International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2021; 6(4): 30-34

www.mathsjournal.com
Received: 11-05-2021
Accepted: 15-06-2021

Amit Kumar Singh
Research Scholar, Department of Mathematics, J.P. Univ. Chapra, Bihar, India

Dr. Ashok Kumar
Associate Professor, Department of Mathematics, DAV PG College, Siwan, Bihar, India

Corresponding Author:
Amit Kumar Singh
Research Scholar, Department of Mathematics, J.P. Univ. Chapra, Bihar, India

Analysis of generalized linear models for functional data

Amit Kumar Singh and Dr. Ashok Kumar

Abstract
In this Paper, we analyze in a regression setting the link between a scalar response and a functional predictor by means of a Functional Generalized Linear Model. We first give a theoretical framework and then discuss identifiability of the model. The functional coefficient of the model is estimated via penalized likelihood with spline approximation.

Keywords: functional generalized, linear model, functional predictor, theoretical framework, discuss identifiability

Introduction

There has been existing for a long time a ''functional'' approach for which models aim at taking into account the functional nature of the data: see for instance the work from [1–2] on Data Analysis in the context of Hilbert spaces theory. Until recently, this approach has been certainly less used than the discrete one in practical studies. The monographs from [3–4] which investigate not only the above regression setting but also a variety of other statistical problems with functional data is an important step for the popularization of these methods. Moreover, an increasing amount of recent papers investigate (functional) models for functional data.

The functional generalized linear model

We adopt in the following form,

$$\exp\{b_1(\eta)y+b_2(\eta)\}v(dy); \qquad (1)$$

Where v is a nonzero measure on R which is not concentrated at a single point and where the function b_1 is twice continuously differentiable and b'_1 is strictly positive on R: Then, the function b_1 is strictly increasing and b_2 is twice continuously differentiable on R: The mean μ of the distribution is

$$\mu = b_3 = -\frac{b'_2(\eta)}{b'_1(\eta)}$$

Where b_3 is continuously differentiable and b'_3 is strictly positive on R: The function b_3^{-1} 3 is called the link function and one has $\eta = b_3^{-1}(\mu)$:
It is also assumed as in Stone's paper that there is an interval S in R such that v is concentrated on S and

$$(H.1)\ b_1^n(\eta)y + b_2^n(\eta) < 0, \quad \forall \eta \in R, \forall y \in S.$$

The reader is referred to [5] for examples of exponential families, such as the Bernouilli or the gamma distribution, satisfying condition (H.1). Let X and Y be two random variables defined on the same probability space with X valued in the separable Hilbert space $H = L^2_{[0,1]}$ and Y valued in R: Let ⟨ϕ, ψ⟩ denote the usual inner product of functions f and c in H; defined by

$\langle\phi, \psi\rangle = \int_0^1 \phi(t)\psi(t)\,dt$ and let $\|\phi\|$ denote the norm associated with this inner product.

We assume that the following functional generalized linear model holds, that is to say we assume the existence of a function $\alpha \in H$ such that

$$E(Y \mid X = x) = b_3(\langle \alpha, x \rangle),\ x \in H \quad . \tag{2}$$

The conditional distribution of Y given X = x is supposed to belong to the exponential family (1) or at least to satisfy Conditions 2–4 of Stone [5]. Without loss of generality we assume that the functional random variable X is centered i.e. EX(t) = 0, for t a.e. We also suppose that X is of second order i.e. $E\|X\|^2 < \infty$: Thus, the covariance operator Γ of the H-valued random variable X is defined as

$$\Gamma x(t) = \int_0^1 E[X(t)X(s)]x(s)ds, \quad x \in H,\ t \in [0,1].$$

The operator Γ is an integral operator whose kernel is the covariance function of X and it is nuclear, self-adjoint and nonnegative [1, 6]. Moreover, the operator Γ is assumed to satisfy the condition

(H:2) The eigenvalues of Γ are nonzero.

