Output Distribution of Choquet Integral

Amit Dhurandhar and Paul Gader

Abstract

In this paper we show how the distribution of the discrete Choquet integral can be analytically computed. The advantage of having an analytical expression is that the value of the cumulative distribution function (cdf) can be computed exactly for the Choquet. We also derive an expression for the density of the Ordered Weighted Average (OWA) operator, which is a special case of the Choquet.

1 Introduction

Fuzzy integrals, including the Choquet integral, have been important tools for fuzzy systems since at least 1990 when Tahani and Keller first proposed them for use in information fusion [9]. It is used in a plethora of application areas viz. image processing, pattern classification, regression analysis etc. [8, 4, 7]. One of the major reasons for its popularity, is its ability to model other commonly used aggregation functions. For eg. the arithmetic mean which is one of the simplest and hence popular aggregation functions, can be modeled by a Choquet integral. The ordered weighted average operators (OWA) used in multicriteria decision making, are also special cases of the Choquet integral. Considering its wide usage, it is interesting to study the statistical behavior of this aggregation function. Although an analytical form for the cumulative distribution of the Sugeno integral has been derived, [6] there has previously not been a form to analytically compute the cdf of the Choquet integral. In this paper, we attempt to derive such an analytical expression.

In the next section we provide definitions of the basic terminology. This is followed by a section which discusses ways of first analytically computing the distributions of the special cases of the Choquet integral i.e. arithmetic mean and OWA and eventually for the full-fledged Choquet integral. Using this analytical expression we plot the cumulative distribution function (cdf) for the Choquet integral and empirically validate it with the help of Monte Carlo estimates.

2 Technical Background

Let $T = \{X_1, X_2, ..., X_n\}$ be a set of n random variables. We now define the following quantities,

1. *Fuzzy measure:* A fuzzy measure $\mu : 2^T \rightarrow [0,1]$[1] is a function such that $\mu(\emptyset) = 0$, $\mu(T) = 1$ and $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$, $\forall\ A, B \in 2^T$.

2. *Additive fuzzy measure:* An additive fuzzy measure is a fuzzy measure with the property that $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$.

3. *Discrete Choquet Integral:* If μ is a fuzzy measure on T, then the discrete Choquet integral is given by,

$$\mathcal{C}_\mu = \sum_{i=1}^{n} X_{\pi(i)}(\mu(A_{\pi(i)}) - \mu(A_{\pi(i+1)})) \tag{1}$$

where π is a permutation defined on $\{1, 2, ..., n\}$ such that $X_{\pi(i)} \leq X_{\pi(j)}$ for $i \leq j$ and $A_{\pi(i)} = \{X_{\pi(i)}, ..., X_{\pi(n)}\}$.

Hereafter, we shall refer to discrete Choquet integral plainly as the Choquet integral.

3 Distribution of Choquet integral

In this section we provide an analytical form to compute the distribution of the Choquet integral.

[1] by fuzzy measure we $\Rightarrow$ normalized fuzzy measure

3.1 Arithmetic Mean

When the fuzzy measure is an additive fuzzy measure, the Choquet integral is independent of the sort and is given by,

$$\mathcal{C}_\mu = \sum_{i=1}^{n} X_i \mu(\{X_i\})$$

The weight $\alpha(i)$, corresponding to X_i, is thus a function of only X_i i.e. $\alpha(i) = \mu(\{X_i\})$. The Choquet integral can be re-written as,

$$\mathcal{C}_\mu = \sum_{i=1}^{n} X_i \alpha(i)$$

The above equation is exactly that of an arithmetic mean (as $\sum_i \alpha(i) = 1$). If the $X_i's$ are independent, then the density of $\mathcal{C}_\mu$ is a convolution of the densities of the $X_i's$ with a change of variable. More clearly, if $p_i()$ is the density of X_i and $Y_i = \alpha(i)X_i$ then the density of $\mathcal{C}_\mu$ is given by,

$$f(\mathcal{C}_\mu) = \frac{1}{\alpha(1)...\alpha(n)} \int_{y_1 \in Y_1} ... \int_{y_{n-1} \in Y_{n-1}} p_1(\frac{y_1}{\alpha(1)})...p_n(\frac{\mathcal{C}_\mu - \sum_{i=1}^{n-1} y_i}{\alpha(n)}) dy_1...dy_{n-1}$$

If the $X_i's$ are normally distributed with mean ν_i and standard deviation σ_i, then the density of $\mathcal{C}_\mu$ is also normal with mean $\sum_i \alpha(i)\nu_i$ and standard deviation $\sqrt{\sum_i \alpha_i^2 \sigma_i^2}$.

