what is the impact of noise corrupting learning data on the fuzzy measure, i.e., the parameters of the model, for a given learning procedure? This second question in fact addresses the problem of robustness of such models.

The paper is organized as follows. Section II recalls basic definitions on fuzzy integrals, while Section III states precisely the kinds of problem we study. Sections IV and V give the result of the empirical studies, concerning respectively the analysis of output when the input vector is corrupted with noise, and the analysis of the fuzzy measure when the learning data is corrupted with noise. Section VI concludes the paper.

All tables and figures showing results of experiments are put in appendix. We give also in appendix an improved version of the HLMS algorithm, whose original version [7] is well known by practitioners of the Choquet integral.

II. Basic definitions

Throughout the paper, we assume that input vectors are n-dimensional nonnegative vectors, and $N := \{1, \ldots, n\}$ is the index set. We call for commodity *attributes* the dimensions of x.

Definition 1: A *fuzzy measure* [31] or *capacity* [3] is a function $\mu : 2^N \to \mathbb{R}_+$ such that $\mu(\emptyset) = 0$, and $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$ (*monotonicity*). A fuzzy measure is *normalized* if $\mu(N) = 1$.

We assume in this paper that fuzzy measures are normalized. A fuzzy measure is *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$. The *uniform additive measure* is the additive measure defined by $\mu(\{i\}) = 1/n$, for $i = 1, \ldots, n$.

Definition 2: Let μ be a fuzzy measure on N, and $x := (x_1, \ldots, x_n) \in \mathbb{R}^n_+$. The *Choquet integral* of x w.r.t. μ is defined by:

$$\mathcal{C}_\mu(x_1, \ldots, x_n) := \sum_{i=1}^{n} (x_{\pi(i)} - x_{\pi(i-1)}) \mu(A_{\pi(i)}) \qquad (1)$$

with π a permutation on N such that $x_{\pi(1)} \leq x_{\pi(2)} \leq \cdots \leq x_{\pi(n)}$, with the convention $x_{\pi(0)} := 0$, and $A_{\pi(i)} := \{\pi(i), \ldots, \pi(n)\}$.
An equivalent expression is

$$\mathcal{C}_\mu(x_1, \ldots, x_n) := \sum_{i=1}^{n} x_{\pi(i)}[\mu(A_{\pi(i)}) - \mu(A_{\pi(i+1)})] \quad (2)$$

with the convention $A_{\pi(n+1)} := \emptyset$. From this formula, we deduce that the Choquet integral reduces to a classical weighted arithmetic mean $\sum_{i=1}^{n} w_i x_i$ when the fuzzy measure is additive, with $w_i := \mu(\{i\})$.

Definition 3: Let μ be a fuzzy measure on N, and $x \in [0,1]^n$. The *Sugeno integral* of x w.r.t. μ is defined by:

$$\mathcal{S}_\mu(x_1, \ldots, x_n) := \bigvee_{i=1}^{n} \left[x_{\pi(i)} \wedge \mu(A_{\pi(i)})\right]$$

with same notations as above.

When these integrals are used to model the relationship between an input vector x and some output y, the parameters of the model are the $2^n - 2$ values of the fuzzy measure μ for all subsets of N, except for $\emptyset$ and N whose values are fixed. The exponential complexity of these models obliges to look for either simpler models or for interpretative tools. The notion of k-additive fuzzy measures [11] provides simpler models, ranging from the purely additive one ($k = 1$), to the general case ($k = n$), and the general notion of interaction [11], closely linked to k-additive fuzzy measures, provides a way to interpret the Choquet integral model. We recall here very briefly the essential notions, since they will be used in the sequel. The interested reader can find more details in e.g., [10], [13], [14], [19].

Definition 4: Let μ be a fuzzy measure on N. The *Möbius transform* of μ is a function $m : 2^N \to \mathbb{R}$ defined by:

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B).$$

The Möbius transform is invertible since one can recover μ from m by:

$$\mu(A) = \sum_{B \subseteq A} m(B).$$

If the fuzzy measure is additive, then m is non null only for singletons, and $m(\{i\}) = \mu(\{i\})$.

Definition 5: A fuzzy measure μ is said to be *k-additive* for some integer $k \in \{1, \ldots, n\}$ if $m(A) = 0$ whenever $|A| > k$, and there exists some A with $|A| = k$ such that $m(A) \neq 0$.
Clearly, 1-additive fuzzy measures are additive measures, and a k-additive fuzzy measure needs only $\sum_{i=1}^{k} \binom{n}{i} - 1$ values to be defined.

Definition 6: Let μ be a fuzzy measure on N. The *Shapley value* of $i \in N$ is defined by:

$$\phi(i) := \sum_{K \subseteq N \setminus i} \frac{(n-k-1)!k!}{n!}[\mu(K \cup i) - \mu(K)]$$

with $k := |K|$.

This notion has been introduced by Shapley [30] in cooperative game theory. It represents the overall importance of attribute i in the model, and it has the property that $\sum_{i=1}^{n} \phi(i) = 1$. For an additive measure, the Shapley value, the Möbius transform and μ coincide in the sense that $\mu(\{i\}) = m(\{i\}) = \phi(i)$, $i = 1, \ldots, n$. It could be said that if $\phi(i) = 1/n$, attribute i is neither important nor unimportant, since the Shapley value of the uniform additive measure satisfies this property for every $i \in N$.

Another useful notion is the notion of interaction between two attributes i, j (originally introduced by Murofushi and Soneda [28] and later generalized to more attributes by Grabisch [8]).

Definition 7: Let μ be a fuzzy measure on N, and $i, j \in N$. The *interaction index* between i and j is defined by:

$$I(i,j) := \sum_{K \subseteq N \setminus \{i,j\}} \frac{(n-k-2)!k!}{(n-1)!}[\mu(K \cup \{i,j\}) - \mu(K \cup i) - \mu(K \cup j) + \mu(K)].$$

If μ is additive, then $I(i,j) = 0$ for all $i, j \in N$. Positive (resp. negative) interaction values represent a kind of positive synergy or complementarity (resp. negative synergy, redundancy) between attributes. Together with the Shapley value, the interaction index is a valuable tool to interpret the model (see for example [15]). It can be proved that if μ is 2-additive, the values $\phi(i)$ and $I(i,j)$ for $i \in N$ and $j \neq i$ uniquely determine μ.

III. Definition of the Empirical Study

The aim of the paper is to study the statistical properties of the Choquet and Sugeno integrals. Two important questions arise in an applied context:

- What is the statistical behavior of $y = \mathcal{C}_\mu(x)$ and $y = \mathcal{S}_\mu(x)$ when input vector x is corrupted with noise whose distribution is known?
- For a given set of learning data supposed to be corrupted with noise whose distribution is known, what are the statistical properties of the parameters of the model, namely μ, for a given learning method?

It is very difficult to give a mathematical answer to these questions, especially the second one; hence our position is to undertake an empirical study. Section IV addresses the first question, while Section V addresses the second one.

We give below some insights to the first question. This will show why a mathematical analysis is difficult to undertake. Let us take the case of the Choquet integral, and consider first an additive fuzzy measure. In this case, the Choquet integral writes:

$$y = \sum_{i=1}^{n} \mu(\{i\}) x_i.$$

Considering $x_1, \ldots, x_n$ as independent random variables with expected values $m_1, \ldots, m_n$, we know that the distribution of y is the convolution product of the distributions of the x_i's, up to multiplicative constants $\mu(\{i\})$. Hence, in the normal case, the result is particularly simple since we know that if $X_1, \ldots, X_n$ are independent Gaussian random variables, with means $m_1, \dots, m_n$ and variances $\sigma_1^2, \dots, \sigma_n^2$, then the distribution of the sum $\alpha_1 X_1 + \cdots + \alpha_n X_n$ is again a Gaussian distribution, with mean and variance given by:

$$m = \alpha_1 m_1 + \cdots + \alpha_n m_n \tag{3}$$
$$\sigma^2 = \alpha_1^2 \sigma_1^2 + \cdots + \alpha_n^2 \sigma_n^2. \tag{4}$$

The additive case is thus solved when the input variables are independent Gaussian. Let us consider the general case where the fuzzy measure is not additive. Rearranging terms in Def. 2, the Choquet integral can be written as in (2). It is known that if $X_1, \dots, X_n$ are identically distributed and independent, with cumulative distribution function $F(t) := P(X \leq t)$, then the distribution function of $X_{\pi(k)}$ (called the kth order statistic) is given by:

$$F_{X_{\pi(k)}}(t) = P(X_{\pi(k)} \leq t) = \sum_{j=k}^{n} \binom{n}{j} F^j(t)(1 - F(t))^{n-j}.$$

Supposing that this result permits to compute the distributions of $X_{\pi(k)}$, $k = 1, \dots, n$, it does not seem obvious to compute y, because the $X_{\pi(k)}$'s are no more statistically independent, and thus the classical results on the sum of independent random variables (e.g., the sum of Gaussian r.v. is still Gaussian, or the p.d.f. of the sum is the convolution product of the p.d.f.'s) do not apply.