Condition (H.2) ensures the identifiability of the model (see below) and for instance is assumed in other settings such as the one described in [7]. It is fulfilled when X(t) is a standard Brownian motion [8]. Let us notice that it can be relaxed when there exist some null eigenvalues by changing the Hilbert space of reference, taking H as the closure of the range of Γ[9]. Let us also remark that if we had supposed that there were only a finite number of non-null eigenvalues then we would be in a classical parametric framework since it would mean that we would have only a finite number of covariates. To get identifiability, let us denote by λ_j, j = 1, 2,... the eigenvalues of Γ and by v_j, j = 1, 2,....a complete orthonormal sequence of Eigen functions and let α_1 and α_2 be two functions in H such that

$b_3(\langle a_1, X\rangle) = b_3(\langle a_2, X\rangle)$.

Since b_3 is strictly increasing one has

$\langle \alpha_1 - \alpha_2, X\rangle = 0$,

and then

$$\begin{aligned} E\langle \alpha_1 - \alpha_2, X\rangle^2 &= \langle \Gamma(\alpha_1 - \alpha_2), \alpha_1 - \alpha_2\rangle \\ &= \sum_{j=1}^{+\infty} \lambda_j \langle \alpha_1 - \alpha_2, v_j\rangle^2 \\ &= 0. \end{aligned}$$

Now, since $\lambda_j \neq 0, \forall_j$, one has

$\langle \alpha_1 - \alpha_2, v_j\rangle^2 = 0, \forall_j$,

and then $\alpha_1 = \alpha_2$ almost everywhere in H.

The expected log-likelihood is defined as

$\Lambda(a) = E(b_1(\langle a, X\rangle)b_3(\langle \alpha, X\rangle) + b_2(\langle a, X\rangle)),\ \alpha \in H$,

Hypothesis (H.1) gives directly

$$b_1^{n}(\eta)b_3(\eta_0) + b_2^{"}(\eta) < 0,\ \forall \eta, \eta_0 \in R, \tag{3}$$

Which implies that the function $\Psi(\eta) = b_1(\eta)b_3(\eta_0) + b_2(\eta)$ is strictly concave and has a unique maximum at η_0. Then, when model (2) holds, the function α is a maximum of Λ which is essentially uniquely determined under (H.2).

Estimation of the functional coefficient

We introduce an estimator of α based on a B-splines expansion maximizing the penalized log-likelihood. First of all, let us describe the space of spline functions defined on [0, 1] with equispaced knots. Suppose that q and k are integers and let S_{qk} be the space of spline functions defined on [0, 1], with degree q and k–1 equispaced interior knots. The set S_{qk} is the set of functions s satisfying:

- s is a polynomial of degree q on each interval [(t–1)/k, t/k], t = 1,....,k;
- s is q–1 times continuously differentiable on [0, 1].

The set S_{qk} is known to be a linear space with dimension q+k and one can derive a basis by means of normalized B-splines $\{B_{k,j}, j=1,\dots,k+q\}$ [10]. In the following we denote as B_k the vector of all the B-splines and as $B_k^{(m)}$ the vector of derivatives of order m of all the B-splines for some integer m (m<q).
Our penalized B-splines estimator of a is thus defined as

$$\hat{\alpha}_{PS} = \sum_{j=1}^{q+k} \hat{\theta}_j B_{kj}$$
$$= B_k^{'}\hat{\theta},$$

Where $\hat{\theta}$ is a solution of the following maximization problem

$$\max_{\theta \in R^{q+k}} \frac{1}{n}\sum_{i=1}^{n}\left(b_i\left(\left\langle B_k^{'}\theta, X_i\right\rangle\right)Y_i + b_2\left(\left\langle B_k^{'}\theta, X_i\right\rangle\right)\right) - \frac{1}{2}\rho \| B_k^{(m)'}\theta \|^2, \tag{4}$$

With smoothing parameter ρ>0. The estimator $\hat{\alpha}_{PS}$ is of the same type as the one introduced by Marx and Eilers [11], with however a different roughness penalty. Indeed, our penalty, borrowed from [12], allows to obtain a given level of smoothness in the smooth representation following ideas from [3]. This penalty can also be modified in order to give local measures of roughness [13]. Computation of the estimator is achieved by means of a slight modification of the scoring algorithm for generalized linear models [14]. Marx and Eilers [15] also give formulas for computing a generalized cross validation criterion that allows to choose reasonable values for ρ.

