SIT718 Real World Analytics

Lecturer: Dr Ye Zhu

School of Information Technology
Deakin University

Week 5: Weighted Averaging

Weighted averaging

For this week we have the following learning aims

- To be able to apply weighted versions of the averaging functions we have studied so far
- To be able to define and interpret weighting vectors of aggregation functions in context

Group decision making

Candidate	Judge 1	Judge 2	Judge 3
Yezi	9	6	4
Jimin	7	7	6
Hyolyn	4	8	8

Group decision making

Candidate	Judge 1	Judge 2	Judge 3
Yezi	9	6	4
Jimin	7	7	6
Hyolyn	4	8	8

$$\frac{1}{3}(x_1 + x_2 + x_3),$$

or alternatively

$$\frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3.$$

In order to give the manager (who provides the score x_1 for each of candidates) more importance, we can instead use the coefficients $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ or $0.5, 0.25, 0.25$.
Now the calculation will look like this.

$$\frac{1}{2}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3.$$

Group decision making

Candidate	Judge 1	Judge 2	Judge 3	AM	WAM
Yezi	9	6	4	6.33	7
Jimin	7	7	6	6.67	6.75
Hyolyn	4	8	8	6.67	6

Regression parameters

You may have studied regression models in statistics, i.e. where we find a model for a situation in terms of a coefficient and an intercept,

$$y = \beta_1 x + \beta_0.$$

The β_1 term is interpreted as the change in y for every unit change in x, i.e. the gradient of the linear regression line. The β_0 parameter represents the intercept or the value of y when $x = 0$.

Aggregation parameters

For aggregation functions, we have boundary conditions $A(0, 0, \ldots, 0) = 0$ and so none of the functions we deal with have an 'intercept'.

Remember that in most cases, we have pre-processed the data so that it all ranges over the same interval - this is where we would usually deal with any features of the data that correspond with an intercept.

A weighted version of the arithmetic mean looks like this:

Formula (1)

$$WAM_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

Aggregation parameters

$$WAM_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

We interpret the parameters similarly to how we do with the regression line. The value of w_i represents how much the output changes with each increase of x_i.
*Note: You probably have heard of the 'WAM' in the calculation of your average study score at university.

Aggregation parameters

Example

Let $\mathbf{w} = \langle 0.5, 0.2, 0.3 \rangle$.
So explicitly, our weighted arithmetic mean is:

$$0.5x_1 + 0.2x_2 + 0.3x_3.$$

Then if we have $\mathbf{x} = \langle 0.6, 0.7, 0.1 \rangle$ *then our output is:*

$$WAM_{\mathbf{w}}(0.6, 0.7, 0.1) = 0.5(0.6) + 0.2(0.7) + 0.3(0.1)$$

$$= 0.3 + 0.14 + 0.03 = 0.47.$$

If we increase x_1 *by 0.1 then the output will go up to* 0.52*, however if we increase* x_2 *by 0.1 then the output only goes up to* 0.49*.*

Aggregation parameters

For aggregation functions we also have a restriction on the weights.

$$w_i \geq 0, \forall\, i$$

$$\sum_{i}^{n} w_i = 1.$$

Aggregation parameters

$$w_i \geq 0, \forall\, i$$

$$\sum_{i}^{n} w_i = 1.$$

We need the first of these so that the function stays monotone increasing. In normal regression it's okay to have negative parameters, it just means the variable is negatively correlated with the output. However we are restricting ourselves to this special class of functions.

Aggregation parameters

$$w_i \geq 0, \forall\, i$$

$$\sum_{i}^{n} w_i = 1.$$

The second condition ensures that the function stays averaging. E.g. if $\sum_{i=1}^{n} w_i > 1$, we could have $\mathbf{w} = (0.6, 0.6, 0.6)$. Then the result of our previous example using $\mathbf{x} = (0.6, 0.7, 0.1)$ would result in an output of 0.84 which is greater than all the outputs.

Weighting vectors

The vector of parameters which we denote as $\mathbf{w}$ is often referred to as a weighting vector. The i-th weight usually corresponds with the i-th input.
The usual interpretation of weights is the *importance* of that variable. However it can refer to other concepts too.

Weighting vectors

- In calculated average study scores in university, the weights are often only increased if a unit is a 2-credit point unit. However for some scholarship applications, they might giver third year units higher priority than 2nd or first year units by making these weights higher.
- Weights could also be used if inputs are reflective of a proportion of the population involved in a decision process. For example, if we were doing peer evaluation and 5 people recommended a score of 7, 2 people recommended a score of 6 and 3 people recommended a score of 5, then rather than aggregate $AM(7, 7, 7, 7, 7, 6, 6, 5, 5, 5)$ we could calculate a weighted arithmetic mean of $\mathbf{x} = \langle 7, 6, 5 \rangle$ with weights $\mathbf{w} = \langle 0.5, 0.2, 0.3 \rangle$.

