Communications of the Korean Statistical Society Vol. 15, No. 6, 2008, pp. 1003–1011

Semiparametric Kernel Poisson Regression for Longitudinal Count Data†

Changha Hwang[1], Jooyong Shim[2]

Abstract

Mixed-effect Poisson regression models are widely used for analysis of correlated count data such as those found in longitudinal studies. In this paper, we consider kernel extensions with semiparametric fixed effects and parametric random effects. The estimation is through the penalized likelihood method based on kernel trick and our focus is on the efficient computation and the effective hyperparameter selection. For the selection of hyperparameters, cross-validation techniques are employed. Examples illustrating usage and features of the proposed method are provided.

Keywords: Longitudinal data; fixed-effect; random-effect; Poisson regression; canonical parameter; kernel trick; penalized likelihood; cross-validation function.

1. Introduction

Count data are increasingly common in fields such as medicine, biology, criminology, political sciences and marketing. The ordinary least squares method for count data results in biased, inefficient and inconsistent estimates (Long, 1997). The Poisson regression model provides an attractive solution for the analysis of count data if observations are independent, *i.e.*, not longitudinal or clustered (Winkelmann, 2003). However, it is often the case that subjects are observed nested within clusters or are repeatedly measured. In this case, the ordinary Poisson regression model assuming independence of observations causes problems since observations from the same cluster or subject are usually correlated.

For data that are clustered and/or longitudinal, mixed-effect regression models are becoming increasingly popular - Hedeker and Gibbons (2006), Long (1997), McCullagh and Nelder (1983), Wu and Zhang (2006). Mixed-effect models constitute both fixed and random effects. In clustered data, subjects are clustered within an organization such as a hospital, school, clinic or firm. In longitudinal data where individuals are repeatedly assessed, measurements are clustered within individuals. For clustered data the random

† This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund)(KRF-2007-313-C00130).

1) Professor, Department of Statistics, Dankook University, Gyeonggido 448-160, Korea.

2) Adjunct professor, Department of Applied Statistics, Catholic University of Daegu, Gyungbuk 712-702, Korea. Correspondence: ds1631@hanmail.net

effects represent cluster effects, while for longitudinal data the random effects represent subject effects. There has been much work done on mixed-effect models for continuous responses and an increasing amount of work has focused on mixed-effects models for non-continuous response data (Hedeker and Gibbons, 2006).

For count data, various types of mixed-effect Poisson regression models have been proposed by Gu and Ma (2005), Long (1997), Wu and Zhang (2006). In this paper we propose a semiparametric mixed-effect kernel Poisson regression model for the analysis of longitudinal count data. The proposed model is derived by employing the penalized likelihood method based on kernel tricks in Vapnik (1995). For the easy selection of appropriate hyperparameters to achieve high generalization performance, we propose cross-validation techniques, which use a quadratic loss function instead of the idea of exponential family unlike Yuan (2005) and Shim *et al.* (2007). The rest of this paper is organized as follows. In Section 2 we describe kernel Poisson regression, which is based on the penalized negative log-likelihood. In Section 3 we consider mixed-effect kernel model with semiparametric fixed effects and parametric random effects. In Section 4 we propose GCV function for the model selection. Section 5 presents simulation study and one real data example to illustrate our method. In Section 6 we give the conclusion.

2. Kernel Poisson Regression

In Poisson regression it is assumed that the response variable $y_i \in \{0, 1, 2, \ldots\}$, number of occurrences of an event, has a Poisson distribution given the input vector $\boldsymbol{x}_i \in R^d$,

$$p(y_i) = \frac{e^{-\mu(\boldsymbol{x}_i)}\mu(\boldsymbol{x}_i)^{y_i}}{y_i!}, \quad i = 1, 2, \ldots, n. \tag{2.1}$$

The negative log-likelihood of the given data set can be expressed as(a constant term is omitted)

$$\ell(\mu) = \frac{1}{n}\sum_{i=1}^{n}\{\mu(\boldsymbol{x}_i) - y_i \log \mu(\boldsymbol{x}_i)\}. \tag{2.2}$$

