fANOVA Theory vs. Empiry

Hooker 2007

A single fANOVA effect f_u is a term that minimizes this squared distance:

2^d effects at once. Nonetheless, from (4.2), we can write an optimization criterion for a single effect, defining $f_u(x_u)$ as the first component of the collection $g_u(x_u), \{g_v(x_v)|v \subset u\}, g_{-u}(x_{-u}), \{g_{v'}(x_{v'})| -u \subset v \subsetneq -i, i \in u\}$ that minimizes:

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

- g_u is the term (or contains the term?) of interest, the argument to be found
- All other terms are terms we have to account for (either terms completely different than the ones in u or interactions with some terms in u we have to account for)
- $F(x)$ is the original function we want to approximate, to which we want to minimize the squared difference to
- $w(x)$ is a weighing function and in the classical fANOVA it is simply the uniform distribution over the unit interval (this is why we don't "notice" it in the classical one)

In practice the integral has to be estimates by the sum over a chosen number of points (e.g. uniform grid points):

We now translate (5.1) into the problem of finding $\{f_v(z_{v,k})|v \in \mathcal{I}_u, k \in 1, \ldots, N^{|v|}\}$ to minimize

$$\sum_{i=1}^{N^{d'+1}} w(z_i) \left(\sum_{v \in \mathcal{I}_u} f_v(z_{v,i}) - F(z_i) \right)^2 \tag{6.2}$$

- $w(z_i)$ is a weighing function, evaluated at a chosen sample point z_i
- $F(z_i)$ is the original function evaluated at a chosen sample point z_i
- $N^{d'+1}$ the number of grid points? d' denotes cardinality of u . (The number is somehow related to the number of terms I think? Because the theoretical generalized fANOVA has $2^d - 1$ terms, and
- And I think this $sum_{v \in \mathcal{I}_u} f_v(z_{v,i})$ contains the fANOVA terms of interest, evaluted at a grid point $z_{v,i}$. But honestly I am unsure how a single term f_v could look like.

The single fANOVA terms have to satisfy the zero mean constraint, which in theory is written as

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

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

and similarly

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

In practice the zero mean constraint looks like this:

under the constraints:

$$\forall v \in \mathcal{I}_u, \forall j \in v, \forall z_{v,k} : \sum_{i=1}^{N} \left(\sum_{l=1}^{d'-|v|+1} w(z_{v,i}, z_{-v,l}) \right) f_v(z_{v\backslash j,k}, z_{j,i}) = 0. \tag{6.3}$$

Evaluate the fANOVA component and the weighing function at at all grid points, sum it up, then this should be zero?

The minimization problem involving the sums can be rewritten as a linear system of equations, a linear least squares problem:

We can regard the effect values at the data points, $f_u(z_i)$, as parameters to be estimated. Then (6.2) can be written as a linear least squares problem:

$$\text{minimize } (Xf - F)^T W(Xf - F), \tag{6.4}$$

where F is a vector whose ith component is $F(z_i)$, the vector f lists the value of $f_u(z_i)$ for each u and i and W is a diagonal matrix with $w(z_i)$ on the diagonal. This is minimized

And the zero mean constraint in matrix form looks like this:

for each u and i and W is a diagonal matrix with $w(z_i)$ on the diagonal. This is minimized under the constraints (6.3), written as

$$Cf = 0. \tag{6.5}$$

Here both X and C perform addition on the components of f. We index the rows of X by $k \in 1 \ldots N^{d'+1}$, corresponding to grid points. The columns are indexed by (v, j), for

Hooker basically establishes a link between the new challenge of finding the generalized fANOVA terms and an established solution method.

Rahman 2014

Disclaimer: still unsure about all of these statement

Rahmans generalized fANOVA is not directly based on projections but tries to stay as close to the classical fANOVA as possible I think. So what he changes is what we subtract (the quantities we account for when calculating a single fANOVA term.) In this example we can recognize the classical fANOVA approach, but instead of only subtracting the constant term y_0 when searching for y_1 he subtracts all interactions in which the first variable is part of.