We briefly cite the recent result of Marichal concerning the Sugeno integral [26]. Consider n independent random variables $X_1, \dots, X_n$, with cumulative distribution functions $F_1, \dots, F_n$, and a fuzzy measure μ on N. Let H be the Heaviside step function defined by $H(x) := 1$ if $x \geq 0$, and 0 otherwise. For any $c \in \mathbb{R}$, we also introduce $H_c(x) := H(x - c)$. Then the cumulative distribution function of $Y := \mathcal{S}_\mu(X_1, \dots, X_n)$ is given by

$$F(y) = 1 - \sum_{A \subseteq N} \left[1 - H_{\mu(A)}(y)\right] \prod_{i \in N \setminus S} F_i(y) \prod_{i \in S} [1 - F_i(y)]. \tag{5}$$

IV. Statistical study of the output value

In this section, we study the statistical properties of $y = \mathcal{C}_\mu(x)$ and $y = \mathcal{S}_\mu(x)$ when the vector x is corrupted with noise, that is, $x = x_0 + \nu$, with ν a Gaussian white noise with zero mean. A first experiment is done where the vector x_0 has components which are scattered on the range $[0, 1]$, and a second one where on the contrary components of x_0 are all equal.

In both cases, we generate 1000 6-dimensional samples. We consider the three following fuzzy measures:

- An additive measure μ_{add} defined by:

$\mu_{\text{add}}(\{1\})$	$\mu_{\text{add}}(\{2\})$	$\mu_{\text{add}}(\{3\})$
0.0931979	0.0203709	0.276359
$\mu_{\text{add}}(\{4\})$	$\mu_{\text{add}}(\{5\})$	$\mu_{\text{add}}(\{6\})$
0.0824232	0.0631751	0.464474

- A 2-additive measure $\mu_{\text{2-add}}$ defined by:

$\mu_{\text{2-add}}(\{1\})$	$\mu_{\text{2-add}}(\{2\})$	$\mu_{\text{2-add}}(\{3\})$
0.130094	0.061056	0.246516
$\mu_{\text{2-add}}(\{4\})$	$\mu_{\text{2-add}}(\{5\})$	$\mu_{\text{2-add}}(\{6\})$
0.214887	0.400155	0.508465
$\mu_{\text{2-add}}(\{1,2\})$	$\mu_{\text{2-add}}(\{1,3\})$	$\mu_{\text{2-add}}(\{1,4\})$
0.19115	0.376611	0.344981
$\mu_{\text{2-add}}(\{1,5\})$	$\mu_{\text{2-add}}(\{1,6\})$	$\mu_{\text{2-add}}(\{2,3\})$
0.530249	0.508466	0.307572
$\mu_{\text{2-add}}(\{2,4\})$	$\mu_{\text{2-add}}(\{2,5\})$	$\mu_{\text{2-add}}(\{2,6\})$
0.275943	0.400156	0.569521
$\mu_{\text{2-add}}(\{3,4\})$	$\mu_{\text{2-add}}(\{3,5\})$	$\mu_{\text{2-add}}(\{3,6\})$
0.246517	0.727001	0.826465
$\mu_{\text{2-add}}(\{4,5\})$	$\mu_{\text{2-add}}(\{4,6\})$	$\mu_{\text{2-add}}(\{5,6\})$
0.615042	0.7555	0.569521

- A fuzzy measure coming from identification on some real data set, denoted μ_0. We do not display the 64 values of μ_0 since this would take up a lot of room, and be little informative, however, fuzzy measures μ_{add} and $\mu_{\text{2-add}}$ come from the same real data set, and so can be considered respectively as additive and 2-additive approximations of μ_0. This real data set comes from experiments done on subjective evaluation of mental workload [21].

We consider both the Choquet and Sugeno integrals in these two experiments.

A. Experiment 1

We choose $x_0 = (0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1)$, with standard deviation of the noise being successively 0.01, 0.05 and 0.1. Table I in appendix gives the mean and standard deviation of the output $y = \mathcal{C}_\mu(x)$ and $y = \mathcal{S}_\mu(x)$ for the different μ and x defined above. Note that if σ is low, then x has a high probability to satisfy $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5 \leq x_6$, i.e., the same permutation applies for x_0 and x. From (2), it is clear that the Choquet integral reduces to a weighted sum whose weights are:

$$\mu_0(\{1,2,3,4,5,6\}) - \mu_0(\{2,3,4,5,6\}) = 0$$
$$\mu_0(\{2,3,4,5,6\}) - \mu_0(\{3,4,5,6\}) = 0$$
$$\mu_0(\{3,4,5,6\}) - \mu_0(\{4,5,6\}) = 0.2818$$
$$\mu_0(\{4,5,6\}) - \mu_0(\{5,6\}) = 0$$
$$\mu_0(\{5,6\}) - \mu_0(\{6\}) = 0.2588$$
$$\mu_0(\{6\}) = 0.4594.$$

B. Experiment 2

We choose $x_0 = (0.5,\ 0.5,\ 0.5,\ 0.5,\ 0.5,\ 0.5)$, with standard deviation of the noise being successively 0.01, 0.05 and 0.1. Table II in appendix gives the mean and standard deviation of the output $y = \mathcal{C}_\mu(x)$ and $y = \mathcal{S}_\mu(x)$ for the different μ and x defined above.

C. Interpretation and comments

1) Choquet integral: For the additive case, it is possible to compute the theoretical mean and variance of the Choquet integral thanks to (3) and (4). Thus the mean is nothing else that the value of the Choquet integral of x_0 without noise (theoretical value in the tables). For both Experiments 1 and 2, these theoretical values are recovered with great precision (error less than 1%).

The variance writes in our case:

$$\sigma_y^2 = \sigma^2(0.0931979^2+0.0203709^2+0.276359^2+0.0824232^2 + 0.0631751^2 + 0.464474^2) = 0.3119959\sigma^2.$$

We obtain the following theoretical values for the standard deviation:

standard deviation of x	0.01	0.05	0.1
standard deviation of y	0.0055857	0.027928	0.055857

Again, we remark that the theoretical values are recovered with great precision in both experiments in the case of the Choquet integral w.r.t. an additive measure.

Strangely enough, even if the fuzzy measure is no more additive, the values for mean and standard deviation remain very similar to the theoretical values of the additive case. This curious result may come from the fact that fuzzy measures $\mu_{\text{2-add}}$ and μ_{add} are in a sense approximations of μ_0, as explained above.

Fig. 1 and 2 show some histograms of y for Experiments 1 and 2. We observe that the histograms are close to a Gaussian distribution, especially for low values of σ and for Experiment 1. This is natural, since in this case the Choquet integral reduces to a weighted sum whose weights are given above (Section IV-A), so that the variance of y is

$$\begin{aligned}\sigma_y^2 &= \sigma^2(0.2818^2 + 0.2588^2 + 0.4594^2)\\ &= 0.3574\sigma^2,\end{aligned}$$

and the standard deviation is $\sigma_y = 0.5979\sigma$. This result is very close to the one obtained in Table I.

For Experiment 2, the shape of the distribution is flatter than a Gaussian distribution.

2) Sugeno integral: Here results are more difficult to interpret. The mean value of y is stable, however less than for the Choquet integral, and the standard deviation is approximately proportional to the standard deviation of the noise in case of Experiment 2, but not for Experiment 1.