Weighting vectors

Example

Evaluate the weighted arithmetic mean when $\mathbf{w} = \langle 0.2, 0.5, 0.3 \rangle$ and the input $\mathbf{x}$ is $\langle 0.6, 0.7, 0.1 \rangle$.

Weighting vectors

Example

Evaluate the weighted arithmetic mean when $\mathbf{w} = \langle 0.2, 0.5, 0.3 \rangle$ and the input $\mathbf{x}$ is $\langle 0.6, 0.7, 0.1 \rangle$.
Answer: 0.5

Example: Welfare functions, Gini index

In economics there is the concept of a welfare function.

$$\text{Welfare}_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i,$$

where x_i represents the income (or income with respect to the poverty line) of the i-th poorest person. Depending on how many people there are, the vector **w** is calculated using

$$w_i = \frac{2n + 1 - 2i}{n^2}$$

For example, if $n = 5$ then $\mathbf{w} = \frac{9}{25}, \frac{7}{25}, \frac{5}{25}, \frac{3}{25}, \frac{1}{25}$. Here the lowest income gets the highest weight.
So the welfare increases more if we increase x_1 (the income of the poorest person) than if we increase x_5.

Example: Welfare functions, Gini index

A related concept, the **inequality** of a population can then be calculated based on the welfare function and it's dual (recall we learnt about dual functions last week). The inequality is

$$\text{Inequality}(\mathbf{x}) = \frac{\text{Welfare}^{d}_{\mathbf{w}}(\mathbf{x}) - \text{Welfare}_{\mathbf{w}}(\mathbf{x})}{2\sum\limits_{i=1}^{n} x_i}$$

Note: a more common-way to express inequality is the average of pairwise differences, $\frac{1}{2n\sum\limits_{i=1}^{n} x_i}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}|x_i - x_j|$, but this ends up being the same.

Weighted power means

We can apply weighting vectors in the case of all power means and quasi-arithmetic means too.

Formula (**Weighted** power mean)

$$PM_{\mathbf{w}}(\mathbf{x}) = \left(\sum_{i=1}^{n} w_i x_i^p \right)^{\frac{1}{p}}$$

Recall that for the **power mean**:
if $p = 1$: AM
if $p = 2$: Quadratic mean
if $p = -1$: Harmonic mean
if $p = 0$: Geometric mean
if $p \to \infty$: Maximum
if $p \to -\infty$: Minimum

Weighted power means

For the special cases of harmonic and geometric mean we have:

$$GM_{\mathbf{w}}(\mathbf{x}) = \prod_{i=1}^{n} x_i^{w_i} = x_1^{w_1} \times x_2^{w_2} \times \cdots \times x_n^{w_n}$$

$$HM_{\mathbf{w}}(\mathbf{x}) = \left(\sum_{i=1}^{n} \frac{w_i}{x_i} \right)^{-1} = \frac{1}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \cdots + \frac{w_n}{x_n}}$$

Weighted power means

For means other than *AM*, we need to be careful in interpreting weights and predicting how the function behaves.

Example

Suppose we have a geometric mean with $\mathbf{w} = \langle 0.5, 0.2, 0.3 \rangle$ *and let* $\mathbf{x} = \langle 0.7, 0.6, 0.3 \rangle$ *Then:*

$$GM_{\mathbf{w}}(\mathbf{x}) = 0.7^{0.5} \times 0.6^{0.2} \times 0.3^{0.3} = 0.5264.$$

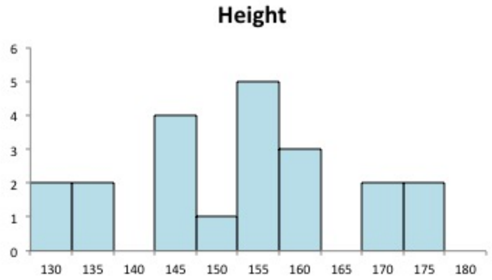
Now, if we increase x_1 *by 0.1, we have* $GM_{\mathbf{w}}(0.8, 0.6, 0.3) \approx 0.5627$*, however if we increase* x_3 *by 0.1, we get* $GM_{\mathbf{w}}(0.7, 0.6, 0.4) \approx 0.5738$*.*

WEIGHTED POWER MEANS

- So we actually are better off increasing x_3, even though x_1 has the highest weight and considered "most important".
- Recall from last week that **geometric means** are **more affected by low inputs** - so this explains our counter-intuitive result.
- The weights still reflect the importance of each variable overall, but we also need to take into account the function's behaviour.

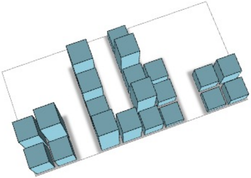
WEIGHTED MEDIAN

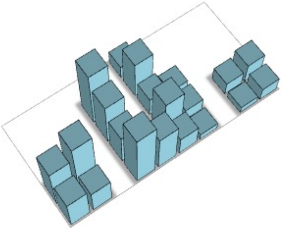
The standard median is the middle value when all of the inputs are arranged in order.