We write the canonical parameter(logarithm of $\mu(\boldsymbol{x}_i)$) as $\eta(\boldsymbol{x}_i)$, then the negative log-likelihood can reexpressed as

$$\ell(\eta) = \frac{1}{n}\sum_{i=1}^{n}\left\{e^{\eta(\boldsymbol{x}_i)} - y_i\eta(\boldsymbol{x}_i)\right\}. \tag{2.3}$$

A nonparametric estimate of the canonical parameter of a Poisson process based on penalized likelihood smoothing spline models was recently studied by Yuan (2005). We now consider a nonlinear "kernelized" variant of Poisson regression model. The canonical parameter given $\boldsymbol{x}_i$ is estimated by a linear model, $\eta(\boldsymbol{x}_i) = b_0 + \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_i)$, conducted in a high dimensional feature space. Here the superscript ^{T} denotes vector or matrix transpose. The feature mapping function $\boldsymbol{\phi}(\cdot) : R^d \to R^{d_f}$ maps the input space to the higher dimensional feature space where the dimension is defined in an implicit way. It suffices to know and use $K(\boldsymbol{x}_k, \boldsymbol{x}_l) = \boldsymbol{\phi}(\boldsymbol{x}_k)^T\boldsymbol{\phi}(\boldsymbol{x}_l)$ instead of defining $\boldsymbol{\phi}(\cdot)$ explicitly. Note that the identity map $\boldsymbol{\phi}$ leads nonlinear model to linear model. We focus on the use of a Gaussian kernel $K(\boldsymbol{x}_k, \boldsymbol{x}_l) = \exp(-\|\boldsymbol{x}_k - \boldsymbol{x}_l\|^2/\sigma^2)$ in the sequel.

Then the estimate of canonical parameter η is obtained by minimizing the penalized negative log-likelihood,

$$\ell(\boldsymbol{w}) = \sum_{i=1}^{n} \left[e^{b_0+\boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_i)} - y_i \left\{ b_0 + \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_i) \right\} \right] + \frac{\lambda}{2}\|\boldsymbol{w}\|^2, \tag{2.4}$$

where λ is a penalty parameter which controls the trade-off between the goodness-of-fit on the data and the smoothness of η. The representation theorem (Kimeldorf and Wahba, 1971) guarantees the minimizer of the penalized negative log-likelihood to be $\eta(\boldsymbol{x}_i) = b_0 + \boldsymbol{k}_i^T\boldsymbol{\alpha}$ for some $n \times 1$ vector $\boldsymbol{\alpha}$, where $\boldsymbol{k}_i$ is the i^{th} column of the $n \times n$ kernel matrix $\boldsymbol{K}$ with elements $K(\boldsymbol{x}_k, \boldsymbol{x}_l),\ k, l = 1, \ldots, n$.

Now the penalized negative log-likelihood (2.4) becomes

$$\ell(b_0, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \left\{ e^{b_0+\boldsymbol{k}_i^T\boldsymbol{\alpha}} - y_i(b_0 + \boldsymbol{k}_i^T\boldsymbol{\alpha}) \right\} + \frac{\lambda}{2}\boldsymbol{\alpha}^T\boldsymbol{K}\boldsymbol{\alpha}. \tag{2.5}$$

It is well known that Newton-Raphson method and the iterative reweighted least squares (IRWLS) procedure yield the maximum likelihood(ML) estimates for general nonlinear regression functions for Poisson distributed data. In this paper we use IRWLS procedure for ease of deriving the ordinary cross validation function for model selection, although it has a little bit slower convergence rate and a little bit larger possibility of convergence failure than Newton-Raphson method. Thus, to obtain the ML estimates of b_0 and $\boldsymbol{\alpha}$, we set the first-order partial derivatives of the penalized negative log-likelihood (2.5) with respect to b_0 and $\boldsymbol{\alpha}$ equal to 0 and $\boldsymbol{0}_n$, *i.e.*, $\partial\ell(b_0, \boldsymbol{\alpha})/\partial b_0 = 0$ and $\partial\ell(b_0, \boldsymbol{\alpha})/\partial\boldsymbol{\alpha} = \boldsymbol{0}_n$. Here $\boldsymbol{0}_n$ and $\boldsymbol{1}_n$ are the $n \times 1$ vectors of zeros and ones, respectively. We then solve the following IRWLS equations in matrix notation:

$$\left(\boldsymbol{V}^T\boldsymbol{W}\boldsymbol{V} + \boldsymbol{U} \right) \boldsymbol{\beta} = \boldsymbol{V}^T\boldsymbol{W}\boldsymbol{y}^*, \tag{2.6}$$

where $\boldsymbol{V} = (\boldsymbol{1}_n, \boldsymbol{K})$, $\boldsymbol{W} = \mathrm{diag}(\exp(b_0 + \boldsymbol{k}_1^T\boldsymbol{\alpha}), \ldots, \exp(b_0 + \boldsymbol{k}_n^T\boldsymbol{\alpha}))$, $\boldsymbol{\beta} = (b_0, \boldsymbol{\alpha}^T)^T$, $\boldsymbol{U}$ is defined as

$$\boldsymbol{U} = \begin{bmatrix} 0 & \boldsymbol{0}_n^T \\ \boldsymbol{0}_n & \lambda\boldsymbol{K} \end{bmatrix}$$

and $\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)^T$ is the working response vector for

$$y_i^* = \frac{y_i - \exp\left(b_0 + \boldsymbol{k}_i^T\boldsymbol{\alpha}\right)}{\exp\left(b_0 + \boldsymbol{k}_i^T\boldsymbol{\alpha}\right)} + \left(b_0 + \boldsymbol{k}_i^T\boldsymbol{\alpha}\right).$$

With the optimal values of b_0 and $\boldsymbol{\alpha}$, the predicted mean function given the input vector $\boldsymbol{x}_0$ is obtained as $\mu(\boldsymbol{x}_0) = e^{b_0+\boldsymbol{k}_0^T\boldsymbol{\alpha}}$, where $\boldsymbol{k}_0 = (K(\boldsymbol{x}_1, \boldsymbol{x}_0), \ldots, K(\boldsymbol{x}_n, \boldsymbol{x}_0))^T$.

3. Semiparametric Mixed-Effect Model

We now consider a semiparametric mixed-effect kernel Poisson regression model for the analysis of longitudinal count data. Let y_{ij} be the j^{th} response variable of the i^{th} subject corresponding to covariate vector $\boldsymbol{x}_{ij}$, where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n_i$. Assuming the Poisson process for y_{ij}, we have a Poisson distribution given the input vector $\boldsymbol{x}_{ij} \in R^p$,

$$p(y_{ij}) = \frac{e^{-\mu(\boldsymbol{x}_{ij})}\mu(\boldsymbol{x}_{ij})^{y_{ij}}}{y_{ij}!}, \quad i = 1, 2, \ldots, n. \tag{3.1}$$

We write the canonical parameter(logarithm of $\mu(\boldsymbol{x}_{ij})$) as $\eta(\boldsymbol{x}_{ij})$. For brevity, we write $\mu(\boldsymbol{x}_{ij})$ and $\eta(\boldsymbol{x}_{ij})$ as μ_{ij} and η_{ij}, respectively. Let $\boldsymbol{x}_{ij} = (\boldsymbol{x}_{1ij}^T, \boldsymbol{x}_{2ij}^T)^T$ be the associated covariate vector with (p_1, p_2) components such that $p = p_1 + p_2$. Let η_{ij} be the regression function given $\boldsymbol{x}_{ij}$ and assume that the η_{ij} is related to covariate vector $\boldsymbol{x}_{ij}$ in a semiparametric form as

$$\eta_{ij} = b_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_{1ij} + \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_{2ij}) + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}, \quad \text{for } i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, n_i, \tag{3.2}$$

where b_0 is the bias, $\boldsymbol{\beta}_1$ is $p_1 \times 1$ regression parameter vector, $\boldsymbol{\phi}(\boldsymbol{x}_{2ij})$ is a nonlinear feature mapping function, $\boldsymbol{z}_{ij}$ is $q \times 1$ random effect covariate vector and $\boldsymbol{b}_i$ is $q \times 1$ random effect parameter vector from $N_q(\boldsymbol{0}, \boldsymbol{B})$. $\boldsymbol{B}$ is generally unknown yet we are not particularly concerned with its estimation. For semiparametric model, we assume the covariates $\boldsymbol{x}_{1ij}$ in the parametric part of the regression function have a linear effect on η_{ij} and the effect of covariates $\boldsymbol{x}_{2ij}$ in the nonparametric part on η_{ij} is not specified.