Fig. 3 and 4 show some histograms of y for Experiments 1 and 2. The Sugeno integral gives rise to very various forms for the histograms. While Experiment 2 shows histograms similar to a Gaussian density, Experiment 1 shows rather curious shapes, but which can be explained for low values of variance. Indeed, in this case the resulting random variable is approximated by, using Def. 2 and values of μ_0:

$$\begin{aligned}Y = \quad & (X_1 \wedge 1) \vee (X_2 \wedge 1) \vee (X_3 \wedge 1) \vee (X_4 \wedge 0.7181)\\ & \vee (X_5 \wedge 0.7181) \vee (X_6 \wedge 0.4594)\\ \approx \quad & X_1 \vee X_2 \vee X_3 \vee [(X_4 \vee X_5) \wedge 0.7181] \vee 0.4594\end{aligned}$$

since the X_0's are centered on 0, 0.2, 0.4, 0.6, 0.8 and 1 respectively, and they have a low variance. Simplifying further, we get:

$$y \approx X_5 \wedge 0.7181$$

which well corresponds to the histogram (see Figure 3 left. For $\sigma = 0.01$, we have seen that the histogram has only one slot in 0.7181).

V. Statistical study of the parameters of the model

A. Description of the experiment

We begin by general considerations. Let $\mathcal{X} := \{x_1^l, \dots, x_n^l; y^l\}_{l \in L}$ be a set of learning data we have at disposal for the identification of the model (in this section we restrict to the case of the Choquet integral), and some given learning procedure M. We obtain after learning a fuzzy measure $\mu_{\mathcal{X},M}$, which represents the set of all parameters of the model. An important question in practice is the robustness of the model, specifically:

> *If a Gaussian noise with zero mean is added to the learning data, what is the influence on the parameters of the model?*

The answer depends on the kind of learning procedure which is used. The most commonly used learning procedures for the Choquet integral are the following:

1) Heuristic Least Mean Squares (HLMS) [7]. This algorithm uses as basic idea the gradient algorithm to minimize the sum of squared errors, under the constraint of monotonicity of the fuzzy measure. The algorithm is not optimal, but is very fast and uses few memory. It has also the property of giving a fuzzy measure as close as possible to the uniform additive fuzzy measure $\mu(\{i\}) = 1/n$, $i = 1, \dots, n$. We use here an improved version of HLMS (described in appendix). Values of parameters for this method are $\alpha = 0.01$ (coefficient of the gradient) and 300 iterations.
2) Quadratic programming (QUAD). The error criterion is again the sum of squared errors, under the constraint of monotonicity of the fuzzy measure. This leads to a quadratic program with linear constraints, and gives a (non unique) optimal solution.
3) k-additive quadratic programming (k-ADD). It is the same as above, but the model is supposed to be a Choquet integral w.r.t. a k-additive fuzzy measure. In this experiment, we use $k = 2$ (2-ADD).

For a careful study of the methods based on quadratic learning, the reader is referred to [27].

We describe our experimental setting and define more precisely our aims. We consider a theoretical model (the Choquet integral) with 4 attributes, built with several different fuzzy measures, and the identification of these models will be done with the above mentionned learning procedures, and also several data sets, each of them illustrating a different practical situation. In order to master the whole set of data of the study, we use only synthetic data.

We consider the following three fuzzy measures:

- an additive measure μ_{add} defined by

$\mu_{\text{add}}(\{1\})$	$\mu_{\text{add}}(\{2\})$	$\mu_{\text{add}}(\{3\})$	$\mu_{\text{add}}(\{4\})$
0.5	0.3	0.15	0.05

- a 2-additive measure $\mu_{\text{2-add}}$ defined by the following Shapley value and interactions (this determines uniquely the fuzzy measure, see Section II):

i	$\phi(i)$
1	0.5
2	0.3
3	0.15
4	0.05

i,j	$I(i,j)$	i,j	$I(i,j)$
1,2	0.2	2,3	0.1
1,3	0	2,4	0.1
1,4	0	3,4	0

This gives the following values for $\mu_{\text{2-add}}$:

$\mu_{\text{2-add}}(\{1\})$	$\mu_{\text{2-add}}(\{2\})$	$\mu_{\text{2-add}}(\{3\})$
0.4	0.1	0.1
$\mu_{\text{2-add}}(\{4\})$	$\mu_{\text{2-add}}(\{1,2\})$	$\mu_{\text{2-add}}(\{1,3\})$
0	0.7	0.5
$\mu_{\text{2-add}}(\{1,4\})$	$\mu_{\text{2-add}}(\{2,3\})$	$\mu_{\text{2-add}}(\{2,4\})$
0.4	0.3	0.2
$\mu_{\text{2-add}}(\{3,4\})$	$\mu_{\text{2-add}}(\{1,2,3\})$	$\mu_{\text{2-add}}(\{1,2,4\})$
0.1	0.9	0.8
$\mu_{\text{2-add}}(\{1,3,4\})$	$\mu_{\text{2-add}}(\{2,3,4\})$	$\mu_{\text{2-add}}(\{1,2,3,4\})$
0.5	0.4	1

- a general (4-additive) fuzzy measure μ_0 defined by the following Möbius transform:

$m_0(\{1\})$	$m_0(\{2\})$	$m_0(\{3\})$
0.4	0.1	0.1
$m_0(\{4\})$	$m_0(\{1,2\})$	$m_0(\{1,3\})$
0	0.1	0
$m_0(\{1,4\})$	$m_0(\{2,3\})$	$m_0(\{2,4\})$
0	-0.1	0.1
$m_0(\{3,4\})$	$m_0(\{1,2,3\})$	$m_0(\{1,2,4\})$
0	0	0
$m_0(\{1,3,4\})$	$m_0(\{2,3,4\})$	$m_0(\{1,2,3,4\})$
0	0.2	0.2

This gives the following values for μ:

$\mu_{\text{2-add}}(\{1\})$	$\mu_{\text{2-add}}(\{2\})$	$\mu_{\text{2-add}}(\{3\})$
0.4	0.1	0.1
$\mu_{\text{2-add}}(\{4\})$	$\mu_{\text{2-add}}(\{1,2\})$	$\mu_{\text{2-add}}(\{1,3\})$
0	0.6	0.5
$\mu_{\text{2-add}}(\{1,4\})$	$\mu_{\text{2-add}}(\{2,3\})$	$\mu_{\text{2-add}}(\{2,4\})$
0.4	0.1	0.2
$\mu_{\text{2-add}}(\{3,4\})$	$\mu_{\text{2-add}}(\{1,2,3\})$	$\mu_{\text{2-add}}(\{1,2,4\})$
0.1	0.6	0.7
$\mu_{\text{2-add}}(\{1,3,4\})$	$\mu_{\text{2-add}}(\{2,3,4\})$	$\mu_{\text{2-add}}(\{1,2,3,4\})$
0.5	0.2	1

The number of free parameters of these three models is respectively 3, 9, and 14.

We consider 9 learning data sets, which we describe below, illustrating three typical situations in practice:

- 100 data $(x^l; y^l)$, with vectors x^l uniformly distributed on $[0,1]^4$, and $y^l = \mathcal{C}_\mu(x^l)$, with μ being one of the three fuzzy measures above defined. We denote by $\mathcal{X}^{100}_{\mu_{\text{add}}}$, $\mathcal{X}^{100}_{\mu_{\text{2-add}}}$ and $\mathcal{X}^{100}_{\mu_0}$ these 3 learning data sets.
  Since data are in sufficient number regarding the number of parameters and uniformly distributed in the whole space, this is an ideal situation for learning, which is not always encountered in practice. We call it Situation S1.
- same as above, but the number of learning data is only 10. We denote these learning data sets by $\mathcal{X}^{10}_{\mu_{\text{add}}}$, $\mathcal{X}^{10}_{\mu_{\text{2-add}}}$ and $\mathcal{X}^{10}_{\mu_0}$.
  The aim is to study the behavior of learning procedures with very few learning data, but still uniformly distributed in the space. We call it Situation S2.
- we select in $\mathcal{X}^{100}_{\mu_{\text{add}}}$, $\mathcal{X}^{100}_{\mu_{\text{2-add}}}$ and $\mathcal{X}^{100}_{\mu_0}$ the data such that $x^l_1 \leq x^l_4$, which gives about 50 data. We denote by $\mathcal{X}^{1\leq 4}_{\mu_{\text{add}}}$, $\mathcal{X}^{1\leq 4}_{\mu_{\text{2-add}}}$ and $\mathcal{X}^{1\leq 4}_{\mu_0}$ these learning data sets.
  This data set illustrates the situation where there are *apparently* enough data regarding the number of parameters, but there is no data in some part of the space, which makes difficult the learning of some parameters. We call this situation S3.
  In our case, imposing $x^l_1 \leq x^l_4$ amounts to cancel all data which would allow the identification of the values of $\mu(A)$ such that $A \ni 1$ but $A \not\ni 4$, i.e., these are $\mu(1)$, $\mu(1,2)$, $\mu(1,3)$ and $\mu(1,2,3)$. Remark that these values are the highest in the three fuzzy measures defined above. This means that we have learning data with a considerable lack of information.