Weighted median

In the case of the weighted median, each input might be associated with a varying degree of importance or density.





WEIGHTED MEDIAN

Example calculation. Let $\mathbf{w} = \langle 0.32, 0.08, 0.20, 0.06, 0.10, 0.24 \rangle$ and the inputs $\mathbf{x} = \langle 0.78, 0.45, 0.03, 0.27, 0.10, 0.45 \rangle$

Weighted median

Example calculation. Let $\mathbf{w} = \langle 0.32, 0.08, 0.20, 0.06, 0.10, 0.24 \rangle$ and the inputs $\mathbf{x} = \langle 0.78, 0.45, 0.03, 0.27, 0.10, 0.45 \rangle$

w	0.32	0.08	0.20	0.06	0.10	0.24
x	0.78	0.46	0.03	0.27	0.10	0.45

Weighted median

Example calculation.

$\mathbf{w}$	0.32	0.08	0.20	0.06	0.10	0.24
$\mathbf{x}$	0.78	0.46	0.03	0.27	0.10	0.45

$\mathbf{w}_{(\mathbf{x}\nearrow)}$	0.20	0.10	0.06	0.24	0.08	0.32
$\mathbf{x}\nearrow$	0.03	0.10	0.27	0.45	0.46	0.78

Now look at the cummulative sum of the weights from left to right and look to the point we have above 50%. This occurs at the w entry that is equal to 0.24, since $0.20 + 0.10 + 0.06 = 0.36$ and then $0.36 + 0.24 = 0.6 > 0.5$. So in this case the weighted median output will be 0.45 (the 4th highest input).

Weighted median

Example calculation:
On the other hand, if we had the same weighting vector but with the input $\mathbf{x} = \langle 0.62, 0.33, 0.26, 0.11, 0.91, 0.12 \rangle$, then we would have

$\mathbf{w}$	0.32	0.08	0.20	0.06	0.10	0.24
$\mathbf{x}$	0.62	0.33	0.26	0.11	0.91	0.12

and when rearranged in order we get

$\mathbf{w}_{(\mathbf{x}\nearrow)}$	0.06	0.24	0.20	0.08	0.32	0.10
$\mathbf{x}_{\nearrow}$	0.11	0.12	0.26	0.33	0.62	0.91

- In this case we have exactly 50% at the weight equal to 0.20, which is dealt with in different ways.
- We can either take the half-way point between the two values on the 50% border $((0.26 + 0.33)/2 = 0.295)$.
- Or we can use the lower value of 0.26 (which is referred to as the 'lower weighted median') or we can take the higher value of 0.33 (referred to as the 'upper weighted median').

Weighted median

Definition
Weighted (lower) median
For an input vector $\mathbf{x} = \langle x_1, x_2, \ldots, x_n \rangle$ and a weighting vector $\mathbf{w}$, the lower weighted median is given by,

$$\mathrm{Med}_{\mathbf{w}}(\mathbf{x}) = x_{(k)},$$

where the $(\cdot)$ notation indicates that the inputs are first reordered in non-decreasing order, $\mathbf{x}_{\nearrow} = \langle x_{(1)}, x_{(2)}, \ldots, x_{(n)} \rangle$ and k is the index such that

$$\sum_{i=1}^{k-1} w_i < \frac{1}{2} \text{ and } \sum_{i=k+1}^{n} w_i \leq \frac{1}{2}.$$

Weighted median

Example. Calculate the weighted median for the input vector $\mathbf{x} = \langle 0.3, 0.7, 0.8, 0.2 \rangle$ and $\mathbf{w} = \langle 0.1, 0.4, 0.3, 0.2 \rangle$.

Other weighting conventions

The concept of dominance (in economics and ecology) is a concept that can also be considered in the framework of weighted means. We let, q_i represent the proportional values, based on $\mathbf{x}$, i.e. $q = \frac{x_i}{\sum_{i=1}^{n} x_i}$.

Formula (Simpson's dominance index)

$$Simp(\mathbf{q}) = \sum_{i=1}^{n} q_i^2$$

We could also write this as $WAM_{\mathbf{w}}(\mathbf{q})$ where $w_i = q_i$, i.e. the weight depends on the population.
This function reaches a minimum when all of the inputs are equal, and reaches a maximum of 1 if one of the $\mathbf{q}_i = 1$, i.e. there is only one species (since this $q_i = 1$ means that x_i is 100% of the total.

Other weighting conventions

Entropy can be thought of in a similar way, although it is not an aggregation function. it is expressed:

$$\text{Entropy}(\mathbf{q}) = \sum_{i=1}^{n} q_i \ln q_i$$

which is similar to the geometric mean however without the inverse performed at the end.

OTHER WEIGHTING CONVENTIONS

- The key difference with these kinds of weighting is that the effective weights will change depending on the inputs, i.e, the smaller the input relative to the remaining inputs, the smaller the weight that will be applied.
- This type of 'weighting convention' does result in non-monotone behaviour, and ***so for the moment we will usually restrict ourself to functions where the weights do not change.*** However next week we will look at situations where the weighting vector remains fixed but the order of the inputs can change.