We study now the performance of estimator $\hat{\alpha}_{PS}$ in terms of the asymptotic behavior of the L^2 norm in H with respect to the distribution of X defined as

$$\| \phi \|_2^2 = \langle \Gamma\phi, \phi \rangle,\ \phi \in H.$$

Note that since for each ϕ in H, there exists a unique element Φ in the space H' of continuous linear operator from H to R such that Φ(X) = ⟨ϕ, X⟩ the corresponding norm in H' is

$$\| \Phi \|_2^2 = E\Phi^2(X),\ \Phi \in H'.$$

To derive L^2 convergence rates for $\hat{\alpha}_{PS}$ we assume moreover the following conditions:

(H.3) $\|X\| \leqslant C_1 < \infty$, a.s.

The function a is supposed to have p' derivatives for some integer p' with $\alpha^{(p')}$ satisfying

(H.4) $|\alpha^{(p')}(y_i) - \alpha^{(p')}(y_2)| \leqslant C_2|y_1 - y_2|^v$, $C_2>0$, $v \in [0, 1]$.

In the following, we note p = p' + v and assume that the degree q of the splines is such that q⩾p.

Theorem 1. Let $\rho \sim n^{(\delta-1)2}$, for some $0<\delta<1$ and suppose that $\rho^{-1}k^{-2p} + \rho k^{2(m-p)} = O(1)$ and $\rho^2 k^{2m} = o(1)$. Under hypothesis (H.1)–(H.4), we have

1. A unique solution to the maximization problem (4) exists except on an event whose probability tends to zero as n→∞.
2. $\| \hat{\alpha}_{PS} - \alpha \|_2^2 = O_P\left(\frac{k}{n}\right) + O(k^{-2p}) + O(\rho k^{2(m-p)}) + O(\rho).$

Corollary 1. Under the assumptions of Theorem 1 and for $k \sim n^{1/(2p-1)}$ and $\rho \sim n^{(\delta-1)/2}$ we get for m⩽p the L^2 rate of convergence

$$\| \hat{\alpha}_{PS} - \alpha \|_2^2 = O_P\left(n^{-2p/(2p+1)}\right) + O(\rho). \tag{5}$$

The particular case of the linear model

For the linear model, we have an explicit expression for our estimator and it can be shown that better bound for the bias occurs if we suppose moreover that the ''projection'' of a onto the space S_{qk} belongs to the range of the covariance operator Γ. It allows us to get Stone's ''optimal'' rates of convergence. Maximizing the expected log-likelihood is equivalent to minimizing the following criterion

$$\min_{\beta\in S_{qk}} E\left(\langle \alpha-\beta, X\rangle^2\right), \quad (6)$$

That is to say $\min_{\beta\in S_{qk}} \|\alpha-\beta\|_2^2$. Let us denote by $\tilde{\alpha}$ the minimizer of (6). Consider to simplify the ridge regression approximation (i.e. m = 0) $\tilde{\alpha}_{PS}$ defined as

$$Arg \min_{\beta\in S_{pk}} E\left(\langle \alpha-\beta, X\rangle^2\right)+\frac{1}{2}\rho\|\beta\|^2. \quad (7)$$

Since S_{qk} is a finite dimensional function space and the eigenvalues of Γ are supposed to be strictly positive, it is easy to show that $\tilde{\alpha}$ and $\tilde{\alpha}_{PS}$ are uniquely determined. Moreover, they satisfy respectively the functional normal equations

$$\Gamma\tilde{\alpha}(t) = E(YX(t)),\ t\in[0,1], \quad (8)$$

and

$$\Gamma\tilde{\alpha}_{PS}(t)+\rho\tilde{\alpha}_{PS}(t) = E(YX(t)),\ t\in[0,1] \quad (9)$$