As said above, we use the following learning procedures: HLMS, QUAD and 2-ADD. We have verified beforehand that the above defined learning data sets (without noise) permit a perfect identification of the 3 fuzzy measures, up to a precision of 10^{-6} for QUAD, and 10^{-2} for HLMS.

For each learning data set and each learning procedure, we add to the data a Gaussian noise with zero mean and standard deviation being successively 0.01, 0.02, and 0.05, and we have done 100 realizations of each data set. We try to answer the following questions for each situation S1, S2 or S3:

What is the behavior of the estimation of each parameter, described in terms of bias and standard deviation? What is the best learning procedure, i.e., with minimum bias and standard deviation?

The results shown are:

- The average bias and average standard deviation of the values of the fuzzy measure, denoted by b_μ, σ_μ, of the Shapley value, denoted by b_ϕ, σ_ϕ, and of the interaction, denoted by b_I, σ_I, obtained on 100 realizations. All these results are given in Tables III to XI.
- In the case of incomplete data sets $\mathcal{X}^{1\leq 4}_{\mu_{\text{add}}}$ and $\mathcal{X}^{1\leq 4}_{\mu_0}$ the values of the fuzzy measure found by the different learning procedures, together with their Shapley values and interactions (Tables XII and XIII). For this case, we have taken $\sigma = 0.01$.

Caution: Shapley values are multiplied by n (hence by 4) in tables, so that a value of 1 indicates a neutral value for overall importance (see Section II). Consequently, for comparing standard deviations of the Shapley value with those of interactions, it is necessary to divide the standard deviation of the Shapley value by 4.

B. Results and comments

We address first Tables III to XI.

- The computation of the overall bias on μ, ϕ, I being done by an arithmetic mean of the biases on each value of μ, ϕ, I, the bias on ϕ is always close to 0, since the sum of the Shapley values is equal to 1. This figure being not significant, we discard it from our analysis.

- Generally speaking, the bias on μ remains very low, and can be considered as being 0, even with a low number of data ($\mathcal{X}^{10}$). However, in the case of learning data which do not cover the whole space $[0,1]^4$ (Tables IX to XI), a non negligible negative bias may occur, in particular for QUAD.

  There is no clear linearity relation between the bias and σ: although the bias is generally increasing with the level of noise, in many cases, it is observed that the bias is smaller for higher levels of noise.

  More or less the same conclusions hold for the bias of the interaction.
- Results are easier to interpret for the standard deviation of μ. For Situations S1 and S2 (that is, with uniform distribution of the data), Tables III to V and VI to VIII show that σ_μ is approximately proportional to σ (although more experiments should be done to confirm this), and is relatively independent of μ, especially for $\mathcal{X}^{10}$. A linear regression $\sigma_\mu = \alpha\sigma$ done for each learning procedure and each situation, and using results for all three fuzzy measures gives the following result:

		α	residual error
Situation S1	HLMS	1.04108	0.003514
	QUAD	0.86981	0.00301
	2-ADD	0.65398	0.001123
Situation S2	HLMS	1.63705	0.004992
	QUAD	2.5612	0.02488
	2-ADD	1.8618	0.01198

  The above table clearly shows that in the case of a large number of uniformly distributed data ($\mathcal{X}^{100}$), 2-ADD, and QUAD in second position give the best results, always better than HLMS. By contrast, when the number of data is insufficient ($\mathcal{X}^{10}$), HLMS gives always the best results, and QUAD the worst ones.

  This tendency is even stronger with $\mathcal{X}^{1\leq 4}$ (see Tables IX to XI), where QUAD and 2-ADD give similar results, far worse than HLMS (for these cases, since σ_μ is no more independent of μ, we did not performed any linear regression).

  Lastly, remark that the linear hypothesis is questionable for Situation S2 and QUAD, 2-ADD.
- The same phenomena can be observed on the standard deviation on the Shapley value, even in a more noticeable way. For the interaction, the behavior of its standard deviation is the same as the one of μ.

Lastly, we comment on Tables XII and XIII, illustrating Situation S3. The effect of a lack of learning data in a half-space can be well observed there. As explained above, this lack forbids a correct estimation of $\mu(\{1\})$, $\mu(\{1,2\})$, $\mu(\{1,3\})$, and $\mu(\{1,2,3\})$. This is exactly what can be observed: these values are very badly estimated, and *only these ones*.

The consequence is that the Shapley values are more or less heavily perturbed. Indeed, in the computation of $\phi(i)$, all terms $\mu(A \cup i) - \mu(A)$, $A \subseteq N \setminus i$, are used. It is easy to see that in our case, for $\phi(1)$ and $\phi(4)$, there are 4 terms among 8 which are spoilt, while there are only 2 among 8 for $\phi(2)$ and $\phi(3)$. This can be well observed on Table XIII (remark that Table XII is not significant for this case since the measure is additive, so that $\phi(i) = \mu(\{i\})$). For the computation of I, the perturbation is even stronger and tends to spread over all values.

Let us come back to the estimation of $\mu(\{1\})$, $\mu(\{1,2\})$, $\mu(\{1,3\})$ et $\mu(\{1,2,3\})$. The results that are closest to the true value are given by HLMS and 2-ADD, while QUAD gives erratic results.

The good performance of HLMS in Situation S3 (and also S2) is explained by the fact that, in the absence of information for values $\mu(A)$ for some subsets $A \subseteq N$, HLMS assign values for those $\mu(A)$'s which are as close as possible (under monotonicity constraints) to the uniform additive measure, i.e., $\mu(A) = \frac{|A|}{n}$. This can be checked in Tables XII and XIII.

VI. Conclusion

The following fundamental conclusions can be drawn from the different experimentations:

- The exact distribution of $y = \mathcal{C}_\mu(x)$ and $y = \mathcal{S}_\mu(x)$ seem to be very complicated to obtain, even if for the latter case, an analytical expression exists. Empirically, it is observed that if x follows a Gaussian distribution, then in general the distribution of y is similar to a Gaussian one (unimodal and symmetric). For the Sugeno integral, the results are more unpredictable, a peak or a Gaussian distribution can be obtained.
- The quality of the learning data set is of considerable influence on the quality of the identification. If the number of data is sufficient with respect to the number of parameters of the model, *and* if they are uniformly distributed in $[0,1]^n$ (Situation S1), then the optimal methods QUAD, 2-ADD give the best results in term of robustness. The standard deviation on μ is roughly proportional to the standard deviation σ of the noise. By contrast, the value of the bias is not clearly related to σ, but remains very low.

  If few learning data are available, but still uniformly distributed (Situation S2), HLMS becomes slightly better than 2-ADD, and significantly better than QUAD. Conclusions on the relation between the standard deviation on μ, the bias and σ remain the same than in Situation S1.

  If some parts of $[0,1]^n$ are not covered (Situation S3), then HLMS gives the most reliable results, then 2-ADD in second position since this method uses a reduced number of parameters. In this case, which is in practice the most frequent one, it seems better to avoid the use of QUAD.

We mention that the above learning procedures, as well as tools to compute the Möbius transform, interaction indices, Shapley values, etc., are all available in the free Kappalab package [17], a package running under the R environment for statistics (see http://www.polytech.univ-nantes.fr/kappalab).

As a final remark, we would like to stress the fact that this work is a first step towards a more complete and theoretical analysis of the statistical properties of these methods, now more and more used in practical applications. In particular, the least square estimation of fuzzy measures can be seen as an extension of classical multiple linear regression estimation, where classical results could be applied here with benefit.

VII. ACKNOWLEDGMENTS

We address all our thanks to INRS (Institut National de Recherche et de Sécurité) for supporting this study (contract N° 5033704). We thank one of the anonymous referees for providing the theoretical result about the Sugeno integral.

REFERENCES

[1] S. Auephanwiriyakul, J. Keller, and P. Gader. Generalized Choquet fuzzy integral fusion. *Information Fusion*, 3:69–85, 2002.

[2] G. Büyüközkan, G. Mauris, O. Feyzioğlu, and L. Berrah. Providing elucidations of web site evaluation based on a multi-criteria aggregation with the Choquet integral. In *Int. Fuzzy Systems Association World Congress (IFSA 2003)*, pages 131–134, Istanbul, Turkey, June 2003.

