Some Basis Functions for Principal Fitted Components

Kofi Placid Adragni

Department of Biostatistics

University of Alabama at Birmingham

November 23, 2009

Abstract

We provide a list of basis functions to be used in conjunction with Principal Fitted Components Models. This document gives the correction of some erratum in my PhD dissertation and add few more details. An R code for the basis functions presented in this document is available.

We assume that the reader of this document knows about Principal Fitted Components models (Cook, 2007). Assuming that $(Y, \mathbf{X})$ has a joint distribution, the randomness of $\mathbf{X}$ is used and, denoting $\mathbf{X}_y$ to be the conditional $\mathbf{X}|Y = y$, a PFC model is written as

$$\mathbf{X}_y = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\beta}\mathbf{f}_y + \boldsymbol{\varepsilon} \tag{1}$$

The term $\mathbf{f}_y$, centered such that $\bar{\mathbf{f}_y} = 0$, is a vector-valued function of the response y. It is constructed under specific basis functions. There is an infinite list of basis functions. We provide a list of some that are useful to be used with PFC models.

The Basis Functions

Given a function $\boldsymbol{\nu} = \boldsymbol{\nu}(y)$, we want to find the transformations $\mathbf{f}_y = (f_1(y), ..., f_r(y))^T$ such that

$$\boldsymbol{\nu}(y) = \sum_{i=1}^{r} \boldsymbol{\beta}_i f_i(y).$$

The known function $\mathbf{f}_y$ constitutes the basis functions to be used. In the thesis (and in this document), polynomial, piecewise continuous and discontinuous polynomial and Fourier basis functions are considered. In all cases, we assume that the response variable is univariate, although there is nothing in the theory that requires this restriction.

The polynomial approach derives from the Taylor theorem: A function $\boldsymbol{\nu}$ at the point y can be approximated in a neighborhood of y by a linear combination of polynomials. In general, one can approximate a nonlinear function by a polynomial. A polynomial basis consists of the powers of y, that is, 1, y, y^2, ... , y^r. For this work, we consider r^{th}-degree polynomial bases. The linear basis $\mathbf{f}_y = y$, $\mathbf{f}_y \in \mathbb{R}$, the quadratic basis $\mathbf{f}_y = (y, y^2)^T$, $\mathbf{f}_y \in \mathbb{R}^2$ and cubic basis $\mathbf{f}_y = (y, y^2, y^3)^T$, $\mathbf{f}_y \in \mathbb{R}^3$ are mentioned in Cook (2007) and are particular cases of polynomial bases.

To determine piecewise basis functions, the range of y is sliced into h slices $\mathrm{H}_1, ..., \mathrm{H}_h$. Within each slice, a constant, linear, quadratic or cubic polynomial basis is used. Except for the constant intra slice basis, we consider two cases: in the first, the curves from adjacent slices are discontinuous. We refer to this as the *piecewise discontinuous basis*. In the second case, the curves are continuous without being necessarily differentiable at the joints. This is the *piecewise continuous basis*. We consider the following notations: for the k^{th} slice, n_k is the number of observations it contains and $n = \sum n_k$. We denote by $J_k(y)$ the indicator function such that $J_k(y) = 1$ if $y \in \mathrm{H}_k$ and $J_k(y) = 0$ otherwise. We also denote by $\tau_0, \tau_1, ...\tau_h$, the end-points of the slices. For example, (τ_0, τ_1) are the end-points of the first slice; (τ_1, τ_2) are the end-points of the second slice, and so on.

For piecewise discontinuous bases, a constant, a linear, a quadratic or a cubic polynomial is fitted within each slice. For a polynomial of degree m, there are $(m+1)h$ parameters to determine. The general form of the components f_{y_i} of $\mathbf{f}_y$ where $\mathbf{f}_y \in \mathbb{R}^{(m+1)h-1}$ is obtained. This yields the relationship between the number of slices and the dimension of $\mathbf{f}_y$. Here $r = (m+1)h - 1$ when h slices are used.

A linear, a quadratic and a cubic polynomial basis within the slices are also considered for the piecewise continuous case. Unlike the discontinuous case, curves from adjacent slices are continuous at each of the $(h-1)$ inner knots. For a piecewise linear polynomial, $2h$ parameters are needed but there is one constraint at each knot. The number of parameters to determine is $2h - (h-1) = h + 1$.

In the piecewise continuous quadratic case, we can set one or two constraints at each of the inner knots. Continuity alone implies one constraint at the knots. Differentiability at the knot gives two constraints. We chose the case with continuity without differentiability at the inner knots. With one constraints at each of the $(h-1)$ knots and 3 parameters for each slice, there are $3h - (h-1) = 2h + 1$ parameters to determine.