Combining equalities (8), (9) and expanding $\tilde{\alpha}$ and $\tilde{\alpha}_{PS}$ in the basis of the orthonormal eigenfunctions of Γ, we get

$$\langle v_j, \tilde{\alpha}_{PS}\rangle = \frac{\lambda_j}{\lambda_j+\upsilon}\langle v_j,\tilde{\alpha}\rangle$$

and

$$\|\tilde{\alpha}-\tilde{\alpha}_{PS}\|_2^2 = \sum_{j=1}^{\infty}\lambda_j\frac{\rho^2}{(\lambda_j+\rho)^2}\langle v_j,\tilde{\alpha}\rangle^2 \quad (10)$$

Suppose now that $\tilde{\alpha}$ belongs to the range of Γ, that is to say there exists a function $g_{\tilde{\alpha}}\in H$ such that $\Gamma g_{\tilde{\alpha}} = \tilde{\alpha}$. We have that

$$\sum_{j=1}^{\infty}\frac{\langle v_j,\tilde{\alpha}\rangle}{\lambda_j^2} < +\infty \quad (11)$$

and thus with (10), we can bound

$$\|\tilde{\alpha}-\tilde{\alpha}_{PS}\|_2^2 = O(\rho^2)\,. \quad (15)$$

Conclusion

First of all, let us note that very few asymptotic results have been proved in the context of linear models for functional variables, except in the setting of (Hilbertian) autoregressive linear processes and for the functional linear regression model. It appears in these papers that the (upper bounds for the) rates of convergence for estimators based on Functional Principal Components are quite poor comparatively to the ones obtained for spline estimators (even with stronger assumptions for the first case). This should be linked with results obtained by means of simulation studies where spline estimators seems to be superior.

References

1. Dauxois J, Pousse A. Les analyses factorielles en calcul des probabilite´s et en statistique: Essai d'etude synthe´tique, The` se, Universite´ Paul sabatier, Toulouse, France.
2. Deville JC. Me´thodes statistiques et nume´riques de l'analyse harmonique, Ann. Insee 2014, 15.
3. Ramsay JO, Silverman BW. Functional Data Analysis, Springer, Berlin 2017.
4. Ramsay JO, Silverman BW, Applied Functional Data Analysis: Methods and Case Studies, Springer, Berlin 2002.
5. Stone CJ. The dimensionality reduction principle for generalized additive models, Ann. Statist 2016;14:590-606.

6. Dauxois J, Pousse A, Romain Y. Asymptotic theory for the principal component analysis of a random vector function: some applications to statistical inference, J Multivariate Anal 2019;12:136-154.
7. Bosq D. Linear processes in function spaces, in: Lecture Notes in Statistics, Springer, Berlin 2000, 149.
8. Cardot H, Ferraty F, Sarda P. Functional linear model, Statist. Probab. Lett 2019;45:11-22.
9. Cardot H, Ferraty F, Mas A, Sarda P. Testing hypotheses in the functional linear model, Scand. J Statist 2003b;30:241-255.
10. De Boor C. A Practical Guide to Splines, Springer, New York 2018.
11. Marx BD, Eilers PH. Generalized linear regression on sampled signals with penalized likelihood, in: A. Forcina, G.M. Marchetti, R. Hatzinger, G. Galmacci (Eds.), Statistical Modelling, Proceedings of the 11th International Workshopon Statistical Modelling, Orvietto 2016.
12. O'Sullivan F. A statistical perspective on ill-posed inverse problems (with discussions), Statist. Sci 2018;4:502-527.
13. Cardot H. Spatially adaptive splines for statistical linear inverse problems, J. Multivariate Anal 2002;81:100-119.
14. McCullagh P, Nelder JA, Generalized Linear Models, 2nd Edition, Chapman & Hall, London 2019.
15. Marx BD, Eilers PH. Generalized linear regression on sampled signals and curves: a P-Spline approach, Technometrics 1999;41:1-13.