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

[4] D. Dubois, J.-L. Marichal, H. Prade, M. Roubens, and R. Sabbadin. The use of the discrete Sugeno integral in decision making: a survey. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(5):539–561, 2001.

[5] P. Gader, W. H. Lee, and X. P. Zhang. Rényi entropy with respect to choquet capacities. In *Proc. IEEE Conf. Fuzzy Systems*, Budapest, Hungary, July 2004.

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

[7] M. Grabisch. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *Int. Joint Conf. of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symposium*, pages 145–150, Yokohama, Japan, March 1995.

[8] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. In *4th Int. Conf. on Soft Computing*, pages 316–319, Iizuka, Japan, September 1996.

[9] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European J. of Operational Research*, 89:445–456, 1996.

[10] M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems*, 5:587–607, 1997.

[11] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.

[12] M. Grabisch. Fuzzy integral for classification and feature extraction. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 415–434. Physica Verlag, 2000.

[13] M. Grabisch. The interaction and Möbius representations of fuzzy measures on finite spaces, k-additive measures: a survey. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 70–93. Physica Verlag, 2000.

[14] M. Grabisch. Set function over finite sets: transformations and integrals. In E. Pap, editor, *Handbook of Measure Theory*, pages 1381–1401. Elsevier Science Publ., 2002.

[15] M. Grabisch, J. Duchêne, F. Lino, and P. Perny. Subjective evaluation of discomfort in sitting position. *Fuzzy Optimization and Decision Making*, 1(3):287–312, 2002.

[16] M. Grabisch and F. Huet. Texture recognition by Choquet integral filters. In *6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, pages 1325–1330, Granada, Spain, July 1996.

[17] M. Grabisch, I. Kojadinovic, and P. Meyer. Using the Kappalab R package for Choquet integral based multi-attribute utility theory. In *Proc. Int. Conf. on Information Processing and Management of Unicertainty (IPMU'06)*, pages 1702–1709, Paris, France, July 2006.

[18] M. Grabisch and Ch. Labreuche. Fuzzy measures and integrals in MCDA. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis*, pages 563–608. Kluwer Academic Publishers, 2005.

[19] M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent representations of set functions. *Mathematics of Operations Research*, 25(2):157–178, 2000.

[20] M. Grabisch, H.T. Nguyen, and E.A. Walker. *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference*. Kluwer Academic, 1995.

[21] M. Grabisch, H. Prade, E. Raufaste, and P. Terrier. Application of the Choquet integral to the evaluation of subjective mental workload. In *9th IFAC Symposium on Automated Systems Based on Human Skill And Knowledge*, Nancy, France, may 2006.

[22] M. Grabisch and M. Schmitt. Mathematical morphology, order filters and fuzzy logic. In *Int. Joint Conf. of the 4th IEEE Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symposium*, pages 2103–2108, Yokohama, Japan, March 1995.

[23] J. M. Keller, P. D. Gader, and A. K. Hocaoğlu. Fuzzy integrals in image processing and recognition. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 435–466. Physica Verlag, 2000.

[24] S. H. Kwon and M. Sugeno. A hierarchical subjective evaluation model using non-monotonic fuzzy measures and the Choquet integral. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, *Fuzzy Measures and Integrals — Theory and Applications*, pages 375–391. Physica Verlag, 2000.

[25] J.-L. Marichal. On Sugeno integral as an aggregation function. *Fuzzy Sets and Systems*, 114:347–365, 2000.

[26] J.-L. Marichal. Cumulative distribution functions and moments of weighted lattice polynomials. In *27th Linz Seminar on Fuzzy Set Theory*, pages 95–100, Linz, Austria, Februray 2006.

[27] P. Miranda and M. Grabisch. Optimization issues for fuzzy measures. *Int. J. of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 7(6):545–560, 1999.

[28] T. Murofushi and S. Soneda. Techniques for reading fuzzy measures (III): interaction index. In *9th Fuzzy System Symposium*, pages 693–696, Sapporo, Japan, May 1993. In Japanese.

[29] T. Murofushi and M. Sugeno. Some quantities represented by the Choquet integral. *Fuzzy Sets & Systems*, 56:229–235, 1993.

[30] L. S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in Annals of Mathematics Studies, pages 307–317. Princeton University Press, 1953.

[31] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[32] A. Verkeyn, D. Botteldooren, B. De Baets, and G. De Tré. Sugeno integrals for the modelling of noise annoyance aggregation. In T. Bilgiç, B. De Baets, and O. Kaynak, editors, *Fuzzy Sets and Systems — IFSA 2003*, pages 277–284. Springer Verlag, 2003.

[33] X. Z. Wang and J. F. Chen. Multiple neural networks fusion model based on choquet fuzzy integral. In *Proc. 3d Int. Conf. Machine Learning and Cybernetics*, Shanghai, August 2004.

[34] Z. Wang, K. S. Leung, M. L. Wong, J. Fang, and K. Xu. Nonlinear nonnegative multiregressions based on Choquet integrals. *Int. J.of Approximate Reasoning*, 25:71–87, 2000.

APPENDIX

An improvement of HLMS

The HLMS algorithm for the identification of a fuzzy measure when the model is a Choquet integral has been proposed by Grabisch in [7]. We propose here a simplification of it, which leads to better performance. We first recall the basic ideas of HLMS. Consider a learning datum $(x_1, \dots, x_n; y)$. We compute the model error $e = \mathcal{C}_\mu(x) - y$. Let us denote by $u(0), u(1), \dots, u(n)$ the values of the fuzzy measure μ used in the computation of $\mathcal{C}_\mu(x)$. The new values are computed as follows (gradient algorithm):

$$u^{\text{new}}(i) = u^{\text{old}}(i) - \alpha \frac{e}{e_{\max}} (x_{\pi(n-i)} - x_{\pi(n-i-1)}), \qquad (6)$$

where $\alpha \in [0, 1]$ is a constant, and $e_{\max}$ is the maximum value of the error. $e_{\max} = 1$ if y takes its values in $[0, 1]$, and π is a permutation on N as in Def. 2.

The problem is then to ensure the monotonicity of μ, while modification of its values are done. In the original HLMS algorithm, this was done in two steps. In a first step, for each datum, after application of Eq. (6), verification of monotonicity was done only for values of μ already modified (for previous data). Then in a second step, when all data have ben used, unmodified values of μ are checked to see if they satisfy monotonicity, and they are modified accordingly. Then a kind of uniformization is performed on the whole set of unmodified values of μ.

It has been observed in experiments that is some rare cases, monotonicity may be violated when HLMS is used. This is due to the intricate way of modifying the values of μ. The new version of the algorithm avoids this drawback, and is much simpler (particularly in step 2). We describe it below.

- **step 0 :** the fuzzy measure is initialized at the uniform additive measure.
- **step 1.1 :** consider a learning datum $(x_1, \ldots, x_n; y)$, inducing a permutation π. Compute $e = \mathcal{C}_\mu(x) - y$, and denote by $u(0) := \mu(\emptyset) = 0, \ldots, u(i) := \mu(A_{\pi(n-i+1)}), \ldots, u(n) := \mu(N) = 1$ the values of the fuzzy measure μ used in the computation of $\mathcal{C}_\mu(x)$.
- **step 1.2 :** compute the new value $u^{\text{new}}(i)$ by Eq. (6).
- **step 1.3 :** verify the monotonicity relations. If $e > 0$, the verification is done for all lower neighbors, i.e., all $\mu(A_{\pi(n-i+1)} \setminus j)$ for all $j \in A$, and if $e < 0$, for all upper neighbors, i.e., all $\mu(A_{\pi(n-i+1)} \cup j)$ for all $j \in N \setminus A$. If a monotonicity relation is violated, say with $\mu(K)$, then $u(i) = \mu(K)$.

  Repeat steps 1.2 and 1.3 for $i = 1, \ldots, n-1$, in the following order:
  - if $e > 0$, we begin by $u(1), u(2), \ldots, u(n-1)$
  - if $e < 0$, we begin by $u(n-1), u(n-2), \ldots, u(1)$

  Repeat steps 1.1 to 1.3 for all learning data.
- **step 2:** for every $\mu(K)$ left unmodified in step 1 (scanning begins by singletons, then pairs, etc.), adjust its value considering the values of its upper and lower neighbors, in order to homogenize the whole. This is done by computing the minimum distance between $\mu(K)$ and its upper (resp. lower) neighbors, denoted $\overline{d}_{\min}$, (resp. $\underline{d}_{\min}$). Then

$$\mu^{\text{new}}(K) := \mu^{\text{old}}(K) + \frac{\overline{d}_{\min} - \underline{d}_{\min}}{2}.$$

Step 1 and 2 form one iteration, and they can be repeated in the same order.