Borda count

Borda count

- As well as weighting the importance of judges, another use of weighting is in vote-based evaluations.
- In relation to our first example, this would require each of the selection panel members to rank the candidates rather than give them an overall weighting. Once the ranks are in, each first place vote is worth a certain number of points, each second place is worth a certain number of points (usually less) and so on. The overall winner is then the one who receives the highest number of points.

Borda count

Example:
What would be the Borda count for each candidate in the opening example if the judges' scores were converted to rankings?
Judge 1's ranking is

$$\text{Yezi} \succ \text{Jimin} \succ \text{Hyolyn},$$

while both Judge 2 and Judge 3 have the ranking

$$\text{Hyolyn} \succ \text{Jimin} \succ \text{Yezi}.$$

Awarding $\omega_1 = 2$ points for first place, $\omega_2 = 1$ point for second place, and $\omega_3 = 0$. The candidate's scores will be,

$$\begin{array}{ll} \text{Yezi} & 2(1) + 1(0) + 0 = 2 \\ \text{Jimin} & 2(0) + 3(1) + 0 = 3 \\ \text{Hyolyn} & 2(2) + 1(0) + 0 = 4 \end{array}$$

Note that these Borda count systems do not technically define weighting vectors since they don't add to 1, however the could be normalised to give equivalent results.

Averaging with interaction

For this week we have the following learning aims

- To be able to apply OWA weights and identify special cases
- To be able to calculate how similar an OWA function is to the maximum function
- To introduce the notion of a fuzzy measure and learn how to interpret its values and calculate with the Choquet integral

ORDERED WEIGHTED AVERAGING

In some Olympic competitions involving judges, the highest and lowest scores given by the judges are removed and then the average of the remaining scores is given as the overall score.

Example

A diver receives the scores $\mathbf{x} = \langle 9, 8.8, 9.6, 4.3, 7.6, 8 \rangle$ *for one of her dives.*
Her final dive score is: $AM(9, 8.8, 7.6, 8) = 8.35$.

Why do you suppose this method is employed?

Ordered weighted averaging

We can represent this as a weighted mean, with $\mathbf{w} = \langle 0.25, 0.25, 0, 0, 0.25, 0.25 \rangle$, however if the next diver has scores $\mathbf{x} = \langle 8.5, 8, 8, 7.6, 7, 3 \rangle$ then now it's the first and last score that would be removed.

This is an example of what we refer to as ordered weighted averaging.

ORDERED WEIGHTED AVERAGING

The ordered weighted averaging (OWA) operator can be defined as follows:

Formula (OWA)

$$OWA_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_{(i)} = w_1 x_{(1)} + w_2 x_{(2)} + \cdots + w_n x_{(n)}$$

where we use the brackets notation $x_{(\cdot)}$ to indicate that the arguments of $\mathbf{x}$ have been rearranged into non-decreasing order.

Ordered weighted averaging

$$OWA_{\mathbf{w}} = \sum_{i=1}^{n} w_i x_{(i)}$$

Example

Suppose $\mathbf{x} = \langle 0.3, 0.6, 0.8, 0.2 \rangle$ *and* $\mathbf{w} = \langle 0.1, 0.3, 0.4, 0.2 \rangle$. *In this case we have,*

$$OWA_{\mathbf{w}}(\mathbf{x}) = 0.1(0.2) + 0.3(0.3) + 0.4(0.6) + 0.2(0.8)$$

i.e., we first reorder the vector (in non-decreasing order), so it becomes $\mathbf{x}_{\nearrow} = \langle 0.2, 0.3, 0.6, 0.8 \rangle$ and then apply the weights.

Ordered weighted averaging

$$OWA_{\mathbf{w}} = \sum_{i=1}^{n} w_i x_{(i)}$$

Example

Calculate the OWA when $\mathbf{x} = \langle 0.2, 0.1, 0.3, 0.7 \rangle$ *and* $\mathbf{w} = \langle 0.7, 0.1, 0, 0.2 \rangle$.

OTHER DEFINITIONS OF ORDERED WEIGHTED AVERAGING

IMPORTANT NOTE:

- In some other books (also in the previous year's' lectures of this unit) the OWA is defined in terms of a '*non-increasing*' permutation. Then interpretation of the OWA values will be just the reverse. For example, (with the 'non-increasing' definition) a weight vector $w =< 0.7, 0.1, 0.2 >$ for OWA will have the interpretation that, it gives weight of 0.7 for the 'highest' input and a weight of 0.2 for the 'smallest' input.

- For this unit, we follow the definition using 'non-decreasing' permutation as defined in the previous slides. [Chapter 4 of the reference book also uses this definition for OWA]
- Hence, in this case (with the 'non-decreasing' definition) the interpretation of a weight vector $w =< 0.7, 0.1, 0.2 >$ for OWA will be: it gives a weight of 0.7 for the 'smallest' input and a weight of 0.2 for the 'largest' input.

