Chapter 2

Solving Multicollinearity in Functional Multinomial Logit Models for Nominal and Ordinal Responses

Ana Aguilera and Manuel Escabias

Abstract Different functional logit models to estimate a multicategory response variable from a functional predictor will be formulated in terms of different types of logit transformations as base-line category logits for nominal responses or cumulative, adjacent-categories or continuation-ratio logits for ordinal responses. Estimation procedures of functional logistic regression based on functional PCA of sample curves will be generalized to the case of a multicategory response. The true functional form of sample curves will be reconstructed in terms of basis expansions whose coefficients will be estimated from irregularly distributed discrete time observations.

2.1 Introduction

The functional logistic regression model is the most used method to explain a binary variable in terms of a functional predictor as can be seen in many applications in different fields. Ratcliffe *et al.* (2002) used this model for predicting if human foetal heart rate responses to repeated vibroacoustic stimulation. The relation between the risk of drought and time evolution of temperatures has been modeled in Escabias *et al.* (2005). With respect to the problem of estimation of this model, Escabias *et al.* (2004) proposed different functional principal component approaches for solving multicollinearity by providing an accurate parameter function estimation. An alternative estimation procedure based on PLS logit regression has been recently considered (Aguilera *et al.*, 2007).

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

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

In the general context of generalized functional linear models, James (2002) assumes that each predictor can be modeled as a smooth curve from a given functional family. Then, the functional model can be equivalently seen as a generalized linear model whose design matrix is given by the unobserved basis coefficients for the predictor and the EM algorithm is used for estimating the model from longitudinal observations at different times for each individual. On the other hand, Müller and StadtMüller (2005) considered an orthonormal representation of sample curves and used as predictor variables of the functional model a finite number of coefficients of such orthonormal expansion. Asymptotic tests and simultaneous confidence bands for the parameter function have been obtained by using this dimension reduction approach. An estimation procedure based on B-splines expansion maximizing the penalized log-likelihood has been studied in Marx and Eilers (1999) for a functional binomial response model and in Cardot and Sarda (2005) for the general case of functional generalized linear models.

The natural generalization of the functional logit model is the functional multinomial regression model where the response variable has a finite set of categories and the predictor is a functional variable. An initial work on this issue has been developed by Cardot *et al.* (2003) where a functional baseline-category logit model has been considered for predicting land use with the temporal evolution of coarse resolution remote sensing data. In this paper we propose a different approach comparing different methods of estimation based on the approximation of the functional predictor and the parameter functions in a finite space generated by a basis of functions what turns the functional model into a multiple one. Model estimation will be improved by developing several functional principal component approaches and selecting the predictor principal components according to their ability to provide the best possible estimation of the parameter functions.

2.2 Functional multinomial response model

Let us consider a functional predictor $\{X(t) : t \in T\}$, whose sample curves belong to the space $L^2(T)$ of square integrable functions on T, and a categorical response random variable Y with S categories.

Given a sample of observations of the functional predictor $\{x_i(t) : t \in T, i = 1, \ldots, n\}$, the sample of observations of the response associated to them is a set of n vectors $(y_{i1}, \ldots, y_{iS})'$ of dimension S defined by

$$y_{is} = \begin{cases} 1 & \text{if category } s \text{ is observed for } X(t) = x_i(t) \\ 0 & \text{other case} \end{cases}$$

so that each observation is generated by a multinomial distribution $M(1; \pi_{i1}, \ldots, \pi_{iS})$ with $\pi_{is} = P[Y = s|X(t) = x_i(t)]$ and $\sum_{s=1}^{S} \pi_{is} = 1 \quad \forall i = 1, \ldots, n$.

Let us observe that y_{iS} is redundant. Then, if we denote by $y_i = (y_{i1}, \ldots, y_{i,S-1})'$ the vector response for subject i, with mean vector $\mu_i = E[Y_i] = (\pi_{i1}, \ldots, \pi_{i,S-1})'$, the multinomial response model is a particular case of generalized linear model $y_{is} = \pi_{is} + \varepsilon_{is}$ with

$$g_s(\mu_i) = \alpha_s + \int_T \beta_s(t) x_i(t) dt, \quad s = 1, \ldots, S-1, \tag{2.1}$$

where the link function components g_s can be defined in different ways, ε_{is} are independent and centered errors and α_s and $\beta_s(t)$ a set of parameters to be estimated. In this paper we are going to generalize the functional logit model for a binary response to the case of a multinomial response. Because of this we will consider as link functions different types of logit transformations $l_{is} = g_s(\mu_i)$ (see Agresti (2002) for a detailed explanation).