APPENDIX

Tables and Figures

	μ_{add}			$\mu_{2-\text{add}}$			μ_0		
	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$
Choquet integral									
theoretical value of y	0.679086			0.779156			0.778908		
mean value of y	0.679012	0.678877	0.674190	0.778995	0.779147	0.775559	0.778984	0.778470	0.775214
standard deviation of y	0.005555	0.028340	0.054858	0.005924	0.029583	0.056919	0.006001	0.030522	0.058501
Sugeno integral									
theoretical value of y	0.600000			0.600000			0.718178		
mean value of y	0.599739	0.586059	0.577585	0.600681	0.607169	0.620187	0.718178	0.716494	0.707472
standard deviation of y	0.008676	0.028394	0.036368	0.010102	0.037656	0.061997	0.000000	0.008465	0.028723

TABLE I

STATISTICS OF THE OUTPUT VALUE OF CHOQUET AND SUGENO INTEGRALS

	μ_{add}			$\mu_{2-\text{add}}$			μ_0		
	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$
Choquet integral : theoretical value of y = 0.5									
mean value of y	0.500125	0.500963	0.497570	0.503197	0.516222	0.530218	0.502822	0.514508	0.527123
standard deviation of y	0.005694	0.027271	0.055070	0.005166	0.025244	0.049286	0.005232	0.025009	0.048361
Sugeno integral: theoretical value of y=0.5									
mean value of y	0.500243	0.500973	0.499047	0.503594	0.514922	0.525977	0.502231	0.509829	0.517041
standard deviation of y	0.007088	0.030868	0.055203	0.006528	0.029340	0.052684	0.006067	0.026390	0.049715

TABLE II

STATISTICS OF THE OUTPUT VALUE OF CHOQUET AND SUGENO INTEGRALS

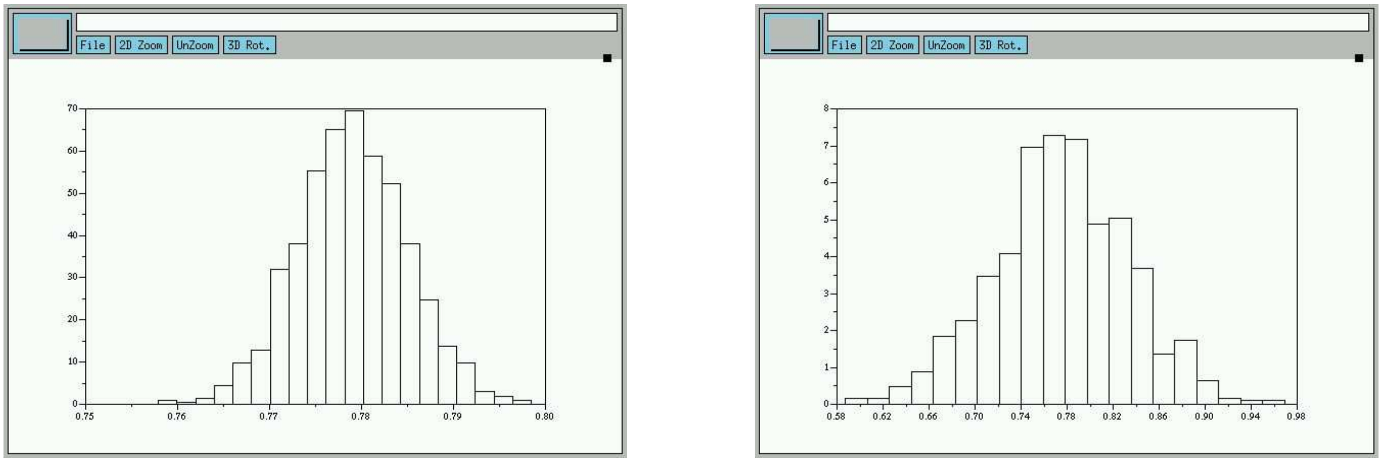

Fig. 1. Histogram of $y = \mathcal{C}_\mu(x)$ for Experiment 1, μ_0 and $\sigma = 0.01$ (left), $\sigma = 0.1$ (right)

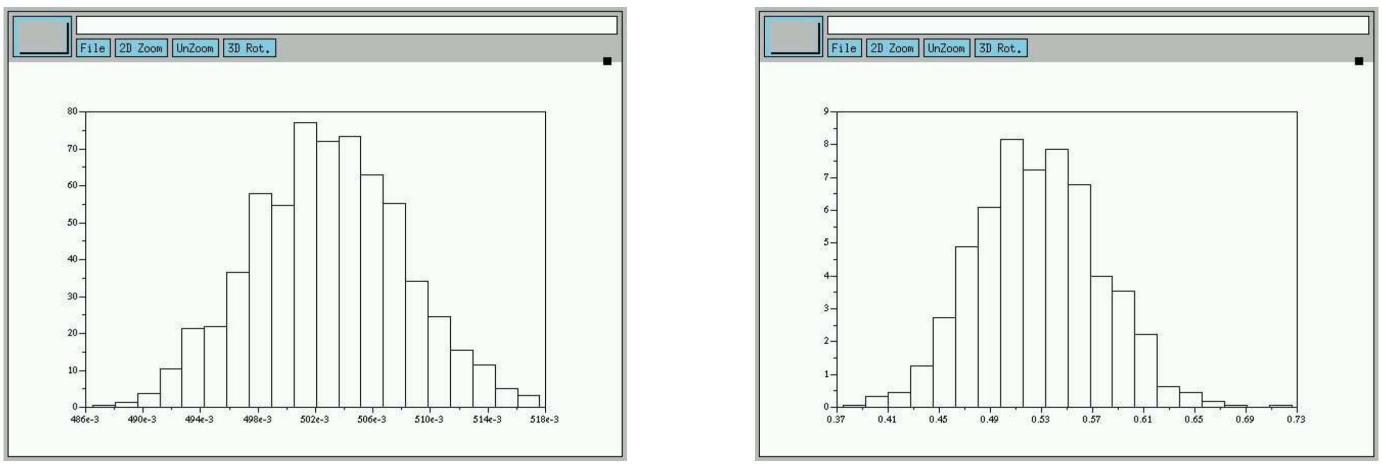

Fig. 2. Histogram of $y = \mathcal{C}_\mu(x)$ for Experiment 2, μ_0 and $\sigma = 0.01$ (left) and $\sigma = 0.1$ (right)

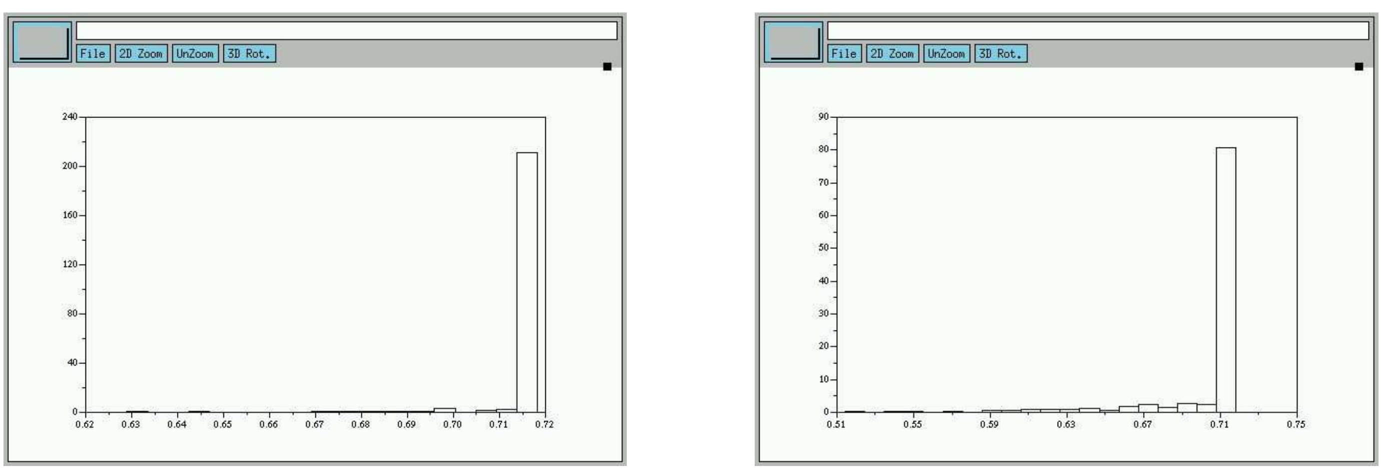

Fig. 3. Histogram of $y = \mathcal{S}_\mu(x)$ for Experiment 1, μ_0 and $\sigma = 0.05$ (left) and $\sigma = 0.1$ (right)