In the piecewise continuous cubic case, continuity and differentiability constraints at the knots yield quadratic and cubic splines. We relax these constraints to allow a continuity without differentiability at the inner knot. A total of $4h - (h-1) = 3h + 1$ parameters need to be estimated.

In all cases, the end-points $\tau_0, ..., \tau_h$ of the slices are obtained such that the slices contain approximately the same number of observations. Following are the expressions of the basis functions considered.

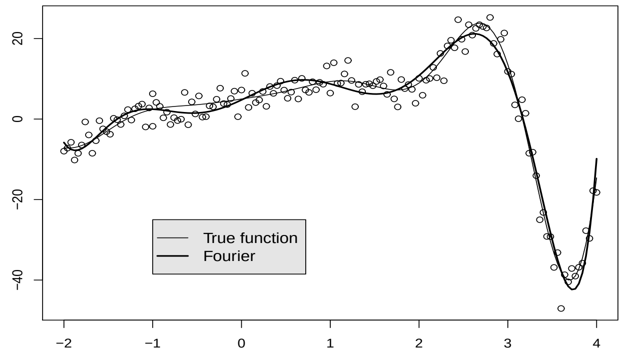

Figure 1: Fourier Basis

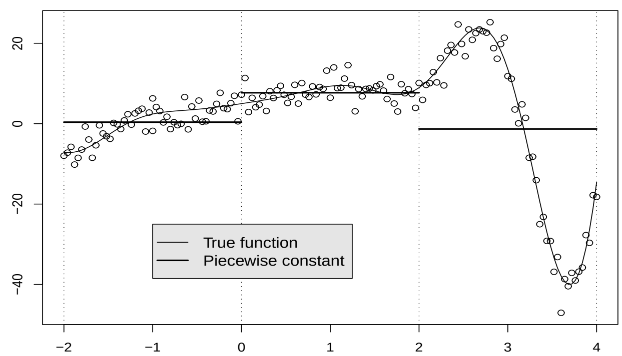

Figure 2: Piecewise Constant Basis

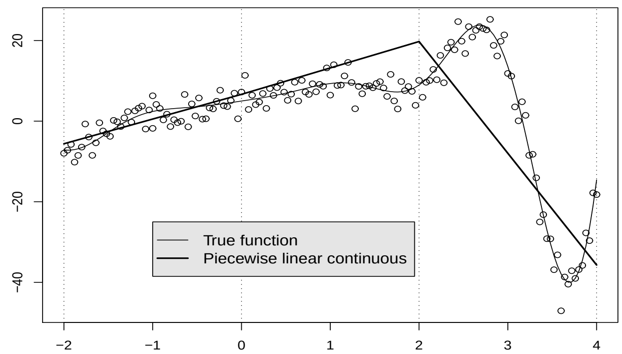

Figure 3: Piecewise Linear Continuous Basis

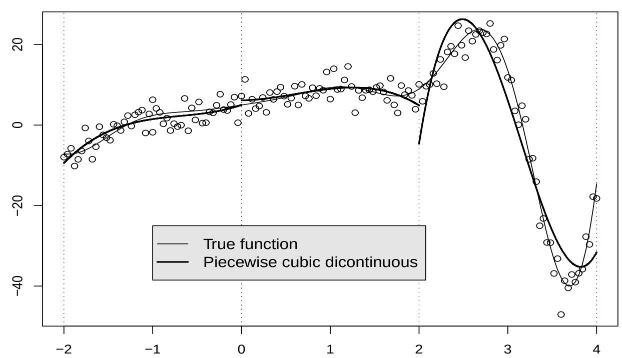

Figure 4: Piecewise Cubic Discontinuous Basis

1. **Fourier bases** are suggested by Cook (2007). They consist of a series of pairs of sines and cosines of increasing frequency. A Fourier basis is given by

$$\mathbf{f}_y = (\cos(2\pi y), \sin(2\pi y), ..., \cos(2\pi k y), \sin(2\pi k y))^T. \quad (2)$$

and $r = 2k$. Fourier bases can also be used within slices but this case is not explored here. Fourier bases are very popular in signal processing. They are mostly used for *periodic functions*. Figure 1 show how Fourier basis can be used to approximate a function.

2. **Piecewise Constant Basis.** This basis is suitable for a categorical response y taking values $1, 2, ..., h$ where h is the number of sub-populations or sub-groups. The k^{th} component f_{y_k} of $\mathbf{f}_y \in \mathbb{R}^{h-1}$ takes a constant value in the slice H_k with $f_{y_k} = J_k(y \in \mathrm{H}_k), k = 1, ..., h-1$. Figure 2 shows an example where a piecewise constant basis is used to approximate a continuous function.