3.2 Ordered Weighted Average

The OWA operator is a special case of the Choquet integral where the weight assigned to a random variable depends on its position in the sorted sequence. In other words, the weights are constant w.r.t. a particular order statistic. Formally,

$$\mathcal{C}_\mu = \sum_{i=1}^{n} X_i \alpha(l_i)$$

where l_i is the position of the i^{th} random variable in a sorted sequence of the n

random variables. Representing the same equation in different manner we have,

$$\mathcal{C}_\mu = \sum_{i=1}^{n} X_{(i)}\alpha(i)$$

where $X_{(i)}$ is the i^{th} order statistic and $\alpha(i)$ is the corresponding weight with $\sum_i \alpha(i) = 1$. The distribution of the Choquet in this case is not as straightforward, since the $X_{(i)}$'s are not independent. Thus the problem lies in finding the joint distribution/density of the $X_{(i)}$'s. Given the joint distribution, the distribution of the Choquet can be found analytically. There is a well known closed form expression for the distribution function of the individual $X_{(i)}$'s [2]. The strategy to derive the joint distribution of the $X_{(i)}$'s is a straightforward extension of the strategy to find the individual distributions. If $p()$ is the density of the original random variables X_i i.e. they are i.i.d., then the joint density of the n order statistics is given by,

$$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = \begin{cases} n!p(x_1)p(x_2)...p(x_n) & if\ x_1 < x_2 < ... < x_n \\ 0 & otherwise \end{cases}$$

This result and the general technique for proving it can be found in [2]. The $n!$ comes from the fact that any of the X_i's can be any order statistic. The density of the Choquet (or OWA in this case) is given by,

$$\begin{aligned} f(\mathcal{C}_\mu) =& \frac{1}{\alpha(1)...\alpha(n)} \int_{y_1 \in Y_1} ... \int_{y_{n-1} \in Y_{n-1}} n!p(\frac{y_1}{\alpha(1)})...p(\frac{\mathcal{C}_\mu - \sum_{i=1}^{n-1} y_i}{\alpha(n)}) \cdot \\ & I[\frac{y_1}{\alpha(1)} \leq \frac{y_2}{\alpha(2)} \leq ... \leq \frac{\mathcal{C}_\mu - \sum_{i=1}^{n-1} y_i}{\alpha(n)}] dy_1...dy_{n-1} \end{aligned} \tag{2}$$

where $Y_i = \alpha(i)X_{(i)}$, $I[]$ is an indicator function (can be removed by introducing the correct limits in the above integral).

The limitation of the result in [2] is that it applies only when the $X's$ are i.i.d. We now derive expressions for the joint density of the order statistics when the $X's$ are not i.i.d.. Moreover, we observe that when the X_i's are in fact i.i.d. the result actually reduces to the known result. The proof we provide here is simpler and more general than that for the previously stated result. With this, we state the following theorem.

Theorem 1. *Let X_1, X_2,..., X_n be n random variables. Let $X_{(i)}$ denote the i^{th} order statistic of these n random variables. Then the joint density $f()$ of the order statistics when $x_1 < x_2 < ... < x_n$* [2] *is given by,*

1. *$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = \sum_{i=1}^{n!} p(x_{\pi(i,1)}, x_{\pi(i,2)}, ..., x_{\pi(i,n)})$ when p() is the joint density of X_1, X_2,..., X_n and $x_{\pi(i,j)}$ is the j^{th} element in the i^{th} permutation of x_1, x_2,..., x_n.*
2. *$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = \sum_{i=1}^{n!} p_{\pi(i,1)}(x_1)p_{\pi(i,2)}(x_2)...p_{\pi(i,n)}(x_n)$ when $p_{\pi(i,j)}()$ is the density of the random variable whose value is x_j in the i^{th} permutation and the $X_i's$ are independent.*
3. *$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = n!p(x_1)p(x_2)...p(x_n)$ when $p()$ is the density of the $X_i's$ i.e. the $X_i's$ are i.i.d.*

Proof. We need to find an expression for $f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n)$ where $x_1 < x_2 < ... < x_n$ else the density is 0.

