

Background

Functional regression is a statistical method used when the predictors (independent variables) are functions rather than scalar values. This type of analysis is common in fields like environmental science, where data might be collected continuously over time, or in medical science, where data like heart rate or brain activity are recorded as functions over time.

Default Formulation

In the simplest form of functional regression, you might have a single functional predictor and a scalar response. The model looks like this:

$$y_i = \alpha + \int_T x_i(t)\beta(t)dt + \epsilon_i$$

- y_i is the scalar response for the i th observation.
- α is the intercept term.
- $x_i(t)$ is the functional predictor for the i th observation.
- $\beta(t)$ is the functional coefficient that needs to be estimated.
- ϵ_i is the error term, often assumed to be Gaussian.
- The integral is taken over the domain T of the function.

Extended Formulation for Multiple Functional Predictors

When you have multiple functional predictors, the model becomes more complex. This is where the FRM you described comes into play:

$$y_i = \beta_0 + \sum_{m=1}^M \int_{T_m} x_{im}(t_m)\beta_m(t_m)dt_m + \epsilon_i$$

- y_i is still the scalar response.
- β_0 is the intercept term.
- $x_{im}(t_m)$ are the functional predictors, where m indexes over different predictors.
- $\beta_m(t_m)$ are the functional coefficients for each predictor.
- ϵ_i is the error term.
- The integral is taken over the domain T_m of each functional predictor.

Representation of Predictors and Coefficients

Both predictors and coefficients are often represented using basis expansions:

1. **Predictors:** $x_{im}(t_m) = \sum_{j=1}^{p_m} \omega_{imj} \phi_{mj}(t_m)$
 - $\phi_{mj}(t_m)$ are basis functions (e.g., B-Splines).
 - ω_{imj} are coefficients of these basis functions.
2. **Coefficients:** $\beta_m(t_m) = \sum_{j=1}^{p_m} b_{mj} \phi_{mj}(t_m)$
 - b_{mj} are coefficients for the basis functions in representing the functional coefficients.

Each predictor $x_{im}(t_m)$ (where m indexes the predictor) is expanded into a linear combination of basis functions $\phi_{mj}(t_m)$ with coefficients ω_{imj} .

This expansion simplifies the functional form into a sum of known, manageable functions (the basis functions).

The functional coefficients $\beta_m(t_m)$ in the model are also expanded in a similar way using the same or a different set of basis functions.

Linear Model Representation

- Substitute these expansions into the original integral:

$$\int_{T_m} x_{im}(t_m) \beta_m(t_m) dt_m = \int_{T_m} \left(\sum_{j=1}^{p_m} \omega_{imj} \phi_{mj}(t_m) \right) \left(\sum_{k=1}^{p_m} b_{mk} \phi_{mk}(t_m) \right) dt_m$$

- Simplify the integral by multiplying the sums and integrating:

$$= \sum_{j=1}^{p_m} \sum_{k=1}^{p_m} \omega_{imj} b_{mk} \int_{T_m} \phi_{mj}(t_m) \phi_{mk}(t_m) dt_m$$

- This is where the cross-product matrix J_{ϕ_m} comes into play. Each element of J_{ϕ_m} is essentially one of these integrals.
- Let J_{ϕ_m} be the matrix with elements $\int_{T_m} \phi_{mj}(t_m) \phi_{mk}(t_m) dt_m$.
 - Let W_{im} be the vector $(\omega_{im1}, \dots, \omega_{imp_m})$.
 - Let b_m be the vector $(b_{m1}, \dots, b_{mp_m})$.

Combining all these, the linearized form of the FRM becomes:

$$y_i = \beta_0 + \sum_{m=1}^M W_{im}^T J_{\phi_m} b_m + \epsilon_i$$

- W_{im} and b_m are vectors of coefficients from the basis expansion.

- J_{ϕ_m} are cross-product matrices of the basis functions.

Matrix J (Cross Product Matrix):

- J_{ϕ_m} represents the cross-product matrices of basis functions.
- It is essentially a matrix of integrals of the product of pairs of basis functions over the domain T_m .
- Mathematically, J_{ϕ_m} is calculated as $\int_{T_m} \phi_m(t_m) \phi_m^T(t_m) dt_m$.
- These matrices are crucial because they encapsulate the interaction of basis functions across the domain, capturing the essence of the functional form of predictors and coefficients.

The Concept of Kernel and Its Applicability:

- In functional data analysis, a kernel often refers to a function used to understand the interaction or similarity between different points in the data.
- In the context of FRM, while the term "kernel" is not explicitly used, the concept is implicitly present in the cross-product matrices J . These matrices, through the integration of products of basis functions, essentially measure the interaction of these functions across the domain, akin to what a kernel does in measuring similarity or interaction.
- The use of B-splines as basis functions also aligns with kernel methods, as they effectively capture the underlying structure of the functional data, similar to how a kernel in machine learning captures patterns or similarities in the data.

Linear Model Final Formulation

Finally, for computational purposes, the model can be expressed in matrix form:

$$Y = Zb + \epsilon$$

- Y is the vector of responses.
- Z is a matrix constructed from the coefficients and basis functions.
- b is a vector of all the coefficients including β_0 .
- ϵ is the vector of error terms.

Matrix Z (Design Matrix):

- Z is the design matrix in the linear model.
- Each row of Z , denoted as Z_i , corresponds to an observation and includes transformed predictor information.
- Specifically, Z_i is composed of the product of the coefficients of basis expansions and the cross-product matrices J_{ϕ_m} .
- This structure makes Z key for linking the functional form of the data to the linear model.

In summary, the transition from the functional form to the linear form involves expanding the functional predictors and coefficients using basis functions, simplifying the integral into a series of summations, and then translating these into vector and matrix products, leveraging linear algebra. This transformation greatly simplifies the computational complexity of the model while preserving the essential characteristics of the functional data.