Ordered weighted averaging - special cases

You might recall some other functions that depend on the order of the inputs, i.e. the median, the maximum and the minimum. These are special cases with weighting vectors:

Maximum $\mathbf{w} = \langle 0, 0, \ldots, 0, 1 \rangle$

Minimum $\mathbf{w} = \langle 1, 0, \ldots, 0, 0 \rangle$

Median

$$\mathbf{w} = \langle \underbrace{0, \ldots, 0}_{k \text{ zeros}}, 1, \underbrace{0, \ldots, 0}_{k \text{ zeros}} \rangle, \quad n = 2k + 1$$

$$\mathbf{w} = \langle \underbrace{0, \ldots, 0}_{k-1 \text{ zeros}}, 0.5, 0.5, \underbrace{0, \ldots, 0}_{k-1 \text{ zeros}} \rangle, \quad n = 2k$$

Ordered weighted averaging - special cases

Trimmed mean:

- In fact the arithmetic mean is also a special case (when all weights are equal, and we also have two robust statistics calculations.
- Let us assume h be a number (multiple of 2) indicating ***how many inputs are discarded in total from top and bottom together***.
- **Trimmed mean** discards the highest $\frac{h}{2}$ inputs and the lowest $\frac{h}{2}$ inputs and then takes the average of those remaining

Trimmed mean $\quad \mathbf{w} = \langle \underbrace{0, \ldots, 0}_{\frac{h}{2} \text{ zeros}}, \frac{1}{n-h}, \frac{1}{n-h}, \ldots, \frac{1}{n-h}, \underbrace{0, \ldots, 0}_{\frac{h}{2} \text{ zeros}} \rangle,$

Winsorized mean:

- The highest and lowest $\frac{h}{2}$ inputs are replaced with the next highest/lowest of the inputs

$$\text{Winsorized mean} \quad \mathbf{w} = \langle \underbrace{0, \ldots, 0}_{\frac{h}{2}\text{ zeros}}, \frac{2+h}{2n}, \frac{1}{n}, \ldots, \frac{1}{n}, \frac{2+h}{2n}, \underbrace{0, \ldots, 0}_{\frac{h}{2}\text{ zeros}} \rangle,$$

[NOTE: In some other situations/books, h may be given in terms of the percentage that is removed from each side. For example, h=0.1, in this case would represent removing 10% from lower side and 10% from upper side. Always check what the h represent]

Example - Winsorized mean

Calculate the Winsorized mean for $h = 4$ and
$\mathbf{x} = \langle 0.3, 0.8, 0.43, 0.2, 0.33, 0.49, 0.7, 0.4 \rangle$.

- We first reorder the inputs so that we can easily identify the highest and lowest.

$$\mathbf{x}_{\nearrow} = \langle 0.2, 0.3, 0.33, 0.4, 0.43, 0.49, 0.7, 0.8 \rangle$$

- We now replace the lowest two inputs 0.2 and 0.3 with the next lowest, i.e. $x_{(3)} = 0.33$, and the highest two inputs with 0.49. We then calculate the arithmetic mean.

$$\mathrm{AM}(\color{red}{0.33, 0.33}, 0.33, 0.4, 0.43, 0.49, \color{red}{0.49, 0.49}) = 0.41125.$$

Ordered weighted averaging

- The OWA is useful because it allows us to define functions that graduate between the minimum and maximum function, favouring higher or lower inputs depending on our preference.
- For example,

$$\mathbf{w} = \langle 0, 0, 0.1, 0.2, 0.3, 0.4 \rangle$$

  takes the *highest 4 inputs* into account and gives the *highest one the most influence*.
- We might have a function like this kind of aggregation with student tests if we want to take their best test scores into account but also reward consistently high performance.

Orness

- Depending on the weighting vector, OWA functions can be described by the level of "orness".
- This measure, (named after the logical term "or") essentially gives how close the OWA is to the maximum, i.e. *how much it favours high inputs*.

ORNESS

- If orness($\mathbf{w}$) $= 1$, then the OWA is the **maximum** with $\mathbf{w} = \langle 0, 0, \ldots, 0, 1\rangle$.
- If orness($\mathbf{w}$) $= 0$ then we obtain the **minimum** function.
- The orness of the arithmetic mean is 0.5, however other weighting vectors like $\mathbf{w} = (0, 0.5, 0.5, 0)$ or $\mathbf{w} = (0.1, 0.2, 0.4, 0.2, 0.1)$ also have orness of 0.5.