Notice that for the n order statistics to be equal to $x_1, x_2, ..., x_n$ each of the random variables $X_1, X_2, ..., X_n$ must equal exactly one of these values. A possible assignment is $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$. There are n! such assignments. Note that any two of these assignments are mutually exclusive. Moreover, all the assignments account for all possible ways in which $X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n$. Using these facts and denoting the joint density of $X_1, X_2, ..., X_n$ by $p()$ we have the following expression,

$$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = \sum_{i=1}^{n!} p(x_{\pi(i,1)}, x_{\pi(i,2)}, ..., x_{\pi(i,n)})$$

where $x_{\pi(i,j)}$ is the j^{th} element in the i^{th} permutation.

[2] if $x_1 < x_2 < ... < x_n$ is not satisfied the density is 0

If the X_i's are independent with density p_i then the joint density factorizes into a product of the individual densities given by,

$$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = \sum_{i=1}^{n!} p_{\pi(i,1)}(x_1)p_{\pi(i,2)}(x_2)...p_{\pi(i,n)}(x_n)$$

where $p_{\pi(i,j)}()$ is the density of the random variable whose value is x_j in the i^{th} permutation.

If the X_i's are i.i.d. with density $p()$ then all the terms in the above sum are identical and hence we have,

$$f(X_{(1)} = x_1, X_{(2)} = x_2, ..., X_{(n)} = x_n) = n!p(x_1)p(x_2)...p(x_n)$$

□

Using the above expressions the expression for the density of the OWA can be found as in equation 2 by inserting the appropriate term for the joint density of the order statistics.

3.3 Choquet Integral

In the previous two subsections we provided formulae for analytically computing the pdf and hence the cdf of special cases of the Choquet integral. The key idea in obtaining these characterizations, is expressing the Choquet integral in a form wherein the $\alpha()$ is fixed w.r.t. the random variable it is multiplied by. This permits the integral to be expressed as a linear combination of the random variables in question. With the knowledge of the joint distribution of these random variables, the distribution of the Choquet integral can be computed. In other words, to compute the distribution of the full-fledged Choquet we have to represent it in a way that the $\alpha()$ remains fixed, and then compute the distribution.

For the arithmetic mean we have seen that $\alpha()$ is a function of the random variable it is multiplied to. Thus fixing the variable fixes the value of $\alpha()$. For the OWA operators the $\alpha()$ is a function of the position of the random variable in the sorted sequence of the random variables. Thus for a particular order statistic the $\alpha()$ here is fixed. For the full-fledged Choquet integral though, the $\alpha()$ is a function of both the random variable it is multiplied by and its position in the sorted sequence. In fact $\alpha()$ does not depend only on the position but also on the random variables that follow the particular random variable in the sort. Thus by fixing the random variable and the random variables that follow it, we fix the $\alpha()$. A stronger condition is fixing the sort, and hence by fixing the sort and the random variable we fix $\alpha()$. The Choquet integral can be split up into the following piecewise linear form based on the sort,

$$\mathcal{C}_\mu = \begin{cases} \sum_{i=1}^{n} \alpha(S^{(1)}, i) X_i & if\ S^{(1)} \\ \sum_{i=1}^{n} \alpha(S^{(2)}, i) X_i & if\ S^{(2)} \\ . \\ . \\ . \\ \sum_{i=1}^{n} \alpha(S^{(n!)}, i) X_i & if\ S^{(n!)} \end{cases}$$

where $S^{(j)}$ denotes the j^{th} permutation of the n random variables. The probability of $\mathcal{C}_\mu < u$ is given by,

$$\begin{aligned} P[\mathcal{C}_\mu < u] &= \sum_{j=1}^{n!} P[\mathcal{C}_\mu < u, S^{(j)}] \\ &= \sum_{j=1}^{n!} P[\sum_{i=1}^{n} \alpha(S^{(j)}, i) X_i < u, S^{(j)}] \end{aligned} \tag{3}$$

