Solution 1:

a) **Step 1. Compute the constant term.**

$$g_\emptyset = \mathbb{E}[\hat{f}(X_1, X_2, X_3)] = \mathbb{E}[4X_1 + 2X_2X_3 - 4] = 4\ \mathbb{E}[X_1] + 2\ \mathbb{E}[X_2]\ \mathbb{E}[X_3] - 4 = -4,$$

because the variables are all independent and centered (i.e., $\mathbb{E}[X_i] = 0$ for all i.).

Alternatively, using all data points:

$$\begin{aligned} g_\emptyset &= \frac{1}{27} \sum_{x_1,x_2,x_3 \in \{-1,0,1\}} \big(4x_1 + 2x_2x_3 - 4\big) = \frac{1}{27}\Big(\underbrace{9 \cdot 4 \cdot (-1+0+1)}_{\text{All possible values of } x_1} \\ &\quad + \underbrace{3 \cdot 2 \cdot \big(-1\cdot(-1) + (-1)\cdot 0 + (-1)\cdot 1 + 0\cdot(-1) + 0\cdot 0 + 0\cdot 1 + 1\cdot(-1) + 1\cdot 0 + 1\cdot 1\big)}_{\text{All possible combinations of values for } x_2 \text{ and } x_3} - 27\cdot 4\Big) \\ &= \frac{1}{27}\Big(0 + 0 - 27\cdot 4\Big) = -4, \end{aligned}$$

because each value of x_1 occurs 9 times and each combination of values of x_2 and x_3 occurs 3 times.

Analogously, for all the PD-functions (which are marginal expectations), we can use the fact that the variables are independent and centered.

Step 2. First-order terms.

$$g_1(x_1) = \mathbb{E}[\hat{f}(x_1, X_2, X_3)] - g_\emptyset = \mathbb{E}[4x_1 + 2X_2X_3 - 4] - g_\emptyset = 4x_1 + 2\ \underbrace{\mathbb{E}[X_2]}_{=0}\ \underbrace{\mathbb{E}[X_3]}_{=0} - 4 - (-4) = 4x_1.$$

$$g_2(x_2) = \mathbb{E}[\hat{f}(X_1, x_2, X_3)] - g_\emptyset = \mathbb{E}[4X_1 + 2x_2X_3 - 4] - g_\emptyset = 4\underbrace{\mathbb{E}[X_1]}_{=0} + 2x_2\underbrace{\mathbb{E}[X_3]}_{=0} - 4 - (-4) = 0.$$

Because $\hat{f}$ is symmetric in x_2 and x_3, we get analogously

$$g_3(x_3) = 0.$$

Step 3. Second-order terms.

$$g_{1,2}(x_1, x_2) = \mathbb{E}[\hat{f}(x_1, x_2, X_3)] - g_1(x_1) - g_2(x_2) - g_\emptyset = \mathbb{E}[4x_1 + 2x_2X_3 - 4] - 4x_1 + 4 = 4x_1 + 2x_2\underbrace{\mathbb{E}[X_3]}_{=0} - 4 - 4x_1 + 4 = 0.$$

Again using that $\hat{f}$ is symmetric in x_2 and x_3, we receive exactly the same result for $S = \{1, 3\}$, so $g_{13}(x_1, x_3) = 0$.

For $\{2, 3\}$, we get:

$$g_{2,3}(x_2, x_3) = \mathbb{E}[\hat{f}(X_1, x_2, x_3)] - g_2(x_2) - g_3(x_3) - g_\emptyset = \mathbb{E}[4X_1 + 2x_2x_3 - 4] + 4 = 4\underbrace{\mathbb{E}[X_1]}_{=0} + 2x_2x_3 - 4 + 4 = 2x_2x_3.$$

Step 4. Third-order term.