2.2.1 Nominal responses

Baseline-category logits for nominal response pair each response with a baseline category

$$l_{is} = \log\left[\pi_{is}/\pi_{iS}\right].$$

Then, the equation that expresses baseline-category logit models directly in terms of response probabilities is $(\alpha_S = 0, \beta_S(t) = 0)$

$$\pi_{is} = \frac{\exp\left\{\alpha_s + \int_T x_i\left(t\right)\beta_s\left(t\right)dt\right\}}{\sum_{s=1}^{S}\exp\left\{\alpha_s + \int_T x_i\left(t\right)\beta_s\left(t\right)dt\right\}}, \; s = 1, \ldots, S, \; i = 1, \ldots, n. \tag{2.2}$$

2.2.2 Ordinal responses

When the response variable is ordinal the logit transformations l_{is} reflect ordinal characteristics such as monotone trend. Next, several types of ordinal logits will be studied.

Cumulative logits

Cumulative logits use category ordering by forming logits of cumulative probabilities. The most popular logit model for ordinal responses is the proportional odds model

$$l_{is} = \log \frac{P[Y \leq s|x_i(t)]}{1 - P[Y \leq s|x_i(t)]} = \frac{\sum_{j=1}^{s} \pi_{ij}}{\sum_{j=s+1}^{S} \pi_{ij}} = \alpha_s + \int_T \beta(t) x_i(t) dt,$$

$s = 1, \ldots, S-1$, that has the same effects $\beta(t) = \beta_s(t) \quad \forall s = 1, \ldots, S-1$ for each cumulative logit. Then, each response probability is obtained as $\pi_{is} = F_{is} - F_{i,s-1}$ with

$$F_{is} = P[Y \leq s|x_i(t)] = \frac{\exp\left(\alpha_s + \int_T \beta(t) x_i(t) dt\right)}{1 + \exp\left(\alpha_s + \int_T \beta(t) x_i(t) dt\right)}.$$

Adjacent-categories logits

Logits for ordinal responses do not need use cumulative probabilities. Alternative logits for ordinal responses are the adjacent-categories logits and the continuation-ratio logits.

Adjacent-categories logits are defined as $l_{is} = \log\left[\pi_{is}/\pi_{i,s+1}\right]$, $s = 1, \ldots, S-1$. Taking into account the relation between baseline-category logits and adjacent-categories, the adjacent-categories logit model with common effect $\beta(t)$ (equal odds model)

$$\log\left[\frac{\pi_{is}}{\pi_{i,s+1}}\right] = \alpha_s + \int_T \beta(t) x_i(t) dt,$$

can be expressed in terms of the response probabilities as

$$\pi_{is} = \frac{\exp\left[\sum_{j=s}^{S-1} \alpha_j + \int_T (S-s)\beta(t) x_i(t) dt\right]}{1 + \sum_{s=1}^{S-1} \exp\left[\sum_{j=s}^{S-1} \alpha_j + \int_T (S-s)\beta(t) x_i(t) dt\right]}.$$

Continuation-ratio logits

Continuation-ratio logits are $l_{is} = \log\left[P(Y = s)/P(Y > s)\right] = \log\left[\pi_{is}/\sum_{j=s+1}^{S} \pi_{ij}\right]$. Denoting by $p_{is} = \frac{\pi_{is}}{\pi_{is} + \cdots + \pi_{iS}}$ the probability of response s, given response s or higher, the continuation-ratio logit models can be seen as ordinary binary logit models

$$l_{is} = \log \frac{\pi_{is}}{\sum_{j=s+1}^{S} \pi_{ij}} = \log \frac{p_{is}}{1 - p_{is}} = \alpha_s + \int_T \beta_s(t) x_i(t) dt,$$

so that these conditional probabilities are modeled as

$$p_{is} = \frac{\pi_{is}}{\pi_{is} + \cdots + \pi_{iS}} = \frac{\exp\left(\alpha_s + \int_T \beta_s(t) x_i(t) dt\right)}{1 + \exp\left(\alpha_s + \int_T \beta_s(t) x_i(t) dt\right)}.$$