We have to characterize the $S^{(j)}$'s and we do this using the following observation. Any sorted ordering of n variables can be represented by $n-1$ linear inequalities. For example, the relations $X_1 \leq X_2 \leq ... \leq X_n$ hold iff $X_2 - X_1 \geq 0$, $X_3 - X_2 \geq 0$,..., $X_n - X_{n-1} \geq 0$ i.e. iff $X_{i+1} - X_i \geq 0\ \forall\ i \in \{1, 2, ..., n-1\}$. Thus the probability in equation 3 can be written as,

$$P[\mathcal{C}_\mu < u] = \sum_{j=1}^{n!} P[\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u, X_{\pi_j(i+1)} - X_{\pi_j(i)} \geq 0,\ \forall\, i \in \{1, 2, ..., n-1\}] \tag{4}$$

where π_j is the permutation of the n variables indexed by j. To compute $P[\mathcal{C}_\mu < u]$, we need to compute $P[\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u, X_{\pi_j(i+1)} - X_{\pi_j(i)} \geq 0,\ \forall\ i \in \{1, 2, ..., n-1\}]$ for each j. Geometrically, the first condition in this probability $\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u$ represents a region bounded by the hyperplane $\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i - u = 0$. The remaining conditions also represent hyperplanes in an n dimensional space. The value of the probability is computed by integrating the joint density of the n random variables $X_1, X_2, ..., X_n$ over the region specified by these hyperplanes. If the random variables $X_1, X_2, ..., X_n$ are independent with densities $p_1(), p_2(), ..., p_n()$, $P[\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u, X_{\pi_j(i+1)} - X_{\pi_j(i)} \geq 0,\ \forall\ i \in \{1, 2, ..., n-1\}]$ is given by,

$$\begin{aligned}
&P[\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u, X_{\pi_j(i+1)} - X_{\pi_j(i)} \geq 0,\ \forall\ i \in \{1, 2, ..., n-1\}] = \\
&\int_{x_1 \in X_1} \int_{x_2 \in X_2} \cdots \int_{x_n \in X_n} p_1(x_1)p_2(x_2)...p_n(x_n) \cdot \\
&I[\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u, X_{\pi_j(i+1)} - X_{\pi_j(i)} \geq 0,\ \forall\ i \in \{1, 2, ..., n-1\}] dx_1...dx_n
\end{aligned} \tag{5}$$

Combining equations 4 and 5 we get,

$$\begin{aligned}
&P[\mathcal{C}_\mu < u] = \sum_{j=1}^{n!} \int_{x_1 \in X_1} \int_{x_2 \in X_2} \cdots \int_{x_n \in X_n} p_1(x_1)p_2(x_2)...p_n(x_n) \cdot \\
&I[\sum_{i=1}^{n} \alpha(S^{(j)}, i)X_i < u, X_{\pi_j(i+1)} - X_{\pi_j(i)} \geq 0,\ \forall\ i \in \{1, 2, ..., n-1\}] dx_1...dx_n
\end{aligned} \tag{6}$$

Instantiating the pdf's of the $X_i's$ in the above equation and choosing a suitable measure we can compute values of $P[\mathcal{C}_\mu < u]$ for different values of u. We illustrate this fact in the following section.

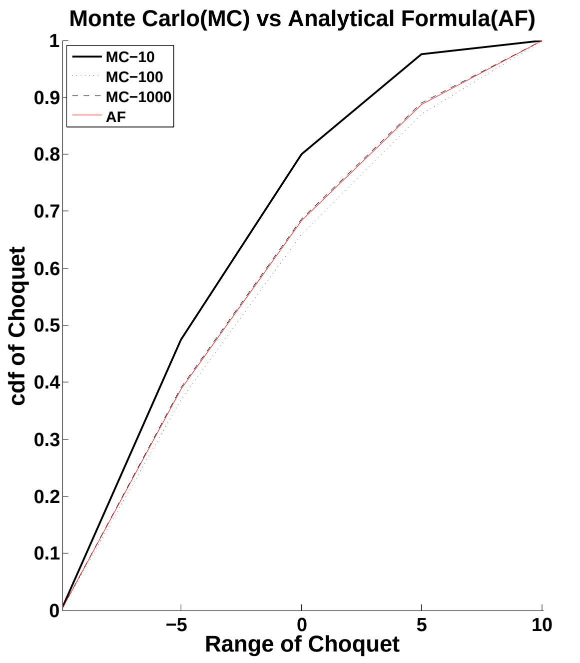

Figure 1: Behavior of MC and AF when $X_1, X_2, X_3 \sim^{i.i.d.}$ N(0,1).

