ARTICLE IN PRESS

Journal of Statistical Planning and Inference xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Nonparametric Poisson regression from independent and weakly dependent observations by model selection

Martin Kroll [1]

Universität Mannheim, Germany

ARTICLE INFO

Article history:
Received 8 August 2017
Received in revised form 9 May 2018
Accepted 9 July 2018
Available online xxxx

MSC:
62G05
62G08

Keywords:
Poisson regression
Nonparametric estimation
Projection estimator
Adaptive estimation
Model selection

ABSTRACT

We consider the non-parametric Poisson regression problem where the integer valued response Y is the realization of a Poisson random variable with parameter $\lambda(X)$. The aim is to estimate the functional parameter λ from independent or weakly dependent observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ in a random design framework.

First we determine upper risk bounds for projection estimators on finite dimensional subspaces under mild conditions. In the case of Sobolev ellipsoids the obtained rates of convergence turn out to be optimal.

The main part of the paper is devoted to the construction of adaptive projection estimators of λ via model selection. We proceed in two steps: first, we assume that an upper bound for $\|\lambda\|_\infty$ is known. Under this assumption, we construct an adaptive estimator whose dimension parameter is defined as the minimizer of a penalized contrast criterion. Second, we replace the known upper bound on $\|\lambda\|_\infty$ by an appropriate plug-in estimator of $\|\lambda\|_\infty$. The resulting adaptive estimator is shown to attain the minimax optimal rate up to an additional logarithmic factor both in the independent and the weakly dependent setup. Appropriate concentration inequalities for Poisson point processes turn out to be an important ingredient of the proofs.

We illustrate our theoretical findings by a short simulation study and conclude by indicating directions of future research.

1. Introduction

We consider the non-parametric estimation of a regression function

$$\lambda : \mathbb{X} \to [0, \infty)$$

defined on some Polish space $\mathbb{X}$ from observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ where, conditional on $X_1, \ldots, X_n$, the Y_i are independent and Poisson distributed with parameter $\lambda(X_i)$. The covariates $X_1, \ldots, X_n$ are drawn from some strictly stationary process $(X_i)_{i\in\mathbb{Z}}$, and we will consider the two cases where either (i) the $X_1, \ldots, X_n$ are independent, or (ii) some adequate condition on the dependence of the underlying process $(X_i)_{i\in\mathbb{Z}}$ is satisfied. Although we will also provide minimax theoretical results for the non-parametric estimation problem at hand, our focus will be on the adaptive estimation of λ, that is, the construction of estimators that depend only on the observations but not on any structural presumptions concerning the regression function.

E-mail address: martin.kroll@ensae.fr.

[1] Present address: École Nationale de la Statistique et de l'Administration Économique (ENSAE), 5 Avenue Henry Le Chatelier, F-91120 Palaiseau.

https://doi.org/10.1016/j.jspi.2018.07.003
0378-3758/

Regression models with count data, i.e., non-negative and integer-valued response, are of interest in a wide range of applications, for instance in economics (Winkelmann, 2008), quantitative criminology (Berk and MacDonald, 2008), and ecology (Ver Hoef and Boveng, 2007). The *Poisson regression* model introduced above is the most natural example of such a count data regression model. Other models with count data response include models based on the negative binomial distribution which can also deal with *overdispersion*. Such more advanced models will not be considered in this paper. Most of the work in the area of count data regression has been devoted to parametric models, see for instance the monograph (Cameron and Trivedi, 1998) for a comprehensive overview of methods. Let us just mention some examples: the paper (Diggle et al., 1998) gives an application of a Poisson regression model in a geostatistical context. It provides a fully parametric approach and suggests MCMC techniques for fitting a model to the given data. The paper (Carota and Parmigiani, 2002) introduces a semi-parametric Bayesian model for count data regression and applies it as a prognostic model for early breast cancer data. The article (Nakaya et al., 2005) considers geographically weighted Poisson regression for disease association mapping.