The penalized negative log-likelihood becomes

$$\begin{aligned}\ell(b_0, \boldsymbol{\beta}_1, \boldsymbol{w}, \boldsymbol{b}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \Big[-y_{ij} \left\{ b_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_{1ij} + \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_{2ij}) + \boldsymbol{b}_i^T \boldsymbol{z}_{ij} \right\} \\ + e^{b_0 + \boldsymbol{\beta}_1^T \boldsymbol{x}_{1ij} + \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_{2ij}) + \boldsymbol{b}_i^T \boldsymbol{z}_{ij}} \Big] + \frac{\lambda_1}{2}\|\boldsymbol{w}\|^2 + \frac{\lambda_2}{2}\boldsymbol{b}^T \tilde{\boldsymbol{B}}^{-1}\boldsymbol{b},\end{aligned} \tag{3.3}$$

where $\boldsymbol{b} = (\boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_N^T)^T$ and $\tilde{\boldsymbol{B}} = \text{diag}(\boldsymbol{B}, \ldots, \boldsymbol{B})$. Write $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_N^T)^T$, $N_n = \sum_{k=1}^{N} n_k$ and $\tilde{\boldsymbol{Z}} = \text{diag}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N)$ with $\boldsymbol{Z}_i = [\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{in_i}]^T$. Let $\boldsymbol{\beta}$ represent the vector of parameters defined as $\boldsymbol{\beta} = (b_0, \boldsymbol{\beta}_1^T, \boldsymbol{\alpha}^T, \boldsymbol{b}^T)^T$ with $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{in_i})^T$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_N^T)^T$. Let $\boldsymbol{X}_1$ represents the $N_n \times p_1$ matrix obtained by stacking in order $\boldsymbol{x}_{1ij}^T$'s. Let $\boldsymbol{K}$ be the $N_n \times N_n$ kernel matrix consisting of $K(\boldsymbol{x}_{2ik}, \boldsymbol{x}_{2il})$, $i = 1, \ldots, N$, $k, l = 1, \ldots, n_i$. Then, using kernel tricks in Vapnik (1995), we can rewrite (3.3) in matrix notation as:

$$\ell(\boldsymbol{\beta}) = -\boldsymbol{y}^T \boldsymbol{V}\boldsymbol{\beta} + \boldsymbol{1}_{N_n}^T \exp(\boldsymbol{V}\boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{U}\boldsymbol{\beta}, \tag{3.4}$$

where $\boldsymbol{V} = [\boldsymbol{1}_{N_n}, \boldsymbol{X}_1, \boldsymbol{K}, \tilde{\boldsymbol{Z}}]$.

And $\boldsymbol{U}$ is defined as

$$\boldsymbol{U}=\begin{bmatrix} 0 & \mathbf{0}_{p_1}^T & \mathbf{0}_{N_n}^T & \mathbf{0}_{Nq}^T \\ \mathbf{0}_{p_1} & \boldsymbol{O}_{p_1\times p_1} & \boldsymbol{O}_{p_1\times N_n} & \boldsymbol{O}_{p_1\times Nq} \\ \mathbf{0}_{N_n} & \boldsymbol{O}_{N_n\times p_1} & \lambda_1\boldsymbol{K} & \boldsymbol{O}_{N_n\times Nq} \\ \mathbf{0}_{Nq} & \boldsymbol{O}_{Nq\times p_1} & \boldsymbol{O}_{Nq\times N_n} & \lambda_2\tilde{\boldsymbol{B}}^{-1} \end{bmatrix}, \tag{3.5}$$

for the $l\times k$ zero matrix $\boldsymbol{O}_{l\times k}$.

We then solve the following IRWLS equations in matrix notation:

$$\left(\boldsymbol{V}^T\boldsymbol{W}\boldsymbol{V}+\boldsymbol{U}\right)\boldsymbol{\beta}=\boldsymbol{V}^T\boldsymbol{W}\boldsymbol{y}^*, \tag{3.6}$$

where $\boldsymbol{W}$ is a diagonal matrix consisting of elements of $\exp(\boldsymbol{V}\boldsymbol{\beta})$ and $\boldsymbol{y}^*=\boldsymbol{W}^{-1}(\boldsymbol{y}-\exp(\boldsymbol{V}\boldsymbol{\beta}))+\boldsymbol{V}\boldsymbol{\beta}$ is the working response vector.