Orness

- The orness is a concept that applies to all averaging functions (the geometric mean when $n = 2$ is $1/3$ while the harmonic mean (when $n = 2$) has an orness of approximately 0.2274).
- However for these functions it is more difficult to calculate. For the OWA there is a simple formula,

Formula (Orness of OWA)

$$\sum_{i=1}^{n} w_i \frac{i-1}{n-1}$$

Orness-Example

- calculate the orness for an OWA with the weighting vector $\mathbf{w} = \langle 0.1, 0.5, 0.4 \rangle$.
- Since the value of $n = 3$, the multipliers for the weights will be

$$\frac{1-1}{3-1} = 0, \frac{2-1}{3-1} = \frac{1}{2}, \frac{3-1}{3-1} = 0$$

  respectively.
- Hence

$$\text{orness}(\langle 0.1, 0.5, 0.4 \rangle) = 0.1(0) + 0.5(\frac{1}{2}) + 0.4(1) = 0.25 + 0.4 = 0.65.$$

- Since this is greater than 0.5, we would say that this particular OWA tends more toward higher inputs.

Orness

$$\sum_{i=1}^{n} w_i \frac{i-1}{n-1}$$

Example

Calculate the orness of an OWA function when $\mathbf{w} = \langle 0.1, 0.2, 0.2, 0.4, 0.1 \rangle$

The Choquet integral

- So far we have looked at applying weights to arguments either depending on their source (weighted power mean) or depending on their relative size (the OWA).
- We now turn to a very useful (but a somewhat complex) operator that can take both into account.
- For it, we need to learn the concept of a fuzzy measure (or capacity).

The Choquet integral

- Basically, a fuzzy measure allocates a weight to a **set** of inputs
- with the property that the *full set* has a value of 1, the *empty set* has a value of zero, and *if we add an element to the set, we can't decrease* the measure (i.e.,, we have a kind of monotonocity in terms of adding elements to the set).
- An example of a fuzzy measure when there are 3 variables

$$
\begin{array}{ccc}
 & v(\{1,2,3\}) = 1 & \\
v(\{1,2\}) = 0.6 & v(\{1,3\}) = 0.8 & v(\{2,3\}) = 0.7 \\
v(\{1\}) = 0.3 & v(\{2\}) = 0.3 & v(\{3\}) = 0.7 \\
 & v(\emptyset) = 0 &
\end{array}
$$

The Choquet integral

- The Choquet integral (and fuzzy measures) associate a weight with every **set of inputs,** rather than just a *single inpu*t.
- It might be the case that two inputs together are worth more or less than their individual components.

The Choquet integral

An example of interpretation: For 3 workers, the values represent the percentage of a project they can achieve in a day in each combination.

$$
\begin{array}{ccc}
 & v(\{1,2,3\}) = 1 & \\
v(\{1,2\}) = 0.6 & v(\{1,3\}) = 0.8 & v(\{2,3\}) = 0.7 \\
v(\{1\}) = 0.3 & v(\{2\}) = 0.3 & v(\{3\}) = 0.7 \\
 & v(\emptyset) = 0 &
\end{array}
$$

THE CHOQUET INTEGRAL

The (discrete) Choquet integral is an averaging aggregation function that is calculated based on the values of the fuzzy measure.

$$C_v(\mathbf{x}) = \sum_{i=1}^{n} x_{[i]}(v(\{[i],[i+1],\ldots,[n]\}) - v(\{[i+1],[i+2],\ldots,[n]\}))$$

the notation $[i]$ means we arrange the inputs in from lowest to highest.

THE CHOQUET INTEGRAL

The formula seems complicated, but actually calculation is not too difficult.

$$C_v(\mathbf{x}) = \sum_{i=1}^{n} x_{[i]}(v(\{[i],[i+1],\ldots,[n]\}) - v(\{[i+1],[i+2],\ldots,[n]\}))$$

$$v(\{1,2,3\}) = 1$$

$$v(\{1,2\}) = 0.6 \quad v(\{1,3\}) = 0.8 \quad v(\{2,3\}) = 0.7$$

$$v(\{1\}) = 0.3 \quad v(\{2\}) = 0.3 \quad v(\{3\}) = 0.7$$

$$v(\emptyset) = 0$$

Example

Using the fuzzy measure given, determine the output for $\mathbf{x} = \langle 0.8, 0.3, 0.4 \rangle$

The Choquet integral

$$\begin{array}{ccc} & v(\{1,2,3\}) = 1 & \\ v(\{1,2\}) = 0.6 & v(\{1,3\}) = 0.8 & v(\{2,3\}) = 0.7 \\ v(\{1\}) = 0.3 & v(\{2\}) = 0.3 & v(\{3\}) = 0.7 \\ & v(\emptyset) = 0 & \end{array}$$

Example

Using the fuzzy measure given, determine the output for $\mathbf{x} = \langle 0.8, 0.3, 0.4 \rangle$

- Arrange the inputs in order (keep track of which position it belonged to)
- Determine the weights to apply to each input

THE CHOQUET INTEGRAL

$$
\begin{array}{ccc}
 & v(\{1,2,3\}) = 1 & \\
v(\{1,2\}) = 0.6 & v(\{1,3\}) = 0.8 & v(\{2,3\}) = 0.7 \\
v(\{1\}) = 0.3 & v(\{2\}) = 0.3 & v(\{3\}) = 0.7 \\
 & v(\emptyset) = 0 &
\end{array}
$$

