fANOVA Generalized fANOVA

Trying to bring Hooker and Rahman together

Rahman

Classical fANOVA decomposition denotes as:

$$y(\mathbf{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_{u,C}(\mathbf{X}_u), \tag{3.1}$$

Generalized fANOVA decomposition denotes as:

$$y(\mathbf{X}) = \sum_{u \subseteq \{1,\ldots,N\}} y_{u,G}(\mathbf{X}_u), \tag{4.1}$$

So far saying nothing. The difference is in how the individual terms y_u look like.

Rahman builds classical fANOVA decomposition on the strong annihilating conditions:

3.1. Strong annihilating conditions. The strong annihilating conditions relevant to the classical ADD require all nonconstant component functions $y_{u,C}$ to integrate to *zero* with respect to the marginal density of each random variable in u, that is [26, 17, 18, 13],

$$\int_{\mathbb{R}} y_{u,C}(\mathbf{x}_u) f_{\{i\}}(x_i) dx_i = 0 \quad \text{for } i \in u \neq \emptyset, \tag{3.2}$$

From the conditions A) zero mean B) orthogonality follow:

Proposition 3.1. *The classical ADD component functions $y_{u,C}$, where $\emptyset \neq u \subseteq \{1,\ldots,N\}$, have zero means, i.e.,*

$$\mathbb{E}\left[y_{u,C}(\mathbf{X}_u)\right] = 0.$$

Proposition 3.2. *Two distinct classical ADD component functions $y_{u,C}$ and $y_{v,C}$, where $\emptyset \neq u \subseteq \{1,\ldots,N\}$, $\emptyset \neq v \subseteq \{1,\ldots,N\}$, and $u \neq v$, are orthogonal; i.e., they satisfy the property*

$$\mathbb{E}\left[y_{u,C}(\mathbf{X}_u) y_{v,C}(\mathbf{X}_v)\right] = 0.$$

The generalized fANOVA is based on the weak annihilating conditions:

4.1. Weak annihilating conditions. The weak annihilating conditions appropriate for the generalized ADD mandate all nonconstant component functions $y_{u,G}$ to integrate to *zero* with respect to the marginal density of $\mathbf{X}_u$ in each coordinate direction of u, that is [11],

$$\int_{\mathbb{R}} y_{u,G}(\mathbf{x}_u) f_u(\mathbf{x}_u) dx_i = 0 \quad \text{for } i \in u \neq \emptyset. \tag{4.2}$$

They are basically the same as the strong ones, but they work with the joint pdf of the variables of interest, to appropriately capture the dependencies that exist.
And from the weak annihilating conditions it follows that the generalized fANOVA components have A) zero mean and B) (a form of) orthogonality (just as the classical fANOVA terms have).

Proposition 4.1. *The generalized ADD component functions $y_{u,G}$, where $\emptyset \neq u \subseteq \{1,\dots,N\}$, have zero means, i.e.,*

$$\mathbb{E}\left[y_{u,G}(\mathbf{X}_u)\right] = 0.$$

where the last line follows from using (4.2).

Proposition 4.2. *Two distinct generalized ADD component functions $y_{u,G}$ and $y_{v,G}$, where $\emptyset \neq u \subseteq \{1,\dots,N\}$, $\emptyset \neq v \subseteq \{1,\dots,N\}$, and $v \subset u$, are orthogonal; i.e., they satisfy the property*

$$\mathbb{E}\left[y_{u,G}(\mathbf{X}_u) y_{v,G}(\mathbf{X}_v)\right] = 0.$$

As you see A) zero mean is formulated exactly equal but for B) orthogonality has a subtle variation, we require that $v \subset u$ which leads to hierarchical orthogonality.

The single classical component functions can be defined as:

$$y_{\emptyset,C} = \int_{\mathbb{R}^N} y(\mathbf{x}) \prod_{i=1}^{N} f_{\{i\}}(x_i) dx_i, \tag{3.3a}$$

$$y_{u,C}(\mathbf{X}_u) = \int_{\mathbb{R}^{N-|u|}} y(\mathbf{X}_u, \mathbf{x}_{-u}) \prod_{i=1, i \notin u}^{N} f_{\{i\}}(x_i) dx_i - \sum_{v \subset u} y_{v,C}(\mathbf{X}_v). \tag{3.3b}$$

And the generalized components look like this:

Theorem 4.4. *The generalized ADD component functions $y_{u,G}$, $u \subseteq \{1,\dots,N\}$, of a square-integrable function $y:\mathbb{R}^N \to \mathbb{R}$ for a given probability measure $f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$ of $\mathbf{X} \in \mathbb{R}^N$ satisfy*

$$y_{\emptyset,G} = \int_{\mathbb{R}^N} y(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \tag{4.5a}$$

$$\begin{aligned} y_{u,G}(\mathbf{X}_u) = &\int_{\mathbb{R}^{N-|u|}} y(\mathbf{X}_u, \mathbf{x}_{-u}) f_{-u}(\mathbf{x}_{-u}) d\mathbf{x}_{-u} - \sum_{v \subset u} y_{v,G}(\mathbf{X}_v) \\ &- \sum_{\substack{\emptyset \neq v \subseteq \{1,\dots,N\} \\ v \cap u \neq \emptyset, v \not\subseteq u}} \int_{\mathbb{R}^{|v \cap -u|}} y_{v,G}(\mathbf{X}_{v \cap u}, \mathbf{x}_{v \cap -u}) f_{v \cap -u}(\mathbf{x}_{v \cap -u}) d\mathbf{x}_{v \cap -u}. \end{aligned} \tag{4.5b}$$

We can study the second-moments statistics of the components to get a better understanding for their behaviour/ properties. For the second-moment statistics of the classical fANOVA terms we can say:

3.2. Second-moment statistics. Applying the expectation operators on $y(\mathbf{X})$ in (3.1) and $(y(\mathbf{X}) - \mu)^2$ and recognizing Propositions 3.1 and 3.2, the mean of y is

$$\mu := \mathbb{E}\left[y(\mathbf{X})\right] = y_{\emptyset,C}, \tag{3.4}$$

whereas its variance

$$\sigma^2 := \mathbb{E}\left[(y(\mathbf{X}) - \mu)^2\right] = \sum_{\emptyset \neq u \subseteq \{1,\dots,N\}} \mathbb{E}\left[y_{u,C}^2(\mathbf{X}_u)\right] \tag{3.5}$$

The second-moment statistics for the generalized terms; for the expected value it is the same while the variance decomposition still works but not as easily as for the classical components.