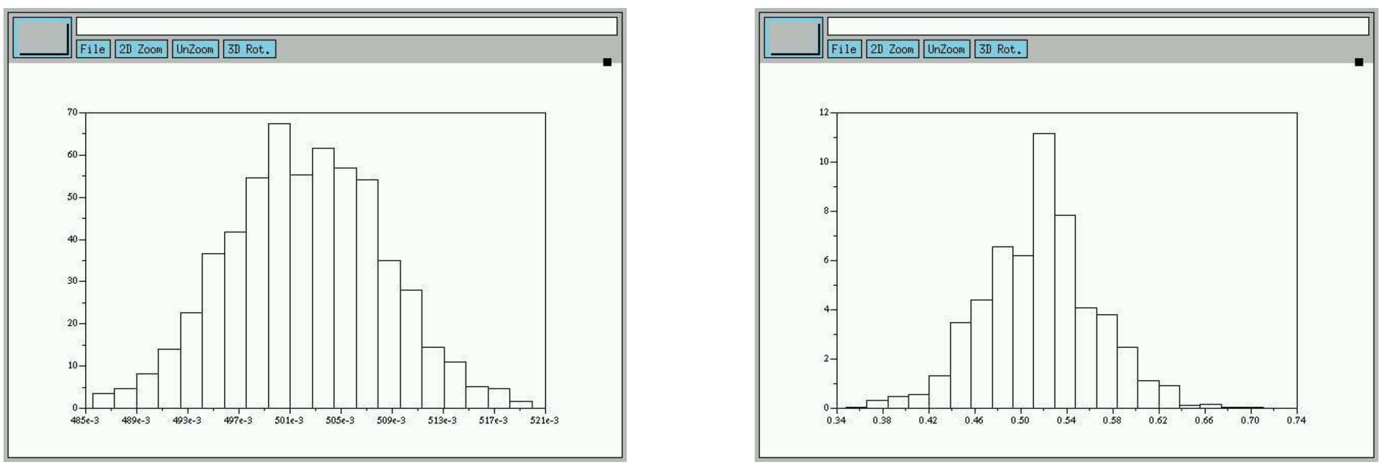

Fig. 4. Histogram of $y = \mathcal{S}_{\mu}(x)$ for Experiment 2, μ_0 and $\sigma = 0.01$ (left) and $\sigma = 0.1$ (right)

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	0.0002359	-0.0001914	0.0008443	0.0005501	-0.0002410	0.0014112	0.0000685	-0.0002020	0.0006598
b_ϕ	-0.0000225	-0.0000125	0.0000025	0.0000025	-0.0000125	0.000005	0.0000025	-0.0000075	-0.00001
b_I	0.0001983	0.0013167	0.0004183	0.0000383	0.0010817	-0.0003783	-0.0000633	0.0001333	-0.00044
σ_μ	0.0150327	0.0256260	0.0515991	0.0116519	0.0222456	0.0459522	0.0075573	0.0152102	0.0337329
σ_ϕ	0.0238063	0.0443845	0.1000121	0.0177995	0.0340447	0.0752618	0.0149487	0.0335126	0.0721673
σ_I	0.0168865	0.0275741	0.0576677	0.0131208	0.0252633	0.0486643	0.0116452	0.0223630	0.0489726

TABLE III

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{100}_{\mu_{\text{add}}}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	0.0010217	0.0018660	0.0064667	0.0004332	0.0011473	0.0030560	0.0004663	-0.0001632	0.0042076
b_ϕ	-0.0000375	0.000015	0.0000125	0.00001	0.0000225	0.000005	-0.000005	0.00001	-0.000005
b_I	0.00066	0.0016067	0.003075	-0.00033	0.0001567	-0.0016917	-0.0001083	0.0001283	-0.0014267
σ_μ	0.0134816	0.0254956	0.0515398	0.0101489	0.0198080	0.0425334	0.0065605	0.0131913	0.0325438
σ_ϕ	0.0222527	0.0422767	0.0972430	0.0170127	0.0317112	0.0766498	0.0150386	0.0292226	0.0761997
σ_I	0.0128859	0.0286317	0.0577767	0.0115245	0.0225192	0.0488963	0.0088639	0.0182380	0.0454115

TABLE IV

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{100}_{\mu_{\text{2-ADD}}}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	0.0005083	0.0015232	0.0091191	0.0001542	0.0006519	0.0046949	0.0010867	0.0013311	0.0039621
b_ϕ	-0.0000075	-0.00002	0.0000275	-2.776E-17	-2.776E-17	0.000005	0.00001	-0.0000025	0.0000125
b_I	-0.00045	-0.0016867	-0.0047	-0.0005217	-0.00107	-0.0043833	-0.0132233	-0.013405	-0.0151533
σ_μ	0.0112302	0.0211791	0.0473869	0.0086041	0.0166677	0.0385111	0.0064366	0.0121429	0.0311108
σ_ϕ	0.0223658	0.0429352	0.0955570	0.0157758	0.0298892	0.0643691	0.0147096	0.0272903	0.0696207
σ_I	0.0119700	0.0251895	0.0540117	0.0098140	0.0194602	0.0454427	0.0097146	0.0188734	0.0483268

TABLE V

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{100}_{\mu_0}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	-0.0202491	-0.0184064	-0.0105161	-0.0102117	-0.0167357	-0.0070003	-0.0007447	0.0000908	-0.0034839
b_ϕ	-0.0000325	-0.00001	0.00002	0.00001	-0.0000025	- 0.000005	0.000015	0.000015	-0.0000025
b_I	0.0024567	0.00284	0.0019733	0.0021733	0.0122533	0.00222	0.0004933	-0.0000617	0.0023133
σ_μ	0.0190233	0.0378587	0.0851044	0.0654708	0.0826742	0.1275689	0.0343353	0.0556703	0.0931236
σ_ϕ	0.0501059	0.0942666	0.2098329	0.1073539	0.1368923	0.2567595	0.0792367	0.1263980	0.2468339
σ_I	0.0222792	0.0462891	0.0969113	0.0741965	0.0945179	0.1463481	0.0522507	0.0783139	0.1161039

TABLE VI

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{10}_{\mu_{\mathrm{add}}}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	0.0012531	0.0030217	0.0162242	-0.0031780	-0.0032977	-0.0019243	0.0013517	0.0059293	0.0084196
b_ϕ	0.0000125	0.0000325	0.000025	-1.388E-17	0.0000075	0.0000075	-0.0000125	-0.000015	0.0000025
b_I	-0.0007267	-0.0002933	-0.0076933	0.0055483	0.00242	-0.0030683	-0.0009167	-0.003965	-0.0056233
σ_μ	0.0199690	0.0393712	0.0754240	0.0521809	0.0743177	0.1140825	0.0297983	0.0506263	0.0907867
σ_ϕ	0.0515670	0.0916310	0.1825009	0.0904207	0.1461868	0.2350475	0.0673101	0.1210044	0.2291758
σ_I	0.0223317	0.0488199	0.0850871	0.0611505	0.0828262	0.1305510	0.0435037	0.0688689	0.1152216

TABLE VII

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{10}_{\mu_{\text{2-ADD}}}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	0.0018604	0.0068653	0.0189240	-0.0075505	-0.0083404	-0.0037872	-0.0022342	-0.0025142	0.0017069
b_ϕ	-0.0000125	-2.776E-17	-0.0000025	0.00002	0.0000025	0.000005	0.0000125	-1.388E-17	0.00001
b_I	-0.005185	-0.00565	-0.013615	0.0027933	0.0021133	-0.0080783	-0.0110183	- 0.0108317	-0.0136267
σ_μ	0.0198964	0.0369956	0.0766734	0.0405841	0.0572769	0.1020067	0.0189423	0.0371503	0.0772251
σ_ϕ	0.0501558	0.0920598	0.1837740	0.0663535	0.1150373	0.2215296	0.0471936	0.0905092	0.2139551
σ_I	0.0222982	0.0405338	0.0867574	0.0475859	0.0677331	0.1179952	0.0321105	0.0595129	0.1027754