$$g_{1,2,3}(x_1, x_2, x_3) = \hat{f}(x_1, x_2, x_3) - \sum_{V \subsetneq \{1,2,3\}} g_V(x_V) = \underbrace{4x_1 + 2x_2x_3 - 4}_{=\hat{f}(x_1,x_2,x_3)} - \Big(\underbrace{-4}_{=g_\emptyset} + \underbrace{4x_1}_{=g_1(x_1)} + \underbrace{2x_2x_3}_{=g_{2,3}(x_2,x_3)}\Big) = 0.$$

So, all in all, we only have three nonzero components in the fANOVA, which are exactly equal to the terms in the given equation that we would expect:

$$g_\emptyset = -4, \quad g_1 = 4x_1, \quad g_{2,3} = 2x_2x_3.$$

b) **Orthogonality check.** As indicated by the hint, we only need to consider pairs of two nonzero components, since otherwise the terms are trivially orthogonal. This is because for any U, V with one component being 0, w.l.o.g. $g_U(\mathbf{X}_U) = 0$, we have:

$$\mathbb{E}_{\mathbf{X}}\left[g_U(\mathbf{X}_U)\, g_V(\mathbf{X}_V)\right] = \mathbb{E}_{\mathbf{X}}\left[0 \cdot\ g_V(\mathbf{X}_V)\right] = \mathbb{E}_{\mathbf{X}}\left[0\right] = 0.$$

For the remaining components, we have three possibilities we need to check:

$$\begin{aligned}
&\text{For } (V,U) = (\emptyset, \{1\}): \ \mathbb{E}_{\mathbf{X}}\left[g_\emptyset\, g_1(X_1)\right] = \mathbb{E}_{\mathbf{X}}\left[-4 \cdot 4X_1\right] = -16\mathbb{E}[X_1] = 0,\\
&\text{For } (V,U) = (\emptyset, \{2,3\}): \ \mathbb{E}_{\mathbf{X}}\left[g_\emptyset\, g_{2,3}(\mathbf{X}_{2,3})\right] = \mathbb{E}_{\mathbf{X}}\left[-4 \cdot 2X_2X_3\right] = -8\ \mathbb{E}[X_2]\ \mathbb{E}[X_3] = 0,\\
&\text{For } (V,U) = (\{1\}, \{2,3\}): \ \mathbb{E}_{\mathbf{X}}\left[g_1(X_1)\, g_{2,3}(\mathbf{X}_{2,3})\right] = \mathbb{E}_{\mathbf{X}}\left[4X_1 \cdot 2X_2X_3\right] = 8\ \mathbb{E}[X_1]\ \mathbb{E}[X_2]\ \mathbb{E}[X_3] = 0,
\end{aligned}$$

again using independence and centering. This shows the desired orthogonality.

One can also see in this exercise, that

- the symmetry of $\hat{f}$ in x_2 and x_3 translates to the fANOVA components. In general, the whole fANOVA decomposition inherits the same symmetries as the original function.
- after computing the expected value $g_\emptyset$ in the first step, one could center the whole function $\hat{f}^c(x_1, x_2, x_3) = \hat{f}(x_1, x_2, x_3) - g_\emptyset$ and then use this centered function $\hat{f}^c$ throughout. In this case, we would not need to subtract $g_\emptyset$ in each step anymore.

Solution 2:

a) First we evaluate f on all points:

	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	2	5
$x_1 = 1$	4	9

Hence $\mu = \frac{1}{4}(2 + 4 + 5 + 9) = 5$. For the one-dimensional PD-functions, i.e. the PDPs, we get

$$\begin{aligned}
&\hat{f}_{1,PD}(0) = \tfrac{1}{2}(2+5) = 3.5, \quad \hat{f}_{1,PD}(1) = \tfrac{1}{2}(4+9) = 6.5, \qquad \text{hence } \hat{f}_{1,PD}(x_1) = 3.5 + 3x_1,\\
&\hat{f}_{2,PD}(0) = \tfrac{1}{2}(2+4) = 3, \qquad \hat{f}_{2,PD}(1) = \tfrac{1}{2}(5+9) = 7 \qquad\ \ \text{hence } \hat{f}_{2,PD}(x_2) = 3 + 4x_2.
\end{aligned}$$