6.1. Mean and variance. Applying the expectation operator on (4.1) and noting Proposition 4.1, the mean

$$\mu := \mathbb{E}[y(\mathbf{X})] = y_{\emptyset,G} \tag{6.1}$$

of $y(\mathbf{X})$ matches the constant component function of the generalized ADD. This is similar to (3.4), the result from the classical ADD, although the respective constants involved are not the same. Applying the expectation operator again, this time on $(y(\mathbf{X}) - \mu)^2$, and recognizing Proposition 4.2 results in the variance

$$\sigma^2 := \mathbb{E}\left[(y(\mathbf{X}) - \mu)^2\right] = \sum_{\emptyset \neq u \subseteq \{1,\dots,N\}} \mathbb{E}\left[y_{u,G}^2(\mathbf{X}_u)\right] + \sum_{\substack{\emptyset \neq u,v \subseteq \{1,\dots,N\} \\ u \not\subseteq v \not\subseteq u}} \mathbb{E}\left[y_{u,G}(\mathbf{X}_u) y_{v,G}(\mathbf{X}_v)\right] \tag{6.2}$$

of $y(\mathbf{X})$, where the first sum represents variance contributions from all nonconstant component functions. In contrast, the second sum in (6.2) typifies covariance contributions from

Hooker

fANOVA decomposition written as:

We can now write $F(x)$ as

$$F(x) = \sum_{u \subseteq \{1,\ldots,d\}} f_u(x_u)$$

Same definition as Rahman (of course) just different notation.
The single fANOVA terms can be defined as follows:

$$f_u(x) = \int_{x_{-u}} \left(F(x) - \sum_{v \subset u} f_v(x) \right) dx_{-u}, \tag{3.1}$$

I think here they deviate. Hooker formulates the single terms via the projection of the differences, while Rahman subtracts the lower order terms from the projection.
And the fANOVA terms have desirable properties A) zero mean B) orthogonality C) variance decomposition, which he just lists

A list of desirable properties for the functional ANOVA can be derived sequentially:

Zero Means: $\int f_u(x_u)dx_u = 0$ for each $u \neq \emptyset$.

Orthogonality: $\int f_u(x_u) f_v(x_v) dx = 0$ for $u \neq v$.

Variance Decomposition: Let $\sigma^2_{(f)} = \int f(x)^2 dx$ then

$$\sigma^2(F) = \sum_{u \subseteq d} \sigma_u^2(f_u). \tag{3.2}$$

So he does not follow A) and B) from the annihilating conditions but states them directly. And he states C) in the same context and not in a separate context of second-moment statistics for the fANOVA terms.

Hookers generalized components are defined via a joint conditions:

Explicitly, we will jointly define all the effects $\{f_u(x_u) | u \subseteq d\}$ as satisfying

$$\{f_u(x_u) | u \subseteq d\} = \underset{\{g_u \in \mathcal{L}^2(\mathbb{R}^u)\}_{u \in d}}{\operatorname{argmin}} \int \left(\sum_{u \subseteq d} g_u(x_u) - F(x) \right)^2 w(x)dx \tag{4.1}$$

Crucial thing, again, the use of $w(x)$ which is the joint density, and assumed to be non-product type measure since we allow for independence between variables.

The generalized components also have to be centered, so Hooker formulates a generalized zero mean condition. He writes his generalized zero mean condition as a Lemma. So he does it differently than Rahman who formulates the weak annihilating conditions from which he follows A) zero mean B) hierarchical orthogonality for the generalized fANOVA components.

Lemma 1. *The orthogonality conditions (4.2) are true over $\mathcal{L}^2(\mathbb{R}^d)$ if and only if the integral conditions*

$$\forall u \subseteq d,\ \forall i \in u \int f_u(x_u) w(x) dx_i dx_{-u} = 0. \tag{4.3}$$

hold.

Hooker also formulates hierarchical orthogonality for the components instead of full orthogonality (like Rahman) but again doesn't follow from the weak annihilating conditions but he sets the hierarchical orthogonality at a constraint to the optimization problem.

under the *hierarchical orthogonality conditions*

$$\forall v \subseteq u,\ \forall g_v : \int f_u(x_u) g_v(x_v) w(x) dx = 0. \tag{4.2}$$

And because his joint optimization problem is difficult to actually solve, he rewrites the problem in a simplified from, which separates the component of interest and treats all other unknown components as nuisance parameters that may be estimated together:

$$\int \left(g_u(x_u) + g_{-u}(x_{-u}) + \sum_{v \subset u} g_v(x_v) + \sum_{i \in u} \sum_{-u \subset v' \subseteq -i} g_{v'}(x_{v'}) - F(x) \right)^2 w(x) dx. \tag{5.1}$$

Here g_{-u} is subject to the relaxed condition

$$\int g_{-u}(x_{-u}) w(x) dx_{-u} = 0,$$

and similarly

$$\int g_{v' \subset -j}(x_{v'}) w(x) dx_j dx_{-u} = 0$$