2.3 Model estimation

As with any other functional regression model, estimation of the parameters of a functional multinomial response model is an ill-posed problem due to the infinite dimension of the predictor space. See Ramsay and Silverman (2005) for a discussion on the functional linear model. In addition, the functional predictor is not observed continuously in time so that sample curves $x_i(t)$ are observed in a set of discrete time points $\{t_{ik} : k = 1, \ldots, m_i\}$ that could be different for each sample individual. The most used solution to this problems is to reduce dimension by performing a basis expansion of the functional predictor.

A first estimation of the parameter functions of a functional multicategory logit model can be obtained by considering that both the predictor curves as parameter functions belong a finite space generated by a basis of functions $x_i(t) = a_i' \Phi(t)$, $\beta_s(t) = \beta_s' \Phi(t)$, with $\Phi(t) = (\phi_1(t), \ldots, \phi_p(t))'$ a vector of basic functions that generate the space where $x(t)$ belong to, and $a_i = (a_{i1}, \ldots, a_{ip})'$ and $\beta_s = (\beta_{s1}, \ldots, \beta_{sp})'$ the vectors of basis coefficients of sample curves and parameter functions, respectively. The sample curves basis coefficients will be computed in a first step by using different approximation methods as interpolation (data observed without error) or least squares smoothing (noisy data).

Then, the functional model turns to a multiple one given by

$$l_{is} = \alpha_s + \int_T x_i(t) \beta_s(t) \, dt = \alpha_s + a_i' \Psi \beta_s \quad s = 1, \ldots, S-1, \quad i = 1, \ldots, n,$$

with $\Psi = (\psi_{uv})$ being the $p \times p$ matrix of inner products $\psi_{uv} = \int_T \phi_u(t) \phi_v(t) \, dt$.

In matrix form each vector of logit transformations $L_s = (l_{1s}, \ldots, l_{ns})'$ can be expressed as $L_s = \alpha_s \mathbf{1} + A\Psi\beta_s$, $s = 1, \ldots, S-1$.

The estimation of this model will be carried out by maximizing the associated multinomial log likelihood, under each of the four different multicategory logits considered in previous section. In the case of baseline-category logits the log likelihood is concave, and the Newton-Raphson method yields the ML parameter estimates. For cumulative logits a Fisher scoring algorithm is used for iterative calculation of ML estimates. The adjacent-categories logit model is fitted by using the same methods for its equivalent baseline-category logit model. In the case of continuation-ratio logit models the simultaneous ML estimation of its parameters can be reduced to separate fitting of model for each different continuation-ratio logit by using ML estimation for binary logit models.

2.4 Principal components approach

The ML estimation of the functional multinomial regression model obtained by the approach in the previous section is affected by high multicollinearity what makes the variances of estimated parameter function increase in an artificial way. This has been proven for the logit model of binary response (Aguilera *et al.*, 2005) and for its functional version (Escabias *et al.*, 2004) through different simulated and real data sets. In this paper this problem will be solved by using as covariates of the multiple multinomial regression model a set of functional principal components of the functional predictor. An alternative way of avoiding excessive local fluctuation in the estimated parameter function would be to use a roughness penalty approach based on maximizing a penalized likelihood function (see Marx and Eilers (1999) for the functional regression model with binary response).

Two different FPCA of the sample curves will be considered after approximating such curves in a finite dimension space generated by a basis of functions. First, we will compute FPCA of the sample paths with respect to the usual inner product in $L^2(T)$ that is equivalent to PCA of the data matrix $A\Psi^{1/2}$ with respect to the usual inner product in $\mathbb{R}^p$. And second, we will perform PCA of the design matrix $A\Psi$ with respect to the usual inner product in $\mathbb{R}^p$ that is equivalent to FPCA of certain transformation of sample curves $x_i(t)$. The results set out in Ocaña *et al.* (2007) allow to demonstrate these equivalences between functional and multivariate PCA. Let us observe that both FPCA match when the basis is orthonormal.