b) The PD-functions have the means

$$\begin{aligned}
\mathbb{E}[\hat{f}_{1,PD}(X_1)] &= \tfrac{1}{2}(3.5 + 6.5) = 5 = \mu,\\
\mathbb{E}[\hat{f}_{2,PD}(X_2)] &= \tfrac{1}{2}(3. + 7) = 5 = \mu,
\end{aligned}$$

as expected. We use the second formula for the H-statistic here. The numerator is equal to:

$$\hat{f}(x_1,x_2) - \hat{f}_{1,PD}(x_1) - \hat{f}_{2,PD}(x_2) + \mu = \begin{array}{c|cc} & x_2 = 0 & x_2 = 1 \\ \hline x_1 = 0 & 0.5 & -0.5 \\ x_1 = 1 & -0.5 & 0.5 \end{array} = g_{1,2}(x_1, x_2),$$

which is equal to the interaction component of the fANOVA decomposition.

We then get for the H-statistic

$$\begin{aligned}
H^2_{1,2} &= \frac{\sum_{i,j=1}^{2} \left(\hat{f}_{12,PD}(x_1^{(i)}, x_2^{(j)}) - \hat{f}_{1,PD}(x_1^{(i)}) - \hat{f}_{2,PD}(x_2^{(j)}) + \mu\right)^2}{\sum_{i,j=1}^{2} \left(\hat{f}(x_1^{(i)}, x_2^{(j)}) - \mu\right)^2}\\
&= \frac{4 \cdot 0.5^2}{(2-5)^2 + (4-5)^2 + (5-5)^2 + (9-5)^2} = \frac{1}{26},
\end{aligned}$$

hence $H_{1,2} = \sqrt{\frac{1}{26}} \approx 0.196$.

c) We first consider two-way interactions for an arbitrary function $\hat{f}$ with p features. W.l.o.g., we consider the two features x_1 and x_2.

We start with the third formula given, which uses the general variance. We denote the numerator as

$$\tilde{g}_{1,2}(x_1, x_2) := \hat{f}_{12,PD}(x_1, x_2) - \hat{f}_{1,PD}(x_1) - \hat{f}_{2,PD}(x_2), \text{ and}$$
$$\tilde{g}^c_{1,2}(x_1, x_2) := \hat{f}^c_{12,PD}(x_1, x_2) - \hat{f}^c_{1,PD}(x_1) - \hat{f}^c_{2,PD}(x_2)$$

for the centered version.

We know that all PD-functions have the same expectation μ (and by subtraction of μ one obtains the centered versions), which implies

$$\mathbb{E}\left[\tilde{g}_{1,2}(X_1, X_2)\right] = \mu - \mu - \mu = -\mu \text{ and } \mathbb{E}\left[\tilde{g}^c_{1,2}(X_1, X_2)\right] = 0,$$

the second equation following from the fact that all centered PD-functions have mean 0.

Using this, we can then calculate that the numerator is equal to

$$\begin{aligned}
&\mathsf{Var}\Big[\hat{f}_{12,PD}(X_1, X_2) - \hat{f}_{1,PD}(X_1) - \hat{f}_{2,PD}(X_2)\Big] \\
&= \mathsf{Var}\Big[\tilde{g}_{1,2}(X_1, X_2)\Big] = \mathbb{E}\Big[\left(\tilde{g}_{1,2}(X_1, X_2) - \mathbb{E}\left[\tilde{g}_{1,2}(X_1, X_2)\right]\right)^2\Big] = \mathbb{E}\Big[\left(\tilde{g}_{1,2}(X_1, X_2) + \mu\right)^2\Big] \\
&= \mathbb{E}\Big[\Big(\hat{f}_{12,PD}(x_1, x_2) - \hat{f}_{1,PD}(x_1) - \hat{f}_{2,PD}(x_2) + \mu\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big((\hat{f}_{12,PD}(X_1, X_2) - \mu) - (\hat{f}_{1,PD}(X_1) - \mu) - (\hat{f}_{2,PD}(X_2) - \mu)\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\hat{f}^c_{12,PD}(x_1, x_2) - \hat{f}^c_{1,PD}(x_1) - \hat{f}^c_{2,PD}(x_2)\Big)^2\Big] \\
&= \mathbb{E}\Big[\left(\tilde{g}^c_{1,2}(X_1, X_2)\right)^2\Big] \overset{\mathbb{E}\left[\tilde{g}^c_{1,2}(X_1, X_2)\right]=0}{=} \mathsf{Var}\Big[\left(\tilde{g}^c_{1,2}(X_1, X_2)\right)^2\Big] \\
&= \mathsf{Var}\Big[\hat{f}^c_{12,PD}(x_1, x_2) - \hat{f}^c_{1,PD}(x_1) - \hat{f}^c_{2,PD}(x_2)\Big],
\end{aligned}$$