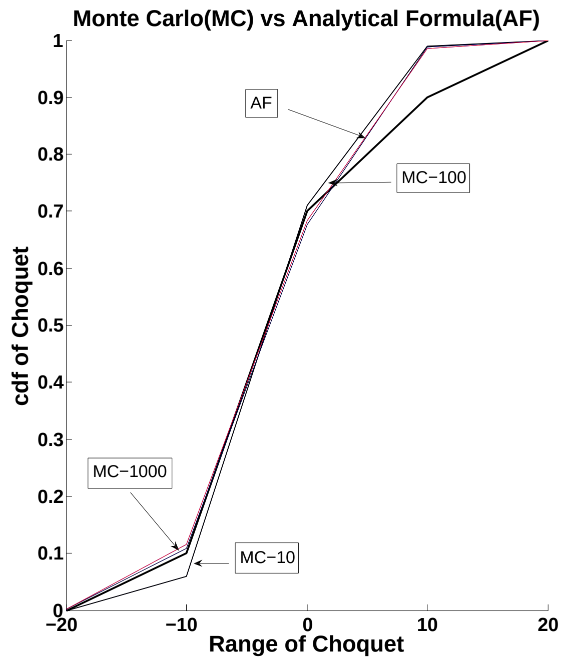

Figure 2: Behavior of MC and AF when $X_1, X_2, X_3 \sim^{i.i.d.}$ N(0,100).

4 Experiments

In the previous section we derived an analytical formula for the cdf of the Choquet integral. In this section we empirically verify on some specific but commonly used distributions that the formula we have obtained is in fact accurate. We accomplish this, by choosing a non-additive measure and comparing the cdf's estimated by Monte Carlo with those estimated by the formula for different distributions of the $X_i's$. We compare Monte Carlo and the analytical formula with $X_i's$ having the following commonly used densities,

1. normal with mean 0 and variance 1 i.e. N(0,1),
2. normal with mean 0 and variance 100 i.e. N(0,100),
3. exponential with mean 1 i.e. Exp(1),
4. uniform in the range $[0, 1]$ i.e. Unif(0,1).

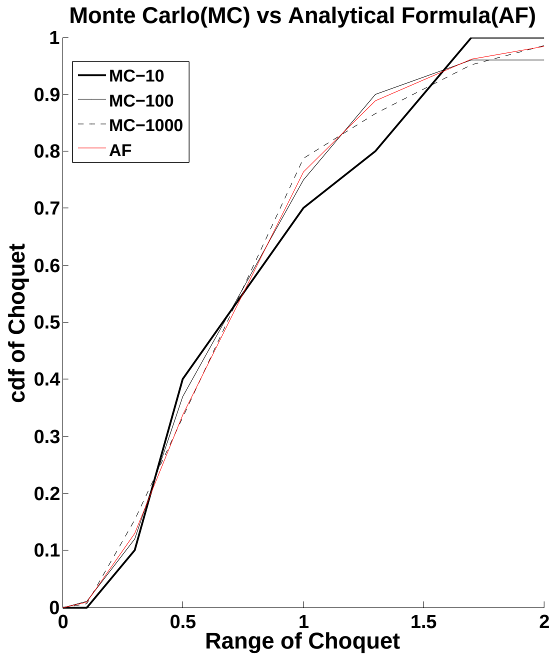

Figure 3: Behavior of MC and AF when $X_1, X_2, X_3 \sim^{i.i.d.}$ Exp(1).