4. Model Selection

The functional structures of the semiparametric mixed-effect kernel Poisson regression is characterized by hyperparameters, the penalty parameters λ_1,λ_2 and the kernel parameter σ^2. For convenience, we rearrange y_{ij}'s using single index and then denote each response by y_k, $k=1,\ldots,N_n$. To determine hyperparameters, we define the leave-one-out cross-validation(CV) function for a set of hyperparameters $\boldsymbol{\theta}$ as follows:

$$\mathrm{CV}(\boldsymbol{\theta})=\frac{1}{N_n}\sum_{k=1}^{N_n}\left\{y_k-\hat{\mu}_\theta^{(-k)}(\boldsymbol{x}_k)\right\}^2, \tag{4.1}$$

where $\boldsymbol{x}_k$ is the covariate vector corresponding to y_k and $\hat{\mu}_\theta^{(-k)}(\boldsymbol{x}_k)$ is the estimate of $\mu_\theta(\boldsymbol{x}_k)$ from data without the k^{th} observation. Since for each candidate of hyperparameter sets, N_n of $\hat{\eta}_\theta^{(-k)}(\boldsymbol{x}_k)$'s should be computed, selecting parameters using CV function is computationally formidable.

By leaving-out-one lemma (Craven and Wahba, 1979), we have

$$\left\{y_k-\hat{\mu}_\theta^{(-k)}(\boldsymbol{x}_k)\right\}-\left\{y_k-\hat{\mu}_\theta(\boldsymbol{x}_k)\right\}=\hat{\mu}_\theta(\boldsymbol{x}_k)-\hat{\mu}_\theta^{(-k)}(\boldsymbol{x}_k)\approx\frac{\partial\hat{\mu}_\theta(\boldsymbol{x}_k)}{\partial y_k}\left\{y_k-\hat{\mu}_\theta^{(-k)}(\boldsymbol{x}_k)\right\}$$

and

$$\frac{\partial\hat{\mu}_\theta(\boldsymbol{x}_k)}{\partial y_k}=\frac{\partial\hat{\mu}_\theta(\boldsymbol{x}_k)}{\partial\hat{\eta}_\theta(\boldsymbol{x}_k)}\frac{\partial\hat{\eta}_\theta(\boldsymbol{x}_k)}{\partial y_k^*}\frac{\partial y_k^*}{\partial y_k}=s_{kk}, \tag{4.2}$$

where s_{kk} is the k^{th} diagonal element of matrix $\boldsymbol{S}\equiv\boldsymbol{V}(\boldsymbol{V}^T\boldsymbol{W}\boldsymbol{V}+\boldsymbol{U})^{-1}\boldsymbol{V}^T\boldsymbol{W}$. Then the ordinary cross-validation(OCV) function can be obtained as follows:

$$\mathrm{OCV}(\boldsymbol{\theta})=\frac{1}{N_n}\sum_{k=1}^{N_n}\left(\frac{y_k-e^{\hat{\eta}_\theta(\boldsymbol{x}_k)}}{1-s_{kk}}\right)^2. \tag{4.3}$$

Replacing s_{kk} by their average $1/N_n \text{tr}(\boldsymbol{S})$, the generalized cross-validation(GCV) function can be obtained as

$$\text{GCV}(\boldsymbol{\theta}) = \frac{N_n \sum_{k=1}^{N_n} \left(y_k - e^{\hat{\eta}_\theta(\boldsymbol{x}_k)} \right)^2}{\{N_n - \text{tr}(\boldsymbol{S})\}^2}. \tag{4.4}$$

5. Numerical Studies

In this section we proceed simulation study to investigate the finite sample behaviors of the proposed semiparametric mixed-effect kernel Poisson regression model for count data. We also use the epileptic seizure data (Thall and Vail, 1990) to evaluate the proposed model.