3. **Piecewise Discontinuous Linear Basis.** It is more elaborate than the piecewise constant basis. Within each slice, we approximate the true function by a linear function. We have $\mathbf{f}_y \in \mathbb{R}^{2h-1}$ and its components are obtained as

$$\begin{aligned} f_{y_{(2i-1)}} &= J(y \in \mathrm{H}_i), \qquad i = 1, 2, ..., h-1 \\ f_{y_{2i}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1}), \qquad i = 1, 2, ..., h-1 \\ f_{y_{(2h-1)}} &= J(y \in \mathrm{H}_h)(y - \tau_{h-1}), \end{aligned} \quad (3)$$

4. **Piecewise Discontinuous Quadratic Basis.** In the pursuit of better approximation of the true trend in the data, this basis may help better than the

piecewise discontinuous linear basis. The components of the $\mathbf{f}_y \in \mathbb{R}^{3h-1}$ are

$$\begin{aligned}
f_{y_{(3i-2)}} &= J(y \in \mathrm{H}_i), \qquad i = 1, 2, ..., (h-1) \\
f_{y_{(3i-1)}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1}), \qquad i = 1, 2, ..., (h-1) \\
f_{y_{(3i)}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1})^2, \qquad i = 1, 2, ..., (h-1) \\
f_{y_{(3h-2)}} &= J(y \in \mathrm{H}_h)(y - \tau_{h-1}) \\
f_{y_{(3h-1)}} &= J(y \in \mathrm{H}_h)(y - \tau_{h-1})^2.
\end{aligned} \tag{4}$$

5. **Piecewise Discontinuous Cubic Basis.** Here we have $\mathbf{f}_y \in \mathbb{R}^{4h-1}$. Figure 4 shows the use of a piecewise cubic basis function to approximate a continuous function.

$$\begin{aligned}
f_{y_{(4i-3)}} &= J(y \in \mathrm{H}_i), \qquad i = 1, 2, ..., (h-1) \\
f_{y_{(4i-2)}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1}), \qquad i = 1, 2, ..., (h-1) \\
f_{y_{(4i-1)}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1})^2, \qquad i = 1, 2, ..., (h-1) \\
f_{y_{4i}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1})^3, \qquad i = 1, 2, ..., (h-1) \\
f_{y_{(4h-3)}} &= J(y \in \mathrm{H}_h)(y - \tau_{h-1}) \\
f_{y_{(4h-2)}} &= J(y \in \mathrm{H}_h)(y - \tau_{h-1})^2 \\
f_{y_{(4h-1)}} &= J(y \in \mathrm{H}_h)(y - \tau_{h-1})^3.
\end{aligned} \tag{5}$$

6. **Piecewise Continuous Linear Basis.** This is also called a *linear spline.* An example of the use of this basis function to approximate a function is shown

on Figure 3. The general form of the components f_{y_i} of $\mathbf{f}_y \in \mathbb{R}^{h+1}$ is

$$\begin{aligned} f_{y_1} &= J(y \in \mathrm{H}_1) \\ f_{y_{i+1}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1}) \qquad i = 1, ..., h. \end{aligned} \tag{6}$$

7. **Piecewise Continuous Quadratic Basis.** Adjacent curves are continuous without being differentiable at the inner knots. The components of $\mathbf{f}_y \in \mathbb{R}^{2h+1}$ are

$$\begin{aligned} f_{y_1} &= J(y \in \mathrm{H}_1) \\ f_{y_{2i}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1}), \qquad i = 1, ..., h. \\ f_{y_{2i+1}} &= J(y \in \mathrm{H}_i)(y - \tau_i)^2, \qquad i = 1, ..., h. \end{aligned} \tag{7}$$

8. **Piecewise Continuous Cubic Basis** In this case also, adjacent curves are continuous at the inner knots but are not differentiable. A piecewise cubic basis can be a good approximation to many functions. The components of $\mathbf{f}_y \in \mathbb{R}^{3h+1}$ are

$$\begin{aligned} f_{y_1} &= J(y \in \mathrm{H}_1) \\ f_{y_{3i-1}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1}), \\ f_{y_{3i}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1})^2, \\ f_{y_{3i+1}} &= J(y \in \mathrm{H}_i)(y - \tau_{i-1})^3, \qquad i = 1, ..., h. \end{aligned}$$

The choice of the basis can be aided by graphical exploration. The inverse response plots (Cook, 1998) of X_{yj} versus y, $j = 1, \ldots, p$, can give a hint about suitable choices for the basis. For example, when the plots show a linear relationship between the predictors and the outcome, then $\mathbf{f}_y = y$ can be used. When quadratic curvature is observed, then $\mathbf{f}_y = (y, y^2)^T$ can be considered. More elaborate basis functions could be useful when it is impractical to apply graphical methods to all of the predictors. It is also possible to develop an automatic mechanism to choose the basis. This can be done by numerically exploring a set of possible bases and choosing the best based on some criterion. For example, prediction performance might be used to select the basis.