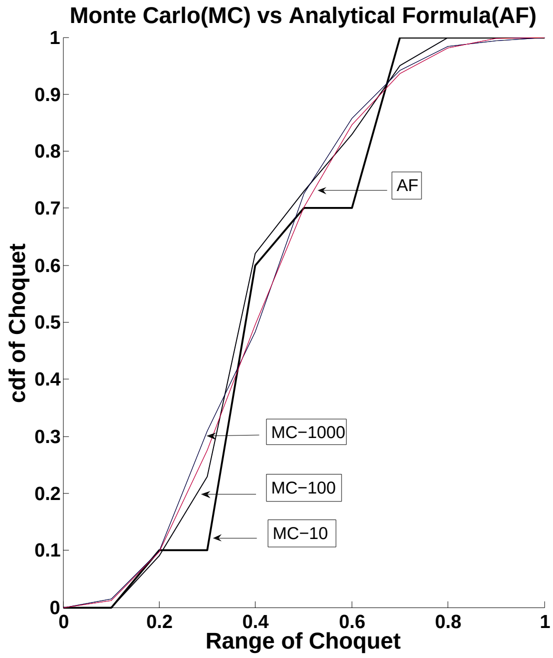

Figure 4: Behavior of MC and AF when $X_1, X_2, X_3 \sim^{i.i.d.}$Unif(0,1).

We have the normal density twice, so as to see if variance affects the accuracy of the estimates. We fix n to 3. Thus, the Choquet integral is determined by X_1, X_2, X_3 distributed i.i.d. w.r.t. some distribution. In the experiments MC-j $\Rightarrow$ Monte Carlo performed with j samples.

Observations: In all the four cases, Figure 1, Figure 2, Figure 3 and Figure 4 we observe that as the sample size for the Monte Carlo increases from 10 to 1000, its estimated cdf approaches the cdf given by instantiating the analytical formula in equation 6. This empirically validates our analytical result.

5 Discussion

In the previous sections we derived an analytical expression for computing the cdf of the Choquet integral and verified it empirically. In this section we discuss issues related to usefullness and scalability of the formula. With the knowledge of an analytical formula, the output distribution of the Choquet integral can be characterized precisely in desired settings. A problem with the full-fledged Choquet integral is that it has exponential number of piecewise linear terms. With increasing n the number of terms that need to be computed increases rapidly. Since, major applications of the Choquet include multicriteria decision making or pattern classification etc. where n is small, the analytical formula can still prove to be effective. In the applications where n is large the optimization in speed depends on the ability to exploit the structure of the measure defined, μ. If μ is additive, the formula reduces to the arithmetic mean with only one term. If μ is non-additive but remains constant w.r.t. an order statistic then the formula reduces to that for a OWA operator with again only a single term. Other existing patterns in the defined μ may also be exploited to reduce the number of terms and enhance performance.

6 Conclusion

In summary, we derived analytical expressions for the density of the OWA operator and for the cdf of the full-fledged Choquet integral. We ascertained the correctness of the expressions for the cdf, by comparing it with Monte Carlo estimates, where we observed that with increasing sample size the empirical cdf approached the analytically estimated cdf. We also discussed issues related to scalability of the expression. It remains to be seen if a similar approach will permit in unveiling an analytical expression for the continuous Choquet integral.

References

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

[2] G. Casella and R. Berger. *Statistical Inference*. Duxbury Advanced Series, 2002.

[3] R. Fuller. Owa operators in decision making.

[4] M. Grabisch, T. Murofushi, and M. Sugeno. *Fuzzy Measures and Integrals.* Springer-Verlag, 3 edition, 2000.

[5] J. Keller and J. Osborn. Training the fuzzy integral. *Int. J. Approx. Reasoning*, 15(1):1–24, 1996.

[6] J. Marichal. Cumulative distribution functions and moments of weighted lattice polynomials. In *In 27th Linz Seminar on Fuzzy Set Theory.*, 2006.

[7] A. Mendez, P. Gader, J. Keller, and K. Chamberlin. Minimum classification error training for choquet integrals with applications to landmine detection. *IEEE Trans. Fuzzy Systems*, 2007.

[8] H. Nguyen and E. Walker. *Fuzzy Logic.* CRC, 3 edition, 2006.

[9] H. Tahani and J. Keller. Information fusion in computer vision using the fuzzy integral. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):733–741, 1990.