Chapter 1
Penalized Spline Approaches for Functional Principal Component Logit Regression

A. Aguilera, M. C. Aguilera-Morillo, M. Escabias, M. Valderrama

Abstract The problem of multicollinearity associated with the estimation of a functional logit model can be solved by using as predictor variables a set of functional principal components. The functional parameter estimated by functional principal component logit regression is often unsmooth. To solve this problem we propose two penalized estimations of the functional logit model based on smoothing functional PCA using P-splines.

1.1 Introduction

The aim of the functional logit model is to predict a binary response variable from a functional predictor and also to interpret the relationship between the response and the predictor variables. To reduce the infinite dimension of the functional predictor and solve the multicollinearity problem associated to the estimation of the functional logit model, Escabias et al. (2004) proposed to use a reduced number of functional principal components (pc’s) as predictor variables. A functional PLS based solution was also proposed by Escabias et al. (2006). The problem associated with these approaches is that in many cases the estimated functional parameter is not smooth and therefore difficult to interpret. Different penalized likelihood estimations with B-spline basis were proposed in the general context of functional generalized linear

Ana Aguilera
Department of Statistics and O. R. University of Granada, Spain, e-mail: aaguiler@ugr.es

Maria del Carmen Aguilera-Morillo
Department of Statistics and O. R. University of Granada, Spain, e-mail: caguilera@ugr.es

Manuel Escabias
Department of Statistics and O. R. University of Granada, Spain, e-mail: escabias@ugr.es

Mariano Valderrama
Department of Statistics and O. R. University of Granada, Spain, e-mail: valderra@ugr.es
models to solve this problem (Marx and Eilers, 1999; Cardot and Sarda, 2005). In this paper we introduce two different penalized estimation approaches based on smoothed functional principal component analysis (FPCA). On one hand, FPCA of P-spline approximation of sample curves is performed. On the other hand, a discrete P-spline penalty is included in the own formulation of FPCA.

1.2 Background

Let us consider a sample of functional observations \( x_1(t), x_2(t), \ldots, x_n(t) \) of a fixed design functional variable and let \( y_1, y_2, \ldots, y_n \) be a random sample of a binary response variable \( Y \) associated to them. That is, \( y_i \in \{0, 1\}, i = 1, \ldots, n \). The functional logistic regression model is given by

\[
y_i = \pi_i + \varepsilon_i, \quad i = 1, \ldots, n,
\]

where \( \pi_i \) is the expectation of \( Y \) given \( x_i(t) \) modeled as

\[
\pi_i = P[Y = 1|\{x_i(t) : t \in T\}] = \frac{\exp\{\alpha + \int_T x_i(t) \beta(t) \, dt\}}{1 + \exp\{\alpha + \int_T x_i(t) \beta(t) \, dt\}}, \quad i = 1, \ldots, n,
\]

\( \alpha \) being a real parameter, \( \beta(t) \) a parameter function, \( \{\varepsilon_i : i = 1, \ldots, n\} \) independent errors with zero mean and \( T \) the support of the sample paths \( x_i(t) \).

The logit transformations can be expressed as

\[
l_i = \ln \left[ \frac{\pi_i}{1 - \pi_i} \right] = \alpha + \int_T x_i(t) \beta(t) \, dt, \quad i = 1, \ldots, n. \quad (1.1)
\]

A way to estimate the functional logit model is to consider that both, the sample curves and the parameter function, admit an expansion in terms of basis functions. Then, the functional logit model turns into a multiple logit model whose design matrix is the product between the matrix of basis coefficients of sample paths and the matrix of inner products between basis functions (Escabias et al., 2004). The estimation of this model is affected by multicollinearity due to the high correlation between the columns of the design matrix. In order to obtain a more accurate and smoother estimation of the functional parameter than the one provided by standard functional principal component logit regression (FPCLR), we present in this paper two penalized estimation approaches based on P-spline smoothing of functional PCA.