TABLE VIII

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{10}_{\mu_0}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	-0.0567233	-0.0571832	-0.0562092	-0.1188823	-0.1042598	-0.0871266	-0.0588714	-0.0720761	-0.0593036
b_ϕ	-0.0000025	0.000005	-0.0000175	0.0000075	-0.0000125	-0.0000225	1.388E-17	-0.000025	0.00001
b_I	0.039065	0.041045	0.04222	0.07986	0.0690617	0.0576067	0.0392233	0.048055	0.0395233
σ_μ	0.0158592	0.0273284	0.0575833	0.037101	0.0622885	0.0983955	0.0420999	0.0649260	0.0889556
σ_ϕ	0.0329080	0.0549783	0.1184147	0.1165405	0.1824432	0.2580886	0.1472002	0.2190381	0.2649518
σ_I	0.0185197	0.0322075	0.0649976	0.0446983	0.0791763	0.1128655	0.0430077	0.0679980	0.1076637

TABLE IX

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{1\leq 4}_{\mu_{\mathrm{add}}}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	-0.0500573	-0.0448485	-0.0355478	-0.1175110	-0.1034209	-0.0716527	-0.0516723	-0.0541870	-0.0535092
b_ϕ	0.0000075	-0.000015	-0.00002	-0.0000275	-0.0000075	0.0000175	-0.00003	0.000015	0.0000225
b_I	0.0356383	0.0335867	0.0332783	0.0795417	0.07009	0.0454717	0.034445	0.03613	0.03566
σ_μ	0.0187834	0.0322533	0.0597895	0.0360110	0.0643024	0.0988410	0.0324026	0.0534711	0.0804050
σ_ϕ	0.0388520	0.0666883	0.1263528	0.1120987	0.1871605	0.2635256	0.1090428	0.1764697	0.2353374
σ_I	0.0205464	0.0347369	0.0637271	0.0434425	0.0861692	0.1187680	0.0338751	0.0569765	0.0965266

TABLE X

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}^{1\leq 4}_{\mu_{\text{2-ADD}}}$

	HLMS			QUAD			2-ADD		
	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.02$	$\sigma = 0.05$
b_μ	-0.0274957	-0.0231342	-0.0136413	-0.1043289	-0.0912383	0.0204555	-0.0240936	-0.0254146	-0.0214275
b_ϕ	0.0000025	0.000015	-0.0000025	-0.00001	0.0000025	-0.00003	0.0000175	0.0000175	0.000015
b_I	0.0106917	0.0071467	0.002985	0.0686683	0.0581317	-0.0186783	0.0035767	0.0044533	0.0017783
σ_μ	0.0162511	0.0293767	0.0590552	0.0386861	0.0616810	0.0575809	0.0260201	0.0478450	0.0681717
σ_ϕ	0.0342420	0.0600811	0.1319089	0.1364497	0.2036908	0.1220339	0.0829561	0.1569252	0.1981763
σ_I	0.0180436	0.0331207	0.0651760	0.0482208	0.0835801	0.0625075	0.0320485	0.0532517	0.0878520

TABLE XI

BIAS AND STANDARD DEVIATION FOR $\mathcal{X}_{\mu_0}^{1\leq 4}$

	reference	HLMS	QUAD	2-ADD
$\mu_{\mathrm{add}}(\{\emptyset\})$	0.	0	0	0
$\mu_{\mathrm{add}}(\{1\})$	0.5	0.233962	0.02742	0.264051
$\mu_{\mathrm{add}}(\{2\})$	0.3	0.29784	0.29769	0.301014
$\mu_{\mathrm{add}}(\{1,2\})$	0.8	0.529422	0.33305	0.564461
$\mu_{\mathrm{add}}(\{3\})$	0.15	0.143562	0.14662	0.148664
$\mu_{\mathrm{add}}(\{1,3\})$	0.65	0.464868	0.16414	0.413156
$\mu_{\mathrm{add}}(\{2,3\})$	0.45	0.456507	0.45211	0.44978
$\mu_{\mathrm{add}}(\{1,2,3\})$	0.95	0.764711	0.47452	0.713669
$\mu_{\mathrm{add}}(\{4\})$	0.05	0.057158	0.04902	0.050786
$\mu_{\mathrm{add}}(\{1,4\})$	0.55	0.549121	0.54962	0.54926
$\mu_{\mathrm{add}}(\{2,4\})$	0.35	0.35709	0.35518	0.351959
$\mu_{\mathrm{add}}(\{1,2,4\})$	0.85	0.845358	0.84869	0.849829
$\mu_{\mathrm{add}}(\{3,4\})$	0.2	0.2074	0.20215	0.200412
$\mu_{\mathrm{add}}(\{1,3,4\})$	0.7	0.695773	0.69809	0.699328
$\mu_{\mathrm{add}}(\{2,3,4\})$	0.5	0.49567	0.49953	0.501688
$\mu_{\mathrm{add}}(\{1,2,3,4\})$	1.0	1	1	1
$\phi_{\mathrm{add}}(1)$	2.000	1.51486	1.04968	1.52473
$\phi_{\mathrm{add}}(2)$	1.200	1.19963	1.20764	1.20337
$\phi_{\mathrm{add}}(3)$	0.600	0.651649	0.59076	0.597669
$\phi_{\mathrm{add}}(4)$	0.200	0.633864	1.15191	0.674235
$I_{\mathrm{add}}(1,2)$	0	0.00172641	-0.001479	-0.001798
$I_{\mathrm{add}}(1,3)$	0	0.0466402	-0.004872	0.0018787
$I_{\mathrm{add}}(1,4)$	0	0.222	0.473771	0.26581
$I_{\mathrm{add}}(2,3)$	0	0.0064852	0.0017322	0.0027885
$I_{\mathrm{add}}(2,4)$	0	-0.0018243	0.0032101	0.0004247
$I_{\mathrm{add}}(3,4)$	0	-0.0420457	0.0075858	-0.001587

TABLE XII

AVERAGE VALUES FOR $\mu_{\mathrm{add}}, \phi_{\mathrm{add}}, I_{\mathrm{add}}$

	reference	HLMS	QUAD	2-ADD
$\mu_0(\{\emptyset\})$	0	0	0	0
$\mu_0(\{1\})$	0.4	0.17436	0.016814	0.142681
$\mu_0(\{2\})$	0.1	0.093464	0.098650	0.083161
$\mu_0(\{1,2\})$	0.6	0.440766	0.132855	0.558729
$\mu_0(\{3\})$	0.1	0.078577	0.086971	0.042494
$\mu_0(\{1,3\})$	0.5	0.333549	0.099548	0.343447
$\mu_0(\{2,3\})$	0.1	0.109621	0.10846	0.132625
$\mu_0(\{1,2,3\})$	0.6	0.720909	0.14176	0.766465
$\mu_0(\{4\})$	0	0.008948	0.009270	0.010288
$\mu_0(\{1,4\})$	0.4	0.397521	0.395168	0.295913
$\mu_0(\{2,4\})$	0.2	0.192852	0.190389	0.137254
$\mu_0(\{1,2,4\})$	0.7	0.699592	0.697977	0.755765
$\mu_0(\{3,4\})$	0.1	0.114172	0.108012	0.089282
$\mu_0(\{1,3,4\})$	0.5	0.492739	0.49685	0.533179
$\mu_0(\{2,3,4\})$	0.2	0.202995	0.201046	0.223217
$\mu_0(\{1,2,3,4\})$	1	1	1	1
$\phi_0(1)$	2.100	1.80051	1.2699	1.83893
$\phi_0(2)$	0.900	1.0206	0.85403	1.09997
$\phi_0(3)$	0.500	0.60101	0.49317	0.57345
$\phi_0(4)$	0.500	0.57788	1.38289	0.48764
$I_0(1,2)$	0.200	0.276207	0.166238	0.332886
$I_0(1,3)$	0.100	0.165956	0.096048	0.158272
$I_0(1,4)$	0.100	0.180482	0.519853	0.142944
$I_0(2,3)$	0.000	0.0519022	0.014073	0.006969
$I_0(2,4)$	0.200	0.0856874	0.224188	0.043803
$I_0(3,4)$	0.100	0.0039727	0.104929	0.036499

TABLE XIII

AVERAGE VALUES FOR μ_0, ϕ_0, I_0