5.1. Simulation study

A Monte Carlo simulation study is conducted to assess the performance of the proposed model. Let y_{ij} be the j^{th} response variable of the i^{th} subject corresponding to covariate t_{ij}, where $i = 1, 2, \ldots, 10$ and $j = 1, \ldots, 7$. Assuming the Poisson process for y_{ij}, we have a Poisson distribution with mean μ_{ij} given the input variable t_{ij}. We shall fit an Poisson regression model of the form

$$\eta_{ij} \equiv \log \mu_{ij} = 1 + 2\sin(\pi t_{ij}) + b_i, \tag{5.1}$$

where t_{ij}'s are generated from a uniform distribution $U(0,1)$ and b_i's are generated from a normal distribution $N(0,\, 0.01)$.

The Gaussian kernel is utilized to estimate the nonparametric component of regression function in this study. The penalty parameters $\lambda_1 = 1$, $\lambda_2 = 100$ and the kernel parameter $\sigma^2 = 0.3$ are obtained by GCV function (4.4). Figure 5.1 shows the scatter plots of data points and the results of the true and fitted regression functions for 4 subjects randomly selected. The data points are denoted by "$\cdot$". The results of the true and fitted regression functions are denoted by the solid and dotted lines, respectively. In Figure 5.1 we can see that the proposed model works well for this simulated data since the results of both regression functions are close. Figure 5.2 shows the histogram of the random effect b_i. The left plot is the histogram for the true b_i and the right plot is the histogram for the estimated b_i. However, these histograms for such small data set do not show quite well that b_i's are normally distributed.

5.2. Treatment of epileptic seizure

Patients suffering from simple or complex partial seizures were randomized to receive either the antiepileptic drug progabide or a placebo, as an adjuvant to standard chemotherapy. The patients were followed up for eight weeks in four biweekly clinic visits and the biweekly seizure counts were collected. Also collected were the baseline seizure

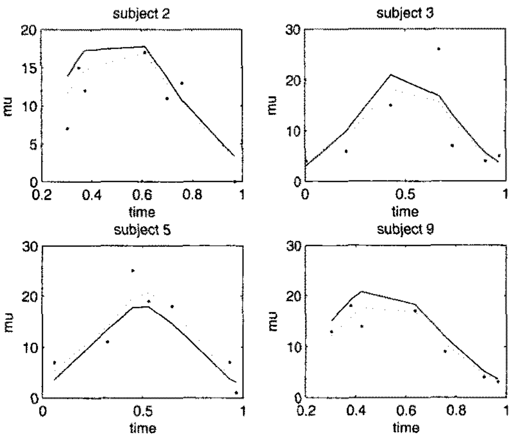

Figure 5.1: Results of regression functions for subjects in simulation study(Observation (dot), true regression function(solid line) and fitted regression function(dotted line))

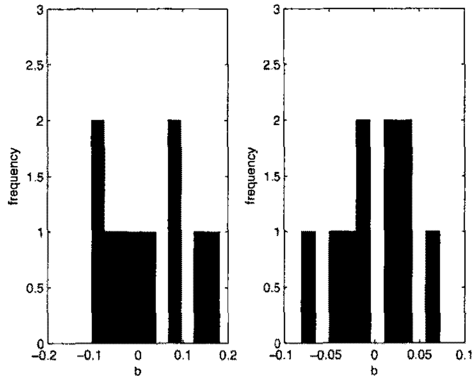

Figure 5.2: Histograms of random effect parameter in simulation study(Left: Histogram of the true b_i, Right: Histogram of the estimated b_i)

counts over the eight weeks prior to the trial and the age of the patients. A total of 59 patients were represented in the data, with 31 receiving progabide and 28 receiving placebo. The data are listed in Thall and Vail (1990), where further details can be found. Gu and Ma (2005) analyzed this data with nonparametric mixed-effect model based on smoothing spline technique.

Let μ_{ij} be the j^{th} seizure intensity of the i^{th} subject corresponding to covariate vector $\boldsymbol{x}_{ij}$, where $\boldsymbol{x}_{ij} = (x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij})^T$ consists of the treatment(2 levels), the time of clinic visit(4 points), the baseline seizure count and the age of patient, in order. We shall fit an Poisson model of the form

$$\eta_{ij} \equiv \log \mu_{ij} = b_0 + \beta_1 x_{1ij} + \boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_{2ij}) + b_i, \tag{5.2}$$