functions $y_{v,G}$, where $v \subset u$, but also on the component functions $y_{v,G}$, where $v \cap u \neq \emptyset$, $v \not\subseteq u$. As an example, consider $u = \{1\}$ and $N = 3$. The classical and generalized component functions depending on x_1 are

$$y_{\{1\},C} = \int_{\mathbb{R}^2} y(x_1, x_2, x_3) f_{\{2\}}(x_2) f_{\{3\}}(x_3) dx_2 dx_3 - y_{\emptyset,C}$$

and

$$\begin{aligned} y_{\{1\},G} = & \int_{\mathbb{R}^2} y(x_1, x_2, x_3) f_{\{2,3\}}(x_2, x_3) dx_2 dx_3 - y_{\emptyset,G} \\ & - \int_{\mathbb{R}} y_{\{1,2\},G}(x_1, x_2) f_{\{2\}}(x_2) dx_2 - \int_{\mathbb{R}} y_{\{1,3\},G}(x_1, x_3) f_{\{3\}}(x_3) dx_3 \\ & - \int_{\mathbb{R}^2} y_{\{1,2,3\},G}(x_1, x_2, x_3) f_{\{2,3\}}(x_2, x_3) dx_2 dx_3, \end{aligned}$$

In practice he implements his generalized fANOVA decomposition approach with Fourier-Polynomial Expansion (don't know yet what this is), and also somehow uses his fANOVA results to estimate marginal densities?? (because his plots show marginal desnities).

I think Rahman plots have sth to do with the following quantity:

The truncations introduced in Corollary 5.2 engender an S-variate, mth-order generalized ADD approximation

$$\tilde{y}_{S,m}(\mathbf{X}) = y_{\emptyset,G} + \sum_{\substack{\emptyset \neq u \subseteq \{1,\ldots,N\} \\ 1 \leq |u| \leq S}} \sum_{k=1}^{m} \sum_{\substack{|\mathbf{j}_{|u|}| = k \\ j_1,\ldots,j_{|u|} \neq 0}} \tilde{C}_{u\mathbf{j}_{|u|}} \psi_{u\mathbf{j}_{|u|}} \tag{5.15}$$

- this quantity approximates a generalized fANOVA decomposition with S variables and m terms?
  I think the following is the theoretical definition of the fANOVA decomposition based on Fourier-Polynomial Expansion

Theorem 5.1. *Let y be a square-integrable function of $\mathbf{X}$, admitting a generalized ADD, where $\mathbf{X} = (X_1, \ldots, X_N)$ is an $\mathbb{R}^N$-valued dependent random vector with an arbitrary non-product-type joint probability density function $f_{\mathbf{X}} : \mathbb{R}^N \to \mathbb{R}_0^+$ and a marginal probability density function f_u of $\mathbf{X}_u$. Given $\emptyset \neq u \subseteq \{1, \ldots, N\}$, let $\{\psi_{v\mathbf{k}_{|v|}}(\mathbf{X}_v), \emptyset \neq v \subseteq u, \mathbf{k}_{|v|} \in \mathbb{N}_0^{|v|}, k_1, \ldots, k_{|v|} \neq 0\}$ be a nested set of measure-consistent orthonormal polynomial basis functions such that $\mathbb{E}[\psi_{u\mathbf{j}_{|u|}}(\mathbf{X}_u)] = 0$ and $\mathbb{E}[\psi_{u\mathbf{j}_{|u|}}(\mathbf{X}_u)\psi_{v\mathbf{k}_{|v|}}(\mathbf{X}_v)] = 0$ for $\emptyset \neq v \subset u$, $j_1, \ldots, j_{|u|} \neq 0$, and $k_1, \ldots, k_{|v|} \neq 0$. Then the expansion coefficients of the polynomial representation of nonconstant component functions of y in (5.2) and (5.3) satisfy*

$$C_{u\mathbf{j}_{|u|}} + \sum_{\substack{\emptyset \neq v \subseteq \{1,\ldots,N\} \\ v \cap u \neq \emptyset, v \not\subseteq u}} \sum_{\substack{\mathbf{k}_{|v|} \in \mathbb{N}_0^{|v|} \\ k_1,\ldots,k_{|v|} \neq 0}} C_{v\mathbf{k}_{|v|}} J_{u\mathbf{j}_{|u|}, v\mathbf{k}_{|v|}} = I_{u\mathbf{j}_{|u|}}, \tag{5.8}$$

But (5.8) is an infinite sum which has to be estimated. Gives the estimate later.