Let Γ be a matrix of functional principal components associated to $x(t)$ so that $\Gamma = A\Psi V$ with $VV' = I$. Then, the multinomial logit model can be equivalently expressed in terms of all principal components as $L_s = \alpha_s\mathbf{1} + A\Psi\beta_s = \alpha_s\mathbf{1} + \Gamma\gamma_s$, and we can give an ML estimation of the parameters of the functional model (coordinates of $\beta_s(t)$) through the estimation of this one, $\widehat{\beta}_s = V\widehat{\gamma}_s$.

Then, we propose to approximate these parameters functions by using a reduced set of principal components. There are different criteria in literature to select principal components in regression methods. Escabias *et al.* (2004) compared in the functional binary logit model the classical one that consist of including principal components in the model in the order given by explained variability with the one of including them in the order given by a stepwise method based on conditional likelihood ratio test. In this work we will compare these two methods for different functional nominal and ordinal logit models. The optimum number of principal components (model order) will be determines by using different criteria based on minimization the leave-one-out prediction error or the leave-one-out misclassification rate via cross-validation.

The model will be tested by different simulated examples and applications with real data. It will be shown that the best parameter function estimation is given by the model that minimizes the mean of the integrated mean squared error of the parameter functions estimates. The relation of this minimum with special trends in other goodness of fit measures will be also investigated. An adequate model selection method based on the results will be proposed.

Acknowledgements This research has been funded by project P06-FQM-01470 from *"Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía, Spain"* and project MTM2007-63793 from *Dirección General de Investigación, Ministerio de Educación y Ciencia, Spain.*

References

[1] Agresti, A.: Categorical Data Analysis. Second edition, Wiley: New York. (2002).
[2] Aguilera, A. M. Escabias, M. and Valderrama, M. J.: Using principal components for estimating logistic regression with high-dimensional multicollinear data. Computational Statistics and Data Analysis. **50**(8), 1905-1924 (2006).
[3] Cardot, H., R. Faivre and Goulard, M.: Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. Journal of Applied Statistics. **30**(10), 1185-1199 (2003).
[4] Cardot, H. and Sarda, P.: Estimation in generalized linear models for functional data via penalized likelihood. Journal of Multivariate Analysis. **92**, 24-41 (2005).
[5] Escabias, M., Aguilera, A. M. and Valderrama, M. J.: Principal component estimation of functional logistic regression: discussion of two different approaches. Journal of Nonparametric Statistics. **16**(3-4), 365-384 (2004).
[6] Escabias, M., Aguilera, A. M. and Valderrama, M. J.: Modelling environmental data by functional principal component logistic regression. Environmetrics. **16** (1), 95-107 (2005).
[7] Escabias, M. Aguilera, A. M. and Valderrama, M. J.: Functional PLS logit regression model. Computational Statistics and Data Analysis. **51**(10), 4891-4902 (2007).
[8] James, G. M.: Generalized linear models with functional predictors. Journal of the Royal Statistical Society, Series B. **64**(3), 411-432 (2002).
[9] Marx, B.D. and P.H.C. Eilers.: Generalized linear regression on sampled signals and curves. A p-spline approach. Technometrics. **41**, 1-13 (1999).
[10] Muller, H-G. and StadtMüller, U.: Generalized functional linear models. The Annals of Statistics. **33**(2), 774-805 (2005).
[11] Ocaña, F.A., Aguilera, A.M. and Escabias, M.: Computational considerations in functional principal component analysis. Computational Statistics. **22**(3), 449-466 (2007).
[12] Ramsay, J. O. and Silverman, B. W.: Functional Data Analysis. Second edition, Springer-Verlag: New York. (2005).
[13] Ratcliffe, S. J., Leader, L. R. and Heller, G. Z.: Functional data analysis with application to periodically stimulated foetal heart rate data. II: functional logistic regression. Statistics in medicine. **21**(8), 1115-1127 (2002).

http://www.springer.com/978-3-7908-2061-4

Functional and Operatorial Statistics
Dabo-Niang, S.; Ferraty, F. (Eds.)
2008, XX, 304 p., Hardcover
ISBN: 978-3-7908-2061-4
A product of Physica-Verlag Heidelberg