Despite its potential utility in many applications, non-parametric Poisson regression has hardly been studied from a theoretical point so far. One possible approach is to apply the so-called Anscombe transform (Anscombe, 1948) to the data and treat the data as if they were Gaussian. Another approach would be to consider the generalized linear model representation of Poisson regression and allow for varying coefficients (Hastie and Tibshirani, 1993; Fan and Zhang, 1999). Recent work has also considered the Poisson regression model in a high-dimensional framework using the LASSO and the group LASSO (Ivanoff et al., 2016). Another interesting reference is Fryzlewicz (2008): in this paper the author considers a very general model with Poisson regression as a special case. In contrast to our model, only regression with deterministic design is considered (note that the distinction between independent and weakly dependent covariates considered by us is not possible in the model with deterministic design). Moreover, the automatic choice of the smoothing parameter is not addressed from a theoretical point of view in Fryzlewicz (2008) whereas this is the major topic of our contribution. Finally, let us mention the work (Besbeas et al., 2004) where an extensive simulation study for count data regression using wavelet methods was performed. That paper contains also further references to Bayesian methods in the context of count data regression. In this paper, we study adaptive non-parametric Poisson regression via the model selection approach. To the best of our knowledge, this approach has not been used for non-parametric Poisson regression so far (in a parametric framework, however, the recent paper (Kamo et al., 2013) considers a model selection approach via a bias-corrected AIC criterion).

Note that a characteristic feature of the non-parametric Poisson regression model is the fact that it naturally incorporates heteroscedastic noise. Besides work on regression in presence of homoscedastic errors (see for instance Baraud, 2000), there already exists research that considers model selection techniques in regression frameworks containing heteroscedasticity (Saumard, 2013). However, in Saumard (2013) the observations are of the form

$$Y = r(X) + \sigma(X)\epsilon$$

where r is the unknown regression function to be estimated, the residuals ϵ have zero mean and variance one, and the function σ models the unknown heteroscedastic noise level. Note that this model does not contain the Poisson regression model to be considered in this paper as a special case.

Our paper is also more general with regard to another aspect: we do not exclusively stick to the case that the covariates X_i are independent but also consider the more general case that the covariates are weakly dependent which seems to be more realistic at least in some real world scenarios. For instance, when studying clutch sizes of bird eggs that are modelled via count data models (that usually go beyond the Poisson model studied here due to over-dispersion of the data) in ornithology (Ridout and Besbeas, 2004), the covariates (e.g., temperature) are not independent when data are collected over a period of time. Concerning mathematical methodology, we will model this by imposing throughout conditions on the decay of the so-called β-mixing coefficients. The class of time series with β-mixing coefficients is sufficiently large to be of interest for applications and includes stationary vector ARMA processes (Mokkadem, 1990) or even more general autoregressive processes of the form $X_t = m(X_{t-1}) + \sigma(X_{t-1})\epsilon_t$ under mild conditions on the functions m and σ (Neumann and Thorarinsdottir, 2006; Doukhan, 1994). Our methodological approach is mainly based on fundamental results from the article (Viennet, 1997) that have also been exploited in a wide variety of other statistical problems: in Baraud et al. (2001) and Asin and Johannes (2016) the authors consider the non-parametric estimation of a regression function in case of β-mixing covariates. The paper (Lacour, 2008) considers adaptive estimation of the transition density of a particular hidden Markov chain under the assumption that the hidden chain is β-mixing. From a methodological point of view our approach was also inspired by the recent work (Asin and Johannes, 2016). However, in contrast to that paper, we build our construction of adaptive estimators on the model selection technique from Barron et al. (1999) only, whereas (Asin and Johannes, 2016) combines the model selection approach with a more recent technique due to Goldenshluger and Lepski (2011).

Let us sketch the organization and summarize the main contributions of the paper. In Section 2 we introduce notations and the general methodology used in the paper. Section 3 is devoted to the case of independent observations: we derive a general minimax upper bound and a matching lower bound over Sobolev ellipsoids. We then consider adaptive estimation of the regression function via model selection which has not been addressed before in the literature. We first consider an estimator based on the a priori knowledge of an upper bound on the regression function (Section 3.3.1), and then, inspired by the approach in Comte (2001), put some effort to develop an estimator that does not depend on this assumption (Section 3.3.2). The risk bound of the adaptive estimators is deteriorated by a logarithmic factor only in comparison with the minimax

Please cite this article in press as: Kroll M., Nonparametric Poisson regression from independent and weakly dependent observations by model selection. J. Statist. Plann. Inference (2018), https://doi.org/10.1016/j.jspi.2018.07.003.