Example

Using the fuzzy measure given, determine the output for $\mathbf{x} = \langle 0.8, 0.3, 0.4 \rangle$

- Calculate

The Choquet integral - Another example

$$C_v(\mathbf{x}) = \sum_{i=1}^{n} x_{[i]}(v(\{[i],[i+1],\ldots,[n]\}) - v(\{[i+1],[i+2],\ldots,[n]\}))$$

Another Example (Example 2):

Example

Let $v(\{1,2,3\}) = 1, v(\{1,2\}) = v(\{1,3\}) = 0.6, v(\{2,3\}) = 0.9, v(\{1\}) = 0.4, v(\{2\}) = v(\{3\}) = 0.1.$ *Then if* $\mathbf{x} = \langle 0.8, 0.3, 0.4 \rangle$*, to calculate the output:*

- *0.3, 0.4, 0.8*
- *to work out the weight allocated to 0.3, we use* $v(\{1,2,3\})$ *minus the set of inputs when the 0.3 is removed. From the original vector we see that 0.3 was the 2nd input, so we will be subtracting* $v(\{1,3\})$*.*

THE CHOQUET INTEGRAL

$$C_v(\mathbf{x}) = \sum_{i=1}^{n} x_{[i]}(v(\{[i], [i+1], \ldots, [n]\}) - v(\{[i+1], [i+2], \ldots, [n]\}))$$

Example

Let $v(\{1,2,3\}) = 1, v(\{1,2\}) = v(\{1,3\}) = 0.6, v(\{2,3\}) = 0.9, v(\{1\}) = 0.4, v(\{2\}) = v(\{3\}) = 0.1.$ *Then if* $\mathbf{x} = \langle 0.8, 0.3, 0.4 \rangle$, *to calculate the output:*

- *0.3, 0.4, 0.8*
- *0.3,* $v(\{1,2,3\})$ *-* $v(\{1,3\})$.
- *to work out the weight allocated to 0.4, we start with the* $v(\{1,3\})$ *and then from this remove the set that is leftover when 0.4 is removed. So we have* $v(\{1,3\}) - v(\{1\})$.

The Choquet integral

$$C_v(\mathbf{x}) = \sum_{i=1}^{n} x_{[i]}(v(\{[i],[i+1],\ldots,[n]\}) - v(\{[i+1],[i+2],\ldots,[n]\}))$$

> **Example**
>
> *Let* $v(\{1,2,3\}) = 1, v(\{1,2\}) = v(\{1,3\}) = 0.6, v(\{2,3\}) = 0.9, v(\{1\}) = 0.4, v(\{2\}) = v(\{3\}) = 0.1.$
> *Then if* $\mathbf{x} = (0.8, 0.3, 0.4)$*, to calculate the output:*

$$0.3(v(\{1,2,3\}) - v(\{1,3\})) + 0.4(v(\{1,3\}) - v(\{1\})) + 0.8(v(\{1\})$$

$$= 0.3(1 - 0.6) + 0.4(0.6 - 0.4) + 0.8(0.4)$$

$$= 0.3(0.4) + 0.4(0.2) + 0.8(0.4) = 0.52$$

THE CHOQUET INTEGRAL

$$0.3(v(\{1,2,3\}) - v(\{1,3\})) + 0.4(v(\{1,3\}) - v(\{1\})) + 0.8(v(\{1\})$$

$$= 0.3(1 - 0.6) + 0.4(0.6 - 0.4) + 0.8(0.4)$$

$$= 0.3(0.4) + 0.4(0.2) + 0.8(0.4) = 0.52$$

Notice that once we have the set of v's to use and do the subtractions, we are left with a weighting vector that sums to 1 (in this case $\langle 0.4, 0.2, 0.4 \rangle$).

The Choquet integral

From our previous scenario with workers - the Choquet integral can be interpreted as the amount of the project that can be completed if $\mathbf{x}$ represents the time commitment of each worker and all workers start at the same time (and can't return once they leave).

THE CHOQUET INTEGRAL

- However, the Choquet integral is not just about modelling worker output.
- It is a function that allocates importance to every subset of inputs.
- In this way, it is able to model interaction and redundancy.
- In some cases arguments may not be worth very much by themselves but in conjunction with others can be very important.

The Choquet integral

- Another often used example is a problem of evaluating students for a scholarship based on their scores in Mathematics, Physics and English.
- One thing that we want to account for is that students who are good at maths are often also good at physics.
- so in combining their scores, we don't necessarily want to 'double count' this ability. On the other hand, we want the importance of maths and physics to be higher than English.