in the middle also showing once again that

$$\tilde{g}^c_{1,2}(x_1, x_2) = \hat{f}^c_{12,PD}(x_1, x_2) - \hat{f}^c_{1,PD}(x_1) - \hat{f}^c_{2,PD}(x_2) = \tilde{g}_{1,2}(x_1, x_2) + \mu,$$

hence $\tilde{g}^c_{1,2}(x_1, x_2)$ is indeed the centered version of $\tilde{g}_{1,2}(x_1, x_2)$.

For the denominator, we have a very similar, but easier, calculation:

$$\begin{aligned}
\mathsf{Var}\Big[\hat{f}_{12,PD}(X_1, X_2)\Big] &= \mathbb{E}\Big[\Big(\hat{f}_{12,PD}(X_1, X_2) - \mathbb{E}\left[\hat{f}_{12,PD}(X_1, X_2)\right]\Big)^2\Big] = \mathbb{E}\Big[\Big(\hat{f}^c_{12,PD}(x_1, x_2)\Big)^2\Big] \\
&= \mathsf{Var}\Big[\hat{f}^c_{12,PD}(x_1, x_2)\Big]
\end{aligned}$$

These calculations show that the two versions of the formula, the one with the uncentered PD-functions and the one with the centered PD-functions, are equal.

If we instead have an empirical distribution from an empirical dataset, or a finite probability distribution, we can replace the variance with the empirical variance and obtain the other two formulae given in the exercise. For example, for the uncentered version we have in this case

$$\mathsf{Var}\Big[\hat{f}_{12,PD}(x_1, x_2) - \hat{f}_{1,PD}(x_1) - \hat{f}_{2,PD}(x_2)\Big] = \sum_{i=1}^{n}\big(\hat{f}_{12,PD}(x_1^{(i)}, x_2^{(i)}) - \hat{f}_{1,PD}(x_1^{(i)}) - \hat{f}_{2,PD}(x_2^{(i)}) + \mu\big)^2$$

Analogous for all the other terms.

In the same way, one can prove that the centered and uncentered versions are equivalent for higher interaction orders. For example, for interaction order 3, we have

One can prove that this pattern of the positive and negative signs always occurs for higher interaction orders.

Solution 3:

The relationship between Friedman's H-statistic and Sobol indices can be understood through their shared foundation in functional decomposition. Both measures quantify interaction strength but from different perspectives.

Foundation: Functional ANOVA Decomposition

Both measures rely on the functional ANOVA decomposition:

$$\hat{f}(\mathbf{x}) = g_\emptyset + \sum_i g_i(x_i) + \sum_{i<j} g_{i,j}(x_i, x_j) + \sum_{i<j<k} g_{i,j,k}(x_i, x_j, x_k) + \dots$$

Two-Way Interactions

For a two-way interaction between features j and k:

Sobol Index: The Sobol index $S_{j,k}$ measures the fraction of total variance explained by the pure interaction component:

$$S_{j,k} = \frac{\mathsf{Var}[g_{j,k}(X_j, X_k)]}{\mathsf{Var}[\hat{f}(\mathbf{X})]}$$

Equivalently, we can write: $\mathsf{Var}[g_{j,k}(X_j, X_k)] = S_{j,k} \cdot \mathsf{Var}[\hat{f}(\mathbf{X})]$, which will be used in the relationship below.