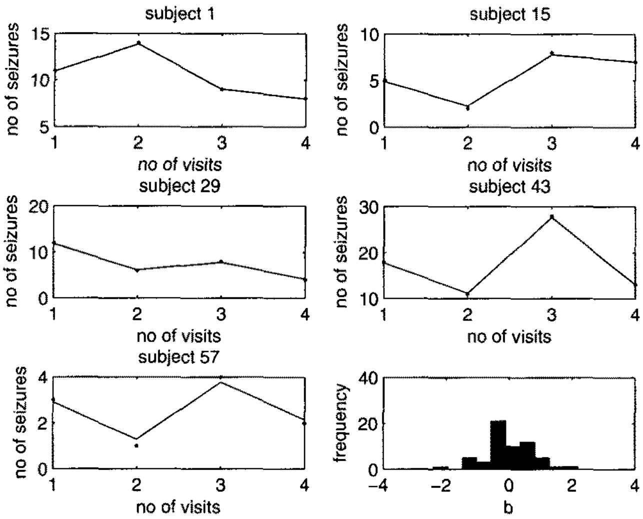

Figure 5.3: Results for treatment of epileptic seizure

where $\boldsymbol{x}_{2ij} = (x_{2ij}, x_{3ij}, x_{4ij})^T$ and $i = 1, \ldots, 59$, $j = 1, \ldots, 4$.

The Gaussian kernel is utilized to estimate the nonparametric component of regression function in this example. The penalty parameters $\lambda_1 = 0.2$, $\lambda_2 = 0.2$ and the kernel parameter $\sigma^2 = 1.1$ are obtained by GCV function (4.4). The intercept estimate is $\hat{b}_0 = 1.6645$ and the slope estimate is $\hat{\beta}_1 = -0.38685$, *i.e.*, the treatment has the negative effect on seizure. Figure 5.3 shows the seizure counts for 5 subjects randomly selected and the histogram of $\hat{b}_i$. As seen from Figure 5.3, the proposed model provides a good fitting performance for this longitudinal data set, which is based on the fact that the proposed model includes not only the penalized term but also subject effects(random effects) in likelihood function.

6. Conclusions

This paper proposes a kernel machine approach to the estimation of semiparametric mixed-effect model in Poisson regression to analyze longitudinal count data. The main advantage of the proposed approach is that its solution is easily obtained from a simple IRWLS technique. This makes it easier to apply the proposal to the analysis of longitudinal count data in practice. The proposal can be applied without heavy computations to high-dimensional covariates settings since it takes after all advantages of kernel machine. An important issue for kernel machine is model selection. Thus, we provide a GCV method for choosing the hyperparameters which affect the performance of the proposed approach.

The numerical studies indicate that the proposal produces good estimates for the parametric component of fixed-effect term and for the nonparametric one and that it works well in general with finite sample. Therefore, we recognize that the proposed method using the idea of kernel machine provides a satisfying solution to Poisson regression model for analysis of longitudinal count data.

References

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation, *Numerische Mathematics,* **31**, 377–403.

Gu, C. and Ma, P. (2005). Generalized nonparametric mixed-effect models: Computation and smoothing parameter selection, *Journal of Computational and Graphical Statistics,* **14**, 485–504.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis,* John Wiley & Sons, New York.

Kimeldorf, G. S. and Wahba, G. (1971). Some results on Tchebycheffian spline functions, *Journal of Mathematical Analysis and its Applications,* **33**, 82–95.

Long, J. S. (1997). Regression models for categorical and limited dependent variables, *Advanced Quantitative Techniques in the Social Sciences,* **7**, Sage Publications.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models (Monographs on Statistics and Applied Probability),* Chapman & Hall/CRC, London.

Shim, J., Hong, D. H., Kim, D. H. and Hwang, C. (2007). Multinomial kernel logistic regression via bound optimization approach, *Communications of the Korean Statistical Society,* **14**, 507–516.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion, *Biometrics,* **46**, 657–671.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory,* Springer, New York.

Winkelmann, R. (2003). *Econometric Analysis of Count Data,* Springer Verlag, Berlin.

Wu, H. and Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches,* John Wiley & Sons, New York.

Yuan, M. (2005). Automatic smoothing for Poisson regression, *Communications in Statistics - Theory and Methods,* **34**, 603–617.

[Received October 2008, Accepted October 2008]