Fuzzy measures make such preferences possible:

$$v(\text{maths}) = v(\text{physics}) = 0.45, v(\text{english}) = 0.3$$

$$v(\text{maths}, \text{physics}) = 0.5, v(\text{physics}, \text{english}) = v(\text{maths}, \text{english}) = 0.9$$

The Choquet integral

$v(\text{maths}) = v(\text{physics}) = 0.45, v(\text{english}) = 0.3$
$v(\text{maths}, \text{physics}) = 0.5, v(\text{physics}, \text{english}) = v(\text{maths}, \text{english}) = 0.9$

Example

Student	*maths*	*physics*	*English*
a	*18*	*16*	*10*
b	*10*	*12*	*18*
c	*14*	*15*	*15*

Student a, we have $10(1 - 0.5) + 16(0.5 - 0.45) + 18(0.45) = 13.9$
Student b, we have $10(1 - 0.9) + 12(0.9 - 0.3) + 18(0.3) = 13.6$
Student c, we have $14(1 - 0.9) + 15(0.9 - 0.3) + 15(0.3) = 14.9$

The Choquet integral

So we were able to assign higher value to maths and physics, but still reward all-round ability.

INTERACTION

In general,

- if $v(\{i,j\}) = v(\{i\}) + v(\{j\})$ we say these two inputs have no interaction or are **additive**,
- if $v(\{i,j\}) > v(\{i\}) + v(\{j\})$ we say these two inputs interact in a **complementary** or superadditive way, (positive synergy)
- if $v(\{i,j\}) < v(\{i\}) + v(\{j\})$ we say these two inputs are somewhat **redundant**, or interact in a subadditive way, (negative synergy).

The Shapley Value

In order to make the fuzzy measure easier to interpret, there is a calculation based on v called the Shapley value.

The Shapley value gives the average importance of a particular variable. For instance, with students a, b and c, the final weight associated with the English score could have been $0.5, 0.3$ or 0.6 depending on whether it was the lowest, highest or middle value respectively.

The Shapley value is the vector $\langle \phi_1, \phi_2, \ldots, \phi_n \rangle$ and is interpreted the same $\mathbf{w}$ is for the WAM.

Special cases

The Choquet integral includes both the WAM and OWA as special cases.

- WAM if $v(S) = \sum_{i \in S} v(\{i\})$.
  - Example: If

$$\begin{array}{ccc} & v(\{1,2,3\}) = 1 & \\ v(\{1,2\}) = 0.8 & v(\{1,3\}) = 0.5 & v(\{2,3\}) = 0.7 \\ v(\{1\}) = 0.3 & v(\{2\}) = 0.5 & v(\{3\}) = 0.2 \\ & v(\emptyset) = 0 & \end{array}$$

    This is equivalent to having a WAM with weights $w =< 0.3, 0.5, 0.2 >$

- OWA if $v(S) = v(T)$ whenever $|S| = |T|$.
  - Example: If

$$\begin{array}{ccc} & v(\{1,2,3\}) = 1 & \\ v(\{1,2\}) = 0.7 & v(\{1,3\}) = 0.7 & v(\{2,3\}) = 0.7 \\ v(\{1\}) = 0.2 & v(\{2\}) = 0.2 & v(\{3\}) = 0.2 \\ & v(\emptyset) = 0 & \end{array}$$

    This is equivalent to having a OWA with weights $w =< 0.3, 0.5, 0.2 >$. The weight vector is calculated as $w =< (triplet - pairs), (pairs - singleton), (singleton\ weight) >$

Example 1

9. The manager of a clothing store forecasts her monthly sales based on an OWA of the previous 4 months. Interpret the vectors $\mathbf{w} = \langle 0.1, 0.2, 0.3, 0.4 \rangle$ and $\mathbf{w} = \langle 0.4, 0.3, 0.2, 0.1 \rangle$ in terms of whether they give a 'conservative' or 'optimistic' estimate of the next month's sales.

Examples 1- Answer

The vector $\mathbf{w} = \langle 0.1, 0.2, 0.3, 0.4 \rangle$ is weighted toward higher inputs. It means that she is basing her forecasts predominantly on the best months of sales. This means the prediction will be optimistic. On the other hand, for $\mathbf{w} = \langle 0.4, 0.3, 0.2, 0.1 \rangle$, the weight is mainly based on the months where sales were low, so we can say this is conservative.

EXAMPLE 2

12. Will the fuzzy measure below define a Choquet integral that is equivalent to either the weighted arithmetic mean or the OWA? Explain why/why not.

$$v(\{1,2,3\}) = 1$$

$$v(\{1,2\}) = 0.7 \quad v(\{1,3\}) = 0.6 \quad v(\{2,3\}) = 0.8$$

$$v(\{1\}) = 0.2 \qquad v(\{2\}) = 0.5 \qquad v(\{3\}) = 0.4$$

$$v(\emptyset) = 0$$

Examples 2- Answer

By looking at the fuzzy measure we see that the values of the singletons are all different, so it can't be the same as an OWA. We also note that the measure of $\{2,3\}$ is not equal to $v(\{2\})+v(\{3\}) = 0.5+0.4 = 0.9$. So in this case, the fuzzy measure is **not** equivalent to either a WAM or an OWA.