H-statistic: The H-statistic measures interaction strength through partial dependence functions:

$$H^2_{j,k} = \frac{\mathsf{Var}[\hat{f}^c_{j,k,PD}(X_j, X_k) - \hat{f}^c_{j,PD}(X_j) - \hat{f}^c_{k,PD}(X_k)]}{\mathsf{Var}[\hat{f}^c_{j,k,PD}(X_j, X_k)]}$$

Key Relationship: Under the standard fANOVA assumptions (independent features), the numerator of $H^2_{j,k}$ equals $\mathsf{Var}[g_{j,k}(X_j, X_k)]$, i.e., the interaction residual in the H-statistic corresponds exactly to the pure interaction component in fANOVA.

Deriving the Clean Relationship:

Step 1 - Decomposition of PD functions: Under feature independence, the standard fANOVA algorithm gives us:

$$\begin{aligned}
\hat{f}^c_{j,k,PD}(X_j, X_k) &= g_{j,k}(X_j, X_k) + g_j(X_j) + g_k(X_k) \\
\hat{f}^c_{j,PD}(X_j) &= g_j(X_j) \\
\hat{f}^c_{k,PD}(X_k) &= g_k(X_k)
\end{aligned}$$

Step 2 - Interaction residual simplifies: The numerator of $H^2_{j,k}$ becomes:

$$\begin{aligned}
&\mathsf{Var}[\hat{f}^c_{j,k,PD}(X_j, X_k) - \hat{f}^c_{j,PD}(X_j) - \hat{f}^c_{k,PD}(X_k)] \\
&= \mathsf{Var}[g_{j,k}(X_j, X_k) + g_j(X_j) + g_k(X_k) - g_j(X_j) - g_k(X_k)] \\
&= \mathsf{Var}[g_{j,k}(X_j, X_k)]
\end{aligned}$$

Note that this is exactly the same numerator as in the Sobol index definition.

Step 3 - Denominator via orthogonality: Under independence, the fANOVA components are orthogonal (uncorrelated), so for the denominator:

$$\begin{aligned}
\mathsf{Var}[\hat{f}^c_{j,k,PD}(X_j, X_k)] &= \mathsf{Var}[g_{j,k}(X_j, X_k) + g_j(X_j) + g_k(X_k)] \\
&= \mathsf{Var}[g_{j,k}(X_j, X_k)] + \mathsf{Var}[g_j(X_j)] + \mathsf{Var}[g_k(X_k)]
\end{aligned}$$

Step 4 - Final H-statistic formula: Combining the results:

$$H_{j,k}^2 = \frac{\mathsf{Var}[g_{j,k}(X_j, X_k)]}{\mathsf{Var}[g_{j,k}(X_j, X_k)] + \mathsf{Var}[g_j(X_j)] + \mathsf{Var}[g_k(X_k)]}$$

Step 5 - Converting to Sobol indices: Using $\mathsf{Var}[g_V(\mathbf{X}_V)] = S_V \cdot \mathsf{Var}[\hat{f}(\mathbf{X})]$:

$$\begin{aligned} H_{j,k}^2 &= \frac{S_{j,k} \cdot \mathsf{Var}[\hat{f}(\mathbf{X})]}{S_{j,k} \cdot \mathsf{Var}[\hat{f}(\mathbf{X})] + S_j \cdot \mathsf{Var}[\hat{f}(\mathbf{X})] + S_k \cdot \mathsf{Var}[\hat{f}(\mathbf{X})]} \\ &= \frac{S_{j,k} \cdot \mathsf{Var}[\hat{f}(\mathbf{X})]}{\mathsf{Var}[\hat{f}(\mathbf{X})] \cdot (S_{j,k} + S_j + S_k)} \\ &= \frac{S_{j,k}}{S_j + S_k + S_{j,k}} \end{aligned}$$

Three-Way Interactions

Similarly, we can derive that for a three-way interaction:

$$H_{i,j,k}^2 = \frac{S_{i,j,k}}{S_i + S_j + S_k + S_{i,j} + S_{i,k} + S_{j,k} + S_{i,j,k}}$$

General Case

For a general k-way interaction among features in set $S \subseteq \{1, 2, \ldots, p\}$, the general relationship is:

$$H_S^2 = \frac{S_S}{\sum_{V \subseteq S} S_V}$$

Solution 4:

TO DO

Solution, 1. part

Step 1: Mean μ.

$$\mu = \int_0^1 \int_0^1 \left(x_1^2 + x_2^2 + \beta\, x_1 x_2\right) dx_1\, dx_2 = \left(\tfrac{1}{3} + \tfrac{1}{3}\right) + \beta\ \tfrac{1}{4} = \tfrac{2}{3} + \tfrac{\beta}{4}.$$

Step 2: Marginal PDs (centered).

$$\begin{aligned} \hat{f}_1(x_1) &= \int_0^1 \left(x_1^2 + t^2 + \beta\, x_1 t\right) dt = x_1^2 + \tfrac{1}{3} + \tfrac{\beta}{2} x_1, \\ \hat{f}_2(x_2) &= \int_0^1 \left(t^2 + x_2^2 + \beta\, t x_2\right) dt = x_2^2 + \tfrac{1}{3} + \tfrac{\beta}{2} x_2. \end{aligned}$$

Subtracting μ gives

$$\hat{f}_{1,c}(x_1) = x_1^2 + \tfrac{\beta}{2} x_1 - \left(\tfrac{\beta}{4} + \tfrac{1}{3}\right), \quad \hat{f}_{2,c}(x_2) = x_2^2 + \tfrac{\beta}{2} x_2 - \left(\tfrac{\beta}{4} + \tfrac{1}{3}\right).$$

Step 3: Variances. Define the *interaction residual*

$$r(X) = f_\beta(X) - \hat{f}_{1,c}(X_1) - \hat{f}_{2,c}(X_2).$$

One computes by direct integration

$$\mathrm{Var}\big(f_\beta(X)\big) = \frac{\beta^2 + 6}{36}, \qquad \mathrm{Var}\big(r(X)\big) = \frac{\beta^2}{36}.$$

Hence by definition

$$H(\beta) = \sqrt{\frac{\mathrm{Var}(r(X))}{\mathrm{Var}(f_\beta(X))}} = \sqrt{\frac{\beta^2/36}{(\beta^2+6)/36}} = \frac{|\beta|}{\sqrt{\beta^2 + 6}}.$$

Step 4: Limits. As $\beta \to 0$, $H(\beta) \to 0$ (no interaction). As $|\beta| \to \infty$, $H(\beta) \to 1$ (interaction dominates). □

Solution, 2. part

(a) $H(\gamma)$**.** Identical to the calculation in Exercise **??** with $\beta \mapsto \gamma$.

(b) $S_{T,x}(\gamma)$**.**

$$\mathrm{Var}(f_\gamma) = \mathrm{Var}(x) + \mathrm{Var}(y) + \mathrm{Var}(\gamma xy) = \tfrac{1}{12} + \tfrac{1}{12} + \tfrac{\gamma^2}{36} = \tfrac{2}{12} + \tfrac{\gamma^2}{36}.$$

The main-effect variance of y is $\mathrm{Var}(\mathbb{E}[f_\gamma \mid Y]) = \mathrm{Var}(y) = 1/12$, so

$$S_{T,x} = 1 - \frac{1/12}{2/12 + \gamma^2/36} = 1 - \frac{1}{1 + \gamma^2/6}.$$

(c) Monotonicity. Compute

$$\frac{d}{d|\gamma|} H(\gamma) = \frac{6}{(\gamma^2 + 6)^{3/2}} > 0, \quad \frac{d}{d|\gamma|} S_{T,x}(\gamma) = \frac{(2/6)\,|\gamma|}{(1 + \gamma^2/6)^2} > 0.$$

Thus both increase in $|\gamma|$, so they rank interaction strength identically. □